Structure of formylglycine-generating enzyme in complex with copper and substrate reveals a pocket for binding and activation of molecular oxygen

Dzmitry A. Miarzlou[1], Florian Leisinger[1] Daniel Joss[1], Daniel Häussinger[1], and Florian P. Seebeck[1]

[1] Department of Chemistry, University of Basel, Mattenstrasse 24a, Basel 4002, Switzerland

SUPPORTING INFORMATION

SUPLEMENTARY FIGURES

Table S1 – Oligonucleotides used for site directed mutagenesis of *tc*FGE variants.

W228F_S	5'-TGGCGGGTAACGTTTTTGAATGGT-3'
W228F_AS	5'-ACCATTCAAAAACGTTACCCGCCA-3'
S266A_S	5'-CAAGGGTGGCGCCTTCCTGT-3'
S266A_AS	5'-ACAGGAAGGCGCCACCCTTG-3'

Sequence of *tc*FGE

GHHHHHHAENLYFQGHMPSFDFDIPRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLSPFLIDRYAVSNRQFA
AFVKATGYVTDAERYGWSFVFHAHVAPGTPVMDAVVPEAPWWVAVPGAYWKAPEGPGSSITDRPNHPVVHVSWNDAVAYATWA
GKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTFPVHDTGEDGYTGTAPVNAFAPNGYGLYNVAGNVWEWCAD
WWSADWHATESPATRIDPRGPETGTARVTKGGSFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

Sequence of *tc*FGE$_{W228F}$

GHHHHHHAENLYFQGHMPSFDFDIPRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLSPFLIDRYAVSNRQFA
AFVKATGYVTDAERYGWSFVFHAHVAPGTPVMDAVVPEAPWWVAVPGAYWKAPEGPGSSITDRPNHPVVHVSWNDAVAYATWA
GKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTFPVHDTGEDGYTGTAPVNAFAPNGYGLYNVAGNVFEWCAD
WWSADWHATESPATRIDPRGPETGTARVTKGGSFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

Sequence of *tc*FGE$_{S266A}$

GHHHHHHAENLYFQGHMPSFDFDIPRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLSPFLIDRYAVSNRQFA
AFVKATGYVTDAERYGWSFVFHAHVAPGTPVMDAVVPEAPWWVAVPGAYWKAPEGPGSSITDRPNHPVVHVSWNDAVAYATWA
GKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTFPVHDTGEDGYTGTAPVNAFAPNGYGLYNVAGNVWEWCAD
WWSADWHATESPATRIDPRGPETGTARVTKGGAFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

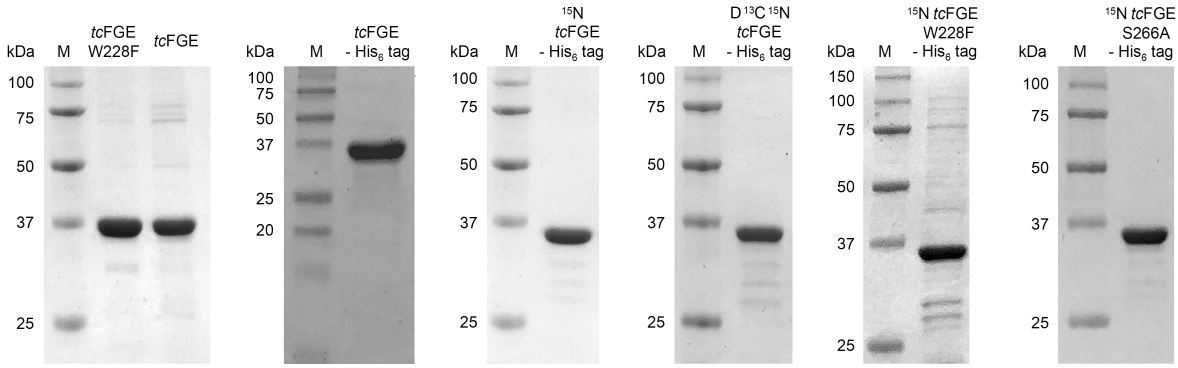


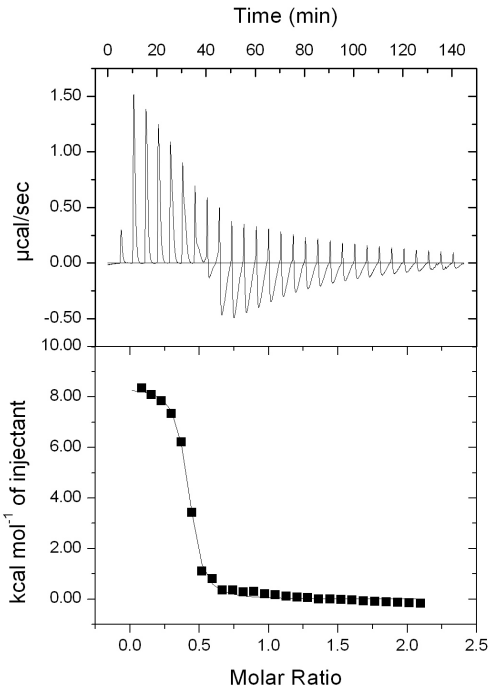
Figure S1 – Characterization of *tc*FGE variants by SDS PAGE.

Table S2 – HR-MS-ESI characterization of *tc*FGE variants and isotopologs.[a]

Enzyme	Sample description	M_w, [Da]			Labeling efficiency, %
		cal., non-labeled	cal., labeled	observed	
*tc*FGE		35216.0	-	35215.1	-
$tc\text{FGE}_{W228F}$		35177.0	-	35175.7	-
*tc*FGE	without His_6 tag	33470.2	-	33468.9	-
*tc*FGE	without His_6 tag	33470.2	33890.9 [b]	33889.8	99.7
$tc\text{FGE}_{W228F}$	without His_6 tag	33431.1	33850.9 [b]	33847.3	99.1
$tc\text{FGE}_{S266A}$	without His_6 tag	33454.2	33874.9 [b]	33871.9	99.3
*tc*FGE	without His_6 tag	33470.2	37035.2 [c]	37012.6	99.4

[a] The labeling efficiency was calculated as a ratio between the obtained increment $\Delta M_{w\,obs.}$ ($M_{w\,obs.}$ (labeled protein) - $M_{w\,cal.}$ (non-labeled protein)) and the expected increment $\Delta M_{w\,cal.}$ ($M_{w\,cal.}$ (labeled protein) - $M_{w\,cal.}$ (non-labeled protein)).

[b] Molecular weight calculated with an assumption, that 99% of N in protein replaced with ^{15}N.
[c] Molecular weight calculated with an assumption, that 99% of N and C in protein replaced with ^{15}N and ^{13}C respectively, and 99% none water exchangeable ^{1}H with ^{2}H.



Model: OneSite
N = 0.708 ± 0.002 Sites
Chi^2/DoF = 6044
K_d = 40 ± 10 nM
ΔH = 5240 ± 30 cal/mol
ΔS = 51.2 cal/mol/deg

Figure S2 - Isothermal titration calorimetry (ITC). A solution containing 100 µM of $tc\text{FGE}_{W228F}$ in 50 mM phosphate buffer pH 8.0, 100 mM NaCl and 600 µM DTT was transferred into the sample cell (cell volume 1.42 mL) of an isothermal titration calorimeter (VP200-ITC system, MicroCal LLC). The 1 mM solution of FGE-27, dissolved in the same buffer, was added to the protein solution by syringe in 1 x 2 µL injection followed by 10 µL injections every five minutes. All ITC measurements were conducted at 25°C. Data were plotted as the power needed to maintain the reference and sample cell at the same temperature against time and as kcal/mol of injectant against the molar ratio of ligand and protein. The program Origin7 (OriginLab Corporation) was used to analyze the data.

Table S3 - Data processing and refinement statistics

Structure	**Reaction complex with substrate and Cu^+**
Wavelength (Å)	1.0000
Space group	P 2_1 2_1 2_1
Cell dimensions	
a, b, c (Å)	58.4, 71.9, 76.7
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)[a]	46.5 – 1.04 (1.04 – 1.06)
R_{merge} [+] (%)[a]	0.054 (0.624)
R_{meas} [++] (%)[a]	0.058 (0.695)
R_{pim} [+++] (%)[a]	0.023 (0.301)
CC ½[a]	0.999 (0.909)
<I/σ(I)>[a]	19.5 (3.5)
Multiplicity[a]	12.1 (10.1)
Completeness (%)[a]	100 (100)
Refinement statistics	
Resolution (Å)	1.04
No. of reflections	155126
R-work * (%)	16.5
R-free ** (%)	17.5
No. atoms	3825
Protein	3401
Ligand/ion	16
Water	408
B-factors (Å^2)	20.2
Protein	19.2
Ligand/ion	30.1
Water	28.0
r.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.59
Ramachandran statistics *** (%)	
Favored	96.68
Allowed	3.32
Outliers	0.00

Numbers in parenthesis belong to outer shell.

[+]$R_{merge} = \Sigma_{hkl} \Sigma_i |(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$ where $I_i(hkl)$ is the observed intensity for a reflection and $\langle I(hkl)\rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

[++]$R_{meas} = \Sigma_{hkl}[N/(N-1)]^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection, $\langle I(hkl)\rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections and N is the number of observations of intensity $\langle I(hkl)\rangle$.

*R-work = ghkl||Fobs| - |Fcalc|| / ghkl|Fobs|

**R-free is the R value calculated for 5 % of the data set that was not included into the refinement.

***Molprobity

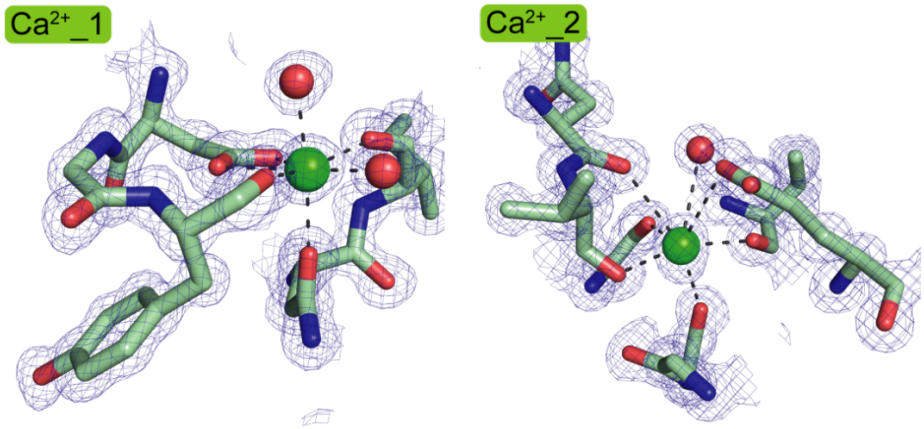


Figure S3 - Calcium bound to conserved calcium-binding site. 2m|F_o|-D|F_c| omit map contoured at σ-level = 1.0.

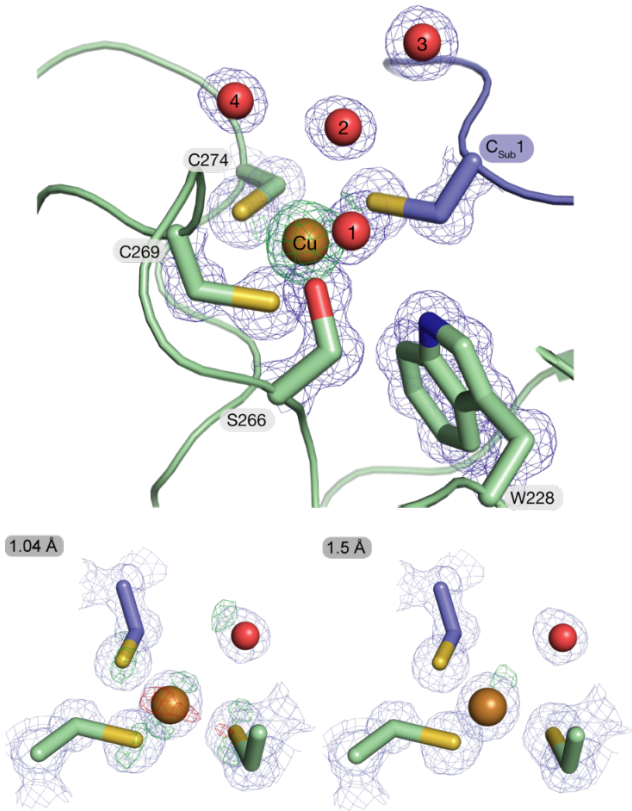


Figure S4 – **Top:** Detailed view into the active site of *tc*FGE_Cu_S with key active site residues displayed in stick representation with electron density (2m|F_o|-D|F_c| omit map contoured at σ-level = 1.0) in mesh representation (blue). **Bottom:** Additional electron density was found in the active site center as displayed in mesh representation (green, (m|F_o|-D|F_c| map contoured at σ-level = 3.0) which could have been modeled by a Cu (I) ion. After several rounds of refinement two positive and negative electron density peaks of same size came up in the m|F_o|-D|F_c| map. Refinement against the same data scaled to lower resolution of 1.5 Å produced a good fit.

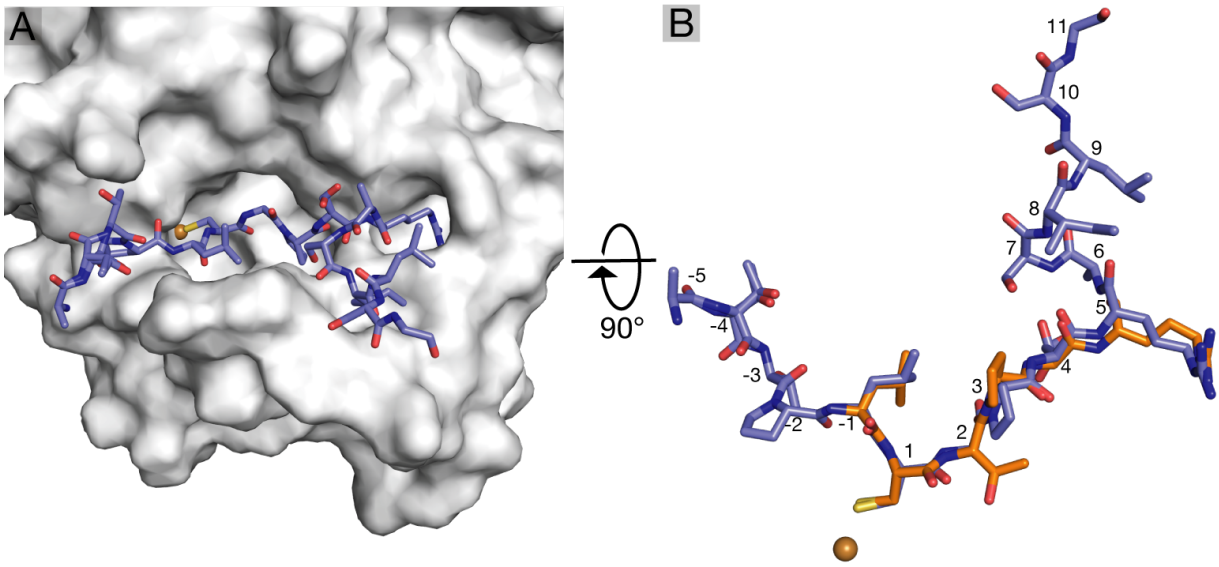


Figure S5 – **A:** Structure of *tc*FGE_Cu (white surface) in complex with substrate (violet, Abz-ATTPLCGPSRASILSGR-NH_2). **B:** Superposition of the substrate that was crystallized in complex with *hs*FGE (orange, LCTPSRA, PDB: 2AIK) shows that the cores of both substrates adopt similar conformations.

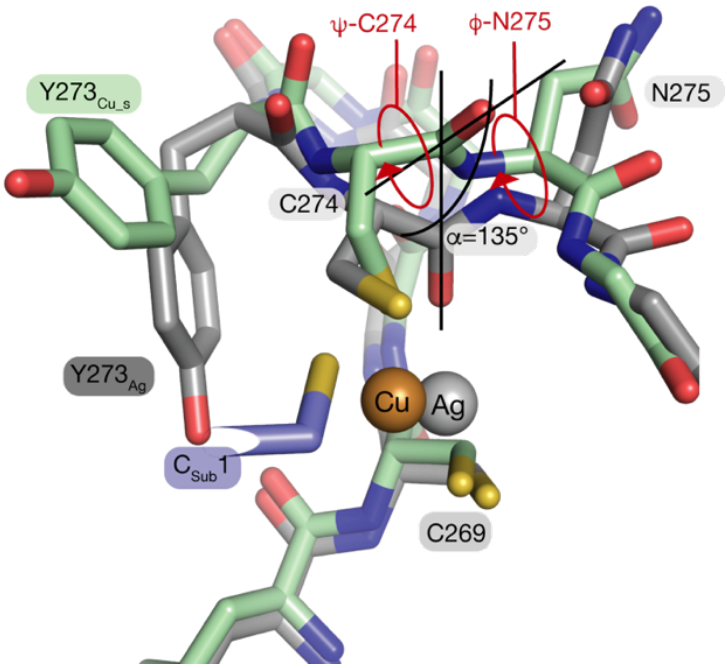


Figure S6 – Rotation of the backbone amide of Cys274 in response to substrate binding. Comparison of the structures from *tc*FGE_Ag and *tc*FGE_Cu_S shows that the plane of this amid bond rotates by 135°. The same rotation was observed in the structure of *tc*FGE_Cd, suggesting that this change is induced by the expansion of the coordination sphere of the metal.

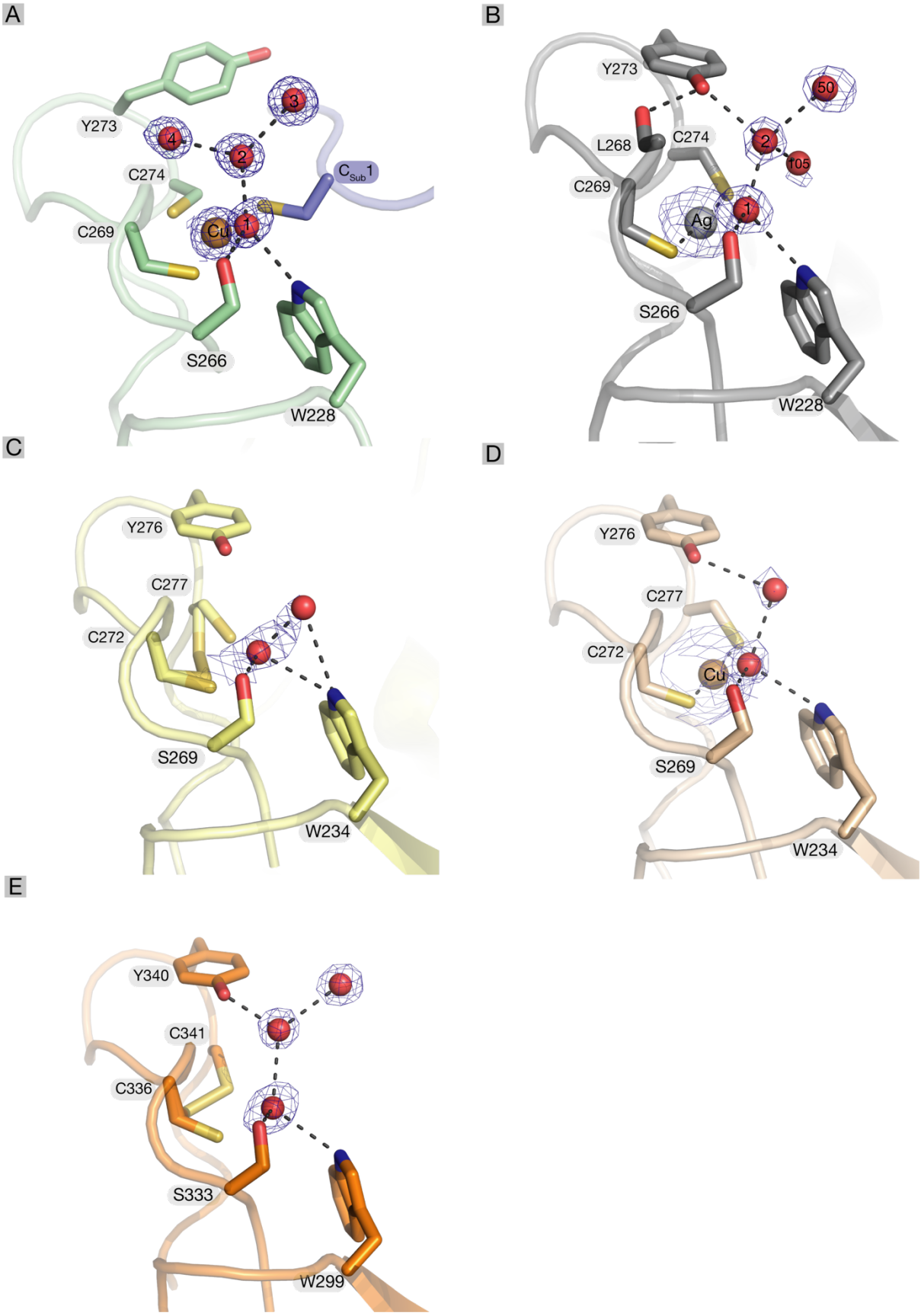


Figure S7 – FGE active sites with crystallographic waters: **A**) *tc*FGE_Cu_S (PDB: XXXX); **B**) *tc*FGE_Ag (PDB: 5NLX); **C**) *sc*FGE (PDB: 2Q17); **D**) *sc*FGE_Cu (PDB: 6MUJ); **E**) *hs*FGE (PDB: 1Y1E). 2m|F_o|-D|F_c| omit maps of the water and metal was contoured at σ-level = 1.1.

Table S4 – The geometry of interactions of conserved active site W and H2O_1.

Structure		Distance	Angle
		O(H$_2$0) – Nε1(W), Å	**O(H$_2$0)-Hε1(W)- Nε1(W)**
S. coelincolor	wt (PDB: 2Q17)	3.1	128.0^o
	wt + Cu (PDB: 6MUJ)	3.1	144.1^o
H. sapiens	wt (PDB: 1Y1E)	3.2	127,6^o
T. curvata	wt + Ag(I) (PDB: 5NLX)	2.9	147.5^o
	wt + Cu(I) + S (PDB: XXXX)	2.9	156.7^o

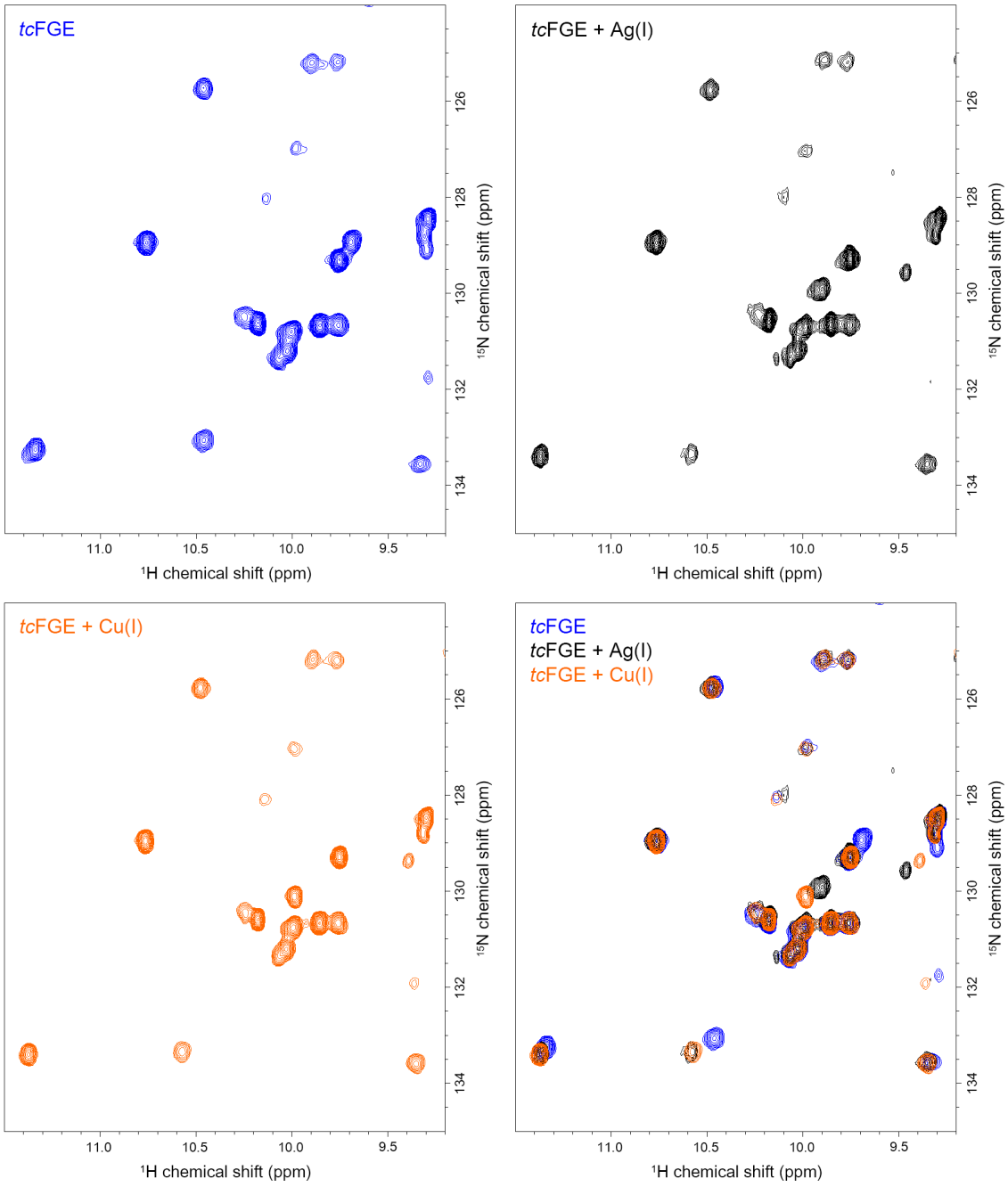


Figure S8 - ^{1}H-^{15}N TROSY HSQC spectra of uniformly ^{15}N-labelled *tc*FGE in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, pH 8.0, 25°C. The binding either Ag(I) or Cu(I) in the active site *tc*FGE causes pronounced shifts in the downfield area of the amide region. **Top left:** HSQC spectrum of apo-form *tc*FGE 300 µM. **Top right:** HSQC spectrum of *tc*FGE 300 µM in the presence of 450 µM $AgNO_3$. **Bottom left:** HSQC spectrum of *tc*FGE 300 µM in the presence of 450 µM $CuSO_4$. **Bottom right:** Overlay of the spectra.

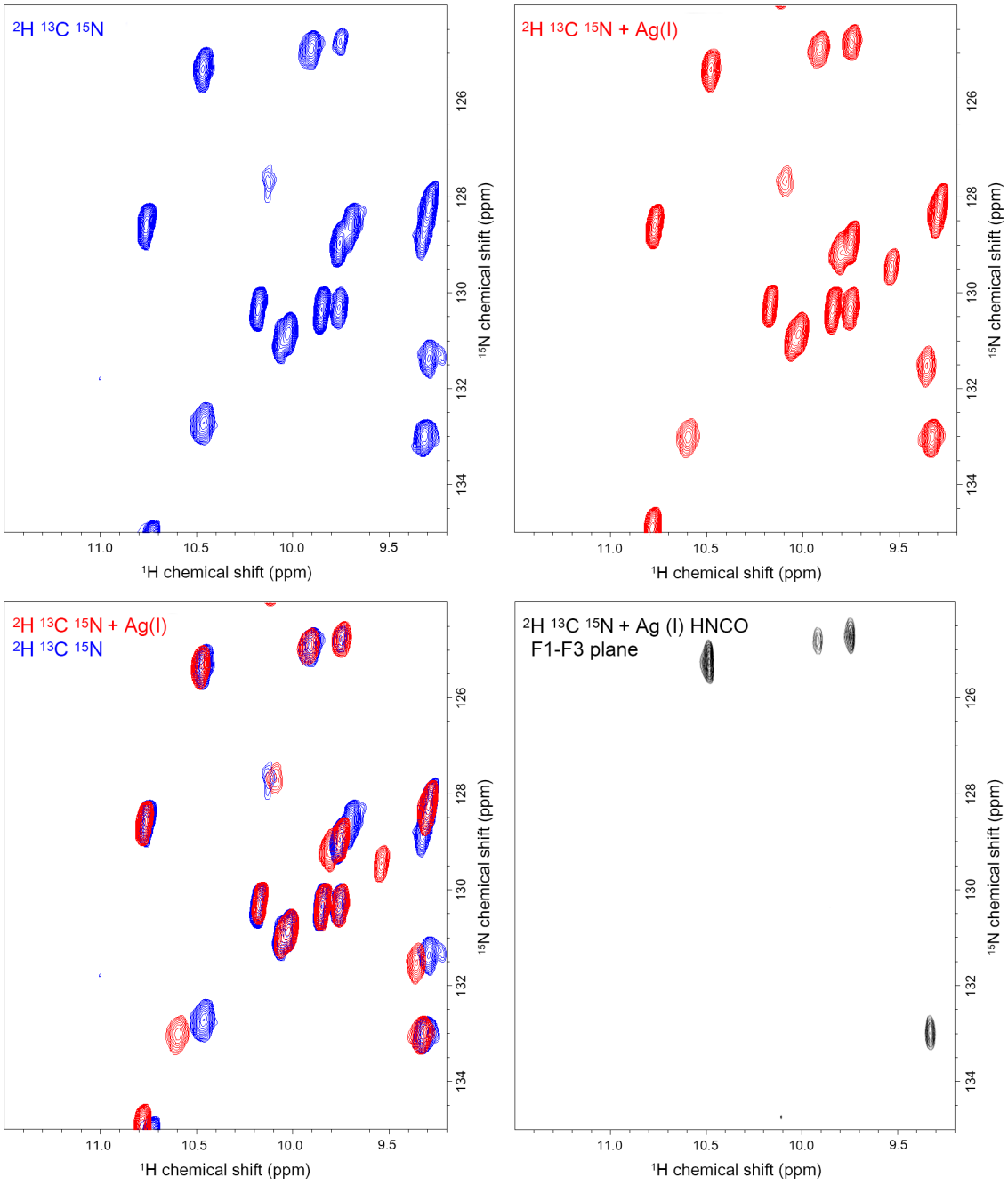


Figure S9 – NMR spectra of ^{2}H ^{13}C ^{15}N-labelled *tc*FGE in 20 mM phosphate buffer, 75 mM NaCl, 2 mM DTT, pH 8.0, 25°C. The downfield area of the amide region is presented. The comparison of ^{1}H-^{15}N TROSY HSQC and F1-F3 (^{1}H-^{15}N) plane from TROSY based HNCO experiment for ^{2}H ^{13}C ^{15}N-labelled *tc*FGE in the presence of 1.5 eq. Ag (I) demonstrated, that significant part of the cross-peaks belongs to ^{1}H-^{15}N correlations in side-chains, most likely to $^{1}H\varepsilon1$-$^{15}N\varepsilon1$ of tryptophan. [2-3] **Top left:** ^{1}H-^{15}N TROSY HSQC spectrum of apo-form ^{2}H ^{13}C ^{15}N-labelled *tc*FGE 860 µM. **Top right:** ^{1}H-^{15}N TROSY HSQC spectrum of ^{2}H ^{13}C ^{15}N-labelled *tc*FGE 860 µM in the presence of 1.29 mM $AgNO_3$. **Bottom left:** The overlay of the spectra of apo- and Ag (I) preloaded ^{2}H ^{13}C ^{15}N-labelled *tc*FGE. **Bottom right:** F1-F3 plane extracted from the HNCO spectrum of ^{2}H ^{13}C ^{15}N-labelled *tc*FGE 860 µM in the presence of 1.29 mM $AgNO_3$.

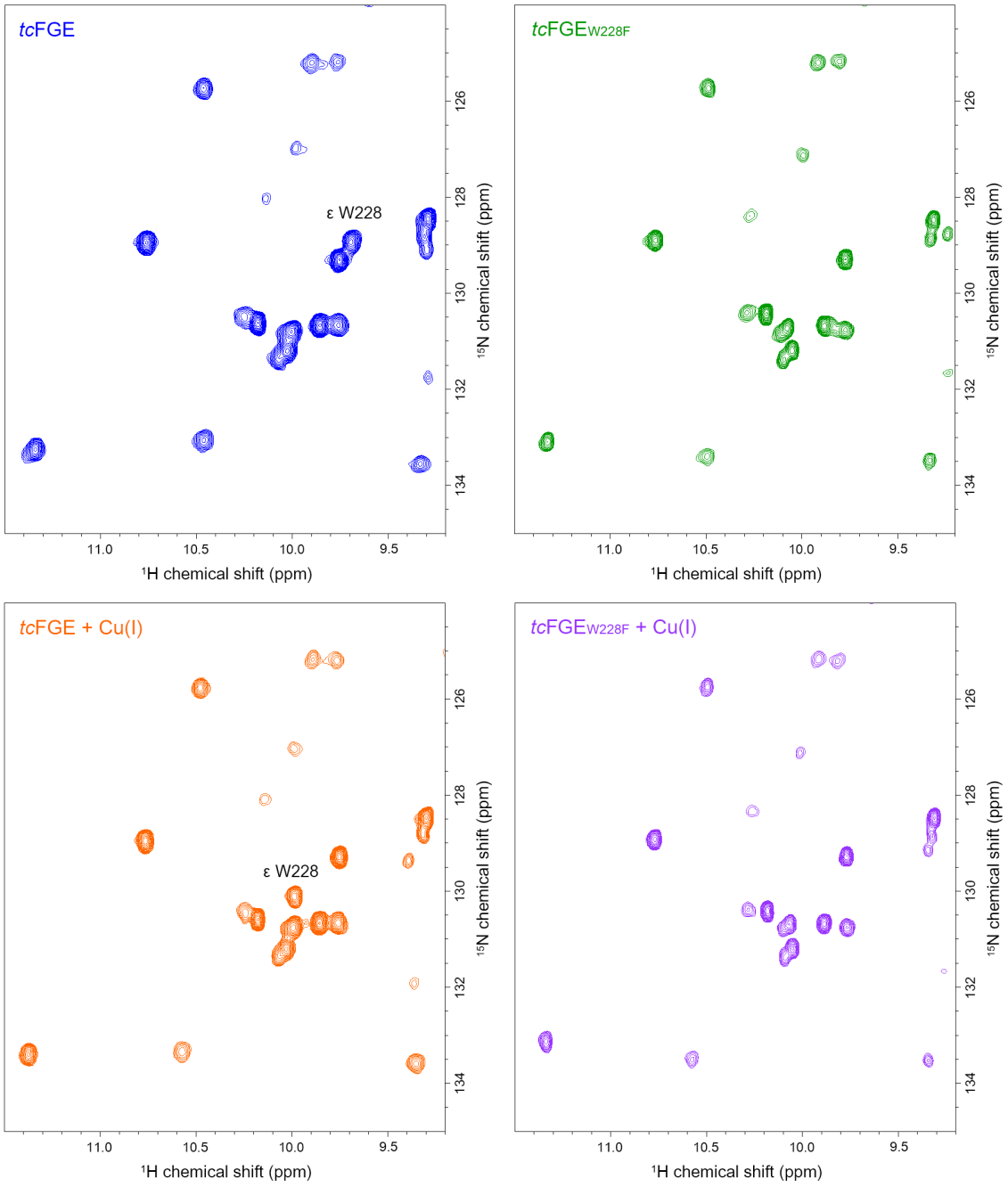


Figure S10 – ^{1}H-^{15}N TROSY HSQC spectra of uniformly ^{15}N-labelled *tc*FGE variants in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, pH 8.0, 25°C. The downfield area of the amide region is presented. The mutation W228F allowed to assign the cross peak, which corresponds to ^{1}Hε1-^{15}Nε1 of W228 side chain. **Top left:** HSQC spectrum of apo-form ^{15}N-labelled *tc*FGE 300 µM. **Top right:** HSQC spectrum of apo-form ^{15}N-labelled *tc*FGE$_{W228F}$ 300 µM. **Bottom left:** HSQC spectrum of ^{15}N-labelled *tc*FGE 300 µM in the presence of 450 µM $CuSO_4$. **Bottom right:** HSQC spectrum of 15N-labelled *tc*FGE$_{W228F}$ 300 µM in the presence of 450 µM $CuSO_4$.

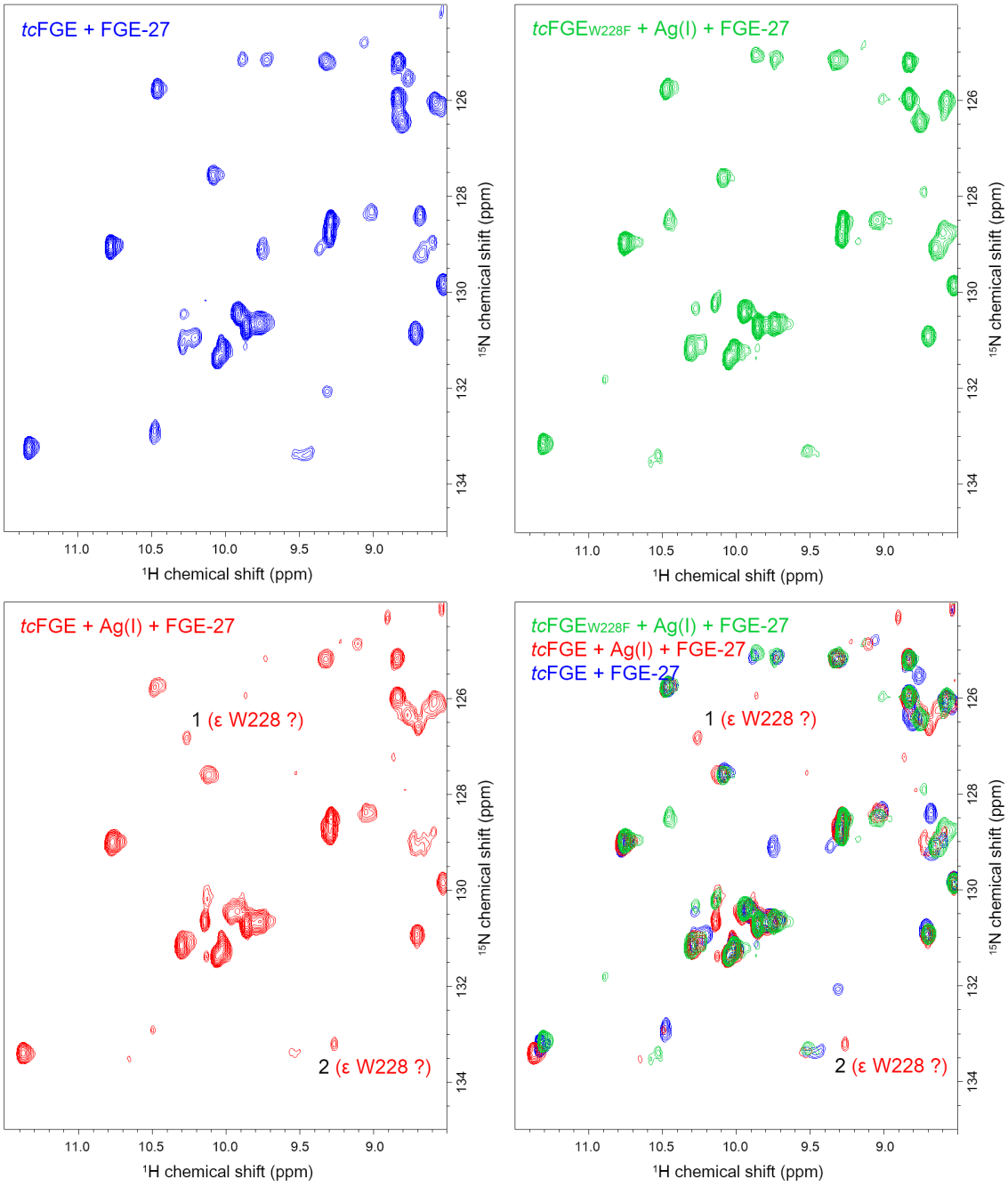


Figure S11 – ^{1}H-^{15}N TROSY HSQC spectra of uniformly ^{15}N-labelled *tc*FGE in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, 2 mM Glc, 2 U GO, 200 U CAT, pH 8.1, 25°C. The downfield area of the amide region is presented. Comparing spectra of three systems *tc*FGE : FGE-27, *tc*FGE$_{W228F}$: Ag(I) : FGE-27 and *tc*FGE : Ag(I) : FGE-27 provides two new cross-peaks 1 (10.26/126.81) and 2 (9.26/133.20), which only appear upon substrate binding to *tc*FGE : Ag (I) complex. These new signals are assigned as possible candidates for the cross-peak ^{1}Hε1-^{15}Nε1 W228, which shifts upon substrate binding. **Tom left:** HSQC spectrum of apo-form *tc*FGE 300 µM in the presence of 1 mM FGE-27. **Top right:** HSQC spectrum of *tc*FGE$_{W228F}$ 300 µM in the presence of 450 µM $AgNO_3$ and 1 mM FGE-27. **Bottom left:** HSQC spectrum of *tc*FGE 300 µM in the presence of 450 µM $AgNO_3$ and 1 mM FGE-27. **Bottom right:** Overlay of the spectra.

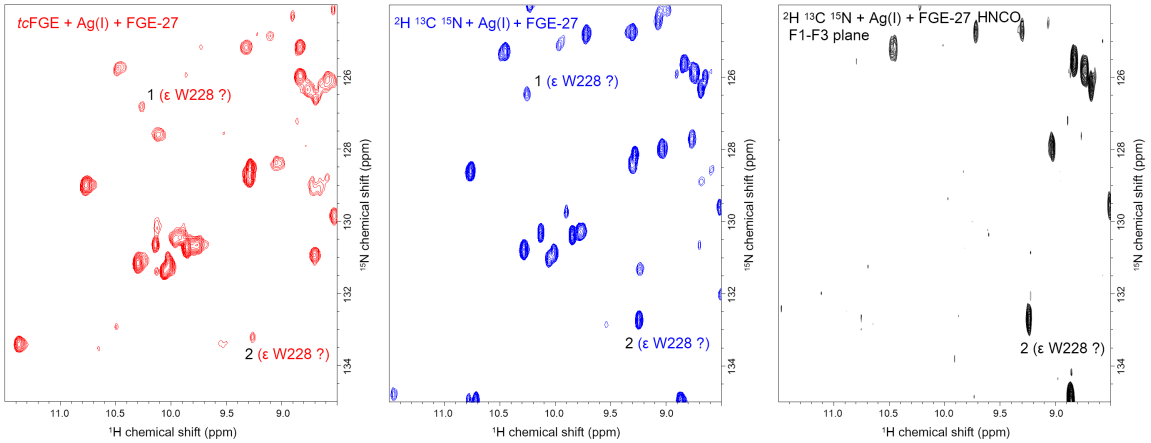


Figure S12 – NMR spectra of *tc*FGE in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, 2 mM glucose, 2 U GO, 200 U CAT, pH 8.1, 25°C. The downfield area of the amide region is presented. The ^{1}H-^{15}N TROSY-HSQC experiment of triple-labelled ^{2}H ^{15}N ^{13}C *tc*FGE in complex with Ag(I) and FGE-27 was compared with the ^{1}H-^{15}N TROSY-HSQC spectrum of ^{15}N labelled *tc*FGE containing Ag(I) and FGE-27. These spectra confirm that both cross peaks 1 and 2 are also visible in the ^{1}H-^{15}N TROSY-HSQC experiment of triple-labelled *tc*FGE and none of them appeared just due to signal to noise issues. The subsequent comparison of ^{1}H-^{15}N TROSY HSQC and F1-F3 plane from the HNCO experiment for ^{2}H ^{13}C ^{15}N labelled *tc*FGE in the presence of Ag(I) and FGE-27 demonstrated, that that cross-peak **2** belongs to a ^{1}H-^{15}N correlation either in the backbone or Asn/Gln side chain [3-4], whereas peak **1** is assigned as ^{1}Hε1-^{15}Nε1 correlation of W228. **Left:** ^{1}H-^{15}N TROSY HSQC spectrum of ^{15}N labelled *tc*FGE$_{W228F}$ 300 µM in the presence of 450 µM $AgNO_3$ and 1 mM of FGE-27. **Middle:** ^{1}H-^{15}N TROSY HSQC spectrum of ^{2}H ^{13}C ^{15}N labelled *tc*FGE 500 µM in the presence of 750 µM $AgNO_3$ and 1.5 mM of FGE-27. **Right:** F1-F3 plane extracted from TROSY based HNCO experiment for ^{2}H ^{13}C ^{15}N labelled *tc*FGE 500 µM in the presence of 750 µM $AgNO_3$ and 1.5 mM of FGE-27.

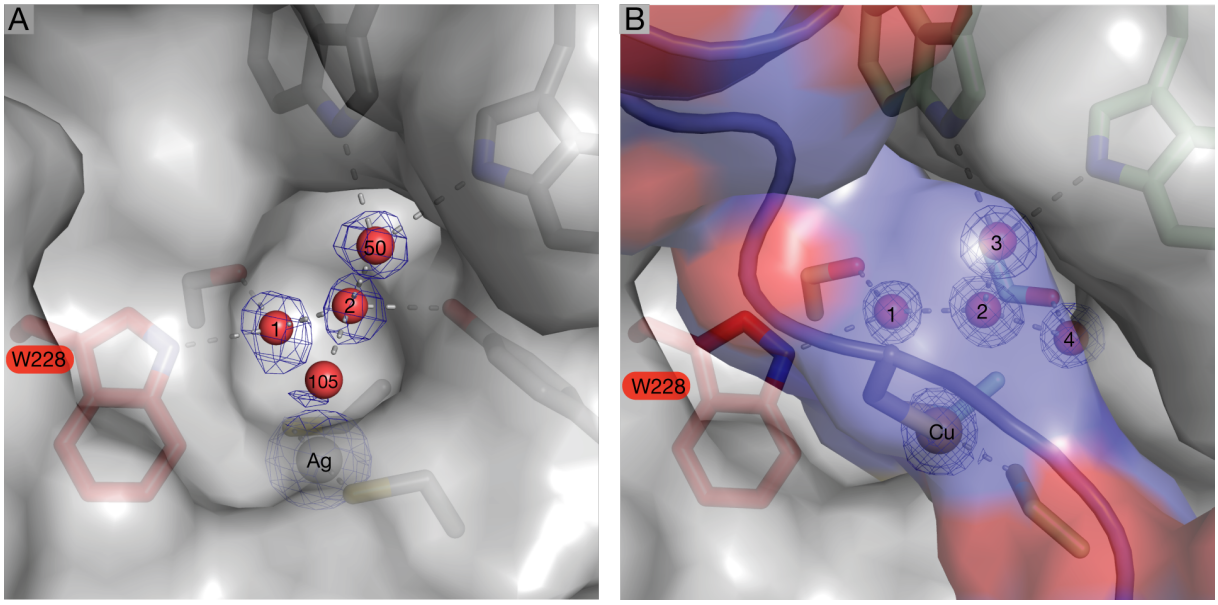


Figure S13 – The views into the active site of *tc*FGE_Ag (**A**) and *tc*FGE_Cu_S (**B**, substrate: violet) show that the water-exposed indol ring of Trp228 in buried upon substrate binding.

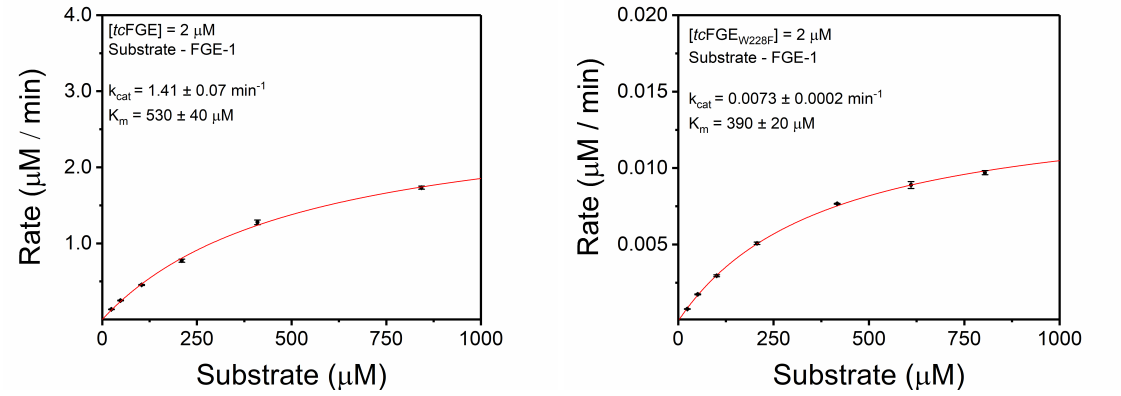


Figure S14 – Michaelis Menten kinetics of *tc*FGE- and *tc*FGE$_{W228}$-catalyzed turnover of FGE-1 (Abz-SALCSPTRA-NH_2). The enzymes were assayed in reactions containing FGE-1 (25-1000 µM), 2 µM $CuSO_4$, 5 mM DTT, 50 mM NaCl and 50 mM Tris Buffer at pH 8.0, 25°C. Reaction aliquots were quenched by addition of 1 volume equivalent of 1.5 % TFA in 4 M urea. Reaction product were quantified by RP-HPLC (column Gemini-NX, 5um C18 250 x 4.6 mm, Phenomenex) [5]. Initial rates (v) from three independent measurements were fitted with the function v = [S] x $k_{cat}/(K_M + [S])$ to obtain the Michaelis-Menten parameters k_{cat} and K_M.

SUPLEMENTARY METHODS

Production of *tc*FGE variants for kinetic and ITC measurements. The nucleotide sequences encoding *tc*FGE variants were cloned into pET19b vector under the control *lac* promoter. The obtained plasmids were incorporated in *E.coli* BL21 DE3 pLyss to overexpress the proteins. The overnight cultures were inoculated into shaker flasks with 750 mL LB medium, supplemented with 100 mg/L Ampicillin and 37 mg/L Chloramphenicol, and were grown at 37°C, 180 rpm until OD_{600} ~ 0.7. The protein expression was induced with 1 mM IPTG for 3 h. Then the cells were pelleted (9000 rpm, 4°C) and stored in the freezer (-20°C). The resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10% Glycerol) cell pellet was lysed by sonication (Sonifer 450, Branson). The cleared lysate (25000 rpm, 4°C, 1 h) was treated with Ni(II)-NTA agarose (Quiagen) according to the standard affinity chromatography protocol. The purified proteins were dialyzed into 50 mM Tris-HCl buffer pH 8.0, 100 mM NaCl, flash frozen in liquid N_2 and stored at -80°C.

Production of ^{15}N labelled *tc*FGE variants. The bacterial strains were cultivated in the fermenter INFORS HT in M9 medium, supplemented with $^{15}NH_4Cl$ (99% ^{15}N) as a source of nitrogen atoms. A single *E. coli* colony was inoculated to 5 mL M9 medium, supplemented with 1 mg/mL $^{15}NH_4Cl$, 3 mg/mL glucose, 100 mg/L ampicillin and 37 mg/L chloramphenicol, and incubated overnight at 37°C. The obtained preculture was transferred to 1 L M9 minimal medium, supplemented with 1 g $^{15}NH_4Cl$ and 1 g glucose. The main culture was grown to OD_{600} ~ 2.3 (~10 h, 37ºC, 300 rpm, pH 7.2) and supplemented with 20 mL the feed solution (7.5 g Glc, 0.5 g $^{15}NH_4Cl$ in 100 mL H_20). Then the protein expression was induced with 1 mM IPTG at 20°C for 20 h. During the expression, the culture was fed with the rest of the feed solution. The consumption of glucose and $^{15}NH_4Cl$ was monitored with QUANTOFIX® test stripes (MACHEREY-NAGEL). The harvested culture (OD_{600} ~ 7.2-8.0) were centrifuged (9000 rpm, 20 min, 4°C) and stored in the freezer (-20°C). The resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10% Glycerol) cell pellet was lysed with microfluidizer Emulsiflex HC3. The cleared lysate (25000 rpm, 4°C, 1 h) was treated with Ni(II)-NTA agarose (Quiagen) according to the standard affinity chromatography protocols. The obtained protein sample was treated with TEV-protease (0.05 w/w) during the dialysis against 100 mM phosphate buffer pH 8.0, 200 mM NaCl, 1 mM DTT at RT overnight. Then enzyme was further purified with size exclusion chromatography (HiLoad 26/600 Superdex 200 pg, 200 mM phosphate buffer pH 8.0, 500 mM NaCl). The obtained fractions concentrated with Amicon Ultra (15 mL centrifugal filter a membrane NMWL of 10 kDa), flash frozen in liquid N_2 and stored at -80°C.

Production of ^{2}H ^{13}C ^{15}N labelled *tc*FGE. The bacterial strain was cultivated in the fermenter INFORS HT in the M9 medium, supplemented with D_2O (99.8% ^{2}H), D-Glucose-$^{13}C_6{}^{2}D_7$ (97% ^{2}D, 99% ^{13}C) and $^{15}NH_4Cl$ (99% ^{15}N) as the sole of carbon and nitrogen atoms. A single colony *E.coli* BL21 DE3 pLyss bearing the plasmid pET19b_*tc*FGE was transferred to 10 mL M9 medium, supplemented with 1 mg/mL NH_4Cl, 3 mg/mL glucose, 100 mg/L ampicillin and 37 mg/L chloramphenicol, and incubated overnight at 37°C. The resulted culture was pelleted (8000 rpm, 5 min) and used to inoculate the shaker flask with 50 mL M9 D_2O medium, supplemented with 99.8% D_2O, 3 mg/mL D-Glucose-$^{13}C_6$ $^{2}D_7$, 1 mg/mL $^{15}NH_4Cl$, 100 mg/L ampicillin and 37 mg/L chloramphenicol, which further was incubated overnight at 37°C, 200 rpm. The obtained cells (OD_{600} ~ 0.8) were transferred in 1 l M9 D_2O medium (99.8% D_2O, 1 g/l D-Glucose-$^{13}C_6{}^{2}D_7$, 0.5 g/l $^{15}NH_4Cl$, 100 mg/L ampicillin and 37 mg/L chloramphenicol). The main culture was grown to OD_{600} ~ 2.3 (~21 h, 37ºC, 300 rpm, pH 7.2) and supplemented with 15 mL the feed solution (7.0 g D-Glucose-$^{13}C_6{}^{2}D_7$, 0.5 g $^{15}NH_4Cl$, in 100 mL

D_2O). The protein expression was induced with 1 mM IPTG at 20°C for 10 h. During the expression, the culture was fed with the rest of the feed solution. The consumption of D-Glucose-$^{13}C_6{}^2D_7$ and $^{15}NH_4Cl$ was monitored with QUANTOFIX® test stripes (MACHEREY-NAGEL). The harvested culture (OD_{600} ~ 8.6) were centrifuged (9000 rpm, 20 min, 4ºC) and stored in the freezer (-20°C). The resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10% Glycerol) cell pellet was lysed with microfluidizer Emulsiflex HC3. The cleared lysate (25000 rpm, 4°C, 1 h) was treated with Ni(II)-NTA agarose (Quiagen) according to the standard affinity chromatography protocols. The obtained protein sample was treated with TEV-protease (0.05 w/w) during the dialysis against 100 mM Tris-HCl buffer pH 8.0, 200 mM NaCl, 1 mM DTT at RT overnight. Then enzyme was further purified with size exclusion chromatography (HiLoad 26/600 Superdex 200 pg, 100 mM Tris-HCl buffer pH 8.0, 200 mM NaCl). To perform the backbone amide hydrogen exchange the protein sample was dialyzed against 40 mM Tris/HCl buffer pH 8.0, 150 mM NaCl in presence 1 mM EDTA at RT for 48 h with subsequent removal EDTA and incubation with 2 mM $CaCl_2$ (3h, RT). Afterwards the Tris buffer was exchanged to 200 mM phosphate buffer pH 8.0, 750 mM NaCl with Amicon Ultra (15 mL centrifugal filter a membrane NMWL of 10 kDa), concentrated, flash frozen in liquid N_2 and stored at -80°C.

Production of *tc*FGE for crystallization. The shaker flasks with 8*1 l ZYM-5052 [6] medium, supplemented with 100 mg/L Ampicillin and 37 mg/L Chloramphenicol, were inoculated with overnight culture of *E.coli* BL21 DE3 pLyss, bearing the plasmid pET19b_*tc*FGE, and the cells were grown at 37°C, 180 rpm until OD_{600} ~ 0.7. Afterwards the temperature was decreased to 20°C and the cultures were cultivated for 20 h. Then the cells were pelleted in the centrifuge (9000 rpm, 4°C) and stored in the freezer (-20°C). The resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10% Glycerol) cell pellet was lysed with microfluidizer Emulsiflex HC3. The cleared lysate (25000 rpm, 4°C, 1 h) was treated with Ni(II)-NTA agarose (Quiagen) according to the standard affinity chromatography protocols. The obtained protein sample was treated with TEV-protease (0.05 w/w) during the dialysis against 100 mM phosphate buffer pH 8.0, 200 mM NaCl, 1 mM DTT at RT overnight. Then enzyme was further purified with size exclusion chromatography (HiLoad 26/600 Superdex 200 pg, 200 mM phosphate buffer pH 8.0, 500 mM NaCl). The obtained fractions concentrated with Amicon Ultra (15 mL centrifugal filter a membrane NMWL of 10 kDa).
The purity of the proteins was confirmed with SDS PAGE and HRMS (ESI). The labeling efficiency for the isotope labeled *tc*FGE variants was estimated on the base of HRMS results. The final protein concentration was estimated from the A_{280} with usage of calculated from amino acid sequences molar absorption coefficients [7].

Crystallization and diffraction data collection. Sitting drop vapor diffusion method was applied for the crystallization of *tc*FGE under strict exclusion of dioxygen in an inert gas chamber containing 95/5 % mixture of N_2/H_2. Using an enzymatic oxygen scavenging system [8] dissolved oxygen was removed from protein and substrate sample. The protein sample *tc*FGE (21.5 mg/mL, 0.61 mM), $CuSO_4$ (0.83 mM), 2.7 mM DTT as well as FGE-27 (1.7 mM) was supplemented 2 mM glucose, 2 U glucose oxidase (0.01 mg/mL, *A. niger*, Sigma) and 200 U catalase (0.04 mg/mL, bovine liber, Sigma) and incubated in the glovebox (N_2 atmosphere, $O_2 \leq 0.1$ ppm) overnight at RT. Prior to crystallization the protein substrate mixture was incubated for 90 min. On a 96-well 2-drop plate (SwissSci) 0.6 µL reservoir solution of crystallization screen was combined with 0.6 µL protein solution. Sealed plates were stored at 23°C for reservoir / protein equilibration and formation of three-dimensional crystals was observed after one week in a drop containing 10 % PEG 8000, Tris-HCl (0.1 M, pH 7.0) and $MgCl_2$ (0.2 M). Crystals were looped and flash frozen in liquid nitrogen approximately 3 weeks past

crystallization. Diffraction data were collected at X06DA (PXIII) and X06SA (PXI) beamline of Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) equipped with a Pilatus 2 M (Dectris) or an Eiger-16M detector (Dectris), respectively. High resolution data sets were collected at an X-ray energy of 12398 eV (λ = 1.000 Å) at the temperature 100 K. Data were indexed and integrated with XDS [9] and scaled and merged with Aimless [10] from the CCP4 program suite [11-12]. *tc*FGE crystals grew into orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 58.4 Å, b = 71.9 Å, c = 76.7 Å, $\alpha = 90°$, $\beta = 90°$, $\gamma = 90°$. A single protein monomer was contained per unit cell.

Data processing and refinement. Due to the excellent diffraction quality the data were scaled to a high-resolution limit of 1.04 Å. A model of the truncated N20FGE silver co-crystallized apo structure (PDB: 5NXL) was used as search model for molecular replacement using Phenix Phaser [13]. Iterative cycles of manual and automated model building into the readily interpretable electron density map and refinement using COOT [14-15] and phenix.refine [16-20]. Hydrogen atoms were generated and used for the refinement but not included into the final model. *tc*FGE polypeptide chain was modeled into continuous electron density for residue 1 – 303 including the so far unresolved residues from Val92 to Val102 with disorder for the C-terminal Leu303. Aside of the protein chain the asymmetric unit cell contained additional electron density that could have readily been modeled by adding the substrate polypeptide chain (NH_2-Abz-AlaThrThrProLeuCysGlyProSerArgAlaSerIleLeuSerGlyArg-CO_2H) resolved from N-terminal *o*-aminobenzoic acid to Gly17. Substrate peptide residues from Leu15 onwards are no more accommodated in a protein surface substrate binding cleft but are found in the solvent exposed area. More remarkable the ultimate C-terminal restudies Gly11 & Arg12 are located at the interface of two crystallographic symmetry mates (-1,0,-1 & -1,-1,-1 molecules). The interaction of the substrate residues with further FGE molecules could be beneficial for crystallization. Data processing and refinement statistics are given in Table S3.

High-resolution mass spectrometry (HRMS). For HRMS measurements protein samples were diluted to concentration 0.2 mg/mL in 1 mM DTT, 1% FA and injected into a reversed-phase C4 column (Jupiter C4 300A; 50 x 2 mm, 5 Microns, Phenomex; Shimadzu Nexera-X2) coupled to a HRMS instrument (maXis II ESI-TOF, Bruker). The obtained MS data were deconvoluted with Bruker Compass DataAnalysis 4.4 SR1 software.

Solid phase peptide synthesis (SPPS). The substrate peptide 17-residue peptide (FGE-27, sequence: Abz-ATTPLCGPSRASILSGR-NH_2) and 9-residue peptide (FGE-1, Abz-SALCSPTRA-NH_2) were synthesized with Syro I peptide synthesizer (Biotage) using standard protocols for Fmoc SPPS.[21] The crude product was purified to homogeneity with C18 reverse phase HPLC and analyzed with ESI-MS (Agilent 6130 Quadrupole LC/MS, C18 column ZORBAX Eclipse Plus RRHD, 1.8 µm, 2.1 x 50 mm) (Figure S17).

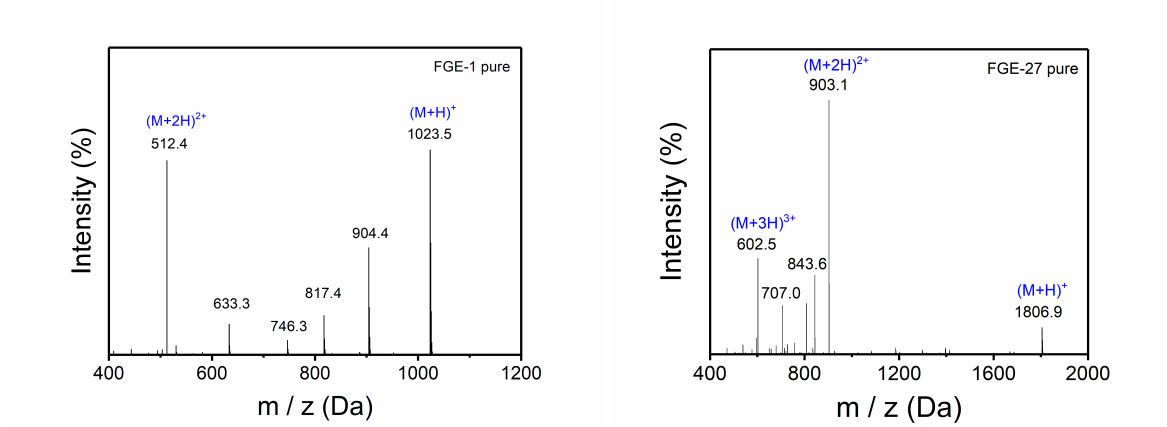


Figure S15 – The mass spectra of substrates after purification with C18 reverse phase HPLC. The analyses were performed with UPLC-MS Agilent 1290 Infinity system equipped with an Agilent 6130 Quadrupole LC/MS using a C18 column (ZORBAX Eclipse Plus RRHD, 1.8 µm, 2.1 x 50 mm) **Left:** The mass spectra purified FGE-1 showed protonated (m/z = 1023.5 Da) and double protonated (m/z = 512.4 Da) peaks. **Right:** The mass spectra purified FGE-27 showed protonated (m/z = 1806.9 Da), double protonated (m/z = 903.1 Da) and triple protonated (m/z = 602.5 Da) peaks.

NMR experiments. NMR spectra of apo forms of *tc*FGE variants as well as metalated one were recorded in oxygen-contained buffers. To prevent substrate oxidation during NMR experiments dissolved oxygen was removed with an enzymatic oxygen scavenging system [8]. The protein as well as FGE-27 were separately diluted in 20 mM phosphate buffer (pH 8.2), supplemented with 50 mM NaCl, 2 mM DTT, 2 mM glucose, 2 U glucose oxidase (*A. niger*, Sigma) and 200 U catalase (bovine liber, Sigma) and 5% D_2O. After addition of the GO and CAT samples were incubated in the glovebox (N_2 atmosphere, $O_2 \leq 0.1$ ppm) overnight at RT. Just before every measurement, the enzyme and the substrate solutions were combined and 350 µL of final solution were transferred to a SHIGEMI micro cell tube in the glovebox. TROSY based experiments were used to record ^{1}H-^{15}N HSQC [22] and triple resonance 3D-HNCO [23] spectra. NMR spectra were recorded on Bruker Avance III NMR spectrometers operating at 600.13 MHz, either equipped with a 5 mm BBFO probe head with z-axis pulsed field gradients or a 5 mm $^{1}H/^{13}C/^{15}N/^{19}F$ QCI-F cryogenic probe head with z-axis pulsed field gradients.

SUPLEMENTARY INFORMATION REFERENCES

1. Meury, M.; Knop, M.; Seebeck, F. P., Structural Basis for Copper-Oxygen Mediated C-H Bond Activation by the Formylglycine-Generating Enzyme. *Angew Chem Int Ed Engl* **2017,** *56* (28), 8115-8119.
2. Mollova, E. T.; Metzler, D. E.; Kintanar, A.; Kagamiyama, H.; Hayashi, H.; Hirotsu, K.; Miyahara, I., Use of 1H-15N heteronuclear multiple-quantum coherence NMR spectroscopy to study the active site of aspartate aminotransferase. *Biochemistry* **1997,** *36* (3), 615-25.
3. John Cavanagh, W. J. F., Arthur G. Palmer III, Mark Rance and Nicholas J. Skelton *Protein NMR Spectroscopy - Principles and Practice*. ACADEMIC PRESS: 2007.
4. Muhandiram, D. R.; Kay, L. E., Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B* **1994,** *103* (3), 203-216.
5. Knop, M.; Engi, P.; Lemnaru, R.; Seebeck, F. P., In Vitro Reconstitution of Formylglycine-Generating Enzymes Requires Copper(I). *Chembiochem* **2015,** *16* (15), 2147-50.
6. Studier, F. W., Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **2005,** *41* (1), 207-234.
7. Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker, J. M., Ed. Humana Press: 2005; pp 571-607.
8. Plumere, N.; Henig, J.; Campbell, W. H., Enzyme-catalyzed O2 removal system for electrochemical analysis under ambient air: application in an amperometric nitrate biosensor. *Anal Chem* **2012,** *84* (5), 2141-6.
9. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.
10. Evans, P. R.; Murshudov, G. N., How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **2013,** *69* (Pt 7), 1204-14.
11. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 1), 72-82.
12. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 235-42.
13. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J Appl Crystallogr* **2007,** *40* (Pt 4), 658-674.
14. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 486-501.
15. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.
16. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **2012,** *68* (Pt 4), 352-67.
17. Terwilliger, T. C., Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2144-9.
18. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 1), 12-21.
19. Headd, J. J.; Echols, N.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Chen, V. B.; Moriarty, N. W.; Richardson, D. C.; Richardson, J. S.; Adams, P. D., Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D Biol Crystallogr* **2012,** *68* (Pt 4), 381-90.
20. Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Adams, P. D.; Read, R. J.; Zwart, P. H.; Hung, L. W., Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr D Biol Crystallogr* **2008,** *64* (Pt 5), 515-24.
21. W. CHAN, P. W., *Fmoc solid phase peptide synthesis A Practical Approach*. Oxford University Press Inc.: New York, 2000.
22. Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K., Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* **1997,** *94* (23), 12366-12371.
23. Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K., TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (23), 13585-13590.