





Cite this: *Org. Biomol. Chem.*, 2021, **19**, 10424

The role of streptavidin and its variants in catalysis by biotinylated secondary amines†

Alexander R. Nödling, ^a Nicolò Santi, ^a Raquel Castillo,^b Magdalena Lipka-Lloyd,^a Yi Jin, ^a Louis C. Morrill, ^c Katarzyna Świderek, ^{*b} Vicent Moliner ^{*b} and Louis Y. P. Luk ^{*a,c}

Here, we combine the use of host screening, protein crystallography and QM/MM molecular dynamics simulations to investigate how the protein structure affects iminium catalysis by biotinylated secondary amines in a model 1,4 conjugate addition reaction. Monomeric streptavidin (M-Sav) lacks a quaternary structure and the solvent-exposed reaction site resulted in poor product conversion in the model reaction with low enantio- and regioselectivities. These parameters were much improved when the tetrameric host T-Sav was used; indeed, residues at the symmetrical subunit interface were proven to be critical for catalysis through a mutagenesis study. The use of QM/MM simulations and the asymmetric dimeric variant D-Sav revealed that both Lys121 residues which are located in the hosting and neighboring subunits play a critical role in controlling the stereoselectivity and reactivity. Lastly, the D-Sav template, though providing a lower conversion than that of the symmetric tetrameric counterpart, is likely a better starting point for future protein engineering because each surrounding residue within the asymmetric scaffold can be refined for secondary amine catalysis.

Received 4th October 2021,
Accepted 14th November 2021

DOI: 10.1039/d1ob01947c

rsc.li/obc

Introduction

Organocatalysts are powerful auxiliaries used in the design of novel artificial enzymes.^{1–3} Various engineering approaches have been applied to facilitate organocatalysis within a protein scaffold including side-chain labeling,^{4–8} genetic code expansion,^{2,9–14} use of N-terminal prolines,^{3,15,16} computational design^{17–20} and the protein–ligand approach.^{1,21–24} To screen a large number of catalysts within a short timeframe, the protein–ligand approach is most frequently used.^{1,21–24} In this approach, ligands are covalently linked with chemical catalysts and then introduced to their binding proteins for activity test.^{25–31} One important consideration is to choose an appropriate protein partner because there are multiple homologues that bind the same ligand with high affinity, but their effect on chemical catalysis can be vastly different.^{32–38} Accordingly, detailed mechanistic studies must be conducted during the design of artificial enzymes.

The importance of choosing an appropriate host has been demonstrated in the use of biotin-binding proteins for artificial enzyme design. Tetrameric streptavidin (T-Sav) has been popularly used, engineered and evolved for hosting reactions that are mediated by biotin-appended chemical catalysts.^{25–29,32,33,39–43} In each T-Sav, there are four biotin-binding sites whose arrangement locates the catalyzed reactions at the interface of two subunits which has a C_2 symmetry (Fig. 1A).^{21,23,36,40,42,44–47} Consequently, changing both the surrounding residues and the degree of binding site occupancy can affect the reaction reactivity and stereoselectivity.^{36,48,49} Furthermore, a switch in the biotin-binding scaffold has also shown to drastically affect the performance of the catalysis.^{36,37} In one scenario, the Ir-catalyzed hydrogen transfer reaction improved significantly when the monomeric subunits of T-Sav were dimerized and transformed into an asymmetrical scaffold (D-Sav) that hosts chemical catalysis without interference caused by (non-)cooperative binding.³⁶ In the other scenario, the Rh(III)-catalyzed lactam formation is enhanced only when T-Sav is replaced by an engineered monomeric counterpart M-Sav; this phenomenon was attributed to enhanced solvent exposure of the bulky catalyst (Fig. 1A).³⁷ These studies indicated that a change of the protein host could greatly affect the outcome of the chemical transformations mediated by metal catalysts, but equivalent studies on protein-hosted organocatalysis have not been conducted.

We have recently reported an organocatalytic 1,4 conjugate addition of nitromethane to cinnamaldehydes within a protein

^aSchool of Chemistry, Main Building, Cardiff University, Cardiff, CF10 3AT, UK. E-mail: LukLY@cardiff.ac.uk

^bDepartment de Química Física i Analítica, Universitat Jaume I, Castellón, 12071, Spain. E-mail: swiderek@uji.es, moliner@uji.es

^cCardiff Catalysis Institute, School of Chemistry, Main Building, Cardiff University, Cardiff, CF10 3AT, UK

†Electronic supplementary information (ESI) available. See DOI: 10.1039/d1ob01947c



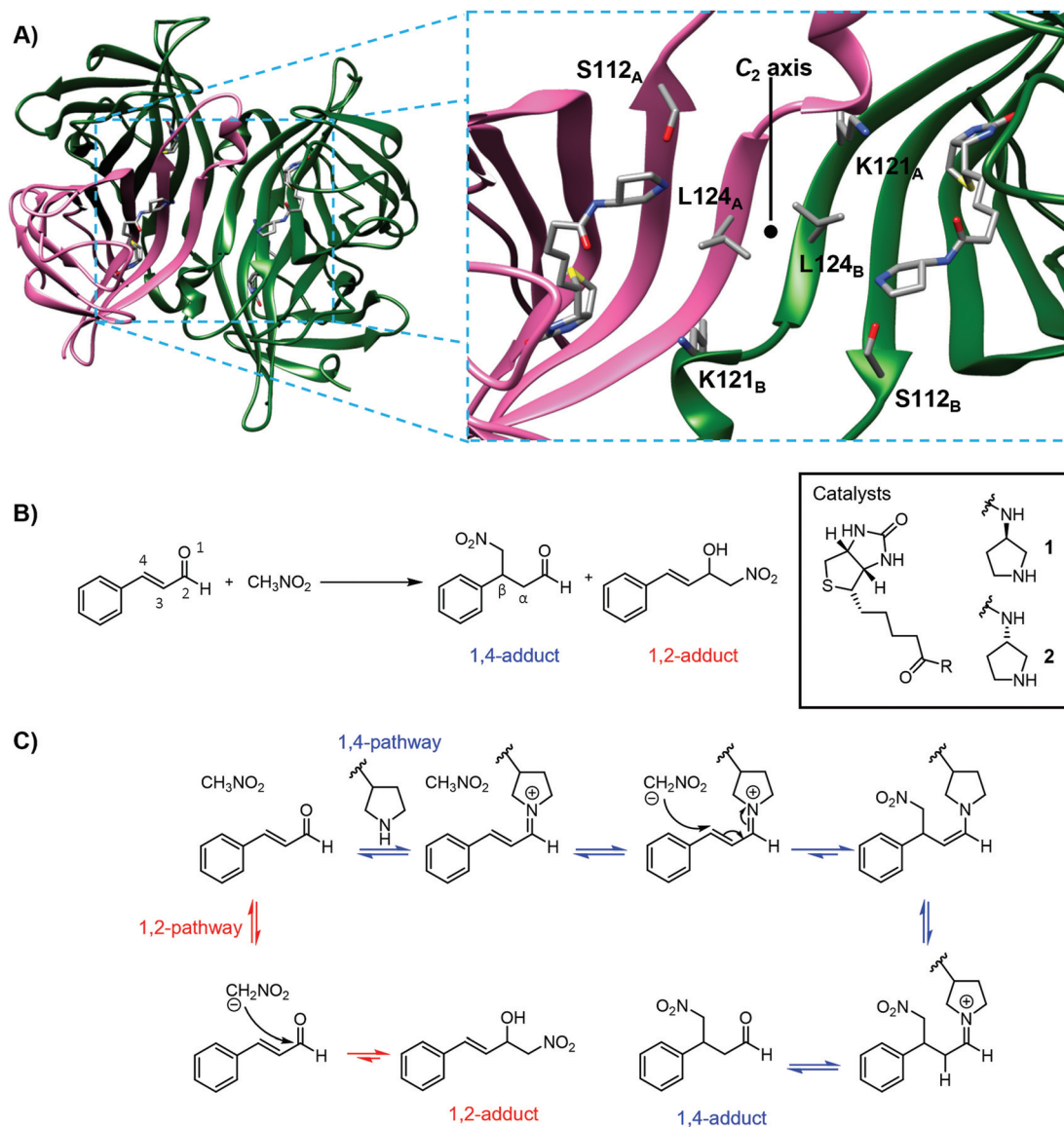


Fig. 1 (A) Left: Cartoon of tetrameric streptavidin (T-Sav) containing ligand 1 (PDB: 6GH7); the monomer unit is shown in pink. Right: Close up of the C₂-symmetric interunit interface with Ser112, Lys121, and Leu124 highlighted; two monomer units in the background are omitted for clarity. (B) Model reaction catalyzed by biotinylated catalysts 1 and 2 within the tetrameric streptavidin (T-Sav), showing the desired 1,4-addition product and 1,2-addition side-product. (C) Mechanistic scheme leading to the desired 1,4-addition and 1,2-addition side-product.

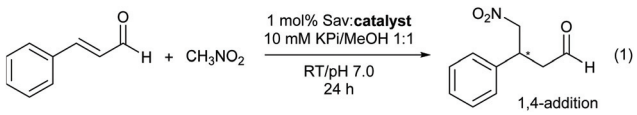
environment through adaptation of the streptavidin–biotin technology (Fig. 1B & C).²¹ In this system, biotin was covalently modified with pyrrolidine groups, resulting in ligands 1 and 2 which differ by one stereocenter (Fig. 1B). The protein T-Sav could not mediate the model reaction and catalysis by the ligands themselves was not stereoselective (Table 1, entries 10 and 11), yet the T-Sav : 1 complex favors the formation of the *S* enantiomer (91 : 9), whereas the T-Sav : 2 complex favors the formation of the *R* counterpart (24 : 76).²¹ Moreover, this protein-hosted system accepts various types of substrates for iminium^{21,50} or enamine catalysis.⁵¹ Whilst promising, the role of the T-Sav residues in proximity to the pyrrolidine has not been investigated. Importantly, whether T-Sav is the most

suitable host for iminium organocatalysis has not been verified, yet such information is fundamental to the creation of a high-performance artificial enzyme.

Here, we performed mechanistic investigations on how a protein host affects iminium catalysis. We began by testing the monomeric M-Sav as the host; the reaction lacks both enantio- and regioselectivity despite numerous optimization attempts, and protein X-ray crystallography analysis revealed that the catalyst is solvent exposed. Site-directed mutagenesis, Quantum Mechanics/Molecular Mechanics (QM/MM) analysis of T-Sav and the use of the asymmetric host D-Sav have indicated that intersubunit residues including Lys121 control both the reactivity and stereoselectivity. Even though the current



Table 1 Organocatalytic reaction (1) hosted by variants of streptavidin (Sav)



Entry	Catalyst	Host	Conversion ^a /%	Ratio 1,4- : 1,2-addition	er (S : R)
1	1	M-Sav (WT) ^b	22	1.4 : 1.0	N.D.
2	2	M-Sav (WT) ^b	27	2.9 : 1.0	
3	2	M-Sav (Y111S) ^b	36	1.3 : 1.0	
4	2	M-Sav (Y111T) ^b	33	1.5 : 1.0	
5	2	M-Sav (Y111V) ^b	28	0.6 : 1.0	
6	2	M-Sav (Y111K) ^b	46	2.1 : 1.0	
7	2	M-Sav (Y111A) ^b	33	1.5 : 1.0	
8	2	M-Sav (E120L) ^b	61	9.2 : 1.0	40 : 60
9	2	M-Sav (Y111S E120L) ^b	41	1.2 : 1.0	N.D.
10	1	—	57	8.5 : 1.0	48 : 52
11	—	T-rSav (WT) ^f	21	1.2 : 1.0	N.D.
12	1 ^c	T-Sav (WT) ^d	30	>5.0 : 1.0	80 : 20
13	1	T-Sav (WT)	80	>15.0 : 1.0	91 : 9
14	1	T-rSav (WT) ^e	87	6.3 : 1.0	90 : 10
15	1	T-rSav (WT) ^f	71	6.9 : 1.0	80 : 20
16	2 ^c	T-Sav (WT)	15	>2.0 : 1.0	24 : 76
17	2	T-Sav (WT)	71	>13.2 : 1.0	33 : 67
18	1	T-rSav S112V ^f	88	3.9 : 1.0	63 : 37
19	1	T-rSav S112E ^f	43	3.3 : 1.0	76 : 24
20	1	T-rSav S112Y ^f	77	4.9 : 1.0	53 : 47
21	1	T-rSav L124E ^f	23	0.8 : 1.0	51 : 49
22	1	T-rSav L124K ^f	80	2.2 : 1.0	53 : 47
23	1	T-rSav L124W ^f	92	>17.4 : 1.0	89 : 11
24	1	T-rSav K121A ^f	58	8.7 : 1.0	66 : 34
25	1	T-rSav K121M ^f	98	11.3 : 1.0	80 : 20
26	1	T-rSav K121R ^f	77	6.0 : 1.0	75 : 25
27	1	D-Sav SARK ^e	38	3.8 : 1.0	75 : 25
28	1	D-Sav SKRA ^e	<5	1.4 : 1.0	N.D.

^a Conversion refers to the overall conversion of cinnamaldehyde to the 1,4- and 1,2-addition products that was performed in triplicate (see Fig. S1 for exemplified conversion evaluation, Tables S1–3 for site-directed protein mutagenesis primer sequences, experimental details on Pg S17–20, Pg S36–38 & Pg S52, and chiral LC results on Pg S43–51†). ^b Additional protein purification using a Ni-affinity column (see Pg S14–16†). ^c The reaction was performed without an organic solvent (MeOH) for 42 h. Its addition most likely helped in preventing cinnamaldehyde from precipitation and affecting the hydration ratio, thus changing the substrate accessibility for reactions and the 1,4- : 1,2-adduct ratio.²¹ ^d Conversion of the reaction catalyzed by T-Sav : biotin was found to be <3%.²¹ ^e Additional protein purification using an iminobiotin column (details on T-Sav and D-Sav purification on Pg S8–13†). ^f Only purified by washing, denaturing and dialysis of the insoluble cell pellet. In line with the reports on different purity grade artificial metalloenzymes;^{47,53} slightly lower yields and enantioselectivities compared to the purified counterparts are observed (see also Pg S8–11†). T-Sav = tetrameric core streptavidin (see Pg S5† for a detailed sequence). T-rSav = tetrameric streptavidin with a “reduced” sequence.³⁵ M-Sav = monomeric streptavidin.³⁴ D-Sav = dimeric streptavidin.³⁶ N.D. = not determined.

D-Sav template provides a lower conversion than that of the symmetrical counterpart, we envisioned that an asymmetric scaffold would facilitate fine-tuning and thus it is likely a better starting point for future protein engineering work.

Results and discussion

Exploration of monomeric Sav hosts

The monomeric variant M-Sav was tested as an alternative host for iminium catalysis. Although it lacks the subunit interface and shares moderate sequence identity with T-Sav (55%), M-Sav has high affinity to biotin ($K_D = 2.8$ nM).³⁴ M-Sav was shown to be minimally active for iminium catalysis,⁵⁰ and its use as a host also resulted in poor chemical transformation in the model organocatalytic reaction (Fig. 1B and Table 1). Both the conversion and regioselectivity dropped when ligand 1 was

introduced to M-Sav; cinnamaldehyde was converted to only a small amount of the product (13% as the 1,4-addition product) with nearly the same amount of the 1,2-addition side-product formed (9%, entries 1 & 10–16; also see Fig. 1B & S1†). Catalysis by 2 in M-Sav was mildly improved (20% 1,4-addition product, entry 2) but there remained a substantial amount of the 1,2-addition side product (7%). The melting temperature of M-Sav increased significantly when ligand 1 or 2 was included (+20 °C) as assessed by circular dichroism spectroscopy (Fig. S2†), implying its binding to the biotinylated catalysts.^{37,51} Although the reaction conditions have been heavily screened (buffer composition and pH, Tables S8–S10†), conversions remained low and regioselectivity (1,4- vs. 1,2-addition) was lacking.

To understand how a switch in the protein host causes a drastic change in catalysis, the X-ray crystallography structures of M-Sav : 2 and T-Sav : 2 were obtained at a resolution of 1.80



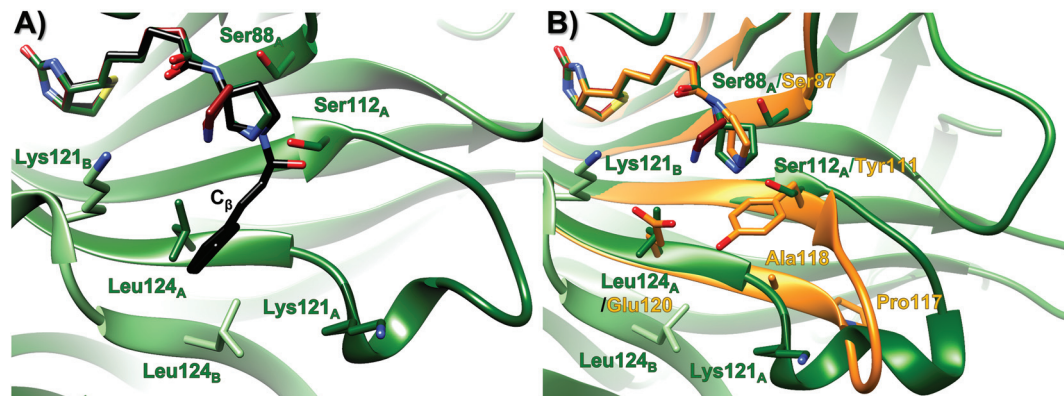


Fig. 2 Overlay of the organocatalytic system T-Sav: 1 (green, PDB 6GH7)²² with (A) ligand 2 (red, PDB 7NLV, 1.29 Å) and the tetrahedral intermediate (black, QM/MM simulation), and (B) ligand 2 (red) and M Sav: 2 (orange, PDB 6ZYT, 1.80 Å). See Tables S4–6† for crystallization conditions and details.

and 1.29 Å respectively (Fig. 2 and Tables S4–S6†), and subsequently site-directed mutagenesis analysis was conducted. Tyr111 in M-Sav: 2 is a relatively bulky residue located in proximity to the secondary amine motif and corresponds to Ser112 in T-Sav: 2. Replacing Tyr111 with other residues including serine, threonine, valine, lysine and alanine did not improve the conversion or the regioselectivity (entries 3–7). Glu120 in M-Sav: 2 is in proximity and spatially corresponds to Leu124 in T-Sav: 2. Though modifying Glu120 slightly improved the overall yield of the 1,4-addition product, the enantioselectivity of the catalysis remained low (entry 8). These observations are in contrast to a recent study which showed that M-Sav is a better host for Rh-catalyzed lactam synthesis.³⁷ One noticeable difference is that the rhodium-based catalyst is bulky and has been observed to prefer an exposed active site for reaction, whereas the organocatalytic iminium intermediate typically requires a hydrophobic environment to drive and steer the selectivity of the catalytic cycle.⁵²

To understand why T-Sav is a superior host for the model organocatalytic reaction,²¹ the ratio of biotinylated ligand 1 to their binding site was investigated. The reaction yield was found to be 46% when an average of one catalyst occupied one of the tetrameric T-Sav hosts (entry 1, Table 2). When two of the binding sites were occupied, the conversion only increased up to 61%, instead of doubling to 92% (entry 2). Eventually, the yield increased to 81% when four of the binding sites were occupied by the catalyst (entries 3 and 4). The non-linear increase in yield implied that each T-Sav binding site cannot

mediate iminium catalysis simultaneously and the intersubunit interface can only host one chemical transformation event at a time.

For further investigations, site-directed mutagenesis was conducted by the use of a truncated variant termed “reduced streptavidin” (T-rSav, see Pg S8–11† for preparation).³⁵ In previous and current crystallographic studies, Ser112 was found to be in a hydrogen-bonding distance with the reacting nitrogen atom of the pyrrolidine motif (Fig. 2A).²¹ When this residue was replaced by valine, glutamate or tyrosine, there was either a drop in conversion (71% to 43%, Table 1, entries 15 vs. 19) or enantioselectivity (80 : 20 to 53 : 47 *S* : *R*, Table 1, entries 15 vs. 20). Notably, the substrate was also converted into the 1,2-addition byproduct by these Ser112 variants, being reminiscent of the reactions hosted by M-Sav and its variants. Altogether, this suggests that Ser112 helps in positioning the intermediate for nucleophilic addition at the C_β position, improving both the regio- and enantioselectivities of the reaction.

Previous MD analysis indicated that Leu124 and Lys121 are located close to the tetrahedral intermediate.^{21,28,29} When Leu124 was changed to charged residues such as lysine and glutamate, the enantioselectivity vanished and a notable amount of the 1,2-addition side-product was observed (entries 21 and 22 in Table 1). In contrast, the conversion and stereoselectivity were mildly improved when the T-rSav-L124W variant was used as a host, indicating the importance of placing a bulky and hydrophobic residue at this position (entry 23). The Lys121 residues in both T-Sav subunits (namely SavA and SavB, see below) are also in close proximity to the reacting centers.^{21,28,29} To test their role(s), Lys121 was converted to alanine, methionine and arginine for analysis. In the reaction hosted by T-rSav-K121A, both the yield and enantioselectivity were vastly reduced (entry 24). In contrast, substituting Lys121 with methionine or arginine does not affect the performance of the reaction, implying that the size of the sidechain, rather than its electrostatic properties, is responsible for the observed selectivity (entries 25 and 26). Indeed, T-rSav-K121M provided

Table 2 Conversion % and er for the 1,4-additions at different T-Sav : 1 ratios

Entry	Equivalent 1	Conversion (%)	er (<i>S</i> : <i>R</i>)
1	1	46	87 : 13
2	2	61	82 : 17
3	3	65	87 : 13
4	4	81	87 : 13



a higher product conversion, yet the enantioselectivity remained similar to that of the wild-type host. Notably, the Lys121 variants resulted in low 1,2-byproduct formation, indicating that this residue does not improve the regioselectivity of the reaction but plays a role in guiding how the nucleophile (deprotonated nitromethane) approaches the C-4 of the iminium intermediate (Fig. 1B).

Use of dimeric Sav as the host

Single chain dimeric streptavidin (D-Sav) is a new family of engineered variants that enables independent mutagenesis of the two neighbouring streptavidin subunits, and it is used here to investigate the role of the intersubunit residues Lys121s in catalyzing the secondary amine organocatalytic reaction. According to a recent report by Wu *et al.*,³⁶ D-Sav was achieved by connecting two subunits of streptavidin through a peptide linker (26-residue), resulting in a single chain monomer that carries two equivalents of biotin-binding modules, SavA and SavB (Fig. 3A and B). An intermolecular di-

sulfide bond formed through an H127C mutation at SavB covalently connects two of the single chain monomers together, yielding the dimeric protein variant. D-Sav enables asymmetric genetic tuning of the biotin-binding vestibule (Fig. 3B), because SavB possesses additional mutations (N23A/S27D/D128A), prohibiting key H-bonding interactions needed for ligand binding. Consequently, each monomer is monovalent with the ligand binding only to the SavA module and avoids (non-)cooperative multiple site occupancy.

For this work, two asymmetric D-Sav variants were tested: SARK and SKRA, where the first two letters code for residues 112 and 121 in SavA and the last two for 112 and 121 in SavB (Fig. 2). Only D-Sav-SARK provides a measurable conversion with significant enantioselectivity, whereas D-Sav-SKRA could not provide a detectable yield (entries 27 and 28). Accordingly, this suggests that Lys121 in the neighbouring unit (*i.e.* Lys121B in SavB) plays a role in improving the reactivity and stereoselectivity; indeed, the QM/MM MD simulations carried out in the present study also suggest that Lys121B in the neigh-

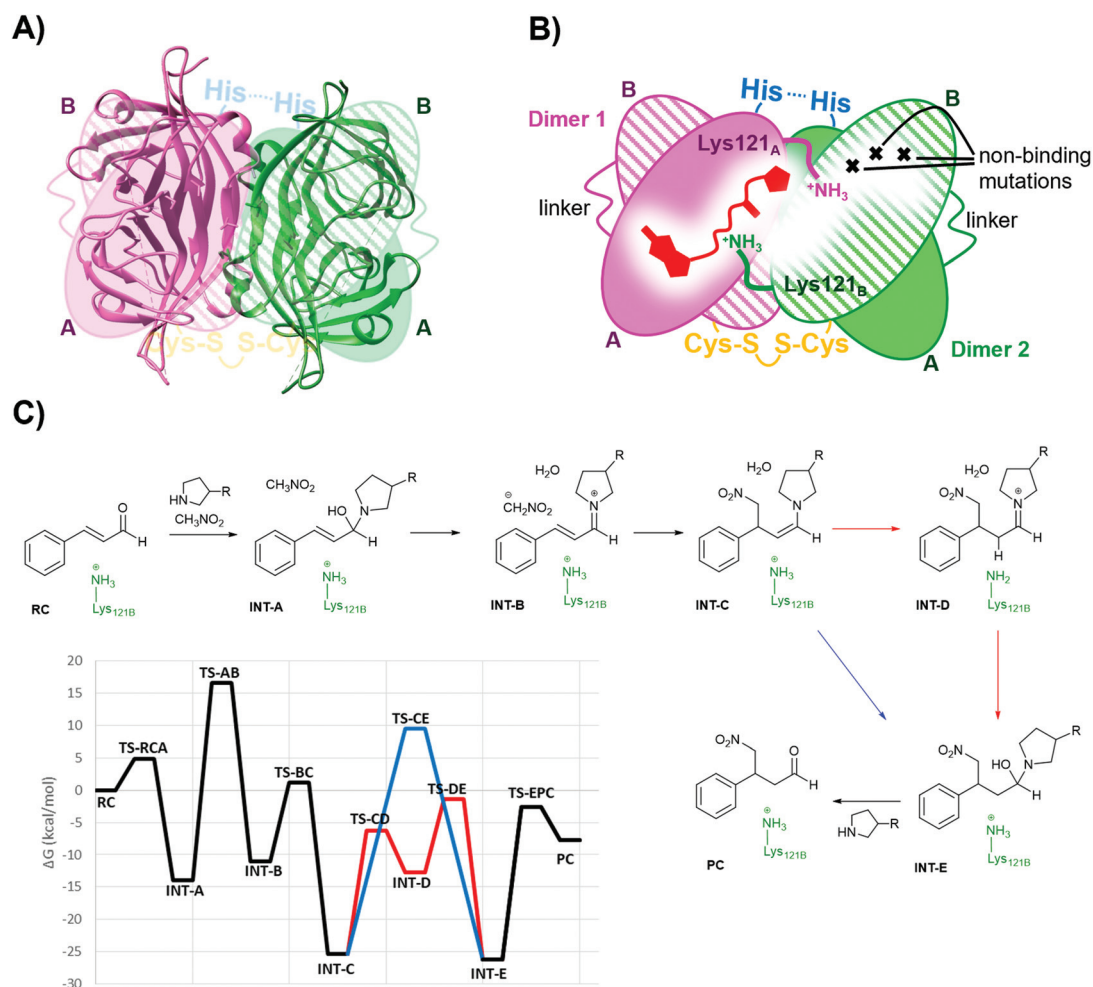


Fig. 3 (A) Overlay of the D-Sav cartoon structure (PDB: 6S50) and a simplified schematic representation. (B) Schematic representation of one inter-subunit interface in D-Sav between dimer A and dimer B including the monovalent binding of **1**. (C) Schematic of the crucial steps in the reaction hosted by the T-Sav derived from the QM/MM free energy surfaces, and the M06-2X:AM1/MM free energy profile derived from the exploration of the free energy surfaces. Steps C to E are assisted by Lys121B (in red) or by a water molecule (in blue).



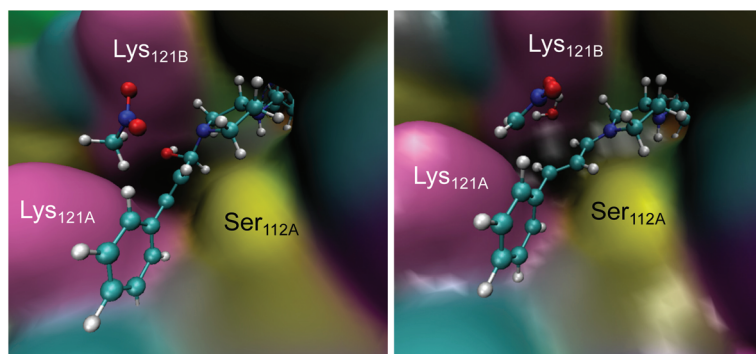


Fig. 4 Representative snapshots of the INT-A (left) and INT-B (right) intermediates derived from the QM/MM MD simulations.

bouring subunit plays a role in guiding how the nucleophile (deprotonated nitromethane) approaches the iminium intermediate (see below). Nevertheless, we noticed that the conversion hosted by D-Sav-SARK is noticeably lower than that by the wild type T-Sav with substantial conversion of the 1,2-byproduct (entry 27). Overlaying the crystal structures of D-Sav and T-Sav indicated that the relative distance between the two binding sites differed moderately (Fig. S3[†]). This structural change is likely a consequence of protein engineering and may contribute to the decrease in reaction yield, and thus further investigations *via* an alternative QM/MM approach were conducted (see below).

Molecular dynamics simulations

Derived from the free energy surfaces explored for every single chemical step at the M06-2X:AM1/MM level (Fig. 3C, S4 & S5 and Page S26–34[†]), the reaction mechanism and the corresponding free energy profile are outlined in Fig. 3C (see the ESI[†]). Our computational studies have generated a model where the geometry of the tetrahedral intermediate INT-A overlays well with the X-ray crystallographic structures. The reaction can take place in five to six steps depending on whether the protein host is included in the chemical transformation. In agreement with the previous work,⁵⁴ the transformation from INT-A to INT-B is rate-limiting requiring the nitromethane to be deprotonated and the iminium ion to be formed before the nucleophilic addition (INT-B to INT-C). Any attempt to look for alternative timing of the events did not render plausible results. The energy required for reaching TS-AB from RC was calculated to be 16.6 kcal mol⁻¹, the highest among all steps while presenting an intrinsic activation free energy barrier of 30.5 kcal mol⁻¹. Hence, this suggests that all the other chemical steps are not or only partially rate-limiting. Upon forming the enamine intermediate INT-C, the reaction can proceed through two paths: a concerted step assisted by a water molecule (and/or a phosphate group) or a step-wise process assisted by Lys121B located in the neighbouring subunit (*i.e.* SavB) (steps in blue and red in Fig. 3C, respectively). Of note, the Lys121B-assisted pathway resulted in lower activation energy barriers.

A dynamical model within a fully solvated protein was constructed to reveal the role of water molecules and residues at

the subunit interface in catalysis (Fig. 4). The averaged structures of INT-A and INT-B obtained in the reaction path indicated that Ser112 and Lys121 assist in the positioning of the iminium intermediate for the nucleophilic addition at the C β position (Table S7 and Fig. 4, S6 & S7[†]). When Ser112 was converted into alanine, there was an enhancement of intermediate mobility. The carbonyl group became more solvent exposed, explaining the formation of the 1,2-addition byproduct in the experimental study (Table 1, entries 18–20). The Lys121Ala variant was found to provoke a significant increase in the mobility of INT-A, and hence the step of nucleophilic attack (INT-B to INT-C) can proceed from different orientations, causing a loss of enantioselectivity that is also observed experimentally (Table 1, entry 23, and Fig. S8[†]). In other words, the QM/MM study indicated that both Lys121A and Lys121B play a critical role in improving the reactivity and selectivity. While the QM/MM study confirmed the role of Lys121B in catalysis, it also contradicts the observation that there was a low but measurable activity when D-Sav-SARK serves as a host in which the corresponding Lys121A is replaced by an alanine residue. A possible explanation is that the role of Lys121A was replaced by neighboring residue(s) in this D-Sav variant which guides the nucleophile to interact with the intermediate in a stereoselective manner.

Conclusions

Our mutational, crystallographic and mechanistic studies have indicated that the protein environment corresponding to the subunit interface of T-Sav is critical for catalysis by biotinylated secondary amines. This microenvironment controls the reaction conversion, regioselectivity (1,4- over 1,2-addition) and stereoselectivity (*S* vs. *R*). In contrast, the monomeric M-Sav contains a solvent-exposed site and was found to be a poor host for iminium catalysis. The use of asymmetric D-Sav and computational analysis have indicated that both Lys121A and Lys121B play a role in improving the stereoselectivity and reactivity. We believe that the D-Sav template is the most ideal starting point for future protein engineering work; although it gives a lower yield and selectivity in comparison to its tetra-



meric counterpart, its asymmetric feature will likely allow fine-tuning of the reaction scaffold, similar to many of the well-evolved artificial enzymes.^{10,43,55–57} Indeed, an initial enzyme often has poor activity but, upon optimization by direct evolution, both the yield and selectivity can be improved exponentially within a few rounds of iterations.⁵⁶ By revealing the protein microenvironment relevant to secondary amine catalysis, the presented mechanistic study will find use in creating high-performance artificial enzymes.

Author contributions

ARN, KS, VM, and LYPL designed the study. NS performed experimental work regarding M-Sav and the T-Sav:1 ratio study. ARN performed all other experimental work regarding T-Sav, D-Sav and M-Sav E120L mutants. Crystallography was performed by ARN, NS, ML-L, and YJ. RC and KS performed MD simulations. ARN, KS, VM, and LYPL wrote the paper. All authors discussed and commented on the manuscript. LCM, KS, VM, and LYPL acquired funding.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We would like to thank Professor Dr Thomas Ward and Ms Alina Stein at the University of Basel for their discussion and donation of the dimeric T-Sav plasmid. We also thank the Diamond Light Source for access to beamlines IO3 and IO4 (MX-14843 and MX-18812) that contributed to the results presented herein. This work was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (Grant PGC2018-094852-B-C21 and PID2019-107098RJ-I00), by Cardiff University through the start-up fund provided by the Cardiff School of Chemistry, Generalitat Valenciana (Grant AICO/2019/195 and SEJI/2020/007) Universitat Jaume I (UJI-B2017-31, UJI-A2019-04), the Leverhulme Trust through a grant to L. Y. P. L. (RPG-2017-195), the Royal Society through a grant to L. C. M. (RG150466), and the UK's Wellcome Trust through grants to L. Y. P. L. (202056/Z/16/Z). The authors thankfully acknowledge the computational resources of the Servei d'Informàtica of Universitat Jaume I.

References

- 1 Y. Cotelle, V. Lebrun, N. Sakai, T. R. Ward and S. Matile, *ACS Cent. Sci.*, 2016, **2**, 388–393.
- 2 I. Drienovská, C. Mayer, C. Dulson and G. Roelfes, *Nat. Chem.*, 2018, **10**, 946–952.
- 3 L. Biewenga, T. Saravanan, A. Kunzendorf, J. Y. van der Meer, T. Pijning, P. G. Tepper, R. van Merkerk, S. J. Charnock, A. W. H. Thunnissen and G. J. Poelarends, *ACS Catal.*, 2019, **9**, 1503–1513.
- 4 Z. P. Wu and D. Hilvert, *J. Am. Chem. Soc.*, 1990, **112**, 5647–5648.
- 5 C. J. Suckling and L.-M. Zhu, *Bioorg. Med. Chem. Lett.*, 1993, **3**, 531–534.
- 6 H. Kuang, M. L. Brown, R. R. Davies, E. C. Young and M. D. Distefano, *J. Am. Chem. Soc.*, 1996, **118**, 10702–10706.
- 7 D. Häring, E. Schüler, W. Adam, C. R. Saha-Möller and P. Schreier, *J. Org. Chem.*, 1999, **64**, 832–835.
- 8 T. L. Williams, A. R. Nödling, Y. H. Tsai and L. Y. P. Luk, *Wellcome Open Res.*, 2018, **3**, 107.
- 9 C. Mayer, C. Dulson, E. Reddem, A. W. H. Thunnissen and G. Roelfes, *Angew. Chem., Int. Ed.*, 2019, **58**, 2083–2087.
- 10 A. J. Burke, S. L. Lovelock, A. Frese, R. Crawshaw, M. Ortmyer, M. Dunstan, C. Levy and A. P. Green, *Nature*, 2019, **570**, 219–223.
- 11 C. Mayer, *ChemBioChem*, 2019, **20**, 1357–1364.
- 12 I. Drienovská and G. Roelfes, *Nat. Catal.*, 2020, **3**, 193–202.
- 13 Z. Zhou and G. Roelfes, *Nat. Catal.*, 2020, **3**, 289–294.
- 14 R. B. Leveson-Gower, Z. Zhou, I. Drienovská and G. Roelfes, *ACS Catal.*, 2021, **11**, 6763–6770.
- 15 C. Guo, M. Saifuddin, T. Saravanan, M. Sharifi and G. J. Poelarends, *ACS Catal.*, 2019, 4369–4373, DOI: 10.1021/acscatal.9b00780.
- 16 J. Y. van der Meer, H. Poddar, B. J. Baas, Y. Miao, M. Rahimi, A. Kunzendorf, R. van Merkerk, P. G. Tepper, E. M. Geertsema, A. M. Thunnissen, W. J. Quax and G. J. Poelarends, *Nat. Commun.*, 2016, **7**, 10911.
- 17 E. A. Althoff, L. Wang, L. Jiang, L. Giger, J. K. Lassila, Z. Wang, M. Smith, S. Hari, P. Kast, D. Herschlag, D. Hilvert and D. Baker, *Protein Sci.*, 2012, **21**, 717–726.
- 18 L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas III, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, *Science*, 2008, **319**, 1387–1391.
- 19 R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths and D. Hilvert, *Nat. Chem.*, 2017, **9**, 50–56.
- 20 B. D. Weitzner, Y. Kipnis, A. G. Daniel, D. Hilvert and D. Baker, *Protein Sci.*, 2019, **28**, 2036–2041.
- 21 A. R. Nödling, K. Świderek, R. Castillo, J. W. Hall, A. Angelastro, L. C. Morrill, Y. Jin, Y.-H. Tsai, V. Moliner and L. Y. P. Luk, *Angew. Chem., Int. Ed.*, 2018, **57**, 12478–12482.
- 22 L. Liu, Y. Cotelle, J. Klehr, N. Sakai, T. R. Ward and S. Matile, *Chem. Sci.*, 2017, **8**, 3770–3774.
- 23 M. M. Pellizzoni, F. Schwizer, C. W. Wood, V. Sabatino, Y. Cotelle, S. Matile, D. N. Woolfson and T. R. Ward, *ACS Catal.*, 2018, **8**, 1476–1484.
- 24 H. Lechner, V. R. Emann, M. Breuning and B. Höcker, *ChemBioChem*, 2021, **22**, 1573–1577.
- 25 M. E. Wilson and G. M. Whitesides, *J. Am. Chem. Soc.*, 1978, **100**, 306–307.
- 26 A. D. Liang, J. Serrano-Plana, R. L. Peterson and T. R. Ward, *Acc. Chem. Res.*, 2019, **52**, 585–595.



- 27 M. T. Reetz, J. J. Peyralans, A. Maichele, Y. Fu and M. Maywald, *Chem. Commun.*, 2006, 4318–4320, DOI: 10.1039/b610461d.
- 28 T. Heinisch and T. R. Ward, *Acc. Chem. Res.*, 2016, **49**, 1711–1721.
- 29 T. R. Ward, *Acc. Chem. Res.*, 2011, **44**, 47–57.
- 30 D. J. Raines, J. E. Clarke, E. V. Blagova, E. J. Dodson, K. S. Wilson and A.-K. Duhme-Klair, *Nat. Catal.*, 2018, **1**, 680–688.
- 31 S. Eda, I. Nasibullin, K. Vong, N. Kudo, M. Yoshida, A. Kurbangalieva and K. Tanaka, *Nat. Catal.*, 2019, **2**, 780–792.
- 32 T. Sano and C. R. Cantor, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 142–146.
- 33 N. Humbert, P. Schürmann, A. Zocchi, J.-M. Neuhaus and T. R. Ward, in *Avidin-Biotin Interactions. Methods In Molecular Biology*, ed. Robert J. McMahon, Humana Press, 2008, vol. 418.
- 34 K. H. Lim, H. Huang, A. Pralle and S. Park, *Biotechnol. Bioeng.*, 2013, **110**, 57–67.
- 35 T. Sano, M. W. Pandori, X. Chen, C. L. Smith and C. R. Cantor, *J. Biol. Chem.*, 1995, **270**, 28204–28209.
- 36 S. Wu, Y. Zhou, J. G. Rebelein, M. Kuhn, H. Mallin, J. Zhao, N. V. Igareta and T. R. Ward, *J. Am. Chem. Soc.*, 2019, **141**, 15869–15878.
- 37 I. S. Hassan, A. N. Ta, M. W. Danneman, N. Semakul, M. Burns, C. H. Basch, V. N. Dippon, B. R. McNaughton and T. Rovis, *J. Am. Chem. Soc.*, 2019, **141**, 4815–4819.
- 38 T. K. Hyster, L. Knörr, T. R. Ward and T. Rovis, *Science*, 2012, **338**, 500–503.
- 39 M. Jeschek, M. O. Bahls, V. Schneider, P. Marliere, T. R. Ward and S. Panke, *Metab. Eng.*, 2017, **40**, 33–40.
- 40 M. Hesticová, T. Heinisch, L. Alonso-Cotchico, J. D. Marechal, P. Vidossich and T. R. Ward, *Angew. Chem., Int. Ed.*, 2018, **57**, 1863–1868.
- 41 T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marliere, S. Panke and T. R. Ward, *Chem. Sci.*, 2018, **9**, 5383–5388.
- 42 M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke and T. R. Ward, *Nature*, 2016, **537**, 661–665.
- 43 F. Christoffel, N. V. Igareta, M. M. Pellizzoni, L. Tiessler-Sala, B. Lozhkin, D. C. Spiess, A. Lledós, J.-D. Maréchal, R. L. Peterson and T. R. Ward, *Nat. Catal.*, 2021, **4**, 643–653.
- 44 J. Zhao, J. G. Rebelein, H. Mallin, C. Trindler, M. M. Pellizzoni and T. R. Ward, *J. Am. Chem. Soc.*, 2018, **140**, 13171–13175.
- 45 A. Chatterjee, H. Mallin, J. Klehr, J. Vallapurackal, A. D. Finke, L. Vera, M. Marsh and T. R. Ward, *Chem. Sci.*, 2016, **7**, 673–677.
- 46 J. M. Zimbron, T. Heinisch, M. Schmid, D. Hamels, E. S. Nogueira, T. Schirmer and T. R. Ward, *J. Am. Chem. Soc.*, 2013, **135**, 5384–5388.
- 47 M. Creus, A. Pordea, T. Rossel, A. Sardo, C. Letondor, A. Ivanova, I. Letrong, R. E. Stenkamp and T. R. Ward, *Angew. Chem., Int. Ed.*, 2008, **47**, 1400–1404.
- 48 V. M. Robles, M. Durrenberger, T. Heinisch, A. Lledós, T. Schirmer, T. R. Ward and J. D. Marechal, *J. Am. Chem. Soc.*, 2014, **136**, 15676–15683.
- 49 A. Loosli, U. E. Rusbandi, J. Gradinaru, K. Bernauer, C. W. Schlaepfer, M. Meyer, S. Mazurek, M. Novic and T. R. Ward, *Inorg. Chem.*, 2006, **45**, 660–668.
- 50 N. Santi, L. C. Morrill, K. Świderek, V. Moliner and L. Y. P. Luk, *Chem. Commun.*, 2021, **57**, 1919–1922.
- 51 N. Santi, L. C. Morrill and L. Y. P. Luk, *Molecules*, 2020, **25**, 2457.
- 52 C. Jimeno, *Org. Biomol. Chem.*, 2016, **14**, 6147–6164.
- 53 H. Mallin, M. Hesticová, R. Reuter and T. R. Ward, *Nat. Protoc.*, 2016, **11**, 835–852.
- 54 K. Świderek, A. R. Nödling, Y. H. Tsai, L. Y. P. Luk and V. Moliner, *J. Phys. Chem. A*, 2018, **122**, 451–459.
- 55 M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke and T. R. Ward, *Nature*, 2016, **537**, 661.
- 56 C. Zeymer and D. Hilvert, *Annu. Rev. Biochem.*, 2018, **87**, 131–157.
- 57 R. Obexer, A. Godina, X. Garrabou, P. R. Mittl, D. Baker, A. D. Griffiths and D. Hilvert, *Nat. Chem.*, 2017, **9**, 50–56.

