Supplementary Material For:

Computational Design of Enone-Binding Proteins with Catalytic Activity for the Morita-Baylis-Hillman Reaction

Sinisa Bjelic[1,#], Lucas G. Nivon[1,#], Nihan Çelebi-Ölçüm[2,3], Gert Kiss[2], Carolyn F. Rosewall[4], Helena M. Lovick[4], Erica L. Ingalls[4], Jasmine Lynn Gallaher[1], Jayaraman Seetharaman[5], Scott Lew[5], Gaetano Thomas Montelione[5], John Francis Hunt[5], Forrest Edwin Michael[4], K. N. Houk[2], David Baker[1,6,*]

Supplementary Methods

Theozyme setup

The theozyme was constructed with the substrates stacking on top of each other to exploit the maximum interaction between cyclohexenone and 4-nitrobenzaldehyde. The theozyme contained three stereocenters, with the nucleophile attacking the beta carbon of cyclohexenone to create an *(S)* configuration at the anomeric carbon. The alpha carbon of the cyclohexenone was in an *(S)* configuration and the carbon of the aldehyde in an *(R)* configuration.

Design filtering

After initial filtering by scaffold and energy terms, designs were also ranked by other factors such as the number of unsatisfied hydrogen bonds, surface complementarity, and active-site rigidity upon repacking without the ligand. Filtered designs were inspected manually to select the best designs for testing. Those final designs were adjusted by manual inspection to place backing-up interactions to the catalytic interactions (backing-up residues often lie outside of the designable shell in Rosetta), or to place a specific residue where Rosetta was not able to clearly distinguish a best residue at a given position. Missing residues from the original construct used in expression of the scaffold were added and a glycine-serine dipeptide was appended before the C-terminal Hexa-His tag.

Gene synthesis and mutagenesis, protein expression and purification

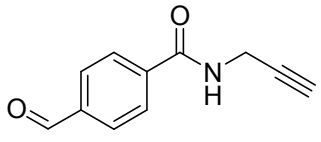
Final designed genes were synthesized by Genscript (Genscript, USA Inc. Piscataway NJ) and cloned into a pET29b+ vector using NdeI and XhoI cleavage sites. This places a leucine-glutamate-hexa-histidine tag at the C-terminus of each polypeptide chain. Genes were transformed into BL21(DE3) pLysS *E. coli* cells and grown in LB-autoinduction medium for 6 hours at 37 deg. C followed by 20-24 hours of induction at 18 deg. C. The lysate was filtered and passed over Qiagen Ni-NTA columns, followed by elution in a 250 mM imidazole solution. Proteins were concentrated and dialyzed into PBS (Phosphate buffer saline; 1X, pH 7.5) buffer for subsequent testing.

The Kunkel mutagenesis protocol was used for single point and saturation mutagenesis (*1*) and mutant gene variants were confirmed by sequencing (Genewiz, Inc.).

Synthesis of 3, 4 and aldol side-product 5

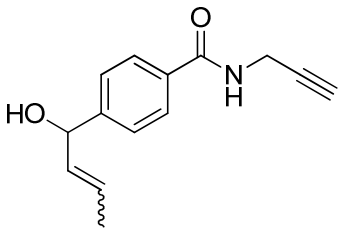
All reactions were performed under nitrogen atmosphere using flame-dried glassware. Infrared spectra were measured on a Perkin Elmer Spectrum RX I Spectrometer. Small-molecular mass spectra were collected on a JEOL HX-110 Mass Spectrometer with FAB or electron impact ionization, or a Bruker Esquire 1100 Liquid Chromatograph - Ion Trap Mass Spectrometer. Column chromatography was performed using silica gel (Sorbent Technologies, 60 Å, 230-400 mesh). NMR spectra were recorded on Bruker AV-300 or AV-500 spectrometers. ^{1}H NMR chemical shifts (δ) are reported in parts per million (ppm) downfield of TMS and are referenced relative to TMS (0.00 ppm) or residual protonated $CHCl_3$ (7.26 ppm). ^{13}C NMR chemical shifts (δ) are reported in parts per million (ppm) relative to the carbon resonance of $CDCl_3$ (77.0 ppm).

Materials. THF and CH_2Cl_2 were degassed and dried on solvent columns of neutral alumina. All other commercial reagents were used as received. Deuterated solvents were purchased from Cambridge Isotope Laboratories, Inc., stored over 4Å molecular sieves, and were used without further purification.



4-Formyl-*N*-(prop-2-ynyl)benzamide (1). 4-Carboxylbenzaldehyde (2.164 g, 14.4 mmol), propargylammonium chloride (1.313 g, 14.1 mmol), 1-hydroxybenzotriazole (1.9 g, 14.1 mmol) *N*,*N*'-diisopropylcarbodiimide (2.33 mL, 14.1 mmol) and triethylamine (2.0 mL, 14.1 mmol) were combined in dichloromethane and stirred at room temperature. The resulting mixture was diluted with dicholoromethane and washed with 1M HCl, 1M $NaHCO_3$ and water, dried over $MgSO_4$ and filtered and concentrated. The product was purified by column chromatography EtOAc/Hex followed by recrystallization from EtOAc/Hex obtaining a white solid.

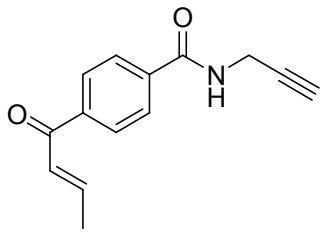
^{1}H NMR (500 MHz, $CDCl_3$): δ 10.08 (s, 1H), 7.97-7.93(m, 4H), 6.43 (br s, 1 H), 4.28 (dd, *J* = 5.0 Hz, 2.4 Hz, 2H), 2.31(t, *J* = 2.4 Hz, 1H). ^{13}C NMR (125 MHz, $CDCl_3$): δ 191.4, 165.9, 138.8, 138.5, 129.9, 127.8, 78.9, 72.3, 29.9. GC/MS (CI, m/z): 187(21), 158(35), 133(100), 105(52), 77(53), 51(38), 39(15). FTIR (KBr, cm^{-1}): 3314, 3242, 2833, 2732, 2117, 1740, 1690, 1643, 1572, 1542, 1499, 1420, 1388, 1352, 1323, 1298, 1258, 1209, 1182, 1154, 1051, 1016, 986, 920, 850, 798, 757, 709, 683.



4-(1-Hydroxybut-2-enyl)-*N*-(prop-2-ynyl)benzamide (**2**). In an oven dried 250 mL 2-necked rbf under N_2, 4-formyl-N-(prop-2-ynyl)benzamide (**1**, 0.317 g, 1.7 mmol) was dissolved in THF (75 mL). The solution was then cooled to 0 °C, and propenylmagnesium bromide (0.5 M, 10.5 mL) was added dropwise over 30 min. The

mixture was immediately quenched at 0 °C by adding water, and was then extracted with ethyl acetate (2 X 75 mL). The organic layers were combined and washed with water (25 mL), and sat. NH_4Cl (25 mL), then dried over $MgSO_4$, filtered and concentrated. The product was obtained as a thick oil (105 mg, 27% yield) of a 3:2 mixture of *E*/*Z* isomers after column chromatography (1-2% MeOH/DCM).

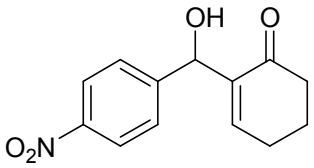
^{1}H NMR (500 MHz, $CDCl_3$, observed as a 3:2 mixture of isomers): δ 7.77 (d, *J* = 8.8 Hz, 4H, both), 7.46 (t, *J* = 9.0 Hz, 4H, both), 6.43 (s, 2H, both), 5.77 (dq, *J* = 15.0, 6.5 Hz, 1H, major), 5.75 – 5.60 (m, 3H, both), 5.20 (d, *J* = 7.0 Hz, 1H, major), 4.24 (dd, *J* = 5.0, 2.5 Hz, 4H, both), 2.29 (t, *J* = 2.5 Hz, 2H, both), 2.26 (br s, 2H, both), 1.83 (d, *J* = 6.9 Hz, 3H, minor), 1.72 (d, *J* = 6.4 Hz, 3H, major). ^{13}C NMR (125 MHz, $CDCl_3$, observed as a 3:2 mixture of isomers): δ 166.9(both), 147.6(minor), 147.3(major), 133.2(major), 132.7(major), 132.7(minor), 132.4(minor), 128.3(major), 127.2(minor), 127.15(major), 127.1(minor), 126.3(major), 126.0(minor), 79.5(both), 74.7(major), 71.9(both), 68.9(minor), 29.8(both), 17.6(major), 13.4(minor). GC/MS (CI, m/z): 229(1), 210(57), 157,(54), 128(100), 77(16), 51(16). FTIR (thin film, cm^{-1}): 3295, 2916, 2100, 1642, 1612, 1570, 1542, 1500, 144, 1421, 1353, 1305, 1152, 1047, 967, 920, 858.



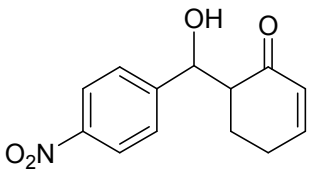
(*E*)-(4-But-2-enoyl-*N*-(prop-2-ynyl)benzamide (3). In a 4 dram vial under N_2 **2** (98 mg, 0.43 mmol) was dissolved in 10 mL of dichloromethane. Dess-Martin periodinane (182

mg, 0.43 mmol) was added to the mixture and was allowed to stir overnight. The mixture was diluted with dichloromethane (20 mL) and then washed with saturated $NaHCO_3$ (2 X 50 mL), water (50 mL) and dried over $MgSO_4$, filtered and concentrated. The crude material was a 1:1 mixture of *E*/*Z* isomers, which were separated by column chromatography (1:4 EtOAc/Hex) obtaining the (*E*) isomer as white solid (32 mg, 33% yield).

^{1}H NMR (500 MHz, $CDCl_3$): δ, 7.96(d, *J* = 8.2 Hz, 2H), 7.87(d, *J* = 8.3 Hz, 2H), 7.10(dq, *J* = 15.5, 7.0 Hz, 1H), 6.89(dd, *J* = 15.5, 1.5 Hz, 1H), 6.54 (br s, 1 H), 4.28 (dd, *J* = 5.0 Hz, 2.4 Hz, 2H), 2.31 (t, J = 2.45 Hz, 1H), 2.02(dd, *J* = 7.0, 1.5 Hz, 3H) ^{13}C NMR (125 MHz, $CDCl_3$): δ 190.1, 166.2, 146.3, 140.5, 137.1, 128.7, 127.4, 127.3, 79.2, 72.1, 29.9, 18.7. GC/MS (CI, m/z): 227(44), 173(100), 115(31), 76(29), 69(52). FTIR (KBr, cm^{-1}): 3568, 3280, 3237, 2947, 2123, 1670, 1637, 1617, 1560, 1533, 1499, 1437, 145, 1354, 1336, 1291, 1224, 1156, 1106, 1015, 993, 964, 923, 875, 823, 765, 697, 668, 647.



2-(Hydroxy(4-nitrophenyl)methyl)cyclohex-2-enone (4): Prepared as previously reported (*2*), spectral data matches literature values. ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.20 (d, *J* = 8.7 Hz, 2H), 7.57 (d, *J* = 8.7 Hz, 2H), 6.87 (t, *J* = 3.9 Hz, 1H), 5.63 (d, *J* = 5.7 Hz, 1H), 3.68 (d, *J* = 5.7 Hz, 1H), 2.46 (m, 4H), 2.01 (m, 2H).



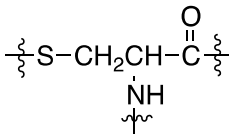
6-(Hydroxy(4-nitrophenyl)methyl)cyclohex-2-enone (5): In a flame-dried 25 mL round-bottomed flask diisopropylamine (0.3 mL, 2.2 mmol) and THF (10 mL) were combined and cooled to -78 °C. Then *n*-butyllithium (2.2 M, 0.9 mL, 2 mmol) was added slowly. The solution was stirred at -78 °C for 30 min then 2-cyclohexenone (0.19 mL, 2 mmol) was added. The reaction mixture was again stirred at -78 °C for 30 min. Then 4-nitrobenzaldehyde (0.302 g, 2 mmol) was added. After one minute, sat. NH_4Cl was added to quench the reaction. The mixture was extracted with Et_2O (3X). The organic layers were combined and washed with water, sat. NaCl, dried ($MgSO_4$), and concentrated. The product was obtained as a 3:1 mixture of diastereomers which were partially separable by column chromatography (EtOAc/Hexanes). Spectral data matched literature values (*3*). *Anti* diastereomer (major); ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.20 (d, 2 H, J = 8.7 Hz), 7.55 (d, 2 H, J = 8.7 Hz), 7.06 (m, 1 H), 6.08 (d, 1 H, J = 9.9 Hz), 4.97 (m, 1 H), 2.8-2.5 (m, 1 H), 2.5-2.2 (m, 2 H), 1.7-1.5 (m, 2 H). *Syn* diastereomer (minor); ^{1}H NMR (300 MHz, $CDCl_3$): 8.24 (dd, 2 H, J = 8.4, 1.8 Hz), 7.54 (d, 2 H, J = 8.4 Hz), 7.00 (m, 1 H), 6.13 (dd, 1 H, J = 9.6, 3.0 Hz), 6.10 (t, 1 H, J = 1.2 Hz), 2.91 (d, 1 H, J = 4.8 Hz), 2.70 (m, 1 H), 2.5-2.2 (m, 2 H), 2.00 (m, 1 H), 1.62 (m, 1 H).

Molecular Dynamics Simulations

Molecular dynamics (MD) were performed with Amber 11 (*4*) using explicit solvent and periodic boundaries to investigate the dynamical behavior of BH25 and

BH32 proteins. Simulations were run for 20-50 ns, or for one microsecond where specified.

System Preparation. Simulation systems were set up by placing the protein, including the co-crystallized water molecules from the scaffold, at the center of the simulation box and solvating the protein with TIP3P (*5*) water molecules ensuring a solvent layer of 10 Å around the protein. This resulted in the addition of ~10,000 – 20,000 solvent molecules depending on the scaffold and a system size of ~72,000 atoms for BH25 and ~32,500 atoms for BH32. The systems were neutralized by addition of explicit counter ions. All systems were parameterized using the Stony Brook modification of the Amber 99 force field (*6*). The parameters for the substrates, cyclohexenone and 4-nitrobenzaldehyde, were generated with the antechamber module of Amber 11 (*4*) using the general Amber force field (GAFF) (*7*) with partial charges set to fit the electrostatic potential generated at HF/6-31G(d) level of theory by RESP (*8*). The charges were calculated according to the Merz-Singh-Kollman scheme (*9, 10*) using Gaussian 03 (*11*). For the parameterization of the covalently bound intermediate, the system for calculating RESP charges consisted of the alkoxide intermediate bound to the nucleophilic cysteine



(int2, where Nu =), assuring a total charge of -1.0 for the unit. LEaP module of Amber 11 was used to split the unit into two, and to generate separate libraries for the alkoxide intermediate (ALK) and the non-standard cysteine residue (CYC).

The systems were initially minimized for the positions of water molecules and ions, with harmonic restraints of 150 kcal/mol applied to the solute. Initial minimization was followed by an unrestrained minimization of all atoms. The systems were heated gently in six steps of 50 K for 50 ps (from 0 K to 300 K) at constant volume with a time step of 1 fs. Each system was equilibrated for 2 ns with a 2 fs time step in the NVT ensemble at 300 K using the Langevin equilibration scheme. The systems were then equilibrated for 2 ns with a 2 fs time step at a constant pressure of 1 atm. Harmonic

restraints of 30 kcal/mol were applied to the solute during the heating and equilibration stages, and water molecules were triangulated using the SHAKE algorithm.

Production MD. Multiple 20 – 50 ns production MD simulations were performed for each system (with and without the substrate bound to the active site) using PMEMD (*12*) in the isothermal-isobaric ensemble (NPT) with a time step of 2 fs. Long-range effects were modeled using the particle-mesh-Ewald method (*13*). This general MD protocol has been recently described for the evaluation and ranking of enzyme designs (*14*).

For microsecond production runs of BH32 we used DESRES's Anton special purpose machine (*15*) at the Pittsburgh Supercomputing Center.

Trajectory analysis. Geometries and velocities were saved every 0.2 ps, resulting in a total of 100,000 frames from each production run. Post-MD data extraction and analysis was performed using the ptraj module of Amber 11.

Crystallization, data collection and structure determination

The crystals of SeMet-BH32 were grown in two different conditions by mixing 1 µl of protein sample with 1 µl of reservoir solution consisting of 0.1M PBS buffer PH 7.5 25% PEG 3350. Condition two is the same as condition one plus the addition of 0.1M cyclohexenone. The crystals were obtained by the hanging drop vapor diffusion method. The crystals of SeMet-BH25 N43Y were obtained by mixing 1 µl of protein sample with 1 µl of reservoir solution consisting of 1.44M potassium acetate, 50mM MES, PH 6 by the micro-batch under-oil method. Both crystals were grown at 18°C, cryo-protected with 20% glycerol and flash-cooled in liquid nitrogen. Diffraction data sets were collected at

the peak of the selenium K edge on a single crystal using the beam line X4A with a Quantum 4R detector at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Data were integrated and scaled with the HKL2000 package (Otwinowski and Minor, 1997). Matthew's coefficient calculations indicated one molecule per asymmetric unit in the monoclinic space group for BH32 and three molecules per symmetric unit in the tetragonal space group for BH25 N43Y.

The structures of BH32 and BH25 N43Y were solved by the single-wavelength anomalous dispersion (SAD) phasing method by SHELX (*16*) using a SeMet-substituted crystal. An experimental electron density map was obtained using ShelxD. After phase refinement we constructed an initial model with resolve (*17*) extended the model using ARP/wARP (*18*) and refined it with Refmac (*19*) and CNS (*20*). Model building was performed using Coot (*21*). Several cycles of simulated annealing and minimization were carried out using the CNS program package (*20*). The R-free was calculated based on 10% of randomly selected data excluding from the refinement. Structure validation was performed with PROCHECK (*22*). Residues in the loop region 169–177 and 231–238 in BH25 N43Y are not defined in the electron density map and are assumed to be disordered. The crystallographic statistics for data collection and refinement are summarized in Table S1.

Analysis of active mutants

We can rationalize improvements in activity in the optimized sequences based on the crystal structures. For BH25 we identified the variant N43Y as the most active point mutant - it was predicted to hydrogen bond with the oxyanion of Int1, and the crystal structure supports this (Figure 4B). Other more active mutants include (with structural

justification): W164Y (creates room for the W166 stacking on 4-nitrobenzaldehyde), G312M (better packing), Y129F (create a more hydrophobic pocket for the nitro group which does not form hydrogen bonds in water).

For BH32 we identified a number of variants with slightly higher activity than the wild-type design: S124A (intended to create a more hydrophobic pocket for the nitro group), S9H (remove hydrogen bonding residue from nitro-group pocket), S91V (Original intended hydrogen-bond may be too long, testing a replacement with hydrophobic packing). The most active point-mutant N14I was tested after examining the crystal structure - the backbone moves too far in the crystal structure for N14 to form the intended hydrogen bond, and the mutation instead forms a hydrophobic pack from the new backbone position. The S9G mutant creates a more hydrophobic pocket for the nitro group. The MD simulation correctly predicts many of the sidechain shifts observed in the crystal structure, but we could not identify simple local mutations to repair these deficiencies and get higher activity.

Appendix A. Composite transition state – a superposition of the transition state (from QM and MD) and int2

```
HETATM    1  C1  LG1 X   1      26.285  45.600  23.012  1.00  0.00           C
HETATM    2  O1  LG1 X   1      23.471  44.469  21.679  1.00  0.00           O
HETATM    3  C2  LG1 X   1      25.504  44.500  23.726  1.00  0.00           C
HETATM    4  O2  LG1 X   1      23.398  44.667  22.276  1.00  0.00           O
HETATM    5  C3  LG1 X   1      24.199  43.557  22.275  1.00  0.00           C
HETATM    6  O3  LG1 X   1      25.729  42.801  25.316  1.00  0.00           O
HETATM    7  C4  LG1 X   1      24.415  43.845  22.785  1.00  0.00           C
HETATM    8  O4  LG1 X   1      27.223  38.764  19.257  1.00  0.00           O
HETATM    9  C5  LG1 X   1      27.556  45.063  22.379  1.00  0.00           C
HETATM   10  O5  LG1 X   1      27.381  40.257  17.776  1.00  0.00           O
HETATM   11  C6  LG1 X   1      28.390  44.284  23.374  1.00  0.00           C
HETATM   12  O6  LG1 X   1      26.031  43.356  25.744  1.00  0.00           O
HETATM   13  C7  LG1 X   1      27.640  43.117  23.985  1.00  0.00           C
HETATM   14  O7  LG1 X   1      22.166  46.667  24.146  1.00  0.00           O
HETATM   15  C8  LG1 X   1      26.239  43.475  24.381  1.00  0.00           C
HETATM   16  C9  LG1 X   1      27.031  39.884  18.868  1.00  0.00           C
HETATM   17  C10 LG1 X   1      26.217  40.075  18.757  1.00  0.00           C
HETATM   18  C11 LG1 X   1      26.319  40.824  19.739  1.00  0.00           C
HETATM   19  C12 LG1 X   1      26.082  42.136  19.334  1.00  0.00           C
HETATM   20  C13 LG1 X   1      25.389  43.031  20.158  1.00  0.00           C
HETATM   21  C14 LG1 X   1      24.913  42.630  21.402  1.00  0.00           C
HETATM   22  C15 LG1 X   1      25.154  41.322  21.819  1.00  0.00           C
HETATM   23  C16 LG1 X   1      25.852  40.430  20.999  1.00  0.00           C
HETATM   24  C17 LG1 X   1      25.743  41.023  19.771  1.00  0.00           C
HETATM   25  C18 LG1 X   1      25.582  42.377  19.467  1.00  0.00           C
HETATM   26  C19 LG1 X   1      25.141  43.284  20.435  1.00  0.00           C
HETATM   27  C20 LG1 X   1      24.856  42.856  21.727  1.00  0.00           C
HETATM   28  C21 LG1 X   1      25.000  41.507  22.036  1.00  0.00           C
HETATM   29  C22 LG1 X   1      25.444  40.601  21.067  1.00  0.00           C
HETATM   30  C23 LG1 X   1      26.361  38.918  19.063  1.00  0.00           C
HETATM   31  C24 LG1 X   1      26.457  40.480  17.653  1.00  0.00           C
HETATM   32  H1  LG1 X   1      25.651  45.975  22.230  1.00  0.00           H
HETATM   33  H2  LG1 X   1      27.322  44.421  21.548  1.00  0.00           H
HETATM   34  H3  LG1 X   1      28.138  45.867  21.967  1.00  0.00           H
HETATM   35  H4  LG1 X   1      29.277  43.922  22.904  1.00  0.00           H
HETATM   36  H5  LG1 X   1      28.720  44.941  24.153  1.00  0.00           H
HETATM   37  H6  LG1 X   1      27.586  42.314  23.281  1.00  0.00           H
HETATM   38  H7  LG1 X   1      28.169  42.751  24.834  1.00  0.00           H
HETATM   39  H8  LG1 X   1      24.771  44.905  24.397  1.00  0.00           H
HETATM   40  H9  LG1 X   1      23.651  43.088  23.072  1.00  0.00           H
HETATM   41  H10 LG1 X   1      23.924  43.293  23.564  1.00  0.00           H
HETATM   42  H11 LG1 X   1      26.432  42.470  18.388  1.00  0.00           H
HETATM   43  H12 LG1 X   1      25.231  44.029  19.828  1.00  0.00           H
HETATM   44  H13 LG1 X   1      24.806  40.994  22.769  1.00  0.00           H
HETATM   45  H14 LG1 X   1      26.024  39.434  21.332  1.00  0.00           H
HETATM   46  H15 LG1 X   1      25.804  42.736  18.492  1.00  0.00           H
HETATM   47  H16 LG1 X   1      25.038  44.314  20.190  1.00  0.00           H
HETATM   48  H17 LG1 X   1      24.778  41.160  23.017  1.00  0.00           H
HETATM   49  H18 LG1 X   1      25.554  39.575  21.327  1.00  0.00           H
```

Figure S1. Product accumulation time course for BH32 and variants.

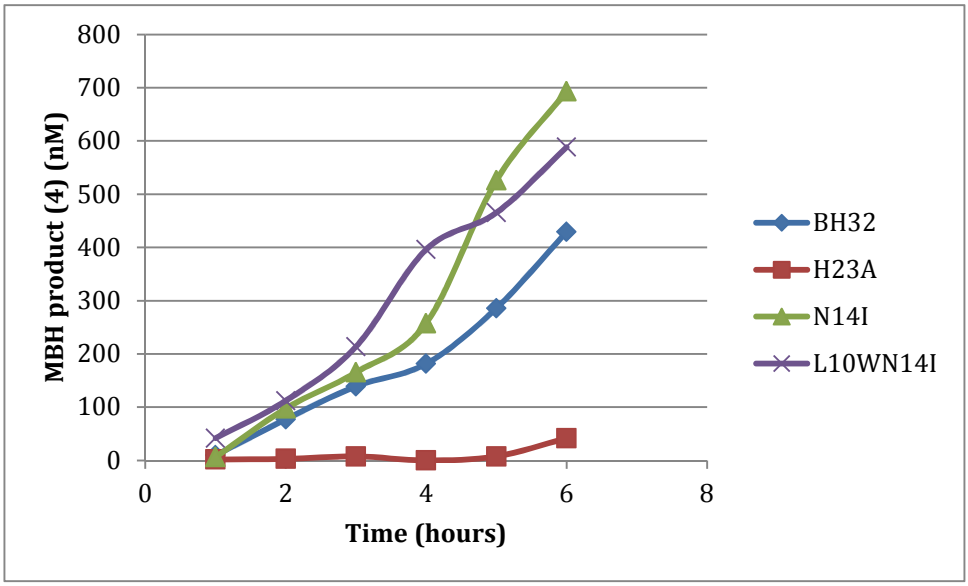


Figure S2. Overlay of representative MD snapshots (30-40 ns, in blue) and the wild-type designs (in gray). The enone substrate (in yellow) is docked in its designed orientation. The nucleophilic cysteine (C39) of BH25 (A) remains preorganized through a strong H-bond network between residues K285, D313 and R363. For BH32, catalytic base E46 is highly solvent exposed (B), and E46-water interactions strongly compete with E46-H23 dyad. Catalytic residues designed to donate H-bonds to the enolate intermediate in BH25 (C) and in BH32 (D) are engaged in alternative binding patterns generating a bottleneck for the later steps of the reaction.

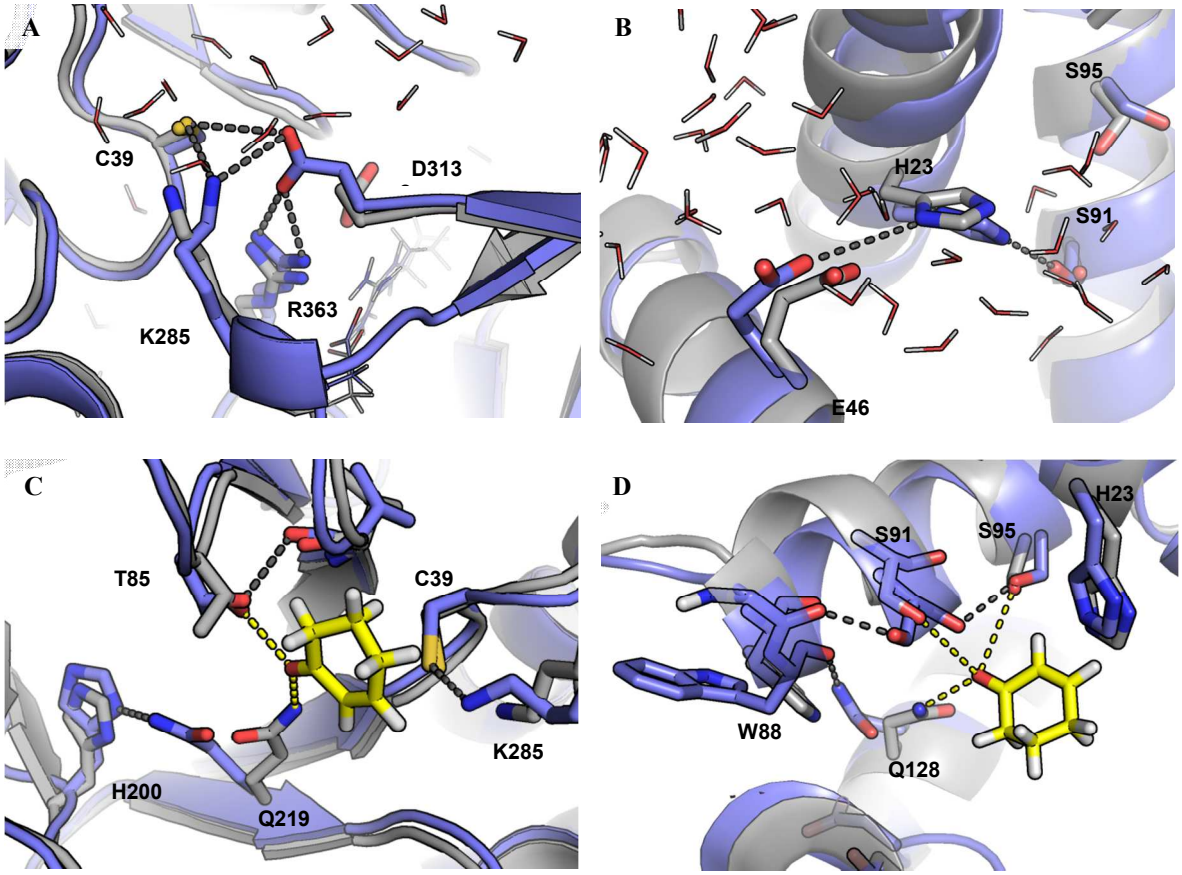


Figure S3. MD on the covalently bound alkoxide intermediate (Int2) for design BH25. (A) The design with docked intermediate; (B) an MD snapshot after 20 ns.

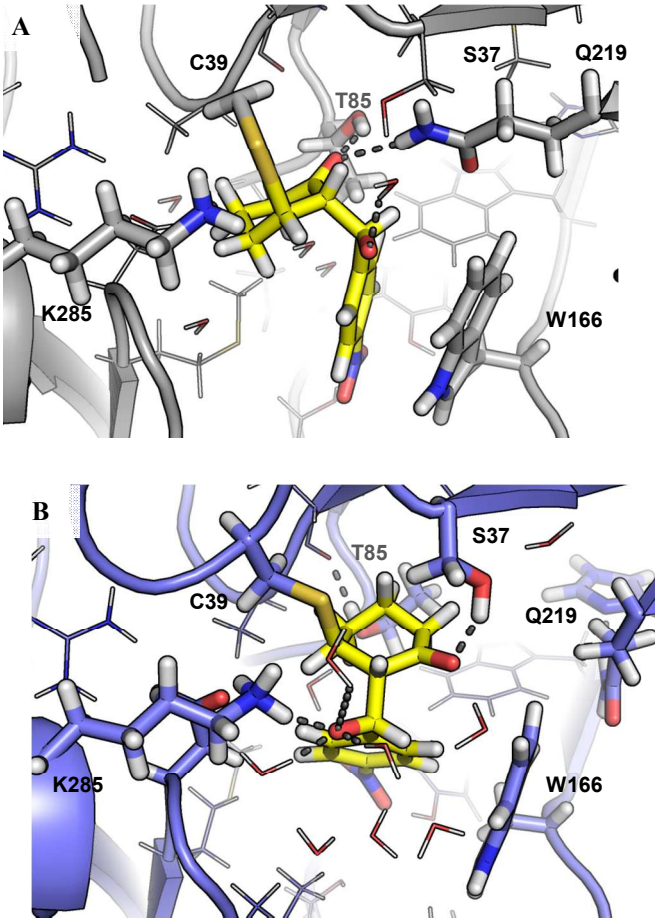


Figure S4. (A) Nucleophilic dyad for BH25. Overlay of an MD snapshot at 22 ns MD (in blue) and the wild-type BH25 design (in gray). The nucleophilic cysteine (C39) remains preorganized through a strong H-bond network between residues K285, D313 and R363. (B) The plot of distances $C39^{HG}$-$D313^{OD}$ versus $C39^{HG}$-$K285^{NZ}$ shows that C39 binds to D313 and K285 in a triangular fashion. The relative populations of strong/moderate H-bond configurations suggest a more tightly bound $C39^{HG}$-$D313^{OD}$ compared to $C39^{HG}$-$K285^{NZ}$.

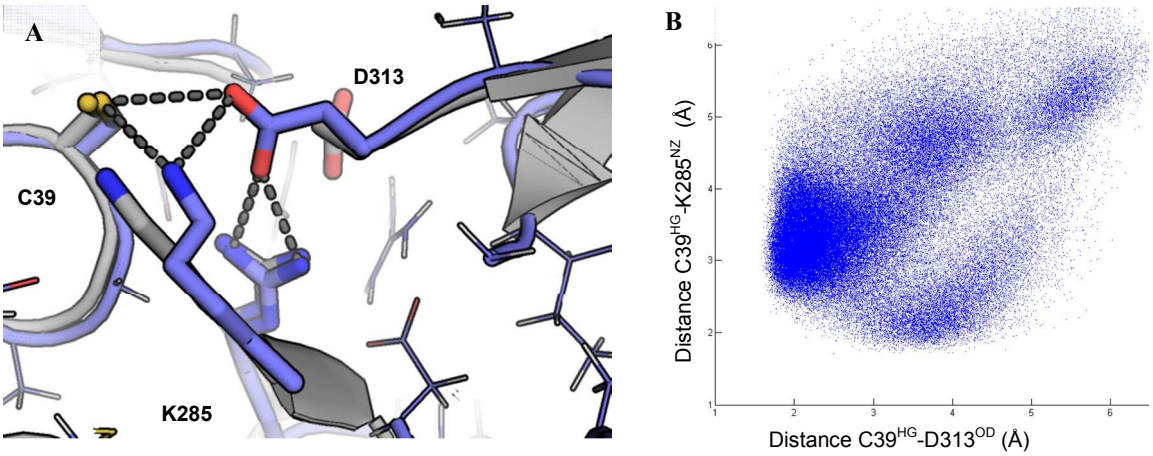


Figure S6. Electron density map 2fofc superimposed on the structure of BH32 with the crystal structure in orange and the design in pink. Residue H23 and the backing up D46 are shown.

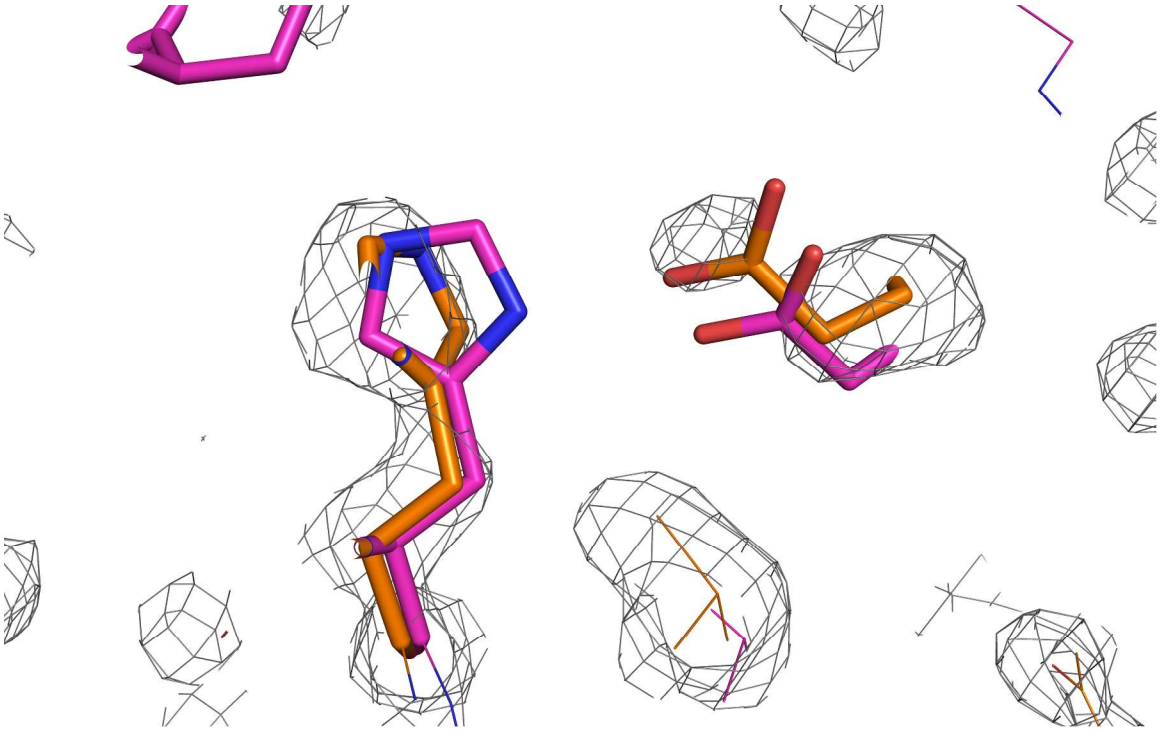
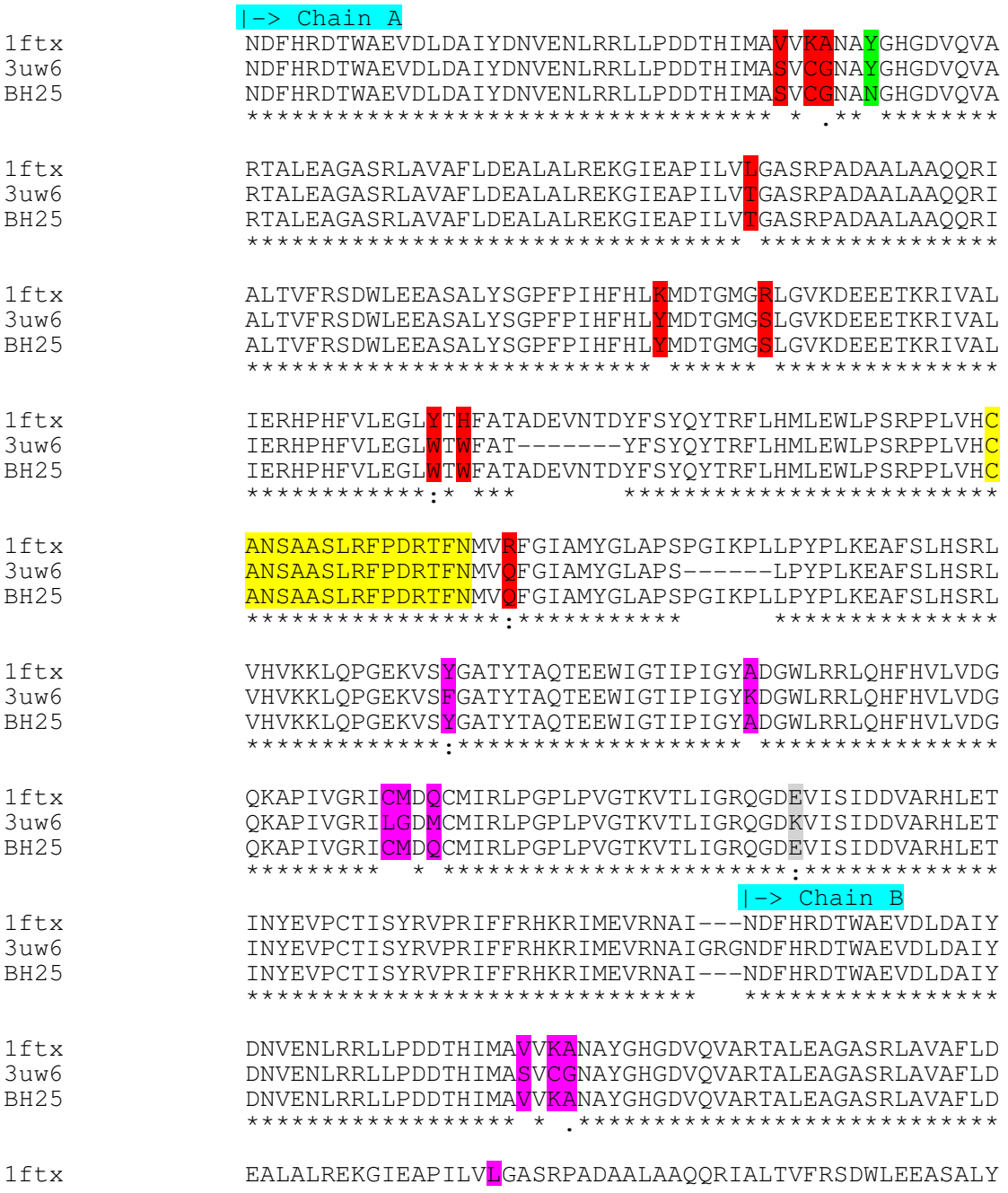


Figure S7. Alignment of BH25 and BH32 designs with X-ray structures and original scaffold, respectively. BH25 is a dimer and red denotes designed residues, which are part of the designed active site. Purple is the second active site, yellow is a large structural change compared to the design and green is the improved variant N42Y.

BH25



```
3uw6            EALALREKGIEAPILVTGASRPADAALAAQQRIALTVFRSDWLEEASALY
BH25            EALALREKGIEAPILVLGASRPADAALAAQQRIALTVFRSDWLEEASALY
                ********************************************************

1ftx            SGPFPIHFHLKMDTGMGRLGVKDEEETKRIVALIERHPHFVLEGLYTHFA
3uw6            SGPFPIHFHLYMDTGMGSLGVKDEEETKRIVALIERHPHFVLEGLWTWFA
BH25            SGPFPIHFHLKMDTGMGRLGVKDEEETKRIVALIERHPHFVLEGLYTHFA
                ********** ****** ***************************:* **

1ftx            TADEVNTDYFSYQYTRFLHMLEWLPSRPPLVHCANSAASLRFPDRTFNMV
3uw6            T-------YFSYQYTRFLHMLEWLPSRPPLVHCANSAASLRFPDRTFNMV
BH25            TADEVNTDYFSYQYTRFLHMLEWLPSRPPLVHCANSAASLRFPDRTFNMV
                *       ****************************************

1ftx            RFGIAMYGLAPSPGIKPLLPYPLKEAFSLHSRLVHVKKLQPGEKVSYGAT
3uw6            QFGIAMYGLAPS------LPYPLKEAFSLHSRLVHVKKLQPGEKVSFGAT
BH25            RFGIAMYGLAPSPGIKPLLPYPLKEAFSLHSRLVHVKKLQPGEKVSFGAT
                :***********      ****************************:***

1ftx            YTAQTEEWIGTIPIGYADGWLRRLQHFHVLVDGQKAPIVGRICMDQCMIR
3uw6            YTAQTEEWIGTIPIGYKDGWLRRLQHFHVLVDGQKAPIVGRILGDMCMIR
BH25            YTAQTEEWIGTIPIGYKDGWLRRLQHFHVLVDGQKAPIVGRILGDMCMIR
                **************** ************************  * *****

1ftx            LPGPLPVGTKVTLIGRQGDEVISIDDVARHLETINYEVPCTISYRVPRIF
3uw6            LPGPLPVGTKVTLIGRQGDKVISIDDVARHLETINYEVPCTISYRVPRIF
BH25            LPGPLPVGTKVTLIGRQGDEVISIDDVARHLETINYEVPCTISYRVPRIF
                ********************:*******************************

1ftx            FRHKRIMEVRNAI---
3uw6            FRHKRIMEVRNAIGRG
BH25            FRHKRIMEVRNAI---
                *************
```

BH32

```
1x42            MIRAVFFDFVGTLLSVEGEAKTHLKIMEEVLGDYPLNPKTLLDEYEKLTR
BH32            MIRAVFFDSLGTLNSVEGAAKSHLKIMEEVLGDYPLNPKTLLDEYEKLTR
                ******** :*** **** **:******************************

1x42            EAFSNYAGKPYRPIRDIEEEVMRKLAEKYGFKYPENFWEIHLRMHQRYGE
BH32            EAFSNYAGKPYRPLRDILEEVMRKLAEKYGFKYPENFWEISLRMSQRYGE
                *************:*** ********************** *** *****

1x42            LYPEVVEVLKSLKGKYHVGMITDSDTEYLMAHLDALGIKDLFDSITTSEE
BH32            LYPEVVEVLKSLKGKYHVGMITDSDTEQAMAFLDALGIKDLFDSITTSEE
                *************************  **.******************

1x42            AGFFKPHPRIFELALKKAGVKGEEAVYVGDNPVKDCGGSKNLGMTSILLD
BH32            AGFFKPHPRIFELALKKAGVKGEEAVYVGDNPVKDCGGSKNLGMTSILLD
                ****************************************************

1x42            RKGEKREFWDKCDFIVSDLREVIKIVDELNGQ
BH32            RKGEKREFWDKCDFIVSDLREVIKIVDELNGQ
                ********************************
```

Table S1. Summary of crystal parameters, data collection and refinement.

Protein Name	BH32	BH32	BH25 N43Y
Space group	P21	P21	P41212
Molecules per asymmetric unit	1	1	3
V_M (Å^3 Da^{-1})	1.59	2.36	2.82
Unit Cell (Å,^{o})	a=33.606,b=67.579,c=48.193,α=90,β=109.30γ=90.0^o	a=34.334,b=71.367 c=52.926,α=90, β= 104.83 ,γ=90	a=b=112.345,c=237.059,α=β=γ=90.0^o
Wavelength(Å)	0.979	0.979	0.979
Resolution (Å)	50-1.59(1.59-1.63)[a]	50-2.3(2.3-2.38)	50-2.3(2.3-2.41)
Temperature(K)	100	100	100
Unique reflections	27144	10847	85640
Mean I/σ(I)	22.8	16.1	23.1
Sigma Cutoff	0	0	0
Completeness	91.0(69.0)	99.1(100.0)	99.5(96.1)
Redundancy	2.0(1.5)	7.5(6.0)	10.0
R_{merge}[#]	0.040(0.128)	0.082(0.179)	0.070(0.058)
R_{cryst}[+]	0.200	0.238	0.215
R_{free}[*]	0.248	0.279	0.259
RMSD Bond lengths (Å) Bond angles (o)	 0.021 2.17	 0.008 1.20	 0.010 1.30
No. of residues No. of ions	230	230 1	1113

[#] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

[+] $R_{\text{cryst}} = \sum_{hkl} \left| |F_{\text{obs}}| - |F_{\text{calc}}| \right| / \sum_{hkl} |F_{\text{obs}}|$

[*] R_{free} is calculated in same manner as R_{cryst} except that it uses 10% of the reflection data omitted from refinement.

[a] Values in parentheses are for the highest resolution bin.

Table S2. Annotation of sidechain dihedral angles in BH25 design. Dihedrals are measured to chain C in 3UW6 X-ray structure. In general only small deviations are observed between different chains in the unit cell. In cases where there is a definite difference between the sidechain chi angels additional values are reported. Sidechain chis are reported in degrees and rounded of to whole numbers. [a] Dihedral measured to CD1 and [b] to OE1, respectively. [c] No density for the sidechain. [d] Dihedral is to CD1 in chain B.

3UW6 numbering	χ_1 Nbb-Ca-Cb-Cg		χ_2		χ_3		χ_4	
3UW6	3UW6	BH25						
S37	71	-179						
C39	50	72						
G40	–							
Y43	–							
T85	60	-58						
Y129	-180	-176	85[a]	17[a]				
S136	79	55						
W164	60	65	-67[a]	-72[a]				
W166	167	-170	-77[a]	-105[a]				
Q219	-75	-178	-178	-179	53[b]	3		
F265	N.D.[c]							
K285	-73	-60	-149	179	-83	-178	-72	180
L311	-64	-61	-37[a] -117[d]	-42				
G312	–							

Q314	63	64	170	−170	98	78		

Table S3. Annotation of sidechain dihedrals in the BH32 design. Dihedrals are measured relative to the X-ray structure 3U26.

3U26 numbering	χ_1 Nbb-Ca-Cb-Cg		χ_2		χ_3		χ_4	
3U26	3U26	BH32	3U26	BH32	3U26	BH32	3U26	BH32
S9	65	179						
L10	−53	−177						
L14	63	−177	133	23				
A19	–	–						
H23	−85	−72	68	179				
E46	−70	−75	−58	−67	−55	−51		
L64	−89	−68	−180	−179				
L68	−60	−62	173	175				
S91	−58	−172						
S95	−62	−66						
Q128	−72	−145	−179	−107	−62	55		
A129	–	–						
F132	176	−179	82	71				

REFERENCES

1. Kunkel, T. A. (1985) Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection, *Proc. Natl. Acad. Sci. U. S. A. 82*, 488-492.
2. Luo, S. Z., Wang, P. G., and Cheng, J. P. (2004) Remarkable rate acceleration of imidazole-promoted Baylis-Hillman reaction involving cyclic enones in basic water solution, *J. Org. Chem. 69*, 555-558.
3. Kataoka, T., Iwama, T., Tsujiyama, S., Iwamura, T., and Watanabe, S. (1998) The chalcogeno-Baylis-Hillman reaction: A new preparation of allylic alcohols from aldehydes and electron-deficient alkenes, *Tetrahedron 54*, 11813-11824.
4. Case, D. A., Darden, T. A., Cheatham, I. T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kocalenko, A., and Kollman, P. A. (2010) AMBER 11, *University of California, San Francisco*.
5. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water, *Journal of Chemical Physics 79*, 926-935.
6. Wang, J. M., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *Journal of Computational Chemistry 21*, 1049-1074.
7. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *Journal of Computational Chemistry 25*, 1157-1174.
8. Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model, *Journal of Physical Chemistry 97*, 10269-10280.
9. Besler, B. H., Merz, K. M., and Kollman, P. A. (1990) Atomic Charges Derived from Semiempirical Methods, *Journal of Computational Chemistry 11*, 431-439.
10. Singh, U. C., and Kollman, P. A. (1984) An Approach to Computing Electrostatic Charges for Molecules, *Journal of Computational Chemistry 5*, 129-145.
11. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T. b., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O.,

Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. (2004) Gaussian 03, Revision C.02, *Gaussian, Inc., Wallingford CT*.

12. Duke, R. E., and Pedersen, L. G. (2003) PMEMD, *University of North Carolina, Chapel Hill*.
13. Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, *Journal of Chemical Physics 98*, 10089-10092.
14. Kiss, G., Rothlisberger, D., Baker, D., and Houk, K. N. (2010) Evaluation and ranking of enzyme designs, *Protein Sci. 19*, 1760-1773.
15. Shaw, D. E. (2009) *Proceedings of the ACM/IEEE Conference on Supercomputing (SC09), ACM Press, New York*.
16. Sheldrick, G. M. (2008) A short history of SHELX, *Acta Crystallographica Section A 64*, 112-122.
17. Terwilliger, T. C. (2003) Automated main-chain model building by template matching and iterative fragment extension, *Acta Crystallographica Section D-Biological Crystallography 59*, 38-44.
18. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458-463.
19. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallographica Section D-Biological Crystallography 53*, 240-255.
20. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallographica Section D-Biological Crystallography 54*, 905-921.
21. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallographica Section D-Biological Crystallography 60*, 2126-2132.
22. Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by C alpha geometry: phi,psi and C beta deviation, *Proteins-Structure Function and Genetics 50*, 437-450.