Bunzel, A. A., Anderson, J. L. R., Hilvert, D., Arcus, V. L., Van Der Kamp, M. W., & Mulholland, A. J. (2021). Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nature Chemistry*, 13(10), 1017-1022. <https://doi.org/10.1038/s41557-021-00763-6>

Peer reviewed version

Link to published version (if available):
[10.1038/s41557-021-00763-6](https://doi.org/10.1038/s41557-021-00763-6)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Nature Research at <https://doi.org/10.1038/s41557-021-00763-6> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Evolution of dynamical networks enhances catalysis in a designer enzyme

H. Adrian Bunzel^{1,2,*}, J. L. Ross Anderson¹, Donald Hilvert³, Vickery L. Arcus⁴, Marc W. van der Kamp^{1,2}, Adrian J. Mulholland^{2,*}

¹ *School of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD, UK*

² *Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK*

³ *Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland*

⁴ *School of Science, University of Waikato, Hamilton, 3216, New Zealand*

**Corresponding author: Adrian.Bunzel@bristol.ac.uk, Adrian.Mulholland@bristol.ac.uk*

Abstract

Activation heat capacity is emerging as a crucial factor in enzyme thermoadaptation, as shown by the non-Arrhenius behaviour of many natural enzymes. However, its physical origin and relationship to the evolution of catalytic activity remain uncertain. Here we show that directed evolution of a computationally designed Kemp eliminase reshapes protein dynamics, which gives rise to an activation heat capacity absent in the original design. These changes buttress transition-state stabilization. Extensive molecular dynamics simulations show that evolution results in the closure of solvent-exposed loops and a better packing of the active site. Remarkably, this gives rise to a correlated dynamical network that involves the transition state and large parts of the protein. This network tightens the transition-state ensemble, which induces a negative activation heat capacity and non-linearity in the activity-temperature dependence. Our results have implications for understanding enzyme evolution and suggest that selectively targeting the conformational dynamics of the transition-state ensemble by design and evolution will expedite the creation of novel enzymes.

Tailor-made enzymes promise to be transformative for the ‘green’ synthesis of pharmaceuticals and fine chemicals, and the achievement of a circular economy¹⁻³. Increasingly, the de novo computational design of catalytic residues and ligand-binding pockets into protein scaffolds can afford incipient, yet generally modest, activity^{4,5}. Such ‘designer enzymes’ can be substantially improved by directed evolution, which may enhance activity by several orders of magnitude in the best cases⁶⁻¹¹. This is now an effective way to develop new protein catalysts, but also demonstrates the limitations of current design protocols⁶. Designed active sites often require fine-tuning by evolution to precisely position catalytic residues and tightly packed ligands⁷⁻¹¹, increase electrostatic preorganization¹²⁻¹⁵, and stabilize catalytically superior conformations¹⁶⁻²³. What is less clear is whether evolution exploits the scaffold’s dynamics to introduce a differential thermodynamic response to the transition state (TS).

Evolution of a negative activation heat capacity

The base-promoted Kemp elimination of benzisoxazoles is a valuable model for the catalysis of an elementary reaction^{24,25}, and has become an exemplar for the designed biological catalysis of a non-natural reaction^{9-11,26,27}. We previously subjected the computationally designed Kemp eliminase 1A53-2 to directed evolution (Fig. 1)¹⁰. Evolution boosted activity 10⁴-fold with the introduction of six mutations during optimization of the first-shell residues only. The binding pocket of the resulting evolved variant, 1A53-2.5, exhibited improved shape complementarity to a transition-state analogue by virtue of several space-filling substitutions¹⁰. The introduction of the A157Y mutation furthermore enhanced activity by restricting the conformational freedom and tuning the pK_a of the catalytic base Glu178.

Fig. 1 | Activity gains during directed evolution of 1A32-2 coincided with the emergence of a negative $\Delta C_p^\ddagger$. **a**, The general base Glu178 deprotonates 6-nitrobenzisoxazole. **b**, Six mutations (blue spheres) were introduced during active site optimization (ligand and base, orange carbons)¹⁰. **c**, The Eyring plot of 1A53-2 (red) is linear, whereas that of the evolved variant 1A53-2.5 (blue) is curved, signalling the emergence of a negative $\Delta C_p^\ddagger$ ¹⁰. Dashed lines indicate a linear Eyring fit, solid lines show the macromolecular rate theory fit including $\Delta C_p^\ddagger$ ²⁹. Data extracted from ref. ¹⁰.

Unexpectedly, activity gains in 1A53-2.5 coincided with the emergence of a curved activity-temperature dependence (Fig. 1c)¹⁰. This curvature is not due to protein unfolding or a change in mechanism (Extended Data Fig. 1)¹⁰. Instead, it can be attributed to an apparent activation heat capacity ($\Delta C_p^\ddagger$, Equation 1) of the chemical step. There is increasing evidence that $\Delta C_p^\ddagger$ is linked to thermoadaptation in natural enzymes due to its effect on an enzyme's temperature optimum (Equation 2)^{28,29}. Notably, directed evolution of 1A53-2 was solely aimed to increase overall activity. The unexpected emergence of a negative $\Delta C_p^\ddagger$ suggests that this effect is more generally related to catalytic proficiency. A negative $\Delta C_p^\ddagger$ indicates that energy fluctuations ($\Delta\langle\delta H^2\rangle^\ddagger$) are reduced in the transition-state ensemble compared with those of the enzyme-substrate complex (Equation 3)³⁰, which potentially strengthens recognition of the chemical TS. Molecular

dynamics (MD) simulations can probe this thermodynamic effect and have yielded $\Delta C_p^\ddagger$ values in good agreement with experiment for two natural enzymes²⁸.

$$\ln\left(\frac{kh}{k_B T}\right) = -\frac{[\Delta H_{T_0}^\ddagger + \Delta C_p^\ddagger(T - T_0)]}{RT} + \frac{[\Delta S_{T_0}^\ddagger + \Delta C_p^\ddagger \ln(T/T_0)]}{R} \quad (1)$$

$$T_{\text{opt}} \approx T_0 - \frac{\Delta H_{T_0}^\ddagger}{\Delta C_p^\ddagger} \quad (2)$$

$$\Delta C_p^\ddagger = \frac{\delta(\Delta H^\ddagger)}{\delta T} = \frac{\Delta(\delta H^2)^\ddagger}{k_B T^2} \quad (3)$$

(k_B : Boltzmann constant, T : Temperature, h : Planck constant, $\Delta H_{T_0}^\ddagger$ and $\Delta S_{T_0}^\ddagger$ activation enthalpy and entropy at the reference temperature T_0 , T_{opt} : Temperature optimum)²⁹.

Here, we show that the evolution of 1A53-2 reshaped its conformational dynamics, which drove the emergence of a negative $\Delta C_p^\ddagger$. Strikingly, emergent correlated movements in the transition-state ensemble of 1A53-2.5 span the protein and link the active site with the protein scaffold through a dynamical network. This network enhances catalytic preorganization and helps to explain the role of distal mutations that arose during further evolution.

Energetic fluctuations

We performed 5 μ s of MD simulations each for the designed and evolved variants in complex with either the substrate (Michaelis complex/ground state (GS)) or TS (Extended Data Fig. 2+3). The designed and evolved variants showed significantly different dynamical responses to the TS. The energy distribution of 1A53-2.5 was notably narrower in the TS than in the GS. In contrast, no significant differences between the GS and TS were observed for 1A53-2 (Extended Data Fig. 4a+b, Supplementary Tab. 1+2).

The change in energy fluctuations between GS and TS determines $\Delta C_p^\ddagger$, which we calculate from a moving average over the simulations using validated methods (Extended Data Fig. 4c)²⁸. Rigidification of the TS ensemble in 1A53-2.5 results in a negative $\Delta C_p^\ddagger$ ($-19.7 \pm 3.1 \text{ kJ}\cdot\text{mol}^{-1}\text{K}^{-1}$), whereas the activation heat capacity of 1A53-2 is zero within error ($0.4 \pm 3.2 \text{ kJ}\cdot\text{mol}^{-1}\text{K}^{-1}$). This is consistent with the experimentally observed linear and non-linear temperature dependence for the designed and evolved catalysts, respectively, and demonstrates the dynamical origin of $\Delta C_p^\ddagger$.

The differences between the designed and evolved catalysts are significant, as demonstrated by the errors for $\Delta C_p^\ddagger$ that were calculated by cross-validation leaving out individual trajectories. Independent simulations using a different TS model in which the transferring proton resides on the base instead of the ligand gave similar results (TS2, see Extended Figures 2-9). We calculate $\Delta C_p^\ddagger$ directly from the difference in energy fluctuations between GS and TS, after removal of the bulk solvent. Our calculations predict a sizable negative $\Delta C_p^\ddagger$ for the evolved protein; the magnitude is overestimated in part because protein solvation reduces $\Delta C_p^\ddagger$ substantially²⁸. Also, slow loop movements, which cannot be exhaustively sampled on a reasonable timescale, may mix into the fluctuations at longer window sizes. In short, comparison of the designed and the evolved catalysts shows that the non-linear temperature dependence of the evolved variant arises from dynamical differences between its GS and the TS complex.

Structural fluctuations

The decrease in energy fluctuations associated with reaction in 1A53-2.5 is accompanied by reduced structural fluctuations in the TS ensemble (Fig. 2 and Extended

Data Fig. 5). In 1A53-2.5, several solvent-exposed loops that cover the active site (residues 53-65, 84-92, and 181-192) become less mobile in the TS ensemble, as indicated by decreased root-mean-square fluctuations (ΔRMSF) between the GS and TS. Cluster analysis (Fig. 2a and Extended Data Fig. 6) and principal component analysis (Extended Data Fig. 7) show that these loops interconvert between an open and a closed state. Both 1A53-2 and 1A53-2.5 both have narrow substrate entry channels with bottleneck radii of 1.8 Å in their X-ray structures³¹. Although loop opening will facilitate ligand binding, several open-closed interconversions were observed for both variants (Extended Data Fig. 6 i+j), which suggests that this motion is faster than turnover.

Loop closure is essential for catalysis in natural enzymes, such as triosephosphate and ketosteroid isomerase³²⁻³⁴, which catalyse proton transfer reactions comparable with the Kemp elimination^{24,35,36}. In these natural enzymes, loop closure desolvates the active site and improves preorganization^{32,33}. Directed evolution of 1A53-2 similarly resulted in tight active-site packing shown by a decrease in solvent-accessible surface area of the ligand and base (Extended Data Fig. 8)¹⁰. Evolution (1) introduced space-filling mutations, such as A157Y and L184F leading to better active site packing, (2) removed water from the active site on loop closure and (3) shifted the open-closed equilibrium towards the closed state. Loop closure in the evolved variant furthermore results in a decrease of the protein radius of gyration (Extended Data Fig. 6), which suggests that pressure may also modulate fluctuations in 1A53-2.5³⁷. These effects – combined with the introduced hydrogen bond to Glu178 from A157Y – restrict the movement between the ligand and base (Extended Data Fig. 8) and preorganize the reactants for catalysis.

Fig. 2. Global dynamics in the evolved enzyme are modulated by a local tightening of the active site associated with reaction. **a**, Distance-based cluster analysis of the closing loops (tubes) in 1A53-2 (red) and 1A53-2.5 (blue) reveals a conformational equilibrium between open (light colours) and closed (dark colours) states, as indicated e.g. by the C_{α} distance between residues 58 and 188 (spheres). The closed state becomes more populated after evolution (see Extended Data Fig. 6). **b**, Evolution introduces six mutations at the active site (red/blue sticks, G211 as sphere), which enhance the interactions of Trp110, Glu178, Arg182 and Trp210 (grey sticks) with the TS in the closed state. As a result, the closed TS ensemble of 1A53-2.5 is more constrained than its GS ensemble, as reflected by reduced fluctuations (Δ RMSF) of many active site residues between the two states.

To understand how evolution affected TS recognition, interaction energies of individual residues with the TS were calculated (Fig. 2b). After evolution, several active-site residues interact more strongly with the TS. Notably, mutations introduced during

evolution do not directly provide additional TS stabilization, but appear to indirectly strengthen interactions of the TS with the catalytic base Glu178, with Trp110 and Trp210 sandwiching the ligand, and with Arg182 stabilising the nascent oxyanion by electrostatic interactions. This tightening is reflected by the ΔRMSF of most active site residues, which are less mobile in the TS than in the GS ensemble in 1A53-2.5 (Fig. 2b). These tight and ordered interactions enhance active-site preorganization; we show here that they are associated with the overall rigidification of the protein scaffold in the TS ensemble that causes the negative $\Delta C_p^\ddagger$.

Evolution of a dynamical network

Dynamical correlations in the evolved TS ensemble connect local reorganization of the active site with the protein scaffold. The change in cross-correlations between the GS and TS was calculated for all backbone C_α atoms and the ligand, with the protein in the closed state (Fig. 3 and Extended Data Fig. 9). These correlations therefore reflect differences between the GS and TS, and do not involve conformational changes between the open and closed states. In the closed state of 1A53-2.5, large parts of the backbone move in a more correlated manner in the TS compared with that in the GS ensemble. Shortest path maps²² show that these increased correlations are communicated via neighbouring residues and indicate that evolution introduced a dynamical network that centres on the ligand and spans the protein (Fig. 3 and Extended Data Fig. 9). Only two mutations introduced during evolution are directly involved in the network. Q211G increases flexibility and potentially tunes the dynamic response of the scaffold, whereas L184F enhances packing by connecting neighbouring solvent-exposed loops. The other evolved mutations apparently facilitate the global dynamical response to the TS by tightening its interactions with

native active site-residues (Trp110, Glu178, Arg182 and Trp210, as described above), rather than being part of the network.

Fig. 3. | Directed evolution introduces an extended correlated network in the transition state ensemble. a+b, Evolution of 1A53-2 (red, left) to 1A53-2.5 (blue, middle) increases the correlated movements in the closed TS ensemble. Cross correlations that increase between GS and TS by $\geq 20\%$ are indicated as lines on the structures (see Extended Data Fig. 9). **c+d,** Shortest path maps²² calculated from the increased cross-correlations demonstrate evolution of a network centred on the chemical TS that spans the closed state of the protein. The sizes of the edges (black lines) and vertices (red/blue spheres: protein, orange spheres: ligand) indicate the strength of the network. **e+f,** The mutations in 1A53-2.5 (blue spheres), those found during further evolution using error-prone PCR (white spheres), and those in the final variant 1A53-2.9 (black spheres) cluster around the correlated movements and networks (yellow mesh) in 1A53-2.5, indicating fine-tuning of its dynamics by remote mutations.

1A53-2.5 was evolved by targeting only first-shell residues. Several distal mutations were identified during its further optimization by error-prone PCR, which provided an

additional threefold activity improvement¹⁰. These mutations cluster around the differential cross-correlations in 1A53-2.5 (Fig. 3e+f), which suggests that evolution altered the dynamical network while slightly increasing activity. The limited improvements achieved during these last rounds of evolution indicate that further fine-tuning would require many more mutations to reprogram the global dynamics encoded in the scaffold. Although this may be challenging for laboratory evolution, nature probably evolved similar networks as indicated by sectors of spatially proximal coevolving residues³⁸⁻⁴⁰ with highly correlated dynamics⁴¹. Dynamical networks^{22,42-46} probably play a crucial role in enhancing preorganization, reducing non-productive conformations^{16-18,22}, and tuning protein flexibility for thermoadaptation^{28,29,47}. In contrast to other methods that often focus on a single state only, our differential analysis directly highlights networks involved in the protein response to the chemical TS. Targeting that response – by mutagenesis or computational redesign of key network vertices – may provide promising opportunities to tailor enzyme activity.

Implications for catalysis

The co-emergence of catalytic power and a negative $\Delta C_p^\ddagger$ raises the intriguing question of the connection between TS tightening (and the associated dynamical network) and catalysis⁴⁸. Although deciphering this cause-and-effect relationship will require further research, exquisite catalytic preorganization certainly requires ordering at the active site¹⁴. The evolved 1A53-2.5 achieves improved catalysis by increased TS stabilization, and its negative $\Delta C_p^\ddagger$ shows that the protein scaffold responds to the chemical TS with significantly altered dynamics. We note that protein dynamics do not ‘drive’ the reaction on the timescale of the chemical bond making and breaking. This is shown by our QM/MM umbrella sampling simulations that reproduce the improvements achieved during

evolution within a TS theory framework (Extended Data Fig. 10). Instead, effective TS stabilization by exquisite active-site preorganization is a function of the cooperative response of the whole protein scaffold. The transition from a flexible GS to a rigid TS could thus lead to activity gains, albeit at the expense of introducing curved temperature dependencies and therefore a loss of activity at elevated temperatures. $\Delta C_p^\ddagger$ is apparently a signature of the superior catalytic efficiency achieved by the more ‘enzyme-like’ 1A53-2.5. Furthermore, $\Delta C_p^\ddagger$ signals the ability of the scaffold to adapt to the chemical TS, and may be an important feature for enzyme evolvability.

Design opportunities

The dynamical network in 1A53-2.5 emerged on introducing only a few active-site mutations, at positions that are typically targeted by current design algorithms. Supplementing conventional design and directed evolution approaches with MD analysis of the most promising candidates will probably allow reliable *in silico* screening for catalytically superior variants. Specifically, we propose to analyse the dynamical differences between the TS and GS ensemble to identify variants that tightly recognize the chemical TS. The dynamical analysis presented here can be performed within days to weeks, on timescales compatible with practical engineering strategies in the laboratory. The emergent TS networks may also be exploited to guide the evolution of distal mutations by suggesting mutagenesis hotspots and targets for computational redesign⁴⁹. Rapid approximate methods to model protein dynamics^{50,51} could, furthermore, reveal fold-specific dynamics to inform design and evolution⁵²⁻⁵⁴. Combining existing design algorithms with atomistically detailed MD²⁸ simulations may help attain tight communication with the scaffold’s dynamics and the exquisite precision required to achieve excellent TS complementarity and catalytic preorganization.

Conclusion

Our MD simulations and statistical thermodynamic analysis demonstrate how the evolution of catalytic activity shapes protein dynamics. The catalytically superior evolved 1A53-2.5 uniquely shows a correlated network and reduced fluctuations in the TS ensemble, which gives rise to a negative $\Delta C_p^\ddagger$. Our findings suggest a connection between the cooperative response of the protein scaffold and TS stabilization at the active site. The results altogether imply that $\Delta C_p^\ddagger$ may be an indicator of successful catalytic preorganization, and perhaps relates to evolvability. Enzyme catalysis is based on selective TS stabilization; targeting the scaffold's dynamical response to the chemical transition state by design and evolution may be essential to create artificial enzymes that truly rival their natural counterparts.

Methods

Methods for all computational experiments are described in detail in the Supplementary Information.

AUTHOR INFORMATION

Corresponding Author

* Adrian.Bunzel@bristol.ac.uk

*Adrian.Mulholland@bristol.ac.uk

ORCID

H. Adrian Bunzel: 0000-0001-6427-368X

Ross Anderson: 0000-0002-6796-0482

Donald Hilvert: 0000-0002-3941-621X

Vickery L. Arcus: 0000-0001-5082-2414

Marc W. van der Kamp: 0000-0002-8060-3359

Adrian J. Mulholland: 0000-0003-1015-4567

COMPETING INTERESTS

The authors declare no competing financial interests.

ACKNOWLEDGEMENTS

HAB and AJM thank EPSRC (EP/M013219/1 and EP/M022609/1) and with JLRA BBSRC (BB/R016445/1) for funding. HAB thanks the Swiss National Science Foundation (Postdoc.Mobility fellowship) for support. MWvdK is a BBSRC David Phillips Fellow (BB/M026280/1). VLA and AJM thank the Marsden Fund of New Zealand (16-UOW-027). VLA is a James Cook Research Fellow (Royal Society of New Zealand). DH thanks the Swiss National Science Foundation. This work was conducted using the computational facilities of the Advanced Computing Research Centre, University of Bristol. We thank Rory Crean and Silvia Osuna for help with and providing a script for performing the shortest-path analysis.

AUTHOR CONTRIBUTIONS STATEMENT

HAB, MWvdK and AJM devised the simulation and analysis. HAB performed the simulations and analysis. HAB, RA, DH, VLA, MWvdK, and AJM wrote the manuscript.

DATA AVAILABILITY

MD simulation raw data required to reproduce this work are available from the authors upon reasonable request.

REFERENCES

- 1 Arnold, F. H. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew. Chem. Int. Ed.* **58**, 14420-14426 (2019).
- 2 Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185-194 (2012).
- 3 Tournier, V. *et al.* An engineered PET depolymerase to break down and recycle plastic bottles. *Nature* **580**, 216-219 (2020).
- 4 Kries, H., Blomberg, R. & Hilvert, D. De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **17**, 221-228 (2013).
- 5 Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed.* **52**, 5700-5725 (2013).
- 6 Bunzel, H. A., Anderson, J. L. R. & Mulholland, A. J. Designing better enzymes: Insights from directed evolution. *Curr. Opin. Struct. Biol.* **67**, 212-218 (2021).
- 7 Obexer, R. *et al.* Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50-56 (2017).
- 8 Preiswerk, N. *et al.* Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8013-8018 (2014).
- 9 Blomberg, R. *et al.* Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418-421 (2013).
- 10 Bunzel, H. A. *et al.* Emergence of a negative activation heat capacity during evolution of a computationally designed enzyme. *J. Am. Chem. Soc.* **141**, 11745-11748 (2019).
- 11 Khersonsky, O. *et al.* Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10358-10363 (2012).
- 12 Fuxreiter, M. & Mones, L. The role of reorganization energy in rational enzyme design. *Curr. Opin. Chem. Biol.* **21**, 34-41 (2014).
- 13 Jindal, G., Ramachandran, B., Bora, R. P. & Warshel, A. Exploring the development of ground-state destabilization and transition-state stabilization in two directed evolution paths of Kemp eliminases. *ACS Catal.* **7**, 3301-3305 (2017).
- 14 Warshel, A. *et al.* Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210-3235 (2006).
- 15 Bhowmick, A., Sharma, S. C. & Head-Gordon, T. The importance of the scaffold for de novo enzymes: A case study with Kemp eliminase. *J. Am. Chem. Soc.* **139**, 5793-5800 (2017).
- 16 Hong, N. S. *et al.* The evolution of multiple active site configurations in a designed enzyme. *Nat. Commun.* **9**, 3900 (2018).
- 17 Bhowmick, A., Sharma, S. C., Honma, H. & Head-Gordon, T. The role of side chain entropy and mutual information for improving the de novo design of Kemp eliminases KE07 and KE70. *Phys. Chem. Chem. Phys.* **18**, 19386-19396 (2016).
- 18 Petrovic, D., Risso, V. A., Kamerlin, S. C. L. & Sanchez-Ruiz, J. M. Conformational dynamics and enzyme evolution. *J. R. Soc. Interface* **15** (2018).
- 19 Campbell, E. C. *et al.* Laboratory evolution of protein conformational dynamics. *Curr. Opin. Struct. Biol.* **50**, 49-57 (2018).
- 20 Otten, R. *et al.* How directed evolution reshapes energy landscapes to boost enzyme catalysis. *Science (New York, N.Y.)* **370**, 1442-1446 (2020).
- 21 Broom, A. *et al.* Evolution of an enzyme conformational ensemble guides design of an efficient biocatalyst. *Nat. Commun.* **11**, 4808 (2020).

- 22 Romero-Rivera, A., Garcia-Borras, M. & Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **7**, 8524-8532 (2017).
- 23 Maria-Solano, M. A., Serrano-Hervás, E., Romero-Rivera, A., Iglesias-Fernández, J. & Osuna, S. Role of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun.* **54**, 6622-6634 (2018).
- 24 Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. Physical organic chemistry of benzisoxazoles. I. Mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.* **38**, 2294-2301 (1973).
- 25 Kemp, D. S. & Casey, M. L. Physical organic chemistry of benzisoxazoles. II. Linearity of the Broensted free energy relation for the base-catalyzed decomposition of benzisoxazoles. *J. Am. Chem. Soc.* **95**, 6670-6680 (1973).
- 26 Privett, H. K. *et al.* Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3790-3795 (2012).
- 27 Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008).
- 28 van der Kamp, M. W. *et al.* Dynamical origins of heat capacity changes in enzyme-catalysed reactions. *Nat. Commun.* **9**, 1177 (2018).
- 29 Arcus, V. L. *et al.* On the temperature dependence of enzyme-catalyzed rates. *Biochemistry* **55**, 1681-1688 (2016).
- 30 Prabhu, N. V. & Sharp, K. A. Heat capacity in proteins. *Annu. Rev. Phys. Chem.* **56**, 521-548 (2005).
- 31 Stourac, J. *et al.* Caver Web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Res.* **47**, W414-W422 (2019).
- 32 Liao, Q. *et al.* Loop Motion in Triosephosphate Isomerase Is Not a Simple Open and Shut Case. *J. Am. Chem. Soc.* **140**, 15889-15903 (2018).
- 33 van der Kamp, M. W., Chaudret, R. & Mulholland, A. J. QM/MM modelling of ketosteroid isomerase reactivity indicates that active site closure is integral to catalysis. *FEBS J.* **280**, 3120-3131 (2013).
- 34 Malabanan, M. M., Amyes, T. L. & Richard, J. P. A role for flexible loops in enzyme catalysis. *Curr. Opin. Struct. Biol.* **20**, 702-710 (2010).
- 35 Hall, A. & Knowles, J. R. Uncatalyzed rates of enolization of dihydroxyacetone phosphate and of glyceraldehyde 3-phosphate in neutral aqueous solution. Quantitative assessment of the effectiveness of an enzyme catalyst. *Biochemistry* **14**, 4348-4352 (1975).
- 36 Houck, W. J. & Pollack, R. M. Activation enthalpies and entropies for the microscopic rate constants of acetate-catalyzed isomerization of 5-androstene-3,17-dione. *J. Am. Chem. Soc.* **125**, 10206-10212 (2003).
- 37 Jones, H. B. L. *et al.* A complete thermodynamic analysis of enzyme turnover links the free energy landscape to enzyme catalysis. *FEBS J.* **284**, 2829-2842 (2017).
- 38 Rivoire, O., Reynolds, K. A. & Ranganathan, R. Evolution-based functional decomposition of proteins. *PLoS Comput. Biol.* **12**, e1004817 (2016).
- 39 Reynolds, K. A., McLaughlin, R. N. & Ranganathan, R. Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564-1575 (2011).
- 40 Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774-786 (2009).
- 41 Lakhani, B., Thayer, K. M., Black, E. & Beveridge, D. L. Spectral analysis of molecular dynamics simulations on PDZ: MD sectors. *J. Biomol. Struct. Dyn.*, 1-10 (2019).
- 42 Sethi, A., Eargle, J., Black, A. A. & Luthey-Schulten, Z. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6620-6625 (2009).

- 43 Rivalta, I. *et al.* Allosteric pathways in imidazole glycerol phosphate synthase. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1428-1436 (2012).
- 44 Goodey, N. M. & Benkovic, S. J. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **4**, 474-482 (2008).
- 45 Melo, M. C. R., Bernardi, R. C., de la Fuente-Nunez, C. & Luthey-Schulten, Z. Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *J. Chem. Phys.* **153**, 134104 (2020).
- 46 Singh, S. & Bowman, G. R. Quantifying Allosteric Communication via Both Concerted Structural Changes and Conformational Disorder with CARDS. *J. Chem. Theory Comput.* **13**, 1509-1517 (2017).
- 47 Åqvist, J., Kazemi, M., Isaksen, G. V. & Brandsdal, B. O. Entropy and enzyme catalysis. *Acc. Chem. Res.* **50**, 199-207 (2017).
- 48 Williams, D. H., Stephens, E., O'Brien, D. P. & Zhou, M. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew. Chem. Int. Ed.* **43**, 6596-6616 (2004).
- 49 Osuna, S. The challenge of predicting distal active site mutations in computational enzyme design. *WIREs Comput. Mol. Sci.*, e1502 (2020).
- 50 Tiwari, S. P. & Reuter, N. Conservation of intrinsic dynamics in proteins-what have computational models taught us? *Curr. Opin. Struct. Biol.* **50**, 75-81 (2018).
- 51 Wells, S. A., van der Kamp, M. W., McGeagh, J. D. & Mulholland, A. J. Structure and Function in Homodimeric Enzymes: Simulations of Cooperative and Independent Functional Motions. *PloS one* **10**, e0133372 (2015).
- 52 Zhang, S., Li, H., Krieger, J. M. & Bahar, I. Shared Signature Dynamics Tempered by Local Fluctuations Enables Fold Adaptability and Specificity. *Mol. Biol. Evol.* **36**, 2053-2068 (2019).
- 53 Fuglebakk, E., Echave, J. & Reuter, N. Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics* **28**, 2431-2440 (2012).
- 54 Echave, J., Spielman, S. J. & Wilke, C. O. Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **17**, 109-121 (2016).

Graphical Abstract

Computationally designed enzymes can be substantially improved by directed evolution. Here, we show that evolution introduces a dynamical network that selectively tightens the transition state ensemble, giving rise to a negative activation heat capacity. Targeting such transition state conformational dynamics may expedite *de novo* enzyme creation.

Supplementary Table 1 | Variances and $\Delta C_p^\ddagger$ of 1A53-2.^a

	Moving Window (ns)															
	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	1A53-2 : GS : Variance (kJ ² ·mol ⁻²)															
1	50910	57655	61152	62914	63994	64789	65588	66334	67011	67640	68196	68760	69337	69846	70341	70810
2	51915	59229	63796	66724	69029	71070	72799	74376	75848	77282	78699	80111	81436	82787	84152	85453
3	52809	61675	66191	69640	71983	73482	74545	75547	76371	76951	77359	77704	78039	78427	78796	79030
4	51217	59484	64315	67783	70043	71676	73080	74268	75405	76501	77454	78334	79206	80120	81031	81884
5	47598	52665	55491	57261	58389	59334	60130	60918	61544	62040	62443	62742	63040	63376	63749	64145
6	48169	56226	61256	64799	67675	70269	72691	74894	76905	78809	80624	82394	84116	85806	87437	88954
7	50943	57331	60427	62257	63658	64896	65835	66591	67241	67817	68357	68805	69217	69620	69958	70208
8	51756	58140	61448	64017	66066	67834	69472	70874	72066	73082	73971	74723	75405	76004	76484	76960
9	48671	56364	60778	63831	66315	68512	70554	72185	73412	74388	75296	76192	77043	77861	78598	79310
10	51151	58329	62287	64308	65603	66617	67582	68490	69353	70080	70729	71205	71483	71775	72071	72243
Average	50514	57710	61714	64353	66275	67848	69228	70448	71516	72459	73313	74097	74832	75562	76262	76900
	1A53-2 : TS : Variance (kJ ² ·mol ⁻²)															
	50084	57964	63380	67196	70074	72196	73766	74863	75488	75918	76332	76903	77631	78431	79175	79809
	49238	57735	61656	63633	64893	65948	66928	67801	68644	69389	70054	70597	70970	71261	71403	71510
3	48644	53988	56753	58618	60137	61462	62607	63649	64604	65381	65962	66449	66831	67069	67201	67311
4	49172	57299	62168	65496	68197	70415	72204	73685	74883	75841	76479	76823	77037	77178	77305	77473
5	47174	54805	58768	61587	63754	65243	66231	67077	67773	68414	68920	69320	69616	69786	69917	70065
6	50956	58731	63370	66687	69442	71728	73633	75243	76295	77047	77541	77982	78367	78686	78998	79306
7	49302	54705	57171	58658	59751	60809	61708	62468	63210	63894	64564	65179	65732	66260	66749	67258
8	49962	58007	62890	66632	69400	71252	72387	73332	74481	75817	77171	78459	79563	80605	81529	82165
9	50341	57941	63019	66479	68663	69968	70943	72078	73427	74763	75973	77099	78280	79433	80557	81555
10	51625	59497	64449	67757	70363	72481	74276	75958	77454	78758	80000	81190	82362	83534	84667	85761
Average	49650	57067	61362	64274	66467	68150	69468	70615	71626	72522	73300	74000	74639	75224	75750	76221
	1A53-2 : TS2 : Variance (kJ ² ·mol ⁻²)															
	52987	62195	67280	70491	72911	74978	76726	78175	79594	80946	82404	83960	85533	87091	88556	89867
	46339	52620	56149	58257	60013	61683	63181	64462	65682	66821	67704	68508	69324	70166	70982	71797
3	47702	55226	59193	61710	63467	64795	65903	66979	68006	69000	70087	71101	72087	73129	74202	75301
4	55094	61686	65016	67325	69009	70109	70902	71400	71857	72362	72864	73402	73953	74338	74735	75121
5	50878	58694	62870	65436	67364	69019	70596	72041	73137	74209	75350	76366	77314	78160	78831	79433
6	48514	54373	57097	58684	60221	61570	62820	63999	65077	66057	66977	67888	68769	69577	70398	71269
7	48849	55733	59649	62575	64627	66032	66940	67763	68577	69283	69817	70275	70728	71194	71672	72109
8	51307	59505	63619	67044	69891	72138	73326	74230	74987	75794	76619	77409	78238	79123	79833	80379
9	47717	55057	60650	65012	68520	71692	74637	77210	79571	82008	84460	86856	89070	91171	93188	95191
10	51667	58007	61288	64023	66464	68554	70378	72021	73498	74834	76103	77176	78044	78826	79463	79985
Average	50105	57310	61281	64056	66249	68057	69541	70828	71999	73132	74239	75294	76306	77277	78186	79045
	1A53-2 : TS – GS : $\Delta C_p^\ddagger$ (kJ·mol ⁻¹ ·K ⁻¹)															
	-864	-643	-352	-79	192	302	241	168	110	63	-13	-97	-193	-338	-512	-679
	$\Delta C_p^\ddagger$	-4.9	-3.6	-2.0	-0.4	1.1	1.7	1.4	0.9	0.6	0.4	-0.1	-0.6	-1.1	-1.9	-2.9
$\Delta\Delta C_p^\ddagger$	0.9	1.3	1.7	2.1	2.4	2.6	2.7	2.9	3.1	3.2	3.3	3.5	3.7	3.9	4.1	4.3
	1A53-2 : TS2 – GS : $\Delta C_p^\ddagger$ (kJ·mol ⁻¹ ·K ⁻¹)															
	-408	-400	-433	-298	-27	209	313	380	483	673	926	1197	1474	1715	1924	2146
	$\Delta C_p^\ddagger$	-2.3	-2.3	-2.5	-1.7	-0.2	1.2	1.8	2.2	2.7	3.8	5.2	6.8	8.3	9.7	10.9
$\Delta\Delta C_p^\ddagger$	1.4	1.7	1.9	2.2	2.4	2.6	2.7	2.9	3.1	3.3	3.5	3.7	4.0	4.2	4.4	4.7

^a Variances and heat capacities calculate for different moving average windows, based on complexes comprising the protein, ligand, and ten water molecules closest to Glu178.

Supplementary Table 2 | Variances and $\Delta C_p^\ddagger$ for 1A53-2.5.^a

	Moving Window (ns)															
	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	1A53-2.5 : GS : Variance (kJ ² ·mol ⁻²)															
1	47802	54132	57931	60672	63142	65404	67296	68901	70238	71456	72619	73668	74624	75549	76490	77428
2	50645	59588	64984	68618	71527	73730	75063	75872	76323	76640	76893	77107	77268	77408	77639	77955
3	52923	61034	65056	67862	70284	72278	73872	75245	76515	77697	78814	79753	80639	81431	82198	83008
4	47297	53118	56709	59841	62750	65254	67074	68173	68930	69396	69735	70095	70497	70960	71497	72028
5	47567	54338	58478	61619	64406	66912	69196	71199	73108	74890	76560	78118	79545	80833	81838	82725
6	47155	53959	57291	59536	61334	62803	64099	65243	66262	67152	68025	68840	69646	70472	71289	72118
7	52035	61936	67582	70893	73190	75231	77096	78767	80034	81097	82128	83043	83875	84725	85579	86412
8	48048	55662	59831	62331	64164	65420	66386	67068	67686	68165	68501	68867	69173	69445	69661	69848
9	48682	53949	57399	60023	62286	64049	65345	66346	67058	67630	68199	68794	69408	70059	70738	71397
10	44837	50595	54482	57614	60012	61913	63491	65000	66431	67656	68670	69393	69967	70333	70547	70685
Average	48699	55831	59974	62901	65309	67299	68892	70182	71259	72178	73014	73768	74464	75121	75748	76360
	1A53-2.5 : TS : Variance (kJ ² ·mol ⁻²)															
	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	1A53-2.5 : TS : Variance (kJ ² ·mol ⁻²)															
1	47671	54854	59147	62004	63978	65377	66510	67460	68436	69344	70169	71005	71878	72735	73479	74182
2	45799	50643	54030	56473	58183	59542	60711	61803	62795	63756	64698	65410	65980	66512	66971	67411
3	49222	56868	60907	62935	64275	65530	66740	67967	69261	70554	71782	72972	74079	75170	76238	77232
4	45963	51755	54561	56497	58020	59216	60156	60874	61489	62053	62583	63081	63542	64043	64600	65190
5	51568	58750	62496	64978	67023	68630	69914	70983	71988	72829	73580	74207	74758	75270	75736	76065
6	45092	50995	53978	55661	56777	57789	58685	59307	59801	60313	60786	61248	61624	61960	62266	62533
7	50525	58326	63271	66639	68959	70754	72503	74333	76104	77725	79087	80221	81280	82241	83174	84049
8	48975	55820	59755	62395	64320	65853	66986	68047	69105	70182	71217	72236	73262	74113	74705	75290
9	50011	56785	60605	63058	64977	66576	67988	69291	70371	71283	72049	72724	73038	73174	73306	73486
10	50150	56308	60235	62661	64283	65700	66813	67581	68252	68903	69551	70278	71067	71860	72609	73284
Average	48498	55110	58898	61330	63080	64497	65701	66765	67760	68694	69550	70338	71051	71708	72309	72872
	1A53-2.5 : TS2 : Variance (kJ ² ·mol ⁻²)															
	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	1A53-2.5 : TS2 : Variance (kJ ² ·mol ⁻²)															
1	51150	59880	65304	69075	71945	74284	76070	77533	78675	79691	80671	81607	82468	83212	83746	84115
2	46482	52393	55830	58598	60909	62824	64516	65951	67211	68303	69225	69962	70473	70941	71404	71850
3	51810	58509	61906	64301	66224	67811	69272	70713	71928	72959	73630	74112	74587	75025	75465	75930
4	50457	56486	59927	62276	64074	65374	66190	66963	67714	68443	69175	69835	70483	71163	71837	72451
5	48669	55851	59778	61458	62056	62635	63360	64037	64598	65148	65636	66104	66536	66958	67354	67717
6	48182	54295	57205	58517	59346	60021	60628	61154	61594	61999	62377	62680	63007	63382	63774	64131
7	50854	57652	61154	63882	65759	66959	67693	68126	68494	68914	69340	69628	69948	70326	70691	71051
8	47770	54890	59516	63103	65714	67850	69641	71379	73198	75105	76984	78730	80393	81962	83363	84581
9	47872	54204	57690	60089	62019	63733	65140	66262	67071	67872	68708	69513	70277	71032	71828	72662
10	49319	56343	60446	62831	64404	65691	66945	68143	69132	70037	70768	71295	71739	72074	72324	72546
Average	49256	56050	59876	62413	64245	65718	66945	68026	68962	69847	70651	71347	71991	72608	73179	73703
	1A53-2.5 : TS – GS : $\Delta C_p^\ddagger$ (kJ·mol ⁻¹ ·K ⁻¹)															
	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	1A53-2.5 : TS – GS : $\Delta C_p^\ddagger$ (kJ·mol ⁻¹ ·K ⁻¹)															
Av. TS1 – GS	-201	-721	-1076	-1571	-2230	-2803	-3191	-3417	-3498	-3484	-3464	-3430	-3413	-3414	-3439	-3488
$\Delta C_p^\ddagger$	-1.1	-4.1	-6.1	-8.9	-12.6	-15.9	-18.1	-19.4	-19.8	-19.7	-19.6	-19.4	-19.3	-19.3	-19.5	-19.8
$\Delta\Delta C_p^\ddagger$	1.4	2.0	2.3	2.5	2.6	2.7	2.8	2.9	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7
	1A53-2.5 : TS2 – GS : $\Delta C_p^\ddagger$ (kJ·mol ⁻¹ ·K ⁻¹)															
	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	1A53-2.5 : TS2 – GS : $\Delta C_p^\ddagger$ (kJ·mol ⁻¹ ·K ⁻¹)															
Av. TS2 – GS	558	219	-99	-488	-1064	-1581	-1946	-2156	-2297	-2331	-2363	-2421	-2473	-2514	-2569	-2657
$\Delta C_p^\ddagger$	3.2	1.2	-0.6	-2.8	-6.0	-9.0	-11.0	-12.2	-13.0	-13.2	-13.4	-13.7	-14.0	-14.3	-14.6	-15.1
$\Delta\Delta C_p^\ddagger$	1.3	1.8	2.2	2.3	2.5	2.6	2.7	2.8	2.9	3.0	3.1	3.3	3.4	3.5	3.6	3.7

^a Variances and heat capacities calculate for different moving average windows, based on complexes comprising the protein, ligand, and ten water molecules closest to Glu178.

Extended Data Fig. 1 | The emergent $\Delta C_p^\ddagger$ in the Kemp eliminase variants predominantly reflects effects on the chemical step. **a**, In the activity-temperature dependence of 1A53-2.9, deviations from the linear Eyring behaviour are slightly more pronounced for k_{cat} than k_{cat}/K_M . This indicates that the $\Delta C_p^\ddagger$ for k_{cat} is of similar magnitude to that of k_{cat}/K_M , though no precise $\Delta C_p^\ddagger$ has been determined for k_{cat} . Thus, the experimentally determined $\Delta C_p^\ddagger$ of $-1.17 \text{ kJ} \cdot \text{mol}^{-1} \text{K}^{-1}$ for k_{cat}/K_M predominantly reflects thermodynamic differences between the TS and GS ensemble. **b**, Comparison of the activity-temperature profiles and CD melt curves of 1A53-2 (red) 1A53-2.5 (blue) and 1A53-2.9 (black) shows that the apparent $\Delta C_p^\ddagger$ is not due to thermal denaturation of the protein. The melting temperatures of all variants are high, and those of the evolved variants are significantly above their catalytic temperature optima. 1A53-2.5, which evolved an activation heat capacity, has an even higher melting temperature than 1A53-2 (82 °C vs. 75 °C). Also, long incubation of all variants at the highest experimental assay temperatures does not affect activity, which shows that there is no significant unfolding at this temperature on the timescale of the experiments. Off note: The Kemp elimination only involves a single deprotonation step, making a change in mechanisms unlikely, and there is no evidence of a change in mechanisms. We furthermore did not observe any kinetic signatures signalling rate-limiting substrate binding or product releases (e.g. burst phases or hysteresis under single turnover condition). All data presented in

Extended Data Fig. **1** were extracted from Bunzel *et al.* JACS 2019.

Extended Data Fig. 2 | MD parameters and restraints. **a**, Ground state and **b**, transition state distances (red) and charges (black) based on gas-phase calculations with 6-nitrobenzisoazole and acetate. To prevent artificial rigidification of the transition state by introduction of a covalent bond between the ligand and catalytic residue,²⁸ the transition state was simulated with the transferring protein residing either on the ligand (TS) or base (TS2). Both transition state models provided similar results, supporting the significance of our findings. *, Because Glu178 has only two γ -hydrogens, their charge is adjusted to the given value according to the corresponding three hydrogens in acetate. **c**, Similar to previous work,²⁸ restraints were used to avoid drifting away from reactive conformations during the extended simulations. Weak harmonic restraints for the ligand-base distance δ (black, $>5.0 \text{ \AA}$, $5.0 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$) and the ligand-base dihedral angle Ψ (red, $>\pm 30^{\circ}$, $5.0 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^2$) were added to keep the ligand in the active site. Furthermore, the χ_2 dihedral angle of Trp110 Φ (blue, $>\pm 60^{\circ}$, $5.0 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^2$) was restrained, preventing its indole sidechain from blocking the active site. To avoid biasing the $\Delta C_p^{\ddagger}$ calculations, the same restraints were added to all variants and ligands.

Extended Data Fig. 3 | MD trajectories. Individual (light colours) and average (dark colours) trajectories of 1A53-2 and 1A53-2.5 with GS (red), TS (blue) and TS2 (grey) bound. **a**, Based on the RMSD plot, the first 50 ns of each simulations were discarded to avoid biasing of the analysis to the starting structures (dotted lines). The ligand-base distance **b**, and dihedral angle **c**, remained within the weak harmonic constraints (dotted lines) for most of the simulations. For a definition of the restraints see Extended Data Fig. 2.

with 10 active-site waters

without active-site waters

Extended Data Fig. 4 | Changes in energetic fluctuations give rise to an activation heat capacity $\Delta C_p^\ddagger$ in the evolved catalyst. a, The energy distributions for 1A53-2 (TS: dark red; GS: light red) are almost identical, indicating a $\Delta C_p^\ddagger$ close to zero. Energetic fluctuations were calculated after removal of all but 10 water molecules, though

removing these only slightly affected $\Delta C_p^\ddagger$. **b**, The distribution for 1A53-2.5 is narrower for the TS (dark blue) than for the GS (light blue) complex, resulting in a negative $\Delta C_p^\ddagger$. Distributions in **(a)** and **(b)** are normalized to a 50 ns moving average and plotted against a log scale to emphasize changes in variance. **c**, $\Delta C_p^\ddagger$ was calculated based on the difference in variance between the TS and GS ensemble for moving windows of various sizes, and converges after 50 ns to negative values for 1A53-2.5 (blue). For 1A53-2, $\Delta C_p^\ddagger$ does not significantly deviate from zero (red). Error bars were obtained by leave-one-out cross-validation. As expected, the statistical errors increase with larger window size, because slow conformational changes that are not exhaustively sampled begin to mix into the calculations, complicating the analysis²⁸. Clearly, the energetic fluctuations in 1A53-2 and 1A53-2.5 are different, which gives rise to the dynamical network that tightens the TS ensemble in the evolved variant²⁸. **d+e**, Heat capacities calculated under various conditions qualitatively agree well with each other. Heat capacities were calculated either dry **(e)**, or with the 10 water molecules closest to Glu178 bound **(d)** based on the difference in energy variance of the transition and ground state for running windows with varying size (5 ns to 80 ns) for 1A53-2 (red) 1A53-2.5 (blue) for TS (filled circles) and TS2 (open circles).

Extended Data Fig. 5 | Structural fluctuations in 1A53-2 (left) and 1A53-2.5 (right). Residues that rigidify (blue) or become more flexible (red) between **a+b**, GS and TS, and **c+d**, GS and TS2. In 1A53-2.5, rigidification is particularly expressed in three solvent exposed loops (residues 53-65, 84-92, and 181-192), which is reproduced by clustering based on the whole protein backbone. These loops were thus employed to cluster the structures in an open and closed state (Extended Data Fig. 6).

Extended Data Fig. 6 | Cluster analysis. Average structures of the **a**, 1A53-2 and **b**, 1A53-2.5 ensembles in the open (light blue) and closed (dark blue) state in complex with the TS. Distance-based clustering was performed using the k-means algorithm based on the three solvent exposed loops (bold tubes, residues 53-65, 84-92, and 181-192). Loop closure is indicated by the C α distance of residues 58 and 188 (spheres). **c+d**, Histograms and **e+f**, distributions of the open (light colours) and closed (dark colours) states bound to GS (red), TS (blue) and TS2 (grey). **g+h**, Loop closure decreases the radius of gyration which became smallest in the closed TS of 1A53-2.5 (blue line). **i+j**, The open and closed state interconvert on the timescale of the simulation. Error bars represent the standard deviation of 10 independent simulations.

Extended Data Fig. 7 | Principal component analysis reproduces two state model. Principal component analysis of the mobile loops (residues 53-65, 84-92, and 181-192) was performed for the **a**, GS (red), **b**, TS (blue), and **c**, TS2 (grey) ensembles of 1A53-2 (left) and 1A53-2.5 (right). The histogram of the first principal component after partition of the trajectory according to the cluster analysis (Extended Data Fig. 6) into open (light) and closed (dark) states reproduces the two-state model.

Extended Data Fig. 8 | Local effects at the active sites of 1A53-2 (left) and 1A53-2.5 (right). **a+b**, The distance-distribution of the ten water molecules closest to the O_ε of Glu178 is shown for the GS (red), TS (blue), and TS2 (grey) in the closed (dark colours) and open (light colours) state. The active site becomes less hydrated in the transition states compared to the ground state for both enzymes. Furthermore, the A157Y mutation in 1A53-2.5 displaces a water molecule in the evolved variant at approximately 3 Å. Remarkably, loop closure expels water from the active site in 1A53-2.5 (arrows) which was not observed for 1A53-2. The **c+d**, SASA of the ligand and **e+f**, RMSF of the ligand and base are furthermore lower in 1A53-2.5 than 1A53-2, indicating tighter packing of the ligand. Error bars represent the standard deviation of 10 independent simulations.

Extended Data Fig. 9 | Dynamical networks arising in the closed ensembles of 1A53-2 and 1A53-2.5 between GS and either TS (left) or TS2 (right). **a**, Backbone movements become highly correlated in the closed TS ensemble of 1A53-2.5, indicating the presence of global vibrations comprising most of the protein. **b**, Cross correlations that increase between GS and TS by $\geq 20\%$ are indicated as lines on the structures. **c**, Based on the cross-correlations, shortest path maps were calculated²². These show that evolution introduced a communication network in 1A53-2.5 that originates from the TS and comprises vast parts of the closed state. The size of the edge (black sticks) and vertices (blue spheres: protein, orange sphere: ligand) indicate the increasing weight of the network between GS and TS.

Extended Data Fig. 10 | QM/MM modelling of the Kemp eliminases. Free energy surfaces of 1A53-2 (top) and 1A53-2.5 (bottom) at the **a**, PM6/CHARMM36, **b**, AM1/CHARMM36 and the **c**, PM6/AMBERff99SB level of theory. The ligand as well as the carboxylate of Glu178 were part of the QM region during umbrella sampling of the C-H and N-O bonds. The transition state (black dot) reproduces the asynchronous Kemp elimination with C-H cleavage preceding N-O cleavage observed for similar systems.⁵⁵⁻⁵⁷ PM6/CHARMM36 gave activation energies of 101 kJ·mol⁻¹ and 90 kJ·mol⁻¹, which reproduce the improvements achieved during evolution (86 and 68 kJ·mol⁻¹ for k_{cat}). Similarly, simulations with AM1/CHARMM36 preserved that trend, but gave higher energies (117 and 105 kJ·mol⁻¹). In contrast, simulations with PM6/AMBERff99SB resulted in an inverse trend for the activation energies (111 and 116 kJ·mol⁻¹). This observation for AMBERff99SB mimics the classical MD, in which the ligand was expelled from the active site within 100 ns.

- 10 Bunzel, H. A. *et al.* Emergence of a negative activation heat capacity during evolution of a computationally designed enzyme. *J. Am. Chem. Soc.* **141**, 11745-11748 (2019).
- 22 Romero-Rivera, A., Garcia-Borras, M. & Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **7**, 8524-8532 (2017).
- 28 van der Kamp, M. W. *et al.* Dynamical origins of heat capacity changes in enzyme-catalysed reactions. *Nat. Commun.* **9**, 1177 (2018).
- 29 Arcus, V. L. *et al.* On the temperature dependence of enzyme-catalyzed rates. *Biochemistry* **55**, 1681-1688 (2016).
- 55 Na, J., Houk, K. N. & Hilvert, D. Transition state of the base-promoted ring-opening of isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody production. *J. Am. Chem. Soc.* **118**, 6462-6471 (1996).
- 56 Alexandrova, A. N., Röthlisberger, D., Baker, D. & Jorgensen, W. L. Catalytic Mechanism and Performance of Computationally Designed Enzymes for Kemp Elimination. *J. Am. Chem. Soc.* **130**, 15907-15915 (2008).
- 57 Świderek, K., Tunon, I., Moliner, V. & Bertran, J. Revealing the origin of the efficiency of the de novo designed Kemp eliminase HG-3.17 by comparison with the former developed HG-3. *Chemistry (Weinheim an der Bergstrasse, Germany)* **23**, 7582-7589 (2017).

Supplementary Information

Evolution of dynamical networks enhances catalysis in a designer enzyme

H. Adrian Bunzel^{1,2}, J. L. Ross Anderson¹, Donald Hilvert³, Vickery L. Arcus⁴, Marc W. van der Kamp^{1,2}, Adrian J. Mulholland^{2,*}

¹ *School of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD, UK*

² *Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK*

³ *Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland*

⁴ *School of Science, University of Waikato, Hamilton, 3216, New Zealand*

**Corresponding author: Adrian.Bunzel@bristol.ac.uk, Adrian.Mulholland@bristol.ac.uk*

Supplementary Methods

Model construction. Ground state (Michaelis complex) and transition state geometries were determined with the GAUSSIAN 16 quantum chemistry software package⁵⁸. The models comprised 6-nitrobenzoxazole and acetate and were optimized at the B3LYP/6-311++g(d,p) level; other methods gave similar results. The transition state was located as a maximum on the potential energy surface for the cleavage of the scissile C-H bond by performing a relaxed distance scan of this bond, similar to previous work⁵⁵. Transition state geometries were subsequently optimized using the Berny algorithm. CHELPG charges (MP2/6-311g++(d,p)) were determined for each state (Extended Data Fig. 2). Ground state (GS) and transition state (TS) models were constructed based on these charges and the General Amber Force-Field (GAFF). An additional transition state model (TS2) was constructed with the same geometry and charge distribution as TS, but with the proton bound to Glu178 instead of to the ligand. Input structures for MD were manually constructed by replacing the 6-nitrobenzotriazole inhibitor in the crystal

structures of 1A53-2 (PDB code: 3NZ1)²⁶ and 1A53-2.5 (PDB: 6NW4)¹⁰ with either the substrate or transition state model. Different forcefields were tested for representing the proteins. For MD simulations with the CHARMM36 protein force-field, the CHARM-GUI was used to parametrize the protein,⁵⁹ and to solvate and neutralize the complex in an octahedral box with a 10 Å edge distance in TIP3P water and 0.15 M sodium chloride. For simulations with the AMBERff99SB protein force-field, tLEAP was used to solvate the complex in an octahedral box with a 10 Å edge distance in TIP3P water and to neutralize with chloride ions⁶⁰. All MD simulations were run with the Amber16 package⁶⁰.

Testing force-fields with MM and QM/MM. The CHARMM36 and AMBERff99SB protein force-fields were tested in MM and QM/MM MD simulations. Simulations with the CHARMM36 force-field gave structures and relative activation energies consistent with experiment. While the GS complex of 1A53-2 was not stable with AMBERff99SB, which apparently underestimated the π -stacking interactions of the GS with Trp110 and Trp210 leading to displacing of the ligand with water within <100 ns, CHARMM36 gave structures consistent with the experimentally observed structures, and allowed sampling on the timescales required for the analysis. The finding that CHARMM36 more accurately describes the bound ligands is mimicked by hybrid QM/MM simulations, in which CHARMM36 reproduced the activity improvements during evolution. To that end, the distances of the scissile C-H and N-O bonds were sampled similar to previously published methods^{56,57}. The QM subsystem was composed of the substrate and the carboxylate and γ -methylene of Glu178 and was described using either the semiempirical AM1 or PM6 Hamiltonian. To saturate the valence of the QM/MM frontier atoms, a link atom was placed between the C $_{\beta}$ and C $_{\gamma}$ atoms of Glu178. A 15 Å cut-off for nonbonding interactions with the QM subsystem was applied. QM/MM MD was started from structures that were

equilibrated with the TS model (see MD simulations). After 1 ns equilibration with 1 fs time steps, umbrella sampling of the scissile C-H and N-O bonds was started. C-H and N-O bonds were sampled in 0.1 Å and 0.05 Å steps with force constants of 200 kcal·mol⁻¹Å⁻¹ and 400 kcal·mol⁻¹Å⁻¹, respectively. Each window was equilibrated by ramping up the force constants from 0 to their final value over 10 ps and simulation over additional 10 ps. Subsequently, 40 ps of sampling were performed. Potential of mean force free-energy landscapes were calculated from these data using the weighted histogram analysis method⁶¹. The PM6/CHARMM36 and AM1/CHARMM36 QM/MM simulations correctly reproduced the activation energy difference between 1A53-2 than 1A53-2.5, while PM6/AMBERff99SB gave higher activation energies for 1A53-2 than 1A53-2.5 (Extended Data Fig. 10), probably because interactions of the ligand with the active site (e.g. π -stacking to Trp110 and Trp210) were not accurately reflected in AMBERff99SB. Thus, CHARMM36 was employed for all subsequent simulations.

MD simulations. All simulations were performed using established and validated protocols²⁸. Briefly, the structures were minimized by 300 step minimization of water, ions and hydrogens only, followed by 50 ps MD simulation in the NVT ensemble at 300 K of the water and ions only (positional restraint on protein atoms and ligand: 25 kcal·mol⁻¹·Å⁻²), and minimization of the whole system with positional restraints on C α atoms (100 kcal·mol⁻¹·Å⁻²) for 300 steps. Ten independent simulations were performed for each protein and ligand combination by assigning random velocities at 25 K followed by heating to 300 K in 20 ps using Langevin dynamics for temperature control with a 1 ps⁻¹ collision frequency (maintaining 100 kcal·mol⁻¹·Å⁻² positional restraints on C α atoms). For each independent simulation, four consecutive 10 ps simulations were subsequently performed under the same conditions, reducing the positional restraints on

the C $_{\alpha}$ atoms to 20, 8, 4, 3, 2, and 1 kcal·mol⁻¹·Å⁻². Subsequently, 1 ns of equilibration was performed in the NPT ensemble at 1 atm, using the Berendsen barostat (10 ps pressure relaxation time) and Langevin dynamics for temperature control (1 ps⁻¹ collision frequency). Production simulations (500 ns) were performed in the NPT ensemble with the Berendsen thermostat and loose temperature coupling and pressure scaling (10 ps time constant each), to limit the influence of the thermostat while avoiding temperature drift. In all MD simulations, the default direct-space cutoff was used for non-bonded interactions, with particle-mesh Ewald summation for long-range electrostatic interactions. MD simulations were run with pmemd.cuda on GPUs, using the SPFP precision model¹². Throughout the simulations, restraints were used to maintain the Michaelis complex, with equivalent restraints in each state (see Extended Data Fig. 2 for a detailed description), similar to previous simulations²⁸. Simulations with the 1A53-2.9 variant were attempted in a similar manner after modelling the additional mutations onto the 1A53-2.5 scaffold with PyMol. Owing to the lack of a structurally accurate experimental model, simulations frequently crashed after >100 ns and could not be analysed in this context.

Heat capacity calculations. Analysis of $\Delta C_p^{\ddagger}$ was performed using procedures developed and tested in previous work on natural enzymes²⁸, based on structures sampled at 10 ps intervals from 50–500 ns windows of the MD simulations, with force-field energies recalculated after stripping of ions and all but 10 water molecules closest to the catalytic base. Energy distributions were calculated by normalization to a moving average (50 ns window). Heat capacities were similarly calculated based on the moving average of the energy variance according to equation 1, where ΔC_{A-B} is the change in heat capacity

between two states, A and B with the variances $\langle \delta H_A^2 \rangle$ and $\langle \delta H_B^2 \rangle$. Error-bars were obtained by leave-one-out cross-validation.

$$\Delta C_{A-B} = \frac{\langle \delta H_A^2 \rangle - \langle \delta H_B^2 \rangle}{k_B T^2} \quad (1)$$

Clustering and principal component analysis. All other analyses were performed with CPPTRAJ⁶², unless otherwise stated. RMSF calculations were performed for the backbone C α atoms of each trajectory, and RMSF values represent the average of the ten independent trajectories. Clustering was performed using the k-means algorithm. Initial clustering based on the whole protein backbone revealed that structural changes are predominantly located in three solvent exposed loops (residues 53-65, 84-92, and 181-192), in agreement with the RMSF calculations. The final cluster analysis presented here was performed for these loops in distance-space with the k-means algorithm. Similarly, the distance-matrix of these loops was analysed by principal component analysis using MDtraj⁶³, and the first principal component aligned well with the two states from the cluster model. For the following analysis, each MD trajectory was partitioned into open and closed state based on the cluster model.

Calculation of local effects. Interaction energies were calculated for each protein residue with the ligand using Sander⁶⁰. SASA and RMSF were calculated for the ligand and base only, after alignment of the C α in the whole trajectory. All values were individually calculated for the ten independent trajectories and subsequently averaged. The width of the substrate entry channel was determined using Caver Web³¹ based on crystal structures of 1A53-2 (PDB code: 3NZ1)²⁶ and 1A53-2.5 (PDB: 6NW4)¹⁰. Bottleneck radii

of 1.8 Å were determined for both structures for tunnels reaching to the general base Glu178.

Dynamic cross correlations and shortest path maps. Shortest path maps were calculated using established protocols⁶⁴, but for the difference in correlations between the TS and GS ensemble (Fig. 3 and Extended Data Fig. 9). The analysis was performed only for frames corresponding to the closed state as determined by the cluster analysis. The analysis was based on the backbone C_α and the ligand centre of mass. Distance matrices were determined from the average structure of the combined ground and transition state frames with CPPTRAJ⁶². Dynamic cross correlations were determined with CPPTRAJ⁶² for both states, and the shortest path maps presented here are based on the difference in correlations ($\Delta corr$) between transition state ($corr_{TS}$) and ground state ($corr_{GS}$). The shortest path maps presented here reflect only correlations that increase in their absolute magnitude (Equation 4). Since the shortest path algorithm requires all input correlation to be above zero, correlations that decreased in magnitude ($|corr_{TS}| \leq |corr_{GS}|$) were replaced with 0.0000000001. Replacement with smaller values closer to zero did not affect the resulting networks.

$$\Delta corr = \begin{cases} |corr_{TS} - corr_{GS}| & \text{if } |corr_{TS}| > |corr_{GS}| \\ 0.0000000001 & \text{if } |corr_{TS}| \leq |corr_{GS}| \end{cases} \quad (4)$$

In contrast to previous work⁶⁴, the distance cut-off was increased from 5.0 Å to 10.0 Å to ensure all active site residues interact with the ligand are included while providing meaningful shortest path maps. For 1A53-2.5, cut-offs between 9.0 Å to 12.0 Å resulted in similar networks. Smaller cut-offs resulted in networks that did not include the ligand, and the connectivity between neighbouring residues were lost with larger cut-offs.

References

- 10 Bunzel, H. A. *et al.* Emergence of a negative activation heat capacity during evolution of a computationally designed enzyme. *J. Am. Chem. Soc.* **141**, 11745-11748 (2019).
- 26 Privett, H. K. *et al.* Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3790-3795 (2012).
- 28 van der Kamp, M. W. *et al.* Dynamical origins of heat capacity changes in enzyme-catalysed reactions. *Nat. Commun.* **9**, 1177 (2018).
- 31 Stourac, J. *et al.* Caver Web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Res.* **47**, W414-W422 (2019).
- 55 Na, J., Houk, K. N. & Hilvert, D. Transition state of the base-promoted ring-opening of isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody production. *J. Am. Chem. Soc.* **118**, 6462-6471 (1996).
- 56 Alexandrova, A. N., Röthlisberger, D., Baker, D. & Jorgensen, W. L. Catalytic Mechanism and Performance of Computationally Designed Enzymes for Kemp Elimination. *J. Am. Chem. Soc.* **130**, 15907-15915 (2008).
- 57 Świderek, K., Tunon, I., Moliner, V. & Bertran, J. Revealing the origin of the efficiency of the de novo designed Kemp eliminase HG-3.17 by comparison with the former developed HG-3. *Chemistry (Weinheim an der Bergstrasse, Germany)* **23**, 7582-7589 (2017).
- 58 Frisch, M. J. *et al.* Gaussian 16. Wallingford, CT (2016).
- 59 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* **29**, 1859-1865 (2008).
- 60 Case, D. A. *et al.* AMBER 2016. University of California, San Francisco (2016).
- 61 Grossfield, A. WHAM: the weighted histogram analysis method, version 2.0.9. http://membrane.urmc.rochester.edu/wordpress/?page_id=126 (2019).
- 62 Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084-3095 (2013).
- 63 McGibbon, Robert T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528-1532 (2015).
- 64 Osuna, S., Jimenez-Oses, G., Noey, E. L. & Houk, K. N. Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc. Chem. Res.* **48**, 1080-1089 (2015).