Supplementary Information

Installing Guest Molecules at Specific Sites Within Scaffold Protein Crystals

Thaddaus R. Huber, Eli C. McPherson, Carolyn E. Keating, Christopher D. Snow

Chemical and Biological Engineering
Colorado State University
1301 Campus Delivery, Fort Collins, Colorado 80523, USA
E-mail: Christopher.Snow@colostate.edu

MKEYTLDKAHTDVGFKIKHLQISNVKGNFKDYSAVIDFDPASAEFKKLDVTIKIASVNTENQTRN

HLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKEI

GFSLNGKIKRSDFKFATSTSTITLSDDINLNIEVKANEKEGGSHHHHHH

Figure S1. Amino acid sequence for CJ without cysteine mutation. The leader peptide on the native putative periplasmic polyisoprenoid-binding protein from *Campylobacter jejuni* was deleted to ease expression and purification.

G34C

MKEYTLDKAHTDV**C**FKIKHLQISNVKGNFKDYSAVIDFDPASAEFKKLDVTIKIASVNTENQTRD

NHLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKE

KIGFSLNGKIKRSDFKFATSTSTITLSDDINLNIEVKANEKEGGSHHHHHH

Figure S2. Amino acid sequence for CJ-G34C.

N48C

MKEYTLDKAHTDVGFKIKHLQISNVKG**C**FKDYSAVIDFDPASAEFKKLDVTIKIASVNTENQTRD

NHLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKE

KIGFSLNGKIKRSDFKFATSTSTITLSDDINLNIEVKANEKEGGSHHHHHH

Figure S3. Amino acid sequence for CJ-N48C.

MKEYTLDKAHTDVGFKIKHLQISNVKGNFKDYSAVIDFDPASAEFKKLDVTIKIASVNTENQTRD

NHLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKE

N182C

KIGFSLNGKIKRSDFKFATSTSTITLSDDINL**C**IEVKANEKEGGSHHHHHH

Figure S4. Amino acid sequence for CJ-N182C.

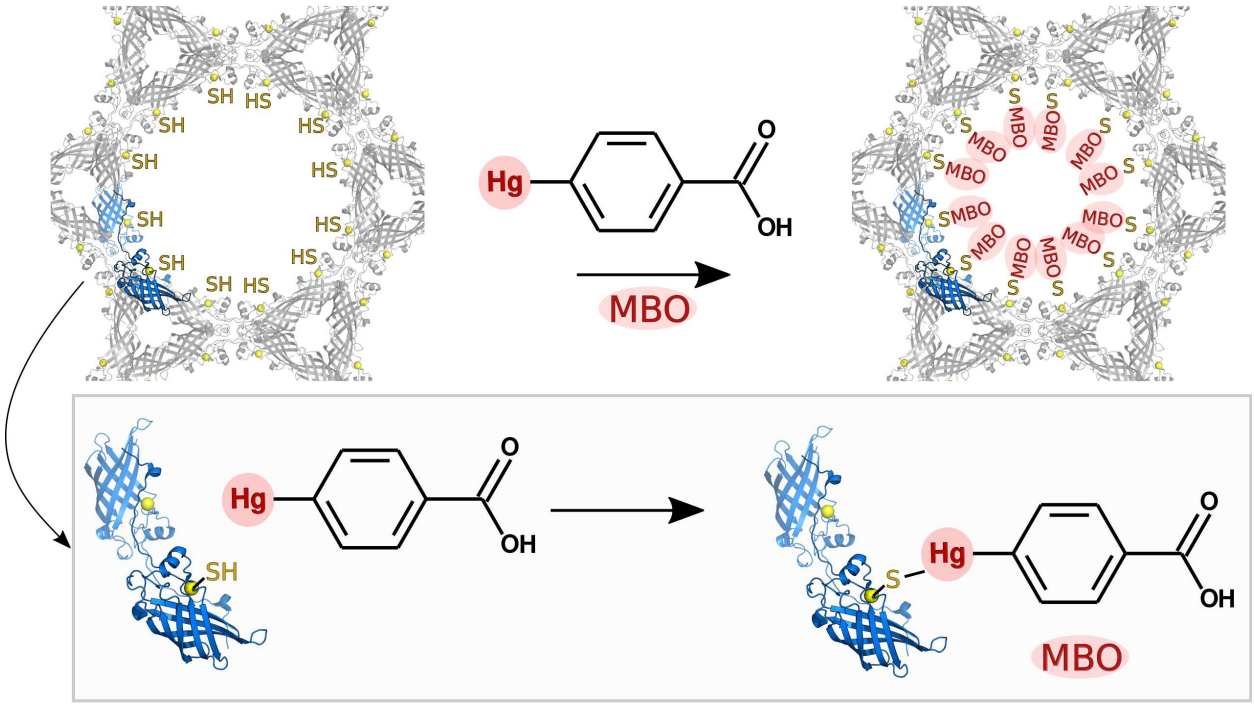

Figure S5. Due to the strong propensity of thiols to covalently bind mercury, mercuribenzoic acid (MBO) was used to demonstrate the accessibility of the engineered cysteines to the solvent channels. The resulting heavy atom adducts were easy to observe via XRD.

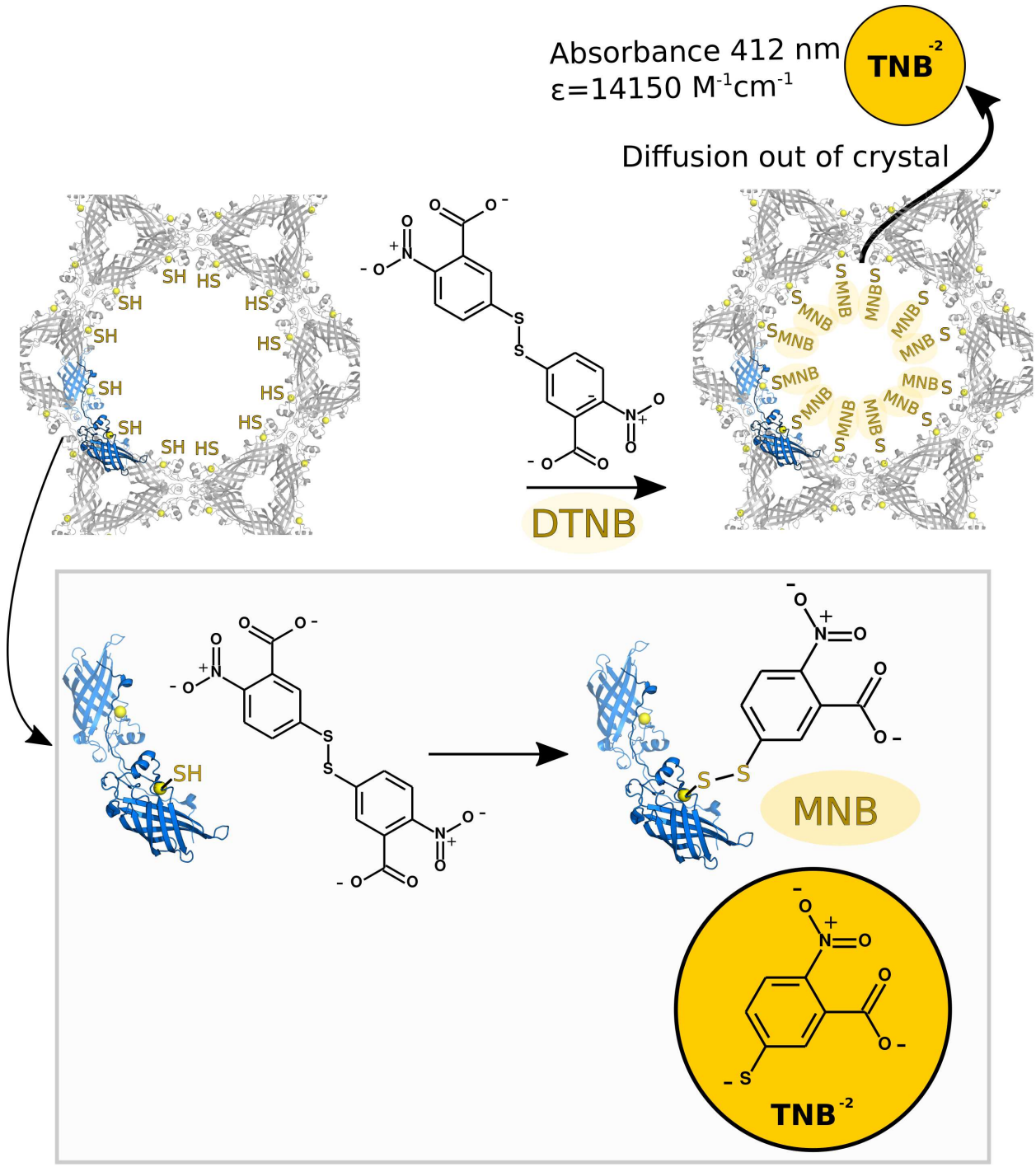

Figure S6. Ellman's reagent (5,5'-dithio-bis-[2-nitrobenzoic acid]) reacts with thiols via an SN_2 reaction that forms a mixed disulfide product with the addition of 5-mercapto-2-nitro-benzoic acid (MNB) to reduced thiols. The reaction can be monitored by measuring the release of 2-nitro-5-thiobenzoate anion (TNB^{-2}). which absorbs strongly at 412 nm due to the resonance stabilized 4-thiolpyridone tautomer. These properties made Ellman's reagent an attractive choice for demonstrating disulfide exchange in CJ cysteine mutant crystals.

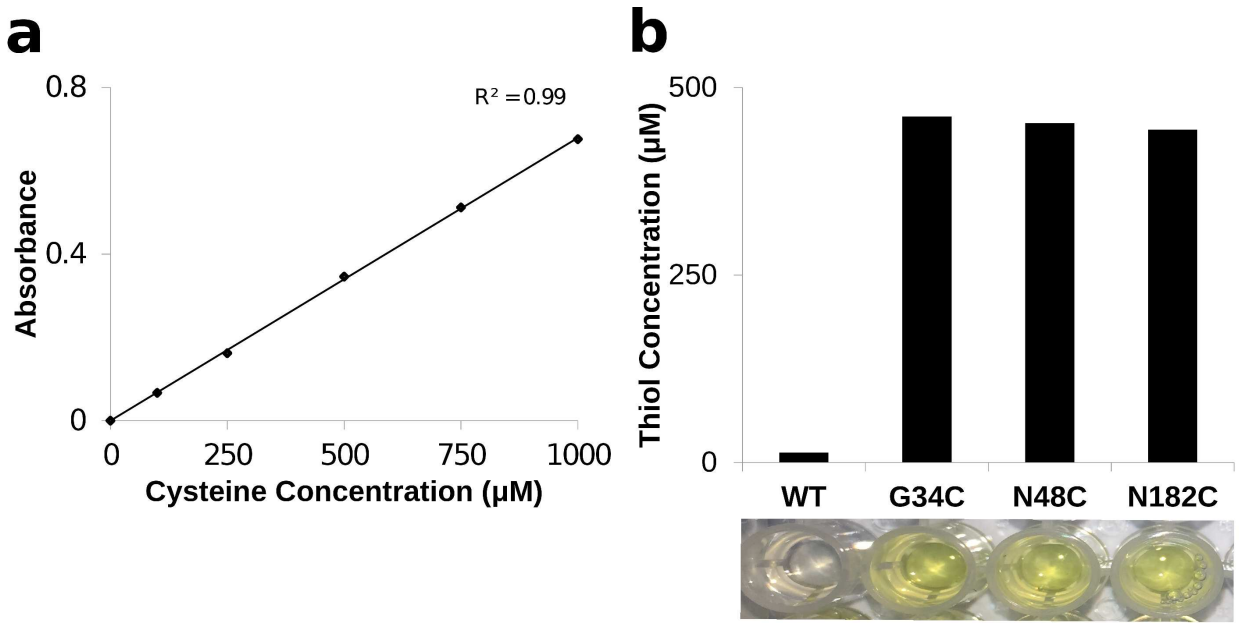

Figure S7. Thiol concentration in solution can be measured by addition of Ellman's reagent and measuring absorbance 412 nm. **(a)** An *in vitro* Ellman's Reagent standard curve was prepared for reduced L-cysteine from 0-1000 µM. **(b)** Purified CJ-variants were diluted to ~10 mg/mL (~500 µM) and Ellman's reagent was added to confirm the presence and accessibility of thiols in solution. Only CJ-variants with engineered cysteines produced a signal at 412 nm.

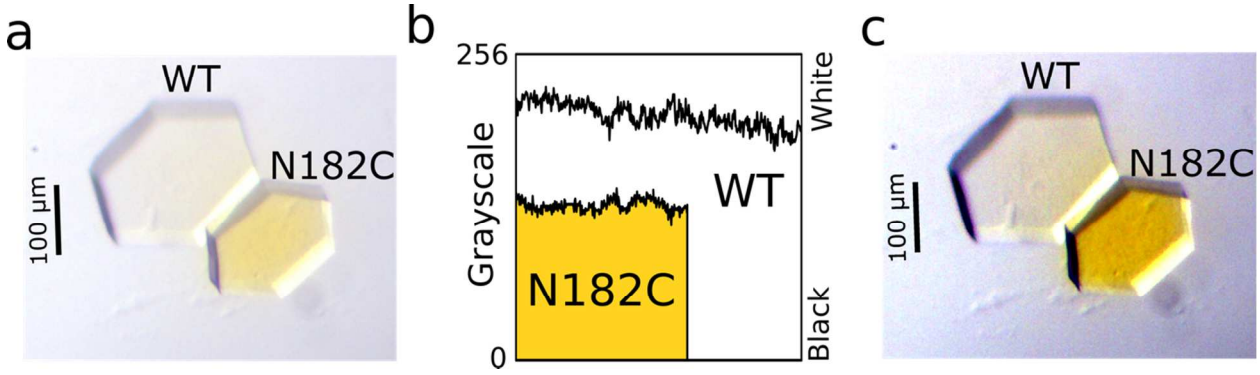

Figure S8. This figure is best observed in the color images available online. Wild-type and CJ cysteine mutant crystals were exposed to Ellman's reagent and extensively washed to remove unreacted Ellman's reagent. **(a)** The raw image above was obtained later, 30 seconds after adding 10 mM 2-mercaptoethanol (BME). Addition of BME produced an intense yellow signal only on CJ cysteine mutant crystals, suggesting release of covalently bound 5-mercapto-2-nitro-benzoic acid (MNB). **(b)** To quantify the difference, we use ImageJ to extract grayscale values across diagonal vectors that span each crystal (avoiding the edge effects). **(c)** The pathlength of these crystals is small, which results in a somewhat subtle signal. Here, a contrast-enhanced image assists with visualizing the difference between the crystals and better represents what is observed by eye. The contrast between the cysteine-bearing crystal and the WT crystal would likely be even larger if the two were not adjacent, since the TNB dye molecule released by the N182C crystal can diffuse into the WT crystal.

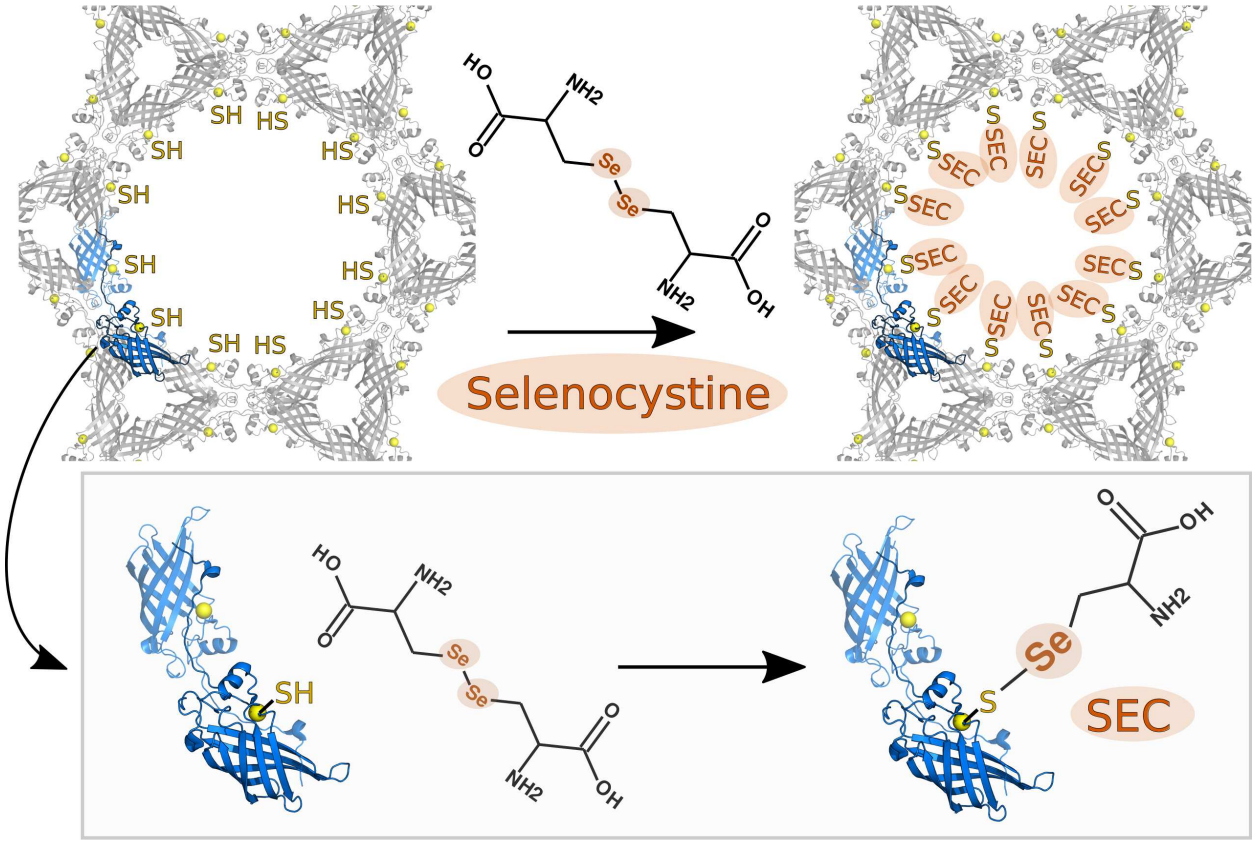

Figure S9. In a reaction analogous to disulfide exchange, thiols can form mixed oxidized products with diselenide compounds. The reaction of thiols with diselenides has the benefit of adding a heavy atom at the attachment point. Selenium addition was readily visible in XRD. However, the remainder of the adduct was too flexible to be clearly resolved in 2 out of 3 cases.

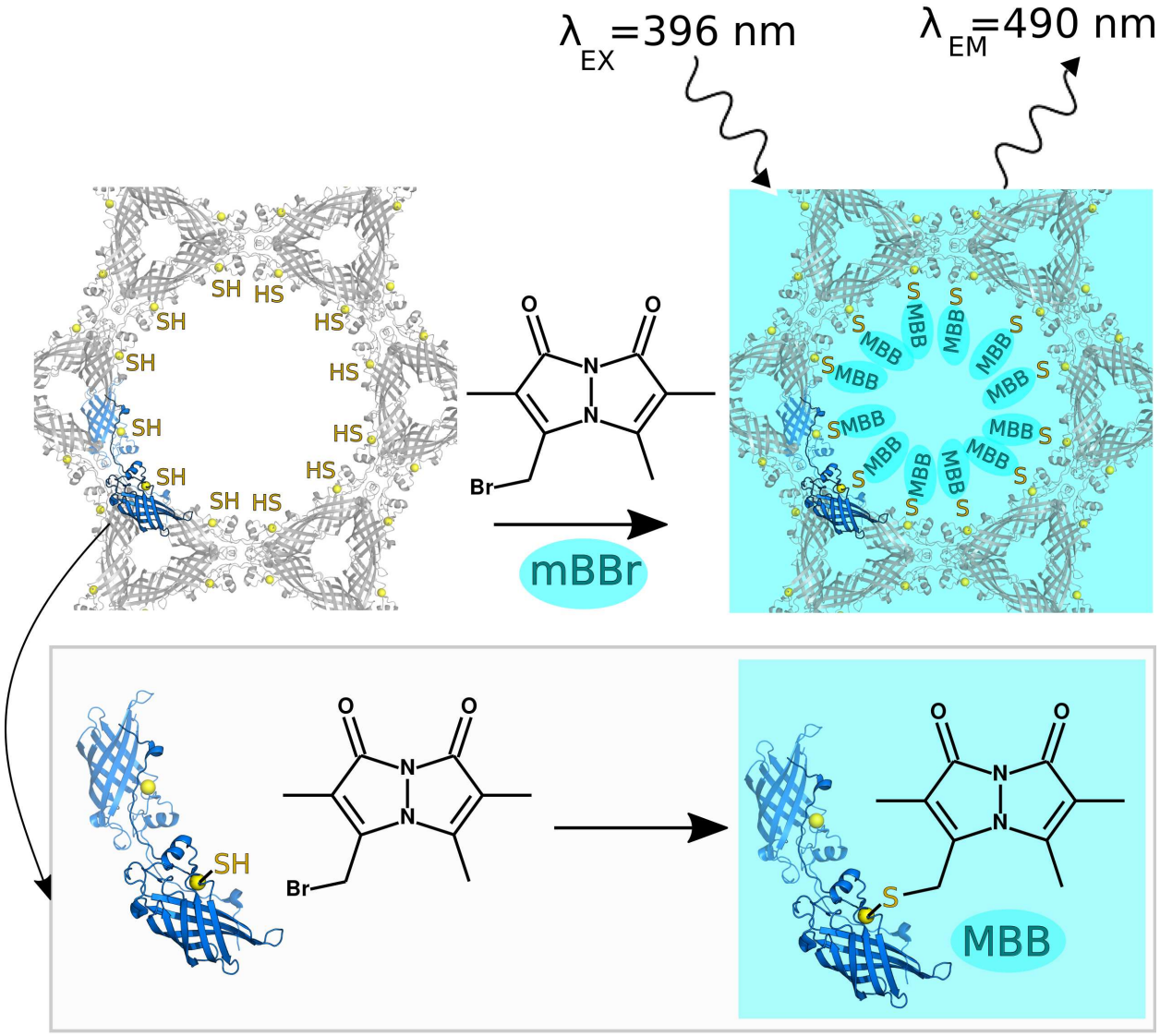

Figure S10. Haloalkyl groups react with thiols via an SN2 reaction resulting in a stable thioether linkage. We used monobromobimane (mBBr) to demonstrate the CJ cysteine crystals ability to react with haloalkyl derivatives. mBBr is essentially non-fluorescent until forming the resulting bimane (MBB) conjugate.

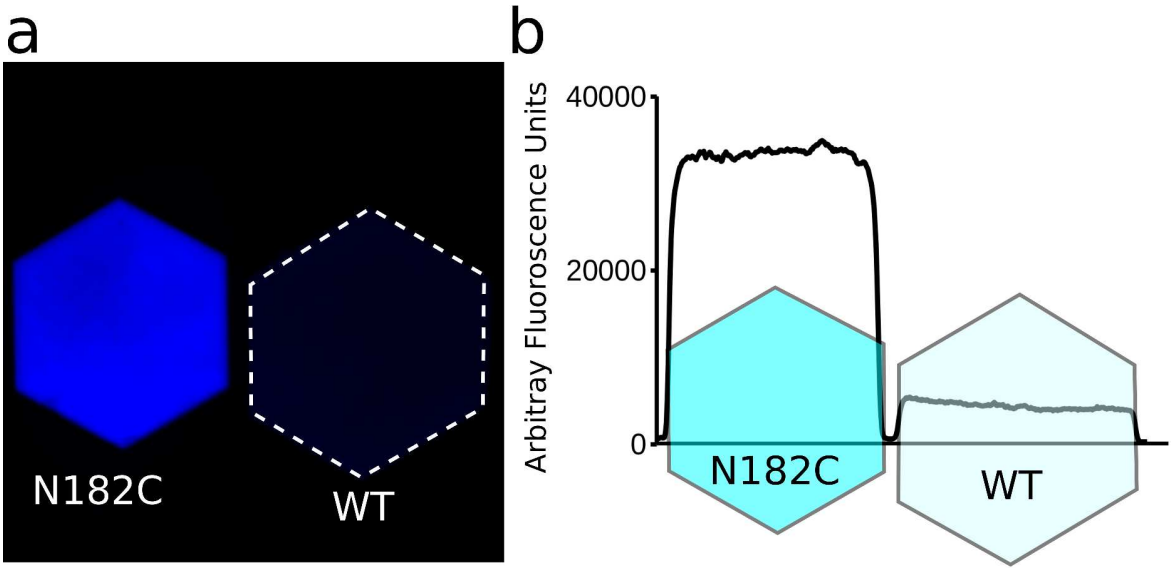

Figure S11. We incubated a CJ without thiol (WT) crystal and N182C crystal in a monobromobimane (mBBr) solution for 30 minutes. The crystals were then washed in 4.2 M trimethylamine-N-oxide (TMAO) pH 7.4 to remove excess reactive species. **(a)** Subsequent confocal microscopy images (λ_{EX}=405 nm) of N182C (left) and WT (right) after exposure to mBBr. N182C was fluorescent, indicative of monobromobimane installation. **(b)** Raw pixel values for N182C (left) and WT (right) show that mBBr was selectively installed on the engineered cysteine. The WT crystal had minimal background fluorescence relative to the N182C with mBBr installed.

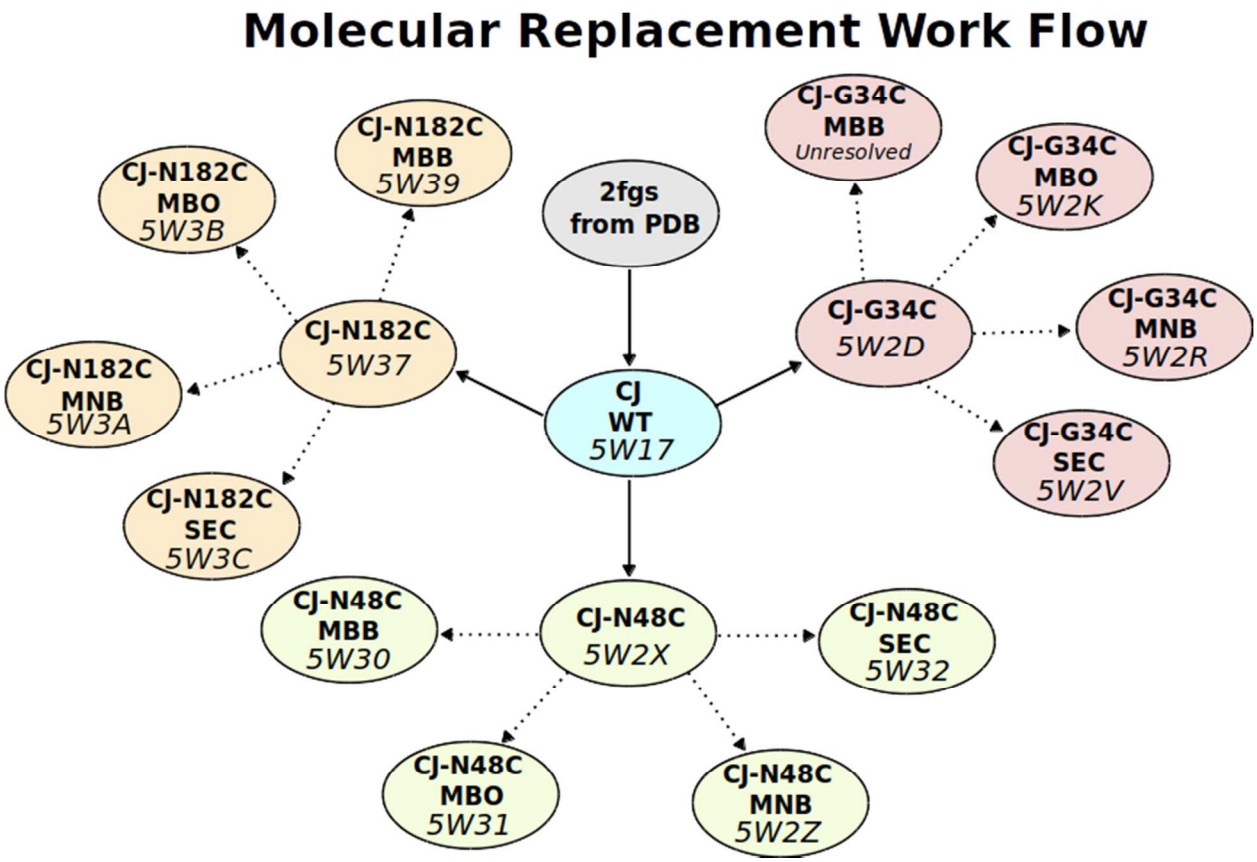

Figure S12. A structure for *Campylobacter jejuni* YceI periplasmic protein at 2.9 Å resolution was deposited by authors at the New York Structure Genomics team (PDB code 2fgs). We obtained an updated model for this protein at improved resolution (2.58 Å). Improved resolution allowed for further refinement of side chains, modeling of ordered water, and placement of a ligand in the hydrophobic core of the protein. The identity of the hydrophobic ligand remains unknown, but a saturated C18 ligand was modeled as a placeholder (**Figure S18**). This improved model (CJ WT) served as the molecular replacement model for the reduced CJ thiol mutant crystals: G34C, N48C, and N182C. The new thiol mutant models further served as models for the resulting CJ thiol crystals conjugated with small molecules. Non-trivial changes from the input molecular replacement model were only made if there was a strong reason such as improved side chain resolution, disrupted hydrogen bond network, modeling new features in the electron density, etc.

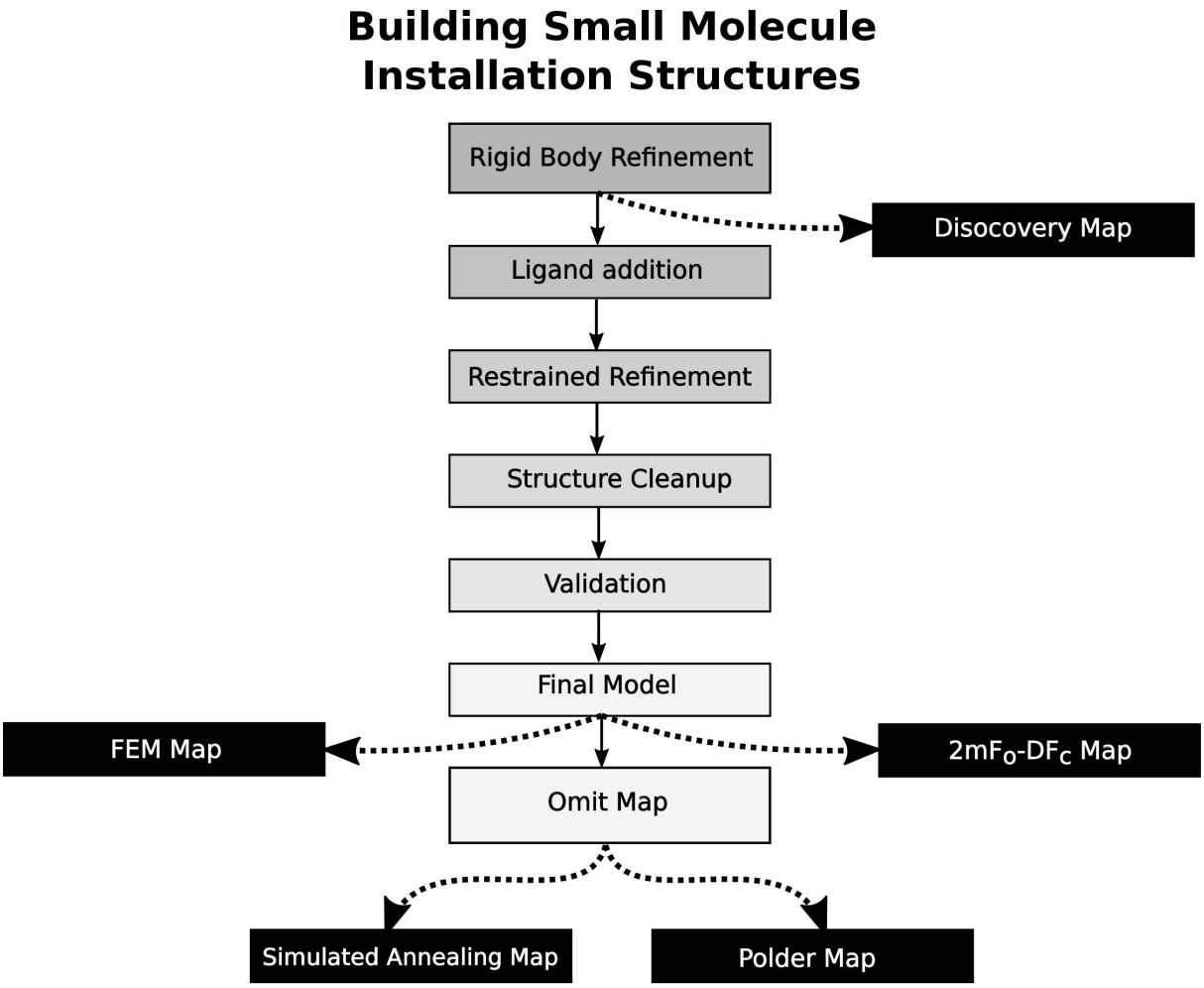

Figure S13. To prevent bias, all structures with installed small molecules were generated with the same work flow. The corresponding reduced thiol structure (without an adduct) was used as a molecular replacement model. An initial "discovery map" was generated by performing rigid body refinement and identifying large map peaks suggestive of small molecule installation. Next, the probable small molecule is added to the observed electron density and a conformation was selected based on the local environment and preferred geometries. Restrained refinement was performed on the resulting structure and the occupancy of the ligand was adjusted to minimize difference map peaks. The remaining structure was refined to improve interactions and fit observed electron density. The structure was then validated on the PDB server and the final model was prepared. From the final model, we generated $2mF_o$-DF_c maps for Figure 2 in the main text. We additionally generated FEM maps[1] to assess adduct placement. To further guard against bias, we generated Polder omit maps using phenix.polder.[2] The Polder maps also exclude bulk solvent from the immediate neighborhood of the adduct. Finally, we used phenix.composite_omit_map to generate conventional simulated annealing[3] omit maps with harmonic restraints. These maps are

compared for each installation site in **Figure S15** (G34C)**, S16** (N48C)**, and S17** (N182C).

	CJ-WT	G34C	N48C	N182C
Data Collection				
Light Source	Synchrotron	Synchrotron	Synchrotron	Synchrotron
Wavelength (Å)	1.0	1.0	1.0	1.0
Spacegroup	*P 6 2 2*	*P 6 2 2*	*P 6 2 2*	*P 6 2 2*
Cell dimensions				
a, b, c (Å)	178.81, 178.81, 50.34	178.88, 178.88, 50.72	179.15, 179.15, 50.77	178.43, 178.43, 50.58
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	38.71-2.58 (2.72-2.58)*	38.73-2.70 (2.85-2.70)*	38.79-2.73 (2.88-2.73)*	38.63-2.52 (2.66-2.52)*
Mesaured reflections	313158	280434	278863	336425
Unique reflections	15452	13638	13248	16542
Completeness (%)	99.9 (99.6)	99.9 (100.0)	99.8 (98.9)	99.9 (99.5)
Redundancy	20.3 (19.9)	20.6 (21.2)	21.0 (21.8)	20.3 (17.4)
Rpim	0.051 (0.690)	0.047 (0.694)	0.044 (1.472)	0.031 (0.733)
$cc_{1/2}$	0.997 (0.897)	0.999 (0.761)	0.999 (0.788)	0.999 (0.784)
$I/\sigma(I)$	10.2 (1.0)	13.2 (1.0)	13.5 (1.0)	22.8 (1.0)
Refinement				
Rwork/Rfree	0.2212/0.2537	0.2299/0.2713	0.2100/0.2393	0.2149/0.2578
No. Atoms				
Protein	1321	1323	1319	1319
LFA	17	17	17	17
SO4	10	10	10	10
Water	21	18	17	21
B-Factors				
Protein	69.69	94.25	83.69	80.47
LFA	69.11	88.35	81.92	75.38
SO4	75.29	101.22	89.43	92.45
Water	57.53	79.65	68.6	72.66
RMSD				
Bond lengths (Å)	0.013	0.013	0.014	0.014
Bond angles (°)	1.856	1.826	1.918	1.935
Ramachandran plot (%)				
Favored region	0.947	0.941	0.935	0.935
Allowed region	0.053	0.059	0.065	0.065
Outlier region	0	0	0	0

* values in parantheses are for high resolution shell

Table S1. Crystallography data collection and refinement statistics for CJ without thiol (CJ-WT), G34C, N48C, and N182C.

	G34C-MBO	N48C-MBO	N182C-MBO
Data Collection			
Light Source	Synchrotron	Synchrotron	Synchrotron
Wavelength (Å)	1.0	1.0	1.0
Spacegroup	*P 6 2 2*	*P 6 2 2*	*P 6 2 2*
Cell dimensions			
a, b, c (Å)	178.81, 178.81, 50.64	178.79, 178.79, 50.81	179.07, 179.07, 50.65
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	38.71-2.78 (2.93-2.78)*	38.71-2.56 (2.70-2.56)*	38.77-2.70 (2.85-2.70)*
Mesaured reflections	244820	330151	245372
Unique reflections	12465	15939	13431
Completeness (%)	99.7 (98.2)	99.9 (99.9)	98.8 (92.2)
Redundancy	19.6 (13.9)	20.7 (19.0)	18.3 (7.4)
Rpim	0.039 (0.770)	0.040 (0.628)	0.025 (0.441)
$cc_{1/2}$	0.999 (0.758)	0.999 (0.857)	0.999 (0.875)
$I/\sigma(I)$	11.4 (1.0)	14.4 (1.0)	20.7 (1.4)
Refinement			
Rwork/Rfree	0.2214/0.2614	0.2133/0.2412	0.2047/0.2281
No. Atoms			
Protein	1323	1315	1316
LFA	17	17	17
SO4	10	10	10
Water	18	17	21
MBO	1	10	11
B-Factors			
Protein	94.44	76.12	73.56
LFA	93.26	75.64	66.64
SO4	106.09	78.6	79.72
Water	80.18	66.85	62.05
MBO	305.21	111.11	74.8
RMSD			
Bond lengths (Å)	0.013	0.014	0.062
Bond angles (°)	1.843	1.94	2.013
Ramachandran plot (%)			
Favored region	0.917	0.935	0.976
Allowed region	0.083	0.065	0.024
Outlier region	0	0	0

* values in parantheses are for high resolution shell

Table S2. Crystallography data collection and refinement statistics for installation of mercuribenzoic acid (MBO) on G34C, N48C, and N182C.

	G34C-MNB	N48C-MNB	N182C-MNB
Data Collection			
Light Source	Synchrotron	Synchrotron	Rigaku Homelab
Wavelength (Å)	1.0	1.0	1.54187
Spacegroup	*P 6 2 2*	*P 6 2 2*	*P 6 2 2*
Cell dimensions			
a, b, c (Å)	178.34, 178.34, 50.32	179.86, 179.86, 50.62	180.33, 180.33, 50.57
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	38.61-2.90 (3.06-2.90)*	38.94-2.80 (2.95-2.80)*	48.11-2.76 (2.91-2.76)*
Mesaured reflections	219589	255445	222910
Unique reflections	10917	12372	12968
Completeness (%)	99.7 (99.1)	99.9 (99.8)	100.0 (100.0)
Redundancy	20.1 (20.0)	20.6 (20.9)	17.2 (9.4)
Rpim	0.059 (0.945)	0.049 (1.050)	0.080 (0.805)
$cc_{1/2}$	0.998 (0.909)	0.998 (0.688)	0.994 (0.390)
I/σ(I)	9.4 (1.0)	11.2 (1.0)	9.0 (1.0)
Refinement			
Rwork/Rfree	0.2158/0.2732	0.2032/0.2358	0.2371/0.2746
No. Atoms			
Protein	1323	1315	1315
LFA	17	17	17
SO4	10	10	10
Water	17	21	22
MNB	13	13	13
B-Factors			
Protein	101.1	79.34	53.78
LFA	94.44	78.64	52.35
SO4	106.39	85.1	57.6
Water	84.9	73.96	40.14
MNB	161.51	163.36	106.77
RMSD			
Bond lengths (Å)	0.012	0.013	0.013
Bond angles (°)	1.688	1.884	1.77
Ramachandran plot (%)			
Favored region	0.917	0.947	0.947
Allowed region	0.083	0.053	0.047
Outlier region	0	0	0.006

* values in parantheses are for high resolution shell

Table S3. Crystallography data collection and refinement statistics for installation of 5-mercapto-2-nitro-benzoic acid (MNB) on G34C, N48C, and N182C.

	G34C-SEC	N48C-SEC	N182C-SEC
Data Collection			
Light Source	Synchrotron	Synchrotron	Rigaku Homelab
Wavelength (Å)	1.0	1.0	1.54187
Spacegroup	*P 6 2 2*	*P 6 2 2*	*P 6 2 2*
Cell dimensions			
a, b, c (Å)	178.02, 178.02, 50.71	178.62, 178.62, 50.81	180.34, 180.34, 50.59
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	38.54-2.90 (3.06-2.90)*	38.67-2.63 (2.77-2.63)*	48.13-2.87 (3.03-2.87)*
Mesaured reflections	203030	287615	192807
Unique reflections	10946	14539	11574
Completeness (%)	99.7 (98.3)	98.8 (92.5)	100.0 (100.0)
Redundancy	18.5 (12.5)	19.8 (12.3)	16.7 (7.5)
Rpim	0.048 (0.675)	0.022 (0.464)	0.074 (0.723)
$cc_{1/2}$	0.999 (0.783)	0.999 (0.888)	0.994 (0.412)
I/σ(I)	10.7 (1.0)	25.5 (1.6)	9.4 (1.0)
Refinement			
Rwork/Rfree	0.2127/0.2608	0.2140/0.2360	0.2334/0.2738
No. Atoms			
Protein	1323	1315	1319
LFA	17	17	17
SO4	10	10	10
Water	16	16	19
SEC	2	7	7
B-Factors			
Protein	100.35	75.55	54.61
LFA	86.03	80.95	49.6
SO4	103.86	83.5	56.38
Water	80.47	65.49	33.78
SEC	199.22	168.95	117.46
RMSD			
Bond lengths (Å)	0.012	0.015	0.012
Bond angles (°)	1.693	2.051	1.816
Ramachandran plot (%)			
Favored region	0.935	0.935	0.953
Allowed region	0.065	0.065	0.047
Outlier region	0	0	0

* values in parantheses are for high resolution shell

Table S4. Crystallography data collection and refinement statistics for installation of selenocysteine (SEC) on G34C, N48C, and N182C.

	N48C-MBB	N182C-MBB
Data Collection		
Light Source	Synchrotron	Synchrotron
Wavelength (Å)	1.0	1.0
Spacegroup	*P 6 2 2*	*P 6 2 2*
Cell dimensions		
a, b, c (Å)	180.25, 180.25, 50.97	178.24, 178.24, 50.60
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	39.03-2.75 (2.90-2.75)*	38.59-2.48 (2.61-2.48)*
Mesaured reflections	253774	339182
Unique reflections	13196	17262
Completeness (%)	100.0 (100.0)	99.6 (97.7)
Redundancy	19.2 (17.9)	19.6 (14.3)
Rpim	0.060 (0.728)	0.029 (0.711)
$cc_{1/2}$	0.998 (0.678)	0.999 (0.772)
$I/\sigma(I)$	11.4 (1.0)	22.4 (1.0)
Refinement		
Rwork/Rfree	0.2114/0.2457	0.2175/0.2448
No. Atoms		
Protein	1319	1319
LFA	17	17
SO4	10	10
Water	18	23
MBB	14	14
B-Factors		
Protein	80.3	72.36
LFA	75.05	71.72
SO4	86.87	65.39
Water	63.13	65.72
MBB	171.48	112.63
RMSD		
Bond lengths (Å)	0.014	0.015
Bond angles (°)	1.898	2.013
Ramachandran plot (%)		
Favored region	0.953	0.941
Allowed region	0.047	0.059
Outlier region	0	0

* values in parantheses are for high resolution shell

Table S5. Crystallography data collection and refinement statistics for installation of monobromobimane (MBB) on N48C and N182C.

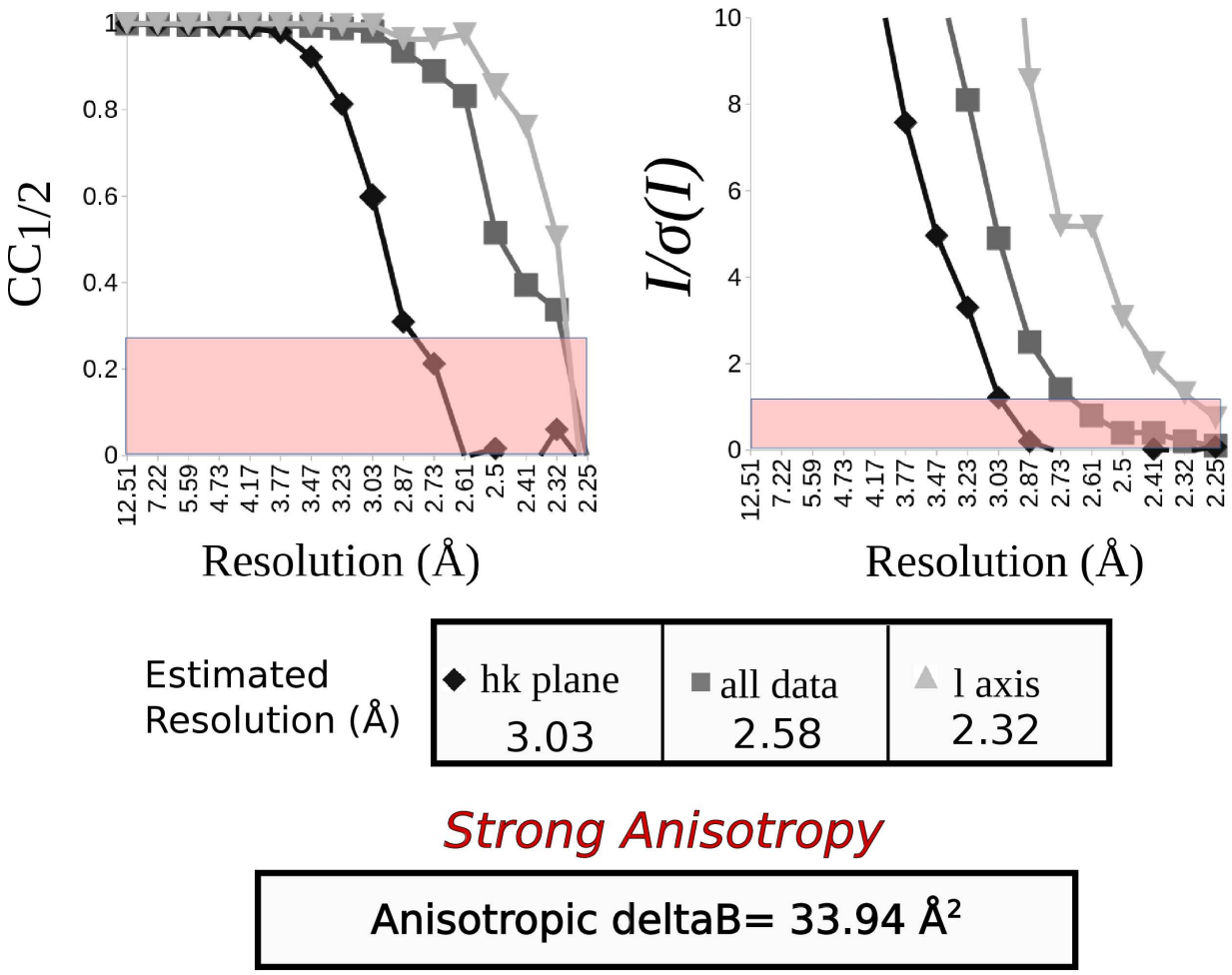

Figure S14. Resolution estimates were determined by identifying the highest resolution shell that satisfied the following criteria: I/σ(I) > 1, completeness > 90%, and $cc_{1/2}$ > 0.3. The resulting data refinement statistics revealed R_{merge} values in the high-resolution shell that were mathematically nonsensical (i.e. >100%). Most likely, this can be attributed to the high anisotropy of the datasets. The data scaling program *aimless* provides anisotropic data refinement statistics.[4] For the P622 symmetry of the CJ crystals, the data were separated for analysis of the hk plane and l axis. Data for CJ WT were scaled using aimless and the decoupling revealed strong anisotropy. Anisotropy is not surprising, given the unusually large solvent nanopores (13 nm diameter) that run through these crystals parallel to the c axis.

Applying the same rules of thumb for resolution estimates, the cutoff for the hk plane and l axis are 3.03 Å and 2.32 Å respectively. For this dataset, the R_{merge} for the full dataset would be an acceptable 66.7% if we were to apply the hk plane derived 3.03 Å high-resolution cutoff for the full dataset. However, such a conservative approach would leave valid and useful data unused. At 3.03 Å, the l-axis data is of very high quality: $cc_{1/2}$>0.996 and I/σ(I)=16.17. Furthermore, Diedrichs and Karplus have demonstrated

that there is no harm to the model R values when weaker higher resolution data is preserved, and that $cc_{1/2}$ is a better indicator for the resulting model quality than R_{merge}.[5,6] Thus, we decided to keep the high-resolution data, rather than maintain traditional R_{merge} values. The resolution cutoff for all but two datasets were determined by the $I/\sigma(I) > 1$ rule of thumb (the others, N48C-SEC and N182C-MBO were determined by completeness >90%) and we maintained very high $cc_{1/2}$ for all high-resolution shells.

Structure refinement programs such as *Refmac*, handle anisotropy in the data to some degree. We were able to obtain acceptable refinement statistics with anisotropic corrections from *Refmac*. However, structure refinement may get "stuck" with highly anistropic data, resulting in high R values and poorly resolved side chains. Because most data refinement programs use circular resolution shells, a crystallographer might be tempted to discard higher resolution data to ensure suitable refinement statistics. Recently Strong and coworkers developed techniques for ellipsoidal scaling, which can keep strong anisotropic data, while truncating weak data in other directions at similar resolutions.[7] If future anisotropic data from CJ structures results in poorly resolved guest molecules or high R factors, we will consider ellipsoidal scaling as an alternative approach.

Protein	Adduct	e/Å^3 equivalent to COOT rmsd 0.8	e/Å^3 equivalent to COOT rmsd 1.0	e/Å^3 equivalent to COOT rmsd 3.0 for Polder map	e/Å^3 equivalent to COOT rmsd 0.8 for Discovery map	e/Å^3 equivalent to COOT rmsd 0.8 for restrained simulated annealing omit map
WT		0.0872	0.1089	NA	NA	NA
N48C		0.0762	0.0952	NA	NA	NA
N48C	MBO	0.0824	0.1031	0.1683	0.0727	0.0776
N48C	MNB	0.0777	0.971	0.1583	0.0699	0.0769
N48C	SEC	0.0833	0.1042	0.1526	0.0685	0.0801
N48C	MBB	0.0785	0.098	0.1592	0.0863	0.0749
G34C		0.0721	0.09	NA	NA	NA
G34C	MBO	0.0713	0.0891	0.1168	0.0793	0.0653
G34C	MNB	0.066	0.0825	0.1618	0.0776	0.0616
G34C	SEC	0.066	0.0825	0.1145	0.0824	0.0603
G34C	MBB	0.0743	0.0928	NA	NA	NA
N182C		0.0784	0.098	NA	NA	NA
N182C	MBO	0.0846	0.1058	0.1583	0.0837	0.0817
N182C	MNB	0.1041	0.13	0.2013	0.1076	0.0989
N182C	SEC	0.1009	0.1262	0.189	0.1071	0.0979
N182C	MBB	0.0855	0.1068	0.1532	0.0854	0.0833

Table S6. To most faithfully represent COOT contours in PyMOL images, we turned off the PyMOL automatic map normalization ("set normalize_ccp4_maps, off") and instead directly set contour values to match the intended COOT contour level. This Table tracks the e/Å^3 values reported by COOT, when contours were set to specific rmsd values. The first two columns are for the primary .mtz file. The Polder map column corresponds to the values found for Polder omit maps (mF_o-DF_c), contoured to 3 rmsd in COOT. The penultimate column has the values found for the rigid body discovery maps ($2mF_o$-DF_c), contoured to 0.8 rmsd in COOT. The final column has the values found for the simulated annealing omit maps ($2mF_o$-DF_c), contoured to 0.8 rmsd in COOT. Unlike the other maps, the FEM approach attempts to eliminate the need to choose (arbitrary) map contouring cutoffs. In PyMOL, FEM maps were always contoured with a cutoff of 1.0.

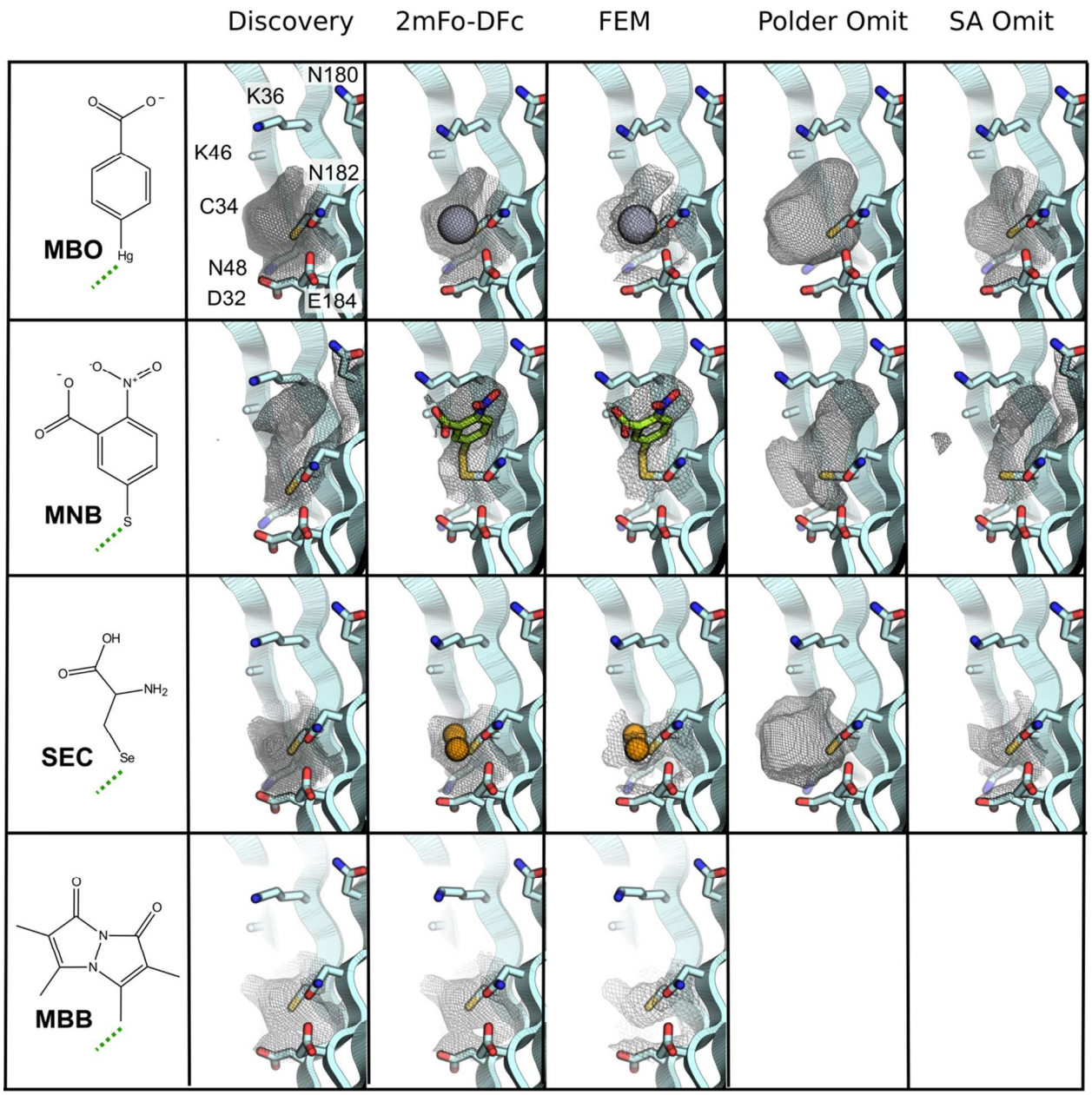

Figure S15. Side by side comparison of maps at the G34C installation site, including the Discovery map (contoured at 0.8 σ), the standard $2mF_o – DF_c$ map (contoured at 0.8 σ), the FEM map (contoured at 1.0), and two different maps generated with all adduct atoms omitted, specifically the Polder omit map (contoured at 3.0 σ) and a simulated annealing omit map with harmonic restraints (contoured at 0.8 σ). In all cases, the illustrated protein coordinates are the same.

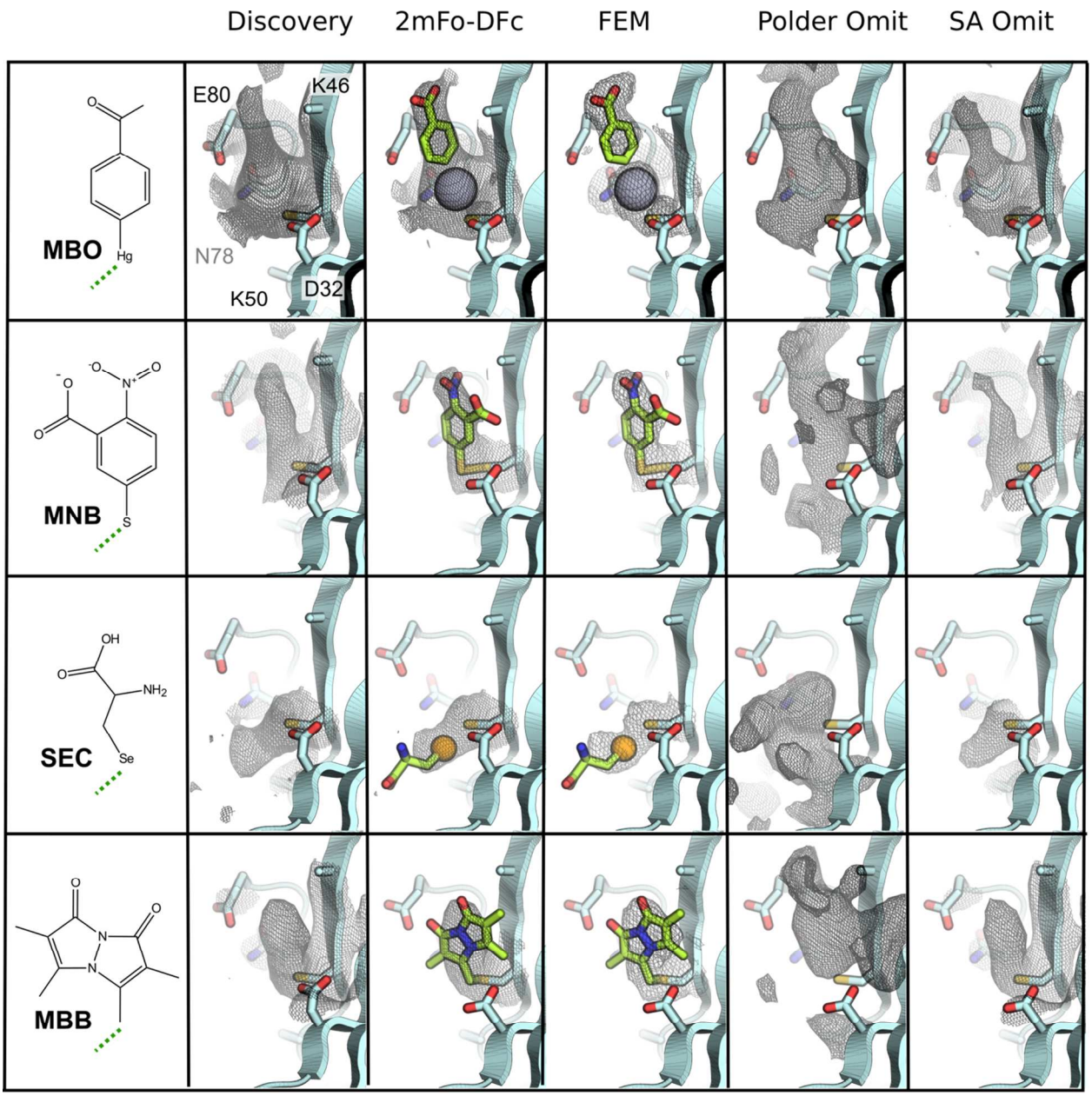

Figure S16. Side by side comparison of maps at the N48C installation site, including the Discovery map (contoured at 0.8 σ), the standard $2mF_o – DF_c$ map (contoured at 0.8 σ), the FEM map (contoured at 1.0), and two different maps generated with all adduct atoms omitted, specifically the Polder omit map (contoured at 3.0 σ) and a simulated annealing omit map with harmonic restraints (contoured at 0.8 σ). In all cases, the illustrated protein coordinates are the same.

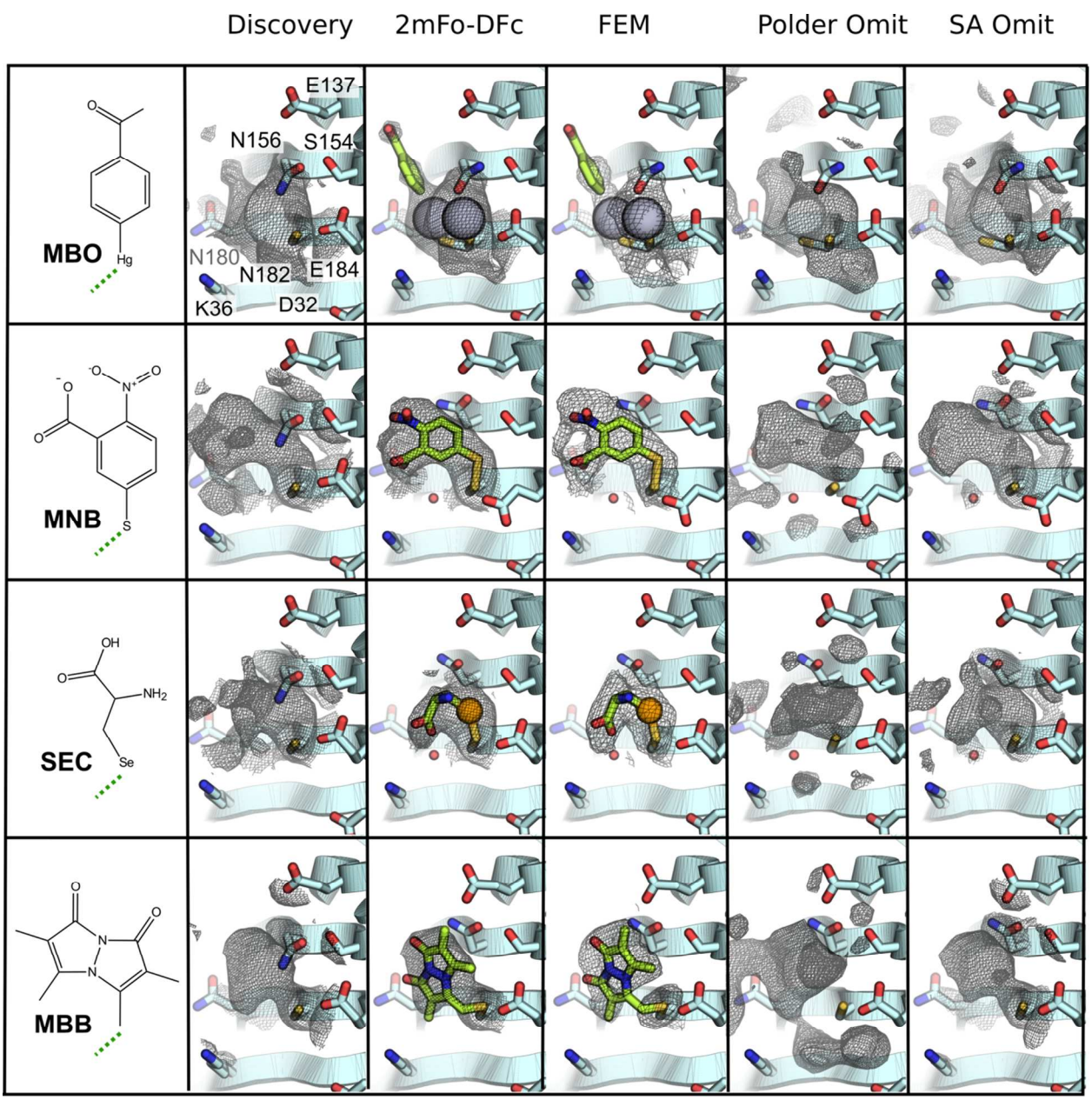

Figure S17. Side by side comparison of maps at the N182C installation site, including the Discovery map (contoured at 0.8 σ), the standard $2mF_o - DF_c$ map (contoured at 0.8 σ), the FEM map (contoured at 1.0), and two different maps generated with all adduct atoms omitted, specifically the Polder omit map (contoured at 3.0 σ) and a simulated annealing omit map with harmonic restraints (contoured at 0.8 σ). In all cases, the illustrated protein coordinates are the same.

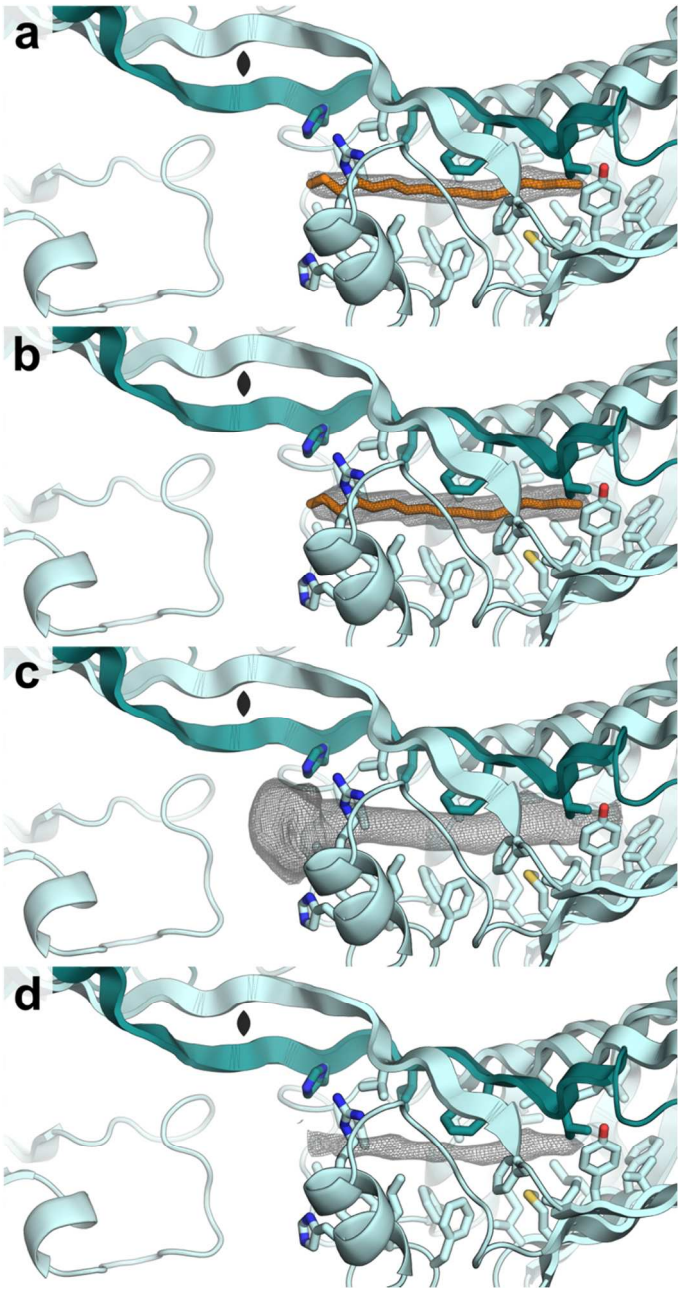

Figure S18. Comparison of WT maps at the internal ligand site, including **(a)** the standard $2mF_o - DF_c$ map (contoured at 0.8 σ), **(b)** the FEM map (contoured at 1.0), and two different maps generated with all adduct atoms omitted, specifically **(c)** the Polder omit map (contoured at 3.0 σ) and **(d)** a simulated annealing omit map with harmonic restraints (contoured at 0.8 σ). In all cases, the illustrated protein coordinates are the same. In this instance, the $2mF_o - DF_c$ map lacks interpretable density near the binding site opening. The Polder map provides a better illustration of the likely placement for a negatively charged ligand head group in the vicinity of His39, Arg84, and His87. A 2-fold symmetry axis relating the halves of the domain swapped dimer (light blue and dark blue chains) is indicated with the curved diamond shape (black).

Set of Surface Residues for Bfactor Comparision			
MET21	VAL55	THR103	ASN156
LYS22	ASP57	THR105	LYS158
GLU23	ASP59	LYS107	LYS160
THR25	SER62	LYS108	SER162
ASP27	GLU64	LYS116	ASP163
LYS28	LYS66	THR120	LYS165
ALA29	LYS67	THR122	THR168
HIS30	ASP69	THR124	THR172
ASP32	THR71	SER129	ASP178
GLY34	LYS73	LYS130	ASN180
LYS36	ASN78	ASP131	ASN182
LYS38	GLU80	ASP135	GLU184
GLN41	ASN81	GLU137	GLU186
SER43	GLN82	ALA142	ASN188
ASN44	THR83	LYS143	GLU189
LYS46	ASN86	LYS145	LYS190
ASN48	GLN90	ASP146	GLU191
LYS50	ASP91	LYS148	
ASP51	ASP92	LYS150	
SER53	LYS97	SER154	

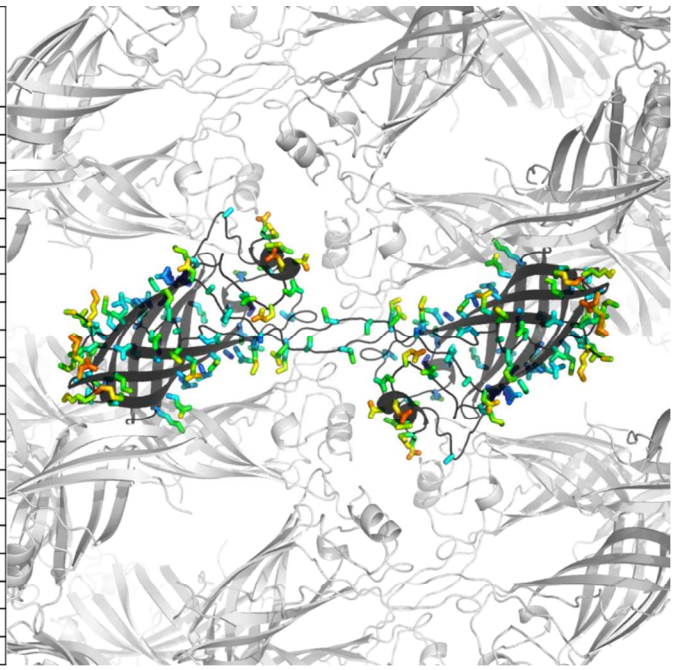

Figure S19. A set of solvent exposed surface residues was manually identified for B-factor comparison to installed guest molecules. A representative image of the selected residues is displayed with B-factor weighted coloring (blue coloring represents relative low B-factor and red coloring represents relative high B-factor).

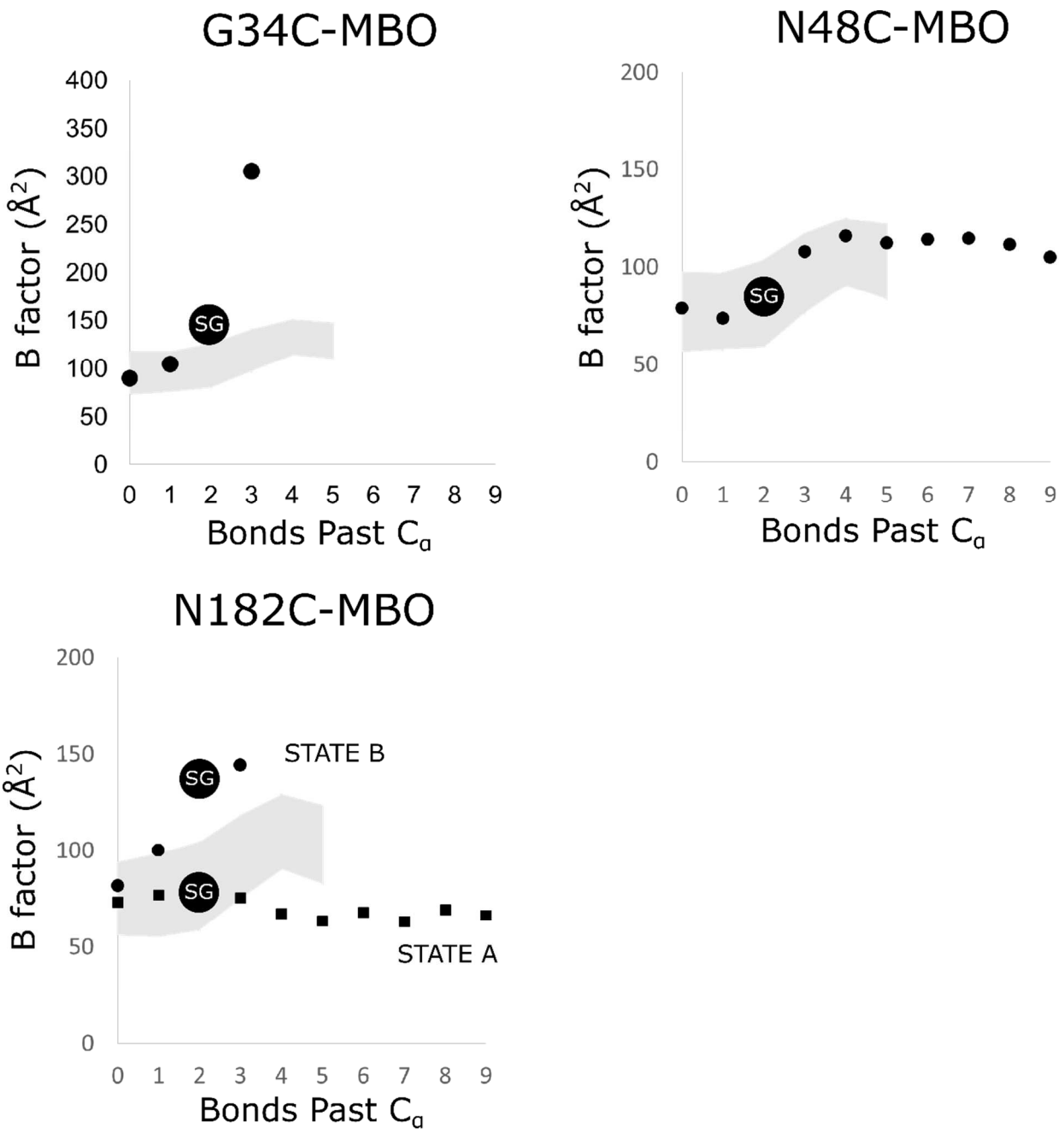

Figure S20. To examine the flexibility of MBO conjugation products, the average B-factor at the attachment site was plotted as a function of bond count from the alpha-carbon (dark circles). The attachment atom is clearly marked (SG). For reference, the mean B-factor (+/- one standard deviation) for a set of solvent exposed amino acid residues (Figure S19) is illustrated (filled light grey).

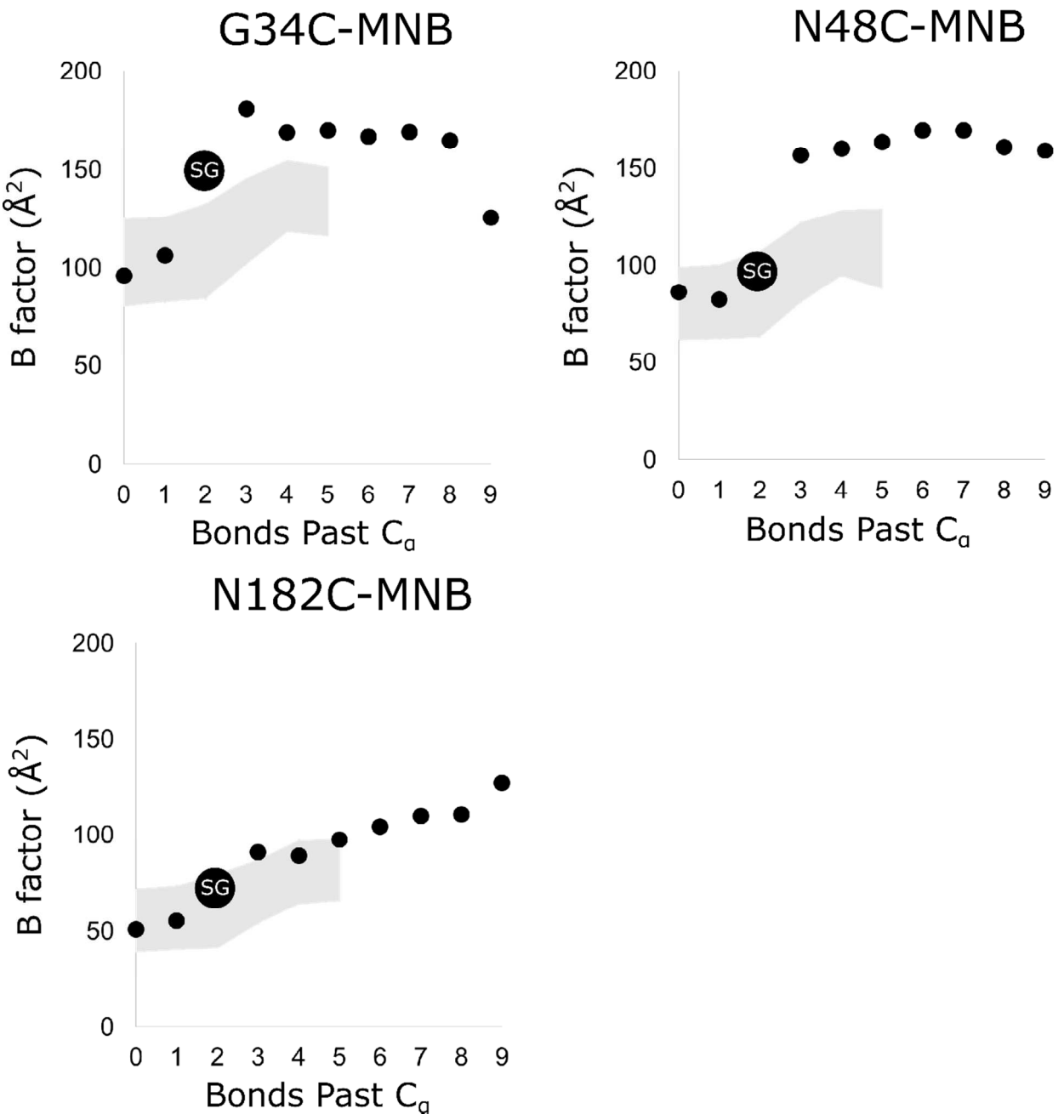

Figure S21. To examine the flexibility of MNB conjugation products the average B-factor at the attachment site was plotted as a function of bond count from the alpha-carbon (dark circles). The attachment atom is clearly marked (SG). For reference, the mean B-factor (+/- one standard deviation) for a set of solvent exposed amino acid residues (Figure S19) is illustrated (filled light grey).

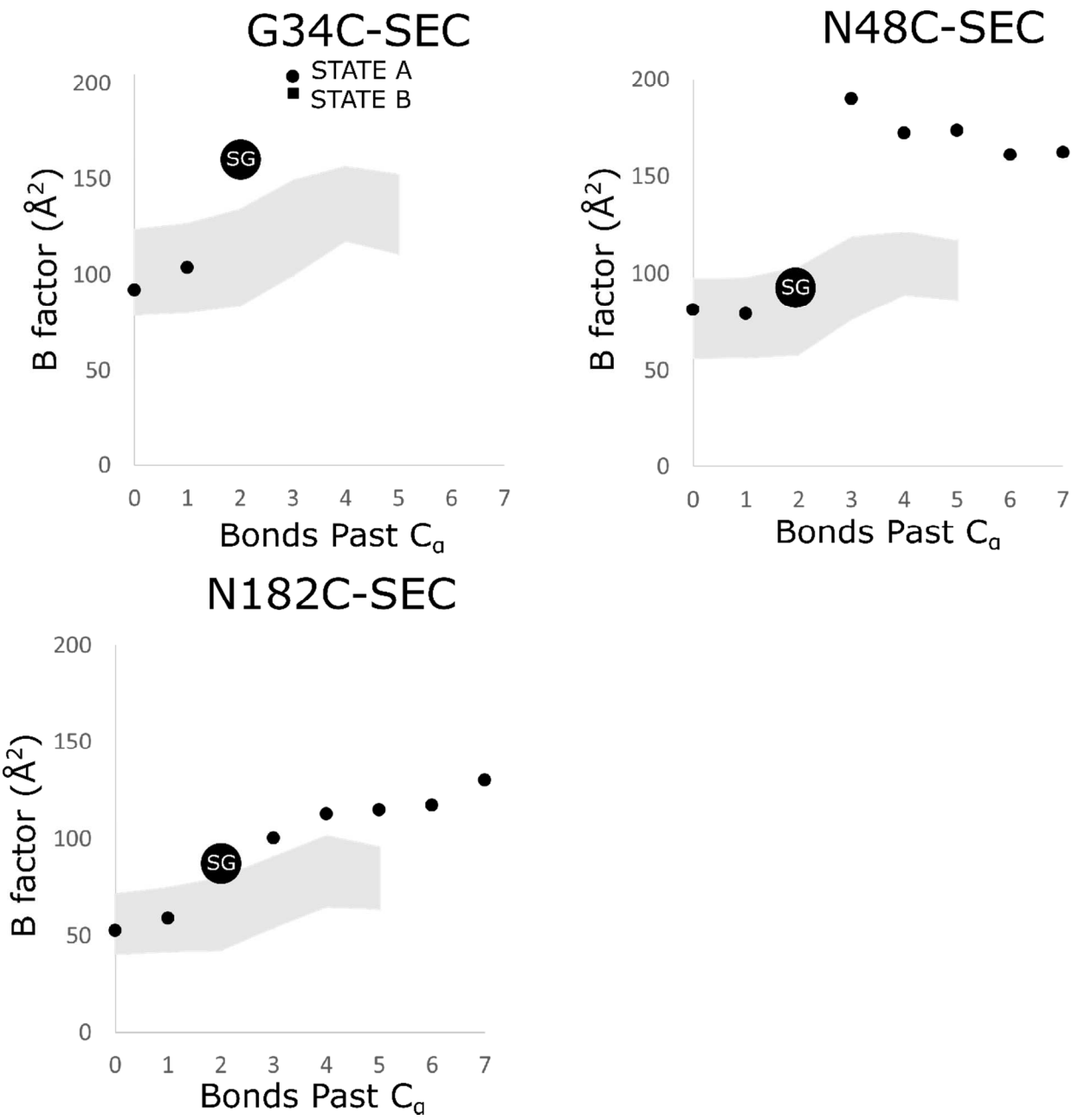

Figure S22. To examine the flexibility of SEC conjugation products, the average B-factor at the attachment site was plotted as a function of bond count from the alpha-carbon (dark circles). The attachment atom is clearly marked (SG). For reference, the mean B-factor (+/- one standard deviation) for a set of solvent exposed amino acid residues (Figure S19) is illustrated (filled light grey).

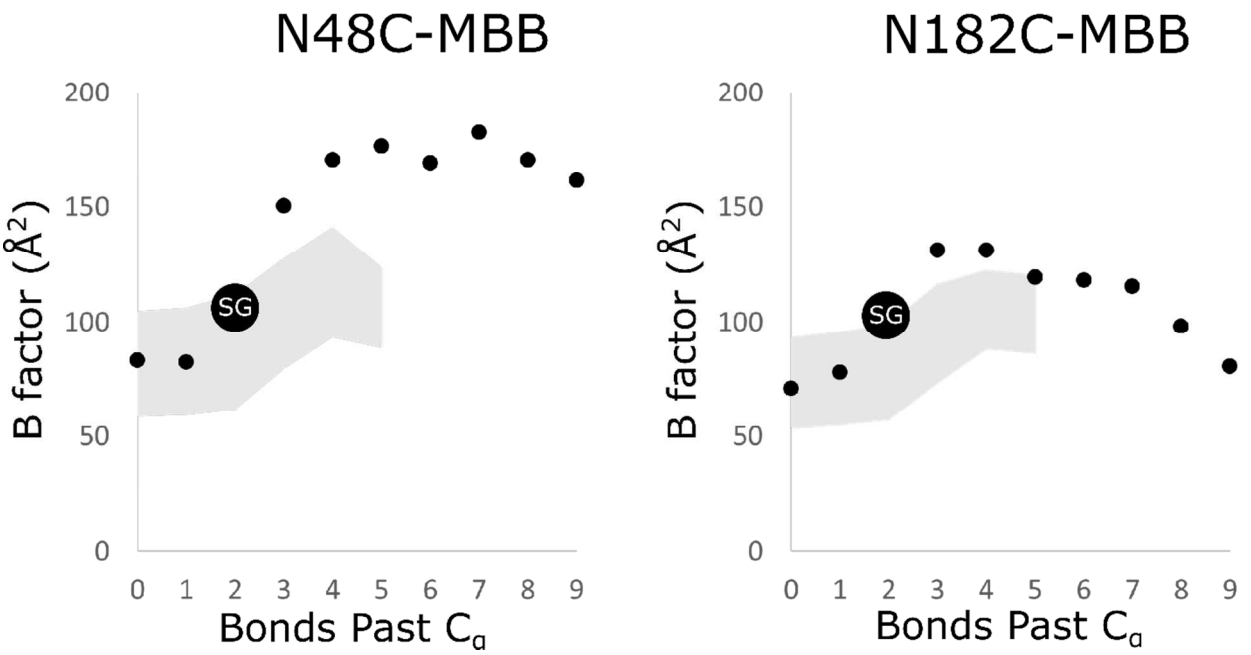

Figure S23. To examine the flexibility of MBB conjugation products, the average B-factor at the attachment site was plotted as a function of bond count from the alpha-carbon (dark circles). The attachment atom is clearly marked (SG). For reference, the mean B-factor (+/- one standard deviation) for a set of solvent exposed amino acid residues (Figure S19) is illustrated (filled light grey).

References

(1) Afonine, P. V., Moriarty, N. W., Mustyakimov, M., Sobolev, O. V., Terwilliger, T. C., Turk, D., Urzhumtsev, A., and Adams, P. D. (2015) FEM: feature-enhanced map. *Acta Crystallogr. D Biol. Crystallogr. 71*, 646–666.
(2) Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., and Adams, P. D. (2017) Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. Struct. Biol. 73*, 148–157.
(3) Hodel, A., Kim, S.-H., and Brünger, A. T. (1992) Model bias in macromolecular crystal structures. *Acta Crystallogr. A 48*, 851–858.
(4) Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr. 69*, 1204–1214.
(5) Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science 336*, 1030–1033.
(6) Diederichs, K., and Karplus, P. A. (2013) Better models by discarding data? *Acta Crystallogr. D Biol. Crystallogr. 69*, 1215–1222.
(7) Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006) Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. 103*, 8060–8065.