Efficient Lewis acid catalysis of an abiological reaction in a de novo protein scaffold

Sophie Basler & Sabine Studer, Yike Zou, Takahiro Mori, Yusuke Ota, Anna Camus, H. Adrian Bunzel, Roger C. Helgeson, K. N. Houk, Anna Camus, Gonzalo Jiménez-Osés and Donald Hilvert

New enzyme catalysts are usually engineered by repurposing the active sites of natural proteins. Here we show that design and directed evolution can be used to transform a non-natural, functionally naive zinc-binding protein into a highly active catalyst for an abiological hetero-Diels–Alder reaction. The artificial metalloenzyme achieves $>10^4$ turnovers per active site, exerts absolute control over reaction pathway and product stereochemistry, and displays a catalytic proficiency ($1/K_{T5} = 2.9 \times 10^{15} \text{M}^{-1}$) that exceeds all previously characterized Diels–Alderases. These properties capitalize on effective Lewis acid catalysis, a chemical strategy for accelerating Diels–Alder reactions common in the laboratory but so far unknown in nature. Extension of this approach to other metal ions and other de novo scaffolds may propel the design field in exciting new directions.

The Diels–Alder reaction has found broad application in natural product synthesis, providing rapid access to complex molecular structures through the formation of two new $\sigma$ bonds and up to four contiguous stereocentres. Despite its synthetic utility, surprisingly few enzymes have been found to catalyse Diels–Alder reactions in nature. Most of these were discovered only recently and promote intramolecular cycloadditions in polyketide biosynthesis. Designer catalysts for even more challenging bimolecular reactions that simultaneously control the reaction pathway, endo/exo ratios and the absolute configuration of the reaction products would therefore be valuable additions to the biocatalytic toolkit. Both catalytic antibodies and computationally designed enzymes have been produced for several Diels–Alder reactions, often with useful stereoselectivities, but their efficiencies are generally modest, even after extensive laboratory evolution, reflecting their reliance on simple hydrogen bonding and hydrophobic binding for transition-state stabilization.

Lewis acid catalysis by metal ions, a well-established strategy for accelerating Diels–Alder reactions in the laboratory, could potentially augment the capabilities of these designer catalysts. In previous attempts to combine the advantages of transition-metal and enzymatic catalysis for Diels–Alder reactions, ligand-chelated copper(II) ions were incorporated into the binding pockets of several natural protein scaffolds, often with useful stereoselectivities, but their efficiencies are generally modest, even after extensive laboratory evolution, reflecting their reliance on simple hydrogen bonding and hydrophobic binding for transition-state stabilization.

In contrast with natural proteins that have been optimized by natural evolution for a specific purpose, de novo protein scaffolds that possess promiscuous binding pockets for metal ions and substrates might be more amenable to functional diversification. Consistent with this hypothesis, we recently transformed a non-natural, functionally naive zinc-binding protein into a highly active catalyst for an abiological hetero-Diels–Alder reaction.

Results and discussion

The target reaction. As our target transformation, we chose the Lewis-acid-catalysed reaction of azachalcone 1 with 3-vinylimidole 2. These compounds are common substrates for Diels–Alder reactions, related to natural units present in a variety of natural products. However, when paired, they can proceed via competing Diels–Alder and hetero-Diels–Alder pathways (Fig. 1). The development of a selective biocatalyst that promotes only one of the many possible reaction pathways thus poses a chemically interesting challenge.

Density functional theory (DFT) calculations suggest that in the absence of Lewis acid, the endo transition state (TS1-endo, 21.4 kcal mol$^{-1}$) is ambimodal, leading to both sets of products (Extended Data Fig. 1a). The inclusion of Zn(ii)(HO$_2$)$_2$ in the calculation shifts the cycloaddition from a slow, concerted reaction to a fast, stepwise reaction with a free energy barrier of 10 kcal mol$^{-1}$; the second step favours the endo hetero-Diels–Alder product (Extended Data Fig. 1b).

Experimentally, 1 and 2 reacted to give cyclohexenes 3 and 3,4-dihydro-2H-pyrans 4 in a 1:3 ratio in dimethyl sulfoxide containing zinc triflate. The product ratio shifted 1:19 in aqueous buffer containing zinc sulfate (Extended Data Fig. 2a–c). $^1$H-NMR analyses of the crude reaction mixtures showed that the endo stereoisomer of the hetero-Diels–Alder adduct 4 is favoured over the exo stereoisomer by a factor of 4.6:1 in dimethyl sulfoxide and 18:1 in buffer.

Design and evolution of a novel enzyme catalyst. The starting MID1sc protein did not detectably catalyse any reaction between 1 and 2. However, Rosetta design using DFT-optimized transition-state geometries predicted that two mutations, E32L and K68W, located on opposite ends of the binding pocket harbouring
the zinc ion, would improve substrate binding and transition-state stabilization while retaining protein stability. These substitutions replace charged side chains that might hinder substrate binding with neutral groups; in addition, Trp68 was expected to shield the relatively open active site from bulk solvent. In fact, introducing these substitutions into MID1sc afforded the enzyme DA0, which exhibited low but detectable activity (approximately twofold over the background level at 5 µM enzyme) and a product profile similar to that observed in the absence of protein (Extended Data Fig. 2c,d).

To increase its catalytic efficacy, DA0 was optimized by laboratory evolution. Eight residues lining the putative substrate-binding site were subjected to cassette mutagenesis, and the resulting variants were screened spectroscopically for consumption of the azachalcone in the presence of 2 and excess zinc. On the basis of the recovered sequences, five of these residues were re-randomized using restricted amino acid alphabets and shuffled (Extended Data Fig. 3a; Supplementary Tables 2 and 3). Screening of the combinatorial library yielded the enzyme DA1, which contained the mutations H35C and I64G and provided detectable activity in cell lysates. This double mutant was subsequently subjected to two consecutive rounds of evolution. DA1, for example, preferentially forms the endo hetero-Diels–Alder product (4R,6R)-3,4-dihydro-2H-pyran with >99% enantioselectivity (Fig. 2i). This was not the case for variants from the early rounds of evolution. DA0, for example, preferentially forms the endo (4S,6S)-3,4-dihydro-2H-pyran isomer, but only with 36% enantioselectivity, and the product additionally contains 5% of the competing exo isomers (Extended Data Fig. 2d). Although the transition state and the product of the DA7-catalysed reaction are structurally homologous, the enzyme catalyses >105 turnovers per active site (Extended Data Fig. 4f). As a result, preparative-scale reactions in aqueous buffer at room temperature afforded quantitative conversion of reactants to the optically pure product, providing a mild and practical alternative to standard chemical synthesis.

Structural and computational characterization. To understand the molecular changes that enabled the emergence of this proficient metalloenzyme, we crystallized DA7 and determined its structure to 1.5 Å resolution (Fig. 3a; Supplementary Table 7). Like the zinc-binding peptide MID1sc from which it originated, the evolved enzyme is a helical bundle consisting of two helix–turn–helix motifs connected by a flexible linker. As anticipated, though, the H3SC and H39V mutations altered the coordination sphere of the catalytic Zn(ii) ion (His39, His61, His65 → Cys35, His61, His65) and induced a major conformational change that reduced the crossover angle of the two helix–turn–helix fragments by >30° (Extended Data Fig. 5a–d). This structural rearrangement created a more enclosed hydrophobic pocket in the core of the bundle to accommodate both diene and dienophile. Because a thiolate is a better Zn(ii) ligand than a neutral histidine, the H3SC mutation early in the evolution likely facilitated the emergence of this innovation.

A similar conformational change was observed when we evolved MID1sc into an efficient metalloesterase. In that case, though, rearrangement occurred midway along the evolutionary trajectory and did not require substitution of His35 with a cysteine. Although the final, evolved esterase has a similar overall topology and Zn(ii) coordination geometry to DA7, its second helix is longer and, because of a Pro mutation, kinked. Like the starting scaffold, the esterase has no detectable Diels–Alderase activity.

Although we were unable to obtain cocrystals of the protein with either substrates or product, docking calculations and molecular
**Fig. 2 | Directed evolution and biophysical characterization of DA7.** a, Reaction of 1 (20 µM) with 2 (250 µM) catalysed by zinc (5 µM, grey), DA0 (0.1 µM, magenta) and DA7 (0.1 µM, orange). b, MIDTsc model showing the zinc ion (yellow), three coordinating histidines (grey and blue) and mutations suggested by computation (magenta) or acquired during evolution (orange). c, Michaelis–Menten plots for the DA7-catalysed reaction of 1 and 2. Data points were measured at 12.5, 25, 50, 100 and 200 µM azachalcone. Error bars indicate the s.d. of three biological replicates. d, Catalytic efficiency increased 140,000-fold over the course of evolution. Error bars indicate the s.d. between three biological replicates for DA0 and DA7 and the error of the fit for DA1 and DA4. e, Double logarithmic plot of effective molarity (EM = \( k_{cat}/(K_m K_d) \) versus catalytic proficiency (1/\( K_{cat}\) = \( (k_{cat}/(K_m K_d))/k_{cat,TS} \)) where \( k_{cat,TS} \) is the apparent transition state binding affinity, \( k_{cat} \) is the catalytic rate constant and \( k_{cat,TS} \) is the rate constant for the reaction in the absence of enzyme). DA7 and its precursors are compared with catalytic antibodies (black triangles) and variants of the computationally designed DA_20_00 Diels–Alderase (grey circles). The \( 1/K_{cat} \) values observed for the natural enzymes IccD, AbyU and SpnF, which catalyse intramolecular Diels–Alder reactions, are shown as dashed lines. f, Chiral HPLC analysis of the hetero-Diels–Alder reaction products obtained in the DA7 (orange) and background (grey) reactions. Inset: X-ray structure of the endo (4R,6R)-3,4-dihydro-2H-pyran isomer produced by DA7.

**Dynamics (MD) simulations provided mechanistic insights into DA7 catalysis (Fig. 3b–d).** As in the solution reaction, the protein-bound zinc ion activates the azachalcone by chelation, facilitating attack of the 3-vinylindole to generate a zwitterionic intermediate, which rapidly cyclizes to the hetero-Diels–Alder product. Consistent with the experimentally observed enantioselectivity, only the transition state leading to the endo (4R,6R)-3,4-dihydro-2H-pyran product fits in the binding pocket with strong coordination to Zn(II) and without steric clashes. A restrained MD simulation of the transition-state complex maintains all these interactions. By contrast, the enantiomeric transition state cannot fit into the binding site and still chelate Zn(II). Restrained simulation of the latter reduces steric clashes at the expense of Zn(II) coordination (Extended Data Fig. 6).

In the transition state leading to the observed product, the N-terminal helix–turn–helix fragment positions the azachalcone via numerous van der Waals contacts and a salt bridge between the guanidinium group of Arg28 and the carboxylate of the diene (Fig. 3b). The dienophile binds against the C-terminal helix–turn–helix fragment, placing its indole ring in a small hydrophobic pocket created by the I64G mutation and its vinyl moiety against the si face of 1 (Fig. 3c; Extended Data Fig. 6). The zwitterionic intermediate and flanking transition states are additionally stabilized by a hydrogen bond between the indole NH and the amide side chain of Gln80 (Fig. 3d; Extended Data Fig. 7). This interaction is supported by a small hydrogen-bonding network consisting of Gln31 and Tyr84, a residue that appeared late in evolution (Fig. 3d). The efficiency of DA7 can thus be ascribed to a combination of Lewis acid catalysis, enthalpic stabilization of the transition state by the close-fitting binding pocket, and a strategically placed hydrogen-bonding interaction.

The hypothesis that the carboxylate substituent of 1 forms a salt bridge with Arg28 is supported by the fivefold lower specific activity observed for the unsubstituted azachalcone 5, which cannot make this interaction (Fig. 4). MD simulations of DA7 further suggested that slight displacement of Met87 at the base of the pyridine binding site would enable the binding of larger heterodienes. In accord with this prediction, the enzyme accepts isoquinoline derivative 6 as efficiently as the original azachalcone (Fig. 4). By contrast, chalcone 7, which can only bind Zn(II) in a monodentate fashion, is a 100-fold poorer substrate than the isosteric azachalcone 1 (Fig. 4). Replacement of the indole NH in the dienophile with an oxygen is even more deleterious. 3-Vinylbenzofuran (8), an isosteric analogue of 2, is completely inactive (Fig. 4), highlighting the importance of the hydrogen bond between the dienophile and Gln80. The 23-fold reduction in efficiency observed when Gln80 was replaced with alanine reinforces this conclusion (Extended Data Fig. 4e).

**Conclusions and prospects**

Evolution has provided complete control over the reaction pathway and product stereochemistry of an abiological hetero-Diels–Alder reaction. Since we only screened for azachalcone disappearance, both properties emerged spontaneously as the activity increased. An important challenge for the future will therefore be to see whether or not equally effective catalysts can be engineered for the other possible isomers of 3 and 4. Judging from the product profile observed for DA0, it should be possible to enrich variants capable
of stabilizing the enantiomeric endo hetero-Diels–Alder transition state by monitoring changes in the product distribution directly by HPLC. Although reversing the inherent reactivity of this system to produce disfavoured exo products or normal-electron-demand Diels–Alder adducts will likely prove more demanding, computational methods 27 should help to design suitable libraries that favour such reaction channels.

In the design of any new enzyme, the choice of an appropriate starting scaffold is a fundamental consideration. Most previous attempts at creating such catalysts have focused on recycling pre-existing proteins from nature 27–29. Our results show that evolutionarily naive metalloproteins like MID1sc are attractive alternatives. Despite its relatively small size (97 amino acids) and simple structure, this de novo helical bundle was able to successfully harness Lewis acid catalysis to produce the valuable dihydropyran moiety with high efficiency and selectivity. Other Lewis-acid-catalysed transformations 30 and other de novo scaffolds 31 may prove similarly amenable to this approach, providing a new avenue for designing and evolving biocatalysts of value for a variety of non-natural functions.
Methods
Materials, methods and data characterization for all biological, chemical and computational experiments are described in detail in the Supplementary Information.

Data availability
Structural data obtained by X-ray crystallography were deposited in the Protein Data Bank (PDB) or the Cambridge Crystallographic Data Centre (CCDC), and are available with the following accession codes: DA7W16W (7BWW), DA7W16G,K58Q,L77R,T78R (6YPI), racemic exo-3 (1972193), 4R,6R-endo-4 (1972197) and R-(+)-phenylethylamine phosphate (1972198). All relevant data are provided in the figures, Extended Data and Supplementary Information; alternatively, the data are available from the corresponding author upon request.

Acknowledgements
We thank M. Solar and N. Trapp from the Small Molecule Crystallography Center at ETH Zürich for structure elucidation. We also thank B. Blattmann from the Protein Crystallization Center at the University of Zürich, the teams from the Mass Spectrometry service and the NMR facility at the Laboratory for Organic Chemistry at ETH Zürich and the staff at the Swiss Light Source at the Paul Scherrer Institute for technical support. This work was supported by the Swiss National Science Foundation (SNSF) and ETH Zürich. G.J.-O. acknowledges grants from the Spanish Ministry of Science and Innovation (MCI) co-financed with FEDER funds (RTI2018-099592-B-C22 and RYC-2013-14706) and the Mizutani Foundation for Glycience (200077).

Author contributions
D.H., G.J.-O. and K.N.H. conceived the project. G.J.-O. and K.N.H. carried out the computational enzyme design. H.A.B. carried out preliminary experiments. S.S. designed the experimental strategy. S.S. and A.C. selected and tested the initial constructs. S.B. and S.S. evolved and characterized the enzyme and analysed the data. S.B., Y.O. and R.C.H. synthesized the substrates. S.B. performed the reaction characterization, protein and small molecule crystallization. T.M. crystallized the enzyme and solved its structure. Y.Z. and G.J.-O. performed computational MD simulations and DFT calculations. D.H. supervised the research. D.H., S.B. and K.N.H. wrote the paper with contributions from all the authors.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41557-020-00628-4.
Supplementary information is available for this paper at https://doi.org/10.1038/s41557-020-00628-4.
Correspondence and requests for materials should be addressed to G.J.-O. or D.H.
Peer review information Nature Chemistry thanks the anonymous reviewers for their contribution to the peer review of this work.
Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | DFT calculations for competing Diels-Alder and hetero-Diels-Alder pathways. a, Diels-Alder and hetero-Diels-Alder reactions of the azachalcone 1 and 3-vinylindole 2 in SMD water [calculated at the ωB97X-D/6-311+ +g(2d,p)/SMD(water)//ωB97X-D/6-311G(d,p)/SMD (water) level]. TS1-endo is ambimodal and affords both 3-endo and 4-endo. Experimentally, only 3-exo and 4-endo are observed. b, The endo hetero-Diels-Alder reaction of 1 and 2 in the presence of [Zn(OH)(H2O)2]+, calculated at the ωB97X-D/6-311+ +G(2d,p)/SMD(water)//ωB97X-D/6-311G(d,p)/ SMD(water) level (the LANL2DZ and SDD effective core potentials were used for Zn in geometry optimisation and single energy calculations, respectively), occurs via a stepwise mechanism that favours formation of 4-endo. Relative enthalpies (H) and free energies (G) are shown in blue and red, and are reported in kcal/mol. Labelled distances are in ångströms.
Extended Data Fig. 2 | Chemo- and stereoselectivity. **a**, HPLC chromatogram (at 350 nm) and crystal structure (Supplementary Table 11) of the racemic exo Diels-Alder product 3 (only one enantiomer is shown for clarity). **b**, HPLC chromatogram of the racemic hetero-Diels-Alder products 4 and crystal structure of the enzymatically synthesised (4R,6R)-endo hetero-Diels-Alder product. The hydrogen atoms on the central six-membered ring are shown for the X-ray crystal structures in **a** and **b**. **c**, HPLC analysis of the Diels-Alder (3.36 min) and hetero-Diels-Alder (3.06 min) product profile under different conditions. **d**, Chiral HPLC chromatograms of the hetero-Diels-Alder reaction products from reactions without protein (grey) and reactions catalysed by DA0 (pink) and DA7 (orange). Retention times vary due to different column temperatures (40 °C vs room temperature).
Extended Data Fig. 3 | Directed evolution of DA7. a, Simplified schematic showing the steps of computational design and directed evolution leading to DA7. b, Amino acid alignment of evolutionary parent DA0, evolved variant DA7, and two intermediate variants, DA1 and DA4. Mutations distinguishing the variants are highlighted in orange. The two mutations suggested by computation are shown in pink.
Extended Data Fig. 4 | Kinetic characterisation of buffer-catalysed reaction, intermediate DA variants, and DA7. a, Non-enzymatic background reaction between Diels-Alder substrates 1 and 2 in buffer. The rates were determined in 20 mM MOPS, pH 8, 3.5% DMSO, 10 µM Zn(II), 1 mg/mL BSA at 25 °C. Error bars denote s.d. b–d, Michaelis-Menten plots for the reaction between 1 and 2 catalysed by the parental scaffold DA0 and evolutionary intermediates DA1 and DA4. An analogous plot for DA7 is shown in Fig. 2c in the main text. For DA0 (b), error bars indicate the standard deviation between three biological replicates. e, Relative activities of DA7 and the variants DA7W16S, DA7C35H and DA7Q80A, which were respectively prepared to aid crystallisation and probe mechanism. Mutation of the metal binding site (DA7C35A/H61A/H65A) or removal of zinc with EDTA leads to complete loss of activity.  

\[ k_{\text{obs}} \approx \frac{v_0[\text{E}]_0}{[E][S]} \]

f, Total turnovers for DA7 were determined using 200 µM 1 and 400 µM 2 and 10 nM enzyme. The consumption of azachalcone was corrected for contributions from the background reaction.
Extended Data Fig. 5 | Change in metal binding site and crossing angle. **a, b**, Comparison of the metal binding sites of MID1 (**a**) and DA7 (**b**). MID1 coordinates the Zn(II) ion (yellow sphere) with three histidine residues (grey sticks). A fourth non-coordinating histidine points away from the metal binding site. H35C and H39V mutations (orange sticks) altered the coordination sphere of the catalytic Zn(II) ion during evolution. **c**, Overlay of parental MID1 (grey) and DA7 (orange) shows the decrease in crossing angle of the two helix-turn-helix fragments by > 30°. **d**, Calculated acute crossover angles \( \theta \) between helices H1–H4 of MID1 and DA7w16s.

|        | H1 \( \rightarrow \) 3 | H1 \( \rightarrow \) 4 | H2 \( \rightarrow \) 3 | H2 \( \rightarrow \) 4 |
|--------|----------------------|----------------------|----------------------|----------------------|
| MID1*  | 81                   | 85                   | 87                   | 84                   |
| DA7w16s (A)† | 50               | 53                   | 50                   | 63                   |
| DA7w16s (B)† | 54               | 54                   | 55                   | 65                   |
| DA7w16s (C)† | 50               | 50                   | 45                   | 56                   |
| DA7w16s (D)† | 49               | 51                   | 50                   | 60                   |

*PDB accession number 3V1C; H1 = H1 chain A; P3–A21, H2 = H2 chain A; M24–F42, H3 = H1 chain B; P3–A21, H4 = H2 chain B; M24–F42. †PDB accession number 7BWW; H1, P3–A21, H2; M24–F42, H3; P52–A70, H4; L73–F91. Acute crossing angles were calculated according to Der et al.18.
Extended Data Fig. 6 | Discrimination of enantiomeric transition states by DA7. a, Whereas the equilibrated rate-limiting transition state leading to the experimentally observed (4R, 6R)-hetero-Diels-Alder product binds productively at the DA7 active site (TS1-endo, green; see also Fig. 3 in the main text), severe steric clashes between the indole moiety of the enantiomeric TS1-endo (yellow sticks) and the protein backbone are observed if the azachalcone moiety chelates the catalytic zinc ion. b, Restrained molecular dynamics simulations show that the clashes between the enantiomeric TS1-endo and DA7 are only relieved at the expense of zinc coordination, explaining why the evolved enzyme does not produce the (4S, 6S)-hetero-Diels-Alder product. Labelled distances are in angstrom. c, d, Surface representation of DA7 showing the binding pocket with the preferred transition state docked at the active site in stick (c) and space-filling (d) representation.
Extended Data Fig. 7 | A catalytically relevant hydrogen bond network in DA7. Molecular dynamics simulations of INT1-endo bound to DA7 show fluctuations in hydrogen bond distances between a, the 3-vinylindole moiety of the ligand and Gln80, b, Gln80 and Gln31, and c, Gln31 and Tyr84. Gln80 maintains hydrogen bonding interactions with both the ligand and Gln31 throughout the simulation, whereas Gln31 and Tyr84 interact for fraction of the time.