Supporting information for

Computational design of catalytic dyads and oxyanion holes for ester hydrolysis

Authors

Florian Richter,[1,4,*] Rebecca Blomberg,[2,7*] Sagar D. Khare,[1] Gert Kiss,[3,10] Alexandre P. Kuzin,[5] Adam J. T. Smith,[3] Jasmine Gallaher,[1] Zbigniew Pianowski,[2] Roger C. Helgeson,[3] Alexej Grjasnow,[11] Rong Xiao,[8] Jayaraman Seetharaman,[5] Min Su,[5] Sergey Vorobiev,[5] Scott Lew,[5] Farhad Forouhar,[5] Gregory J. Kornhaber,[8] John F. Hunt,[5] Gaetano T. Montelione,[8,9] Liang Tong,[5] K.N. Houk,[3] Donald Hilvert[2,†] and David Baker[1,4,6,†]

[1] *Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA*

[2] *Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland*

[3] *Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA*

[4] *Interdisciplinary Program in Biomolecular Structure and Design, University of Washington, Seattle, Washington 98195, USA*

[5] *Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027*

[6] *Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA*

[7] *present address: Firmenich SA, Corporate R&D Division, CH-1211 Geneva, Switzerland*

[8] *Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854 and Northeast Structural Genomics Consortium*

[9] *Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854*

[10] *present address: Department of Chemistry, Stanford University, Stanford, CA 94305*

[11] *Department of Biochemistry, Ludwig-Maximilians-Universität München, 81377 München, Germany*

*These authors contributed equally to this work.

[†]To whom correspondence should be addressed. Email:

hilvert@org.chem.ethz.ch, dabaker@u.washington.edu

Methods

Protein Production. The cell pellets were resuspended in sonication buffer (25 mM Hepes (pH 7.5), 300 mM NaCl, 1 mM TCEP) containing 10 mM imidazole. Cell lysis was achieved by the addition of 1 mg/mL lysozyme and subsequent sonication. The soluble fraction was applied to Ni-NTA slurry (Qiagen), washed with 20 mM imidazole and then with 32.5 mM imidazole before elution with 250 mM imidazole in sonication buffer. The proteins were dialyzed into 20 mM Tris, 20 mM NaCl (pH 8.0) for 16 hours and then purified by anion exchange chromatography (MonoQ column, GE Healthcare) in the same buffer, eluting with a salt gradient (20 mM to 1,000 mM NaCl). The proteins were concentrated using Amicon Ultra-15 units (Millipore). Protein concentrations were determined by measuring the absorbance at 280 nm using the calculated extinction coefficients (ε (ECH13; ECH13 C45A; ECH13 C45A/H100A) = 30940 M^{-1} cm^{-1}, ε (ECH14; ECH14 C132A; ECH14 C132A/H104A) = 42860 M^{-1} cm^{-1}, ε (ECH19; ECH19 C161A; ECH19 C161A/H226A) = 95800 M^{-1} cm^{-1}, ε (ECH19 K354P/P364W) = 101300 M^{-1} cm^{-1}, ε (FR29; FR29 C10A/H126A; FR29 A44S/T112L/V151L) = 41830 M^{-1} cm^{-1}). Protein purity was confirmed by SDS-PAGE.

Initial activity screening. For all 55 designs, a mixture containing 20 μM purified protein and 100 μM of coumarin ester **2** was prepared and product formation was monitored in a fluorimeter at room temperature ((λ_{ex} = 340 nm, λ_{em} = 452 nm). Designs that showed at least a 20-fold increase over the background rate (100 μM substrate **2** in buffer taken from the dialysis bucket) were considered active, and the 4 designs that did were characterized further.

Kinetic measurements. For substrate **2**, product formation was monitored in a fluorimeter (Photon Technology International) at 29°C (λ_{ex} = 340 nm, λ_{em} = 452 nm). The signal was calibrated using a concentration series of 7-hydroxycoumarine (1 μM to 50 μM final concentration) in 25 mM Hepes buffer (pH 7.5), containing 100 mM NaCl and 5% acetonitrile. For substrate **3**, release of para-nitrophenol was monitored at 405 nm in a Lambda 35 UV/Vis spectrometer (PerkinElmer) at 29 °C.

***K*$_i$ determination.** The esterase variants (2 μM final concentrations) were pre-incubated with varying concentrations of the tyrosine ester **1** (0.001 – 300 μM final concentration) in 25 mM HEPES buffer (pH 7.5), 100 mM NaCl, 5% acetonitrile and the reactions were initiated by addition of the coumarin ester **2** (50 μM final concentration). Product formation was monitored as described previously. The IC_{50} value was determined by curve fitting (Hill-Slope model, v_i at infinite inhibitor concentration was set to zero) and subsequently converted into the corresponding K_i value using the Cheng-Prusoff equation $K_i = [IC_{50}]/(1+[S]/K_m)$ (*42*). The K_m value was determined independently under identical reaction conditions using the previously described protocol.

CD spectroscopy. The far-UV spectra of the protein samples (c (ECH13; ECH13 C45A; ECH13 C45A/H100A) = 10 μM, c (ECH19; ECH19 C161A; ECH19 C161A/H226A; ECH19 K354P/P364W) = 5 μM, c (ECH14; ECH14 C132A; ECH14 C132A/H100A) = 5 μM, c (FR29; FR29 C10A H126A; FR29 A44S/T112L/V151L) = 5 μM) were measured at 20 °C using an Aviv 202 spectropolarimeter (Aviv Associates, Lakewood, NJ). Thermal denaturation spectra of the proteins were monitored at 222 nm.

Generation of the optimized variants. The genes for the modified designs were generated by the Kunkel method[1] and the corresponding proteins were produced as described for the original designs. To confirm their identity, all variants were characterized by mass spectrometry (Tables 6S-9S). Their activities were evaluated by incubationg 5 or 10 μM of each enzyme with a tenfold excess of coumarin ester **2** and by subsequently monitoring fluorophore release as described above (Figures 1S and 2S).

Aminolysis experiments. The best third-generation ester hydrolase design, FR29 A44S/T112L/V151, was pre-incubated with a tenfold excess of the amines shown in Figure 6S (and one thiol) at pH 7.5 prior to addition of coumarin substrate **2**.

[1] Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods Enzymol 154*, 367-382

Unfortunately, the deacylation rates remained identical to those observed in the absence of amines. Hence, we conclude that none of the added nucleophiles facilitates cleavage of acyl-enzyme intermediate**.**

Mass spectrometry. Protein samples for the characterization of the acyl-enzyme intermediates were prepared by incubating the designed hydrolases (10 μM final concentration in 25 mM Hepes (pH 7.5), 100 mM NaCl) with a 10-fold excess of tyrosine ester **1** or coumarin ester **2** (or no ester for the negative control) at room temperature for 6 hours or 60 min, respectively. The samples were then subjected to a shock-freeze in liquid nitrogen. Immediately before injection on the mass spectrometer the samples were thawed. For ESI-MS studies, the samples were desalted using a C_4 ZipTip and measured in 50% acetonitrile/0.2% formic acid (pH 2.0) on a Q-TOF Ultima mass spectrometer (Waters). For MALDI-MS/MS studies, 30 μl of the sample were digested with 10-20 μl trypsin (10 ng/ul in 10 mM Tris, 2 mM $CaCl_2$, pH 8.2) at 50° C for 2 hours and the resulting peptides were analyzed by MALDI. Comparison of the treated und untreated samples was used to identify the modified peptides, which were then further fragmented by collision-induced dissociation (CID). The resulting fragments were analyzed to identify the modified amino acid.

Molecular Dynamics Simulations. Molecular dynamics simulations were carried out on designs FR25 through FR32. Each was prepared for three independent MD runs: one apo-MD, one with substrate **1** bound to the active site, and one with substrate **2**. MD simulations were also carried out on ECH13, ECH14, and ECH19, which were prepared for apo-MD and with substrate **2**, only. FR26, FR29 and ECH14 were set up as dimeric systems Substrate parameters were generated within the antechamber module of AMBER 10[2] using the general AMBER force field, with partial charges set to

[2] Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008), AMBER 10, University of California, San Francisco.

fit the electrostatic potential generated at HF/6-31G* by RESP.[3] The charges were calculated according to the Merz-Singh-Kollman scheme[4,5] using Gaussian 03.[6] Structures were immersed in a truncated octahedral box with a 10 Å buffer of TIP3P[7] water molecules. The systems were neutralized by addition of explicit counter ions. Water molecules were triangulated with the SHAKE algorithm such that the angle between the hydrogen atoms is kept fixed. After equilibration (SI), a 20 ns production MD simulation was performed for each of the systems using pmemd.[8] Geometries and velocities were saved every 100 steps (0.2 ps) which resulted in a total of 100,000 frames from each production run. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method.[9] Post-MD data-extraction and analysis was performed using the ptraj module of AMBER 10 and the statistical analysis software OriginPro8.[10]

The active sites of FR25, FR26, and FR31 showed substantial instabilities that were deemed irreparable. FR27 and FR30 were designated as borderline cases that were expected to have weak activity at most due to high solvent accessibility (active site of

---

[3] Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J. Phys. Chem. 97:10269-10280.

[4] Besler BH, Merz KM, Kollman PA (1990) Atomic charges derived from semiempirical methods. J. Comput. Chem. 11:431-439.

[5] Singh UC, Kollman PA (1984) An approach to computing electrostatic charges for molecules. J. Comput. Chem. 5:129-145.

[6] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery Jr JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford CT.

[7] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79:926-935.

[8] Duke RE, Pedersen LG (2003) PMEMD. University of North Carolina, Chapel Hill.

[9] Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98:10089-10092.

[10] Origin, OriginLab, Northampton, MA.

FR30) or missing catalytic contacts (oxyanion hole of FR27). FR28 and FR32, the two most promising designs from MD analysis, turned out to express poorly. Efforts to improve their solubility were met with no success.

Structure determination

Sample preparation for crystallization experiments. The production of the four active designs was carried out as part of the high-throughput protein-production process of the Northeast Structural Genomics Consortium (NESG)[11]. Each of the designs was assigned and NESG identifier: OR49 for ECH19, OR51 for ECH13, OR52 for FR29, and OR54 for ECH14. *E. coli* BL21-GOLD (DE3) were transformed with those expression vectors. A single isolate was cultured in MJ9 minimal medium[12] supplemented with selenomethionine, lysine, phenylalanine, threonine, isoleucine, leucine and valine for the production of selenomethionine-labeled versions of the designs[13]. Initial growth was carried out at 37^{O}C until the OD_{600} of the culture reached 0.6–0.8. The incubation temperature was then decreased to 17^{O}C and protein expression was induced by the addition of IPTG (isopropyl β-D-1-thiogalactopyranoside) to a final concentration of 1 mM. Following overnight incubation, the cells were harvested by centrifugation. Cell pellets were resuspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 40 mM imidazole, 1 mM TCEP and 0.02% NaN_3) containing protease inhibitors (Complete, Mini, EDTA-free, Roche) and disrupted by sonication. The resulting lysate was clarified by centrifugation at 27,000 x g for 30 min at 4^{O}C, followed by filtering through a 0.2 μm filter. The supernatant was loaded onto an AKTAxpress system (GE Healthcare) with a two-step protocol consisting of IMAC (HisTrap HP) and gel-filtration (HiLoad 26/60

[11] Xiao, R.; Anderson, S.; Aramini, J.M.; Belote, R.; Buchwald, W.; Ciccosanti, C.; Conover, K.; Everett, J.K.; Hamilton, K.; Huang, Y.J.; Janjua, H.; Jiang, M.; Kornhaber, G.J.; Lee, D.Y.; Locke, J.Y.; Ma, L.-C.; Maglaqui, M.; Mao, L.; Mitra, S.; Patel, D.; Rossi, P.; Sahdev, S.;Sharma, S.; Shastry, R.; Swapna, G.V.T.; Tong, S.N.; Wang, D.; Wang, H.; Zhao, L.; Montelione, G.T.; Acton, T.B. J. Struct. Biol. 2010, 172: 21 - 33

[12] Jansson, M.; Li, Y.-C.; Jendeberg, L.; Anderson, S.; Montelione, G. T.; Nilsson, B. J. Biomol. NMR. 1996, 7: 131-141

[13] Doublie´, S.; Kapp, U.; Aberg, A.; Brown, K.; Strub, K.; Cusack, S. FEBS Lett. 1996, 384: 219-221

Superdex 75) chromatography. The purified designs were in a buffer containing 10 mM Tris-HCl , 100 mM NaCl, 5 mM DTT, pH 7.5, and concentrated to 6–10.5 mg/m1. Samples were flash-frozen in 50 μL aliquots using liquid nitrogen and stored at -80^{O}C until crystallization. The sample purity (>98%), molecular weight, oligomerization state were verified by SDS–PAGE, MALDI–TOF mass spectrometry, and analytic gel filtration followed by static light scattering, respectively.

Crystallization and data collection. Initial crystallization conditions for all four proteins were found by high-throughput robotic screening of 1536 different conditions at the Hauptmann-Woodward Institute, Buffalo, NY[14]. Crystallization and refinement statistics for all designs can be found in Table 10S.

ECH19 / OR49: Manual optimization of the initial crystallization conditions for ECH19 was performed by mixing of 10.1 mg/ml protein in buffer containing 100 mM NaCl, 5 mM DTT, 0.02% NaN_3, 10 mM Tris-HCl, pH 7.5 with 1 μl of the precipitant. The final precipitant solution contained 18% (w/v) PEG-3350, 0.15 M ammonium sulfate, 0.1 M Tris-HCl, pH 8.0. The crystals were cryoprotected with 15-20 % (w/v) ethylene glycol in the well solution before flash-freezing in liquid nitrogen. X-ray data were collected at the Brookhaven National Laboratory (BNL), beamline X4C. Crystal diffracted to 2.5Å and belongs to space group $P2_12_12$, with one molecule in the asymmetric unit [Table 10S].

ECH13/ OR51: Crystals of ECH13 protein were grown by microbatch method under mineral oil by mixing protein solution containing 100 mM NaCl, 5 mM DTT, 0.02% NaN_3, 10 mM Tris-HCl, pH 7.5 with reservoir solution consisted of 100 mM $NaHPO_4$, 12% PEG20K in 100 mM MES buffer pH 7.5. The complex with coumarin was prepared by a cocrystallization method. As fluorogenic coumarin ester is not water soluble, it was dissolved in DMSO at 200 mM, then mixed with protein solution at final concentration of 10 mM. The mixture was incubated on the ice overnight, and formed supernatant was used for crystallization by macrobatch method. Obtained crystals were flash-cooled at

[14] Luft, J.R., Collins, R.J., Fehrman, N.A., Lauricella, A.M., Veatch, C.K., and DeTitta, G.T. (2003). "A deliberate approach to screening for initial crystallization conditions of biological macromolecules.", Journal of Structural Biology, 142, 170-179

100K in a nitrogen stream with cryoprotectant, whereas the crystals were briefly soaked in solution containing 20% (v/v) ethylene glycol. X-ray data for native protein were collected on BL9-2 beamline, SSRL and were processed with the program DENZO.[15] Crystals diffracted to 1.6Å. Diffraction data for liganded form were collected on RAXIS-IV image plate detector, Rigaku rotating anode, at 1.5418Å wavelength. The crystals of unliganded form belong to space group $P4_32_12$ and diffracted to 1.6Å, but crystals of complex belong to space group C2, diffracted to 2.0Å. It's interesting that attempts to grow complex crystals by soaking of native crystals in a solution containing coumarin failed.

FR29 / OR52: The same approach as for ECH13 was used to get crystals for liganded/unliganded forms of FR29. The final concentration of crystallization cocktail for unliganded form was: 0.17M of NH_4acetate, 0.085M of Na_3citrate, 25% of PEG4K, 15% glycerol, pH 5.6. The crystal for complex with coumarin were grown from solution containing 0.5M of ammonium sulfate, 0.1M HEPES, 30% MPD, pH 7.5. Data for both crystal forms were collected at Brookhaven National Laboratory (BNL), beamlines X4A and X4C [Table 10S]. Both crystal forms belong to orthorhombic space group $P2_12_12_1$, 4 copies of the molecule in the asymmetric unit and diffracted to 2.8Å.

ECH14 / OR54: The crystals for ECH14 were grown by sitting drop under oil method at the 4^0C. Crystallization solution contained 0.1M of KH_2PO_4, 40% PEG 4000, 0.1M Tris, pH 8.0. X-ray data were collected from crystals maintained at 100 K using a wavelength of 0.979Å on beamline X4C at the National Synchtrotron Light Source at BNL. The diffraction images were integrated and merged using HKL2000 and SCALEPACK[14]. Orthorombic crystals (space group $P2_12_12_1$) diffracted to 3.2Å, with 2 molecules in the asymmetric unit [Table 10S].

Structure determination and refinement. ECH19 / OR49: Structures for unliganded cysteine esterases were determined by Molecular Replacement (MR) method (program BALBES[16]) using appropriate scaffold. The scaffold for ECH19 is periplasmic

[15] Otwinowski,Z. & Minor, W. (1997). Methods Enzymol. 276, 307-326
[16] F.Long, A.Vagin, P.Young & G.N.Murshudov (2008).

oligogalacturonide binding unsaturated hexuronate sugars from *Yersinia enterocolitica* (PDB id 2UVH) in its closed form, 430 amino acids residues length. Program automatically made corresponding mutations and refined preliminary model with REFMAC.[17] Further refinement was done with program package PHENIX[18] and refitting/remodeling was performed with program COOT.[18] Details of data collection and refinement are summarized in the Table 10S. The asymmetric unit contains two copies of the molecule (A and B labeled). Subunit A (residues 2-412) and subunit B (residues 3-409) are packed by 'face-to-face' mode. The RMS in CA atom position of two subunits is 0.20Å and maximum deviation is 1.89Å is observed in vicinity of residues Ser229. Overlay of the scaffold molecule A on the subunit A of target molecule showed bigger differences: 4.009Å for corresponding CA atoms with maximum 9.125Å for Asp47. When comparing to the open, unliganded form of the scaffold, PDB code 2UVG, the RMS deviation for all CA-atoms is 1.351Å, with a maximum of 7.045Å for Ser229 in subunit A, located close to mutated M231H. Subunit B is better matched: 1.33Å for all CA-atoms, and maximum difference 6.66Å for Phe226.

ECH13/ OR51: Three-dimensional structure of the apo enzyme ECH13 was solved by MR, using coordinates of the scaffold, human mitochondrial deoxyribonuletidase (PDB entry 1Q92) as search model for program BALBES. It was found 1 molecule (space group $P4_32_12$), and preliminary refined with REFMAC generated Rfree=0.287, R=0.255. After manual inspection and rebuilding with program COOT[19] and anisotropic refinement with solvent molecules with PHENIX, Rfree factor dropped to 0.195, and standard crystallographic R factor dropped to 0.178. The RMS deviation of the coordinates from ideal is 0.008Å. No amino acid residues are in the disallowed region,

---

"BALBES: a Molecular Replacement Pipeline", Acta. Cryst., D64, 125-132

[17] G.Murshudov, A.Vagin, & E.Dodson. (1997). " Refinement of Macromolecular Structures by the Maximum-Likelyhood Method", Acta Cryst, D53, 240-255

[18] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J.S. Richardson, T.C. Terwilliger & P.H.Zwart. (2010), Acta Cryst., D66, 213-221.

[19] Emsley, P.and Cowtan, K.D. (2004) "Coot: Model-building tools for molecular graphics". Acta Cryst., D 60, 2126-2132

91.8% are located in the most favored region on the Ramachandran map. The RMS deviation for CA atoms for target-scaffold molecules is 0.280Å. Biggest deviation 2.815Å is observed for the first amino acid in C-term, second one is for SeMet105, which is located in the vicinity of mutated amino acid I103H. It is interesting that the scaffold and unliganded target molecule have the same space group and almost the same cell parameters: a=b=73.758Å c=105.981Å (1Q92) vs a=b=73.436Å c=105.083Å (3U13, target molecule). The crystallization conditions are also similar: PEG 8K/20K, potassium/sodium phosphate, pH 5.3/6.0. The crystal of the search molecule diffracted to 1.4Å against the current resolution 1.6Å, but the completeness of the scaffold is low 64.8%, whereas the target completeness is 100%. The complex with ester coumarin was solved by MR using coordinates of the unliganded enzyme and was rebuilt/refined with COOT/PHENIX correspondingly.

FR29 / OR52: The three-dimensional structure of the apo-enzyme was determined by MR with the program BALBES using coordinates of the scaffold, tryptophanyl–tRNA synthase from *Bacillus Stearothermophilus* (PDB code 3FHJ). It was found 4 molecules and refined with program PHENIX [9]. Model was rebuilt/corrected with program COOT. Final model (Rfree=0.290, R=0.214 at the resolution 2.8Å, RMS deviation for bond lengths 0.009Å) was deposited to PDB (id 3U1V). Complex with ester coumarin was solved by MR using structure of the apo-enzyme. The RMS deviations from original molecule A are 1.536Å, 1.636Å, 1.618Å, and 1.737Å for target molecules A, B, C, D respectively.

ECH14 / OR54: The coordinates of the design scaffold, aspartate aminotransferase from *E.coli* (PDB id 1TOI, sequence identity 94.7%) was used as search model for MR to solve the structure of the ECH14. The initial R_{free}/R were 0.364/0.249 at the resolution 3.2 Å. After several cycles of rebuilding with program COOT and refinement with PHENIX, R_{free} reduced to 0.287, and R=0.197. 85% of the amino acid residues lie in the favored region on the Ramachandron plot, and 0.1% is in the disallowed region. RMS deviation of the protein atoms from ideal values is 0.009Å. The high average B-factor for protein atoms 74.2Å2 is correlated to poor diffraction of the crystals (3.2Å). Despite of low resolution data, all chains for both subunits (A and B) was traced,

whereas the high resolution 1.9Å scaffold molecule has gap 125-129. RMS deviation for subunit A (2616 atoms) between final model and scaffold, calculated with PHENIX (command Phenix.Superpose_pdbs) is 2.022Å, and 2.030Å for molecule B.

Substrate synthesis

Tyrosine ester 1 (SI, Figure 4S). *Benzyl 2-(S)-(N-fluorenyl-methoxycarbonyl)-amino-3-(p-hydroxyphenyl)-propanoate (Fmoc-Tyr(OH)-OBn)* (***1c***) – Fmoc-Tyr(O-*t*Bu)-OH (1.0 g, 2.18 mmol, 1.0 eq.) and HBTU (908 mg, 2.39 mmol, 1.1 eq.) were dissolved in anhydrous DMF (5 mL) and stirred under N_2 at 0^0C (ice/water bath). DIPEA (431 µL, 5.66 mmol, 2.6 eq.) was added drop wise. After stirring for 5 min, benzyl alcohol (338 µL, 3.26 mmol, 1.5 eq.) was added. After another 40 min, the reaction mixture was warmed up to room temperature and quenched with saturated aqueous NH_4Cl solution (100 mL) after incubation for another hour. Extraction with ethyl acetate (3x50 mL) and washing the combined organic fractions with saturated aqueous NH_4Cl (1x100 mL), brine (2x100 mL) followed by drying with anhydrous Na_2SO_4 and evaporation of solvents yielded the crude product. Purification was performed using flash chromatography on silicagel (most of benzyl alcohol elutes with hexane:ethyl acetate 20:1, the product was then eluted with hexane:ethyl acetate 3:1) resulting in 1.0 g (83% yield) of the benzyl ester.

TLC: R_f = 0.5 (cyclohexane:ethyl acetate 3:1); **LCMS (ESI):** RT = 13.93 min., *m/z* calculated for $C_{35}H_{36}NO_5$: 550.26 $[M+H]^+$, found: 550.24 $[M+H]^+$, *m/z* calculated for $C_{35}H_{35}NO_5Na$: 572.24 $[M+Na]^+$, found: 572.26 $[M+Na]^+$; **^{1}H NMR ($CDCl_3$) δ(ppm):** 7.78 (2H, d), 7.58 (2H, d), 7.29-7.44 (9H, m), 6.85-6.94 (4H, m), 5.32 (1H, d, *J* = 8.4 Hz), 5.17 (2H, m), 4.71 (1H, m), 4.34-4.47 (2H, m), 4.22 (1H, t, *J* = 7.2 Hz), 3.10 (2H, m), 1.34 (9H, s); **^{13}C NMR ($CDCl_3$) δ(ppm):** 171.0, 155.2, 154.2, 143.6, 143.5, 141.1, 134.9, 130.1, 129.6, 128.4, 128.3, 127.5, 126.9, 124.9, 123.9, 119.8, 78.4, 67.3, 67.0, 55.0, 47.3, 37.7, 29.0.

485 mg of the benzyl ester (Fmoc-Tyr(O-*t*Bu)-OBn, 0.882 mmol) was dissolved in trifluoroacetic acid (20 mL) and stirred at room temperature for 1h. The solvent was then

evaporated under reduced pressure and the resulting crude product was purified using flash chromatography (SiO_2, elution of impurities with hexane:ethyl acetate 5:1, the product's elution begins with hexane:ethyl acetate 3:1 (122 mg, 28% of the expected yield), however it is completely removed from the column with pure ethyl acetate followed by ethyl acetate:methanol 4:1). Fractions containing the product were combined and evaporated resulting in 422 mg (97% of the expected yield) of the pure Fmoc-Tyr(OH)-OBn.

TLC: R_f = 0.25 (cyclohexane:ethyl acetate 3:1); **LCMS (ESI):** RT = 12.35 min., *m/z* calculated for $C_{31}H_{28}NO_5$: 494.20 $[M+H]^+$, found: 494.24 $[M+H]^+$, *m/z* calculated for $C_{31}H_{27}NO_5Na$: 516.18 $[M+Na]^+$, found: 516.20 $[M+Na]^+$; **^{1}H NMR ($CDCl_3$) δ(ppm):** 7.67 (2H, d), 7.48 (2H, d), 7.19-7.33 (9H, m), 6.78 (2H, d, *J* = 8.4 Hz), 6.62 (2H, m), 5.05 (2H, m), 4.52 (1H, t, *J* = 5.7 Hz), 4.23-4.35 (2H, m), 4.10 (1H, t, *J* = 6.6 Hz), 2.72 – 3.00 (2H, m); **^{13}C NMR ($CDCl_3$) δ(ppm):** 171.4, 155.5, 143.4, 143.3, 140.9, 134.7, 130.0, 128.3, 128.3, 127.4, 126.8, 126.2, 124.8, 119.7, 115.2, 67.2, 66.9, 55.1, 47.1, 37.2.

*Benzyl 2-(S)-[(N-fluorenylmethoxycarbonyl)-amino]- 3-[p-(2'-phenyl-2'-methyl-acetyloxy)-phenyl] -propanoate(Fmoc-Tyr(O-2-Me-phenylacetyl)-OBn) (**1d**)* – 2-methyl-2-phenyl-acetic acid (107 mg, 0.71 mmol, 1.0 eq.) and HATU (283 mg, 0.74 mmol, 1.05 eq.) were dissolved in anhydrous DMF (3 mL) and stirred under N_2 at 0^0C (ice/water bath) until homogenization. DIPEA (129 µL, 0.78 mmol, 1.1 eq.) was added drop wise at 0^0C. After stirring for 5 min, the reaction mixture was warmed up to room temperature. Fmoc-Tyr(OH)-OBn (350 mg, 0.71 mmol, 1.0 eq.) was dissolved in anhydrous DMF (4 mL) and added to the activated acid. The mixture was then stirred overnight at room temperature in an N_2 atmosphere. The reaction mixture was quenched with saturated aqueous NH_4Cl solution (100 mL). Extraction with ethyl acetate (3x50 mL) and washing the combined organic fractions with saturated aqueous NH_4Cl (1x100 mL), brine (2x100 mL) followed by drying with anhydrous Na_2SO_4 and evaporation of solvents yielded the crude product. Purification was performed using flash chromatography (SiO_2, elution with hexane:ethyl acetate 10:1, then 5:1) resulting in 235 mg (53% yield) of the ester **1d**.

TLC: R_f = 0.65 (hexane:ethyl acetate 3:1); **^{1}H NMR ($CDCl_3$) δ(ppm):** 7.78 (2H, d), 7.58 (2H, d), 7.20-7.43 (14H, m), 6.97 (2H, d), 6.87 (2H, d), 5.33 (1H, d, *J* = 8.1 Hz), 5.16 (2H, m), 4.71 (1H, m), 4.35-4.49 (2H, m), 4.22 (1H, t, *J* = 6.9 Hz), 3.99 (1H, q^4, *J* = 6.9 Hz), 3.11 (2H, m), 1.65 (3H, d, *J* = 6.9 Hz); **^{13}C NMR ($CDCl_3$) δ(ppm):** 172.5, 170.8, 155.2, 149.6, 143.6, 141.1, 139.8, 132.9, 130.1, 128.4, 127.5, 126.9, 124.9, 121.3, 119.9, 67.4, 67.0, 54.8, 47.2, 45.9, 37.7, 18.7.

*Benzyl 2-(S)-amino- 3-[p-(2'-phenyl-2'-methyl-acetyloxy)-phenyl] -propanoate(H-Tyr(O-2-Me-phenylacetyl)-OBn) (**1e**)* – Compound **1d** (235 mg, 0.38 mmol, 1.0 eq.) was dissolved in DMF (4 mL). Piperidine (1 mL) was added and the mixture was stirred 20 min at room temperature. It was then evaporated to dryness under high vacuum. The residue was dissolved in ethyl acetate and then a 3-fold excess of hexane was added. Purification was performed using flash chromatography (SiO_2, elution with hexane:ethyl acetate 3:1 (the fluorenone-containing byproduct), then hexane ethyl acetate 1:3) resulting in 130 mg (86% yield) of the amine **1e**.

TLC: R_f = 0.15 (hexane:ethyl acetate 1:3); **LCMS (ESI):** RT = 8.30 min., *m/z* calculated for $C_{25}H_{26}NO_4$: 404.19 $[M+H]^+$, found: 404.16 $[M+H]^+$; **^{1}H NMR ($CDCl_3$) δ(ppm):** 7.28-7.44 (10H, m), 7.10 (2H, m), 6.90 (2H, m), 5.12 (2H, s), 3.96 (1H, q^4, *J* = 7.2 Hz), 3.74 (1H, t, *J* = 5.4 Hz), 2.85-3.15 (2H, m), 1.82 (2H, br s), 1.62 (3H, d, *J* = 6.9 Hz); **^{13}C NMR ($CDCl_3$) δ(ppm):** 174.7, 173.0, 149.8, 140.1, 135.5, 134.6, 130.4, 130.2, 128.9, 128.6, 128.5, 128.4, 127.6, 127.4, 121.4, 66.9, 55.8, 45.7, 40.4, 18.5;

*2-(S)-amino- 3-[p-(2'-phenyl-2'-methyl-acetyloxy)-phenyl]propanoic acid (H-Tyr(O-2-Me-phenylacetyl)-OH) (**1**)* **–** Compound **1e** (185 mg, 0.46 mmol, 1.0 eq.) was dissolved in methanol (10 mL). The solution was degassed. 5% Pd on charcoal (37 mg, 1 weight percent) was added. The flask was degassed and flushed with nitrogen (3 cycles). Then the reaction mixture was degassed again and filled with hydrogen gas under atmospheric pressure. After 90 min total substrate consumption was confirmed using TLC. The flask was degassed and flushed with nitrogen gas (3 cycles) before the mixture was filtered through a celite pad under reduced pressure. The pad was

subsequently washed with methanol (3 x 50 mL). The organic fractions were combined and evaporated yielding 141 mg (98% yield) of the expected compound **1**.

TLC: R_f = 0 (hexane:ethyl acetate 1:3); **LCMS (ESI):** RT = 7.12 min., *m/z* calculated for $C_{18}H_{20}NO_4$: 314.14 $[M+H]^+$, found: 314.08 $[M+H]^+$, *m/z* calculated for $C_{36}H_{39}N_2O_8$: 627.27 $[2M+H]^+$, found: 626.98 $[2M+H]^+$; **^{1}H NMR (d^6-DMSO+1 drop TFA) δ(ppm):** 8.27 (3H, br s), 7.39 (5H, m), 7.26-7.33 (3H, m), 6.99 (2H, m), 4.20 (1H, m), 4.08 (1H, q^4, *J* = 6.9 Hz), 3.08 (1H, d, *J* = 5.4 Hz), 1.50 (3H, d, *J* = 7.2 Hz); **^{13}C NMR (d^6-DMSO+1 drop TFA) δ(ppm):** 173.1, 152.5, 138.0, 135.2, 133.3, 131.5, 130.2, 130.0, 124.4, 55.8, 47.3, 37.9, 21.3;

Synthesis of coumarin ester 2. 2-Methyl-2-phenylacetic acid (250 mg, 1.66 mmol, 1.0 eq.) was dissolved in dry acetonitrile (2.5 mL). Dicyclohexylcarbodiimide (DCC, 377 mg, 1.83 mmol, 1.1 eq.) was added and the mixture cooled to 0^0C in ice/water bath. 7-hydroxycoumarin (324 mg, 2.00 mmol, 1.2 eq.) was added and stirred 15 min at 0^0C. The cooling bath was then removed and the mixture stirred at room temperature for 65 h. Next, the reaction mixture was quenched with saturated aqueous $NaHCO_3$ solution (50 mL). Extraction with ethyl acetate (3x25 mL) and washing the combined organic fractions with saturated aqueous $NaHCO_3$ (4x50 mL), brine (2x50 mL) followed by drying with anhydrous Na_2SO_4 and evaporation of solvents yielded the crude product. It was purified using flash chromatography on silicagel (25 g SiO_2, elution with mixture of cyclohexane:ethyl acetate 10:1) resulting in 303 mg (62% yield) of the expected product.

TLC: R_f = 0.3 (developed twice in cyclohexane:ethyl acetate 3:1); **LCMS (ESI):** RT = 13.04 min., *m/z* calculated for $C_{18}H_{15}O_4$: 295.10 $[M+H]^+$, found: 295.04 $[M+H]^+$; **^{1}H NMR ($CDCl_3$) δ(ppm):** 7.65 (1H, d, *J* = 9.3 Hz), 7.43 (1H, d, *J* = 8.4 Hz), 7.38 (4H, m), 7.33 (1H, m), 7.02 (1H, d, *J* = 2.4 Hz), 6.94 (1H, dd, *J* = 8.4 Hz, *J'* = 2.1 Hz), 6.37 (1H, d, *J* = 9.6 Hz), 4.00 (1H, q^4, *J* = 7.2 Hz), 1.63 (3H, d, *J* = 7.2 Hz); **^{13}C NMR ($CDCl_3$) δ(ppm):** 172.4, 160.3, 154.7, 153.4, 142.8, 139.6, 129.0, 128.5, 127.7, 127.5, 118.3, 116.7, 116.1, 110.3, 45.7, 18.4;

Synthesis of para-nitrophenol ester 3. 2-Methyl-2-phenylacetic acid (250 mg, 1.66 mmol, 1.0 eq.) was dissolved in dry acetonitrile (2.5 mL). Dicyclohexylcarbodiimide (DCC, 377 mg, 1.83 mmol, 1.1 eq.) was added and the mixture cooled to 0^0C in ice/water bath. *p*-Nitrophenol (278 mg, 2.00 mmol, 1.2 eq.) was added and stirred 15 min at 0^0C. The cooling bath was then removed and the mixture stirred at room temperature for 65 h. Next, the reaction mixture was quenched with saturated aqueous $NaHCO_3$ solution (50 mL). Extraction with ethyl acetate (3x25 mL) and washing the combined organic fractions with saturated aqueous $NaHCO_3$ (4x50 mL), brine (2x50 mL) followed by drying with anhydrous Na_2SO_4 and evaporation of solvents yielded the crude product. It was purified using flash chromatography on silicagel (25 g SiO_2, elution with mixture of cyclohexane:ethyl acetate 14:1) resulting in 407 mg (90% yield) of the expected product.

TLC: R_f = 0.35 (cyclohexane:ethyl acetate 10:1); **^{1}H NMR ($CDCl_3$) δ(ppm):** 8.22 (2H, m), 7.40 (4H, m), 7.34 (1H, m), 7.18 (2H, m), 4.00 (1H, q^4, *J* = 7.2 Hz), 1.64 (3H, d, *J* = 7.2 Hz); **^{13}C NMR ($CDCl_3$) δ(ppm):** 172.2, 155.6, 145.3, 139.4, 129.0, 127.7, 127.5, 125.2, 122.3, 45.7, 18.4

Supporting figures and tables

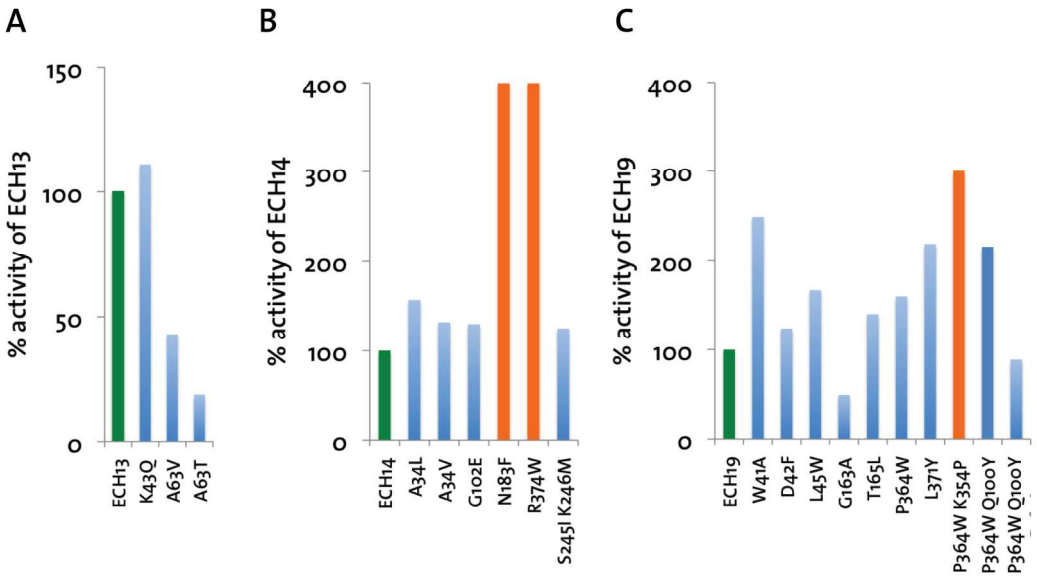


Figure 1S: Esterase activity of soluble **A)** ECH13, **B)** ECH14 and **C)** ECH19 variants. Coumarin release was measured in triplicate after addition of 100 µM substrate **2** to 5 µM of purified protein. Individual measurements deviated by approximately 5-15%. The plots show the relative activity of the variants compared to the parental design (green bar: parental enzyme, blue bars: 0 – 250 %, orange bars: 250 – 500 % of the parental design activity).

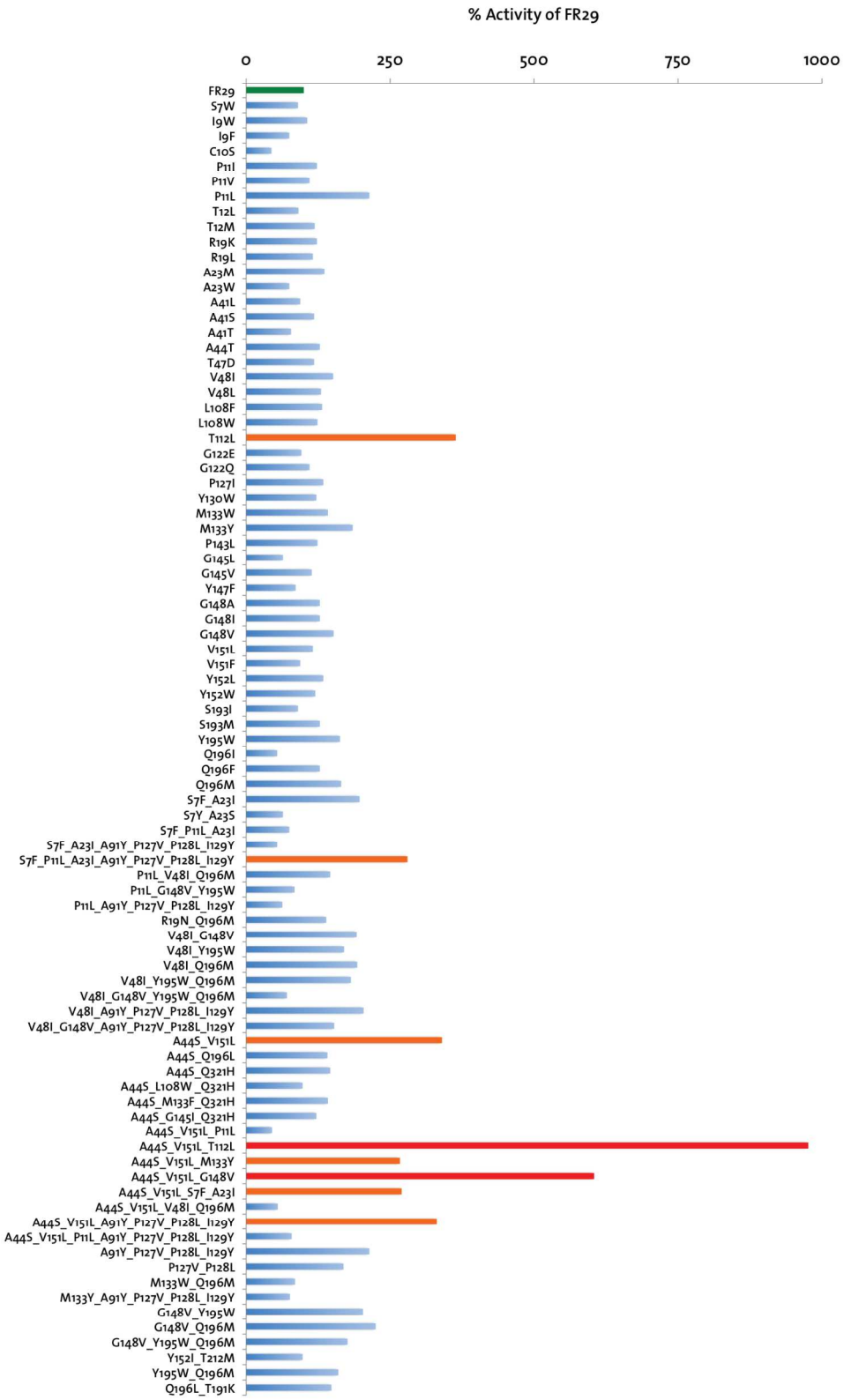

% Activity of FR29
0
250
500
750
1000
FR29
S7W
I9W
I9F
C10S
P11I
P11V
P11L
T12L
T12M
R19K
R19L
A23M
A23W
A41L
A41S
A41T
A44T
T47D
V48I
V48L
L108F
L108W
T112L
G122E
G122Q
P127I
Y130W
M133W
M133Y
P143L
G145L
G145V
Y147F
G148A
G148I
G148V
V151L
V151F
Y152L
Y152W
S193I
S193M
Y195W
Q196I
Q196F
Q196M
S7F_A23I
S7Y_A23S
S7F_P11L_A23I
S7F_A23I_A91Y_P127V_P128L_I129Y
S7F_P11L_A23I_A91Y_P127V_P128L_I129Y
P11L_V48I_Q196M
P11L_G148V_Y195W
P11L_A91Y_P127V_P128L_I129Y
R19N_Q196M
V48I_G148V
V48I_Y195W
V48I_Q196M
V48I_Y195W_Q196M
V48I_G148V_Y195W_Q196M
V48I_A91Y_P127V_P128L_I129Y
V48I_G148V_A91Y_P127V_P128L_I129Y
A44S_V151L
A44S_Q196L
A44S_Q321H
A44S_L108W_Q321H
A44S_M133F_Q321H
A44S_G145I_Q321H
A44S_V151L_P11L
A44S_V151L_T112L
A44S_V151L_M133Y
A44S_V151L_G148V
A44S_V151L_S7F_A23I
A44S_V151L_V48I_Q196M
A44S_V151L_A91Y_P127V_P128L_I129Y
A44S_V151L_P11L_A91Y_P127V_P128L_I129Y
A91Y_P127V_P128L_I129Y
P127V_P128L
M133W_Q196M
M133Y_A91Y_P127V_P128L_I129Y
G148V_Y195W
G148V_Q196M
G148V_Y195W_Q196M
Y152I_T212M
Y195W_Q196M
Q196L_T191K

Figure 2S: Esterase activity of soluble FR29 variants. Coumarin release was measured in triplicate after addition of 100 µM substrate **2** to 5 µM of purified protein. Individual measurements deviated by approximately 5-15%. The plot shows the relative activity of the individual variants compared to the parental design FR29 (green bar: parental design, blue bars: 0 – 250 %, orange bars: 250 – 500 %, red bars: > 500% of FR29 activity).

Table 1S: Overview of expressed designs. The theozyme numberings correspond to the theozymes as presented in Figure S10

Design	scaffold	# mutations to scaffold	soluble	Initial activity
Theozyme I				
ECH01	1eyn	6	yes	-
ECH02	1jcm	20	no	-
ECH03	1fp2	10	no	-
ECH04	1cil	12	yes	-
ECH05	1fp2	9	no	-
ECH06	1b4p	4	yes	-
ECH07	1ftx	17	yes	-
ECH08	1cjw	11	no	-
ECH09	1h1d	11	yes	-
ECH10	1dzu	11	yes	-
ECH11	1n9l	10	no	-
ECH12	1q11	14	yes	-
ECH13	1q91	9	yes	++
ECH14	1toi	13	yes	+
ECH15	1vhn	11	yes	-
ECH16	1vhn	11	yes	-
ECH17	1xjd	14	no	-
ECH18	2uvh	12	yes	-
ECH19	2uvh	11	yes	++
ECH20	4fua	14	no	-
FR25	1nah	14	yes	-
FR26	1qpr	13	yes	-
FR27	1pa9	11	no	-
FR28	1r5l	18	no	-
FR29	1mau	20	yes	+
FR30	1q91	21	yes	-
FR31	1q91	22	no	-
FR32	1cil	14	no	-
EA22	1eus	14	yes	-
EA29	1uyp	18	yes	-

EA30	2h13	16	no	-
Theozyme II				
EA23	1pt2	15	no	-
EA24	1v04	18	yes	-
EA27	1dzu	8	yes	-
EA34	2dri	19	no	-
EA35	1thf	15	yes	-
EA36	1is3	13	no	-
EA37	1dl3	13	no	-
EA28	1pii	13	yes	-
EA38	1c9u	31	no	-
EA39	1f5j	14	no	-
EA40	1st8	13	no	-
EA41	1yna	15	yes	-
Theozyme III				
1ajk_2	1ajk	15	yes	-
1ajk_3	1ajk	17	yes	-
1ukr_1	1ukr	20	no	-
1mac_3	1mac	14	yes	-
2ayh_1	2ayh	18	yes	-
1h0b_1	1h0b	18	yes	-
1dyp_1	1dyp	17	no	-
1mve_1	1mve	13	yes	-
1gbg_1	1gbg	15	yes	-
1f5j_1	1f5j	22	no	-
2jen_1	2jen	11	no	-
1mac_2	1mac	20	yes	-

Table 2S: Identification of the artificial ester hydrolases by ESI-MS.

Variant	$Mass_{calc}$ [Da]	$Mass_{exp}$ [Da]
FR29	38284.9	38284.7
C10A /H126A	38186.7	38186.0
A44S/T112L/V151L	38326.9	38326.5
ECH13	23836.1	23834.9
ECH13 C45A	23804.1	23803.3
ECH13 C45A/H100A	23738.0	23736.9
ECH14	44849.7	44851.6/44892.1
ECH14 C132A	44817.6	44817.1
ECH14 C132A/H104A	44751.5	44752.3
ECH19	47479.1	47479.7
ECH19 C161A	47447.1	47445.5
ECH19 C161A/H226A	47381.0	47379.7
ECH19 K354P/P364W	47537.2	47535.6

Table 3S: Melting temperatures of ester hydrolase designs were determined by curve fitting of the temperature dependent CD signals at 222nm with Sigma plot. Denaturation was irreversible in all cases.

ECH13		**ECH14**		**ECH19**		**FR29**	
Variant	[°C]	Variant	[°C]	Variant	[°C]	Variant	[°C]
wt-design	46	wt-design	42/50	wt-design	52	wt-design	71
C45A	45/49	C132A	47	C161A	52		
C45A/H100A	49	C132A/H104A	46	C161A/H226A	52	C10A/H126A	70
				K354P/P364W	47	A44S/T112L/V151L	69

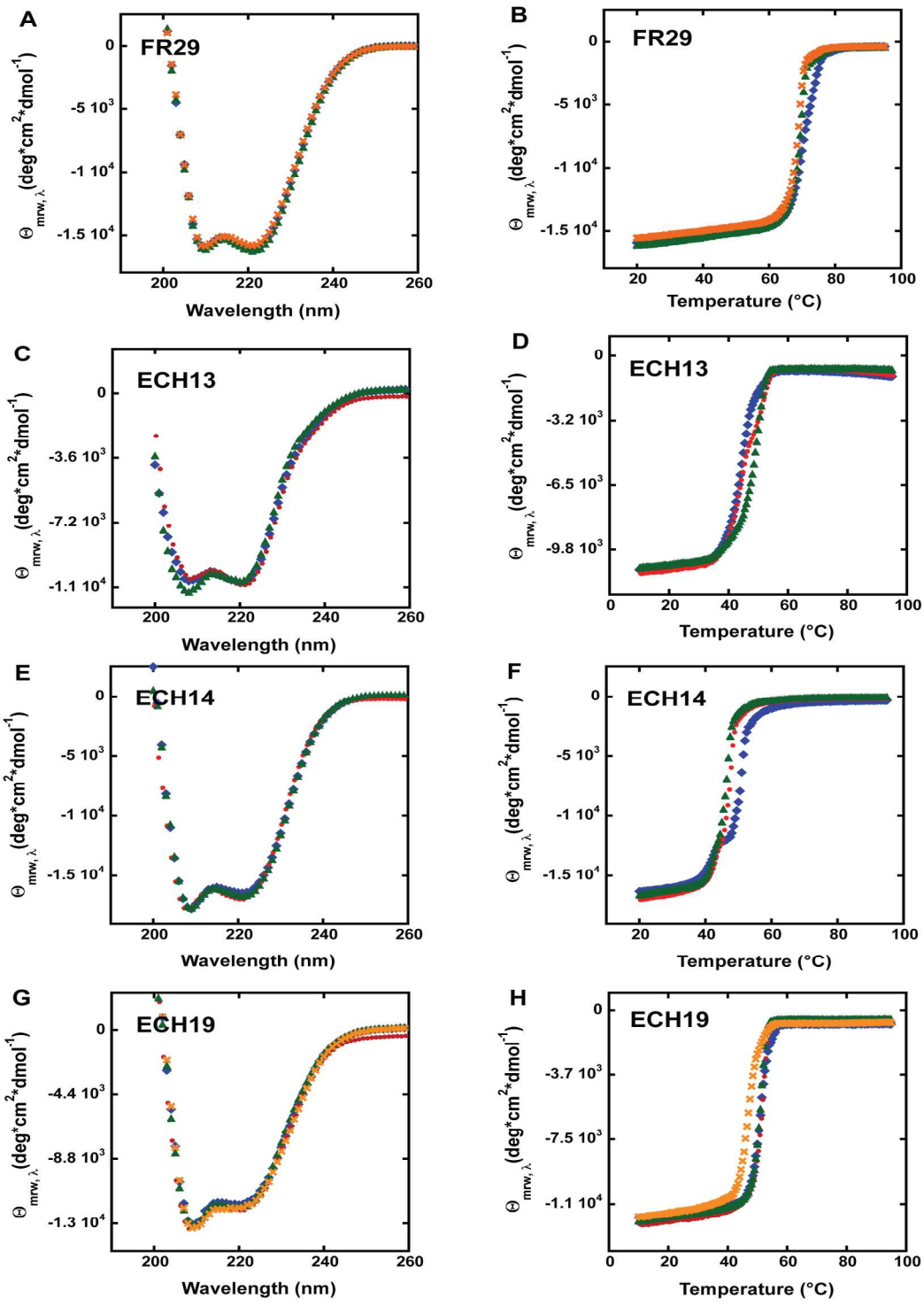


Figure 3S: CD spectra and melting curves of the *de novo* designed ester hydrolases, of their respective knockout mutants, and of the improved FR29 and ECH19 variants. The curves of the parental designs are depicted in blue, the traces of the single knockout variants (cysteine) are shown red, the traces of the double knockout variants (cysteine and histidine) are illustrated in green and the curves of the improved variants are highlighted in orange: **A-B)** FR29; **C-D)** ECH13; **E-F)** ECH14; **G-H)** ECH19.

Table 4S: Results of the mass spectrometric analysis of the *in silico* designs incubated with coumarin ester **2**.

	Before Incubation		After Incubation		
Design	Mass (Da) (ESI)		No. of modifications (ESI)	Location of modification (MALDI –MS/MS)	
	exp	calc		Peptide	Residue
FR29	38286.8	38284.9	1		
FR29 C10A/H126A	38187.8	38186.7	0		
FR29 A44S/V151L/M133Y	38346.3	38346.9	2	TIFSAICPTGVITIGR	Cys10 (AS)
ECH13	23837.1	23836.1	2	ACEQYGR RPCGSLEHHHHH	Cys45 (AS) Cys195
ECH13 C45A/H100A	23737.2	23738.0	n.d.		
ECH14	44851.6/ 44892.1	44849.7	7	NFGLYNESVGACT RGSAVAYVGFEERL TAQPGGHGALR (AS-His)* VWVYNPSSNCSK (AS-Cys)*	
ECH14 C132A/H104A	44752.3	44751.5	7		
ECH19	47480.0/ 47512.4	47479.1	1-2	MSWWGGNGR* AEYETGWDGHLSR* TVQETAEYFNKQGD*	
ECH19 C161A/H100A	47381.9	47381.0	0		

* Due to the low intensity of these peptides, the interpretations are speculative (AS stands for active site).

Table 5S: Results of the mass-spectroscopic analysis of the *in silico* designs after incubation with tyrosine substrate **3**.

	Before Incubation		After Incubation
Design	Mass (Da) (ESI)		No. of modifications (ESI)
	exp	calc	
FR29 A44S/T112L/V151L	38326.5	38326.9	0
ECH19 K354P/P364W	47537.2	47535.6	0

Table 6S: Analysis of FR29 variants.

Mutation(s)	$Mass_{calc}$ [g/mol]	$Mass_{exp}$ [g/mol]	Δ	ε [M^{-1} cm^{-1}]	Activity (% FR29)
S7W	38384.9			47330	90
I9W	38357.9			47330	106
I9F	38318.9			41830	75
C10S	38268.8	38268.2	0.6	41830	44
P11I	38300.9	38300.4	0.5	41830	123
P11V	38286.9			41830	110
P11L	38300.9			41830	214
T12L	38296.9			41830	91
T12M	38314.9	38314.6	0.3	41830	119
R19K	38256.8			41830	123
R19L	38241.8	38241.5	0.3	41830	116
A23M	38345.0	38344.9	0.1	41830	136
A23W	38400.0			47330	75
A41L	38326.9			41830	94
A41S	38300.9	38301.3	0.4	41830	118
A41T	38314.9	38315.4	0.5	41830	78
A44T	38314.9	38314.5	0.4	41830	128
T47D	38298.8	38298.4	0.4	41830	118
V48I	38298.9	38299.4	0.5	41830	151
V48L	38298.9	38299.7	0.8	41830	130
L108F	38318.9	38319.7	0.8	41830	132
L108W	38357.9	38357.4	0.5	47330	124
T112L	38296.9	38296.3	0.3	41830	364

G122E	38356.9	38357.7	0.8	41830	96
G122Q	38355.9	38355.0	0.9	41830	110

Mutation(s)	$Mass_{calc}$ [g/mol]	$Mass_{exp}$ [g/mol]	Δ	ε [M^{-1} cm^{-1}]	Activity (% FR29)
P127I	38300.9	38301.5	0.6	41830	134
Y130W	38307.9	38308.8	0.9	45840	122
M133W	38339.9	38340.6	0.7	47330	142
M133Y	38316.8	38315.2	1.6	43320	185
P143L	38300.9	38299.9	1.0	41830	124
G145L	38341.0	38341.0	0	41830	64
G145V	38326.9	38327.1	0.2	41830	114
Y147F	38268.9	38267.6	1.3	40340	86
G148A	38298.9	38299.8	0.9	41830	128
G148I	38341.0	38342.1	1.1	41830	128
G148V	38326.9	38326.3	0.6	41830	152
V151L	38298.9	38298.2	0.7	41830	116
V151F	38332.9	38333.6	0.7	41830	94
Y152L	38234.8	38235.2	0.4	40340	134
Y152W	38307.9	38308.6	0.7	45840	120
S193I	38310.9	38309.4	1.5	41830	90
S193M	38329.0	38327.9	1.1	41830	128
Y195W	38307.9	38308.5	0.6	45840	163
Q196I	38269.9	38269.5	0.4	41830	54
Q196F	38303.9	38303.7	0.2	41830	128
Q196M	38287.9	38287.0	0.9	41830	165
S7F_A23I	38387.0	38385.9	1.1	41830	197

S7Y_A23S	38377.0	38376.4	0.6	43320	64
S7F_P11L_A23I	38403.1	38401.8	1.3	41830	75
S7F_A23I_A91Y_P127V_P128L_I129Y	38547.2	38546.8	0.4	44810	54
S7F_P11L_A23I_A91Y_P127V_P128L_I129Y	38563.2	38366.7	3.5	44810	280
P11L_V48I_Q196M	38318.0	38317.2	0.8	41830	146

Mutation(s)	**$Mass_{calc}$ [g/mol]**	**$Mass_{exp}$ [g/mol]**	**Δ**	**ε [M^{-1} cm^{-1}]**	**Activity (% FR29)**
P11L_G148V_Y195W	38366.0	38365.2	0.8	45840	84
P11L_A91Y_P127V_P128L_I129Y	38461.1	38459.5	1.6	44810	63
R19N_Q196M	38245.8	38244.7	1.1	41830	139
V48I_G148V	38341.0	38340.5	0.5	41830	192
V48I_Y195W	38321.9	38320.9	1.0	45840	170
V48I_Q196M	38301.9	38301.5	0.4	41830	193
V48I_Y195W_Q196M	38325.0	38323.6	1.4	45840	182
V48I_G148V_Y195W_Q196M	38367.1	38365.4	1.7	45840	71
V48I_A91Y_P127V_P128L_I129Y	38459.1	38457.7	1.4	44810	204
V48I_G148V_A91Y_P127V_P128L_I129Y	38501.1	38500.6	0.5	44810	153
A44S_V151L	38314.9	38314.4	0.5	41830	340
A44S_Q196L	38285.9	38285.4	0.5	41830	141
A44S_Q321H	38309.9	38310.6	0.7	41830	146
A44S_L108W _Q321H	38383.8	38383.4	0.4	47330	98
A44S_M133F_Q321H	38325.8	38324.9	0.9	41830	142
A44S_G145I_Q321H	38366.0	38364.7	1.3	41830	122
A44S_V151L_P11L	38330.9	38330.6	0.3	41830	45
A44S_V151L_T112L	38326.9	38326.5	0.4	41830	976
A44S_V151L_M133Y	38346.9	38346.6	0.3	43320	267

A44S_V151L_G148V	38357.0	38356.5	0.5	41830	604
A44S_V151L_S7F_A23I	37417.1	38416.9	0.2	41830	270
A44S_V151L_V48I_Q196M	38332.0	38332.0	0	41830	55
A44S_V151L_A91Y_P127V_P128L_I129Y	38475.1	38474.0	1.1	44810	331
A44S_V151L_P11L_A91Y_P127V_P128L_I129Y	38491.1	38490.9	0.2	44810	79
A91Y_P127V_P128L_I129Y	38445.0	38444.8	0.2	44810	214
P127V_P128L	38302.9	38301.6	1.3	41830	169
M133W_Q196M	38342.9	38342.4	0.5	47330	85
Mutation(s)	**$Mass_{calc}$ [g/mol]**	**$Mass_{exp}$ [g/mol]**	**Δ**	**ε [M^{-1} cm^{-1}]**	**Activity (% FR29)**
M133Y_A91Y_P127V_P128L_I129Y	38477.0	38477.0	0	46300	76
G148V_Y195W	38350.0	38349.5	0.5	45840	203
G148V_Q196M	38330.0	38329.6	0.4	41830	225
G148V_Y195W_Q196M	38353.0	38352.6	0.4	45840	176
Y152I_T212M	38264.9	38265.3	0.4	40340	98
Y195W_Q196M	38311.0	38309.9	1.1	45840	160
Q196L_T191K	38297.0	38297.9	0.9	41830	148

Table 7S: Analysis of ECH13 variants.

Mutation(s)	Soluble (y/n)	$Mass_{calc}$ [g/mol]	$Mass_{exp}$ [g/mol]	Δ	ε [M^{-1} cm^{-1}]	Activity (% ECH13)
D10W	n	23907.2			36440	
V16I	n	23850.1			30940	
E17M	n	23838.2			30940	
E17W	n	23893.2			36440	
K43Q	y	23836.1	23834.8	1.3	30940	111
A63V	y	23864.2	23864.0	0.2	30940	43
A63T	y	23866.1	23865.9	0.2	30940	19
F68W	n	23875.1			36440	

Table 8S: Analysis of ECH14 variants.

Mutation(s)	Soluble (y/n)	$Mass_{calc}$ [g/mol]	$Mass_{exp}$ [g/mol]	Δ	ε [M^{-1} cm^{-1}]	Activity (% ECH14)
A34L	y	44891.7	44892.6/ 44932.6	0.9/ 40.9	42860	157
A34V	y	44877.7	44877.2/ 44919.4	0.5/ 41.7	42860	132
G102E	y	44921.7	44921.8/ 44961.1	0.1/ 39.4	42860	130
N183F	y	44882.7	44881.0/ 44924.9	1.7 42.2	42860	400
R374W	y	44879.7	44879.5/ 44921.2	0.2/ 41.5	48360	400
S245I K246M	y	44878.8	44879.3	0.5	42860	125

The mass spectra of the ECH14 variants always exhibited a second peak of approximately +42 Da corresponding to acetylation of the proteins.

Table 9S: Analysis of ECH19 variants.

Mutation(s)	Soluble (y/n)	$Mass_{calc}$ [g/mol]	$Mass_{exp}$ [g/mol]	Δ	ε [M^{-1} cm^{-1}]	Activity (% ECH19)
W41A	y	47364.0	47364.0	0	90300	249
D42F	y	47511.2	47511.2	0	95800	124
L45W	y	47552.2	47552.2	0	101300	167
G163A	y	47493.2	47479.7	13.5	95800	50
T165L	y	47491.2	47490.1	1.1	95800	140
T215W	n	47564.3			101300	
P364W	y	47568.2	47567.4	0.8	101300	160
L371Y	y	47529.2	47528.5	0.7	97290	219
P364W K354P	y	47537.2	47535.6	1.6	101300	300
P364W Q100Y	y	47603.3	47601.8	1.5	101300	216
P364W Q100Y G163A	y	47617.3	47616.2	1.1	102790	89

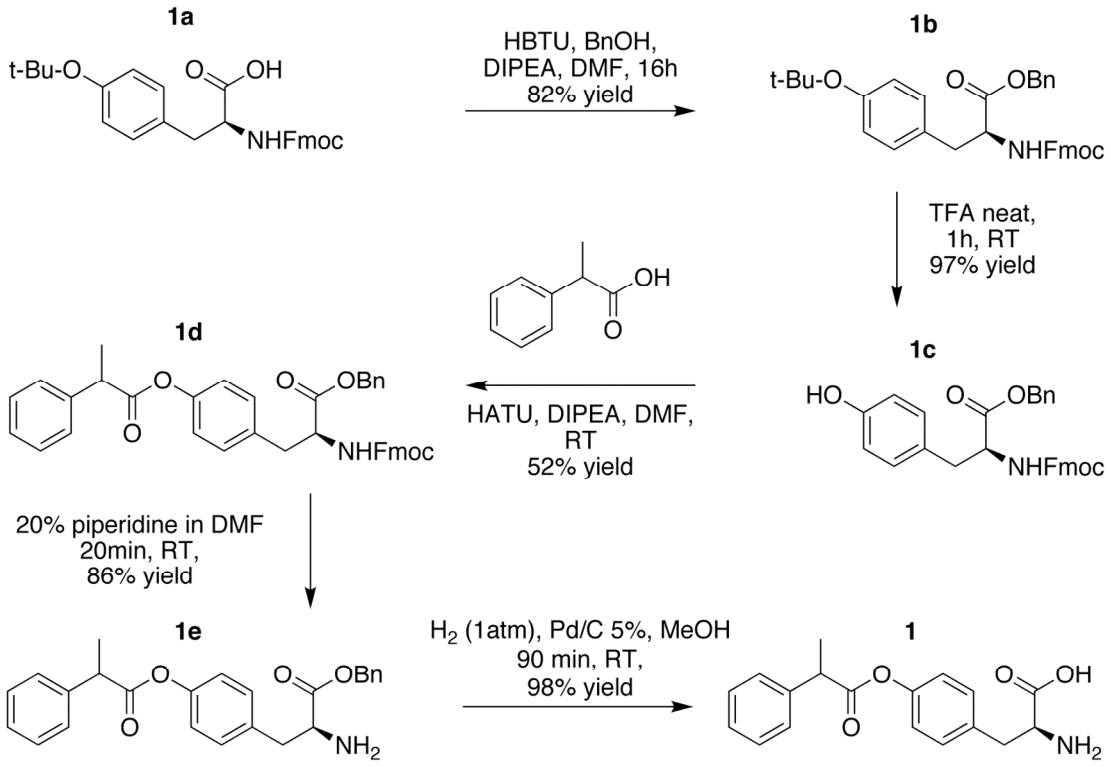


Figure 4S: Synthesis scheme of tyrosine ester **1.**

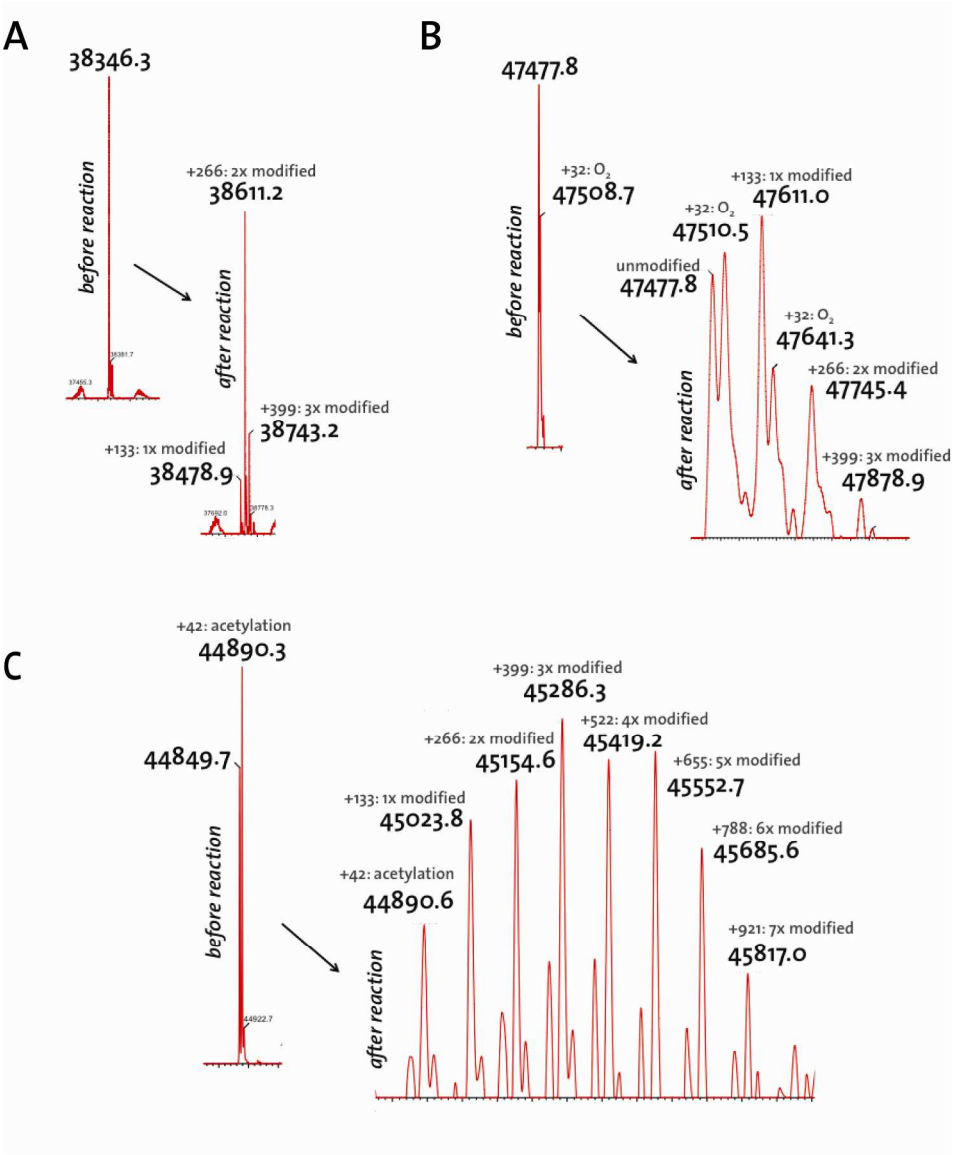


Figure 5S: Exemplary mass spectra before and after the addition of the coumarin ester 1 to **A)** FR29 A44S M133Y V151L, **B)** ECH19 and **C)** ECH 14.

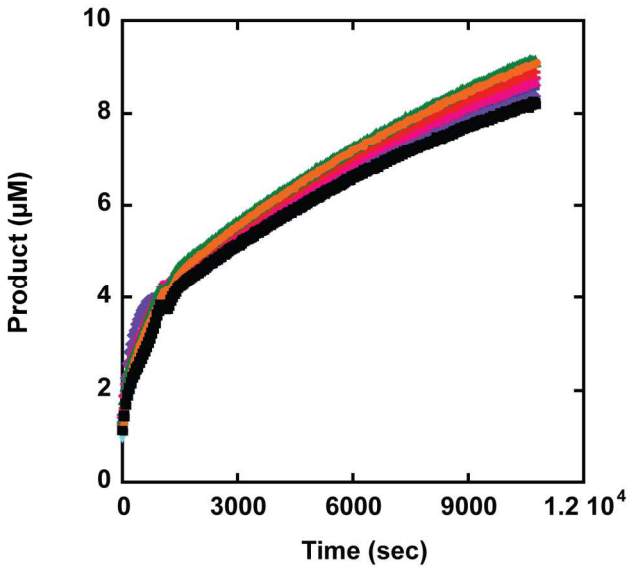


Figure 6S: Progress curve of the conversion of coumarin ester **2** by a FR29 variant in the presence of small nucleophiles. FR29 A44S/T112L/V151L (5 µM) was incubated with 50 µM of methoxyamine (cyan, pK_a = 4.6), aniline (blue; pK_a = 4.6), 2-aminobenzoic acid (purple; pK_a = 2.0), 4-aminobenzoic acid (pink; pK_a = 2.3), (*R*)-1-phenylethanamine (red), (*S*)-1-phenylethanamine (green), benzylthiol (orange) and no nucleophile (black) before 50 µM of coumarin ester **2** was added to the reaction mixture. The coumarin release was monitored in a plate reader at 29°C and pH 7.5. For all progress curves the corresponding background reactions without enzyme were subtracted. In the case of hydroxylamine the background reaction became dominating (data not shown).

Table 10S: X-ray crystallography and structure refinement statistics

General information				
PDB id	3U13	3U1O	3U1V	3UAK
NESG id	OR51	OR49	OR52	OR54
Gene	ECH13	ECH19	FR29	ECH14
Type	apo	apo	apo	apo
Crystal parameters				
Space group	$P4_32_12$	$P2_12_12$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	73.44	109.14	97.75	67.65
b (Å)	73.44	129.20	100.76	81.81
c (Å)	105.08	72.18	188.23	159.71
α (°)	90	90	90	90
β (°)	90	90	90	90
γ (°)	90	90	90	90
Z (number mols./au)	1	2	4	2
Data quality				
Beam line/X-ray source	BL9-2	X4C	X4C	X4C
Resolution range (Å)	30-1.6	30-2.6	30-2.8	30-3.2
Total reflections	2254145	394063	600893	375667
Observed reflections	72527	36229	42465	26157
R_{merge}	0.05/0.054	0.146/0.433	0.124/0.417	0.122/0.645
Mean redundancy	18.0/17.1	4.9/4.9	2.5/2.4	3.1/2.6
Completenes (%)	100.0/100.0	99.5/100.0	90/9/63.8	84.8/80.7
<I>/<σI>	48.7/4.1	11.4/3.5	8.0/1.8	8.9/1.7
Refinement				
Resolution range (Å)	30-1.6	30-2.5	30-2.8	30-3.2
Number of reflections	36618	36117	42365	26157
R_{work}	0.178	0.203	0.214	0.210
R_{free}	0.195	0.258	0.290	0.287
Number of prot. atoms	1642	6583	9937	6146
Number of waters	248	441		
Number of ligand atoms				
Overall mean B factors:				
-protein	24.2	19.3	48.4	74.2
-water	36.5	23.4	28.1	
-ligand				
RMSD bond length (Å)	0.008	0.008	0.009	0.009
RMSD bond angles (o)	1.21	1.18	1.17	1.20
Ramachandran plot				
-most favoured (%)	91.8	93.8	88.3	85.0
-additional allowed (%)	7.6	5.8	11.1	13.7
-generously allowed (%)	0.6	0.3	0.6	1.3
-disallowed (%)	0	0.1	0	0

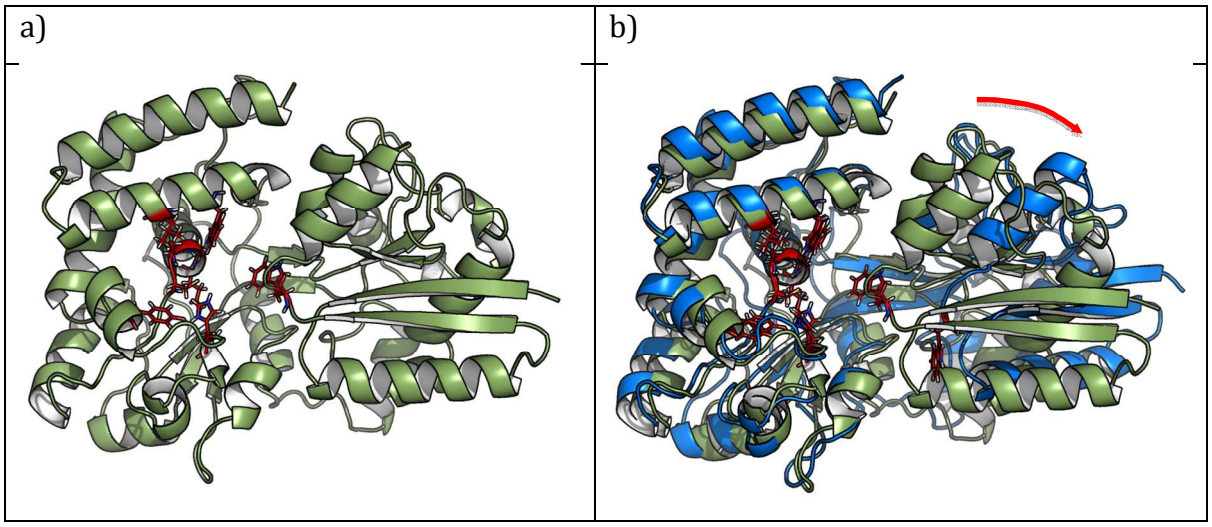


Figure S7. Molecular Dynamics of ECH19 in cartoon representations with the active site residues highlighted (red sticks). (a) The computational design. (b) The equilibrated MD structure superimposed onto the computational design in (a). The backbone RMSD is 3.5 Å.

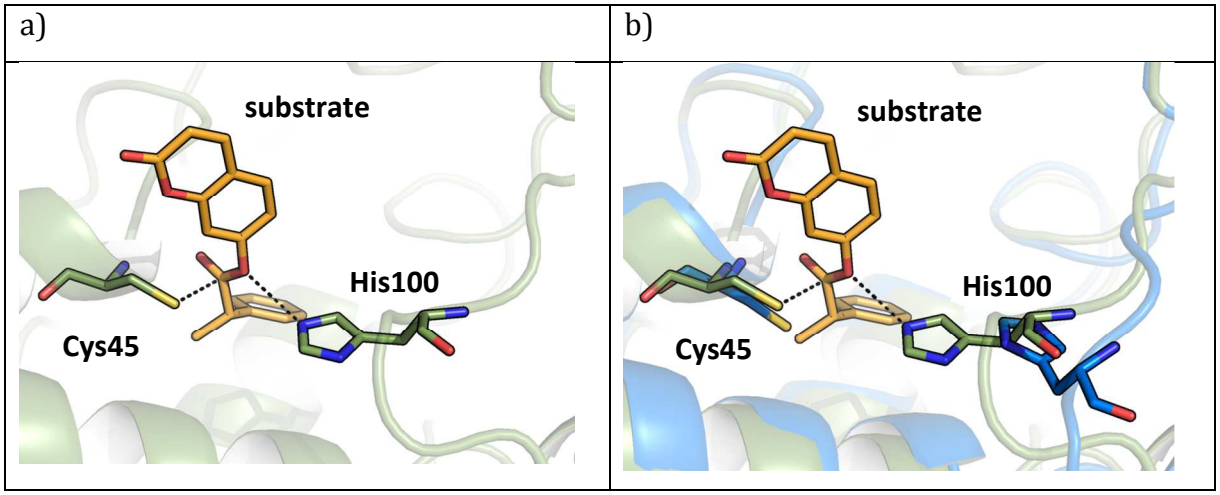


Figure S8. Molecular Dynamics of ECH13 (active site shown). (a) Computational design with the docked substrate in orange. (b) MD (blue) over computational design: His100 is part of a flexible loop and does not remain pre-oriented at the designed position.

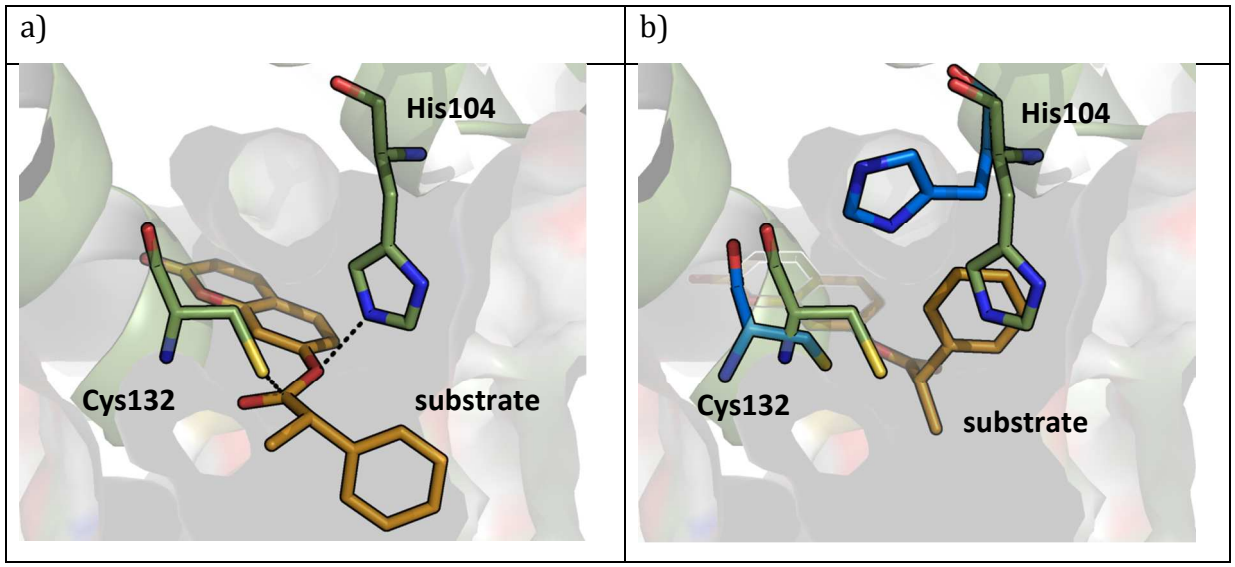
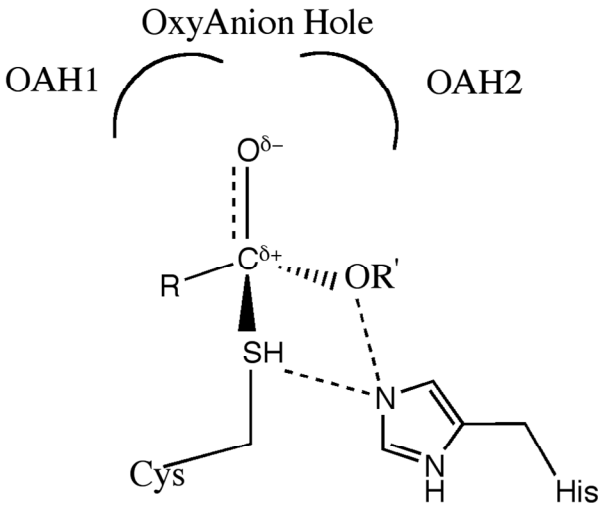


Figure S9. Molecular Dynamics of ECH14 (active site shown). (a) Computational design with the docked substrate in orange. (b) MD (blue) over computational design: His104 occupies a catalytically incompetent conformation.



Theozyme I: OAH1 = backbone-NH OAH2 = none

TheozymeII: OAH1 = sidechain Hbond donor OAH2 = sidechain Hbond donor

Theozyme III: OAH1 = explicit H_2O OAH2 = explicit H_2O/sidechain Hbond donor

Figure S10: Schematic representation of the three theozymes used during the matching stage of the computational design process. A sidechain Hbond donor could be any one of the following atoms: Ser/Thr-Og; Tyr-OH; Asn-Nd; Gln-Ne; His-Nd; His-Ne;