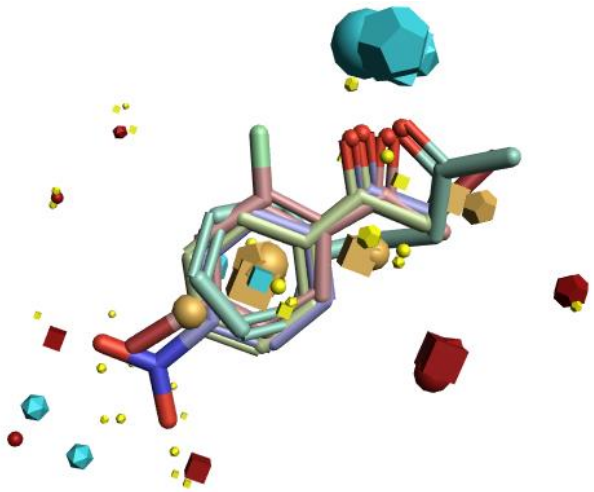
A pharmacophore-based approach to demonstrating the scope of alcohol dehydrogenases

Katrina S. Madden,*[a,b,c] Peter M.T. Todd,[a] Kouji Urata,[a] Angela J. Russell,[d,e] Kylie A. Vincent[a] and Holly A. Reeve*[a]

[a]*Department of Chemistry, University of Oxford, Inorganic Chemistry Laboratory, South Parks Road, Oxford, OX1 3QR UK*
[b]*School of Natural and Environmental Sciences, Bedson Building, Newcastle University, Newcastle Upon Tyne, NE1 7RU UK*
[c]*Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Victoria, Australia*
[d]*Department of Chemistry, University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford, OX1 3TA*
[e]*Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT UK*

holly.reeve@chem.ox.ac.uk, kate.madden@newcastle.ac.uk



ABSTRACT: Barriers to the ready adoption of biocatalysis into asymmetric synthesis for early stage medicinal chemistry are addressed, using ketone reduction by alcohol dehydrogenase as a model reaction. An efficient substrate screening approach is used to show the wide substrate scope of commercial alcohol dehydrogenase enzymes, with a high tolerance to chemical groups employed in drug discovery (heterocycle, trifluoromethyl and nitrile/nitro groups) observed. We use our screening data to build a preliminary predictive pharmacophore-based screening tool using Forge software, with a precision of 0.67/1, demonstrating the potential for developing substrate screening tools for commercially available enzymes without publically available structures. We hope that this work, combined with our simple protocols for scaleable H_2-driven biocatalytic ketone reduction, will facilitate a culture shift towards adopting biocatalysis alongside traditional chemical catalytic methods.

Introduction

Biocatalysis is now well established in late stage medicinal chemistry, providing elegant solutions to asymmetric synthesis of complex molecules, achieving high yields and selectivity under mild reaction conditions.[1–5] Such biotransformations have been incorporated into numerous large-scale drug syntheses, but typically rely on time- and resource-intensive methods for evolution or design of a highly optimised enzyme.[6–11] Development of greener, more efficient procedures throughout early stage drug discovery is highly desirable, potentially saving time and resources down the line when optimising a synthetic route for large scale production.[1–3,7,12–14]

In early stage drug discovery, the medicinal chemist's focus is on rapidly accessing a large range of chemical space to develop Structure Activity Relationships (SARs). The initial synthetic method of choice will likely be a 'tried-and-tested' route that has high versatility and can be employed to rapidly furnish a large number of structurally diverse analogues. Current preferred methods are rich in organic and metal catalysts supported by a wealth of literature detailing their use across a broad range of functionalities.[15]

Commercial biocatalysts, such as alcohol dehydrogenases, are readily available and have often been engineered to achieve broad substrate scope for ketone reduction, with high activity, stability and industrial suitability with respect to solvent tolerance and simple production at large scale. However, the sequences of commercial enzymes are typically not made available. This, combined with limited information published in the literature on the breadth of substrate scope tolerated by a given commercial enzyme represents a barrier to straightforward incorporation of commercial enzymes into a medicinal chemist's synthetic repertoire.[16] Additionally, enzymes for ketone reduction rely on biological reducing agents, such as NADH, for their catalytic activity. Due to the high expense of NADH, a cofactor recycling system must be included, adding complexity to the overall process. Consequently, biocatalysis can be viewed as a 'black box', perceived to require a good deal of expertise to both identify a suitable enzyme then to optimise and implement a biocatalytic process.

However, expanding the application of biocatalysis throughout drug discovery is timely,[17–19] as there is an increasing appreciation of the importance of 3D character and chirality in medicinal chemistry design,[20,21] with biocatalytic technologies becoming increasingly sophisticated.[2,13,22–26] Such methods could provide ready access to enantioenriched building blocks. Recent work by Turner and co-workers demonstrated the potential for transaminases to catalyse the synthesis of chiral amines from ketones, showing a promising substrate scope for a pharmaceutically-relevant transformation.[27]

Our approach is to use standard activity screening techniques, applying medicinal chemistry considerations towards substrate selection and prediction modelling to interrogate the potential of the enzyme to catalyst chemical reactions relevant to medicinal chemistry. Here, we focus on ketone reductions by alcohol dehydrogenase enzymes. We probe the substrate scope of commercial (*R*)-alcohol dehydrogenase (ADH) and (*S*)-ADH, selecting test substrates to cover a wide range of chemical space and incorporate medicinal chemistry-relevant functional groups e.g. heterocycles, F atoms or other halogens for further derivatisation.[28] The ADH enzymes used within this study are commercial and have been engineered for a broad substrate scope. Whereas substrate scope for a native enzyme or its genetic variant can be assessed by in silico docking of substrates into the crystallographically-determined 3D protein structure or a homology model of the structure, this is not possible for commercial enzymes where sequences and structures are generally not made publically available. Predicting substrate scope under these conditions is therefore reminiscent of phenotypic drug discovery, where molecule design proceeds in the absence of a protein structure. Here, compound design follows a pharmacophore-based approach, with design rationale based on interpreting trends between active compounds rather than how individual compounds interact directly with the protein of interest. Modelling approaches in this field focus on predicting the relationship between compound properties e.g. electrostatic field and activity, rather than target-based docking methods. This provides us with an opportunity to employ the tools available for medicinal chemistry, in this case the pharmacophore-based modelling software Forge, to develop a screening tool which predicts the likely reactivity of a substrate based on its chemical features, without requiring any structural knowledge of the enzyme.[29] This strategy represents an accessible approach to evaluating substrate scope, that we hope will encourage the uptake of biocatalysis in early-stage medicinal

chemistry, offering a degree of reassurance that an enzyme will display desirable activity for a given set of reactions.

Results and Discussion

Our focus was to address three key areas of biocatalysis with relevance to medicinal chemistry:

- Suitability for use in diverse synthesis
- Ease of use without specialist equipment
- Confidence in biocatalysis as a go-to method

We sought to assess the substrate scope of two commercially available ADH enzymes with opposite enantioselectivities, (*R*)-ADH (Johnson Matthey, ADH101) and (*S*)-ADH (Johnson Matthey, ADH105), as defined with respect to their reduction of acetophenone **1**, in order to evaluate their potential tolerance for drug-like motifs. The descriptors (*R*) and (*S*) used throughout this manuscript refer to the ADH enzyme used and not to the absolute configuration of products formed. These enzymes are known to have broad substrate scope and, as is typical for alcohol dehydrogenases (ketone reductases), use NAD(P)H cofactor as a source of hydride for the C=O reduction. Standard NADH-consumption screening techniques were used to determine the rate of NADH oxidation, which was taken as a read-out for ketone reduction. Product verification was carried out for a subset of reactions, and should be completed before commencing scale up work. Taking insight from the drug discovery field, here we push the substrate screens further into medicinal chemistry space to access their practicality. Further information on the activity and selectivity of alcohol dehydrogenases is widely reported.[30,31]

Suitability for and ease of use in diverse synthesis

In order to rapidly build a picture of the substrate scope of these ADHs, we elected to screen substrates in a 96-well plate format and monitor the reactions via consumption of NADH over time, using UV-Vis spectroscopy (see **Supplementary Information** for full details on assay set up). Our rationale was to use the specific enzyme activity (SEA, nmol min^{-1} mg^{-1} of enzyme) for NADH consumption as a readout for reactivity with the substrate. As both ADHs are well characterised for reduction of acetophenone **1**, it was included in substrate and condition screens to provide a positive control and benchmark for reactivity.

Initially, we screened over 40 substrates with (*S*)-ADH and (*R*)-ADH in 50 mM Tris HCl buffer (**Figure 1a**, reaction conditions (i)), selecting acetophenone analogues to cover a range of chemical space and beginning to include drug-like motifs. We immediately observed a surprisingly good tolerance for a range of motifs, including fluoropyridine **23**, halobenzenes **2**-**5**, homologation of the alkyl chain to include longer chain arylated ketones such as **31**-**33**, and non-aromatic ketones **35**-**36**.

The substrate reactivity of the two ADHs differed, exemplified by the unreactivity of nitrobenzene **20** with (*S*)-ADH, but high activity with (*R*)-ADH. Higher specific enzyme activity values were observed in general for (*S*)-ADH, along with a higher level of background NADH-oxidising activity (see **Supplementary Information** for details), which was taken into account when evaluating the reactivity towards the various substrates.

Solubility in water was a challenge, particularly for substrates incorporating increasingly lipophilic groups, where heavy precipitation was often observed. We therefore investigated the tolerance of the (*R*)/(*S*)-ADH system to varying percentages of a range of organic solvents with the aim of identifying a bilayer or miscible co-solvent system, suitable for both the enzyme and for improving the solubility of organic compounds (**Figure 1b**). Low concentrations of dimethylsulfoxide, ethyl acetate and dichloromethane were tolerated, as well as hexane in a bilayer comprising up to 80% of the total solvent volume. Both ethyl acetate and dichloromethane damaged the integrity of the 96-well plates, therefore their use would require an alternative screening method. However, hexane appeared consistently well tolerated and was compatible with the plasticware, therefore 20% hexane was selected as a balance between minimising the proportion of organic solvent and maximising substrate availability.

We then rescreened the substrates in **Figure 1a** utilising the 20% hexane bilayer (conditions (ii)), and extended the substrate scope to include those in **Figure 2**, incorporating drug-like motifs such as CF_3 pyridine **49**, nitrile **58**, pyrazine **52**, pyrazole **55**-**56**, methanesulfonamide **57**, and pyrimidine **51**. We found that solubility

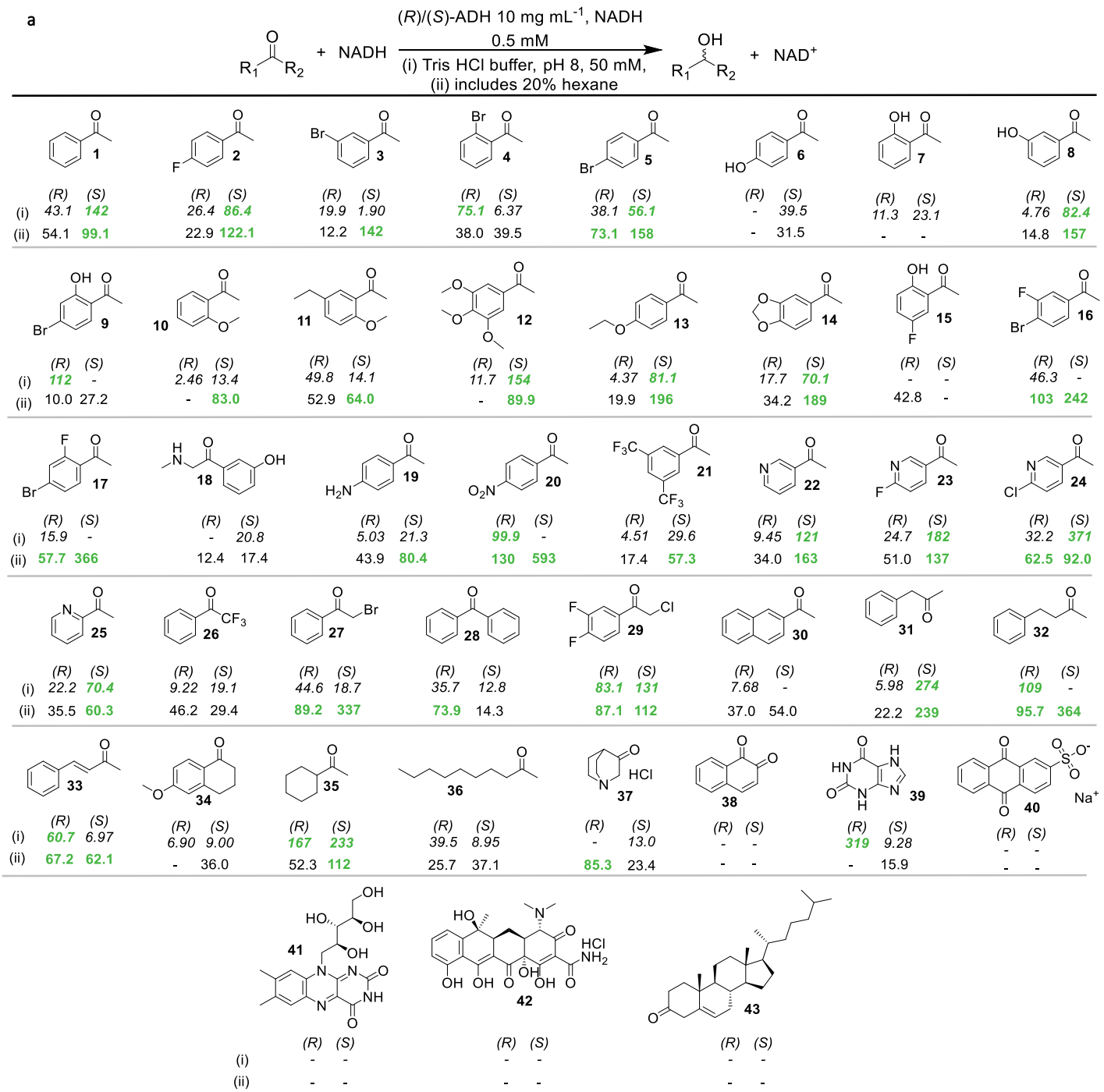


Figure 1 (a) Specific enzyme activity (SEA) values as nmol min^{-1} mg^{-1} for (*R*)/(S)-ADH, relating to consumption of NADH cofactor over time. **The descriptors (*R*) and (*S*) within the figure refer to the ADH enzyme used.** Values are associated with different conditions: (i) Tris HCl buffer, pH 8, 50 mM and (ii) 20% hexane in Tris HCl buffer, pH 8, 50 mM. UV-vis spectra recorded every 30 seconds for 30 minutes, SEA calculated at 364 nm with ε 3158 M^{-1} cm^{-1}, with each substrates run at 3 concentrations of 2.5, 5 and 10 mM and the highest SEA value observed of the 3 concentrations reported (concentrations giving the highest SEA value are detailed in the **Supplementary Information**), results reported to 3 s.f., '-' denotes no consumption of NADH observed, values in green (> 50 nmol min^{-1} mg^{-1}) highlight substrates displaying higher levels of cofactor turnover. (b) Effect of various solvents on the specific enzyme activity (SEA) for (*R*)/(S)-ADH, relating to consumption of NADH cofactor over time. UV-vis spectra recorded every 30 seconds for 30 minutes, SEA calculated at 364 nm with ε 3158 M^{-1} cm^{-1}, using 10 mM acetophenone as substrate. Abbreviations: dimethylsulfoxide (DMSO), ethyl acetate (EtOAc), acetonitrile (MeCN), dichloromethane (DCM), methanol (MeOH), isopropanol (IPA). *wrt average SEA value obtained for 10 mM acetophenone using 2% v/v DMSO in 50 mM Tris HCl buffer.

presented much less of an issue, with improved specific enzyme activity values observed for a number of substrates previously showing poor solubility in water.

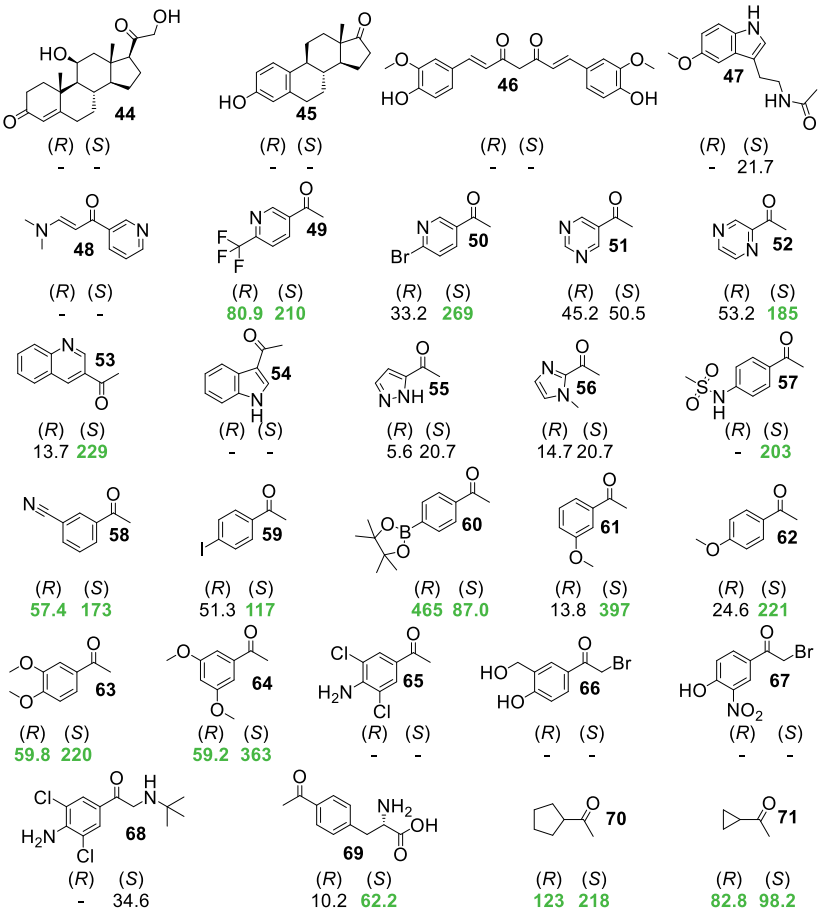


Figure 2 Specific enzyme activity values measured in nmol min^{-1} mg^{-1} for (*R*)/(S)-ADH in 20% hexane bilayer in 50 mM Tris HCl buffer, relating to consumption of NADH cofactor over time (conditions as detailed in Figure 1 a) (ii)). **The descriptors (*R*) and (*S*) within the figure refer to the ADH enzyme used.** UV-vis spectra recorded every 30 seconds for 30 minutes, SEA calculated at 364 nm with ε 3158 M^{-1} cm^{-1}, with each substrates run at 3 concentrations of 2.5, 5 and 10 mM and the highest SEA value observed of the 3 concentrations reported (concentrations giving the highest SEA value are detailed in the **Supplementary Information**), results reported to 3 s.f., '-' denotes no consumption of NADH observed, values in green (> 50 nmol min^{-1} mg^{-1}) highlight substrates displaying higher levels of cofactor turnover.

Gratifyingly, both enzymes tolerated a diverse scope of functional groups spanning a range of electronics and sterics. Compatible groups also included those useful for further derivatisation e.g. boronate ester/halogen for further cross-coupling reactions, halogenated heterocycles for subsequent nucleophilic aromatic substitution (S_NAr). We recognised that our substrate scope screens indicate reactivity by measuring consumption of cofactor, rather than tracking product production, which means that structural information about the products is not obtained during screening. In order to shed some light on the integrity of potentially labile groups following (*R*)/(*S*)-ADH reduction, we tracked the reduction of a series of halogenated acetophenones using (*R*)/(*S*)-ADH (for more details, see **Supplementary Information**), comparing the reactions to samples of both the halogenated products and the de-halogenated alcohol using chiral GC. Here we observed that the halogens were retained in the reduction products, with *ee*s of 98% or higher measured after 2 hours.

These results, in addition to analogous experiments detailed in our previously published work, give us confidence that sensitive groups such as halogens will be retained in the reduced products.[32] There were some trends discernible both individually for (*R*)/(*S*)-ADH, and some general trends in reactivity across both enzymes. For example, 2-OH **7** was not tolerated by either enzyme, and 4-OH **6** was poorly tolerated by (*S*)-ADH whilst unreactive with (*R*)-ADH. Likewise, pyrazole groups did not appear to be tolerated by either enzyme (**55**-**56**).

In general (*S*)-ADH displayed higher specific enzyme activity values than (*R*)-ADH, and was more tolerant of a wider range of substrates eg. phenylmethanesulfonamide **57**. However, (*R*)-ADH appeared to better tolerate more chemically diverse substrates such as benzophenone **28** and quinuclidone HCl **37**. Neither enzyme tolerated large and complex bioactive molecules such as steroids **43**-**45** or the antibiotic tetracycline **42**.

With some preliminary structure-reactivity relationships emerging, and a good level of reactivity for both enzymes observed across chemical space, this screening effort showed clear potential for use of ADHs in diverse chemical synthesis as appropriate to early stage medicinal chemistry.

In order to demonstrate the translation of the above cofactor consumption screens to real-world biocatalytic reduction, a cofactor recycling system is required. This ensures a continuous supply of NADH, and allows sub-stoichiometric quantities of the expensive cofactor to be used. Such cofactor recycling strategies are well established, usually relying on glucose dehydrogenase, or a substrate-couple approach with alcohol dehydrogenase.[33] Here we use H_2-driven biocatalytic cofactor recycling which affords 100 % atom efficient reactions, and simplifies product analysis and isolation, and translates easily into standard hydrogenation laboratory equipment, used on the bench.[34]

We paired ADH-catalysed reduction of 1-(6-(trifluoromethyl)pyridin-3-yl)ethanone **49** and 3-acetylbenzonitrile **58** with H_2-driven cofactor recycling (**Figure 3**). The experimental set up considerations for these biocatalytic reductions were similar to that of Pd/C hydrogenations: reactions could be performed in a fume cupboard under a balloon of hydrogen with solvent degassing, without the requirement for specialist, as detailed in the **Supplementary Information**. We followed the reductions using NMR and chiral GC, observing excellent conversions, particularly for the (*S*)-ADH catalysed reductions, where alcohols **73** and **75** were produced with near perfect yield and enantioselectivity. We have recently demonstrated H_2-driven biocatalytic hydrogenations in the H-cube, utilising enzymes in CatCarts™, highlighting this approach as a convenient method for scale up.[35] These results, combined with our previous batch studies of halogenated acetophenones suggest a great deal of promise for the ability to perform biocatalytic reductions on a range of medicinally relevant substrates, whilst retaining high levels of enantioselectivity. This lends strength to the substrate screens previously performed for their translatability to the lab.

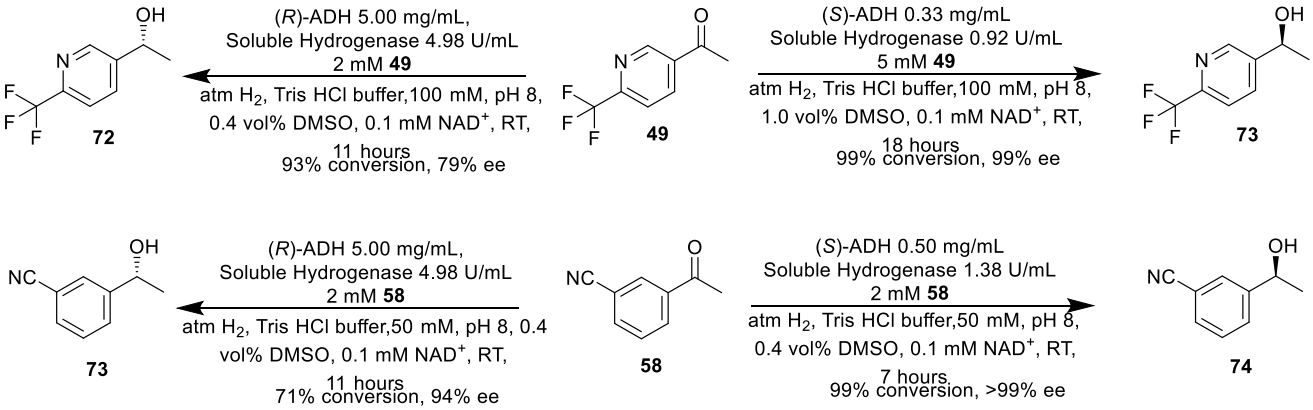
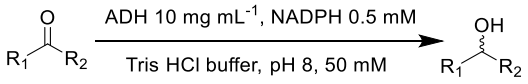


Figure 3 Reduction of **49** and **58** using H_2-driven cofactor recycling.

With these results in hand we sought to investigate whether we could identify an alternative ADH enzyme for the reduction of select substrates which displayed poor reactivity with (*R*)/(*S*)-ADH. These included adrenergic receptor drug precursors **18** and **68** (**Table 1**), the reduced versions of which are (*R*)-phenylephrine and clenbuterol. Whilst pyrazole **55** and 4-hydroxybenzene derivative **6** did not seem to cause cofactor (in this case NADPH) turnover in any of our alternative ADHs, we were pleased to observe that ADH20 appeared to tolerate **18** and **68**, suggesting that it will be possible to develop scaleable biocatalytic reductions of adrenergic receptor drug precursors. A similar approach to substrate screening of these alternative ADHs could open up even more chemical space, in particular allowing for late-stage reductions to furnish drugs such as (*R*)-phenylephrine and clenbuterol, along with a range of other structurally related drugs within this class e.g. levalbuterol ((*R*)-salbutamol).

Table 1 Exploring reactivity of problematic substrates with alternative ADHs (from Johnson Matthey library).



Compound	Specific Enzyme Activity/ nmol min^{-1} mg^{-1}			
	ADH19	ADH20	ADH61	ADH150
6	12.8	-	12.7	-
18	6.78	**67.2**	-	38.6
28	-	41.1	-	**69.9**
34	14.3	31.3	**56.4**	7.83
55	10.3	24.1	10.7	11.8
68	7.76	**55.7**	**123**	33.7
69	10.4	18.8	29.0	6.55

Specific enzyme activity values measured in 50 mM Tris HCl buffer, relating to consumption of NADPH cofactor over time. UV-vis spectra recorded every 30 seconds for 30 minutes, SEA as nmol min^{-1} mg^{-1} calculated at 364 nm with ε 3158 M^{-1} cm^{-1}, with each substrates run at 2 concentrations of 2.5 and 10 mM and the highest SEA value observed of the 2 concentrations reported (concentrations giving the highest SEA value are detailed in the ***Supplementary Information****)results reported to 3 s.f., '-' denotes no consumption of NADPH observed, values in green (> 50 nmol min^{-1} mg^{-1}) highlight substrates displaying higher levels of cofactor turnover.*

Confidence in biocatalysis as a go-to method

With the sequence of (*R*)/(*S*)-ADH not publically disclosed, structure-based docking studies into the enzyme active sites were not available to us, and indeed we sought to identify a solution which was more accessible to synthetic chemists, that did not rely on advanced computational skills to assess the potential reactivity of a given substrate. To this end, we instead applied a pharmacophore-based approach to modelling the reactivity of (*S*)-ADH, using Cresset's Forge software.[29] The intention behind this was to develop an end user-friendly screening tool which could predict the likely reactivity of a given substrate without the need for knowing the enzyme sequence. We ultimately envisioned an easy to use interface where a synthetic chemist could enter their substrate query and receive a reactivity prediction without needing to manipulate the model themselves. Our strategy was analogous to building 3D Quantitative Structure Activity Relationship (QSAR) models in drug discovery.[17,36] By aligning low energy conformations of the tested substrates, then using their field properties to generate similarity scores and compare those to their reactivity in the NADH consumption assay, we aimed to generate a pharmacophore-based screening tool which could return an assessment of the likely reactivity of a substrate molecule naïve of the active site sequence. We adopted a pragmatic approach to the development of this tool, focussing on the key information needed by a prospecting synthetic chemist asking the question "will this substrate likely be reduced by (*S*)-ADH?"

We therefore categorised the experimentally obtained specific enzyme activity data into 3 categories:

1. Poor reactivity with (S)-ADH (SEA <30 nmol min^{-1} mg^{-1})
2. Moderate reactivity with (S)-ADH (30<SEA<90 nmol min^{-1} mg^{-1})
3. Good reactivity with (S)-ADH (SEA >90 nmol min^{-1} mg^{-1})

This would allow synthetic chemists to filter out potential substrates that have not been tolerated in the past by (*S*)-ADH whilst minimising the likelihood of missing a possible substrate, something we felt to be important in early stage medicinal chemistry where diverse exploration of chemical space was a key goal.

We partitioned the data obtained using 20% hexane in **Figures 1a** and **2** into a training set comprising 80% of the input compounds and an activity stratified test set comprising 20% of the input compounds. In order to develop the most relevant tool we also used the FieldTemplater module in Forge to identify a general reference pharmacophore from a small number of structurally diverse but reactive substrates (**Figure 4**).[29] This process provides a basis for more accurate model building, as it builds up a picture of the different structural features within the substrates that allow for high specific enzyme activity values, providing a general structural reference for reactivity from the data set.

Figure 4 Alignment of substrates used to generate a 'reactive' pharmacophore from a number substrates displaying high SEA values with (*S*)-ADH with different structural features in FieldTemplater, Forge, Cresset.

Even with our relatively small data set, we were able to generate a screening tool which could correctly assign a reactivity category to the test set compounds with precision and recall values of 0.67/1.00 (for full details see the **Supplementary Information**). A comparison of predicted and experimentally-determined reactivity categories is shown in **Table 2**. The screening tool correctly predicted the reactivity category for 9 of the 13 test compounds, with the remaining 4 being assigned to a category directly adjacent i.e. predicted category 1, experimental category 2. None of the test compounds were predicted to be unlikely to react when the experimental data showed a large specific enzyme activity, or vice versa. Compounds **4** and **10**, both *ortho*-substituted acetophenones, were predicted to be unlikely to react with (*S*)-ADH- this is perhaps unsurprising when considering the poor reactivity of substrates containing *ortho*-substituents in the training set (**Figure 1a**, substrates **7**, **9** and **15**). Compound **17**, however, had a very similar substrate match within the training set in substrate **16**, which likely contributed to its correct predicted activity despite the presence of an *ortho*-F group. In cases where there were more closely matched substrates in the training set, the tool performed well e.g. substrate **5**.

Table 2 Predicted vs experimentally-determined activity category and measured specific enzyme activity (SEA) of test set compounds with (S)-ADH used in the development of a pharmacophore-based screening tool.

Compound	Reactivity with (S)-ADH		
	Predicted activity category using Forge 3D QSAR model	Experimental activity category	Measured Specific Enzyme Activity/ nmol min^{-1} mg^{-1}
4	1	2	39.5
5	3	3	158
10	1	2	83.0
17	3	3	366
21	2	2	57.3
22	3	3	163
24	3	3	92.0
28	1	1	14.3
37	1	1	23.4
50	3	3	269
58	2	3	173
65	2	1	-
68	2	2	34.6

This indicates that a pharmacophore-based model could be developed to act as a highly predictive early stage screening tool, helping the synthetic chemist decide the likelihood of successful reduction of a target ketone by (*S*)-ADH. We anticipate that a more extensive screening campaign to increase the data set will furnish such a screening tool. This premise could be extended beyond our model ADH system, to develop tools for other

biocatalysts and considerably facilitate the use of enzymes in synthetic chemistry, particularly where the enzyme sequence is not publically known.

Conclusions

Within this work, we combined multiple approaches (broad and high-throughput substrate screening, targeted enzyme screening and pharmacophore-based model building) to demonstrate the prowess of biocatalysis as a complementary method to traditional organic chemistry. In this case study, we have demonstrated the versatility and tolerance of (*R*)/(*S*)-ADH to a wide variety of chemical substrates which could allow for easy access to diverse libraries of compounds with high conversions and enantioselectivity. We then demonstrated facile batch biocatalytic hydrogenation procedures and progress towards a screening tool able to filter out substrates unlikely to be reduced by (*R*)/(*S*)-ADH. This shows the clear potential of our approach for predicting the likely reactivity of other enzymes that catalyse an array of highly useful reactions.

We believe this work represents a significant step towards making biocatalysis more accessible, providing the medicinal chemistry community with new ways to incorporate biocatalysts into early stage medicinal chemistry. Our goal is to build on this research, developing robust experimental procedures for the synthesis of milligram quantities of chiral drug-like fragments, focusing on access to diverse chemical space.

Conflicts of interest

There are no conflicts to declare

Acknowledgements

K.M., K.U., K.V. and H.R are supported by Engineering and Physical Sciences Research Council (EPSRC) IB Catalyst award EP/N013514/1. We are grateful to Dr Beatriz Dominguez and Johnson Matthey for providing the ADH enzymes used within this study, and Cresset for granting a free evaluation of Forge which was used to perform the modelling aspects of this work. We also thank Dr Sarah Cleary and Dr Jack Rowbotham for their extremely insightful comments and discussions throughout preparation of this manuscript.

References

1 A. R. Alcántara, *Biocatalysis and pharmaceuticals: A smart tool for sustainable development*, MDPI AG, 2019, vol. 9.
2 N. J. Turner and R. Kumar, *Curr. Opin. Chem. Biol.*, 2018, 43, A1–A3.
3 B. C. Buckland, D. K. Robinson and M. Chartrain, *Metab. Eng.*, 2000, **2**, 42–48.
4 W. Jiang and B. Fang, *Appl. Biochem. Biotechnol.*, 2020, 1–34.
5 A. J. Burke, C. S. Marques, N. J. Turner and G. J. Hermann, *Active Pharmaceutical Ingredients in Synthesis*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2018.
6 C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman and G. J. Hughes, *Science*, 2010, **329**, 305–309.
7 A. A. Desai, *Angew. Chemie, Int. Ed. English*, 2011, **50**, 1974–1976.
8 F. Parmeggiani, A. Rué Casamajo, D. Colombo, M. C. Ghezzi, J. L. Galman, R. A. Chica, E. Brenna and N. J. Turner, *Green Chem.*, 2019, **21**, 4368–4379.
9 M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra-Garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. McLaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel, A. Rodriguez-Granillo, S. A. Robaire, E. C. Sherer, M. D. Truppo, A. M. Whittaker, D. Verma, L. Xiao, Y. Xu and H. Yang, *Science*, 2019, **366**, 1255–1259.
10 W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz and J. Brands, *Science*, 2010, **329**, 305–310.
11 B. A. Anderson, M. M. Hansen, A. R. Harkness, C. L. Henry, J. T. Vicenzi and M. J. Zmijewski, *J. Am. Chem. Soc.*, 1995, **117**, 12358–12359.
12 R. A. Sheldon and J. M. Woodley, *Chem. Rev.*, 2018, **118**, 801–838.
13 P. N. Devine, R. M. Howard, R. Kumar, M. P. Thompson, M. D. Truppo and N. J. Turner, *Nat. Rev. Chem.*, 2018, 2, 409–421.
14 E. M. M. Abdelraheem, H. Busch, U. Hanefeld and F. Tonin, *React. Chem. Eng.*, 2019, **4**, 1878–1894.
15 C. H. Senanayake, D. R. Fandrick, J. J. Song, C. Busacca, H. C. Shen, J. Yin, W. A. Szabo, V. Yeh, O. R. Thiel, C. K. Chung, L. Terrell and H.-U. Blaser, *Applications of Transition Metal Catalysis in Drug Discovery and Development: An Industrial Perspective*, John Wiley & Sons, Inc., New York, 2012.
16 N. C. Goodwin, J. P. Morrison, D. E. Fuerst and T. Hadi, *ACS Med. Chem. Lett.*, 2019, **10**, 1363–1366.
17 C. Liu, J. Yin, J. Yao, Z. Xu, Y. Tao and H. Zhang, *Front. Cell. Infect. Microbiol.*, 2020, **10**, 118.
18 M. Hönig, P. Sondermann, N. J. Turner and E. M. Carreira, *Angew. Chem. Int. Ed.*, 2017, **56**, 8942–8973.

19 R. O. M. A. de Souza, L. S. M. Miranda and U. T. Bornscheuer, *Chem. - A Eur. J.*, 2017, 23, 12040–12063.
20 W. R. J. D. Galloway, A. Isidro-Llobet and D. R. Spring, *Nat. Commun.*, 2010, 1, 1–13.
21 A. W. Hung, A. Ramek, Y. Wang, T. Kaya, J. A. Wilson, P. A. Clemons and D. W. Young, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 6799–6804.
22 K. Faber, W.-D. Fessner and N. J. Turner, *Adv. Synth. Catal.*, 2019, **361**, 2373–2376.
23 K. Chen and F. H. Arnold, *Nat. Catal.*, 2020, 3, 203–213.
24 C. K. Prier and F. H. Arnold, *J. Am. Chem. Soc.*, 2015, **137**, 13992–14006.
25 M. J. Abrahamson, E. Vazquez-Figueroa, N. B. Woodall, J. C. Moore and A. S. Bommarius, *Angew. Chemie-International Ed.*, 2012, **51**, 3969–3972.
26 G.-D. Roiban, M. Kern, Z. Liu, J. Hyslop, P. L. Tey, M. S. Levine, L. S. Jordan, K. K. Brown, T. Hadi, L. A. F. Ihnken and M. J. B. Brown, *ChemCatChem*, 2017, **9**, 4475–4479.
27 J. Mangas-Sanchez, M. Sharma, S. C. Cosgrove, J. I. Ramsden, J. R. Marshall, T. W. Thorpe, R. B. Palmer, G. Grogan and N. J. Turner, *Chem. Sci.*, 2020, **11**, 5052–5057.
28 A. A. Koesoema, D. M. Standley, T. Senda and T. Matsuda, *Appl. Microbiol. Biotechnol.*, 2020, **104**, 2897–2909.
29 T. Cheeseright, M. Mackey, S. Rose and A. Vinter, *J. Chem. Inf. Model.*, 2006, **46**, 665–676.
30 H. G. Naik, B. Yeniad, C. E. Koning and A. Heise, *Org. Biomol. Chem.*, 2012, **10**, 4961–4967.
31 Y. G. Zheng, H. H. Yin, D. F. Yu, X. Chen, X. L. Tang, X. J. Zhang, Y. P. Xue, Y. J. Wang and Z. Q. Liu, *Appl. Microbiol. Biotechnol.*, 2017, 101, 987–1001.
32 J. S. Rowbotham, M. A. Ramirez, O. Lenz, H. A. Reeve and K. A. Vincent, *Nat. Commun.*, 2020, **11**, 1454.
33 A. Weckbecker, H. Groeger and W. Hummel, in *Biosystems Engineering I: Creating Superior Biocatalysts*, eds. C. Wittmann and W. R. Krull, 2010, vol. 120, pp. 195–242.
34 L. Lauterbach, O. Lenz and K. A. Vincent, *FEBS J.*, 2013, **280**, 3058–3068.
35 B. Poznansky, L. Thompson, H. Reeve and K. Vincent, *ChemRxiv*, , DOI:10.26434/CHEMRXIV.12532301.V1.
36 Q. Gao, L. Yang and Y. Zhu, *Curr. Comput. Aided-Drug Des.*, 2010, **6**, 37–49.

Supplementary information

Biocatalysis for medicinal chemists; a pharmacophore-based approach to demonstrating the scope of alcohol dehydrogenases

Katrina S. Madden ,*[a,b] Peter M.T. Todd,[a] Kouji Urata,[a] Angela J. Russell,[c,d] Kylie A. Vincent[a] and Holly A. Reeve*[a]

Contents

Substrate selection for screening

Substrate selection for screening against (*R*)/(*S*)-ADH was heavily influenced by our experience in medicinal chemistry, in addition to commercial availability of precursor ketones. Our experience in phenotypic medicinal chemistry, where compound optimisation is performed without knowledge of the molecular target, was used to select a range of acetophenone-based substrates aiming to cover as wide a range of chemical space as possible. We particularly focussed here on including different types of functional groups in order to see how this might impact reactivity with (*R*)/(*S*)-ADH. Following initial exploratory substrate selection, we then moved towards assessing substrates with more medicinally relevant functional groups, as governed by our experience in previous drug discovery programmes. Our aim was not to be comprehensive, but more to assess different classes of 'drug-like' motifs to inform pharmacophore-based modelling, and provide a set of substrates familiar to a medicinal chemist as being useful. We also deliberately incorporated a range of substrates with a high degree of complexity and structural disparity to our main substrate set, using bioactive compounds such as steroids. Through these substrate selection processes we aimed to generate a set of data that established a good picture of (*R*)/(*S*)-ADH substrate tolerance, including examples which would not show any reactivity due to their complexity.

General experimental details

Reagents and solvents

Reagents, including substrates for screening and buffer salts, were purchased from a range of commercial suppliers, including Merck, Fluorochem and AlfaAesar. These were used as received without further purification. Substrates were stored as 500 mM stock solutions in DMSO at 4 °C, and were reused for further screening experiments where compound stability allowed. New stock solutions were made up where necessary. NADH and NAD+ cofactors were sourced from Prozomix, stored at -18 °C, and stock solutions were made up fresh no earlier than 1-2 hours before each experiment. All solutions used in the hydrogenation work were prepared using MilliQ water (Millipore, 18 MΩcm)

Catalysts

Commercial samples of (*R*)/(*S*)-ADH (ADH101 and ADH105) were obtained from Johnson Matthey (JM) in lyophilised form, stored at -18 °C, and used without further purification. Stock solutions were made up fresh immediately before each experiment. Soluble hydrogenase used in this work was from Ralstonia eutropha and prepared in-house.

Analytical methods

NADH consumption was monitored by **UV-Vis spectroscopy**, measured at 364 nm. UV-Vis absorbances were either plotted against time (min) as line graphs in Origin, or first converted into M concentration values using ε 3158 M^{-1} cm^{-1}, as calculated within our laboratory, and then plotted as line graphs of concentration over time (min) in Origin. A linear curve fit was applied to the first 5 minutes of the graph in order to obtain an initial rate (either change in absorbance min^{-1} or change in NADH concentration min^{-1}), and this was then converted into a specific enzyme activity value of mmol min^{-1} mg^{-1}, incorporating ε for NADH at 364 nm if this had not already been done.

Chiral phase GC-FID was used to monitor reaction conversion and enantiomeric purity, comparing product and reactant retention times against commercial standards. 300 µL of the reaction mixture was extracted with 600 µL of EtOAc, before being transferred to the glass vial for chiral phase GC-FID analysis, using the following method:

Column: CP-Chirasil-Dex CB (Agilent), 25 m length, 0.25 mm diameter, 0.25 µm (film thickness), fitted with a guard of 10 m undeactivated fused silica of the same diameter

Carrier: He (CP grade), 170 kPa (constant pressure)

Inlet temperature: 200 °C

Injection conditions: Splitless with split flow 60 mL/min, splitless time 0.8 mins, purge 5 mL/min. Injection volume = 0.5 µL.
Detection: FID (H_2 = 35 mL/min, air = 350 mL/min, makeup N_2 = 40 mL/min, temp = 200 °C)
Oven heating profile:
0- 5 min, hold at 70°C
5 – 30 min, ramp to 120 °C at 2 °C/min
30 -36 min, ramp to 180 °C at 10 °C/min
36 -45 min, hold at 180 °C for 5 minutes

^{1}H NMR analysis was carried out as follows:

450 μL of the reaction mixture was extracted with 800 µL of Chloroform-^{2}H, and 600 µL was transferred to a Norell® SelectSeries™ 5 mm 400 MHz NMR sample tube.

^{1}H NMR spectroscopy was carried out on Bruker Avance III (500 MHz) at 298 K, using the following parameters.

Nucleus	^{1}H
RF pulse energy (MHz)	499.9
Temperature (K)	298 ± 2
Number of scans	*As required*
Pulse width (μs)	10.3
Spectral width (Hz)	8000
Acquisition time (s)	2.04
Relaxation delay (s)	2.00

General procedures

Preparation of Tris Buffer

The required weight of Trizma® base (e.g. 6.06 g for 500 mL 100 mM Tris buffer), was dissolved in the corresponding volume of milliQ water and mixed until complete dissolution. The solution was then corrected to pH 8 with 3M HCl, using a pH meter, and sparged overnight with N_2 to give the final buffer.

NADH consumption substrate screening assay

NADH consumption substrate screening assays were performed in ThermoFisher 96-well clear plastic plates, using single and multi-channel pipettes. The reaction components were added to the wells in the following order: Bulk solvent(s), DMSO where necessary to standardise DMSO concentration across the plate, NADH in Tris buffer, substrate stock solution in DMSO, enzyme. On each addition the component to be added was mixed by pipetting up and down a few times, then the reaction mixture pipetted up and down in the same way after addition. Throughout plate preparation, care was taken to replace pipette tips as necessary to prevent any potential contamination or diluting of

the components. The plate was swirled in between each addition, ensuring the plate was quickly and carefully swirled on the final addition of enzyme before placing in the reader. For the most consistent results it was important to begin plate reading within 1 minute of adding the enzyme.

NB: For some substrates, particularly within the Tris/water reaction system, a precipitate was formed in the well on addition of substrate stock. For best results it was important to redissolve the substrate as much as possible by more extensive pipetting *before* addition of the enzyme. Where substrates displayed solubility issues in DMSO, extensive pipetting was performed to homogenise the suspension as much as possible before adding to the well plate. Often in the case of the 20% hexane:Tris/water bilayer, the substrate would then dissolve in the reaction mixture.

H_2-driven cofactor recycling batch reductions

All reactions were set up in a fume hood at room temperature and were conducted on a 1 mL scale in sealed 2 mL-glass vials under H_2 balloon (**Supplementary Figure 1**). All buffers were pre-saturated by bubbling with H_2 gas for at least 30 mins. In order to improve the solubility, dimethylsulfoxide was added, with the final DMSO concentration of 0.1-0.4 vol%. Tris-HCl (100 mM, pH 8.0) buffer was used for all hydrogenation reactions. Reaction compositions were prepared as specified in the specific experimental procedures, and were stirred at 200 rpm. Control experiments were set up in the absence of SH to ensure there was no background substrate loss. Reaction mixtures were then analysed by Chiral phase GC-FID and ^{1}H NMR to determine conversion and ee.

Supplementary Figure 1: The set-up used for batch H_2-driven biocatalytic reductions

Specific experimental details

Tris HCl buffer NADH consumption screening

The general procedure for NADH consumption substrate screening assay was followed, using 50 mM Tris HCl buffer as the solvent system, prepared by diluting the 100 mM Tris HCl stock with milliQ water. Each compound was screened at concentrations of 2.5, 5 and 10 mM, with the SEA value reported being the highest value obtained out of the 3 concentrations.

Results

Compound	Specific Enzyme Activity (*R*)-ADH/ nmol mg^{-1} min^{-1}	Concentration/ mM	Specific Enzyme Activity (*S*)-ADH/ nmol mg^{-1} min^{-1}	Concentration/ mM
1	43.1	10	142	2.5
2	26.4	10	86.4	2.5
3	19.9	2.5	1.90	5

4	75.1	5	6.37	2.5
5	38.1	10	56.1	10
6	-	-	39.5	5
7	11.3	2.5	23.1	2.5
8	4.76	10	82.4	10
9	112	10	-	-
10	2.46	10	13.4	5
11	49.8	5	14.1	10
12	11.7	10	154.7	10
13	4.37	5	81.1	10
14	17.7	10	70.1	2.5
15	-	-	-	-
16	46.3	10	-	-
17	15.9	10	-	-
18	-	-	20.8	2.5
19	5.03	5	21.3	10
20	99.9	2.5	-	-
21	4.51	5	29.6	5
22	9.45	10	121	10
23	24.7	10	182	10
24	32.2	5	371	2.5
25	22.2	10	70.4	10
26	9.22	5	19.1	10
27	44.6	5	18.7	5
28	35.7	5	12.8	5
29	83.1	10	131	10
30	7.68	5	-	-
31	5.98	10	274	10
32	109	10	-	-
33	60.7	10	6.97	2.5
34	6.90	2.5	9.00	10
35	167	10	233	10
36	39.5	10	8.95	5
37	-	-	13.0	10
38	-	-	-	-
39	319	10	9.28	2.5
40	-	-	-	-
41	-	-	-	-
42	-	-	-	-
43	-	-	-	-

Co-solvent screening

The general procedure for NADH consumption substrate screening assay was followed, using varying concentrations of a range of solvents as specified below in 50 mM Tris HCl buffer as the solvent system, and AcPh **1** throughout the experiment at a concentration of 10 mM.

Results

	Specific enzyme activity/ nmol min^{-1} mg^{-1}						
Enzyme and % solvent v/v	DMSO	EtOAc	Hexane	MeCN	DCM	MeOH	IPA
(*R*)-ADH 2%	26.0	12.6	42.5	-	19.8	22.4	-
(*R*)-ADH 5%	57.7	14.7	26.5	-	43.9	22.6	-
(*R*)-ADH 10%	46.2	10.2	43.0	18.3	21.1	-	-
(*R*)-ADH 20%	75.3	42.6	29.4	4.56	35.2	-	-
(*R*)-ADH 40%	27.6	15.1	68.4	-	31.9	-	-
(*R*)-ADH 80%	-	-	-	-	-	-	-
(*S*)-ADH 2%	54.9	47.8	114	15.1	90.5	40.6	12.3
(*S*)-ADH 5%	69.5	39.2	120	21.8	84.1	27.6	-
(*S*)-ADH 10%	55.3	24.0	82.9	22.1	94.3	35.1	-
(*S*)-ADH 20%	25.9	9.02	104	-	50.2	37.7	-
(*S*)-ADH 40%	4.52	73.6	169	-	59.3	-	-
(*S*)-ADH 80%	3.49	-	147	-	-	-	-

Controls

Specific enzyme activity/ nmol min^{-1} mg^{-1}	**Conditions**	**Enzyme**
24.7	AcPh **1**, 10 mM, 2% DMSO	**(*R*)-ADH**
24.6	AcPh **1**, 10 mM, 2% DMSO	
17.6	AcPh **1**, 10 mM, 2% DMSO	
80.3	AcPh **1**, 10 mM, 2% DMSO	**(*S*)-ADH**
39.1	AcPh **1**, 10 mM, 2% DMSO	
52.1	AcPh **1**, 10 mM, 2% DMSO	

	Average AcPh specific enzyme activity/ nmol min^{-1} mg^{-1}
(*R*)-ADH	23.2
stdev	3.81
(*S*)-ADH	56.6
stdev	17.2

Effect on specific enzyme activity determined by calculating the percentage change with respect to the average AcPh control values.

20% hexane in Tris HCl buffer NADH consumption screening

The general procedure for NADH consumption substrate screening assay was followed, using 20% hexane in 50 mM Tris HCl buffer as the solvent system. Each compound was screened at concentrations of 2.5, 5 and 10 mM, with the SEA value reported being the highest value obtained out of the 3 concentrations.

Results

Compound	Specific Enzyme Activity (*R*)-ADH/ nmol mg^{-1} min^{-1}	Concentration/ mM	Specific Enzyme Activity (*S*)-ADH/ nmol mg^{-1} min^{-1}	Concentration/ mM
1	54.1	10	99.1	2.5
2	22.9	5	122.1	10
3	12.2	5	142	10
4	38.0	5	39.5	10
5	73.1	10	158	5
6	-	-	31.5	2.5
7	-	-	-	-
8	14.8	5	157	2.5
9	10.0	2.5	27.2	2.5
10	-	-	83.0	2.5
11	52.9	10	64.0	10
12	-	-	89.9	2.5
13	19.9	5	196	10
14	34.2	5	189	2.5
15	42.8	10	-	-
16	103	10	242	10
17	57.7	10	366	5
18	12.4	2.5	17.4	10
19	43.9	2.5	80.4	5
20	130	5	593	10
21	17.4	5	57.3	10
22	34.0	10	163	10
23	51.0	10	137	10
24	62.5	5	92.0	2.5
25	35.5	10	60.3	10
26	46.2	10	29.4	2.5
27	89.2	2.5	337	10
28	73.9	10	14.3	5
29	87.1	2.5	112	10
30	37.0	10	54.0	2.5
31	22.2	5	239	2.5
32	95.7	2.5	364	10
33	67.2	10	62.1	10
34	-	-	36.0	5
35	52.3	10	112	10
36	25.7	2.5	37.1	10
37	85.3	2.5	23.4	2.5
38	-	-	-	-
39	-	-	15.9	-
40	-	-	-	-
41	-	-	-	-
42	-	-	-	-

43	-	-	-	-
44	-	-	-	-
45	-	-	-	-
46	-	-	-	-
47	-	-	21.7	2.5
48	-	-	-	-
49	80.9	10	210	2.5
50	33.2	2.5	269	5
51	45.2	10	50.5	10
52	53.2	10	185	10
53	13.7	5	229	10
54	-	-	-	-
55	5.62	10	20.7	10
56	14.7	2.5	20.7	2.5
57	-	-	203	2.5
58	57.4	2.5	173	2.5
59	51.3	10	117	2.5
60	465	5	87.0	10
61	13.8	2.5	397	10
62	24.6	2.5	221	2.5
63	59.8	2.5	220	10
64	59.2	10	363	10
65	-	-	-	-
66	-	-	-	-
67	-	-	-	-
68	-	-	34.6	2.5
69	10.2	2.5	62.2	10
70	123	2.5	218	10
71	82.8	10	98.2	2.5

Background NADH consumption measurements for ADH

Background consumption of NADH was generally consistent for each ADH, and was used as a control to provide confidence in the ability to compare results from different screens. Background NADH oxidation was also used as guideline threshold for treating a substrate as having any reactivity with ADH. For (*R*)-ADH (ADH 101), background SEA was usually below 10 nmol mg^{-1} min^{-1}, whereas for (*S*)-ADH (ADH 105), background SEA was generally around 20 nmol mg^{-1} min^{-1}.

GC studies to investigate *ee* and functional group retention of halogenated acetophenone reductions

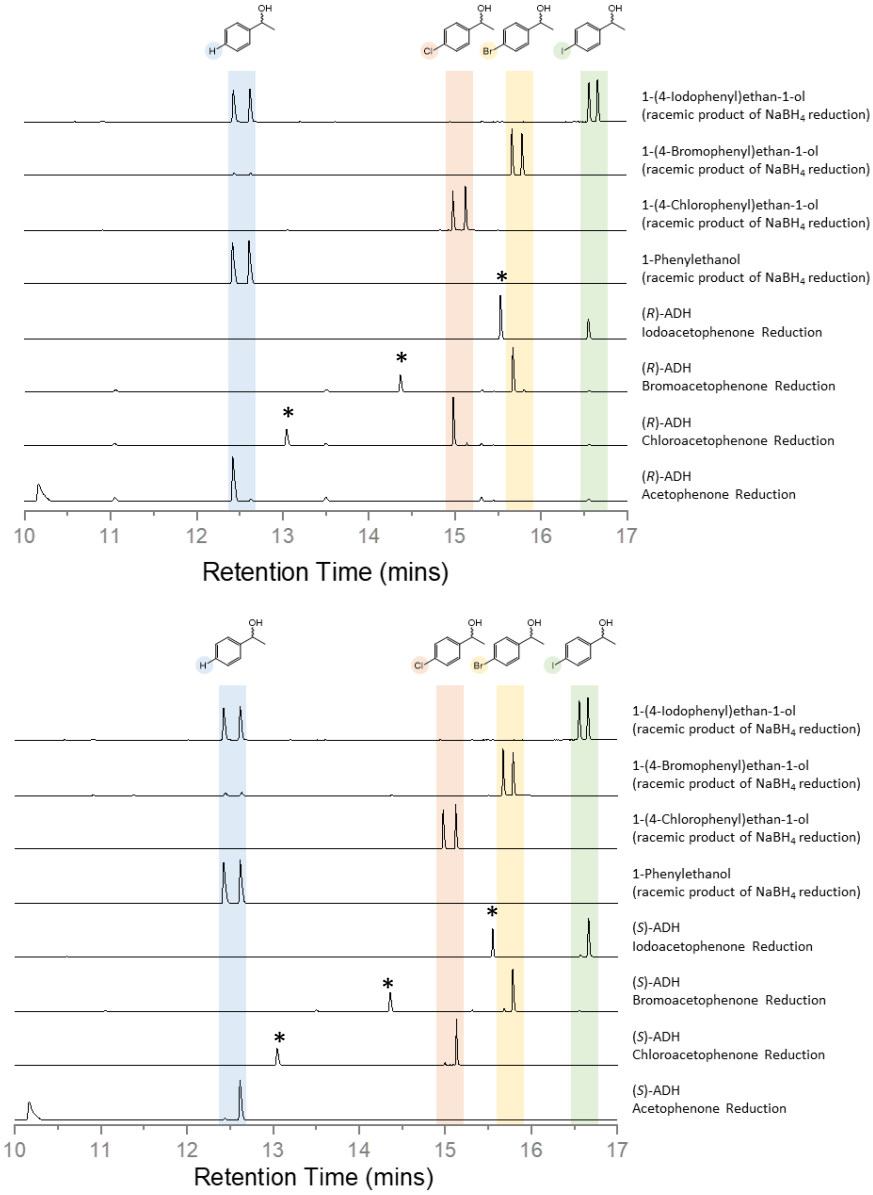
Reductions were performed using (*R*)/(*S*)-ADH on acetophenone (AcPh), 4-iodoacetophenone (IAcPh), 4-bromoacetophenone (BrAcPh) and 4-chloroacetophenone (ClAcPh) to assess whether the halogen groups in IAcPh/BrAcPh/ClAcPh were retained.

In a glove box under N_2, AcPh/ClAcPh/BrAcPh/IAcPh (20 µL, 100 mM in DMSO), NADH (40 µL, 5 mM in Tris HCl buffer), (*R*)/(*S*)-ADH (50 µL, 10 mg mL^{-1} in Tris HCl buffer) and Tris HCl Buffer (390 µL, 100 mM, pH 8.0) were added to 1 mL Eppendorf tubes and left on a shaker to react. After 2 and 28

hours, 250 µL of each reaction mixture was then extracted with ethyl acetate (750 µL, containing 2 mM undecane). The organic layer was dried with $MgSO_4$, then transferred into a GC vial for analysis.

Separately, in a glove box under N_2, AcPh/ClAcPh/BrAcPh/IAcPh (20 µL, 100 mM in DMSO), $NaBH_4$ (5 mg), and Tris HCl buffer (480 µL, 100 mM, pH 8.0) were added to 1 mL Eppendorf tubes and left on a shaker to react for 2 hours. 250 µL of each reaction mixture was then extracted with ethyl acetate (750 µL, containing 2 mM undecane). The organic layer was dried with $MgSO_4$, then transferred into a GC vial for analysis.

Comparison of reaction mixtures to the racemic product standards produced chemically by $NaBH_4$ reduction, including the standard for 1-phenylethanol (the product of reduction and dehalogenation), showed that the biocatalytic reductions did not cause dehalogenation for the IAcPh, BrAcPh and ClAcPh (**Supplementary Figure 2**). Dehalogenation was observed for the $NaBH_4$ reductions of IAcPh and BrAcPh, with considerable deiodination and a modest amount of



Supplementary Figure 2: Stacked GC chromatograms comparing reductions of AcPh, IAcPh, BrAcPh and ClAcPh by (*R*)/(*S*)-ADH after 28 hours and $NaBH_4$, *denotes unreacted substrate peak

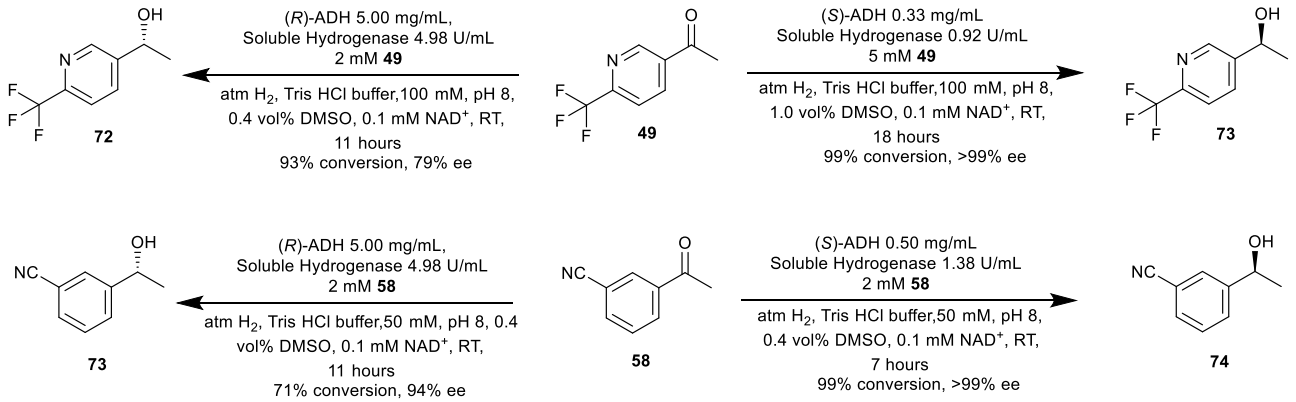
debromination observed. The *ee* values calculated for each reaction at the two time points are shown below. We observed that *ee* values generally decreased after 28 hours, which we believe to be due to instability of the chiral centres within the products in aqueous solution.

Reaction	Undecane Peak Area	Retention Time Substrate Peak/ min	Area Substrate Peak	Retention Time *R*-Product Peak/ min	Area *R*-Product Peak	Retention Time *S*-Product Peak/ min	Area *S*-Product Peak	Dehalogenation	GC Conversion	Enantiomeric Excess
AcPh + (*R*)-ADH (2h)	6.6884	10.167	1.2408	12.422	1.434	12.633[a]	0.0053	0%	54%	**99%**
AcPh + (*R*)-ADH (28h)	7.3108	10.163	1.2355	12.417	1.5381	12.63[a]	0.0537	0%	54%	**93%**
AcPh +(*S*)-ADH (2h)	3.9491	10.182	0.7368	12.443	0.0047	12.622[a]	0.8103	0%	52%	**99%**
AcPh + (*S*)-ADH (28h)	6.5514	10.165	1.1978	12.438	0.0419	12.615[a]	1.1845	0%	49%	**93%**
ClAcPh + (*R*)-ADH (2hrs)	5.3995	13.047	0.7086	14.982	1.2275	15.137[b]	0.0054	0%	63 %	**99%**
ClAcPh + (*R*)-ADH (28h)	7.2288	13.045	0.7668	14.98	1.5585	15.137[b]	0.0703	0%	65%	**91%**
ClAcPh + (*S*)-ADH (2h)	7.4903	13.043	1.1225	14.995	0.0119	15.122[b]	1.7392	0%	61%	**99%**
ClAcPh + (*S*)-ADH (28h)	6.0722	13.043	0.6968	14.993	0.0552	15.125[b]	1.0554	0%	58%	**90%**
BrAcPh + (*R*)-ADH (2h)	5.0468	14.362	0.7005	15.668[c]	1.1376	15.798	0.0034	0%	62%	**99%**
BrAcPh + (*R*)-ADH (28h)	6.6892	14.36	0.7324	15.667[c]	1.5132	15.795	0.0563	0%	66%	**93%**
BrAcPh + (*S*)-ADH (2h)	4.8686	14.362	0.7127	15.682[c]	0.0109	15.785	1.1345	0%	61%	**98%**
BrAcPh + (*S*)-ADH (28h)	9.0587	14.355	1.0842	15.68[c]	0.1124	15.78	1.8713	0%	61%	**89%**
IAcPh + (*R*)-ADH (2h)	5.9736	15.547	0.411	16.557	0.9112	16.625	0.0017	0%	69%	**>99%**
IAcPh + (*R*)-ADH (28h)	8.3244	15.532	2.6099	16.55	1.3783	16.667	0.0192	0%	34 %	**97%**
IAcPh + (*S*)-ADH (2h)	5.4465	15.537	1.8618	16.568	0.0083	16.658	0.8038	0%	30%	**98%**
IAcPh + (*S*)-ADH (28h)	5.8249	15.543	0.4145	16.563	0.0366	16.658	0.7126	0%	61%	**90%**

[a]consistent with retention time reported for commercial product standard in previous work, (*S*)-product observed at 12.6 min.[1] [b]consistent with retention time reported for commercial product standard in previous work, (*S*)-product observed at 15.1 min.[1] [c]consistent with retention time reported for commercial product standard in previous work, (*R*)-product observed at 15.7 min.[1]

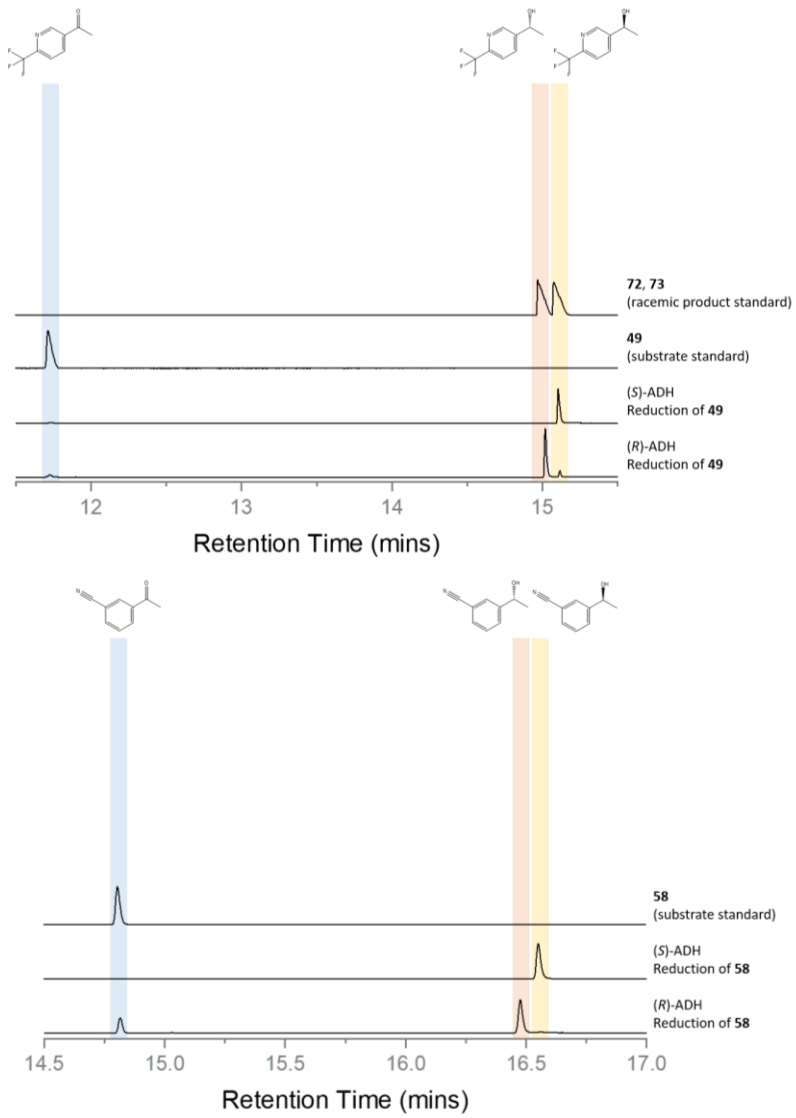
Reduction of 49 and 58 using H_2-driven cofactor recycling

Reduction of **49** and **58** was performed as described in the H_2-driven cofactor recycling batch reductions general procedure, with the reaction mixtures as detailed below in **Supplementary Figure 4**. GC (**Supplementary Figure 5**) and ^{1}H NMR (**Supplementary Figure 6**) was used to confirm reduction and calculate *ee*. We observed that the racemic product standard contained peaks roughly 0.1 min apart, with reductions of **49** giving an earlier peak with (*R*)-ADH and later peak with (*S*)-ADH, corresponding with those in the racemic product standard. This was consistent with our previous observations that reduction of acetophenone analogues results in a peak approximately 0.1 min later for the (*S*)-product and gave us confidence in our assignments.

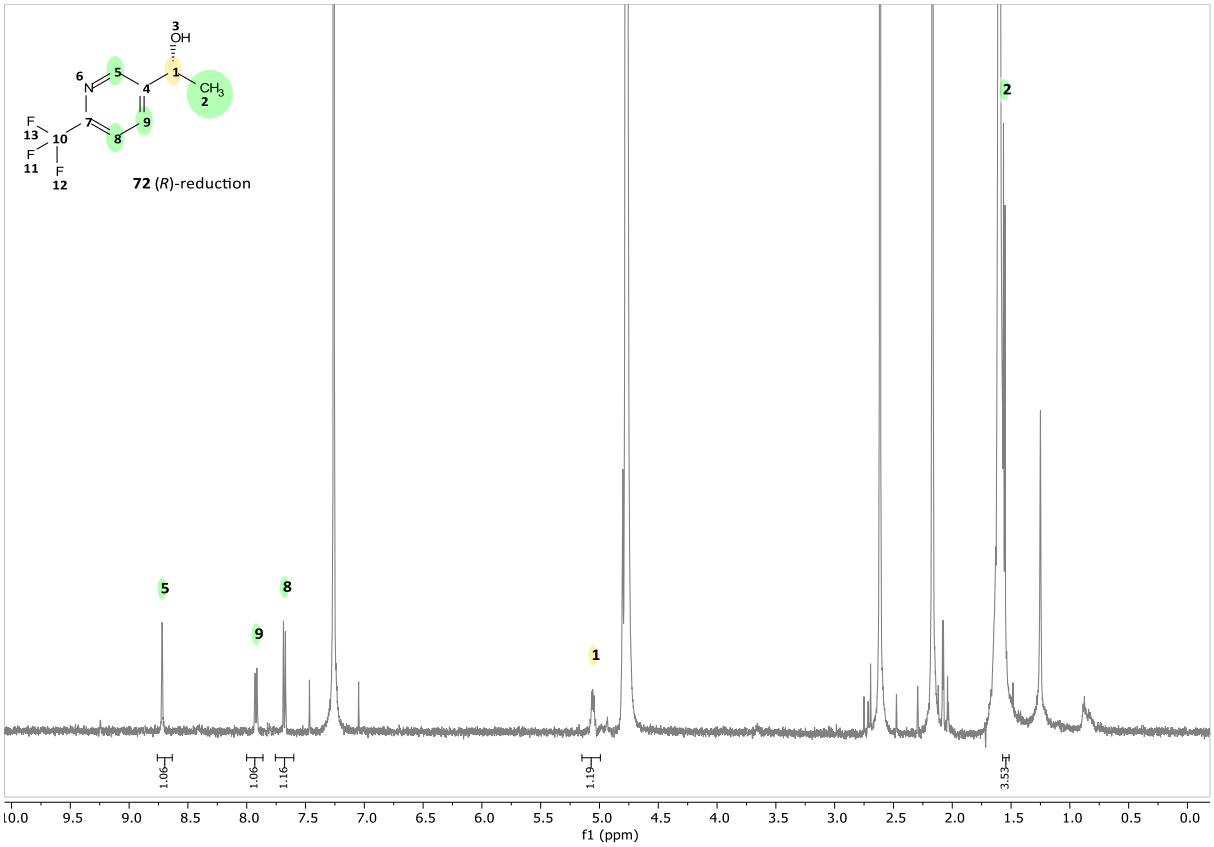


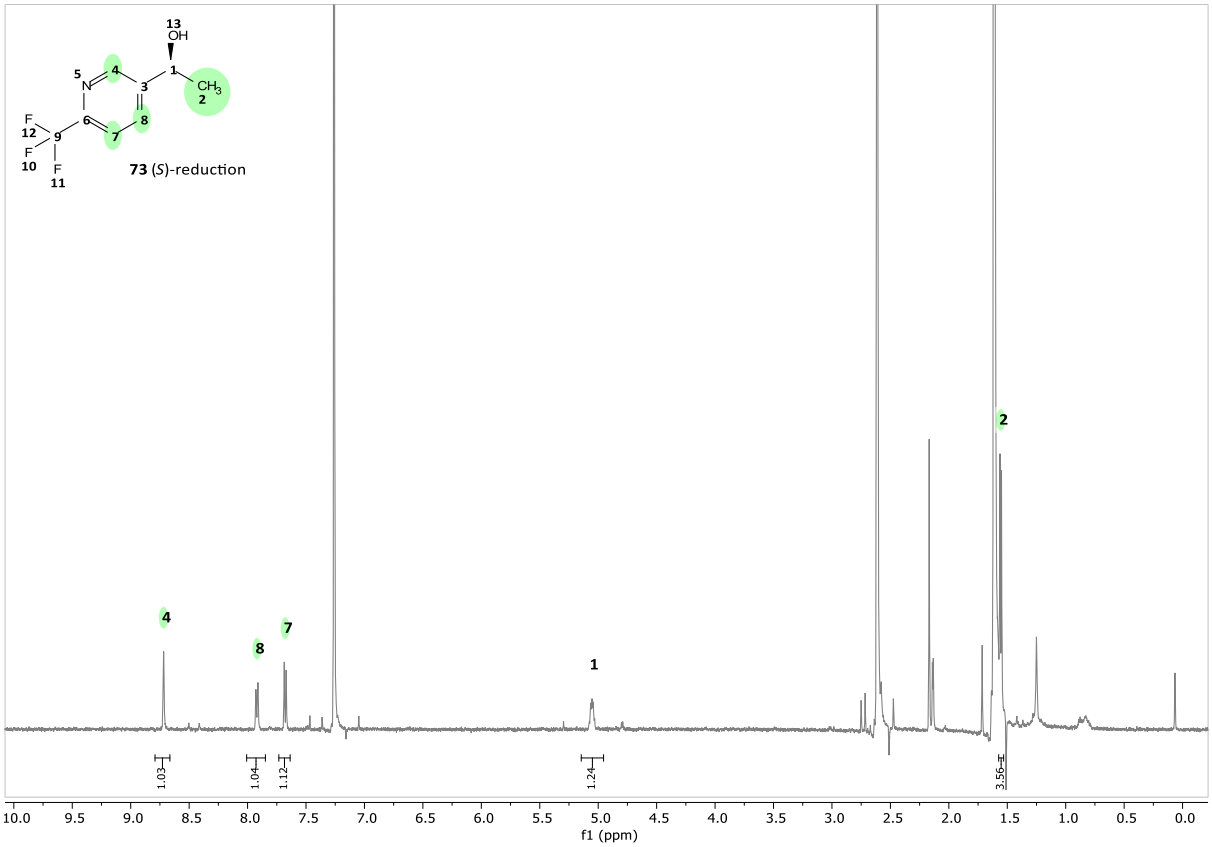
Supplementary Figure 4: Reaction compositions for the reduction of **49** and **58** using H_2-driven cofactor recycling

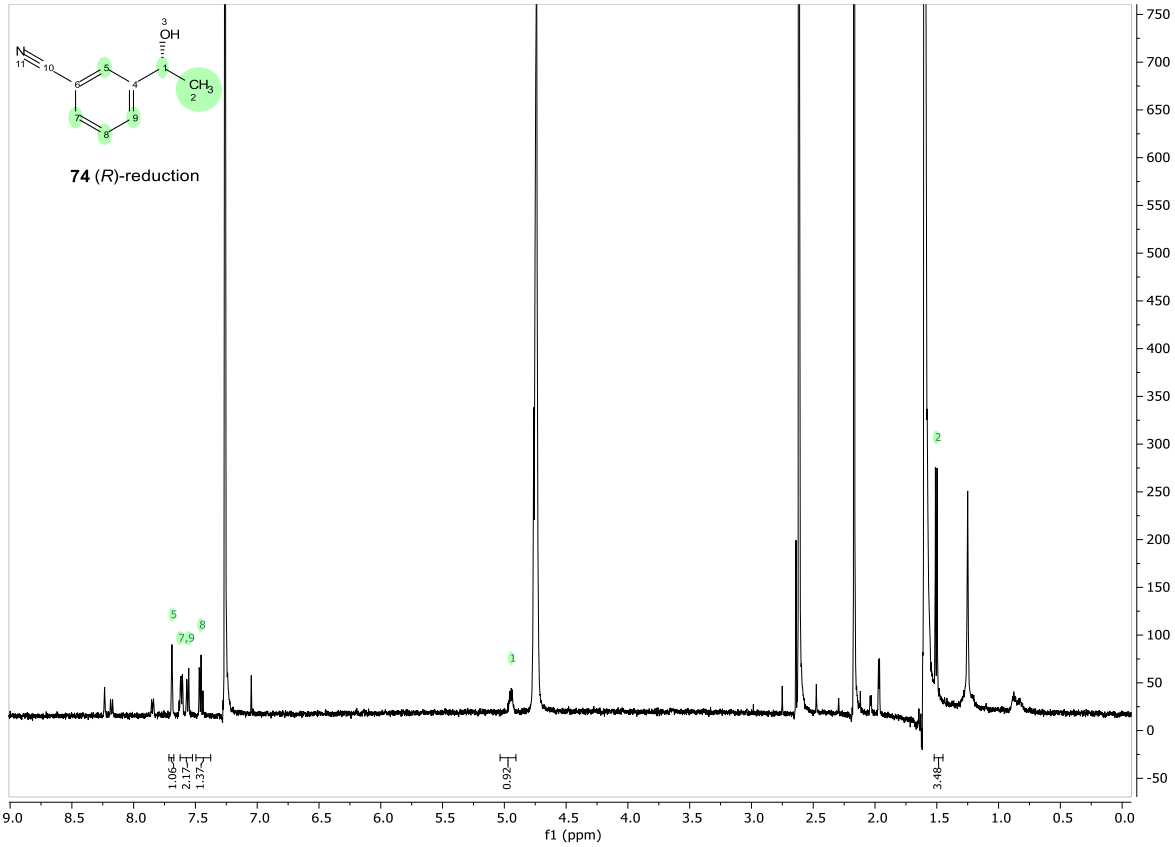
Reaction	Undecane Peak Area	Retention Time Substrate Peak/ min	Area Substrate Peak	Retention Time R-Product Peak/ min	Area R-Product Peak	Retention Time S-Product Peak/ min	Area S-Product Peak	GC Conversion	Enantiomeric Excess
49 + (*R*)-ADH (11h)	6.9175	11.725	0.0812	15.018	1.0102	15.116	0.1104	93%	**80%**
49 + (*S*)-ADH (18h)	5.8377	11.728	0.0053	–	–	15.103	1.2171	99%	**>99%**
58 + (*R*)-ADH (11h)	5.9764	14.815	0.3039	16.475	0.7227	16.562	0.0205	71%	**94%**
58 + (*S*)-ADH (7h)	6.2662	14.822	0.0040	–	–	16.552	0.9564	99%	**>99%**

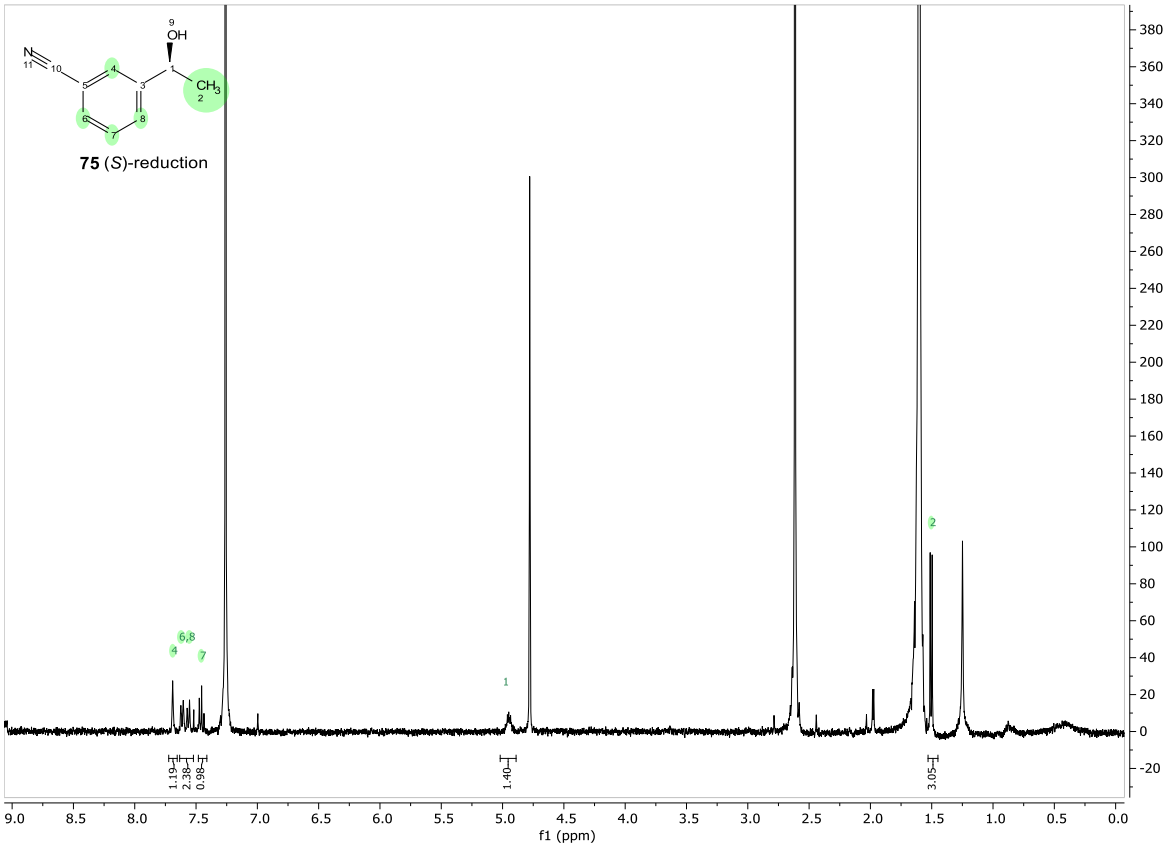


Supplementary Figure 5: Chiral GC FID data for the reduction of **49** and **58**


72 (R)-reduction
f1 (ppm)


73 (S)-reduction
f1 (ppm)





Supplementary Figure 6: ^{1}H NMR spectra for the reduction of **49** and **58**

NADPH consumption screening for new ADHs

This screening followed the same protocol as for Tris HCl buffer NADH consumption screening, using ADHs 19, 20, 61 and 150, NADPH as the cofactor, and substrate concentrations of 10 mM and 2.5 mM. The reported SEA value is the higher value obtained out of the two concentrations.

Compound	**Specific Enzyme Activity ADH19/ nmol mg^{-1} min^{-1}**	**Concentration/ mM**	**Specific Enzyme Activity ADH20/ nmol mg^{-1} min^{-1}**	**Concentration/ mM**	**Specific Enzyme Activity ADH61/ nmol mg^{-1} min^{-1}**	**Concentration/ mM**	**Specific Enzyme Activity ADH150/ nmol mg^{-1} min^{-1}**	**Concentration/ mM**
69	10.4	10	18.8	2.5	29.0	10	6.55	10
18	6.78	2.5	67.2	2.5	-	-	38.6	2.5
68	7.76	2.5	55.7	10	123	10	33.7	2.5
34	14.3	2.5	31.3	2.5	56.4	2.5	7.83	2.5
28	-	-	41.1	2.5	-	-	69.9	2.5
55	12.8	2.5	-	-	12.7	2.5	-	-
6	45.6	10	-	-	50.8	10	-	-

Screening tool development

Rationale

With the sequence of the commercial enzymes used within this study not publically available, our approach was to build a screening tool which could predict substrate reactivity naïve of the enzyme active site. To do this we applied Cresset's Forge software, which was originally developed as a drug discovery solution to allow for pharmacological modelling (modelling bioactivity naïve of the structure of the enzyme being targeted), based on the nature of the active molecules themselves in terms of electronic fields and molecule alignments. Our rationale was that the principles to model the reactivity of a substrate with a given enzyme should be analogous to those governing bioactivity of a ligand in a biologically relevant enzyme. By modelling NADH consumption activity values and conformational alignments of the substrates, we hoped to generate a model of reactivity analogous to a Quantitative Structure Activity Relationship (QSAR) and use this to generate a predictive screening tool to indicate the likely reactivity of a substrate with a specific ADH. We envisioned this would greatly help the synthetic chemist evaluate the potential utility of the ADH enzyme in question for their intended reduction.

Structures and compound identifiers

Substrates were imported in Forge as .mol files, with their lowest energy conformations generated in Chem3D. The identifiers as used in Forge for each compound were as follows:

Compound	**Molecule title in Forge**
1	Substrate 31:1
2	Substrate 3:1
3	Substrate 4:1
4	Substrate 39:1
5	Substrate 25:1
6	Substrate 2:1
7	Substrate 34:1
8	Substrate 1:1
9	Substrate 8:1
10	Substrate 10:1
11	Substrate 15:1

Compound	Molecule title in Forge
12	Substrate 16:1
13	Substrate 12:1
14	Substrate 11:1
15	Substrate 5:1
16	Substrate 6:1
17	Substrate 7:1
18	Substrate 33:1
19	Substrate 9:1
20	Substrate 13:1
21	Substrate 23:1
22	Substrate 17:1
23	Substrate 19:1
24	Substrate 32:1
25	Substrate 36:1
25	Substrate 35:1
26	Substrate 22:1
27	Substrate 28:1
28	Substrate 29:1
29	Substrate 24:1
30	Substrate 37:1
31	Substrate 26:1
32	Substrate 27:1
33	Substrate G:1
34	Substrate 14:1
36	Substrate 30:1
37	Substrate F:1
38	Substrate C:1
40	Substrate D:1
45	Substrate AC:1
47	Substrate R:1
48	Substrate P:1
49	Substrate L:1
50	Substrate AH:1
51	Substrate 18:1
52	Substrate 20:1
53	Substrate J:1
54	Substrate 21:1
55	Substrate S:1
56	Substrate I:1
57	Substrate M:1
58	Substrate N:1
59	Substrate AJ:1
60	Substrate AG:1
61	Substrate O:1
62	Substrate Z:1
63	Substrate AA:1
64	Substrate AB:1
66	Substrate AD:1
67	Substrate AI:1
68	Substrate Y:1
69	Substrate K:1
70	Substrate AF:1
71	Substrate AE:1

Imported activity values

Specific Enzyme Activity values for NADH consumption were imported into Forge as a .csv file. Inactive substrates were given the SEA value 0 mUnits.

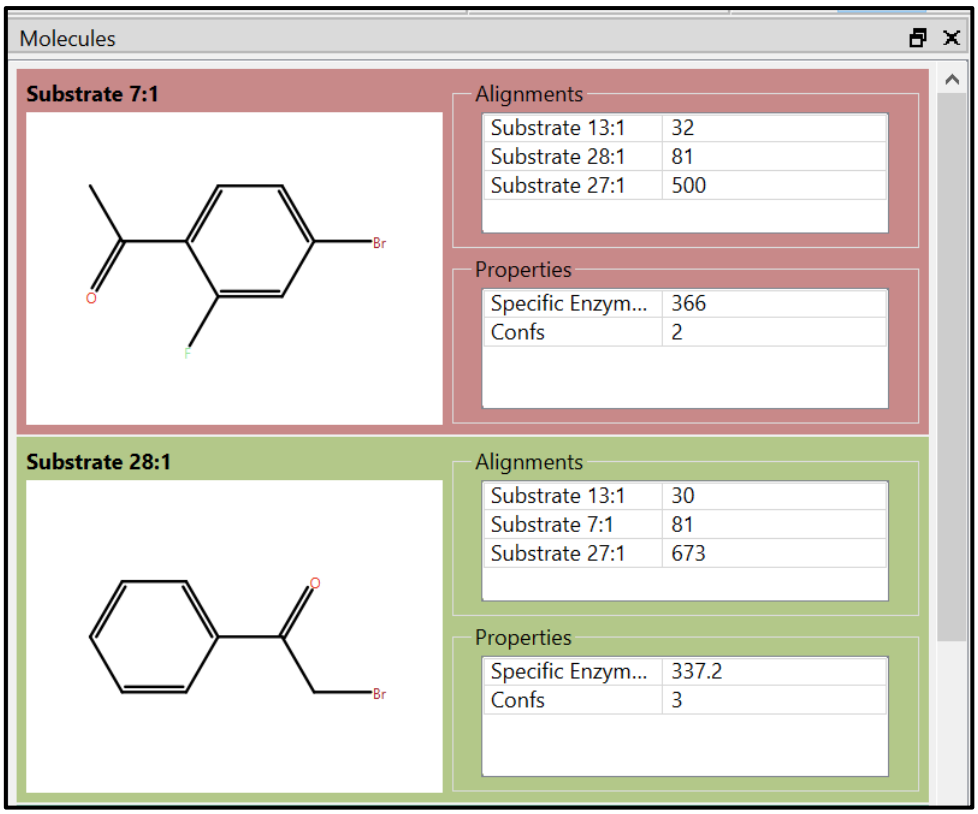
Substrate	Specific Enzyme Activity (*S*)-ADH/ mUnits
acetophenone:1	99.1
Substrate 2:1	31.5
Substrate 3:1	122.1
Substrate 4:1	141.6
Substrate 8:1	27.2
Substrate 13:1	593.1
Substrate 14:1	36
Substrate 16:1	89.9
Substrate 17:1	162.6
Substrate 18:1	50.5
Substrate 19:1	136.6
Substrate 20:1	185.1
Substrate 21:1	0
Substrate 22:1	29.4
Substrate 23:1	57.3
Substrate 24:1	111.7
Substrate 26:1	239.4
Substrate 27:1	364
Substrate 28:1	337.2
Substrate 29:1	14.3
Substrate 30:1	37.1
Substrate 31:1	99.1
Substrate 32:1	92
Substrate 33:1	17.4
Substrate 35:1	60.3
Substrate 36:1	111.6
Substrate 37:1	54
Substrate 38:1	0
Substrate 39:1	39.5
Substrate C:1	0
Substrate D:1	0
Substrate F:1	23.4
Substrate G:1	62.1
Substrate I:1	20.7
Substrate J:1	229.2
Substrate K:1	62.2
Substrate L:1	210.2
Substrate M:1	203.3
Substrate N:1	173.4
Substrate O:1	396.7
Substrate R:1	21.7
Substrate S:1	20.7
Substrate P:1	0
Substrate 9:1	80.4

Substrate 10:1	83
Substrate 11:1	188.5
Substrate 12:1	195.8
Substrate 15:1	64
Substrate 25:1	158.2
Substrate 34:1	0
Substrate 1:1	156.6
Substrate 5:1	0
Substrate 6:1	242.4
Substrate 7:1	366
Substrate Y:1	34.6
Substrate Z:1	222.1
Substrate AA:1	220
Substrate AB:1	362.6
Substrate AC:1	0
Substrate AD:1	0
Substrate AE:1	98.2
Substrate AF:1	218.3
Substrate AG:1	87
Substrate AH:1	268.7
Substrate AI:1	0
Substrate AJ:1	116.9

Reference generation using FieldTemplater

Method

Molecules selected for generating a pharmacophore-based reference in FieldTemplater were selected from those displaying a high specific enzyme activity, but being as structurally different as possible.



Log for molecule 'Substrate 7:1'
==========

Molecule 'Substrate 7:1' read from file 'Q:/Kate/Modelling/Training set hexane tris water/Substrate 7.mol'

New conformation hunt process started at Sun Mar 29 14:07:06 2020

Conformation hunt settings for molecule 'Substrate 7:1'
==========
Process config '[Custom]'
Acyclic secondary amide handling: Force trans
Add field points to conformations: true
Bypass RMS filter on rotatable bond twist of: 90
Dielectric: 2
Filter duplicate conformers at RMS: 0.5
Generate conformational enantiomers: true
Gradient cutoff for conformer minimization: 0.1
Keep conformations within an energy window of: 3
Maximum no. of chiral centres to enumerate: 4
Maximum number of conformations: 200
No. of high-T dynamics runs for flexible rings: 5
Number of bond randomisations: 300
Perform the conformation hunt: true
Perform the conformation hunt using an external tool: false
Process proto-confs in random order: true
Remove conformations containing boat and twistboat rings: false
Turn off Coulombic and attractive vdW forces: true
Use ring conformation library: true

A total of 2 confs were generated.
Conformation energies:
12.23 13.05
Aligning to molecule Substrate 13:1 using settings 'Normal'.
Aligning to molecule Substrate 28:1 using settings 'Normal'.
Aligning to molecule Substrate 27:1 using settings 'Normal'.

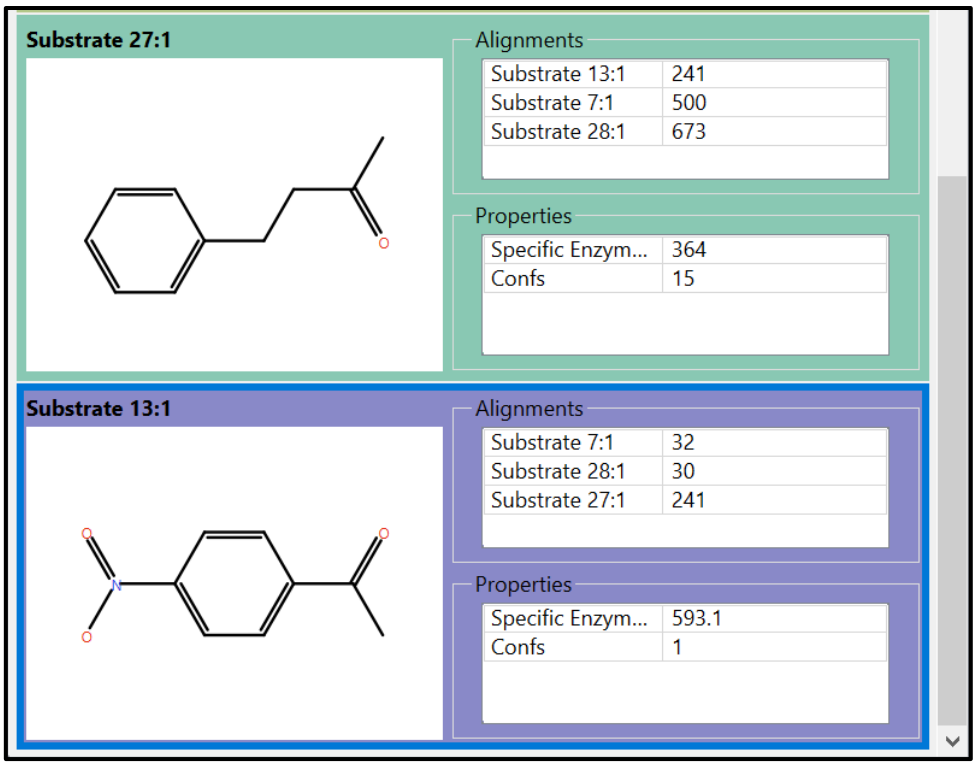
Log for molecule 'Substrate 28:1'
==========

Molecule 'Substrate 28:1' read from file 'Q:/Kate/Modelling/Training set hexane tris water/Substrate 28.mol'

New conformation hunt process started at Sun Mar 29 14:07:06 2020

Conformation hunt settings for molecule 'Substrate 28:1'
==========
Process config '[Custom]'
Acyclic secondary amide handling: Force trans
Add field points to conformations: true
Bypass RMS filter on rotatable bond twist of: 90
Dielectric: 2
Filter duplicate conformers at RMS: 0.5
Generate conformational enantiomers: true
Gradient cutoff for conformer minimization: 0.1
Keep conformations within an energy window of: 3
Maximum no. of chiral centres to enumerate: 4
Maximum number of conformations: 200
No. of high-T dynamics runs for flexible rings: 5
Number of bond randomisations: 300
Perform the conformation hunt: true
Perform the conformation hunt using an external tool: false
Process proto-confs in random order: true
Remove conformations containing boat and twistboat rings: false
Turn off Coulombic and attractive vdW forces: true
Use ring conformation library: true

A total of 3 confs were generated.
Conformation energies:
11.43 11.43 12.19
Aligning to molecule Substrate 7:1 using settings 'Normal'.
Aligning to molecule Substrate 13:1 using settings 'Normal'.
Aligning to molecule Substrate 27:1 using settings 'Normal'.



Log for molecule 'Substrate 27:1'
==========

Molecule 'Substrate 27:1' read from file 'Q:/Kate/Modelling/Training set hexane tris water/Substrate 27.mol'

New conformation hunt process started at Sun Mar 29 14:07:06 2020

Conformation hunt settings for molecule 'Substrate 27:1'
==========
Process config '[Custom]'
Acyclic secondary amide handling: Force trans
Add field points to conformations: true
Bypass RMS filter on rotatable bond twist of: 90
Dielectric: 2
Filter duplicate conformers at RMS: 0.5
Generate conformational enantiomers: true
Gradient cutoff for conformer minimization: 0.1
Keep conformations within an energy window of: 3
Maximum no. of chiral centres to enumerate: 4
Maximum number of conformations: 200
No. of high-T dynamics runs for flexible rings: 5
Number of bond randomisations: 300
Perform the conformation hunt: true
Perform the conformation hunt using an external tool: false
Process proto-confs in random order: true
Remove conformations containing boat and twistboat rings: false
Turn off Coulombic and attractive vdW forces: true
Use ring conformation library: true

A total of 15 confs were generated.
Conformation energies:
10.60 10.89 10.89 11.09 11.09 11.86 11.86 12.62
12.62 12.81 12.81 13.01 13.01 13.73 13.73
Aligning to molecule Substrate 7:1 using settings 'Normal'.
Aligning to molecule Substrate 28:1 using settings 'Normal'.
Aligning to molecule Substrate 13:1 using settings 'Normal'.
Log for molecule 'Substrate 13:1'
==========

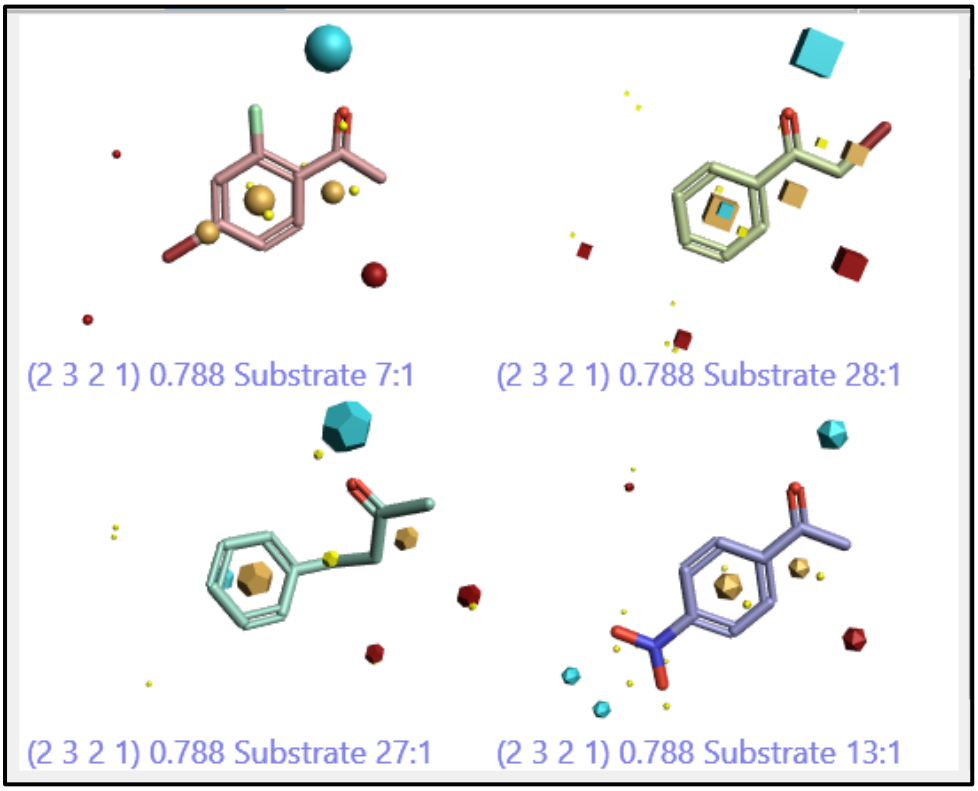
Molecule 'Substrate 13:1' read from file 'Q:/Kate/Modelling/Training set hexane tris water/Substrate 13.mol'

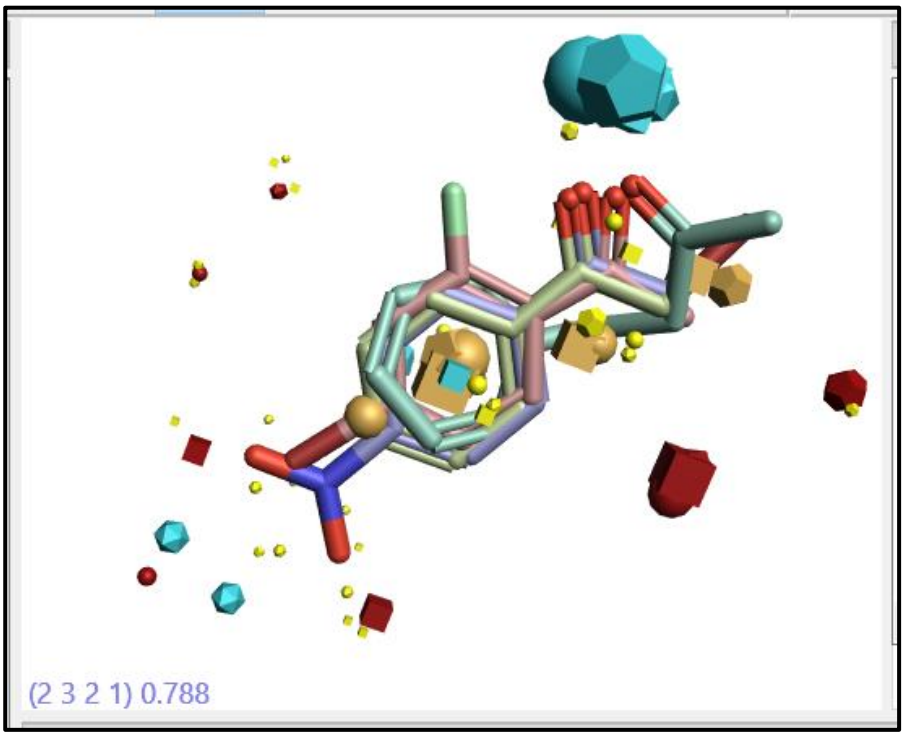
New conformation hunt process started at Sun Mar 29 14:07:06 2020

Conformation hunt settings for molecule 'Substrate 13:1'
==========
Process config '[Custom]'
Acyclic secondary amide handling: Force trans
Add field points to conformations: true
Bypass RMS filter on rotatable bond twist of: 90
Dielectric: 2
Filter duplicate conformers at RMS: 0.5
Generate conformational enantiomers: true
Gradient cutoff for conformer minimization: 0.1
Keep conformations within an energy window of: 3
Maximum no. of chiral centres to enumerate: 4
Maximum number of conformations: 200
No. of high-T dynamics runs for flexible rings: 5
Number of bond randomisations: 300
Perform the conformation hunt: true
Perform the conformation hunt using an external tool: false
Process proto-confs in random order: true
Remove conformations containing boat and twistboat rings: false
Turn off Coulombic and attractive vdW forces: true
Use ring conformation library: true

A total of 1 confs were generated.
Conformation energies:
31.67
Aligning to molecule Substrate 7:1 using settings 'Normal'.
Aligning to molecule Substrate 28:1 using settings 'Normal'.
Aligning to molecule Substrate 27:1 using settings 'Normal'.

Results


(2 3 2 1) 0.788 Substrate 7:1
(2 3 2 1) 0.788 Substrate 28:1
(2 3 2 1) 0.788 Substrate 27:1
(2 3 2 1) 0.788 Substrate 13:1


(2 3 2 1) 0.788

Template	Isomer Substrate 7:1	Conf Substrate 7:1	Isomer Substrate 28:1	Conf Substrate 28:1	Isomer Substrate 27:1	Conf Substrate 27:1	Isomer Substrate 13:1	Conf Substrate 13:1	Similarity	Field Similarity	Raw Field Score	Penalised Field Score	Shape Similarity	Raw Shape Score	Penalised Shape Score	Total Penalty	Excluded Volume Clash Penalty	Atom Distance Constraint Penalty	Field Value Penalty	Template Pairwise RMS Similarity	Template RMS Angle Deviation	Template RMS Distance Deviation	Template Density
1	1	2	1	3	1	2	1	1	0.787702	0.766795	-40.4434	-40.4434	0.808609	83.4774	83.4774	0	0	0	0	0.791029	3.73037	0.114744	1
2	1	2	1	3	1	4	1	1	0.78606	0.757551	-42.9015	-42.9015	0.814569	84.0913	84.0913	0	0	0	0	0.789876	3.1746	0.12006	1
3	1	2	1	3	1	3	1	1	0.786006	0.762346	-40.1652	-40.1652	0.809667	83.5876	83.5876	0	0	0	0	0.788356	3.73055	0.119032	1
4	1	2	1	3	1	5	1	1	0.785679	0.757976	-42.7453	-42.7453	0.813381	83.9769	83.9769	0	0	0	0	0.788664	3.87503	0.134596	1
5	1	2	1	1	1	2	1	1	0.780322	0.753852	-41.3972	-41.3972	0.806791	83.2941	83.2941	0	0	0	0	0.780725	4.58577	0.105556	1
6	1	2	1	2	1	3	1	1	0.779181	0.749347	-40.5138	-40.5138	0.809015	83.5246	83.5246	0	0	0	0	0.776552	2.4159	0.063249	1
7	1	2	1	2	1	4	1	1	0.777653	0.74665	-43.4127	-43.4127	0.808656	83.4912	83.4912	0	0	0	0	0.778473	3.34259	0.100822	1
8	1	2	1	1	1	5	1	1	0.776007	0.741364	-43.5073	-43.5073	0.810651	83.6941	83.6941	0	0	0	0	0.779106	4.37173	0.148756	1
9	1	2	1	2	1	2	1	1	0.774727	0.743792	-40.232	-40.232	0.805663	83.1761	83.1761	0	0	0	0	0.77611	8.24494	0.217227	1
10	1	1	1	3	1	3	1	1	0.77203	0.729639	-38.7034	-38.7034	0.814419	84.0931	84.0931	0	0	0	0	0.77582	2.78238	0.162524	1
11	1	2	1	1	1	3	1	1	0.771528	0.737213	-40.3883	-40.3883	0.805843	83.1873	83.1873	0	0	0	0	0.774369	6.80866	0.180823	1
12	1	1	1	3	1	2	1	1	0.770589	0.723996	-38.4557	-38.4557	0.817182	84.3728	84.3728	0	0	0	0	0.771814	3.33889	0.152389	1
13	1	1	1	3	1	4	1	1	0.770503	0.71951	-41.0275	-41.0275	0.821496	84.8258	84.8258	0	0	0	0	0.773191	2.99743	0.171653	1
14	1	1	1	3	1	5	1	1	0.769859	0.719744	-40.863	-40.863	0.819975	84.6683	84.6683	0	0	0	0	0.773084	1.76724	0.104139	1
15	1	2	1	1	1	4	1	1	0.768695	0.725822	-42.6498	-42.6498	0.811568	83.7922	83.7922	0	0	0	0	0.773015	6.89815	0.139252	1
16	1	2	1	2	1	5	1	1	0.765098	0.721484	-41.6589	-41.6589	0.808712	83.5033	83.5033	0	0	0	0	0.769832	6.86491	0.164416	1
17	1	2	1	3	1	8	1	1	0.756399	0.702673	-40.8484	-40.8484	0.810125	83.6196	83.6196	0	0	0	0	0.786014	4.57015	0.433548	1
18	1	2	1	2	1	7	1	1	0.755152	0.678904	-39.3827	-39.3827	0.8314	85.8236	85.8236	0	0	0	0	0.756513	2.72599	0.066852	1
19	1	2	1	3	1	9	1	1	0.753922	0.694329	-39.9285	-39.9285	0.813515	83.9769	83.9769	0	0	0	0	0.781246	8.67767	0.448992	1

Template	Isomer Substrate 7:1	Conf Substrate 7:1	Isomer Substrate 28:1	Conf Substrate 28:1	Isomer Substrate 27:1	Conf Substrate 27:1	Isomer Substrate 13:1	Conf Substrate 13:1	Similarity	Field Similarity	Raw Field Score	Penalised Field Score	Shape Similarity	Raw Shape Score	Penalised Shape Score	Total Penalty	Excluded Volume Clash Penalty	Atom Distance Constraint Penalty	Field Value Penalty	Template Pairwise RMS Similarity	Template RMS Angle Deviation	Template RMS Distance Deviation	Template Density
20	1	2	1	3	1	7	1	1	0.747885	0.682539	-38.5503	-38.5503	0.813232	83.9705	83.9705	0	0	0	0	0.693111	4.04981	0.100744	0.833333
21	1	2	1	3	1	2	1	1	0.746961	0.747381	-39.3684	-39.3684	0.746542	77.1443	77.1443	0	0	0	0	0.700092	5.88437	0.194258	0.833333
22	1	2	1	1	1	6	1	1	0.746755	0.661885	-39.1366	-39.1366	0.831625	85.8501	85.8501	0	0	0	0	0.649287	8.86666	0.230698	0.666667
23	1	2	1	2	1	8	1	1	0.746525	0.696335	-41.4304	-41.4304	0.796714	82.2888	82.2888	0	0	0	0	0.697849	5.09696	0.397956	0.833333
24	1	2	1	1	1	9	1	1	0.745788	0.688252	-41.0681	-41.0681	0.803323	82.9412	82.9412	0	0	0	0	0.694372	5.3758	0.319415	0.833333
25	1	2	1	3	1	6	1	1	0.744789	0.676742	-38.3892	-38.3892	0.812836	83.9296	83.9296	0	0	0	0	0.647019	5.08475	0.216013	0.666667
26	1	2	1	3	1	4	1	1	0.743381	0.73322	-41.4942	-41.4942	0.753543	77.8667	77.8667	0	0	0	0	0.696937	5.47889	0.07335	0.833333
27	1	1	1	3	1	8	1	1	0.743294	0.666529	-38.943	-38.943	0.82006	84.6798	84.6798	0	0	0	0	0.77688	4.01849	0.533438	1
28	1	1	1	3	1	9	1	1	0.74294	0.657755	-38.0789	-38.0789	0.828125	85.5042	85.5042	0	0	0	0	0.768678	5.50698	0.47804	1
29	1	2	1	3	1	3	1	1	0.742808	0.738871	-38.8697	-38.8697	0.746744	77.1664	77.1664	0	0	0	0	0.696643	8.46004	0.182252	0.833333
30	1	2	1	3	1	1	1	1	0.741641	0.670192	-39.5989	-39.5989	0.813091	83.9271	83.9271	0	0	0	0	0.649762	8.77954	0.185787	0.666667
31	1	2	1	2	1	9	1	1	0.737902	0.676175	-39.7979	-39.7979	0.799628	82.5662	82.5662	0	0	0	0	0.693071	5.64458	0.325691	0.833333
32	1	2	1	1	1	8	1	1	0.737885	0.692042	-41.7454	-41.7454	0.783727	80.9405	80.9405	0	0	0	0	0.699141	5.08048	0.413521	0.833333

Template	Isomer Substrate 7:1	Conf Substrate 7:1	Isomer Substrate 28:1	Conf Substrate 28:1	Isomer Substrate 27:1	Conf Substrate 27:1	Isomer Substrate 13:1	Conf Substrate 13:1	Similarity	Field Similarity	Raw Field Score	Penalised Field Score	Shape Similarity	Raw Shape Score	Penalised Shape Score	Total Penalty	Excluded Volume Clash Penalty	Atom Distance Constraint Penalty	Field Value Penalty	Template Pairwise RMS Similarity	Template RMS Angle Deviation	Template RMS Distance Deviation	Template Density
33	1	1	1	3	1	9	1	1	0.737381	0.665077	-38.4272	-38.4272	0.809684	83.6202	83.6202	0	0	0	0	0.672376	6.65704	0.147939	0.833333
34	1	2	1	2	1	11	1	1	0.733532	0.66379	-38.4331	-38.4331	0.803275	82.9384	82.9384	0	0	0	0	0.632936	8.19721	0.170567	0.666667
35	1	2	1	1	1	2	1	1	0.733498	0.729874	-39.9509	-39.9509	0.737122	76.1955	76.1955	0	0	0	0	0.69178	6.00942	0.214425	0.833333
36	1	2	1	2	1	3	1	1	0.732757	0.729449	-39.34	-39.34	0.736065	76.0805	76.0805	0	0	0	0	0.686861	7.31339	0.169168	0.833333
37	1	2	1	3	1	8	1	1	0.732734	0.667783	-38.5424	-38.5424	0.797684	82.389	82.389	0	0	0	0	0.692073	6.4797	0.222892	0.833333
38	1	2	1	2	1	4	1	1	0.731789	0.717718	-41.6909	-41.6909	0.745861	77.0932	77.0932	0	0	0	0	0.687587	5.96852	0.084012	0.833333
39	1	1	1	3	1	1	1	1	0.731032	0.64852	-38.5293	-38.5293	0.813545	83.9953	83.9953	0	0	0	0	0.753472	7.52728	0.532886	1
40	1	2	1	2	1	1	1	1	0.73047	0.661907	-39.9814	-39.9814	0.799033	82.5242	82.5242	0	0	0	0	0.689315	3.4342	0.091151	0.833333
41	1	2	1	3	1	9	1	1	0.730422	0.649412	-37.453	-37.453	0.811433	83.7786	83.7786	0	0	0	0	0.707731	9.05332	0.455144	0.833333
42	1	2	1	3	1	14	1	1	0.729906	0.673397	-37.9049	-37.9049	0.786414	81.2036	81.2036	0	0	0	0	0.756762	8.95202	0.323136	1
43	1	2	1	1	1	7	1	1	0.729634	0.660579	-38.7287	-38.7287	0.798689	82.4759	82.4759	0	0	0	0	0.682673	3.56574	0.047918	0.833333
44	1	2	1	1	1	1	1	1	0.72943	0.656628	-40.1452	-40.1452	0.802232	82.8534	82.8534	0	0	0	0	0.690623	3.45638	0.082052	0.833333
45	1	2	1	3	1	8	1	1	0.728399	0.650016	-37.8897	-37.8897	0.806781	83.2882	83.2882	0	0	0	0	0.711694	4.23746	0.419736	0.833333

Template	Isomer Substrate 7:1	Conf Substrate 7:1	Isomer Substrate 28:1	Conf Substrate 28:1	Isomer Substrate 27:1	Conf Substrate 27:1	Isomer Substrate 13:1	Conf Substrate 13:1	Similarity	Field Similarity	Raw Field Score	Penalised Field Score	Shape Similarity	Raw Shape Score	Penalised Shape Score	Total Penalty	Excluded Volume Clash Penalty	Atom Distance Constraint Penalty	Field Value Penalty	Template Pairwise RMS Similarity	Template RMS Angle Deviation	Template RMS Distance Deviation	Template Density
46	1	1	1	3	1	6	1	1	0.727572	0.62936	-35.9123	-35.9123	0.825783	85.2763	85.2763	0	0	0	0	0.639907	4.77606	0.247522	0.666667
47	1	2	1	2	1	2	1	1	0.727027	0.715707	-38.607	-38.607	0.738348	76.3173	76.3173	0	0	0	0	0.686795	8.34881	0.253294	0.833333
48	1	1	1	3	1	4	1	1	0.726717	0.694543	-39.5651	-39.5651	0.75889	78.4312	78.4312	0	0	0	0	0.685168	8.99329	0.227211	0.833333
49	1	2	1	3	1	1	1	1	0.726683	0.658583	-38.751	-38.751	0.794784	82.0802	82.0802	0	0	0	0	0.763811	4.12981	0.477363	1
50	1	1	1	3	1	14	1	1	0.725303	0.657983	-37.2999	-37.2999	0.792622	81.8706	81.8706	0	0	0	0	0.745657	6.80884	0.404889	1
51	1	2	1	1	1	3	1	1	0.725255	0.709798	-38.7825	-38.7825	0.740711	76.5574	76.5574	0	0	0	0	0.684163	8.41145	0.236024	0.833333
52	1	2	1	1	1	14	1	1	0.724419	0.682103	-39.8898	-39.8898	0.766736	79.2101	79.2101	0	0	0	0	0.683373	7.54781	0.149045	0.833333
53	1	1	1	3	1	8	1	1	0.722284	0.64315	-37.3771	-37.3771	0.801419	82.7863	82.7863	0	0	0	0	0.687283	4.03096	0.157147	0.833333
54	1	2	1	1	1	4	1	1	0.72216	0.702348	-41.2095	-41.2095	0.741971	76.6927	76.6927	0	0	0	0	0.681172	7.46824	0.134101	0.833333
55	1	2	1	1	1	10	1	1	0.720315	0.652627	-38.4087	-38.4087	0.788003	81.3692	81.3692	0	0	0	0	0.637382	9.16799	0.235013	0.666667
56	1	2	1	1	1	13	1	1	0.719628	0.646418	-38.3293	-38.3293	0.792837	81.8758	81.8758	0	0	0	0	0.63771	9.94777	0.273477	0.666667
57	1	2	1	2	1	14	1	1	0.718878	0.67708	-39.0814	-39.0814	0.760677	78.5786	78.5786	0	0	0	0	0.682051	8.0946	0.16225	0.833333
58	1	2	1	3	1	8	1	1	0.716962	0.658441	-38.4283	-38.4283	0.775483	80.0912	80.0912	0	0	0	0	0.682188	8.11201	0.516459	0.833333

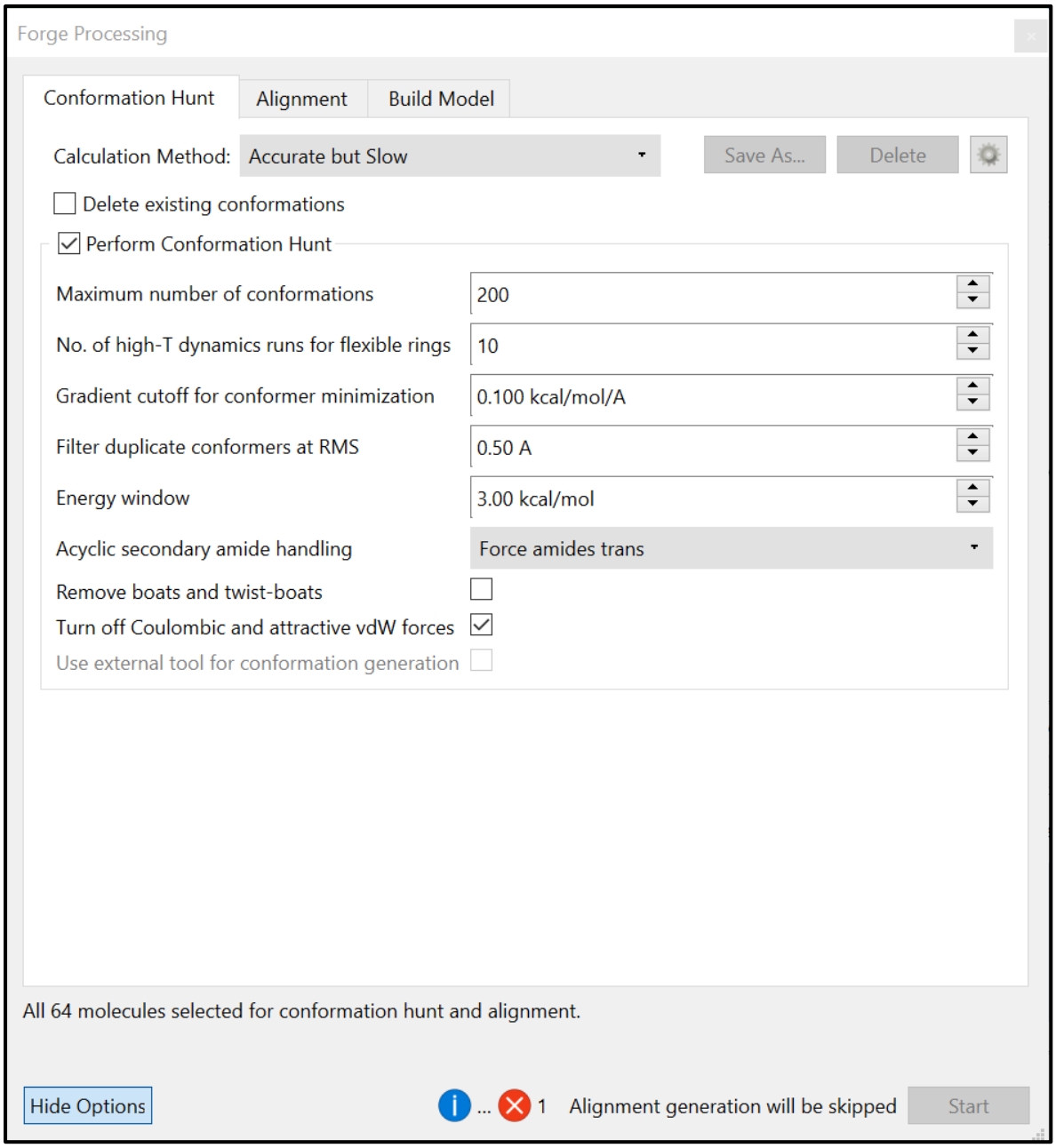
Template	Isomer Substrate 7:1	Conf Substrate 7:1	Isomer Substrate 28:1	Conf Substrate 28:1	Isomer Substrate 27:1	Conf Substrate 27:1	Isomer Substrate 13:1	Conf Substrate 13:1	Similarity	Field Similarity	Raw Field Score	Penalised Field Score	Shape Similarity	Raw Shape Score	Penalised Shape Score	Total Penalty	Excluded Volume Clash Penalty	Atom Distance Constraint Penalty	Field Value Penalty	Template Pairwise RMS Similarity	Template RMS Angle Deviation	Template RMS Distance Deviation	Template Density
59	1	2	1	3	1	14	1	1	0.713248	0.656947	-37.0565	-37.0565	0.76955	79.4811	79.4811	0	0	0	0	0.678434	7.97792	0.270486	0.833333
60	1	1	1	3	1	8	1	1	0.708965	0.615367	-36.1443	-36.1443	0.802563	82.891	82.891	0	0	0	0	0.703898	4.34522	0.587335	0.833333
61	1	2	1	3	1	1	1	1	0.708947	0.629733	-37.3424	-37.3424	0.78816	81.4087	81.4087	0	0	0	0	0.680983	3.45218	0.435768	0.833333
62	1	1	1	3	1	9	1	1	0.70046	0.60183	-35.0049	-35.0049	0.799089	82.5334	82.5334	0	0	0	0	0.696163	3.78372	0.559736	0.833333
63	1	1	1	3	1	1	1	1	0.699919	0.624139	-37.2087	-37.2087	0.7757	80.1265	80.1265	0	0	0	0	0.682261	8.84066	0.597077	0.833333
64	1	1	1	3	1	7	1	1	0.698663	0.606538	-34.2882	-34.2882	0.790788	81.6935	81.6935	0	0	0	0	0.671496	9.94607	0.515126	0.833333
65	1	1	1	3	1	14	1	1	0.688496	0.612851	-34.8175	-34.8175	0.764141	78.9267	78.9267	0	0	0	0	0.671677	9.23378	0.486492	0.833333
66	1	1	1	3	1	1	1	1	0.679877	0.584101	-34.8305	-34.8305	0.775653	80.1311	80.1311	0	0	0	0	0.671583	2.76069	0.594714	0.833333
67	1	2	1	3	1	2			0.811113	0.836062	-43.5187	-43.5187	0.786163	82.1453	82.1453	0	0	0	0	0.814122	2.33513	0.079002	1
68	1	2	1	3	1	5			0.810493	0.828863	-47.1211	-47.1211	0.792122	82.7685	82.7685	0	0	0	0	0.812567	3.32677	0.106607	1
69	1	2	1	3	1	4			0.809453	0.826259	-47.2329	-47.2329	0.792648	82.8295	82.8295	0	0	0	0	0.813209	2.93106	0.078756	1
70	1	2	1	3	1	3			0.807047	0.828287	-43.0644	-43.0644	0.785808	82.1146	82.1146	0	0	0	0	0.810687	3.63525	0.079445	1
71	1	2	1	2	1	4			0.80003	0.815801	-48.2449	-48.2449	0.784259	81.9815	81.9815	0	0	0	0	0.797132	2.97648	0.079912	1
72	1	2	1	1	1	5			0.799617	0.820621	-49.0783	-49.0783	0.778613	81.387	81.387	0	0	0	0	0.799716	4.23137	0.169013	1
73	1	2	1	1	1	2			0.79729	0.814151	-44.5889	-44.5889	0.780429	81.5678	81.5678	0	0	0	0	0.799785	2.74327	0.109549	1

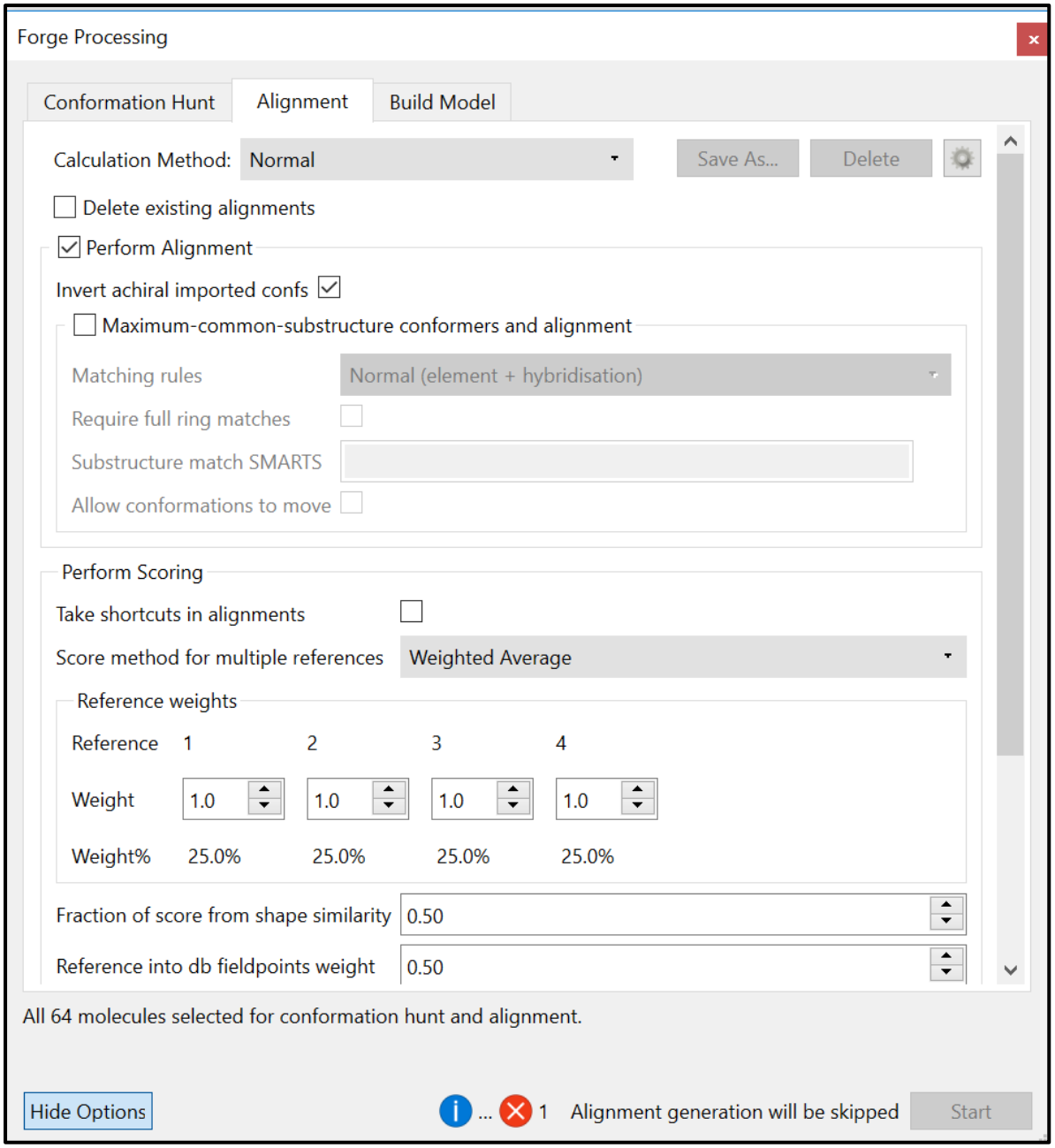
Template	Isomer Substrate 7:1	Conf Substrate 7:1	Isomer Substrate 28:1	Conf Substrate 28:1	Isomer Substrate 27:1	Conf Substrate 27:1	Isomer Substrate 13:1	Conf Substrate 13:1	Similarity	Field Similarity	Raw Field Score	Penalised Field Score	Shape Similarity	Raw Shape Score	Penalised Shape Score	Total Penalty	Excluded Volume Clash Penalty	Atom Distance Constraint Penalty	Field Value Penalty	Template Pairwise RMS Similarity	Template RMS Angle Deviation	Template RMS Distance Deviation	Template Density
74	1	2	1	2	1	3			0.795202	0.810535	-43.555	-43.555	0.779869	81.5128	81.5128	0	0	0	0	0.793817	1.93182	0.04804	1
75	1	2	1	3	1	8			0.790185	0.766556	-45.4255	-45.4255	0.813813	85.0083	85.0083	0	0	0	0	0.818849	2.80693	0.46356	1
76	1	2	1	2	1	2			0.786205	0.788356	-42.3482	-42.3482	0.784053	81.9392	81.9392	0	0	0	0	0.791147	7.79474	0.210381	1
77	1	2	1	3	1	9			0.783841	0.754845	-44.1232	-44.1232	0.812838	84.9045	84.9045	0	0	0	0	0.815252	9.78308	0.515525	1
78	1	2	1	1	1	3			0.783111	0.783086	-42.818	-42.818	0.783136	81.8222	81.8222	0	0	0	0	0.789144	7.86352	0.201624	1
79	1	1	1	3	1	3			0.779306	0.767201	-40.2044	-40.2044	0.791412	82.695	82.695	0	0	0	0	0.786095	1.093	0.187613	1
80	1	2	1	1	1	4			0.778368	0.770293	-46.098	-46.098	0.786444	82.1998	82.1998	0	0	0	0	0.78604	8.15452	0.153512	1
81	1	1	1	3	1	4			0.777969	0.757811	-43.7144	-43.7144	0.798127	83.401	83.401	0	0	0	0	0.78045	3.02921	0.220551	1
82	1	1	1	3	1	5			0.777656	0.762255	-43.7193	-43.7193	0.793056	82.8695	82.8695	0	0	0	0	0.782017	1.37306	0.113381	1
83	1	2	1	2	1	5			0.777112	0.771978	-45.2453	-45.2453	0.782245	81.7623	81.7623	0	0	0	0	0.781948	5.50333	0.100196	1
84	1	1	1	3	1	2			0.776751	0.761884	-40.0177	-40.0177	0.791619	82.7191	82.7191	0	0	0	0	0.776329	3.73515	0.195238	1
85	1	2	1	2	1	7			0.76214	0.698027	-41.2516	-41.2516	0.826252	86.2818	86.2818	0	0	0	0	0.767289	2.18931	0.057213	1
86	1	2	1	3	1	1			0.754996	0.723489	-43.825	-43.825	0.786502	82.2146	82.2146	0	0	0	0	0.777674	2.98114	0.512013	1
87	1	2	1	3	1	14			0.7548	0.747905	-42.54	-42.54	0.761695	79.6238	79.6238	0	0	0	0	0.76495	6.39223	0.312024	1
88	1	1	1	3	1	9			0.749714	0.68816	-40.6816	-40.6816	0.811269	84.7504	84.7504	0	0	0	0	0.790966	2.22825	0.616205	1
89	1	1	1	3	1	8			0.746727	0.683689	-40.9207	-40.9207	0.809764	84.6133	84.6133	0	0	0	0	0.801195	2.85338	0.647778	1
90	1	1	1	3	1	14			0.72364	0.693829	-39.821	-39.821	0.753452	78.8153	78.8153	0	0	0	0	0.748584	8.07904	0.560183	1
91	1	1	1	3	1	1			0.721179	0.645772	-39.4903	-39.4903	0.796586	83.2323	83.2323	0	0	0	0	0.763799	7.69629	0.697052	1
92	1	2	1	3			1	1	0.809883	0.780073	-42.2796	-42.2796	0.839693	87.1449	87.1449	0	0	0	0	0.809653	3.02744	0.109719	1

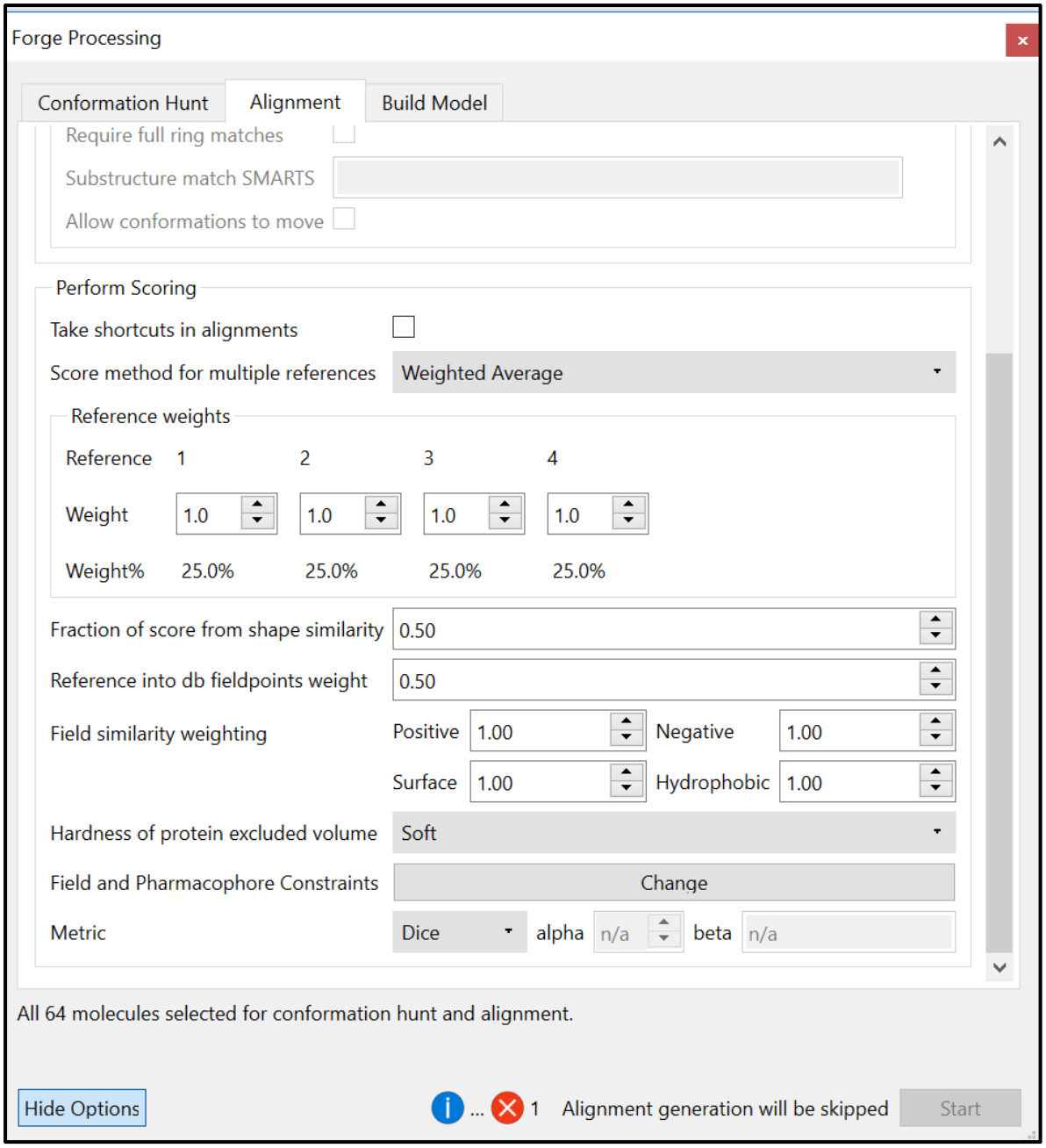
Template	Isomer Substrate 7:1	Conf Substrate 7:1	Isomer Substrate 28:1	Conf Substrate 28:1	Isomer Substrate 27:1	Conf Substrate 27:1	Isomer Substrate 13:1	Conf Substrate 13:1	Similarity	Field Similarity	Raw Field Score	Penalised Field Score	Shape Similarity	Raw Shape Score	Penalised Shape Score	Total Penalty	Excluded Volume Clash Penalty	Atom Distance Constraint Penalty	Field Value Penalty	Template Pairwise RMS Similarity	Template RMS Angle Deviation	Template RMS Distance Deviation	Template Density
93	1	1	1	3			1	1	0.796786	0.740915	-40.5027	-40.5027	0.852656	88.5228	88.5228	0	0	0	0	0.798269	1.29187	0.070962	1
94	1	2	1	1			1	1	0.79464	0.74669	-42.6158	-42.6158	0.842589	87.4622	87.4622	0	0	0	0	0.791719	2.49893	0.035716	1
95	1	2	1	2			1	1	0.79442	0.745947	-41.8048	-41.8048	0.842894	87.5005	87.5005	0	0	0	0	0.789436	2.01151	0.05933	1

Conformation hunting and molecule alignment

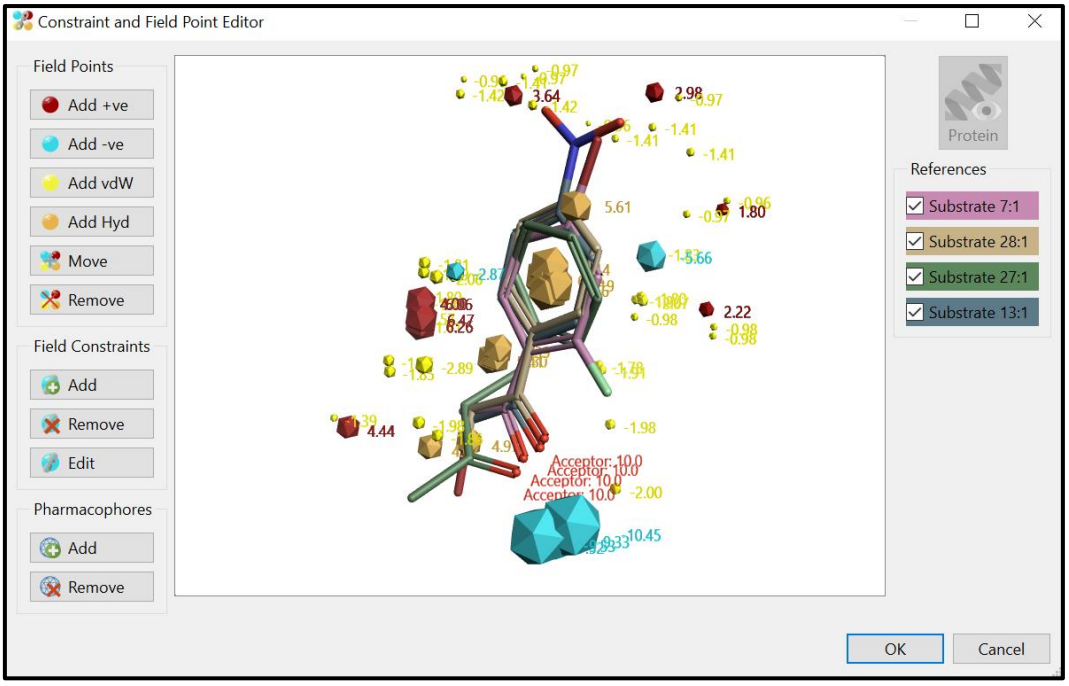
Conformations and alignments were then calculated and compared to the pharmacophore reference generated.




Forge Processing
Conformation Hunt
Alignment
Build Model
Calculation Method:
Normal
Save As...
Delete
Delete existing alignments
Perform Alignment
Invert achiral imported confs
Maximum-common-substructure conformers and alignment
Matching rules
Normal (element + hybridisation)
Require full ring matches
Substructure match SMARTS
Allow conformations to move
Perform Scoring
Take shortcuts in alignments
Score method for multiple references
Weighted Average
Reference weights
Reference 1 2 3 4
Weight 1.0 1.0 1.0 1.0
Weight% 25.0% 25.0% 25.0% 25.0%
Fraction of score from shape similarity 0.50
Reference into db fieldpoints weight 0.50
All 64 molecules selected for conformation hunt and alignment.
Hide Options
... 1 Alignment generation will be skipped
Start



The following pharmacophore constraint was applied to ensure alignments scored highest if they overlapped the carbonyl with the reference pharmacophore, acceptor 10.0 strength.



All alignments were manually checked and adjusted in the Edit Molecule tool if necessary

Categorising activity

Method

Data set partitioned (activity stratified) and specific enzyme activity categorised as below

Results

Category	Specific Enzyme Activity range/ mUnits	Number of molecules
1	<30	14
2	30<x<90	12
3	>90	25

51 molecules in training set

13 molecules in test set

Model building

Method

Forge's automatic model building settings were used to find the best fit initially, with small adjustments made to test set size, number of folds for cross-validation and global optimisation time limit in order to find the best fit model, with the best fit model detailed in the logs below.

Results

Model log

Building model at Thu Apr 16 13:08:09 2020

Building sample point list from 51 training set molecules

Pre-cluster: 884 points

Post-cluster: 140 points

Process config 'SVM Classification'

Sample point minimum distance threshold: 1

Type of machine learning model: Support Vector Machine (SVM) Classification

Use electrostatic fields in model building: true

Use hydrophobic (bulk grease) field in model building: false

Use references in generating field samples: false

Use steric (vdW) field in model building: false

Use volume indicator field in model building: true

Weighting: weight molecules according to their similarity: false

Cache size (MB): 200

Global optimization - maximum C: 1000

Global optimization - maximum gamma: 0.1

Global optimization - maximum number of iterations: 50

Global optimization - minimum C: 0.1

Global optimization - minimum gamma: 1e-5

Global optimization - time limit (s): 3600

Number of folds for cross-validation: 5

Perform parameter optimization: true

Finished building model at Thu Apr 16 13:08:10 2020

Support Vector Machine (SVM) classification model (radial basis function kernel) parameters:

Gamma: 6.14060e-3

C: 7.14352e+1

Number of basis vectors for 1-2 classifier: 26

Number of basis vectors for 1-3 classifier: 30

Number of basis vectors for 2-3 classifier: 35

Statistics for predictions from full model on 51 compds:

Confusion matrix:

Predicted

1 2 3

Actual 1 14 0 0

2 0 12 0

3 0 0 25

Informedness (Bookmaker's): 1.000

F1 statistic: 1.000

Mean precision: 1.000

Mean recall: 1.000

Precision Recall Youden's J

Class 1 1.000 1.000 1.000

Class 2 1.000 1.000 1.000

Class 3 1.000 1.000 1.000

Statistics for predictions from cross-validation on 51 compds:

Confusion matrix:

Predicted

1 2 3

Actual 1 6 1 3

2 3 3 4

3 1 3 21

Informedness (Bookmaker's): 0.442

F1 statistic: 0.586

Mean precision: 0.593

Mean recall: 0.580

Precision Recall Youden's J

Class 1 0.600 0.600 0.486

Class 2 0.429 0.300 0.186

Class 3 0.750 0.840 0.490

Note that the sum of the elements of the confusion matrix from k-fold cross-validation may be less than the number of training molecules if the number of molecules in a class is not an even multiple of the number of folds

Statistics for test set predictions from full model on 13 compds:

Confusion matrix:

Predicted

1 2 3

Actual 1 2 1 0

2 2 2 0

3 0 1 5

Informedness (Bookmaker's): 0.550

F1 statistic: 0.667

Mean precision: 0.667

Mean recall: 0.667

Precision Recall Youden's J

Class 1 0.500 0.667 0.467

Class 2 0.500 0.500 0.278

Class 3 1.000 0.833 0.833

Molecule list for creation of QSAR descriptor matrix uses training set molecules:

Molecule 'Substrate 9:1' activity 2 similarity 0.768662

Molecule 'Substrate 28:1 (1)' activity 3 similarity 0.802714

Molecule 'Substrate 24:1' activity 3 similarity 0.775898

Molecule 'Substrate R:1' activity 1 similarity 0.505279

Molecule 'Substrate 3:1' activity 3 similarity 0.804482

Molecule 'Substrate 36:1' activity 3 similarity 0.713799

Molecule 'Substrate 30:1' activity 2 similarity 0.667505

Molecule 'Substrate 13:1 (1)' activity 3 similarity 0.774402

Molecule 'Substrate Z:1' activity 3 similarity 0.773946

Molecule 'Substrate 21:1' activity 1 similarity 0.692405

Molecule 'Substrate 15:1' activity 2 similarity 0.718686

Molecule 'Substrate M:1' activity 3 similarity 0.709612

Molecule 'Substrate 33:1' activity 1 similarity 0.718807

Molecule 'Substrate 27:1 (1)' activity 3 similarity 0.764832

Molecule 'Substrate AD:1' activity 1 similarity 0.733445

Molecule 'Substrate 34:1' activity 1 similarity 0.76425

Molecule 'Substrate 22:1' activity 1 similarity 0.761697

Molecule 'Substrate 16:1' activity 2 similarity 0.672556

Molecule 'Substrate AA:1' activity 3 similarity 0.716728

Molecule 'Substrate L:1' activity 3 similarity 0.733287

Molecule 'Substrate 2:1' activity 2 similarity 0.774585

Molecule 'Substrate 20:1' activity 3 similarity 0.700799

Molecule 'Substrate AF:1' activity 3 similarity 0.708141

Molecule 'Substrate D:1' activity 1 similarity 0.56784

Molecule 'Substrate K:1' activity 2 similarity 0.524355

Molecule 'Substrate S:1' activity 1 similarity 0.654308

Molecule 'Substrate 5:1' activity 1 similarity 0.755082

Molecule 'Substrate 35:1' activity 2 similarity 0.71247

Molecule 'Substrate AI:1' activity 1 similarity 0.722646

Molecule 'Substrate AJ:1' activity 3 similarity 0.783419

Molecule 'Substrate 19:1' activity 3 similarity 0.772806

Molecule 'Substrate 31:1' activity 3 similarity 0.800635

Molecule 'Substrate AB:1' activity 3 similarity 0.679284

Molecule 'Substrate 11:1' activity 3 similarity 0.75562

Molecule 'Substrate 4:1' activity 3 similarity 0.768086

Molecule 'Substrate AG:1' activity 2 similarity 0.595191

Molecule 'Substrate I:1' activity 1 similarity 0.431222

Molecule 'Substrate 18:1' activity 2 similarity 0.706099

Molecule 'Substrate 1:1' activity 3 similarity 0.751198

Molecule 'Substrate 6:1' activity 3 similarity 0.770132

Molecule 'Substrate G:1' activity 2 similarity 0.683298

Molecule 'Substrate J:1' activity 3 similarity 0.724922

Molecule 'Substrate 12:1' activity 3 similarity 0.748498

Molecule 'Substrate P:1' activity 1 similarity 0.663946

Molecule 'Substrate 14:1' activity 2 similarity 0.725131

Molecule 'Substrate AE:1' activity 3 similarity 0.597768

Molecule 'Substrate O:1' activity 3 similarity 0.738688

Molecule 'Substrate C:1' activity 1 similarity 0.694372

Molecule 'Substrate 26:1' activity 3 similarity 0.648241

Molecule 'Substrate 8:1' activity 1 similarity 0.785681

Molecule 'Substrate 37:1' activity 2 similarity 0.740522

Data export from Forge

Molecule data export

Title	Radial Plot	Activity (Specific Enzyme Activity)	Support Vector Machine Classification (Specific Enzyme Activity)	Alignment Chosen	Sim	FSim	FScore	FScore+P	SSim	SScore	SScore+P	Excl Vol Pen	Field Pen	Confs	Conf#	Alns	MW	#Atoms	2D Sim	SlogP	TPSA	Flexibility	#RB	Rof5	Exp. Specific Enzyme Activity
Substrate 7:1	0.482	3		n/a	0	0	0	0	0	0	0	0	0	0	n/a	0	217	11	0.422	3.1	17.1	0.3	1	0	366
Substrate 28:1	0.495	3		n/a	0	0	0	0	0	0	0	0	0	0	n/a	0	199	10	0.455	2.3	17.1	1.3	2	0	337.2
Substrate 27:1	0.495	3		n/a	0	0	0	0	0	0	0	0	0	0	n/a	0	148.2	11	0.404	2.2	17.1	3	3	0	364
Substrate 13:1	0.744	3		n/a	0	0	0	0	0	0	0	0	0	0	n/a	0	165.1	12	0.426	1.7	62.9	0.5	2	0	593.1
Substrate 28:1 (1)	0.687	3	3	1	0.803	0.813	-47.096	-45.017	0.792	87.724	81.488	0	0	2	1	2	199	10	0.455	2.3	17.1	1.3	2	0	337.2
Substrate 24:1	0.652	3	3	1	0.776	0.755	-40.977	-38.923	0.797	91.71	85.548	0	0	2	2	2	190.6	12	0.244	2.9	17.1	1.3	2	0	111.7
Substrate 36:1	0.576	3	3	1	0.714	0.659	-35.279	-33.075	0.769	79.895	73.281	0	0	2	1	10	126.2	9	0.047	2.2	17.1	2	1	0	111.6
Substrate 30:1	0.48	2	2	1	0.668	0.693	-43.748	-41.507	0.642	75.874	69.152	0	0	108	101	10	156.3	11	0.049	3.3	17.1	7	7	0	37.1
Substrate Z:1	0.702	3	3	1	0.774	0.743	-47.745	-45.633	0.805	87.659	81.323	0	0	2	2	2	150.2	11	0.31	1.9	26.3	0.8	2	0	222.1
Substrate 21:1	0.654	1	1	1	0.692	0.626	-36.301	-34.103	0.759	86.257	79.662	0	0	3	3	3	159.2	12	0.287	2.4	32.9	0.3	1	0	0
Substrate 15:1	0.633	2	2	1	0.719	0.738	-45.742	-43.614	0.699	84.598	78.216	0	0	3	1	3	178.2	13	0.235	2.5	26.3	1.8	3	0	64
Substrate M:1	0.818	3	3	1	0.71	0.7	-48.353	-46.273	0.72	88.297	82.058	0	0	5	2	5	213.3	14	0.25	1.3	63.2	1	3	0	203.3
Substrate 33:1	0.825	1	1	1	0.719	0.649	-45.175	-43.159	0.789	88.028	81.98	0	0	5	4	5	165.2	12	0.277	0.8	49.3	2.8	4	0	17.4
Substrate 27:1 (1)	0.639	3	3	1	0.765	0.761	-40.518	-38.252	0.768	85.456	78.659	0	0	2	2	2	148.2	11	0.404	2.2	17.1	3	3	0	364
Substrate 34:1	0.782	1	1	1	0.764	0.741	-41.098	-38.965	0.788	81.24	74.843	0	0	4	4	4	136.2	10	0.316	1.6	37.3	0.8	2	0	0
Substrate 16:1	0.728	2	2	1	0.673	0.651	-41.438	-39.393	0.694	88.618	82.484	0	0	6	1	6	210.2	15	0.154	1.9	44.8	1.8	4	0	89.9
Substrate L:1	0.681	3	3	1	0.733	0.701	-40.434	-38.37	0.766	88.008	81.816	0	0	2	2	2	189.1	13	0.192	2.3	30	0.3	2	0	210.2
Substrate 2:1	0.796	2	2	1	0.775	0.729	-47.371	-45.323	0.82	83.976	77.831	0	0	2	2	2	136.2	10	0.347	1.6	37.3	0.8	2	0	31.5
Substrate 20:1	0.75	3	3	1	0.701	0.615	-31.302	-29.276	0.786	76.041	69.961	0	0	2	2	2	122.1	9	0.214	0.7	42.8	0.3	1	0	185.1
Substrate K:1	0.74	2	2	1	0.524	0.471	-33.606	-29.816	0.578	78.104	66.732	0	0	1	1	1	207.2	15	0.239	0.8	80.4	4.5	6	0	62.2

Title	Radial Plot	Activity (Specific Enzyme Activity)	Support Vector Machine Classification (Specific Enzyme Activity)	Alignment Chosen	Sim	FSim	FScore	FScore+P	SSim	SScore	SScore+P	Excl Vol Pen	Field Pen	Confs	Conf#	Alns	MW	#Atoms	2D Sim	SlogP	TPSA	Flexibility	#RB	Rof5	Exp. Specific Enzyme Activity
Substrate S:1	0.716	1	1	1	0.654	0.599	-31.05	-28.23	0.71	69.315	60.853	0	0	1	1	1	110.1	8	0.169	0.6	45.8	0.3	1	0	20.7
Substrate AI:1	0.815	1	1	1	0.723	0.693	-43.518	-41.377	0.753	92.733	86.31	0	0	3	3	3	260	14	0.256	1.8	83.1	2	4	0	0
Substrate AJ:1	0.662	3	3	1	0.783	0.768	-46.202	-44.141	0.799	91.231	85.049	0	0	2	1	2	246	10	0.347	2.8	17.1	0.3	1	0	116.9
Substrate 19:1	0.733	3	3	1	0.773	0.728	-38.784	-36.769	0.818	82.973	76.928	0	0	2	1	2	139.1	10	0.275	1.7	30	0.3	1	0	136.6
Substrate AB:1	0.659	3	3	1	0.679	0.667	-40.294	-38.262	0.691	81.984	75.886	0	0	5	5	5	180.2	13	0.174	1.9	35.5	1.3	3	0	362.6
Substrate 11:1	0.756	3	3	1	0.756	0.739	-42.193	-40.101	0.773	85.701	79.424	0	0	2	2	2	164.2	12	0.264	1.6	35.5	0.3	1	0	188.5
Substrate 4:1	0.643	3	3	1	0.768	0.793	-48.674	-46.639	0.743	82.518	76.412	0	0	2	1	2	199	10	0.361	2.7	17.1	0.3	1	0	141.6
Substrate AG:1	0.621	2	2	1	0.595	0.573	-46.252	-44.208	0.617	89.646	83.515	0	0	2	1	2	246.1	18	0.148	3	35.5	0.8	2	0	87
Substrate I:1	0.62	1	1	1	0.431	0.336	-20.446	-16.33	0.526	60.942	48.595	0	0	1	1	1	124.1	9	0.167	0.6	34.9	0.3	1	0	20.7
Substrate F:1	0.498	1	1	1	0.554	0.525	-29.559	-27.486	0.584	61.223	55.004	0	0	1	1	10	125.2	9	0.039	0.3	20.3	1.4	0	0	23.4
Substrate 39:1	0.652	2	1	1	0.775	0.798	-43.547	-41.531	0.753	83.615	77.567	0	0	2	2	2	199	10	0.352	2.7	17.1	0.3	1	0	39.5
Substrate AC:1	0.774	1	2	1	0.72	0.719	-47.941	-45.886	0.721	87.289	81.123	0	0	2	2	2	204.1	12	0.19	2.8	43.1	0.3	1	0	0
Substrate AH:1	0.708	3	3	1	0.753	0.695	-41.128	-39.087	0.81	88.355	82.233	0	0	2	2	2	200	10	0.296	2.1	30	0.3	1	0	268.7
Substrate 7:1 (1)	0.679	3	3	1	0.807	0.812	-43.683	-41.616	0.803	91.771	85.569	0	0	2	1	2	217	11	0.422	3.1	17.1	0.3	1	0	366
Substrate 17:1	0.717	3	3	1	0.76	0.724	-39.187	-37.112	0.796	78.457	72.234	0	0	2	1	2	121.1	9	0.292	1.3	30	0.3	1	0	162.6
Substrate 10:1	0.676	2	1	1	0.753	0.756	-43.545	-41.486	0.751	82.035	75.859	0	0	2	2	2	150.2	11	0.288	1.9	26.3	0.8	2	0	83
Substrate 25:1	0.675	3	3	1	0.794	0.776	-43.077	-41.002	0.812	89.85	83.626	0	0	2	2	2	199	10	0.393	2.7	17.1	0.3	1	0	158.2
Substrate N:1	0.773	3	2	1	0.733	0.701	-38.575	-36.541	0.765	81.633	75.529	0	0	2	1	2	145.2	11	0.283	1.8	40.9	1.3	1	0	173.4
Substrate 6:1	0.632	3	3	1	0.77	0.752	-42.815	-40.752	0.788	90.147	83.957	0	0	2	2	2	217	11	0.338	3.1	17.1	0.3	1	0	242.4
Substrate G:1	0.536	2	2	1	0.683	0.656	-43.716	-41.383	0.711	78.846	71.848	0	0	3	1	3	146.2	11	0.289	2.2	17.1	0.5	2	0	62.1
Substrate 12:1	0.669	3	3	1	0.748	0.726	-48.61	-46.517	0.771	88.129	81.849	0	0	4	4	4	164.2	12	0.312	2.3	26.3	1.8	3	0	195.8
Substrate P:1	0.621	1	1	1	0.664	0.622	-40.443	-38.399	0.706	83.46	77.33	0	0	2	1	2	176.2	13	0.215	1.2	33.2	0.8	3	0	0
Substrate AE:1	0.497	3	3	1	0.598	0.586	-29.897	-27.819	0.609	55.208	48.974	0	0	2	1	10	84.1	6	0.062	1	17.1	1	1	0	98.2

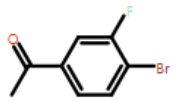
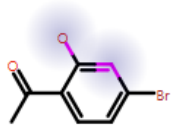
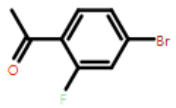
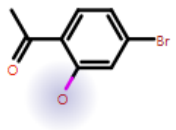
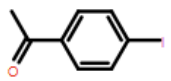
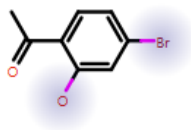
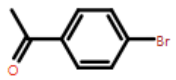
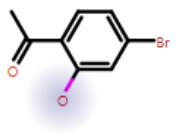
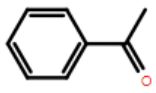
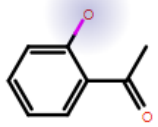
Title	Radial Plot	Activity (Specific Enzyme Activity)	Support Vector Machine Classification (Specific Enzyme Activity)	Alignment Chosen	Sim	FSim	FScore	FScore+P	SSim	SScore	SScore+P	Excl Vol Pen	Field Pen	Confs	Conf#	Alns	MW	#Atoms	2D Sim	SlogP	TPSA	Flexibility	#RB	Rof5	Exp. Specific Enzyme Activity
Substrate C:1	0.667	1	1	1	0.694	0.62	-36.545	-34.141	0.768	84.975	77.762	0	0	2	2	2	158.2	12	0.284	1.4	34.1	0.2	0	0	0
Substrate 8:1	0.809	1	1	1	0.786	0.766	-43.138	-41.108	0.806	91.355	85.263	0	0	3	3	3	215	11	0.345	2.4	37.3	0.8	2	0	27.2
Substrate 37:1	0.608	2	2	1	0.741	0.708	-48.004	-45.948	0.773	90.664	84.498	0	0	2	2	2	170.2	13	0.309	3	17.1	0.3	1	0	54
Substrate 23:1	0.394	2	2	1	0.662	0.689	-40.906	-38.873	0.635	85.058	78.959	0	0	3	1	3	256.1	17	0.114	3.9	17.1	0.3	3	0	57.3
Substrate 32:1	0.721	3	3	1	0.764	0.712	-41.115	-39.078	0.816	87.07	80.961	0	0	3	1	3	155.6	10	0.26	2	30	0.3	1	0	92
Substrate 29:1	0.518	1	1	1	0.67	0.656	-43.704	-41.673	0.684	83.304	77.21	0	0	3	1	3	182.2	14	0.215	2.9	17.1	0.5	2	0	14.3
Substrate 9:1	0.837	2	2	1	0.769	0.72	-46.844	-44.794	0.817	84.046	77.897	0	0	2	2	2	135.2	10	0.347	1.5	43.1	0.3	1	0	80.4
Substrate R:1	0.742	1	1	1	0.505	0.449	-38.884	-32.61	0.561	90.439	71.618	0	0	82	71	10	232.3	17	0.176	1.9	54.1	3.5	4	0	21.7
Substrate 3:1	0.688	3	3	1	0.804	0.79	-38.509	-36.433	0.819	84.477	78.247	0	0	2	2	2	138.1	10	0.363	2.3	17.1	0.3	1	0	122.1
Substrate 13:1 (1)	0.899	3	3	1	0.774	0.736	-42.096	-40.062	0.813	88.38	82.277	0	0	2	1	2	165.1	12	0.426	1.7	62.9	0.5	2	0	593.1
Substrate AD:1	0.848	1	1	1	0.733	0.707	-46.327	-44.276	0.76	92.774	86.621	0	0	12	1	10	245.1	13	0.245	1.8	57.5	3.8	5	0	0
Substrate 22:1	0.634	1	1	1	0.762	0.746	-42.76	-40.57	0.777	86.369	79.799	0	0	2	2	2	174.1	12	0.231	2.4	17.1	0.3	2	0	29.4
Substrate AA:1	0.707	3	3	1	0.717	0.686	-46.776	-44.662	0.747	88.465	82.125	0	0	4	1	4	180.2	13	0.225	1.9	35.5	1.3	3	0	220
Substrate AF:1	0.569	3	3	1	0.708	0.66	-35.541	-33.382	0.756	74.844	68.366	0	0	2	1	10	112.2	8	0.051	1.8	17.1	1.5	1	0	218.3
Substrate D:1	0.687	1	1	1	0.568	0.476	-34.842	-32.781	0.659	94.686	88.504	0	0	8	8	8	288.3	20	0.146	1.7	88.5	2	2	0	0
Substrate 5:1	0.77	1	1	1	0.755	0.73	-43.61	-41.564	0.78	82.876	76.738	0	0	3	1	3	154.1	11	0.278	2	37.3	0.8	2	0	0
Substrate 35:1	0.657	2	2	1	0.712	0.631	-34.607	-32.56	0.794	78	71.861	0	0	2	2	2	121.1	9	0.293	1.3	30	0.3	1	0	60.3
Substrate 31:1	0.685	3	3	1	0.801	0.798	-45.027	-42.993	0.804	80.1	73.998	0	0	2	2	2	120.2	9	0.391	1.9	17.1	0.3	1	0	99.1
Substrate 18:1	0.756	2	2	1	0.706	0.623	-29.659	-27.629	0.789	76.64	70.552	0	0	2	2	2	122.1	9	0.198	0.7	42.8	0.3	1	0	50.5
Substrate Y:1	0.717	2	2	1	0.624	0.62	-44.554	-42.458	0.628	92.76	86.471	0	0	7	1	7	275.2	17	0.115	3.2	55.1	2.3	4	0	34.6
Substrate 1:1	0.766	3	3	1	0.751	0.712	-41.378	-39.355	0.791	81.058	74.988	0	0	3	3	3	136.2	10	0.32	1.6	37.3	0.8	2	0	156.6
Substrate J:1	0.671	3	3	1	0.725	0.68	-44.74	-42.632	0.77	89.263	82.94	0	0	2	2	2	171.2	13	0.287	2.4	30	0.3	1	0	229.2
Substrate 14:1	0.64	2	2	1	0.725	0.682	-40.664	-38.596	0.768	90.368	84.163	0	0	3	3	3	176.2	13	0.27	2.2	26.3	1	1	0	36

Title	Radial Plot	Activity (Specific Enzyme Activity)	Support Vector Machine Classification (Specific Enzyme Activity)	Alignment Chosen	Sim	FSim	FScore	FScore+P	SSim	SScore	SScore+P	Excl Vol Pen	Field Pen	Confs	Conf#	Alns	MW	#Atoms	2D Sim	SlogP	TPSA	Flexibility	#RB	Rof5	Exp. Specific Enzyme Activity
Substrate O:1	0.659	3	3	1	0.739	0.73	-42.281	-40.245	0.748	81.608	75.499	0	0	3	3	3	150.2	11	0.301	1.9	26.3	0.8	2	0	396.7
Substrate 26:1	0.495	3	3	1	0.648	0.618	-37.595	-35.356	0.678	72.661	65.943	0	0	4	1	4	134.2	10	0.362	1.8	17.1	2	2	0	239.4

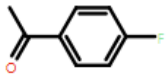
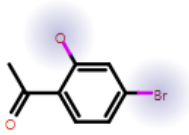
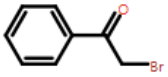
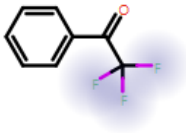
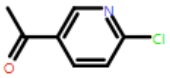
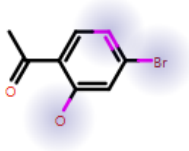
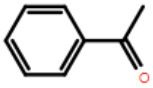
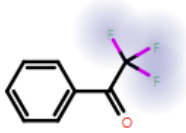
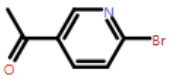
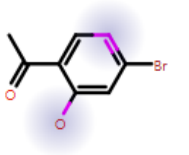
Activity Miner compound pairings

Substrate pairs showing greatest disparity between field similarity and specific enzyme activity with (*S*)-ADH

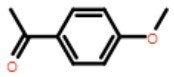
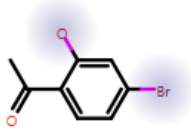
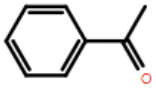
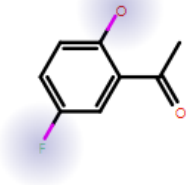
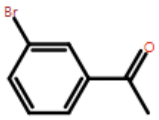
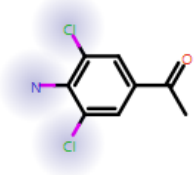
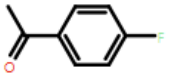
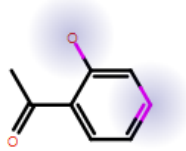
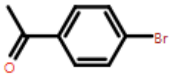
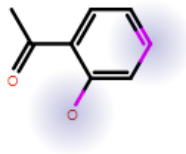
Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate 6:1 	3.000	Substrate 8:1 	1.000	-40.0	0.951
Substrate 7:1 (1) 	3.000	Substrate 8:1 	1.000	-40.0	0.957
Substrate AJ:1 	3.000	Substrate 8:1 	1.000	-40.0	0.961
Substrate 25:1 	3.000	Substrate 8:1 	1.000	-40.0	0.974
Substrate 31:1 	3.000	Substrate 34:1 	1.000	-24.1	0.917

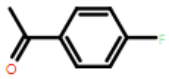
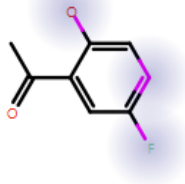
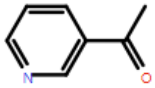
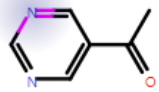
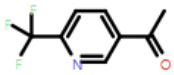
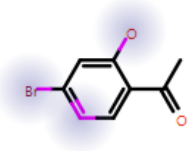
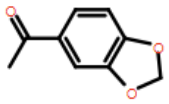
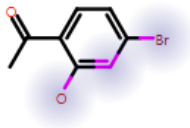
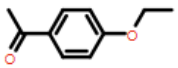
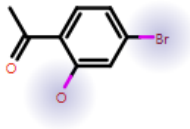
Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate 3:1 	3.000	Substrate 8:1 	1.000	-23.4	0.915
Substrate 28:1 (1) 	3.000	Substrate 22:1 	1.000	-23.3	0.914
Substrate 32:1 	3.000	Substrate 8:1 	1.000	-20.2	0.901
Substrate 31:1 	3.000	Substrate 22:1 	1.000	-20.2	0.901
Substrate AH:1 	3.000	Substrate 8:1 	1.000	-20.1	0.900

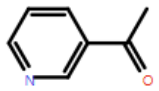
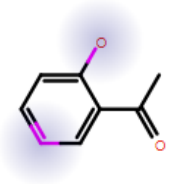
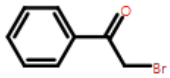
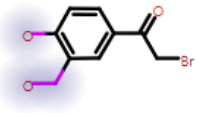
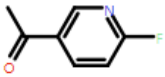
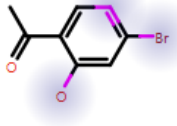
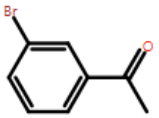
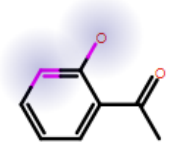
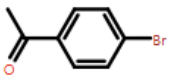
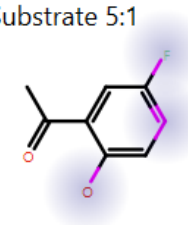
Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate Z:1 	3.000	Substrate 8:1 	1.000	-20.0	0.900
Substrate 31:1 	3.000	Substrate 5:1 	1.000	-19.5	0.897
Substrate 4:1 	3.000	Substrate AC:1 	1.000	-19.3	0.897
Substrate 3:1 	3.000	Substrate 34:1 	1.000	-18.3	0.891
Substrate 25:1 	3.000	Substrate 34:1 	1.000	-18.2	0.890

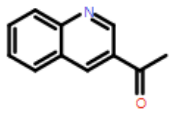
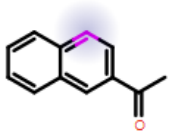
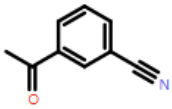
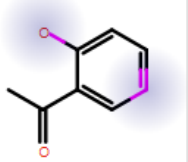
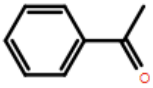
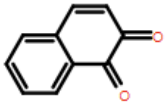
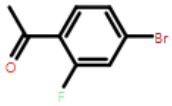
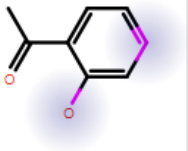
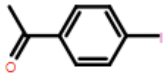
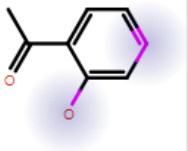
Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate 3:1 	3.000	Substrate 5:1 	1.000	-17.6	0.887
Substrate 17:1 	3.000	Substrate 18:1 	2.000	-17.4	0.942
Substrate L:1 	3.000	Substrate 8:1 	1.000	-17.2	0.884
Substrate 11:1 	3.000	Substrate 8:1 	1.000	-17.1	0.883
Substrate 12:1 	3.000	Substrate 8:1 	1.000	-16.4	0.878

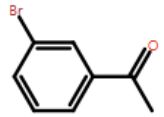
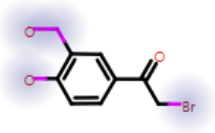
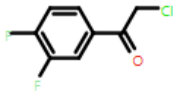
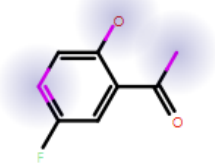
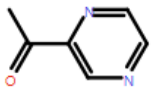
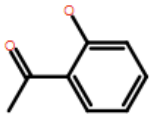
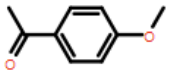
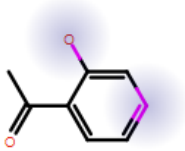
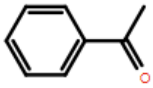
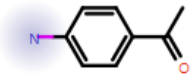
Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate 17:1 	3.000	Substrate 34:1 	1.000	-16.2	0.877
Substrate 28:1 (1) 	3.000	Substrate AD:1 	1.000	-16.0	0.875
Substrate 19:1 	3.000	Substrate 8:1 	1.000	-15.7	0.872
Substrate 4:1 	3.000	Substrate 34:1 	1.000	-15.4	0.870
Substrate 25:1 	3.000	Substrate 5:1 	1.000	-15.3	0.869

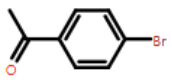
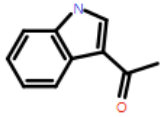
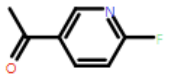
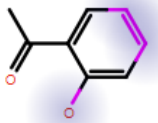
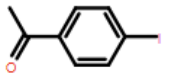
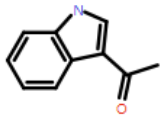
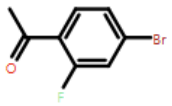
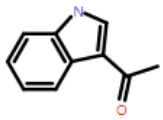
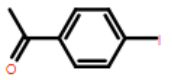
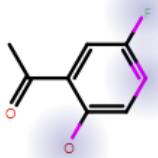
Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate J:1 	3.000	Substrate 37:1 	2.000	-15.2	0.934
Substrate N:1 	3.000	Substrate 34:1 	1.000	-15.1	0.868
Substrate 31:1 	3.000	Substrate C:1 	1.000	-14.7	0.864
Substrate 7:1 (1) 	3.000	Substrate 34:1 	1.000	-14.5	0.862
Substrate AJ:1 	3.000	Substrate 34:1 	1.000	-14.4	0.861

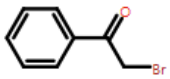
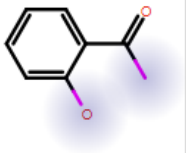
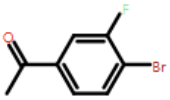
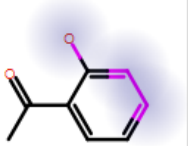
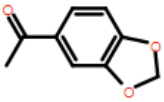
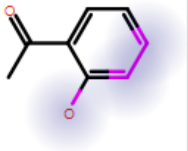
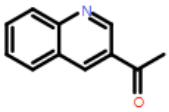
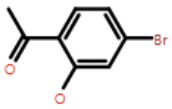
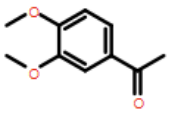
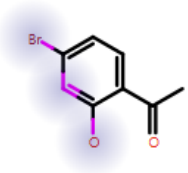
Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate 4:1 	3.000	Substrate AD:1 	1.000	-14.3	0.861
Substrate 24:1 	3.000	Substrate 5:1 	1.000	-14.3	0.860
Substrate 20:1 	3.000	Substrate 34:1 	1.000	-14.2	0.859
Substrate Z:1 	3.000	Substrate 34:1 	1.000	-14.1	0.858
Substrate 31:1 	3.000	Substrate 9:1 	2.000	-14.1	0.929

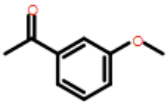
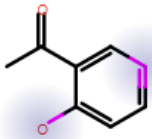
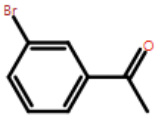
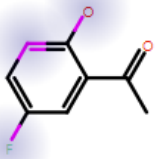
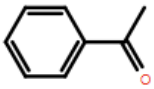
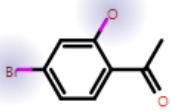
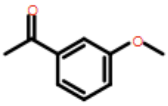
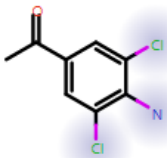
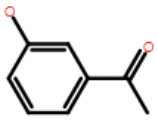
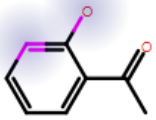
Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate 25:1 	3.000	Substrate 21:1 	1.000	-14.0	0.858
Substrate 19:1 	3.000	Substrate 34:1 	1.000	-13.9	0.857
Substrate AJ:1 	3.000	Substrate 21:1 	1.000	-13.9	0.856
Substrate 7:1 (1) 	3.000	Substrate 21:1 	1.000	-13.9	0.856
Substrate AJ:1 	3.000	Substrate 5:1 	1.000	-13.7	0.854

Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate 28:1 (1) 	3.000	Substrate 34:1 	1.000	-13.7	0.854
Substrate 6:1 	3.000	Substrate 34:1 	1.000	-13.6	0.853
Substrate 11:1 	3.000	Substrate 34:1 	1.000	-13.6	0.852
Substrate J:1 	3.000	Substrate 8:1 	1.000	-13.5	0.852
Substrate AA:1 	3.000	Substrate 8:1 	1.000	-13.4	0.850

Top Pairs (Specific Enzyme Activity)

Greater	Greater Activity	Lesser	Lesser Activity	Disparity	Similarity
Substrate O:1 	3.000	Substrate 34:1 	1.000	-13.2	0.848
Substrate 4:1 	3.000	Substrate 5:1 	1.000	-12.9	0.845
Substrate 31:1 	3.000	Substrate 8:1 	1.000	-12.9	0.845
Substrate O:1 	3.000	Substrate AC:1 	1.000	-12.9	0.845
Substrate 1:1 	3.000	Substrate 34:1 	1.000	-12.7	0.843

References

1 J. S. Rowbotham, M. A. Ramirez, O. Lenz, H. A. Reeve and K. A. Vincent, *Nat. Commun.*, 2020, **11**, 1–7.