Chemo- and Regioselective Lysine Modification on Native Proteins

Maria J. Matos, Bruno L. Oliveira, Nuria Martínez-Sáez, Ana Guerreiro, Pedro M. S. D. Cal, Jean Bertoldo, María Maneiro, Elizabeth Perkins, Julie Howard, Michael J. Deery, Justin M. Chalker, Francisco Corzana, Gonzalo Jiménez-Osés* & Gonçalo J. L. Bernardes*

Table of contents

1. Computational methods and data

Quantum Mechanical calculations

Full geometry optimizations and transition structure (TS) searches were carried out with the Gaussian 09 package[1] using the M06-2X hybrid functional[2] and 6-31+G(d,p) basis set. Bulk solvent effects in water were considered implicitly through the IEF-PCM integral equation formalism polarizable continuum model.[3] The possibility of different conformations was taken into account for all structures. Frequency analyses were carried out at the same level used in the geometry optimizations. Thermal and entropic corrections to energy were calculated from vibrational frequencies. The nature of the stationary points was determined in each case according to the appropriate number of negative eigenvalues of the Hessian matrix. Scaled frequencies were not considered. The quasiharmonic approximation reported by Truhlar *et al.* was used to replace the harmonic oscillator approximation for the calculation of the vibrational contribution to enthalpy and entropy.[4] Scaled frequencies were not considered. Mass-weighted intrinsic reaction coordinate (IRC) calculations were carried out by using the Gonzalez and Schlegel scheme[5,6] in order to ensure that the TSs indeed connected the appropriate reactants and products. Gibbs free energies (ΔG) were used for the discussion on the relative stabilities of the considered structures. Free energies calculated using the gas phase standard state concentration (1 atm = 1/24.5 M) were converted to reproduce the standard state concentration in solution (1 M) by adding or subtracting 1.89 kcal mol^{-1} for bimolecular additions and decompositions, respectively. The lowest energy conformer for each calculated stationary point (**Supporting Figure 1**) was considered in the

discussion; all the computed structures can be obtained from authors upon request.
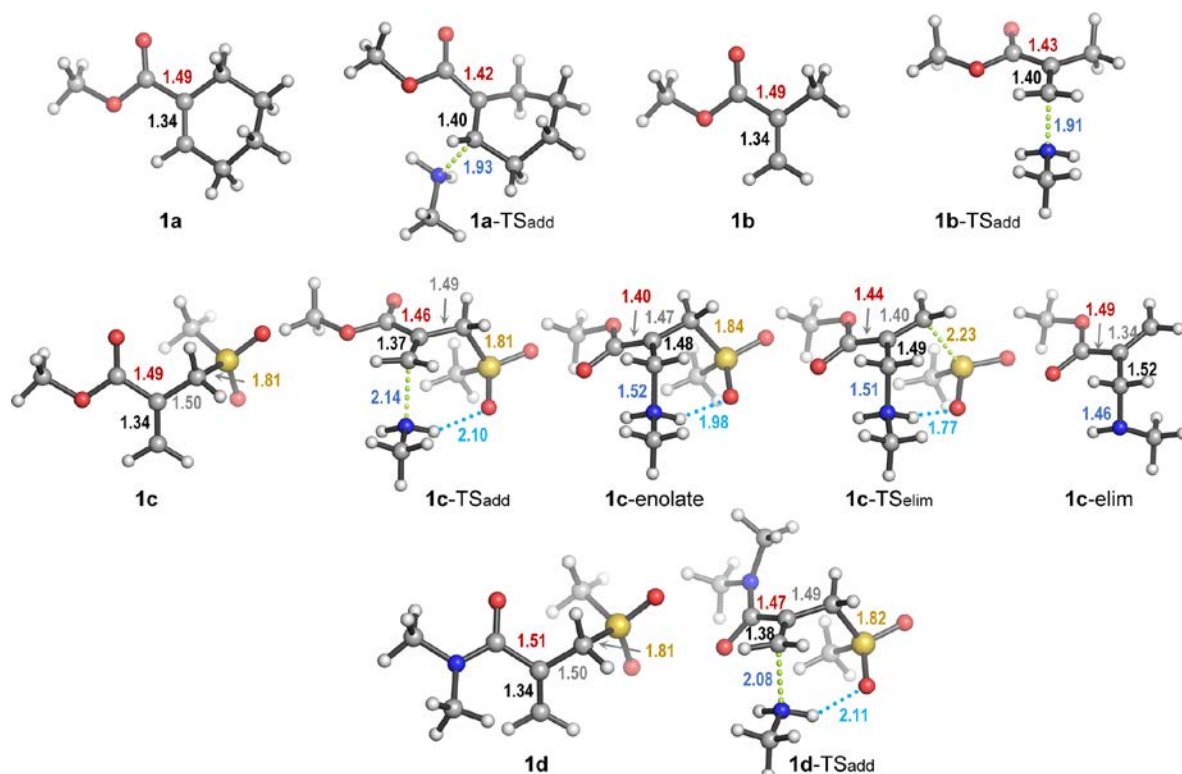
Supporting Figure 1. Guide to compound numbering of structures calculated with PCM(H_2O)/M06-2X/6-31+G(d,p) (only the lowest energy conformers are shown). Cartesian coordinates have been inverted with respect to those shown in **Figure 2** for clarity.
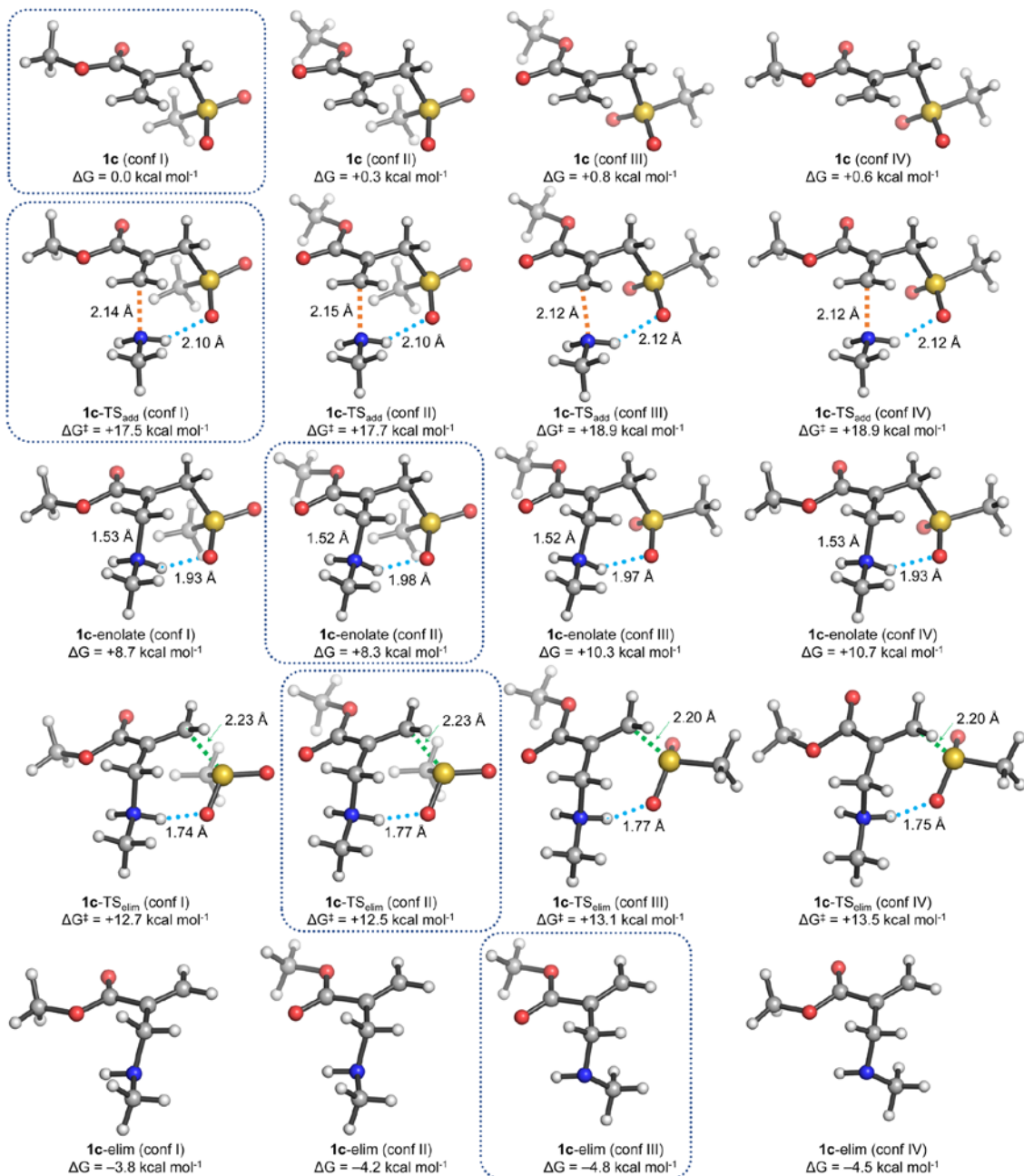
Cartesian coordinates, electronic energies, entropies, enthalpies, Gibbs free energies, and lowest frequencies of the calculated structures are summarized in the following table.

Supporting Table 1. Energies, entropies, and lowest frequencies of the lowest energy calculated structures discussed in the manuscript.[a]
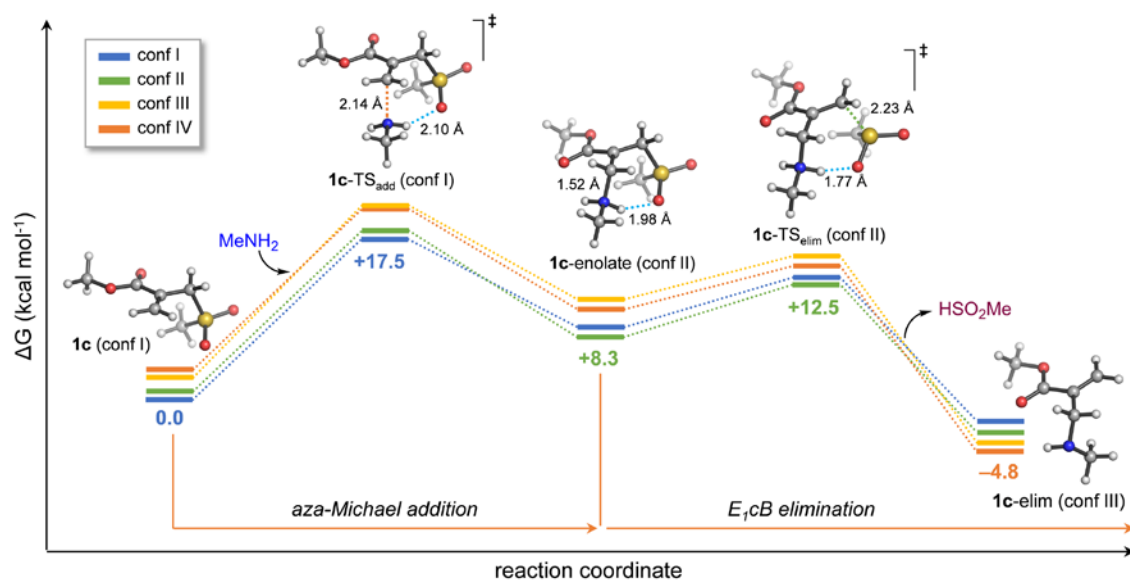
Structure	E_{elec} (Hartree)	E_{elec} + ZPE (Hartree)	H (Hartree)	S (cal mol^{-1} K^{-1})	G (Hartree)	Lowest freq. (cm^{-1})	# of imag freq.
MeNH2	–95.815844	–95.751237	–95.746870	57.4	–95.774165	302.4	0
1a	–462.354967	–462.164130	–462.153344	97.5	–462.198872	48.9	0
1a-TSadd	–558.142546	–557.884681	–557.870408	114.8	–557.924953	–337.7	1
1b	–345.656374	–345.531857	–345.522941	87.0	–345.563773	58.7	0
1b-TSadd	–441.451992	–441.260358	–441.247824	106.8	–441.297309	–349.8	1
1c (conf I)	-933.456980	-933.293216	-933.279950	110.9	-933.330768	39.4	
1c (conf II)	-933.456305	-933.292700	-933.279402	111.2	-933.330226	36.8	
1c (conf III)	-933.455493	-933.291836	-933.278481	111.6	-933.329497	40.6	
1c (conf IV)	-933.455948	-933.292206	-933.278861	112.1	-933.329791	31.9	
1c-TSadd (conf I)	-1029.268686	-1029.038188	-1029.021313	128.6	-1029.080022	-241.8	
1c-TSadd (conf II)	-1029.268298	-1029.037986	-1029.021037	130.0	-1029.079744	-240.7	
1c-TSadd (conf III)	-1029.266219	-1029.035902	-1029.018881	130.7	-1029.077754	-247.5	
1c-TSadd (conf IV)	-1029.266464	-1029.035932	-1029.019035	129.4	-1029.077755	-246.8	
1c-enolate (conf I)	-1029.287218	-1029.052871	-1029.036500	124.4	-1029.094124	48.8	
1c-enolate (conf II)	-1029.287863	-1029.053465	-1029.037116	124.2	-1029.094678	53.2	
1c-enolate (conf III)	-1029.284934	-1029.050336	-1029.034049	124.0	-1029.091510	47.3	
1c-enolate (conf IV)	-1029.284094	-1029.049731	-1029.033390	124.5	-1029.090872	40.1	
1c-TSelim (conf I)	-1029.278669	-1029.046321	-1029.029929	126.6	-1029.087773	-241.8	
1c-TSelim (conf II)	-1029.279478	-1029.046803	-1029.030511	125.6	-1029.088100	-244.7	
1c-TSelim (conf III)	-1029.278514	-1029.045781	-1029.029509	126.2	-1029.087010	-233.7	
1c-TSelim (conf IV)	-1029.277459	-1029.045013	-1029.028679	126.3	-1029.086357	-232.0	
1c-elim (conf I)	-440.273118	-440.101816	-440.090528	99.9	-440.137173	51.8	
1c-elim (conf II)	-440.273596	-440.102325	-440.090998	100.2	-440.137713	46.2	
1c-elim (conf III)	-440.274533	-440.103215	-440.091949	100.2	-440.138244	43.3	
1c-elim (conf IV)	-440.274893	-440.103560	-440.092276	100.2	-933.330768	39.4	
1d	–952.884761	–952.680345	–952.665451	117.5	–952.719654	42.2	0
1d-TSadd	–1048.689543	–1048.418309	–1048.400008	133.8	–1048.461350	–266.1	1
HSO2Me	–589.002931	–588.946092	–588.940008	71.3	–588.973866	216.0	0

[a]Energy values calculated at the PCM(H$_2$O)/M06-2X/6-31+G(d,p) level. 1 Hartree = 627.51 kcal $^{-1}$ mol^{-1}. Thermal corrections at 298.15 K.

1c (conf I)
ΔG = 0.0 kcal mol⁻¹

1c (conf II)
ΔG = +0.3 kcal mol⁻¹

1c (conf III)
ΔG = +0.8 kcal mol⁻¹

1c (conf IV)
ΔG = +0.6 kcal mol⁻¹

1c-TS$_{add}$ (conf I)
ΔG‡ = +17.5 kcal mol⁻¹

1c-TS$_{add}$ (conf II)
ΔG‡ = +17.7 kcal mol⁻¹

1c-TS$_{add}$ (conf III)
ΔG‡ = +18.9 kcal mol⁻¹

1c-TS$_{add}$ (conf IV)
ΔG‡ = +18.9 kcal mol⁻¹

1c-enolate (conf I)
ΔG = +8.7 kcal mol⁻¹

1c-enolate (conf II)
ΔG = +8.3 kcal mol⁻¹

1c-enolate (conf III)
ΔG = +10.3 kcal mol⁻¹

1c-enolate (conf IV)
ΔG = +10.7 kcal mol⁻¹

1c-TS$_{elim}$ (conf I)
ΔG‡ = +12.7 kcal mol⁻¹

1c-TS$_{elim}$ (conf II)
ΔG‡ = +12.5 kcal mol⁻¹

1c-TS$_{elim}$ (conf III)
ΔG‡ = +13.1 kcal mol⁻¹

1c-TS$_{elim}$ (conf IV)
ΔG‡ = +13.5 kcal mol⁻¹

1c-elim (conf I)
ΔG = −3.8 kcal mol⁻¹

1c-elim (conf II)
ΔG = −4.2 kcal mol⁻¹

1c-elim (conf III)
ΔG = −4.8 kcal mol⁻¹

**1c**-elim (conf IV)
ΔG = −4.5 kcal mol⁻¹

5

Supporting Figure 2. Four different conformers (top) and Minimum Energy Pathway (bottom) calculated with PCM(H_2O)/M06-2X/6-31+G(d,p) for the whole reaction pathway (aza-Michael addition followed by sulfone elimination) between sulfonyl acrylate **1c** and methylamine (model for lysine sidechain). Rotation around the C–C ester bond (conformers I and II) has virtually no effect on the activation barriers ($\Delta G^{\ddagger}$), while rotation around the C–S sulfone bond (conformers III and IV) slightly increases $\Delta G^{\ddagger}$. Cartesian coordinates have been inverted with respect to those shown in **Figure 2** for clarity.
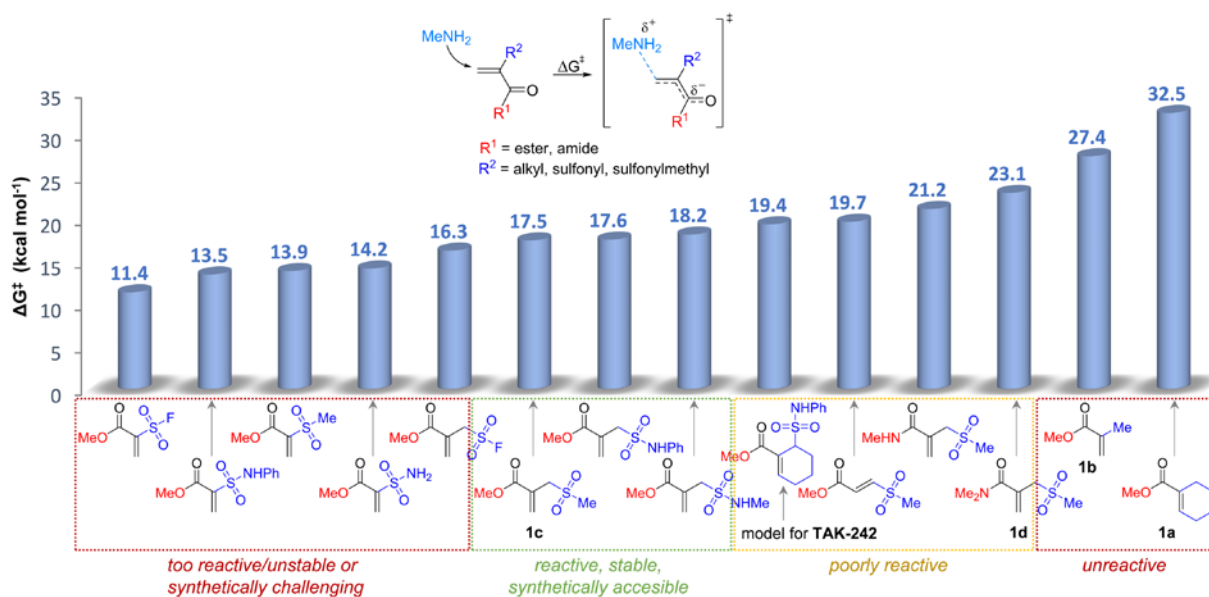
Supporting Figure 3. Computational screening of acrylic acid derivatives as potential warheads for lysine conjugation. Activation barriers ($\Delta G^{\ddagger}$) were calculated with PCM(H_2O)/M06-2X/6-31+G(d,p). Sulfonyl acrylates were predicted to be exceedingly reactive ($\Delta G^{\ddagger} \sim$ 11-14 kcal mol^{-1}), as verified experimentally by their high instability in aqueous solution. Alkyl acrylates were predicted (and then verified experimentally) to be too unreactive towards aza-Michael addition ($\Delta G^{\ddagger} \sim$ 27–33 kcal mol^{-1}). Sulfonylmethyl acrylates were predicted to have the adequate reactivity ($\Delta G^{\ddagger} \sim$ 16-21 kcal mol^{-1}) towards aza-Michael addition, while being synthetically accessible and stable enough in aqueous solution. Among these, sulfonylmethyl acrylate **1c** was predicted to have a superior reactivity compared to its amide analogues.
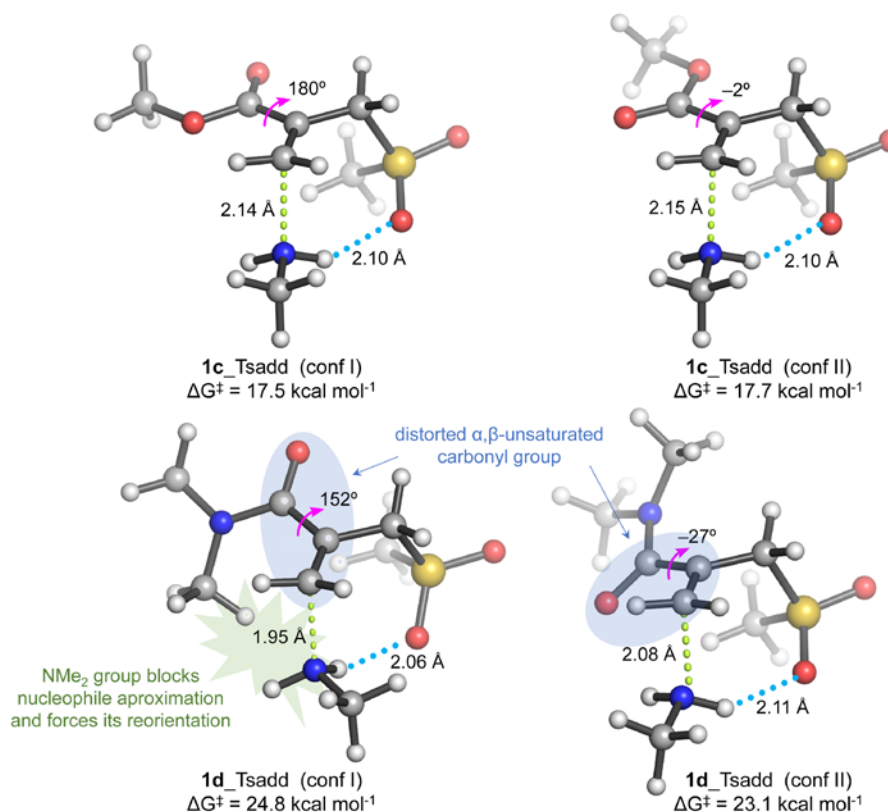
Supporting Figure 4. Structural differences between the transition states calculated with PCM(H_2O)/M06-2X/6-31+G(d,p) for methylamine (model for lysine sidechain) aza-Michael addition to sulfonyl acrylate **1c** (top) and sulfonyl acrylamide **1d** (bottom). Two different carbonyl rotamers: s-trans (left) and s-cis (right) are shown. The much bulkier N,N-dimethylamide group in **1d** severely distorts the α,β-unsaturated carbonyl group deviating it for planarity (CCCO dihedral angles in degrees shown as magenta arrows), thus reducing the electrophilic character of the Michael acceptor as reflected by the higher calculated activation barriers (ΔG‡) . Additionally, the NMe2 group blocks the nucleophilic attack of methylamine through steric hindrance, further increasing ΔG‡ in certain conformations. Cartesian coordinates have been inverted with respect to those shown in **Figure 2** for clarity.

Cartesian coordinates of the lowest energy structures calculated with PCM(H$_2$O)/M06-2X/6-31+G(d,p). All the calculated structures can be obtained from the authors upon request.

Structure **MeNH2**

N	-0.75080	-0.00000	-0.12570
H	-1.14450	0.81190	0.34080
C	0.70770	-0.00000	0.01730
H	1.11820	0.88160	-0.48080
H	1.06170	-0.00010	1.05650
H	1.11820	-0.88150	-0.48100
H	-1.14450	-0.81190	0.34080

Structure **1a**

C	0.98750	1.32250	0.03810
C	0.10360	0.09920	-0.00790
C	0.58440	-1.15220	-0.06830
C	2.05120	-1.47410	-0.05320
C	2.90350	-0.27420	0.36310
C	2.42730	0.98820	-0.35560
H	0.96000	1.75010	1.04950
H	2.22160	-2.32740	0.61280
H	2.34020	-1.81770	-1.05680
H	2.81560	-0.12510	1.44710
H	3.95800	-0.47180	0.14900
H	3.08020	1.83580	-0.12710
H	2.47820	0.82410	-1.44000
C	-1.36000	0.35640	0.02640
O	-1.84260	1.47050	0.12220
O	-2.11590	-0.74790	-0.05130
C	-3.53190	-0.54010	-0.01450
H	-3.84470	0.08120	-0.85560
H	-3.97550	-1.53080	-0.08580
H	-3.81890	-0.05810	0.92180
H	-0.11290	-1.98390	-0.12730
H	0.57240	2.08900	-0.62410

Structure **1a-TSadd**

C	0.77820	-1.61870	0.92490
C	-0.06550	-0.61530	0.18100
C	0.59770	0.48220	-0.38820
C	2.10710	0.36280	-0.44770
C	2.46740	-1.05220	-0.92760
C	1.95340	-2.12780	0.05430
H	0.15840	-2.45690	1.25310
H	2.53750	1.12370	-1.10430
H	2.53200	0.51510	0.55370
H	2.02280	-1.20150	-1.91770
H	3.55150	-1.13360	-1.05400
H	1.63340	-3.00820	-0.51290
H	2.76970	-2.45420	0.70860
C	-1.48090	-0.73670	0.12230
O	-2.17450	-1.67770	0.52780
O	-2.09070	0.36350	-0.44080
C	-3.50230	0.27820	-0.60040
H	-3.99790	0.15760	0.36580
H	-3.80750	1.21650	-1.06260
H	-3.77270	-0.56140	-1.24580
H	0.12150	0.99250	-1.22470
N	0.37140	2.04370	0.72330
H	-0.64500	2.10950	0.74840
C	0.99010	3.28120	0.23730
H	0.72200	3.41480	-0.81370
H	0.66440	4.15870	0.80170
H	2.07440	3.19030	0.31440

H	0.69120	1.79120	1.65670
H	1.19380	-1.16800	1.83990

Structure **1b**

C	2.24060	-0.76980	-0.00000
C	1.14550	0.25820	-0.00000
C	1.35280	1.57830	-0.00000
H	2.16580	-1.41570	-0.87970
H	0.52840	2.28230	-0.00000
C	-0.24300	-0.29580	0.00000
O	-0.48380	-1.48800	0.00010
O	-1.20120	0.63690	-0.00000
C	-2.54870	0.15070	0.00000
H	-2.73170	-0.45120	0.89190
H	-3.17970	1.03660	0.00010
H	-2.73180	-0.45110	-0.89180
H	3.21790	-0.28440	-0.00010
H	2.36230	1.97810	-0.00010
H	2.16590	-1.41570	0.87970

Structure **1b-TSadd**

C	0.09330	2.44850	0.06550
C	0.10210	0.98690	-0.30730
C	-1.03490	0.34310	-0.79940
H	0.93620	2.98900	-0.37980
C	1.28270	0.24280	-0.00470
O	2.34140	0.71020	0.43140
O	1.17630	-1.10670	-0.23280
C	2.35390	-1.87070	0.00160
H	2.67000	-1.79020	1.04450
H	2.09090	-2.90220	-0.23070
H	3.17080	-1.53850	-0.64390
H	-0.91750	-0.62510	-1.27900
N	-2.15930	-0.34760	0.57480
H	-1.56720	-0.99930	1.08600
C	-3.38020	-0.98760	0.07350
H	-3.09750	-1.79340	-0.60650
H	-3.99510	-1.39690	0.87830
H	-3.96080	-0.24760	-0.48020
H	-2.36380	0.43480	1.19320
H	-0.82850	2.92280	-0.28550
H	-1.81970	0.96180	-1.23290
H	0.15470	2.61110	1.15000

Structure **1c (conf I)**

C	0.67580	0.46460	-1.07020
C	-0.58410	0.91340	-0.39220
C	-0.73320	2.13880	0.11900
H	0.50390	-0.35770	-1.76730
C	-1.66690	-0.11350	-0.30390
O	-1.52250	-1.25920	-0.68440
O	-2.79320	0.34800	0.23300
C	-3.86440	-0.59960	0.35480
H	-3.55770	-1.42910	0.99370
H	-4.68800	-0.05100	0.80510
H	-4.14360	-0.97560	-0.63050
S	1.94340	-0.11760	0.07960
O	2.22960	0.96010	1.03830
O	3.05550	-0.63180	-0.73450
H	-1.65300	2.43160	0.61220
H	1.17580	1.28760	-1.58760
H	0.06840	2.86770	0.05290

```
C     1.20680  -1.47590   0.96260          C    -3.13180  -0.96130   0.52540
H     1.99160  -1.87300   1.60870          H    -3.59880  -0.28140   1.23820
H     0.37510  -1.09910   1.55870          H    -3.83110  -1.22440  -0.26990
H     0.87300  -2.22220   0.24190          H    -2.74910  -1.85660   1.01530
                                           O    -1.06260  -1.07060  -1.12770
Structure 1c (conf II)
C     0.53030   0.45280  -1.07330   Structure 1c-TSadd (conf I)
C    -0.58400   1.15410  -0.35300          C    -0.79660  -0.64800   1.25770
C    -0.50430   2.43410   0.02010          C     0.55310  -0.10520   0.92480
H     0.18860  -0.38150  -1.68720          C     0.88120   1.20940   1.13740
C    -1.83390   0.39810  -0.02330          H    -0.77560  -1.71880   1.47530
O    -2.80410   0.87710   0.52390          C     1.49240  -1.03930   0.31900
O    -1.75600  -0.88170  -0.39800          O     1.25940  -2.22540   0.11260
C    -2.89840  -1.69810  -0.10050          O     2.67690  -0.48350  -0.01040
H    -3.77960  -1.30250  -0.60730          C     3.64450  -1.36880  -0.57900
H    -2.65260  -2.69020  -0.47120          H     3.26290  -1.80830  -1.50260
H    -3.06880  -1.71830   0.97700          H     4.51960  -0.75510  -0.78370
S     1.79410  -0.22200   0.03150          H     3.89450  -2.16470   0.12530
O     2.29870   0.87070   0.87590          S    -2.01980  -0.47070  -0.06740
O     2.75630  -0.94450  -0.81480          O    -2.17430   0.96630  -0.37920
H    -1.32850   2.90280   0.54730          O    -3.23640  -1.20380   0.32710
H     1.10490   1.14590  -1.69420          H     1.91430   1.52670   1.06610
H     0.37860   3.02740  -0.19380          N     0.44020   2.47410  -0.53410
C     0.96240  -1.39800   1.07730          H     1.02840   2.17170  -1.30550
H     1.73770  -1.82820   1.71370          C     0.60360   3.89390  -0.23520
H     0.22170  -0.87330   1.68170          H     1.64230   4.08450   0.04580
H     0.49950  -2.16100   0.45180          H     0.34340   4.54820  -1.07430
                                           H    -0.03280   4.15170   0.61500
Structure 1c (conf III)                    H    -0.52420   2.21650  -0.73540
C    -0.52870   0.38240   1.09710          H    -1.27360  -0.11680   2.08760
C     0.61820   1.11670   0.46680          H     0.22510   1.83180   1.73930
C     0.61500   2.43960   0.28700          C    -1.31470  -1.26680  -1.49620
H    -0.20630  -0.50570   1.64320          H    -2.06940  -1.19680  -2.28140
C     1.80190   0.33330  -0.01130          H    -0.40740  -0.73410  -1.78290
O     2.66430   0.77580  -0.74110          H    -1.09720  -2.30410  -1.24320
O     1.81580  -0.90690   0.47830
C     2.90670  -1.73690   0.05680   Structure 1c-TSadd (conf II)
H     3.85440  -1.29700   0.37120          C     0.30180   0.71740  -1.28750
H     2.74820  -2.69710   0.54160          C     0.12350  -0.71300  -0.89610
H     2.89210  -1.84500  -1.02900          C    -1.06560  -1.37010  -1.07350
S    -1.64080  -0.18000  -0.21760          H     1.33210   0.97200  -1.54440
O    -2.41130   0.98040  -0.69000          C     1.21230  -1.43580  -0.24860
H     1.45590   2.92900  -0.19300          O     1.15420  -2.57940   0.18350
H    -1.12740   1.02950   1.74080          O     2.34550  -0.70230  -0.15890
H    -0.21750   3.05060   0.62040          C     3.45160  -1.33260   0.49100
C    -2.75390  -1.31150   0.58670          H     3.74840  -2.23340  -0.04940
H    -3.26590  -0.78550   1.39280          H     4.25620  -0.60000   0.48010
H    -3.46530  -1.63090  -0.17660          H     3.18810  -1.59500   1.51770
H    -2.18110  -2.16000   0.96060          S    -0.16640   1.90800  -0.00190
O    -0.84260  -0.91800  -1.20840          O    -1.59100   1.70160   0.33370
                                           O     0.22640   3.25370  -0.45800
Structure 1c (conf IV)                     H    -1.09890  -2.44590  -0.94540
C    -0.67990   0.34870   1.09530          N    -2.38730  -1.20010   0.61470
C     0.60750   0.86170   0.52190          H    -1.92490  -1.64280   1.40370
C     0.79060   2.14180   0.18710          C    -3.71150  -1.76050   0.36230
H    -0.52230  -0.54880   1.69690          H    -3.61320  -2.82430   0.13210
C     1.68140  -0.16370   0.34270          H    -4.40060  -1.64860   1.20640
O     1.60790  -1.29030   0.78970          H    -4.14660  -1.26240  -0.50780
O     2.73100   0.29040  -0.34070          H    -2.39940  -0.19330   0.76470
C     3.81400  -0.63500  -0.51170          H    -0.35490   1.00680  -2.11470
H     3.47240  -1.51110  -1.06480          H    -1.84320  -0.92590  -1.68770
H     4.57160  -0.09590  -1.07540          C     0.79550   1.49440   1.43950
H     4.20170  -0.94090   0.46120          H     0.51910   2.22520   2.20140
S    -1.77820  -0.11890  -0.26470          H     0.53280   0.48560   1.76070
O    -2.28910   1.11370  -0.88400          H     1.85290   1.56650   1.18610
H     1.72830   2.47990  -0.23870
H    -1.21370   1.10880   1.66860   Structure 1c-TSadd (conf III)
H     0.00390   2.87360   0.34110          C    -0.17840   0.79410   1.21230
```

```
C    -0.34670  -0.65060   0.87330
C     0.65900  -1.56170   1.08080
H    -1.12880   1.30270   1.38670
C    -1.55760  -1.10830   0.20600
O    -1.73400  -2.22980  -0.25590
O    -2.52500  -0.16580   0.15640
C    -3.73990  -0.54930  -0.48850
H    -4.19940  -1.39300   0.03030
H    -4.38810   0.32380  -0.44070
H    -3.54880  -0.82440  -1.52780
S     0.61020   1.68320  -0.16030
O     2.03300   1.28250  -0.21320
H     0.43840  -2.61860   0.97990
N     1.99690  -1.73130  -0.55090
H     1.44380  -2.01000  -1.35660
C     3.09650  -2.65490  -0.28500
H     2.68950  -3.64920  -0.08500
H     3.80990  -2.72950  -1.11250
H     3.63250  -2.31910   0.60610
H     2.30580  -0.76960  -0.68270
H     0.49090   0.95300   2.06250
H     1.49910  -1.29920   1.71740
C     0.54320   3.40520   0.29110
H     1.05580   3.53670   1.24430
H     1.05760   3.95160  -0.50110
H    -0.50170   3.70880   0.35480
O    -0.18090   1.48520  -1.38660

Structure 1c-TSadd (conf IV)
C    -0.79830  -0.75010   1.14820
C     0.56840  -0.20830   0.88850
C     0.89140   1.09970   1.15460
H    -0.79470  -1.84020   1.22880
C     1.52490  -1.12100   0.27940
O     1.34300  -2.31980   0.10140
O     2.68640  -0.52450  -0.07140
C     3.69090  -1.38530  -0.61080
H     3.33260  -1.86570  -1.52320
H     4.54220  -0.74330  -0.82960
H     3.96970  -2.15110   0.11600
S    -1.92430  -0.36330  -0.22060
O    -2.19730   1.09010  -0.19810
H     1.92860   1.41190   1.13240
N     0.51030   2.41370  -0.46120
H     1.03940   2.06290  -1.25480
C     0.82990   3.80390  -0.14850
H     1.89350   3.88290   0.08940
H     0.60240   4.49600  -0.96610
H     0.25900   4.10770   0.73240
H    -0.48170   2.26110  -0.63430
H    -1.27500  -0.30800   2.02760
H     0.22230   1.69620   1.76890
C    -3.44160  -1.21010   0.17170
H    -3.79970  -0.85430   1.13810
H    -4.14960  -0.95650  -0.61880
H    -3.24830  -2.28260   0.18810
O    -1.38440  -0.91740  -1.47400

Structure 1c-enolate (conf I)
C     0.94510  -0.45880  -1.29210
C    -0.45280  -0.13890  -0.95100
C    -0.94200   1.25720  -1.02120
H     1.10470  -1.51610  -1.52340
C    -1.25790  -1.13900  -0.38750
O    -0.98820  -2.34200  -0.21790
O    -2.48500  -0.64660   0.03840
C    -3.38910  -1.60630   0.57370
H    -2.96540  -2.09350   1.45500
```

```
H    -4.28560  -1.05120   0.84870
H    -3.63600  -2.36790  -0.17010
S     2.10590  -0.12010   0.09050
O     1.79080   1.23200   0.62880
O     3.50290  -0.34420  -0.33460
H    -1.96260   1.34860  -1.40240
N    -0.99700   1.90200   0.36490
H    -1.62150   1.31280   0.92520
C    -1.45550   3.31210   0.37780
H    -2.45180   3.35880  -0.06020
H    -1.47890   3.67430   1.40410
H    -0.75980   3.90570  -0.21400
H    -0.04800   1.80970   0.75890
H     1.37230   0.15520  -2.09380
H    -0.28130   1.89710  -1.61200
C     1.69610  -1.30270   1.36090
H     2.30930  -1.05070   2.22750
H     0.63310  -1.21230   1.58530
H     1.93000  -2.29940   0.98730

Structure 1c-enolate (conf II)
C    -0.81180  -0.24040   1.33100
C     0.60140  -0.02800   0.96790
C     1.19160   1.33320   0.98570
H    -1.06090  -1.27550   1.57880
C     1.37400  -1.00600   0.33420
O     2.49510  -0.80750  -0.19370
O     0.80870  -2.25840   0.28850
C     1.52740  -3.24170  -0.44720
H     2.50620  -3.43070  -0.00030
H     0.91970  -4.14540  -0.40720
H     1.66490  -2.92980  -1.48610
S    -1.95720   0.18910  -0.04130
O    -1.53010   1.50390  -0.59230
O    -3.36180   0.08690   0.40470
H     2.20720   1.36040   1.39330
N     1.34310   1.86380  -0.43480
H     1.89600   1.14500  -0.92150
C     1.98740   3.19380  -0.54610
H     2.98050   3.13810  -0.10160
H     2.06200   3.47010  -1.59650
H     1.37870   3.92070  -0.00920
H     0.39670   1.86740  -0.84140
H    -1.17600   0.41440   2.13150
H     0.57020   2.05700   1.51840
C    -1.67620  -1.02890  -1.31480
H    -2.31810  -0.74800  -2.15140
H    -0.62460  -0.99820  -1.60070
H    -1.94620  -2.00890  -0.92160

Structure 1c-enolate (conf III)
C     0.72060   0.51990   1.21860
C    -0.61480  -0.02290   0.90680
C    -0.87680  -1.48080   0.99340
H     0.72970   1.59920   1.38800
C    -1.62460   0.72770   0.29530
O    -2.68180   0.24550  -0.18080
O    -1.40640   2.08110   0.23640
C    -2.37800   2.83840  -0.47450
H    -3.36060   2.76220  -0.00280
H    -2.03050   3.87080  -0.44440
H    -2.45510   2.50120  -1.51160
S     1.86960   0.26510  -0.20090
O     1.75490  -1.16080  -0.60840
H    -1.84130  -1.72080   1.45260
N    -0.96540  -2.09140  -0.39920
H    -1.68740  -1.53530  -0.87760
C    -1.29340  -3.53570  -0.43240
```

```
H     -2.25140   -3.68730    0.06360
H     -1.35090   -3.86870   -1.46730
H     -0.51070   -4.08060    0.09450
H     -0.06000   -1.89900   -0.85130
H      1.22440    0.01020    2.04790
H     -0.08390   -2.02250    1.51480
C      3.52580    0.51270    0.41380
H      3.72040   -0.21860    1.19840
H      4.20110    0.36460   -0.43020
H      3.60200    1.53200    0.79400
O      1.61400    1.26350   -1.25780
```

Structure **1c-enolate (conf IV)**
```
C     -0.88970   -0.65180    1.15250
C      0.49100   -0.22560    0.85720
C      0.88910    1.19110    1.02160
H     -0.98340   -1.73480    1.26910
C      1.39480   -1.14900    0.30890
O      1.25040   -2.37050    0.12860
O      2.58930   -0.54260   -0.06350
C      3.60240   -1.41170   -0.55490
H      3.26780   -1.93300   -1.45480
H      4.45550   -0.77490   -0.78780
H      3.88420   -2.15040    0.19970
S     -2.02010   -0.23520   -0.24000
O     -1.81700    1.20570   -0.55340
H      1.89050    1.32060    1.44060
N      0.94560    1.92210   -0.32070
H      1.60080    1.39140   -0.90370
C      1.35090    3.34600   -0.23890
H      2.34150    3.40170    0.21110
H      1.36780    3.77390   -1.23970
H      0.62900    3.87420    0.38280
H      0.00630    1.82090   -0.73530
H     -1.34630   -0.14080    2.00830
H      0.17190    1.75390    1.62530
C     -3.68860   -0.41500    0.36480
H     -3.83470    0.27240    1.19800
H     -4.35160   -0.16450   -0.46480
H     -3.83330   -1.45060    0.67450
O     -1.83060   -1.17450   -1.36240
```

Structure **1c-TSelim (conf I)**
```
C     -0.66470   -0.46460    1.57980
C      0.59100   -0.10660    1.08930
C      1.02960    1.31990    1.05250
H     -0.87030   -1.52020    1.74140
C      1.38660   -1.10740    0.42630
O      1.13970   -2.30670    0.32970
O      2.50580   -0.57140   -0.14590
C      3.37120   -1.49050   -0.81550
H      2.84230   -1.98760   -1.63110
H      4.19340   -0.89280   -1.20460
H      3.74530   -2.24030   -0.11570
S     -2.15800   -0.17880   -0.05710
O     -1.69000    1.07460   -0.78020
O     -3.63560   -0.19420    0.20470
H      2.07150    1.46290    1.34920
N      0.91580    1.88060   -0.34770
H      1.56240    1.34630   -0.93660
C      1.19980    3.33270   -0.45260
H      2.20190    3.52660   -0.07150
H      1.12990    3.63350   -1.49650
H      0.46270    3.87090    0.14230
H     -0.05520    1.65690   -0.66670
H     -1.17830    0.21010    2.26160
H      0.39010    1.94310    1.68200
C     -1.84350   -1.49950   -1.24240
```

```
H     -2.39240   -1.25210   -2.15380
H     -0.76940   -1.54480   -1.42970
H     -2.19690   -2.43800   -0.81370
```

Structure **1c-TSelim (conf II)**
```
C     -0.38270   -0.42100    1.57880
C      0.74980    0.17640    1.02370
C      0.90290    1.66020    0.96610
H     -0.38140   -1.49520    1.74390
C      1.71140   -0.56460    0.25510
O      2.60640   -0.04060   -0.41880
O      1.56020   -1.90820    0.29850
C      2.47480   -2.66810   -0.49350
H      3.50020   -2.50850   -0.15430
H      2.18850   -3.70910   -0.35530
H      2.39440   -2.38860   -1.54620
S     -1.99320   -0.40620    0.03210
O     -1.82530    0.93270   -0.66690
O     -3.41970   -0.72730    0.36970
H      1.90980    1.99670    1.22840
N      0.65820    2.15420   -0.44040
H      1.36680    1.70060   -1.02910
C      0.71010    3.62750   -0.59610
H      1.68460    3.98520   -0.26520
H      0.55380    3.87910   -1.64370
H     -0.07700    4.06680    0.01570
H     -0.27410    1.77510   -0.71640
H     -0.97280    0.14520    2.29690
H      0.17590    2.16460    1.60610
C     -1.49220   -1.60470   -1.21790
H     -2.15320   -1.46500   -2.07600
H     -0.45470   -1.40700   -1.49260
H     -1.59600   -2.60710   -0.80090
```

Structure **1c-TSelim (conf III)**
```
C     -0.38520   -0.53170    1.45960
C      0.76030    0.12510    0.99180
C      0.86260    1.61360    0.99770
H     -0.34830   -1.61030    1.58840
C      1.78810   -0.54950    0.25500
O      2.71360    0.03140   -0.32930
O      1.68260   -1.89850    0.22710
C      2.67510   -2.58910   -0.53220
H      3.66990   -2.40910   -0.11960
H      2.42100   -3.64500   -0.45700
H      2.65300   -2.26760   -1.57560
S     -1.89720   -0.53600   -0.13870
O     -1.58350    0.70050   -0.95380
H      1.82700    1.97980    1.36140
N      0.73110    2.14460   -0.41130
H      1.51830    1.74200   -0.93440
C      0.72380    3.62240   -0.52190
H      1.64380    4.01390   -0.08910
H      0.65570    3.89970   -1.57250
H     -0.13820    4.00580    0.02290
H     -0.14060    1.71850   -0.79150
H     -1.01000   -0.01460    2.18650
H      0.06120    2.07120    1.58230
C     -3.59240   -0.27610    0.41320
H     -3.89450   -1.13130    1.01850
H     -3.63080    0.64640    0.99360
H     -4.21050   -0.19680   -0.48390
O     -1.91680   -1.79920   -0.94520
```

Structure **1c-TSelim (conf IV)**
```
C     -0.63440   -0.62860    1.44470
C      0.62340   -0.17580    1.02890
C      0.97750    1.27220    1.07530
```

```
H    -0.77680   -1.70130    1.55390        H     2.10130    1.41190    0.52200
C     1.50400   -1.09540    0.36240
O     1.33980   -2.30220    0.20000    Structure 1c-elim (conf III)
O     2.62170   -0.46960   -0.12220        C     0.77110    2.15260    0.06170
C     3.57400   -1.30610   -0.78010        C     0.38630    0.90690    0.35030
H     3.12290   -1.78890   -1.64930        C     1.27730   -0.10870    1.03820
H     4.38310   -0.64710   -1.09010        H     0.10120    2.84250   -0.44180
H     3.94870   -2.06990   -0.09580        C    -0.99510    0.49620   -0.05480
S    -2.06770   -0.37170   -0.20630        O    -1.79310    1.19830   -0.64240
O    -1.54880    0.82920   -0.97080        O    -1.26940   -0.76260    0.31210
H     1.98700    1.46550    1.44630        C    -2.57420   -1.24870   -0.03020
N     0.92340    1.88310   -0.30940        H    -2.71130   -1.22130   -1.11230
H     1.65000    1.42030   -0.86480        H    -2.61170   -2.27180    0.33650
C     1.10420    3.35450   -0.34140        H    -3.33900   -0.63820    0.45250
H     2.06400    3.60350    0.11040        H     0.76790   -0.50620    1.92100
H     1.07720    3.69420   -1.37530        N     1.69130   -1.23980    0.21010
H     0.29450    3.81380    0.22430        C     2.42310   -0.84690   -0.98980
H     0.00300    1.59360   -0.70970        H     3.33370   -0.31540   -0.69400
H    -1.19480   -0.02160    2.15380        H     2.71760   -1.74290   -1.54070
H     0.26200    1.83280    1.68170        H     1.85700   -0.19160   -1.66840
C    -3.72680    0.11860    0.29650        H     1.76350    2.50380    0.32870
H    -4.17100   -0.69730    0.86770        H     2.17960    0.40940    1.38150
H    -3.65340    1.02350    0.90060        H     0.87110   -1.78350   -0.04380
H    -4.29390    0.30710   -0.61780
O    -2.23950   -1.59470   -1.05540    Structure 1c-elim (conf IV)
                                            C    -0.21690    2.05390    0.03970
Structure 1c-elim (conf I)                  C    -0.29390    0.78560   -0.37200
C    -0.14680    2.44820   -0.03870        C    -1.54210    0.16650   -0.96700
C    -0.00390    1.13420    0.14840        H     0.69230    2.46090    0.46880
C     1.30110    0.48650    0.53860        C     0.87180   -0.13960   -0.24700
H    -1.10970    2.86920   -0.30960        O     0.83870   -1.30110   -0.61110
C    -1.21690    0.27480   -0.02810        O     1.95530    0.41760    0.29940
O    -2.31460    0.67110   -0.36310        C     3.09760   -0.43700    0.43890
O    -0.95810   -1.01330    0.23200        H     3.42070   -0.79490   -0.54010
C    -2.06270   -1.91790    0.09540        H     3.86910    0.17850    0.89570
H    -2.44290   -1.89120   -0.92690        H     2.85450   -1.28710    1.07850
H    -1.66680   -2.90230    0.33370        H    -1.30280   -0.25870   -1.94670
H    -2.85750   -1.64310    0.79060        N    -2.15660   -0.88630   -0.16090
H     1.16940   -0.06740    1.48410        C    -2.56040   -0.44080    1.16880
N     1.82950   -0.37090   -0.52330        H    -3.30160    0.35900    1.06710
H     1.17180   -1.13230   -0.67240        H    -3.02950   -1.27210    1.69970
C     3.12330   -0.93680   -0.15020        H    -1.73410   -0.05650    1.78540
H     3.10160   -1.47200    0.81260        H    -1.07490    2.71410   -0.05170
H     3.45560   -1.63010   -0.92570        H    -2.28200    0.95980   -1.11920
H     3.86190   -0.13310   -0.07290        H    -1.50430   -1.66230   -0.08570
H     0.69400    3.12430    0.07960
H     2.03320    1.27870    0.72630    Structure 1d
                                            C     0.78990    0.41540   -0.98200
Structure 1c-elim (conf II)                 C    -0.40330    0.70750   -0.11370
C    -0.28680    2.18880   -0.13490        C    -0.46260    1.77420    0.68750
C     0.06610    0.93120    0.14330        H     0.61280   -0.41990   -1.66150
C     1.47310    0.51590    0.48820        C    -1.43640   -0.38740   -0.08060
H    -1.31260    2.45180   -0.36780        O    -1.03560   -1.55860   -0.07370
C    -0.94320   -0.16960    0.14450        C    -3.70760   -1.14400    0.18960
O    -0.66400   -1.32450    0.41130        H    -3.24420   -1.94920    0.75640
O    -2.17840    0.22290   -0.17190        H    -4.56140   -0.74800    0.74320
C    -3.17970   -0.80330   -0.18150        H    -4.05800   -1.53720   -0.77150
H    -3.26730   -1.25110    0.80970        S     2.28360   -0.01580   -0.06160
H    -4.10530   -0.30530   -0.46070        O     2.62960    1.11350    0.81700
H    -2.92110   -1.57270   -0.91070        O     3.28500   -0.44280   -1.05300
H     1.48350    0.06930    1.49810        H    -1.31320    1.94400    1.34050
N     2.03120   -0.39540   -0.51100        H     1.10750    1.29230   -1.55390
H     1.47000   -1.24390   -0.50460        H     0.35270    2.49000    0.72500
C     3.41420   -0.74310   -0.19610        C     1.85880   -1.40760    0.95980
H     3.53910   -1.14970    0.82020        H     2.77360   -1.67000    1.49420
H     3.77690   -1.48270   -0.91300        H     1.07110   -1.10540    1.64960
H     4.04140    0.14970   -0.28170        H     1.51870   -2.21940    0.31850
H     0.45300    2.98350   -0.12970        N    -2.74560   -0.06740   -0.00980
```

```
C     -3.32890    1.22540   -0.35390
H     -3.74520    1.71280    0.53320
H     -2.58770    1.87650   -0.81050
H     -4.13660    1.05940   -1.07240
```

Structure **1d-TSadd**
```
C      0.11000    0.86500   -1.23420
C      0.05650   -0.59660   -0.93320
C     -1.06110   -1.33570   -1.24770
H      1.10690    1.27550   -1.41080
C      1.01990   -1.21880   -0.00670
O      0.64500   -2.06730    0.82580
C      3.19880   -1.22420    1.05800
H      3.63520   -2.21000    0.85320
H      4.00710   -0.49600    1.16060
H      2.63760   -1.27720    1.99070
S     -0.54180    1.92240    0.09830
O     -1.96030    1.58310    0.33900
O     -0.23920    3.32670   -0.23830
H     -1.01110   -2.41720   -1.17560
N     -2.41880   -1.38680    0.32770
H     -1.78570   -1.73310    1.04530
C     -3.59040   -2.23350    0.13590
H     -3.26330   -3.23440   -0.15890
H     -4.21410   -2.32300    1.03200
H     -4.19900   -1.82150   -0.67240
H     -2.63810   -0.41160    0.51750
H     -0.52830    1.14040   -2.08070
H     -1.78980   -0.93830   -1.94930
C      0.37870    1.48250    1.56210
H      0.02820    2.15390    2.34800
H      0.16550    0.44300    1.81740
H      1.44120    1.63310    1.36520
N      2.32030   -0.79260   -0.01670
C      2.99680   -0.31030   -1.21270
H      2.33390   -0.37380   -2.07520
H      3.34500    0.72150   -1.08830
H      3.86770   -0.94450   -1.41050
```

Structure **HSO2Me**
```
S     -0.13800    0.15420   -0.42660
O     -0.77650   -1.20270    0.29060
O     -0.57060    1.34460    0.36860
H     -1.61770   -1.42280   -0.14070
C      1.51900   -0.22170    0.15120
H      1.47110   -0.32630    1.23700
H      1.86380   -1.14110   -0.32240
H      2.15450    0.61850   -0.12960
```

Constant pH Molecular Dynamics (CpHMD) simulations

The pK_a of titratable residues was calculated using the method implemented by MacCammon[7] in the Amber 16 package supplemented with Ambertools 17[8] for the following target proteins: Hen white egg lysozyme (PDB 1G7H), Synaptotagmin I C2A domain, C2Am (PDB 3F04; the S95C mutation was modelled using PyMol), Annexin V (PDB 1AVH) and Trastuzumab/Herceptin® (PDB 1N8Z). This method works only in Generalized Born implicit solvent (*igb* = 2).[9,10]

For the generation of the topology and input coordinate files, the specifically developed *leaprc.constph* file containing all the necessary variables for a CpHMD simulation was used in combination with the *leap* utility. The underlying force field was *ff10* (equivalent to *ff99SB* for proteins) and the atomic radii (*PBRadii*) were set to *mbondi2*.[11] The salt concentration was set to 0.1 M. The *cpinutil.py* program was used to generate the constant pH input file (*cpin*) containing the definition of the target titratable residues (lysine and cysteine). An initial geometry optimization (1000 steepest-descent steps) in which the positions of the protein backbone atoms (CA, C, O and N) were restrained using a 10 kcal/mol·A^2 restraint force constant, was performed with no change in the protonation states (*icnstph* = 1, *ntcnstph* = 100000) and no cut-off for Lennard-Jones and electrostatic interactions (*cut* = 1000). The systems were then gently heated with no change in the protonation state by linearly incrementing the temperature from 10 to 300 K using the *nmropt* = 1 option. Harmonic restraints of 2 kcal/mol/A^2 were applied to the protein backbone, and temperature was controlled and equalized through Langevin dynamics (*ntt* = 3) with a collision frequency of γ = 5 ps^{-1} (*gamma_ln* = 5) and a time constant for heat bath coupling of τ_p = 2 ps (*tautp* = 2.0). A random seed (*ig* = -1) was used to initialize velocities to avoid synchronization artifacts.[12,13]

No constant pressure periodic boundary conditions were used (*ntp* = 0) and the particle-mesh-Ewald method[14] to model long-range electrostatic effects was turned off (*ntb* = 0) with

no cut-off for Lennard-Jones and electrostatic interactions. The SHAKE algorithm was used (*ntc* = 2, *ntf* = 2) with a relative geometrical tolerance for coordinate resetting of 1E-6 Å (*tol* = 0.000001), such that the angle between the hydrogen atoms is kept fixed. The time step was kept at 2 fs (*dt* = 0.002) during the 2 ns heating stage. Each system was then equilibrated for 2 ns with a 2 fs timestep at a constant temperature of 300 K, using Langevin dynamics under the same conditions described above. At this point constant pH in implicit solvent is turned on (*icnstph* = 1) and changing protonation states starting from physiological pH (*solvph* = 7.5) is attempted every 2 or 5 steps (*ntcnstph* = 2 or 5) depending on the number of the titratable residues of the target protein (**Supporting Table 2**). The original paper suggested that each residue should attempt to swap states every ~100 steps at least. Production trajectories were then run under the same simulation conditions for additional 40 or 100 ns depending on the target protein to facilitate proper conformational sampling, in a range spanning pH = 5 to pH 14 (**Supporting Table 2**). The program *cphstats* was used to analyse the results obtained from the CpHMD simulations. From these data, the pK_a values were computed using the Hill equation:

$$pK_a(Lys) = pH - n\log\frac{[LysNH_2]}{[LysNH_3^+]} \qquad pK_a(Cys) = pH - n\log\frac{[CysS^-]}{[CysSH]}$$

where $[LysNH_2]$ and $[LysNH_3^+]$ are the concentrations of the neutral and protonated forms of lysine, respectively, $[CysS^-]$ and $[CysSH]$ are the concentrations of the anionic and neutral forms of cysteine, respectively, and *n* is the Hill coefficient.

Knowing that the fractions of each deprotonated residues (*f_d*) are:

$$f_d(Lys) = \frac{[LysNH_2]}{[LysNH_3^+]+[LysNH_2]} \qquad f_d(Cys) = \frac{[CysS^-]}{[CysSH]+[CysS^-]}$$

Hill equation can be rearranged to:

$$f_d = \frac{1}{1 + 10^{n(pK_a - pH)}}$$

The calculated pK_a values and Hill coefficients for each residue were derived by fitting the protonated fraction (1 – f_d) at each considered pH using a non-linear, least-squares Marquardt-Levenberg algorithm as implemented in the Gnuplot program (http://gnuplot.sourceforge.net). The calculated values and CpHMD simulations parameters are summarized in **Supporting Table 2**. Note that different replicas of identical CpHMD simulations gave slightly different results. Thus, reported values are intended to reflect global trends more than exact numbers and must not be over interpreted; this is especially relevant for residues with very similar pK_a values. All graphical representations of proteins were performed using the PyMol software (http://www.pymol.org).

Supporting Table 2. CpHMD simulation parameters, calculated pK_a values and Hill coefficients (*n*)

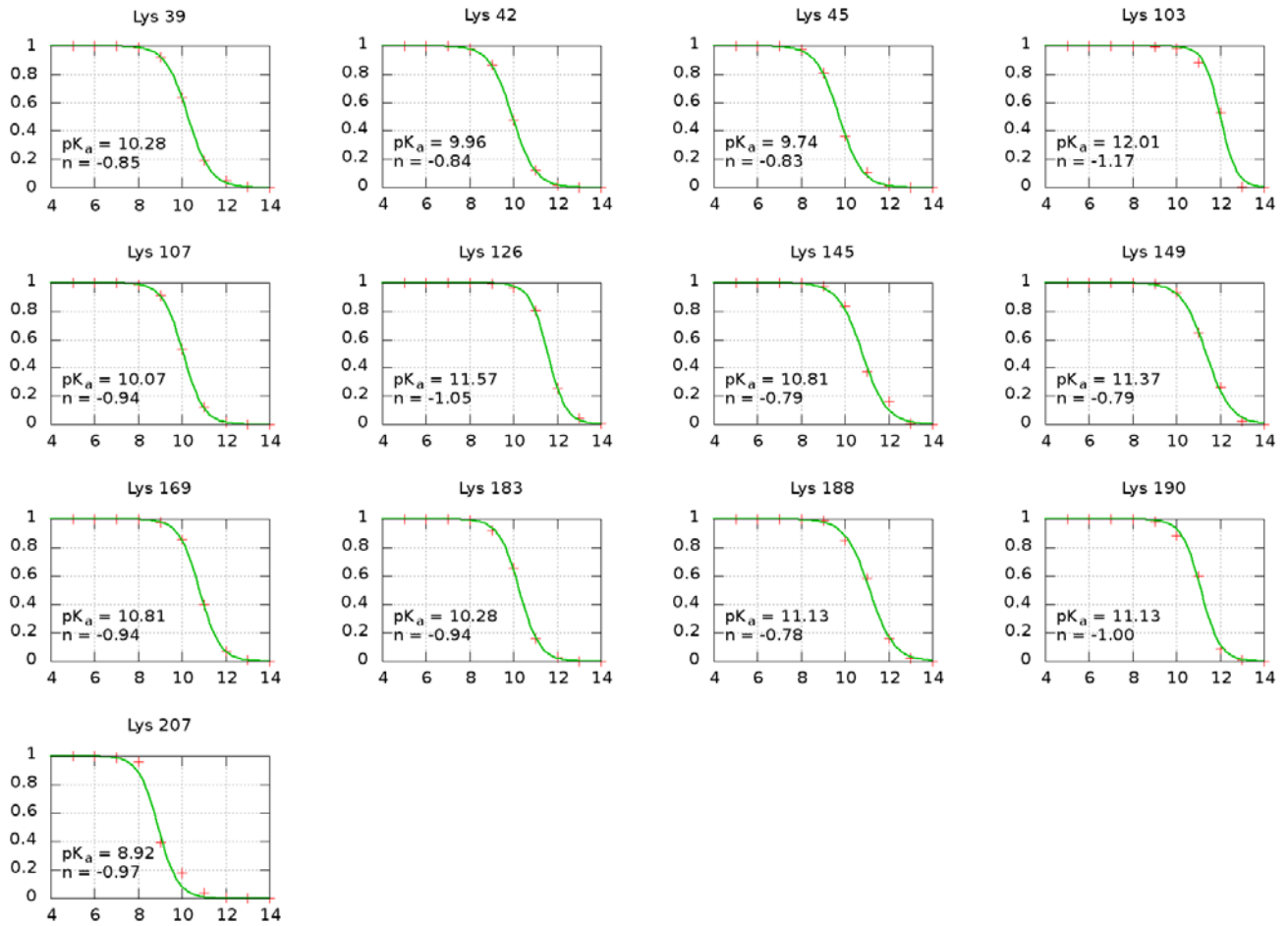Target protein: **Human Serum Albumin**
Production run: **40 ns**
Protonation state change attempted every **5** simulation steps

Protonated fraction (1 – f_d) for each residue as a function of the pH

pH	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
LYS64	0.999999	1.000000	0.999926	0.999474	0.994477	0.961369	0.504778	0.158012	0.010136	0.001034
LYS573	0.999997	0.999981	0.999758	0.998131	0.976351	0.627082	0.261659	0.103270	0.008036	0.000628
LYS574	0.999998	0.999982	0.999943	0.999358	0.988068	0.903824	0.416039	0.117707	0.010979	0.000925

Marquardt-Levenberg fitting results (protonated fraction vs. pH)
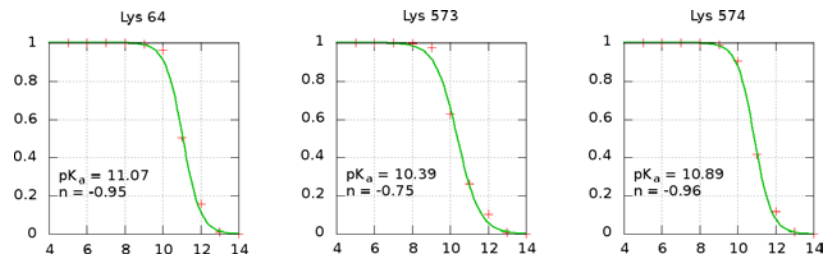
Residue	pKa	sigma[a]	n	sigma[a]
LYS64	11.070	0.042	-0.946	0.084
LYS573	10.390	0.055	-0.746	0.062
LYS574	10.890	0.030	-0.963	0.061

[a] sigma = asymptotic standard error

Target protein: **Lysozyme**
Production run: **40 ns**
Protonation state change attempted every **5** simulation steps

pH	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
LYS13	0.999993	0.999947	0.997636	0.992302	0.945720	0.398367	0.188996	0.012162	0.001551	0.000046
LYS33	0.999820	0.999562	0.997475	0.949310	0.775744	0.208294	0.032031	0.004438	0.000303	0.000018
LYS96	0.999989	0.999846	0.998788	0.990135	0.900597	0.547084	0.099162	0.011172	0.001762	0.000174
LYS97	0.999971	0.999822	0.997938	0.978243	0.900403	0.554339	0.091912	0.011426	0.001499	0.000092
LYS116	0.999996	0.999961	0.998962	0.986439	0.898668	0.542344	0.079672	0.010680	0.000642	0.000143

Protonated fraction ($1 - f_d$) for each residue as a function of the pH

Marquardt-Levenberg fitting results (protonated fraction vs. pH)
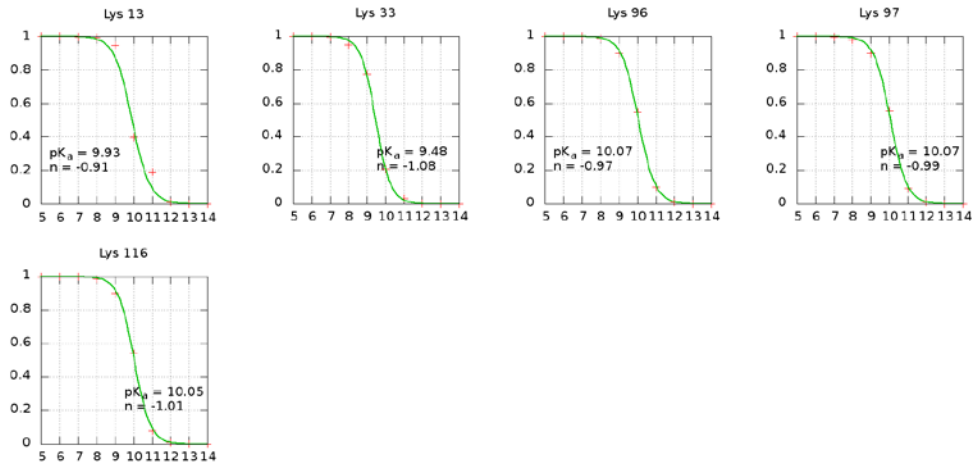
Residue	pK_a	sigma[a]	n	sigma[a]
LYS13	9.928	0.079	-0.911	0.142
LYS33	9.479	0.017	-1.082	0.035
LYS96	10.068	0.012	-0.971	0.026
LYS97	10.073	0.018	-0.986	0.039
LYS116	10.050	0.018	-1.012	0.042

[a] sigma = asymptotic standard error

Target protein: **C2Am** (S95C mutant)
Production run: **40 ns**
Protonation state change attempted every **5** simulation steps

Protonated fraction (1 – f_d) for each residue as a function of the pH

pH	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
LYS19	1.000000	0.999977	0.999898	0.999531	0.988915	0.931835	0.675904	0.194791	0.029217	0.003675
LYS22	1.000000	0.999999	0.999925	0.998385	0.990926	0.896486	0.528679	0.201681	0.063395	0.003145
LYS60	0.999985	0.999841	0.999491	0.991961	0.951109	0.723018	0.479659	0.098363	0.019021	0.001713
LYS67	0.999999	0.999906	0.999127	0.991474	0.910448	0.695823	0.243321	0.049635	0.006052	0.000679
LYS68	0.999980	0.999974	0.999640	0.994905	0.968965	0.287336	0.659747	0.266191	0.052265	0.000114
LYS69	0.999996	0.999856	0.999536	0.995759	0.955872	0.630256	0.315406	0.147217	0.013932	0.000697
LYS70	0.999979	0.999879	0.999067	0.987455	0.882378	0.972845	0.302036	0.063909	0.010939	0.007149
LYS74	1.000000	0.999996	0.999825	0.997620	0.978340	0.867355	0.390863	0.081598	0.007923	0.000928
LYS78	1.000000	0.999992	0.999933	0.997187	0.990792	0.860491	0.452550	0.079294	0.017320	0.001621
LYS91	0.999995	0.999935	0.999400	0.994906	0.949227	0.759636	0.301780	0.047998	0.007298	0.000572
CYS95	0.999997	0.999923	0.999327	0.992679	0.934039	0.761220	0.090126	0.088244	0.008090	0.000371
LYS100	0.999997	0.999436	0.999828	0.983693	0.913829	0.507527	0.216423	0.079761	0.010230	0.000182
LYS114	0.999996	0.999984	0.999537	0.996382	0.967262	0.810289	0.323950	0.051809	0.007660	0.000854
LYS122	1.000000	0.999992	0.999811	0.998107	0.979860	0.902447	0.530330	0.132242	0.013915	0.002051

Marquardt-Levenberg fitting results (protonated fraction vs. pH)
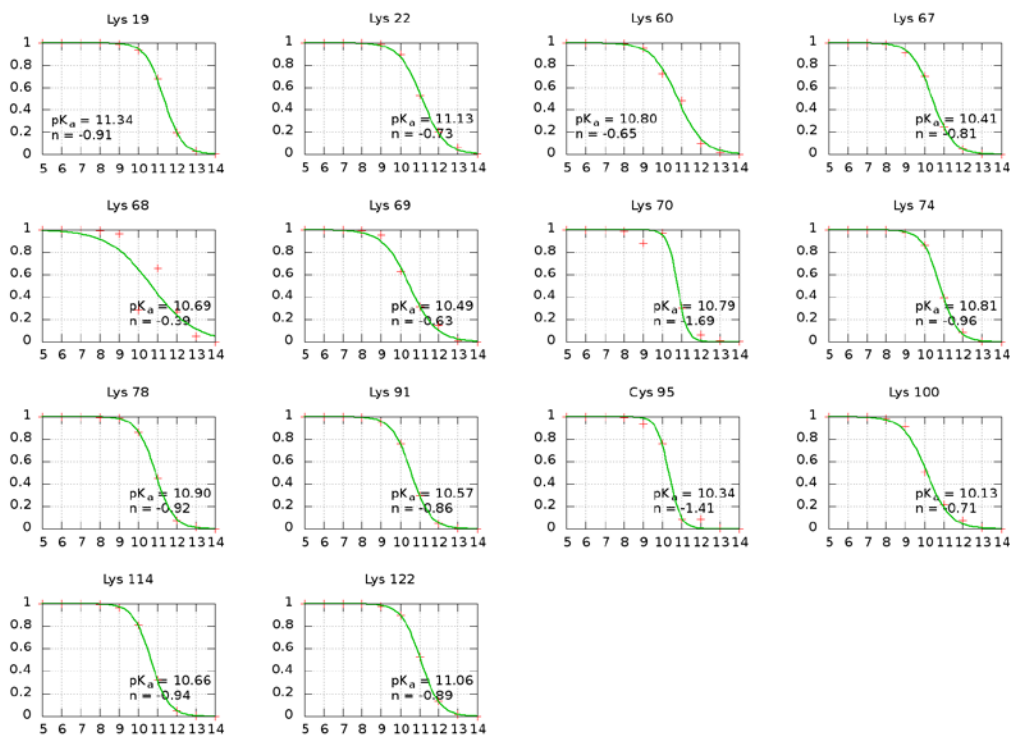
Residue	pK_a	sigma[a]	n	sigma[a]
LYS19	11.337	0.008	-0.913	0.014
LYS22	11.130	0.033	-0.733	0.036
LYS60	10.801	0.061	-0.654	0.053
LYS67	10.412	0.017	-0.810	0.022
LYS68	10.692	0.431	-0.387	0.133
LYS69	10.486	0.064	-0.634	0.052
LYS70	10.791	0.073	-1.688	0.478
LYS74	10.815	0.011	-0.963	0.021
LYS78	10.898	0.011	-0.916	0.019
LYS91	10.574	0.007	-0.864	0.011
CYS95	10.341	0.056	-1.411	0.181
LYS100	10.130	0.054	-0.713	0.056
LYS114	10.663	0.004	-0.942	0.007
LYS122	11.064	0.006	-0.888	0.010

[a] sigma = asymptotic standard error

Target protein: **Annexin V**
Production run: **100 ns**
Protonation state change attempted every **2** simulation steps

Protonated fraction (1 – f$_d$) for each residue as a function of the pH

pH	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
LYS25	0.999996	0.999948	0.999673	0.997490	0.978944	0.793121	0.305885	0.067508	0.006662	0.001083
LYS28	0.999998	0.999956	0.999462	0.996571	0.986707	0.812887	0.245752	0.084735	0.003041	0.001110
LYS57	0.999995	0.999971	0.999470	0.992680	0.957480	0.735778	0.105843	0.031285	0.003435	0.000529
LYS69	1.000000	0.999993	0.999950	0.998790	0.987482	0.929378	0.382953	0.250516	0.049093	0.002490
LYS75	1.000000	0.999992	0.999905	0.999405	0.997635	0.851481	0.553351	0.177397	0.014344	0.003059
LYS78	1.000000	0.999953	0.999882	0.999057	0.987745	0.832652	0.477981	0.136719	0.010647	0.002370
LYS85	0.999995	0.999896	0.998530	0.995209	0.940092	0.884963	0.296664	0.020648	0.001814	0.000514
LYS96	1.000000	0.999987	0.999766	0.997647	0.979229	0.924373	0.419093	0.021911	0.005596	0.000685
LYS100	0.999999	0.999972	0.999093	0.999878	0.968393	0.957536	0.682092	0.062910	0.010007	0.002846
LYS107	0.999997	0.999960	0.999620	0.998796	0.965795	0.880879	0.671688	0.076478	0.004650	0.000804
LYS125	0.999998	0.999997	0.999739	0.993832	0.995378	0.946398	0.142631	0.058165	0.011365	0.000146
LYS185	1.000000	0.999992	0.999953	0.999623	0.997870	0.884702	0.405456	0.152294	0.011232	0.001713
LYS192	1.000000	0.999994	0.999845	0.998393	0.985538	0.930947	0.737522	0.061074	0.012103	0.000636
LYS207	1.000000	0.999974	0.999575	0.997010	0.968484	0.979864	0.404945	0.075382	0.006683	0.001120
LYS211	1.000000	0.999981	0.999710	0.995396	0.984675	0.944704	0.447048	0.076113	0.003034	0.000549
LYS241	1.000000	0.997738	0.998655	0.999614	0.995021	0.989731	0.932309	0.379514	0.005434	0.001363
LYS259	0.999995	0.999971	0.999657	0.998894	0.920533	0.843513	0.198770	0.042238	0.004267	0.000744
LYS285	1.000000	0.999998	0.999976	0.997776	0.988658	0.911289	0.319186	0.049277	0.016819	0.000672
LYS289	1.000000	0.999992	0.999933	0.999192	0.992822	0.865509	0.530841	0.135377	0.016731	0.000743
LYS300	0.999988	0.999923	0.999885	0.997910	0.981220	0.619392	0.235598	0.102750	0.006448	0.000569
LYS308	0.999988	0.999984	0.999855	0.996220	0.951217	0.706026	0.213480	0.043284	0.007560	0.000655
LYS309	0.999996	0.999988	0.999911	0.999699	0.996482	0.938890	0.743154	0.348963	0.038861	0.003989
CYS315	1.000000	1.000000	1.000000	1.000000	1.000000	0.999975	1.000000	0.869211	0.385601	0.654927

Marquardt-Levenberg fitting results (protonated fraction vs. pH)

Lys 25 — pKa = 10.63, η = -0.92
Lys 28 — pKa = 10.58, η = -1.07
Lys 57 — pKa = 10.33, η = -1.33
Lys 69 — pKa = 10.97, η = -0.73
Lys 75 — pKa = 11.10, η = -0.75

Lys 78 — pKa = 10.94, η = -0.77
Lys 85 — pKa = 10.70, η = -1.23
Lys 96 — pKa = 10.88, η = -1.27
Lys 100 — pKa = 11.22, η = -1.44
Lys 107 — pKa = 11.22, η = -1.18

Lys 125 — pKa = 10.61, η = -2.00
Lys 185 — pKa = 10.88, η = -0.87
Lys 192 — pKa = 11.28, η = -1.55
Lys 207 — pKa = 10.90, η = -1.43
Lys 211 — pKa = 10.94, η = -1.17

Lys 241 — pKa = 11.84, η = -1.40
Lys 259 — pKa = 10.54, η = -1.27
Lys 285 — pKa = 10.75, η = -1.29
Lys 289 — pKa = 11.05, η = -0.83
Lys 300 — pKa = 10.35, η = -0.78

Lys 308 — pKa = 10.40, η = -0.93
Lys 309 — pKa = 11.61, η = -0.79
Cys 315 — pKa = 13.97, η = -0.32
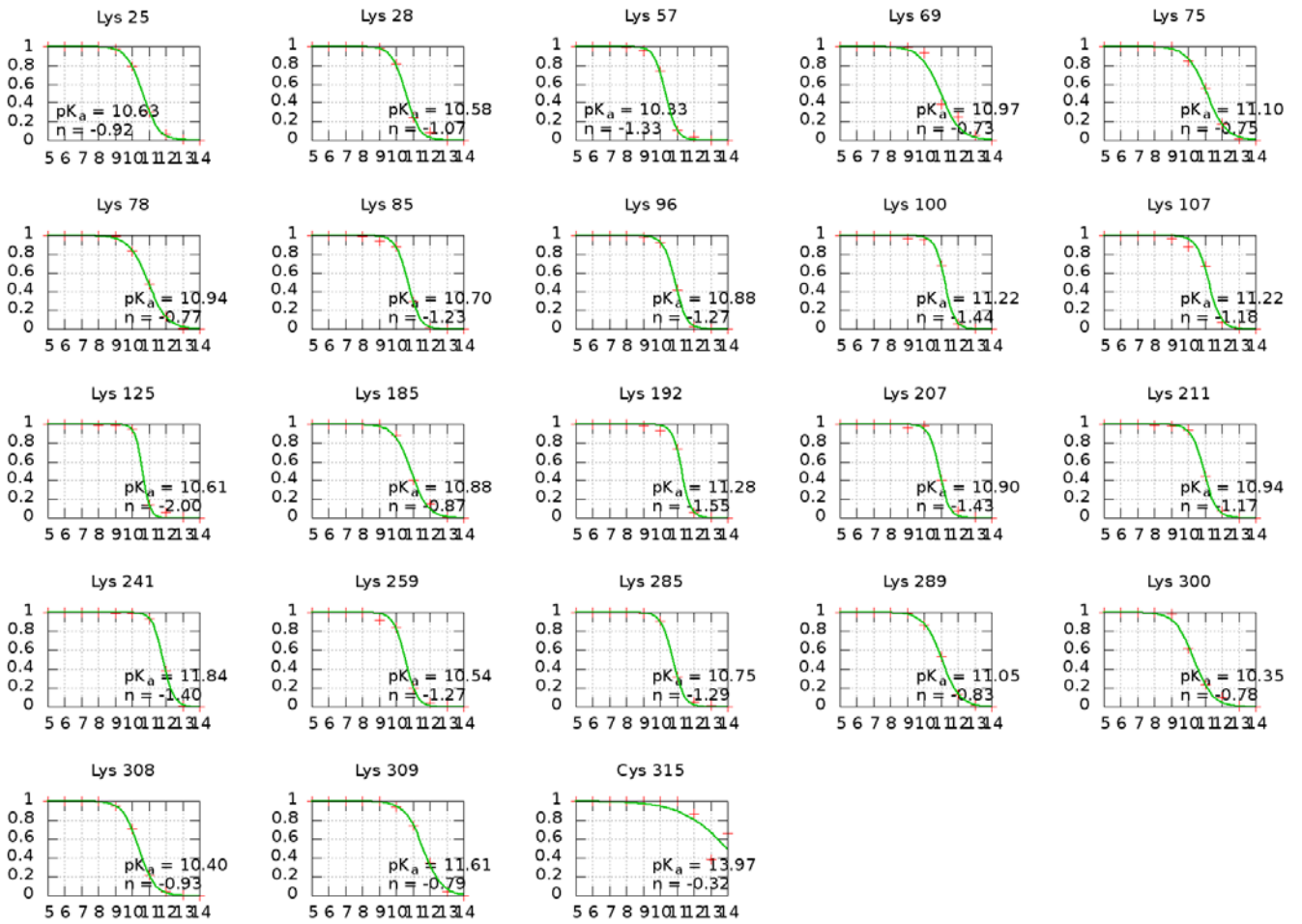
Residue	pK_a	sigma[a]	n	sigma[a]
LYS25	10.628	0.012	-0.924	0.020
LYS28	10.577	0.034	-1.066	0.068
LYS57	10.326	0.021	-1.327	0.063
LYS69	10.973	0.116	-0.735	0.129
LYS75	11.104	0.025	-0.754	0.030
LYS78	10.944	0.017	-0.773	0.021
LYS85	10.697	0.029	-1.229	0.080
LYS96	10.883	0.012	-1.272	0.049
LYS100	11.224	0.022	-1.443	0.098
LYS107	11.221	0.053	-1.182	0.156
LYS125	10.614	0.039	-2.004	0.187
LYS185	10.884	0.046	-0.868	0.074
LYS192	11.280	0.035	-1.553	0.151
LYS207	10.897	0.030	-1.434	0.193
LYS211	10.936	0.016	-1.175	0.059
LYS241	11.844	0.010	-1.404	0.054
LYS259	10.543	0.042	-1.270	0.106
LYS285	10.754	0.017	-1.288	0.057
LYS289	11.048	0.014	-0.825	0.021
LYS300	10.347	0.058	-0.783	0.071
LYS308	10.404	0.008	-0.930	0.013
LYS309	11.608	0.028	-0.792	0.035
CYS315	13.969	0.592	-0.316	0.136

[a] sigma = asymptotic standard error

Target protein: **Trastuzumab (light chain)**
Production run: **100 ns**
Protonation state change attempted every **5** simulation steps

Protonated fraction (1 – f$_d$) for each residue as a function of the pH

pH	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
LYS39	0.999996	0.999892	0.999236	0.991277	0.923060	0.637617	0.189059	0.052338	0.006406	0.000645
LYS42	0.999981	0.999764	0.998194	0.984401	0.864930	0.475390	0.123055	0.018472	0.002084	0.000234
LYS45	0.999981	0.999761	0.997641	0.979143	0.812119	0.364165	0.108210	0.011350	0.001249	0.000148
LYS103	1.000000	0.999993	0.999974	0.998538	0.993760	0.985845	0.880296	0.531261	0.007728	0.001569
LYS107	0.999991	0.999907	0.999018	0.990234	0.911655	0.533222	0.120320	0.017501	0.002087	0.000217
LYS126	1.000000	0.999997	0.999976	0.999709	0.994069	0.968379	0.805997	0.257210	0.047075	0.003483
LYS145	0.999996	0.999975	0.999816	0.997782	0.977905	0.837335	0.373043	0.161995	0.007603	0.000822
LYS149	0.999992	0.999995	0.999911	0.999260	0.993174	0.933815	0.646709	0.260840	0.023327	0.002044
LYS169	0.999993	0.999974	0.999760	0.997505	0.978265	0.854972	0.397284	0.074578	0.011551	0.001598
LYS183	0.999989	0.999917	0.999280	0.992986	0.924252	0.653617	0.163461	0.029612	0.002551	0.000308
LYS188	1.000000	0.999985	0.999888	0.999124	0.990895	0.851242	0.581562	0.162060	0.019454	0.001899
LYS190	1.000000	0.999972	0.999771	0.998861	0.985398	0.882037	0.599929	0.089587	0.010277	0.001269
LYS207	0.999977	0.999357	0.988881	0.963288	0.392089	0.179068	0.038634	0.005008	0.000588	0.000040

Marquardt-Levenberg fitting results (protonated fraction vs. pH)

Lys 39
pK$_a$ = 10.28
n = -0.85

Lys 42
pK$_a$ = 9.96
n = -0.84

Lys 45
pK$_a$ = 9.74
n = -0.83

Lys 103
pK$_a$ = 12.01
n = -1.17

Lys 107
pK$_a$ = 10.07
n = -0.94

Lys 126
pK$_a$ = 11.57
n = -1.05

Lys 145
pK$_a$ = 10.81
n = -0.79

Lys 149
pK$_a$ = 11.37
n = -0.79

Lys 169
pK$_a$ = 10.81
n = -0.94

Lys 183
pK$_a$ = 10.28
n = -0.94

Lys 188
pK$_a$ = 11.13
n = -0.78

Lys 190
pK$_a$ = 11.13
n = -1.00

Lys 207
pK$_a$ = 8.92
n = -0.97

Residue	pK_a	sigma[a]	n	sigma[a]
LYS39	10.282	0.013	−0.849	0.019
LYS42	9.956	0.006	−0.837	0.009
LYS45	9.742	0.021	−0.833	0.029
LYS103	12.006	0.043	−1.169	0.155
LYS107	10.066	0.003	−0.935	0.006
LYS126	11.574	0.012	−1.049	0.024
LYS145	10.806	0.050	−0.787	0.063
LYS149	11.368	0.025	−0.787	0.031
LYS169	10.811	0.004	−0.938	0.006
LYS183	10.275	0.013	−0.942	0.023
LYS188	11.134	0.030	−0.779	0.037
LYS190	11.130	0.036	−0.996	0.078
LYS207	8.918	0.081	−0.970	0.169

[a] sigma = asymptotic standard error
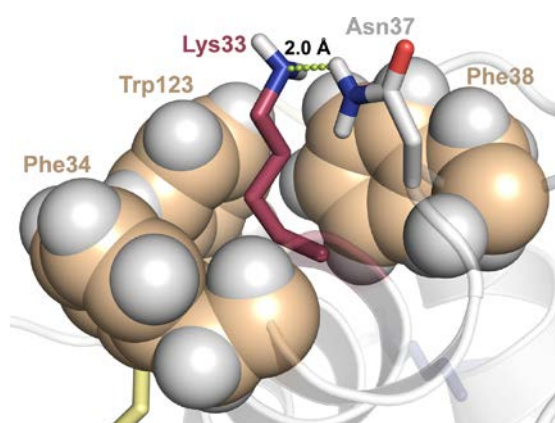
Supporting Figure 5. Detail of the X-ray structure of Hen white egg lysozyme (PDB 1G7H), showing the interaction between Lys33 (in red) and Asn37 (in white). This close contact suggests that Lys33 is in its neutral form, acting as a hydrogen bond acceptor. Hydrogens were modelled with PyMol. The hydrophobic pocket in which Lys33 is enclosed (Phe34-Trp123-Phe38) is shown in orange spheres.

2. Synthetic chemistry

General remarks

All solvents were commercially available grade. All reactions were carried out under argon atmosphere unless otherwise mentioned. All reagents were purchased from either Sigma-Aldrich, Alfa-Aesar or FluorChem and used without further purification. Reaction mixtures were analyzed by analytical thin-layer chromatography and flash column chromatography was performed on Merck TLC Silica gel 60 F254 glass plates and Silica Gel high purity grade (Merck grade 9385 pore size 60Å, 230-400 mesh particle size), respectively. Visualization was accomplished with UV light (254 nm), ninhydrin or $KMnO_4$. ^1H NMR and ^{13}C NMR spectra were recorded on a Bruker 400 MHz DPX-400 Dual Spectrometer and Bruker 500 MHz AVIII HD Smart Probe in the stated solvents ($CDCl_3$, DMF-*d7* or D_2O) using tetramethylsilane as an internal standard. Chemical shifts were reported in parts per million (ppm) on the δ scale from an internal standard (NMR descriptions: s, singlet). Mass spectroscopy was performed using a Waters micromass ZQ (LCMS) with Waters 2795 HPLC and a Waters 2996 photodiode array detector. This system is an automated service utilizing electrospray (ESI) ionization. The mobile phases are 95% aqueous acetonitrile with 0.05% formic acid and 10 mM ammonium acetate with 0.1% formic acid. The separation technology is based on a 50x4.6 mm C18 column (currently a Phenomenex Kinetix solid core column). There are several methods available enabling the user to produce mass spectra for compounds up to 2k Da in positive and negative modes of ionization. In some cases, a Waters LCT Premier combined with an Agilent 1100 autosampler was also used. The system

runs using 50% aqueous acetonitrile with 0.25% formic acid as mobile phase and can measure accurate masses from 150 Da to 1500 Da.

Synthesis of methyl 2-(methylsulfonyl)acrylate (1c)
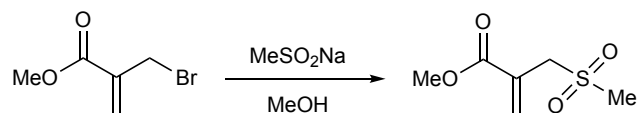
A solution of methyl 2-(bromomethyl)acrylate (0.18 mL, 1.5 mmol) in 5 mL of methanol was treated with sodium methanesulfinate (0.18 g, 1.5 mmol) portion wise over 10 min at room temperature. The solution was then stirred at room temperature for a period of 1.25 h and the solution concentrated *in vacuo*. The residue was then taken up in water and extracted four times with ethyl acetate. The combined ethyl acetate solution was washed with saturated sodium chloride, dried over anhydrous magnesium sulfate, filtered and concentrated to give a white solid. The residue was then purified by silica gel column chromatography eluting with hexane/ethyl acetate 8:2, to give desired product as a white solid (**1c**, 0.175 g, 65% yield).

M.p. = 64.9–66.7 °C. ^{1}H NMR (400 MHz, CDCl$_3$): 2.89 (s, 3H, CH$_3$), 3.81 (s, 3H, OCH$_3$), 4.05 (s, 2H, CH$_2$), 6.15 (s, 1H, CH$_2$), 6.62 (s, 1H, CH$_2$). ^{13}C NMR (100 MHz, CDCl$_3$): 40.5, 52.7, 56.4, 128.8, 134.2, 165.8. DEPT: 40.5, 52.7, 56.4, 134.2. IR (ATR): 1712 (CO) cm^{-1}. LC-MS m/z (ESI): 179 (MH$^+$). HRMS calcd. for C$_6$H$_{10}$O$_4$S (MH$^+$): 179.0300; found: 179.0347. Data was consistent with that previously reported.[15]
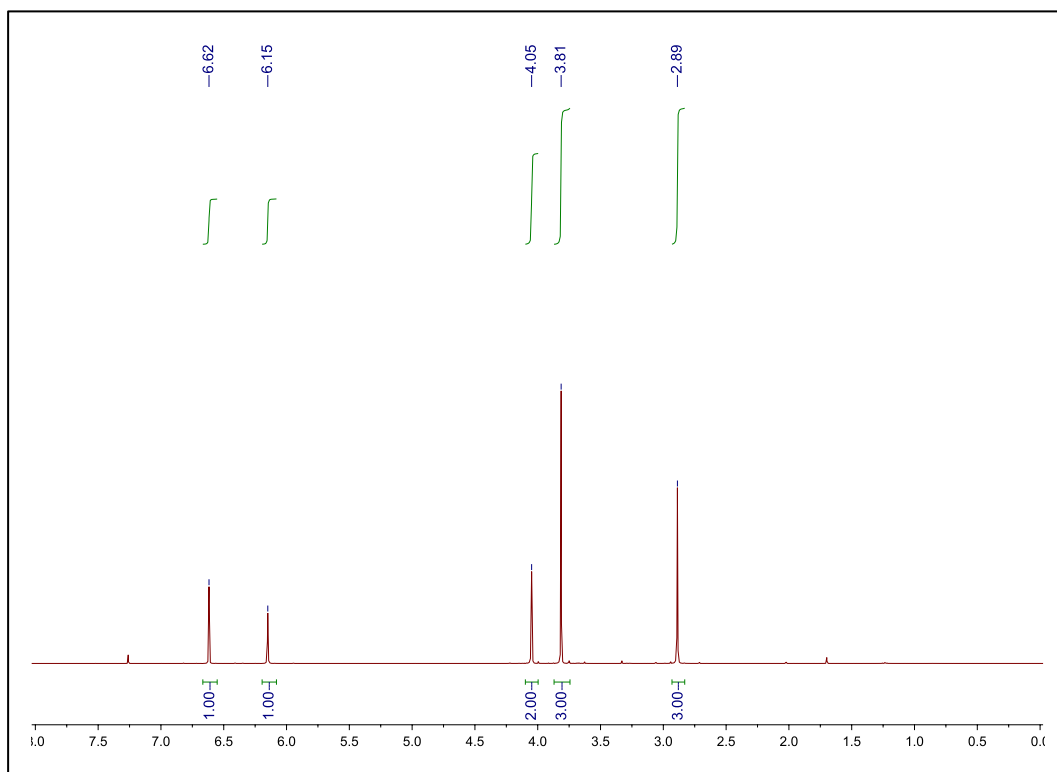
Supporting Figure 6. ^{1}H NMR of **1c**.

Synthesis of *N,N*-dimethyl-2-[(methylsulfonyl)methyl]acrylamide (1d)
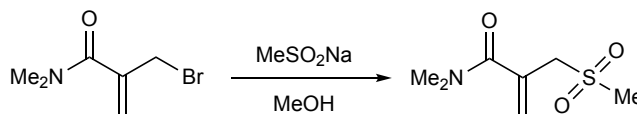
A solution of 2-(bromomethyl)-*N,N*-dimethylacrylamide[16] (222 mg, 1.16 mmol) in dry methanol (0.8 ml) under inert atmosphere was treated with sodium methanesulfinate (118 mg, 1.16 mmol, 85%) in portions for 10 min. and the resulting solution was allowed to stir at room temperature for 1 h. An additional portion of sodium methanesulfinate (118 mg, 1.16 mmol, 85%) was added, and the reaction was stirred for 15 h at room temperature. The solvent was then evaporated and the residue was partitioned between ethyl acetate and water. The organic layer was separated and the aqueous layer was extracted three times with ethyl acetate. The combined organic layers were washed with saturated sodium chloride, dried over anhydrous magnesium sulfate and concentrated. The residue obtained was purified by flash chromatography eluting with 100% ethyl acetate to afford the methyl sulfone (**1d**, 142 mg, 64%) as a white solid.

^{1}H NMR (400 MHz, CDCl$_3$): 2.90 (s, 3H, SCH$_3$), 2.94 (s, 3H, NCH$_3$), 3.09 (s, 3H, NCH$_3$), 3.99 (s, 3H, OCH$_3$), 5.51 (s, 1H, CH*H*), 5.72 (s, 1H, C*H*H). ^{13}C NMR (100 MHz, CDCl$_3$): 35.3 (NCH$_3$), 39.4 (NCH$_3$), 40.9 (SCH$_3$), 59.0 (CH$_2$), 124.5 (CH$_2$), 131.5 (C), 169.2 (CO). IR (ATR): 1645 (CO) cm^{-1}. LC-MS m/z (ESI): 192 (MH$^+$). HRMS calcd for C$_7$H$_{14}$NO$_3$S (MH$^+$): 192.0694; found: 192.0707.
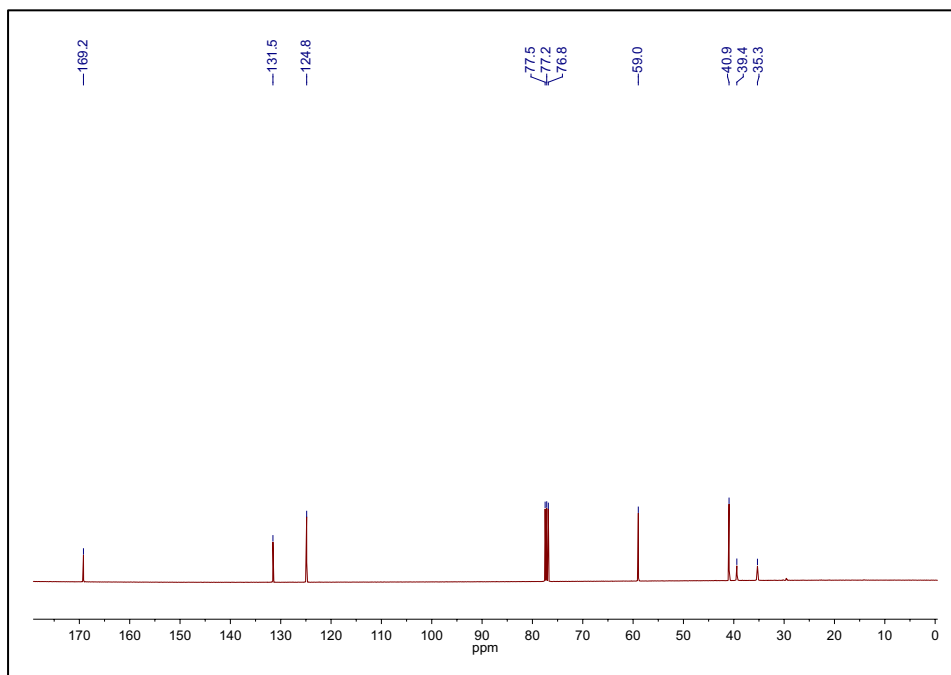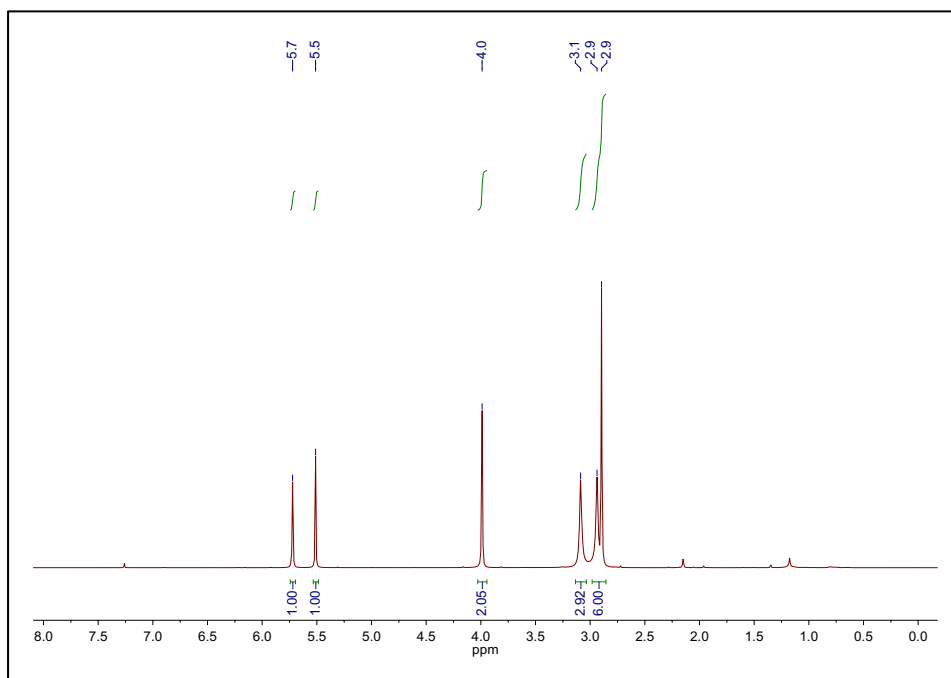
Supporting Figure 7. ^{1}H and ^{13}C NMR of **1d**.

Synthesis of Ac-GKAT-NH$_2$
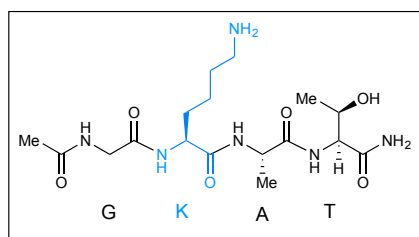
The peptide was synthesized by a stepwise solid-phase peptide synthesis using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). The Fmoc amino acids (10 molar equivalents) were automatically coupled on an Applied Biosystems 433A peptide synthesizer using HBTU. The acetylation step was carried out with Ac$_2$O/pyridine. The peptide was then released from the resin and all acid sensitive side chain protecting groups were simultaneously removed using TFA 95%, TIS 2.5%, H$_2$O 2.5%, followed by precipitation with diethyl ether. Finally, the peptide was purified by HPLC on a Waters Delta Prep 4000 reverse phase HPLC and Waters 2987 Dual Absorbance Detector, using a Phenomenex Luna C18(2) column (10 µ, 250 mm × 21.2 mm), 2% (v/v) CH$_3$CN in H$_2$O (containing 0.1% v/v TFA) gradient to 13% CH$_3$CN (t = 27 min) and then to 60% CH$_3$CN (t = 33 min), 10 mL/min and registered at λ = 212 nm.

Reaction of 1c with Ac-GKAT-NH₂ followed by ¹H NMR

An Ac-GKAT-NH$_2$ solution was prepared in 300 μL of NaPi (sodium phosphate buffer) in D$_2$O (50 mM, pH 8.0) and added to a NMR tube. Afterwards, 1 equiv. of the acrylate derivative **1c** was dissolved in 300 μL of DMF-*d7* and added to the previous solution (total concentration of 8 mM in 600 μL). The reaction was left at 37 ºC for 24 h. ¹H NMR spectra of the reaction were collected every hour over a 24 h period. Conversions were based on the disappearance of the double bond protons of the acrylate in **1c** and formation of two new peaks corresponding to the new acrylate formed after aza-Michael addition followed of lysine followed by spontaneous elimination of methanesulfinic acid (40% at 2 h). The comparison of the signal from α-protons of the lysine were also compared. ¹H NMR spectra were recorded on a Bruker 400 MHz DPX-400 Dual Spectrometer in the stated solvents using D$_2$O standard.
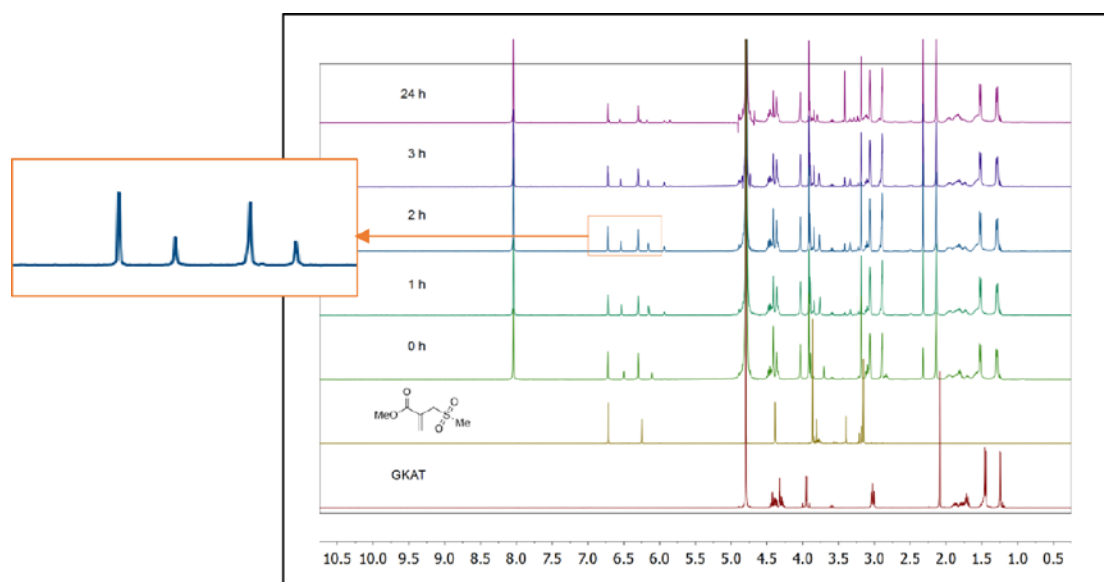
Supporting Figure 8. Monitoring of the reaction progress of peptide Ac-GKAT-NH$_2$ and **1c** (1 equiv.) by ¹H NMR [400 MHz, 37 ºC, NaP$_i$ in D$_2$O (50 mM, pH 8.0):DMF-d$_7$, 1:1].

3. General procedures and characterization methods

General procedure for protein and antibody conjugation with sulfonyl acrylates

To an eppendorf tube with TrisHCl (20 mM, pH 8.0) and DMF (10% of total volume), an aliquot of a stock solution of protein (final concentration 10 μM) was added. Afterwards, a solution of the acrylate derivative **1c** (1 equiv.) in DMF was added and the resulting mixture was vortexed for 10 seconds. The reaction was mixed for 1, 2 or 24 h at 37 ºC. A 10 μL aliquot of each reaction time was analysed by LC–MS and conversion to the expected product was observed.

LC–MS method for analysis of protein conjugation

LC–MS was performed on a Xevo G2-S TOF mass spectrometer coupled to an Acquity UPLC system using an Acquity Q6 UPLC BEH300 C4 column (1.7 mm, 2.1 × 50 mm). Solvents A, a water with 0.1% formic acid and B, 71% acetonitrile, 29% water and 0.075% formic acid were used as the mobile phase at a flow rate of 0.2 mL min^{-1}. The gradient was programmed as follows: 72% A to 100% B after 25 min then 100% B for 2 min and after that 72% A for 18 min. The electrospray source was operated with a capillary voltage of 2.0 kV and a cone voltage of 40 V. Nitrogen was used as the desolvation gas at a total flow of 850 L h^{-1}. Total mass spectra were reconstructed from the ion series using the MaxEnt algorithm preinstalled on MassLynx software (v. 4.1 from Waters) according to the manufacturer's instructions. To obtain the ion series described,

the major peak(s) of the chromatogram were selected for integration and further analysis.

Analysis of protein conjugation by LC–MS

A typical analysis of a conjugation reaction by LC–MS is described below. The total ion chromatogram, combined ion series and deconvoluted spectra are shown for the product of the reaction. Identical analyses were carried out for all the conjugation reactions performed in this work.
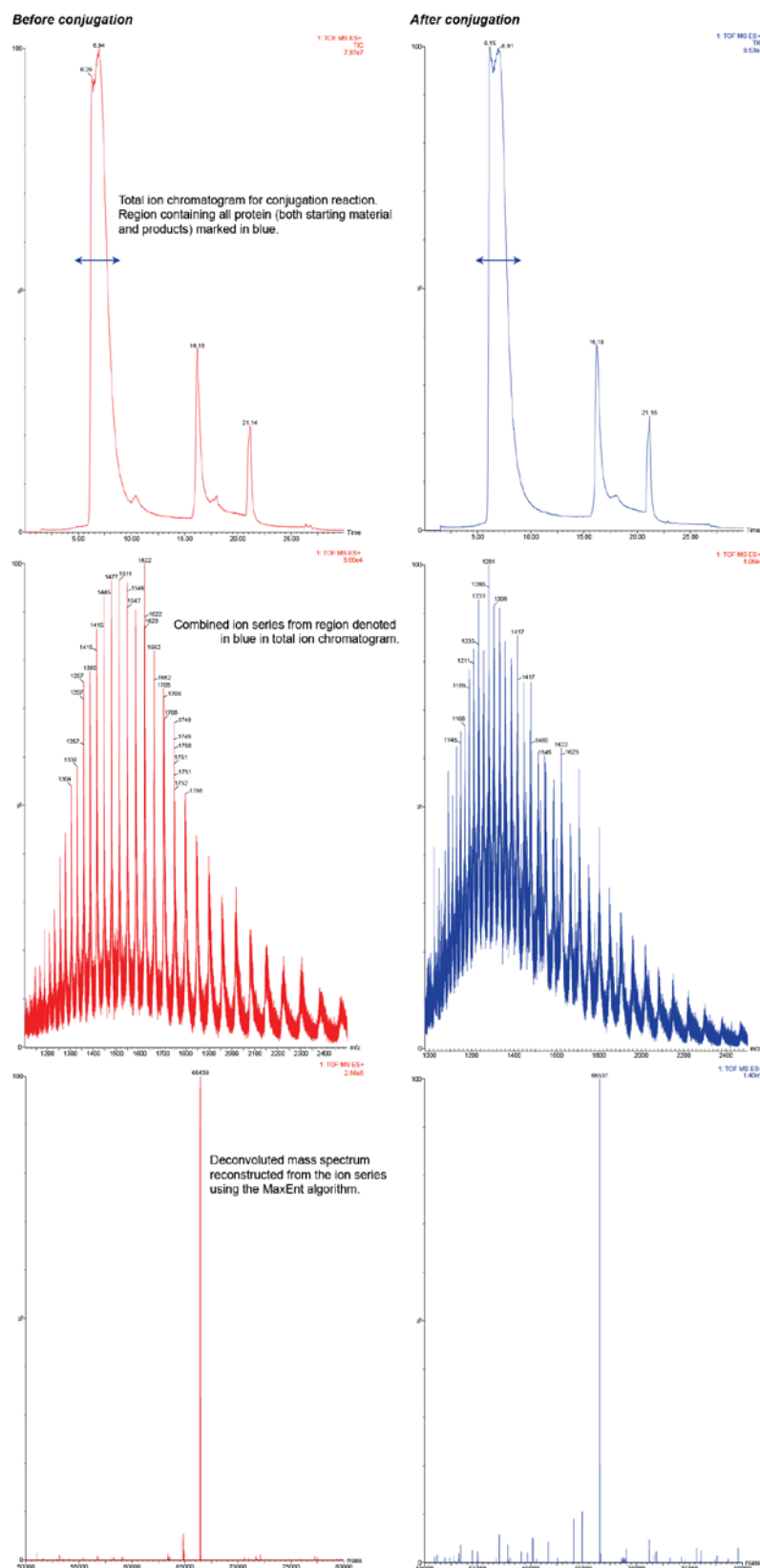
Supporting Figure 9. A typical analysis of a conjugation reaction by LC–MS is described for the reaction of rHSA protein with the acrylate derivative **1c**. The total ion chromatogram, combined ion series and deconvoluted spectra are shown for the starting material and the product of the reaction of rHSA with 1 equiv. of **1c**. Identical analyses were carried out for all the conjugation reactions performed in this work.

Enzymatic digestion and LC–MS/MS analysis

Lysozyme was subjected to reduction (DTT) and alkylation (iodoacetamide) prior to overnight solution digestion with chymotrypsin (Promega, Winsconsin, USA) at 25 ºC and pH 8.0. Albumin was reduced and alkylated before overnight solution digestion with *Staphylococcus aureus* Protease V8 (New England BioLabs, Massachusetts, USA) at 37 ºC and pH 8.0. C2Am and Annexin V were not reduced and alkylated and were digested in solution overnight with chymotrypsin at 25 ºC and pH 8.0.

All LC–MS/MS experiments were performed using a nanoAcquity UPLC (Waters Corp., Milford, MA) system and an LTQ Orbitrap Velos hybrid ion trap mass spectrometer (Thermo Scientific, Waltham, MA). Separation of peptides was performed by reverse-phase chromatography using a Waters reverse-phase nano column (HSS T3 C18, 75 μm i.d. x 250 mm, 1.7 μm particle size) at flow rate of 300 nL/min. Peptides were initially loaded onto a pre-column (Waters UPLC Trap Symmetry C18, 180 μm i.d x 20 mm, 5 μm particle size) from the nanoAcquity sample manager with 0.1% formic acid for 3 min at a flow rate of 5 μL/min. After this period, the column valve was switched to allow the elution of peptides from the pre-column onto the analytical column. Solvent A was water + 0.1% formic acid and solvent B was acetonitrile + 0.1% formic acid. The linear gradient employed was 5-40% B in 40 min.

The LC eluent was sprayed into the mass spectrometer by means of a nanospray ion source. All *m/z* values of eluting ions were measured in the Orbitrap Velos mass analyzer, set at a resolution of 30000. Data dependent scans (Top 20) were employed to automatically isolate and generate fragment

ions by collision-induced dissociation in the linear ion trap, resulting in the generation of MS/MS spectra. Ions with charge states of 2+ and above were selected for fragmentation. Post-run, the data was processed using Protein Discoverer (version 1.4., ThermoFisher). Briefly, all MS/MS data were converted to mgf files and these were submitted to the Mascot search algorithm (Matrix Science, London UK) and searched against custom databases containing the lysozyme, albumin, C2Am and Annexin V protein sequences along with common contaminant sequences, such as keratins, trypsin etc., (http://www.thegpm.org/crap/) and applying variable modifications of oxidation (M), deamidation (NQ), carbamidomethyl (albumin and lysozyme only) and custom modifications of **1c** (K) (lysozyme, C2Am, albumin and Annexin V), and using peptide tolerances of 25 ppm (MS) and 0.8 Da (MS/MS). Peptide identifications were accepted if they could be established at greater than 95.0% probability. Significant hits that suggested that the expected Lys modifications were bound to peptides were then verified by manual inspection of the MS/MS data.

Protein gels

The incubation solution (5.0 mL) was transferred to tube, and NuPAGE LDS Sample Buffer (4x, 2.5 mL), NuPAGE Reducing Agent (10x, 1 mL), and H_2O (1.5 mL) were added to the tube. The solution was heated at 70 ºC for 10 min. The heated solution was loaded to NuPAGE Bis-Tris mini gel (10x 10 cm) with 4–12% gradient polyacrylamide concentration, and then the conjugation reaction was analysed by electrophoresis (200 V). The buffering system employed was 1x SDS Running Buffer (NuPAGE MES SDS Running Buffer,

20x, pH 7.3, 50 to 950 mL deionized water). For reduced samples, 500 mL of NuPAGE antioxidant was added to each 200 mL 1x SDS running buffer. After 35 min, the intensities of fluorescence were analysed. Then, the gel was stained with 0.5% of Ruby. The gel was mixed overnight at room temperature and read the day after. After wash the gel, coomassie (0.5%) was added and the gel was read 2 h after mixing at room temperature.

Stability of bioconjugates in human plasma

A 20 μL aliquot of the bioconjugate (10 μM) in TrisHCl buffer (20 mM, pH 8.0) was thawed. 1 μL of reconstituted human plasma was added at room temperature and the resulting mixture vortexed for 10 seconds. The resulting reaction mixture was then mixed at 37 ºC. After 1 and 48 h, a 10 μL aliquot of each reaction mixture was analysed by LC–MS.

Stability of bioconjugates in the presence of GSH (1 mM)

A 20 μL aliquot of the bioconjugate (10 μM) in TrisHCl buffer (20 mM, pH 8.0) was thawed. 1 μL of a 20 mM glutathione solution (6 mg glutathione dissolved in 1 mL of TrisHCl buffer (20 mM, pH 8.0)) was added at room temperature and the resulting mixture vortexed for 10 seconds. The resulting reaction mixture was then mixed at 37 °C. After 1 and 48 h, a 10 μL aliquot of each reaction mixture was analysed by LC–MS.

Circular dichroism

Circular dichroism (CD) spectroscopy was used to analyse protein secondary structure in solution. Samples were concentrated to 10 nM in NaP$_i$ buffer (50 mM, pH 8.0). CD measurements were recorded using a Chirascan spectrophotometer equipped with a Quantum TC125 temperature control unit (25 ºC). The data was acquired in a 0.1 cm path length with a response time of 1 s, a per-point acquisition delay of 5 ms and a pre- and post-scan delay of 50 ms. Spectra were averaged over three scans, in a wavelength range from 200 nm to 260 nm, and the spectrum from a blank sample containing only buffer was subtracted from the averaged data.[17]

4. Proteins and antibodies used in this study

Recombinant human serum albumin (rHSA) was kindly provided by Albumedix Limited; C2Am was provided by Dr. André Neves and Prof. Kevin Brindle;[18] Lysozyme was purchased from Sigma-Aldrich; and finally Annexin V was expressed and purified as previously described.[19]

The Trastuzumab antibody used in this study was purchased from commercial supplier (Carbosynth Limited).

rHSA (59 Lys, 1 free Cys, 17 disulfides)

Sequence:

```
DAHKSEVAHR  FKDLGEENFK  ALVLIAFAQY  LQQCPFEDHV  KLVNEVTEFA   50
KTCVADESAE  NCDKSLHTLF  GDKLCTVATL  RETYGEMADC  CAKQEPERNE   100
CFLQHKDDNP  NLPRLVRPEV  DVMCTAFHDN  EETFLKKYLY  EIARRHPYFY   150
APELLFFAKR  YKAAFTECCQ  AADKAACLLP  KLDELRDEGK  ASSAKQRLKC   200
ASLQKFGERA  FKAWAVARLS  QRFPKAEFAE  VSKLVTDLTK  VHTECCHGDL   250
LECADDRADL  AKYICENQDS  ISSKLKECCE  KPLLEKSHCI  AEVENDEMPA   300
DLPSLAADFV  ESKDVCKNYA  EAKDVFLGMF  LYEYARRHPD  YSVVLLLRLA   350
KTYETTLEKC  CAAADPHECY  AKVFDEFKPL  VEEPQNLIKQ  NCELFEQLGE   400
YKFQNALLVR  YTKKVPQVST  PTLVEVSRNL  GKVGSKCCKH  PEAKRMPCAE   450
DYLSVVLNQL  CVLHEKTPVS  DRVTKCCTES  LVNRRPCFSA  LEVDETYVPK   500
EFNAETFTFH  ADICTLSEKE  RQIKKQTALV  ELVKHKPKAT  KEQLKAVMDD    550
FAAFVEKCCK  ADDKETCFAE  EG**K**KLVAASQ  AALGL  585
```

Isotopically Averaged Molecular Weight = 66472 Da;

with 17 internal disulfides: 66439 Da
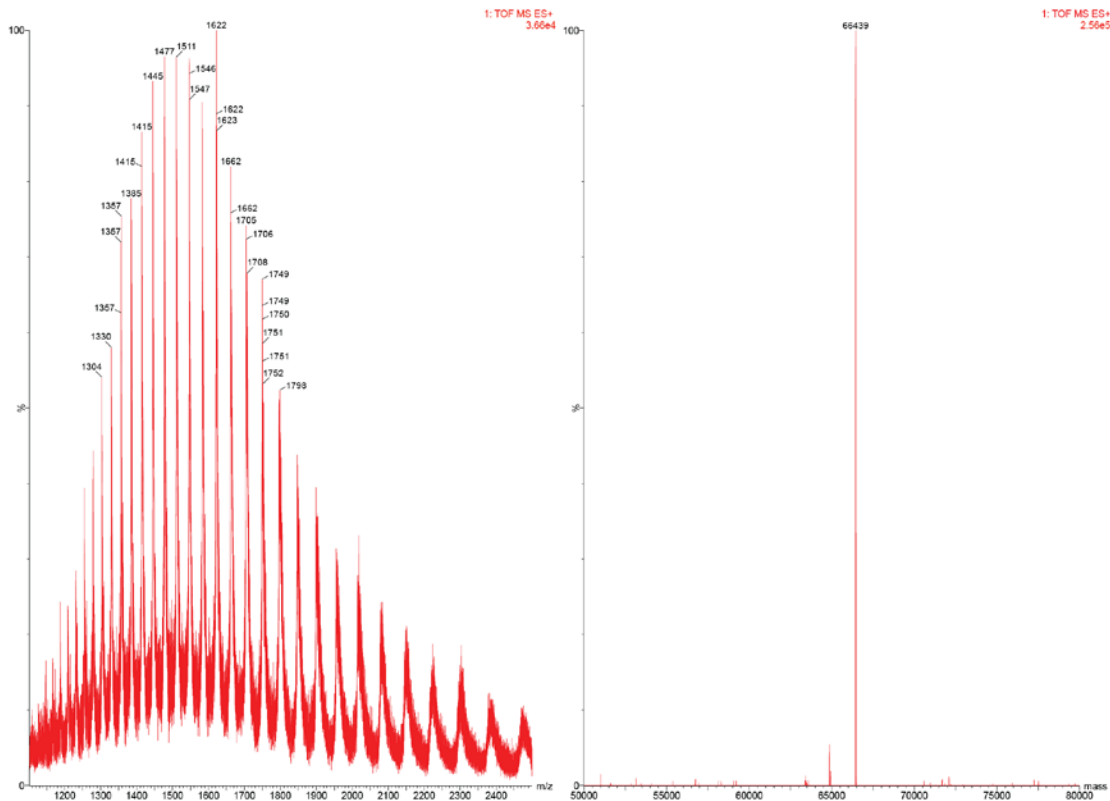
Supporting Figure 10. Combined ion series and deconvoluted mass spectrum of rHSA.

rHSA-K573P (58 Lys, 1 free Cys, 17 disulfides)

Sequence:

```
DAHKSEVAHR  FKDLGEENFK  ALVLIAFAQY  LQQCPFEDHV  KLVNEVTEFA  50
KTCVADESAE  NCDKSLHTLF  GDKLCTVATL  RETYGEMADC  CAKQEPERNE  100
CFLQHKDDNP  NLPRLVRPEV  DVMCTAFHDN  EETFLKKYLY  EIARRHPYFY  150
APELLFFAKR  YKAAFTECCQ  AADKAACLLP  KLDELRDEGK  ASSAKQRLKC  200
ASLQKFGERA  FKAWAVARLS  QRFPKAEFAE  VSKLVTDLTK  VHTECCHGDL  250
LECADDRADL  AKYICENQDS  ISSKLKECCE  KPLLEKSHCI  AEVENDEMPA  300
DLPSLAADFV  ESKDVCKNYA  EAKDVFLGMF  LYEYARRHPD  YSVVLLLRLA  350
KTYETTLEKC  CAAADPHECY  AKVFDEFKPL  VEEPQNLIKQ  NCELFEQLGE  400
YKFQNALLVR  YTKKVPQVST  PTLVEVSRNL  GKVGSKCCKH  PEAKRMPCAE  450
DYLSVVLNQL  CVLHEKTPVS  DRVTKCCTES  LVNRRPCFSA  LEVDETYVPK  500
EFNAETFTFH  ADICTLSEKE  RQIKKQTALV  ELVKHKPKAT  KEQLKAVMDD  550
FAAFVEKCCK  ADDKETCFAE  EG PKLVAASQ  AALGL  585
```

Isotopically Averaged Molecular Weight = 66441 Da;

with 17 internal disulfides: 66408 Da
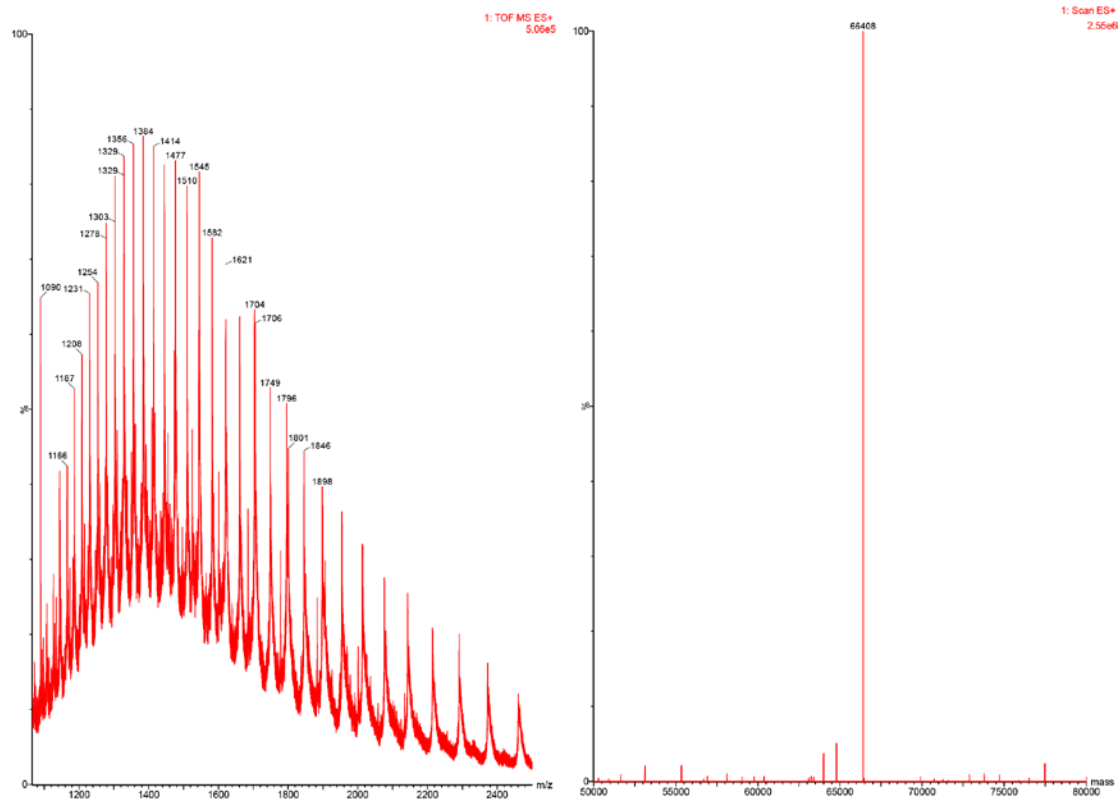
Supporting Figure 11. Combined ion series and deconvoluted mass spectrum of rHSA-K573P.

Lysozyme (6 Lys, no free Cys, 4 disulfides)

Sequence:

KVFGRCELAA AMKRHGLDNY RGYSLGNWVC AA**K**FESNFNT QATNRNTDGS   50
TDYGILQINS RWWCNDGRTP GSRNLCNIPC SALLSSDITA SVNCAKKIVS   100
DGNGMNAWVA WRNRCKGTDV QAWIRGCRL 129

Isotopically Averaged Molecular Weight = 14313 Da;

with 4 internal disulfides: 14305 Da
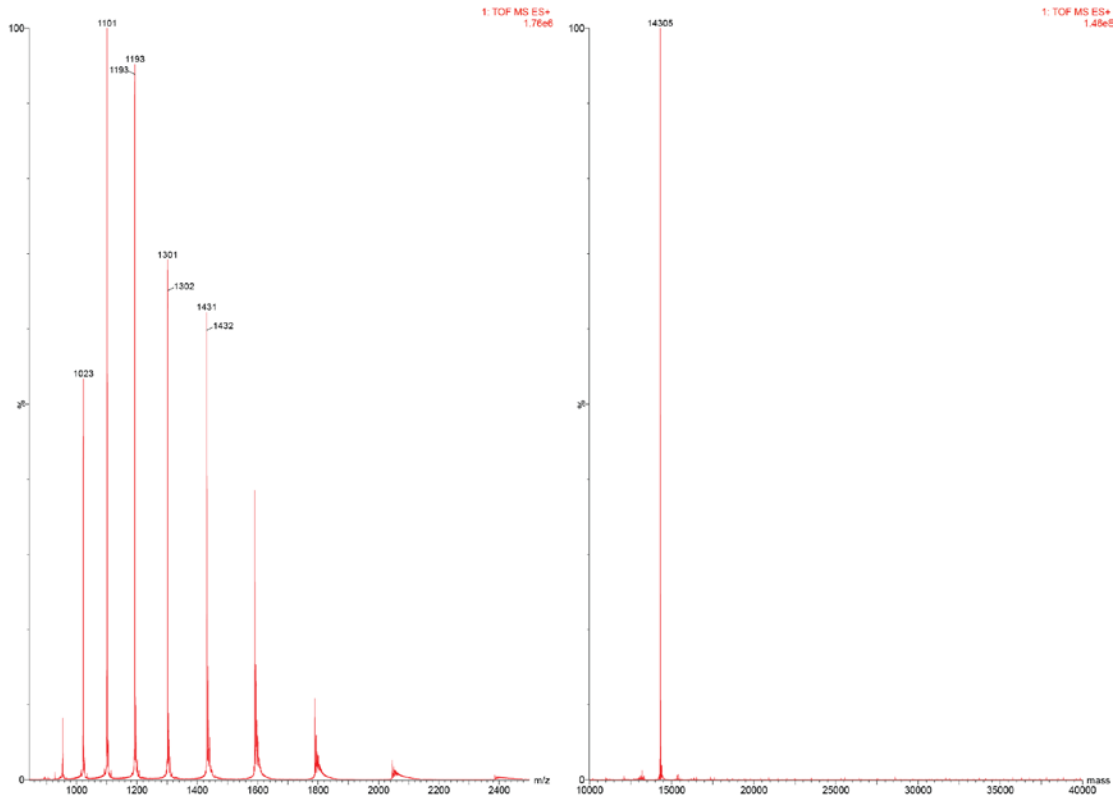
Supporting Figure 12. Combined ion series and deconvoluted mass spectrum of lysozyme.

C2Am (14 Lys, 1 free Cys, no disulfides)

Sequence:

```
GSPGISGGGG GILDSMVEKL GKLQYSLDYD FQNNQLLVGI IQAAELPALD   50
MGGTSDPYVK VFLLPDKKKK FETKVHRKTL NPVFNEQFTF KVPYCELGGK  100
TLVMAVYDFD RFSKHDIIGE FKVPMNTVDF GHVTEEWRDL QSAEK 145
```

Isotopically Averaged Molecular Weight = 16222 Da
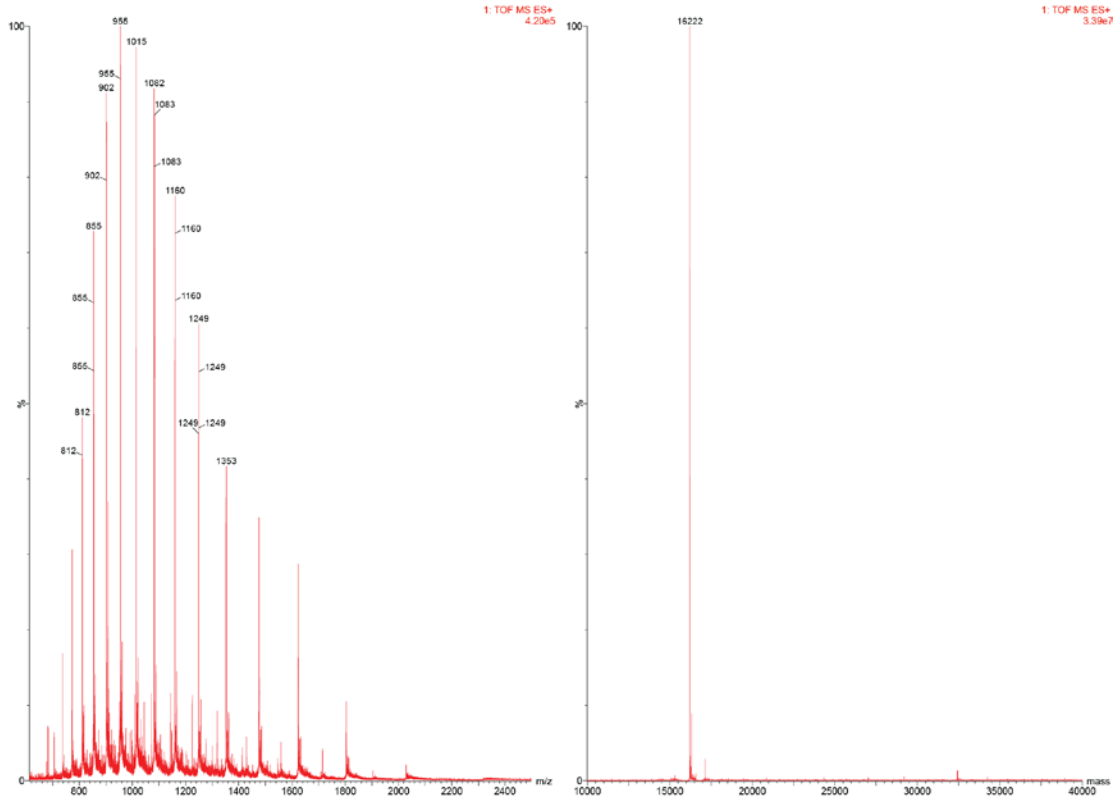
Supporting Figure 13. Combined ion series and deconvoluted mass spectrum of C2Am.

Annexin V (22 Lys, 1 free Cys, no disulfides)

Sequence:

```
AQVLRGTVTD  FPGFDERADA  ETLRKAMKGL  GTDEESILTL  LTSRSNAQRQ   50
EISAAFKTLF  GRDLLDDLKS  ELTGKFEKLI  VALMKPSRLY  DAYELKHALK   100
GAGTNEKVLT  EIIASRTPEE  LRAIKQVYEE  EYGSSLEDDV  VGDTSGYYQR   150
MLVVLLQANR  DPDAGIDEAQ  VEQDAQALFQ  AGELKWGTDE  EKFITIFGTR   200
SVSHLRKVFD  KYMTISGFQI  EETIDRETSG  NLEQLLLAVV  KSIRSIPAYL   250
AETLYYAMKG  AGTDDHTLIR  VMVSRSEIDL  FNIRKEFRKN  FATSLYSMI**K**   300
GDTSGDYKKA  LLLLCGEDD  319
```

Isotopically Averaged Molecular Weight = 35805 Da
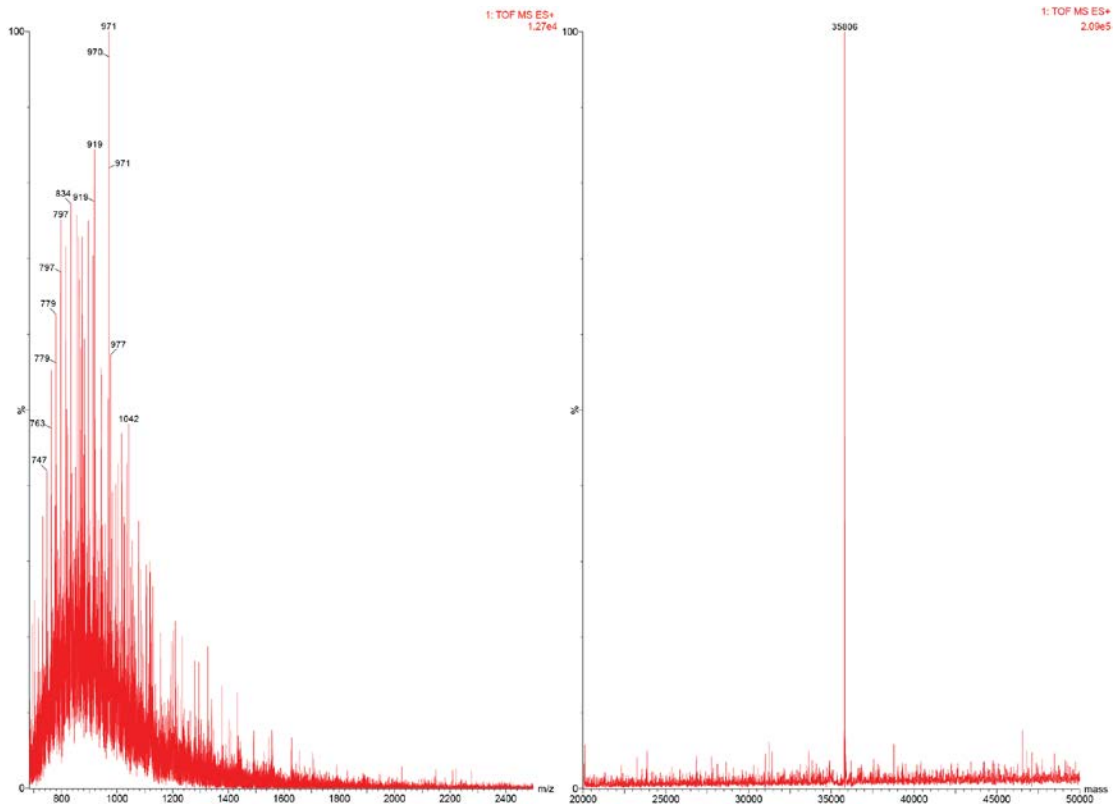
Supporting Figure 14. Combined ion series and deconvoluted mass spectrum of Annexin V.

Trastuzumab
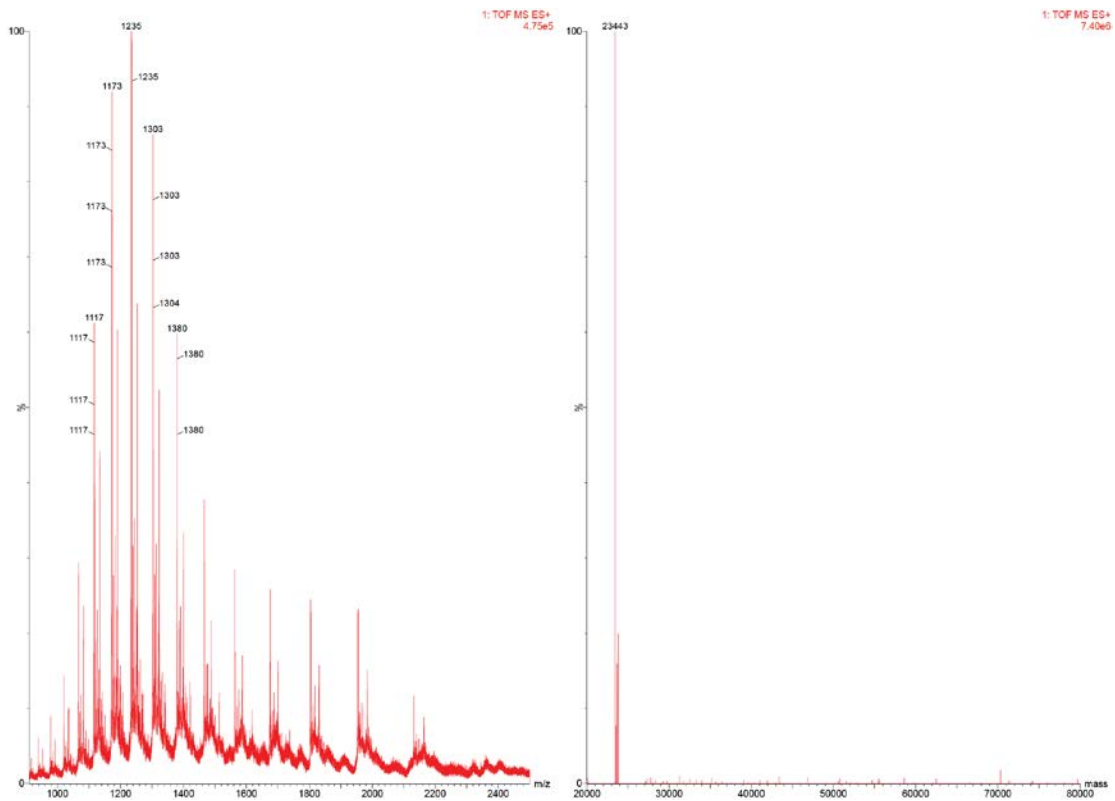
Supporting Figure 15. Combined ion series and deconvoluted mass spectrum of the light-chain of Trastuzumab.
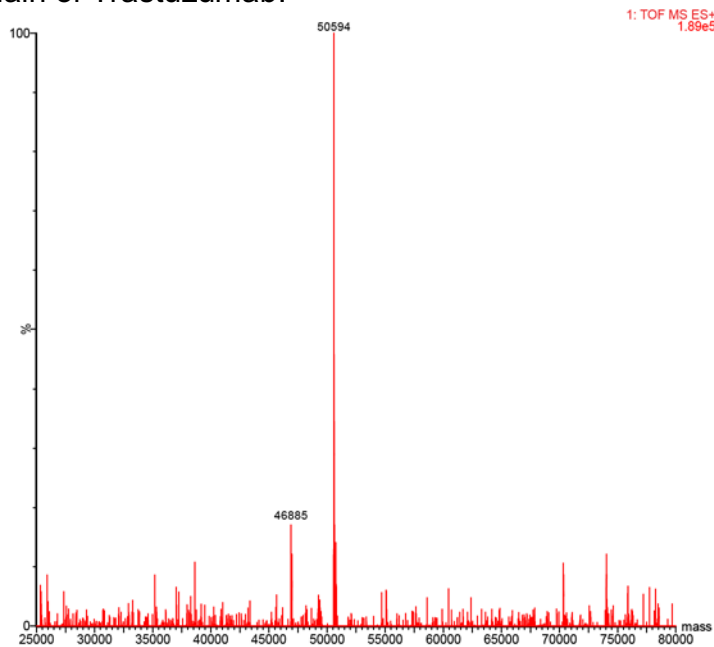
Supporting Figure 16. Deconvoluted mass spectrum of the heavy-chain of Trastuzumab.

5. Reactions and characterization of rHSA-conjugates

Reaction of rHSA with 1a
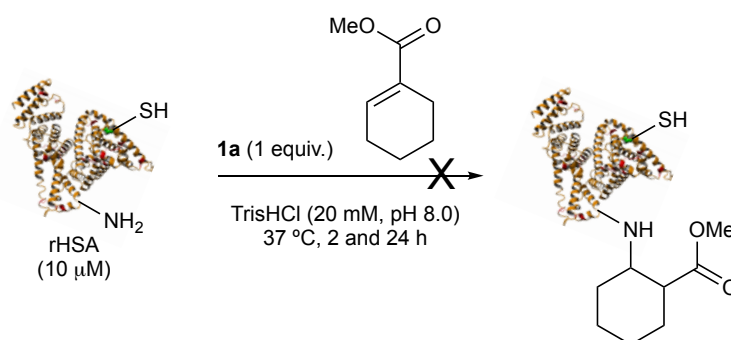
The reaction was performed according to the general procedure. To an eppendorf with 9.4 μL of TrisHCl (20 mM, pH 8.0) and 3.4 μL of DMF, was added a 26.6 μL aliquot of a stock solution of rHSA (15.05 μM) and the resulting mixture was vortexed for 10 seconds. Afterwards, a 0.71 mM solution of methyl cyclohex-1-ene-1-carboxylate **1a** (0.6 μL, 1 equiv.) in DMF was added and the reaction mixed for 2 and 24 h at 37 °C. At each reaction time, a 10 μL aliquot was analysed by LC–MS and no conversion to the expected product was observed (calculated mass, 66580 Da; observed mass, 66440 Da). The same occurred when using 2, 10 and 50 equiv. of **1a** and leaving the reaction for 24 h.
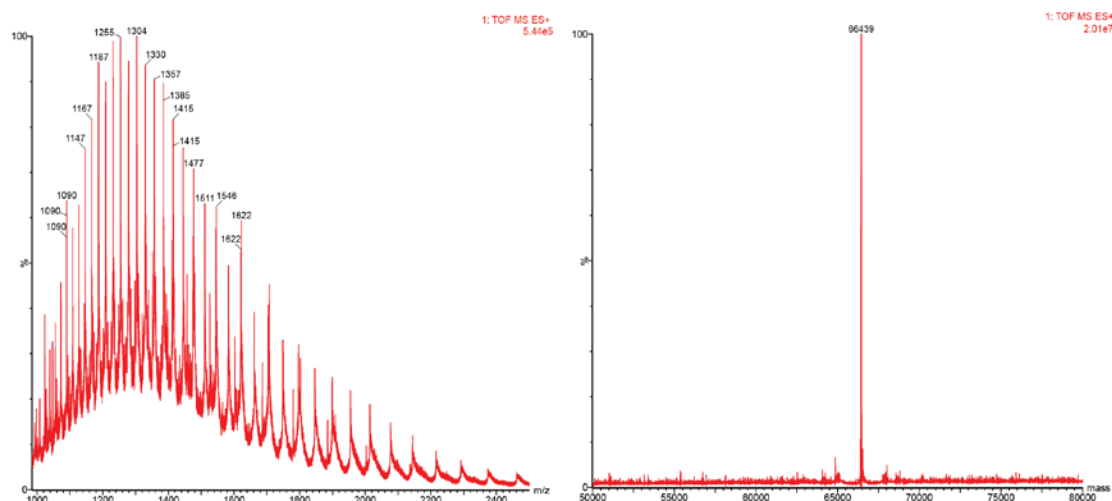
Supporting Figure 17. Combined ion series and deconvoluted mass spectrum of the reaction of rHSA (10 μM) with 1 equiv. of **1a** after 2 h at 37 °C. Identical data was obtained when using 2, 10 or 50 equiv. of **1a** even after 24 h.
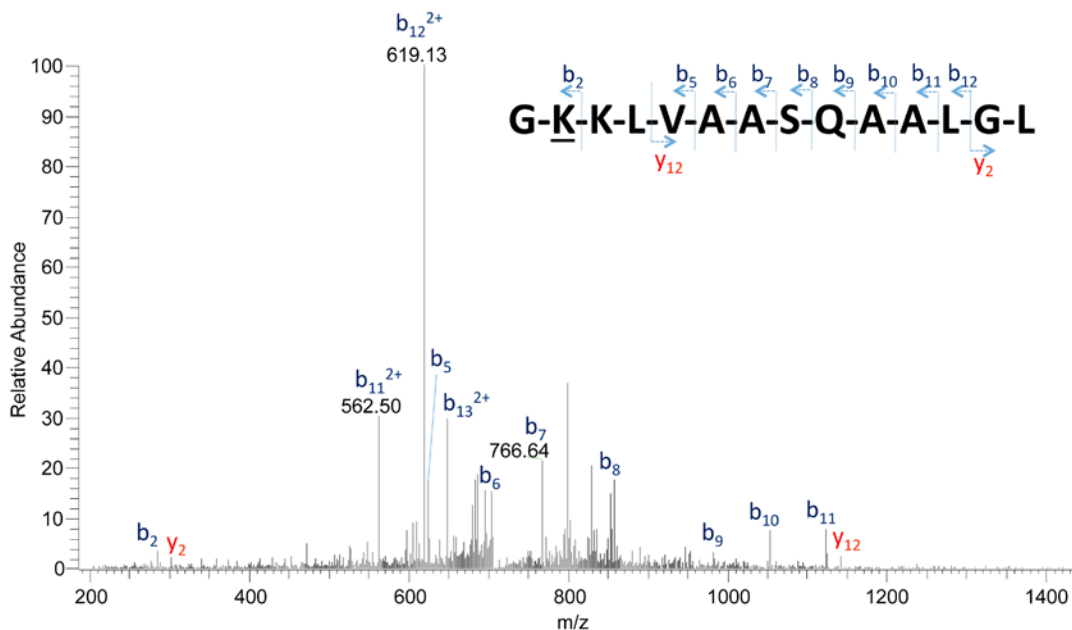
Reaction of rHSA with 1b
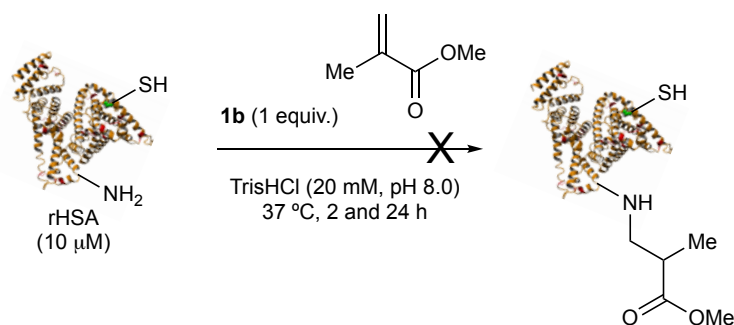
The reaction was performed according to the general procedure. To an eppendorf with 9.4 μL of TrisHCl (20 mM, pH 8.0) and 3.6 μL of DMF, was added a 26.6 μL aliquot of a stock solution of rHSA (15.05 μM) and the resulting mixture was vortexed for 10 seconds. Afterwards, a 0.99 mM solution of methylmethacrylate **1b** (0.4 μL, 1 equiv.) in DMF was added and the reaction mixed for 2 and 24 h at 37 °C. At each reaction time, a 10 μL aliquot was analysed by LC–MS and no conversion to the expected product was observed (calculated mass, 66541 Da; observed mass, 66439 Da). The same occurred when using 2, 10 and 50 equiv. of **1b** and leaving the reaction for 24 h.
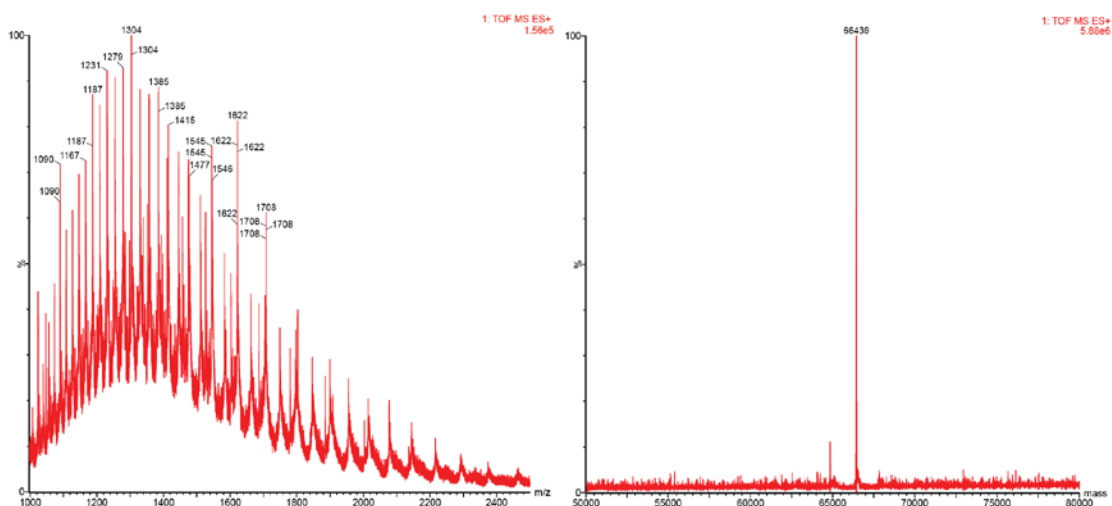
Supporting Figure 18. Combined ion series and deconvoluted mass spectrum of the reaction of rHSA (10 μM) with 1 equiv. of **1b** after 2 h at 37 °C.

Reaction of rHSA with 1c
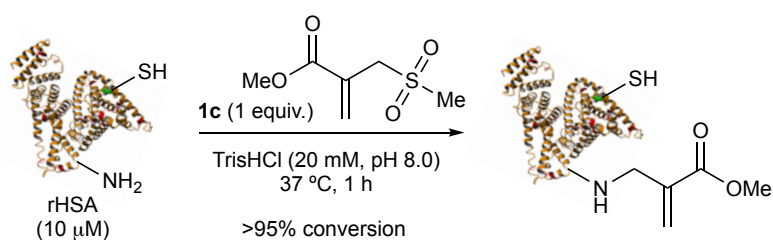
The reaction was performed according to the general procedure. To an eppendorf with 9.4 μL of TrisHCl (20 mM, pH 8.0) and 3.3 μL of DMF, was added a 26.6 μL aliquot of a stock solution of rHSA (15.05 μM) and the resulting mixture was vortexed for 10 seconds. Afterwards, a 0.56 mM solution of methyl 2-(methylsulfonyl)acrylate **1c** (0.7 μL, 1 equiv.) in DMF was added and the reaction mixed for 1, 2 and 24 h at 37 °C. At each reaction time, a 10 μL aliquot was analysed by LC–MS and complete conversion to the expected product was observed (calculated mass, 66536 Da; observed mass, 66537 Da).
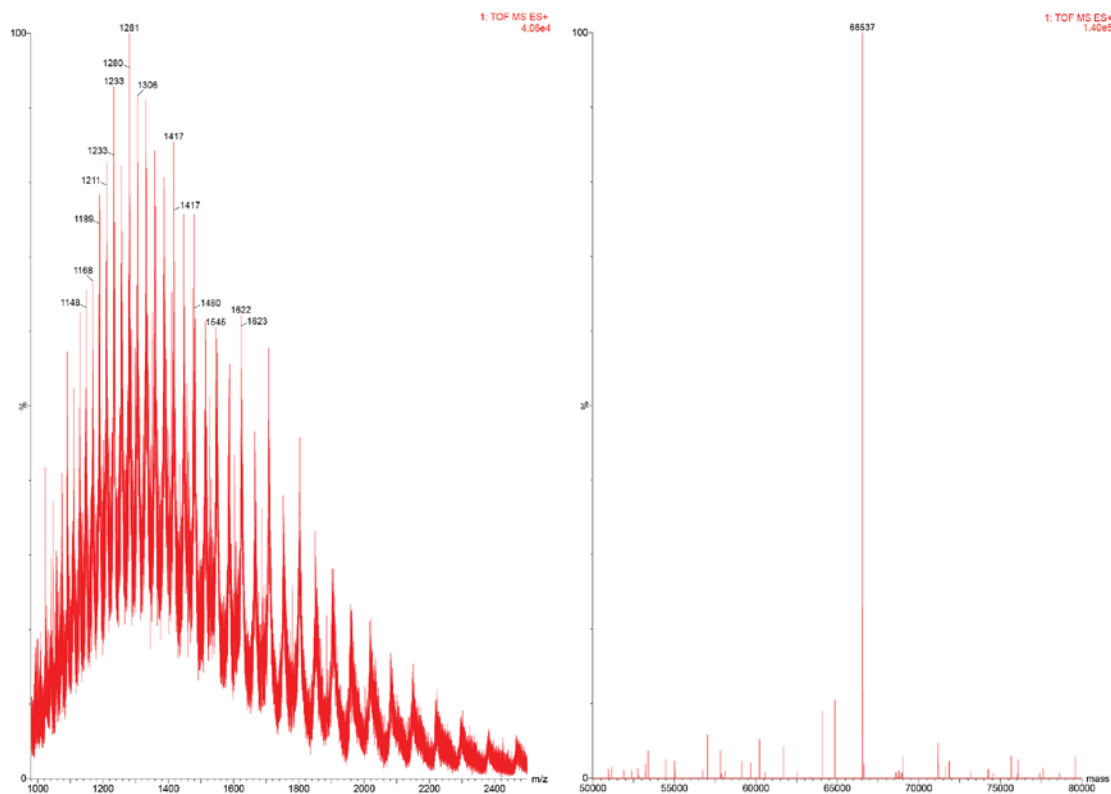
Supporting Figure 19. Combined ion series and deconvoluted mass spectrum of the reaction of rHSA (10 μM) with 1 equiv. of **1c** after 1 h at 37 ºC. Identical data was obtained at 2 and 24 h.

The same procedure was repeated using 10 and 50 equiv. of **1c**. Masses compatible with multiple modifications were found, as shown in **Supporting Figures 20 and 21**.
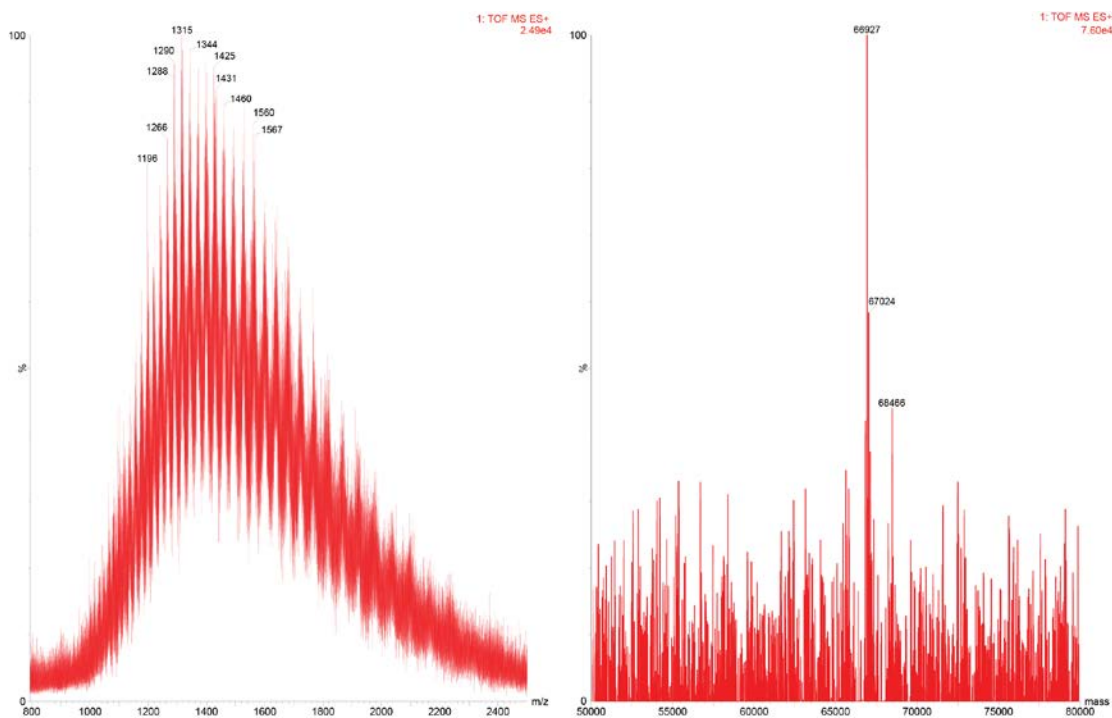
Supporting Figure 20. Combined ion series and deconvoluted mass spectrum of the reaction of rHSA (10 μM) with 10 equiv. of **1c** after 1 h at 37 ºC.
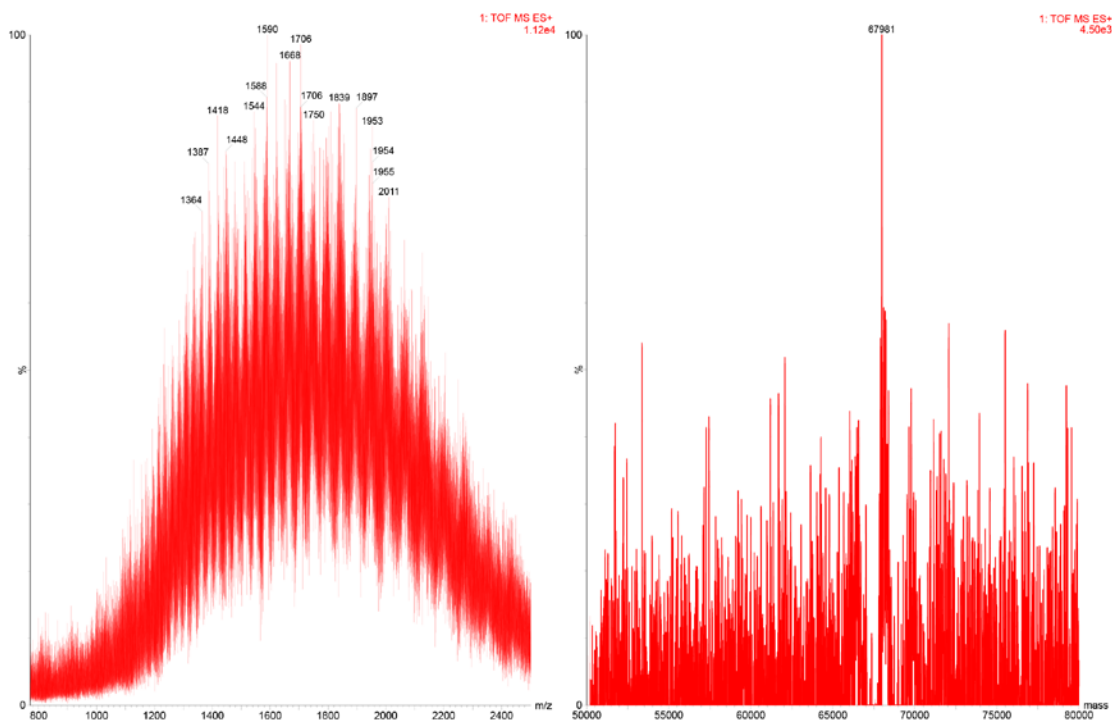
Supporting Figure 21. Combined ion series and deconvoluted mass spectrum of the reaction of rHSA (10 μM) with 50 equiv. of **1c** after 1 h at 37 ºC.

Supporting Table 3. Optimisation of reaction conditions between rHSA and

1c with respect to pH, buffer and time.

Reaction conditions		Conversion (%) rHSA–1c (10 μM, 37 ºC)				
Buffer	pH	Time (h)				
		½	1	2	3	4
Tris HCl 20	6.0	30	30	30	-	-
mM	7.0	30	40	40	-	-
	8.0	80	**100**	100	100	100
	9.0	37	42	55	-	-
NaPi 50 mM	8.0	40	50	65	70	95
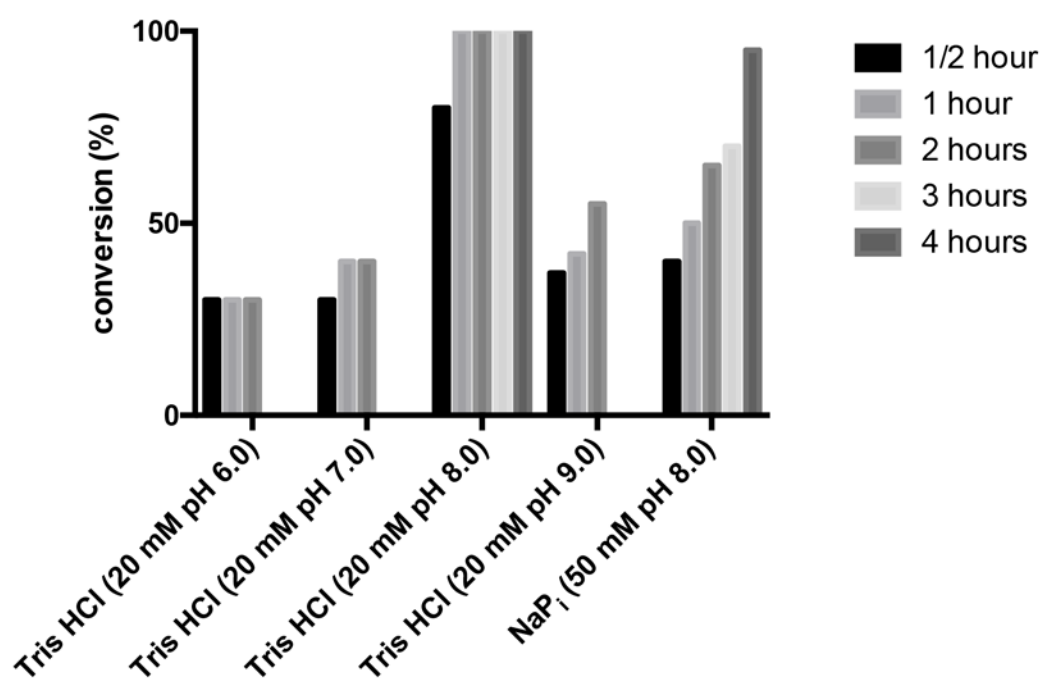
Supporting Figure 22. Graphical representation of the reaction of rHSA (10 μM) and 1 equiv. of **1c** using TrisHCl, pH from 6.0 to 10.0 (10% DMF total volume), or NaP$_i$ at pH 8.0, 37 °C, after 30 min, 1 and 2 h, respectively.
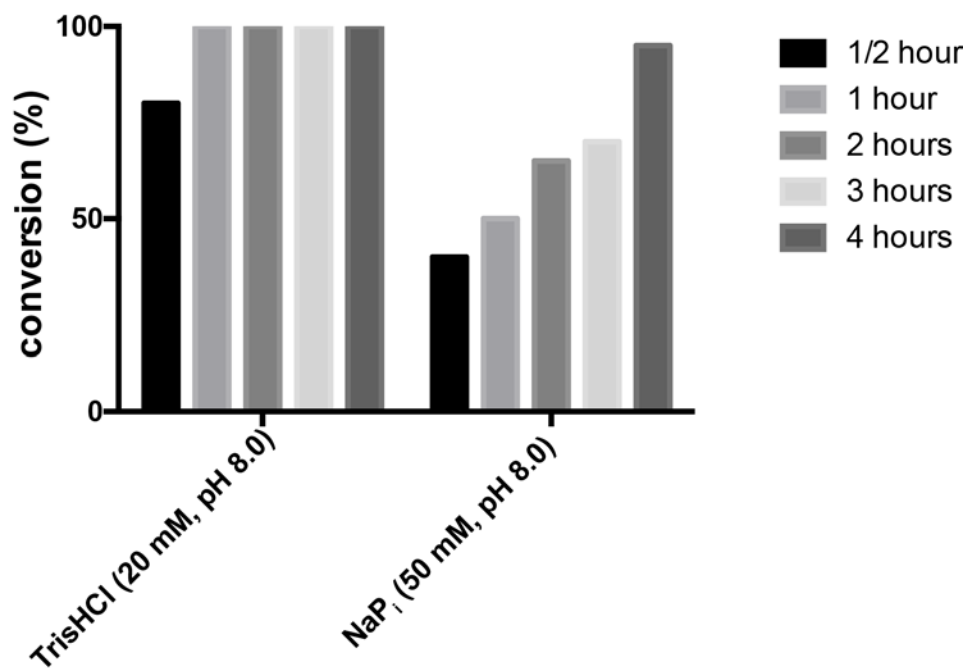
Supporting Figure 23. Graphical representation of the reaction of rHSA (10 μM) and 1 equiv. of **1c**, NaP$_i$ at pH 8.0 (10% DMF total volume), 37 °C, after 30 min, 1, 2, 3 and 4 h, respectively. Reactivity was compared to the reaction in TrisHCl (20 mM, pH 8.0).

Reaction of rHSA with 1d
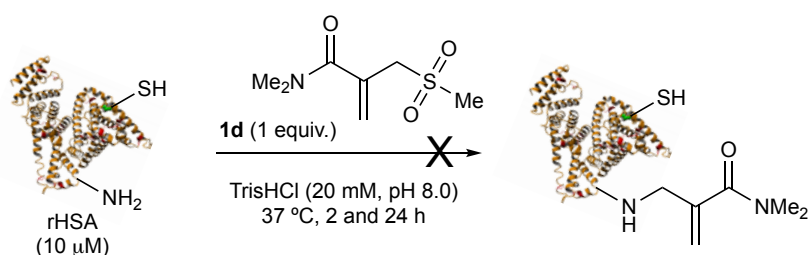
The reaction was performed according to the general procedure. To an eppendorf with 9.4 μL of TrisHCl (20 mM, pH 8.0) and 3.24 μL of DMF, was added a 26.6 μL aliquot of a stock solution of rHSA (15.05 μM) and the resulting mixture was vortexed for 10 seconds. Afterwards, a 0.56 mM solution of *N,N*-dimethyl-2-[(methylsulfonyl)methyl]acrylamide **1d** (0.76 μL, 1 equiv.) in DMF was added and the reaction mixed for 2 and 24 h at 37 ºC. At each reaction time, a 10 μL aliquot was analysed by LC–MS and no conversion to the expected product was observed (calculated mass, 66553 Da; observed mass, 66441 Da). The same occurred when using 2, 10 and 50 equiv. of **1d** and leaving the reaction for 24 h.
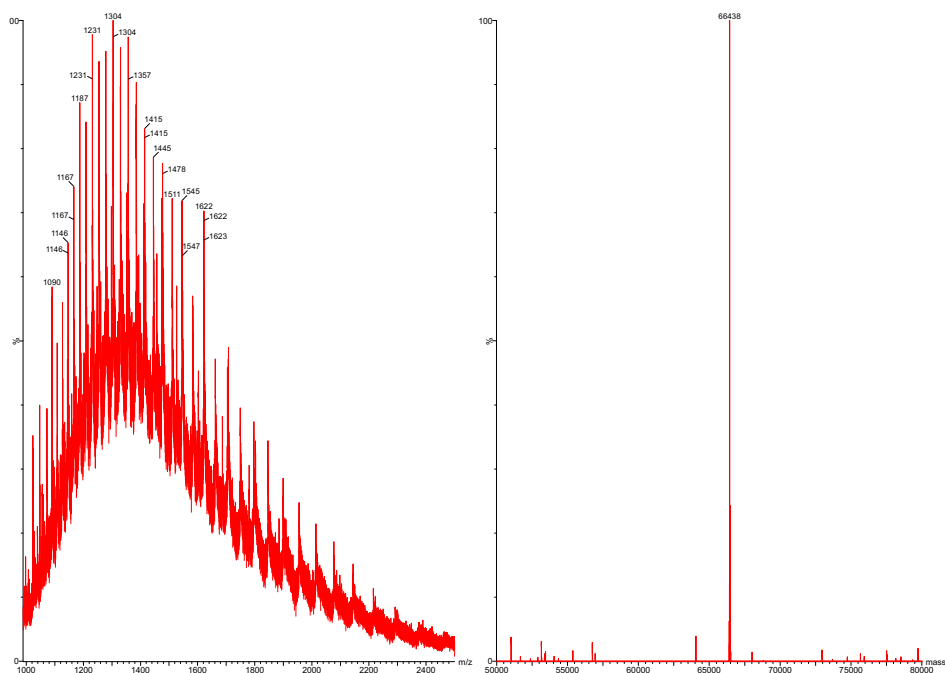
Supporting Figure 24. Combined ion series and deconvoluted mass spectrum of the reaction of rHSA (10 μM) with 1 equiv. of **1d** after 2 h at 37 ºC.

Stability of rHSA–1c in human plasma

A 20 μL aliquot of rHSA–**1c** (10 μM) in 20 mM TrisHCl buffer at pH 8.0 was thawed. 1 μL of reconstituted human plasma (Sigma-Aldrich) was added at room temperature and the resulting mixture vortexed for 10 seconds. The resulting reaction mixture was then mixed at 37 °C. After 1 and 48 h, a 10 μL aliquot of each reaction mixture was analysed by LC–MS. No significant degradation of the adduct was observed at either time point.
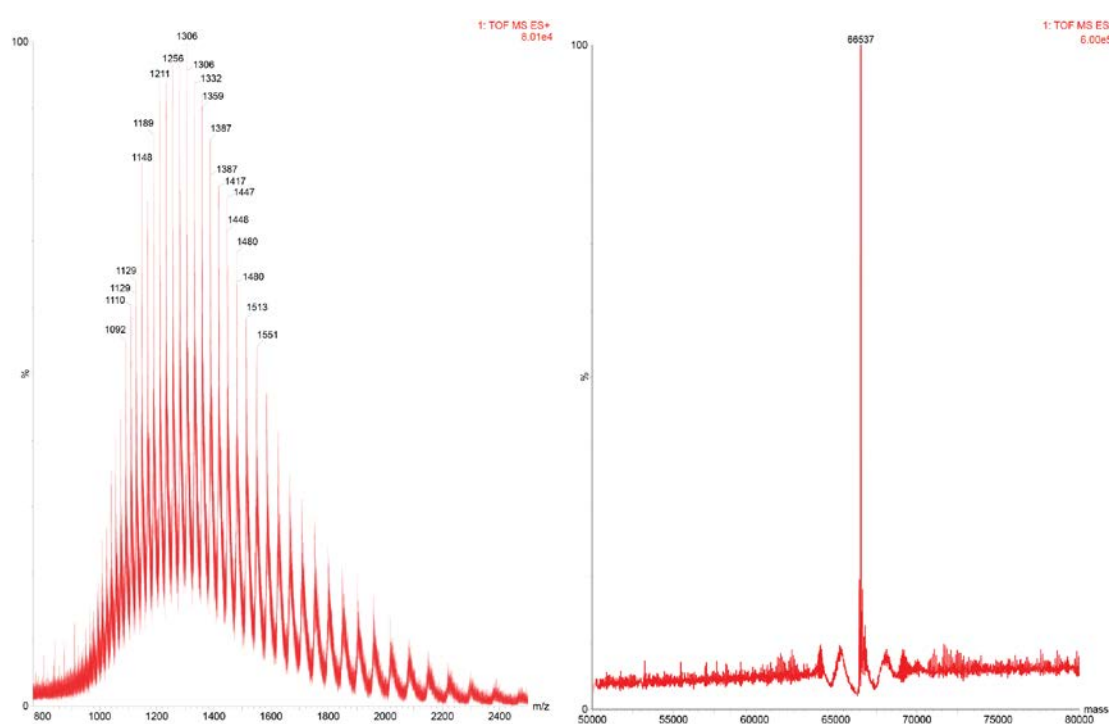
Supporting Figure 25. Combined ion series and deconvoluted mass spectrum of rHSA–**1c** after incubation in human plasma for 48 h at 37 ºC.

Stability of rHSA–1c in the presence of GSH (1 mM)

A 20 μL aliquot of rHSA–**1c** (10 μM) in TrisHCl (20 mM, pH 8.0) was thawed. 1 μL of a 20 mM glutathione solution (6 mg glutathione dissolved in 1 mL of TrisHCl (20 mM, pH 8.0)) was added at room temperature and the resulting mixture vortexed for 10 seconds. The resulting reaction mixture was then mixed at 37 ºC. After 1 and 48 h, a 10 μL aliquot of each reaction mixture was analysed by LC–MS. No significant degradation of the adduct was observed at either time points.
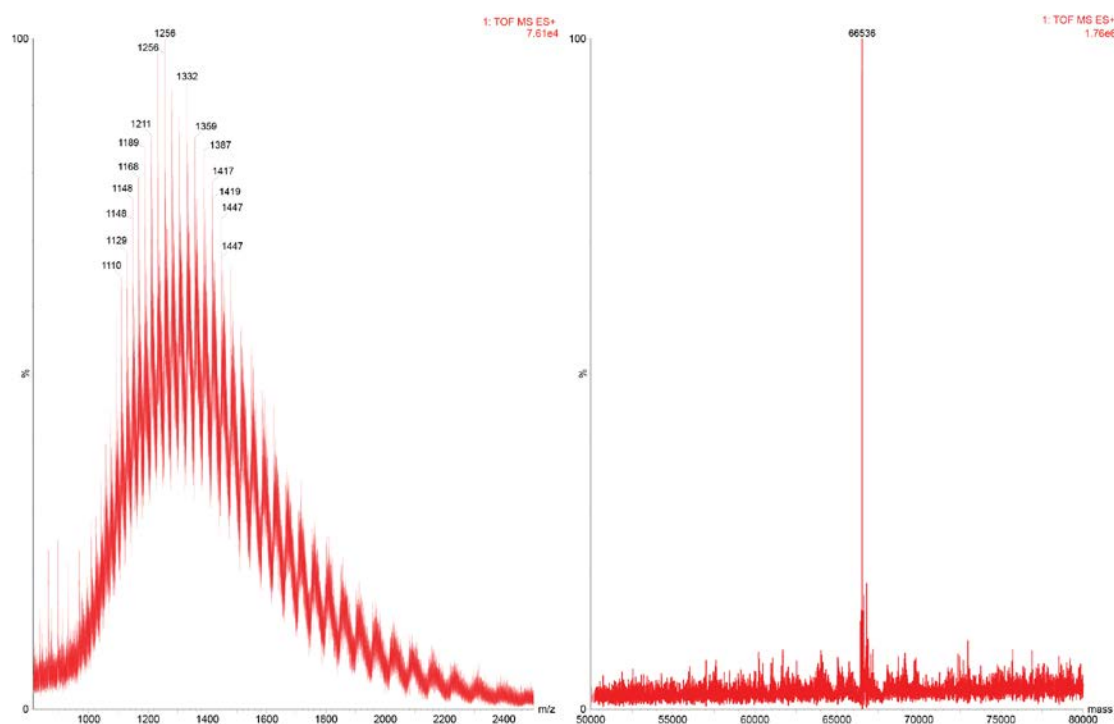
Supporting Figure 26. Combined ion series and deconvoluted mass spectrum of rHSA–**1c** after incubation with GSH for 48 h at 37 ºC.

Reaction of rHSA–1c with Ellman's reagent

rHSA-**1c**
(10 μM)
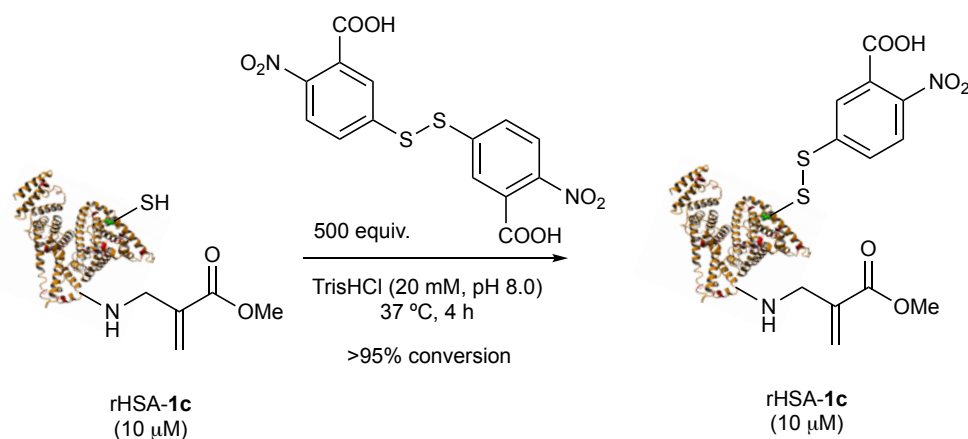
rHSA-**1c**
(10 μM)

A 40 μL aliquot of rHSA–**1c** (10 μM) was transferred to a 0.5 mL eppendorf tube. An aliquot of 8.0 μL (500 equiv.) of a stock suspension of Ellman's reagent (50.5 μM) was added and the resulting mixture vortexed for 10 seconds. After 4 h of additional mixing, at 37 °C, small molecules were removed from the reaction mixture by loading the sample into a Zeba Spin Desalting Column previously equilibrated with TrisHCl (20 mM, pH 8.0). The sample was eluted *via* centrifugation (2 min, 1000xg). A 10 μL aliquot was analysed by LC–MS and full conversion to the expected mixed disulfide conjugate was observed (calculated mass, 66736 Da; observed mass, 66735 Da).

The use of same conditions but performing the reactions in reverse order (i.e., Ellman's mixed disulfide formation followed by reaction with **1c**) gave identical results.
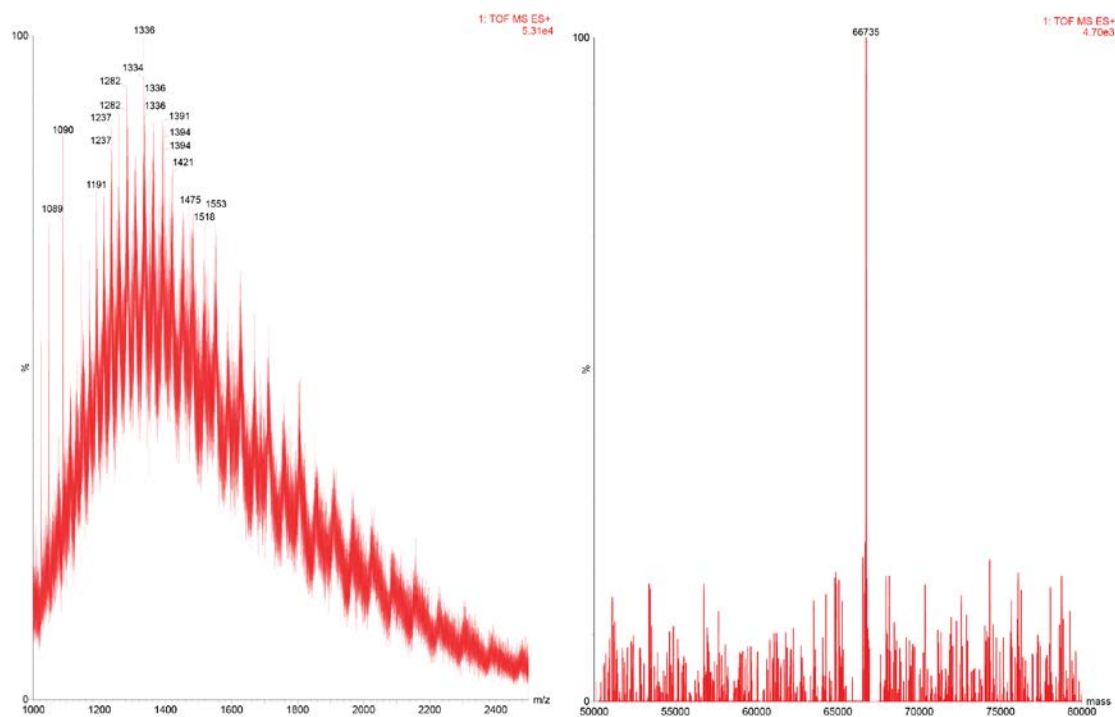
Supporting Figure 27. Combined ion series and deconvoluted mass spectrum of rHSA–**1c** with Ellman's reagent (500 equiv.) for 4 h at 37 ºC.

Reaction of rHSA-K573P with 1c
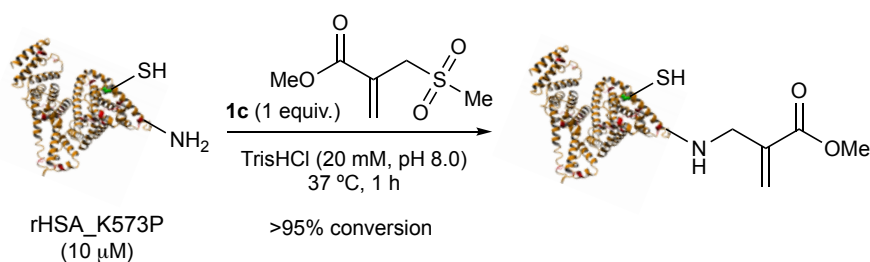
The reaction was performed according to the general procedure. To an eppendorf with 87.35 μL of TrisHCl (20 mM, pH 8.0) and 9.64 μL of DMF, was added a 2.65 μL aliquot of a stock solution of rHSA-K573P (375 μM) and the resulting mixture was vortexed for 10 seconds. Afterwards, a 5.59 mM solution of methyl 2-(methylsulfonyl)acrylate **1c** (0.18 μL, 1 equiv.) in DMF was added and the reaction mixed for for 1, 2 and 24 h at 37 ºC. At each reaction time, a 10 μL aliquot was analysed by LC–MS and complete conversion to the expected product was observed after 1 h (calculated mass, 66507 Da; observed mass, 66506 Da).
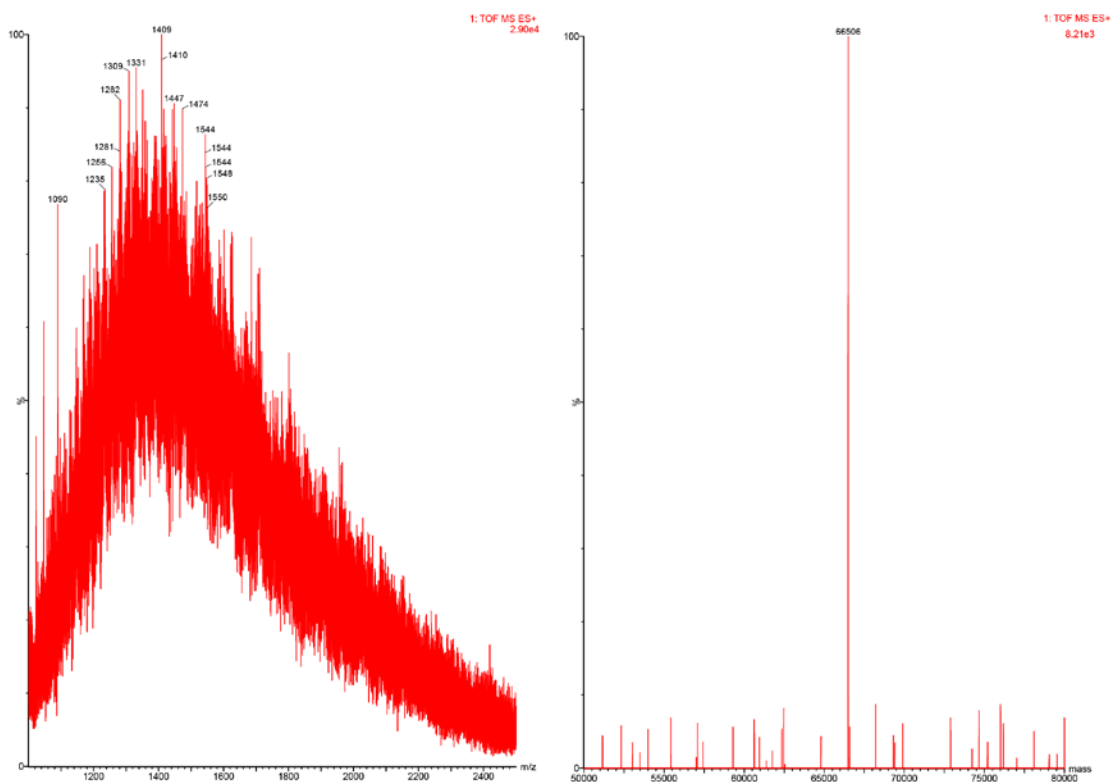
**Supporting Figure 28.** Combined ion series and deconvoluted mass spectrum of the reaction of rHSA-K573P (10 µM) with 1 equiv. of **1c** after 1 h at 37 ºC. Identical data was obtained at 2 and 24 h.

Enzymatic digestion and LC–MS/MS analysis of rHSA–1c

rHSA (59 Lys, 1 free Cys, 17 disulfides)

Lys 573 – found experimentally

Lys64 – reported by Barbas and co-workers.[20]

Sequence:

```
DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV KLVNEVTEFA  50
KTCVADESAE NCDKSLHTLF GDKLCTVATL RETYGEMADC CAKQEPERNE  100
CFLQHKDDNP NLPRLVRPEV DVMCTAFHDN EETFLKKYLY EIARRHPYFY  150
APELLFFAKR YKAAFTECCQ AADKAACLLP KLDELRDEGK ASSAKQRLKC  200
ASLQKFGERA FKAWAVARLS QRFPKAEFAE VSKLVTDLTK VHTECCHGDL  250
LECADDRADL AKYICENQDS ISSKLKECCE KPLLEKSHCI AEVENDEMPA  300
DLPSLAADFV ESKDVCKNYA EAKDVFLGMF LYEYARRHPD YSVVLLLRLA  350
KTYETTLEKC CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE  400
YKFQNALLVR YTKKVPQVST PTLVEVSRNL GKVGSKCCKH PEAKRMPCAE  450
DYLSVVLNQL CVLHEKTPVS DRVTKCCTES LVNRRPCFSA LEVDETYVPK  500
EFNAETFTFH ADICTLSEKE RQIKKQTALV ELVKHKPKAT KEQLKAVMDD  550
FAAFVEKCCK ADDKETCFAE EG**K**KLVAASQ AALGL 585
```


Supporting Figure 29. MS/MS spectrum of the *m/z* 712.93 doubly charged ion of the lysine modified peptide GKKLVAASQAALGL from rHSA. The underscore relates to the modified amino acid.

Enzymatic digestion and LC–MS/MS analysis of rHSA-K573P–1c

rHSA-K573P (58 Lys, 1 free Cys, 17 disulfides)

Lys 4 – found experimentally

Sequence:

```
DAHKSEVAHR  FKDLGEENFK  ALVLIAFAQY  LQQCPFEDHV  KLVNEVTEFA  50
KTCVADESAE  NCDKSLHTLF  GDKLCTVATL  RETYGEMADC  CAKQEPERNE  100
CFLQHKDDNP  NLPRLVRPEV  DVMCTAFHDN  EETFLKKYLY  EIARRHPYFY  150
APELLFFAKR  YKAAFTECCQ  AADKAACLLP  KLDELRDEGK  ASSAKQRLKC  200
ASLQKFGERA  FKAWAVARLS  QRFPKAEFAE  VSKLVTDLTK  VHTECCHGDL  250
LECADDRADL  AKYICENQDS  ISSKLKECCE  KPLLEKSHCI  AEVENDEMPA  300
DLPSLAADFV  ESKDVCKNYA  EAKDVFLGMF  LYEYARRHPD  YSVVLLLRLA  350
KTYETTLEKC  CAAADPHECY  AKVFDEFKPL  VEEPQNLIKQ  NCELFEQLGE  400
YKFQNALLVR  YTKKVPQVST  PTLVEVSRNL  GKVGSKCCKH  PEAKRMPCAE  450
DYLSVVLNQL  CVLHEKTPVS  DRVTKCCTES  LVNRRPCFSA  LEVDETYVPK  500
EFNAETFTFH  ADICTLSEKE  RQIKKQTALV  ELVKHKPKAT  KEQLKAVMDD  550
FAAFVEKCCK  ADDKETCFAE  EGPKLVAASQ  AALGL  585
```


Supporting Figure 30. MS/MS spectrum of the *m/z* 416.55 triply charged ion of the lysine modified peptide DAHKDEVAHR from rHSA-K573P–**1c**. The underscore relates to the modified amino acid.

Circular dichroism of rHSA and rHSA–1c
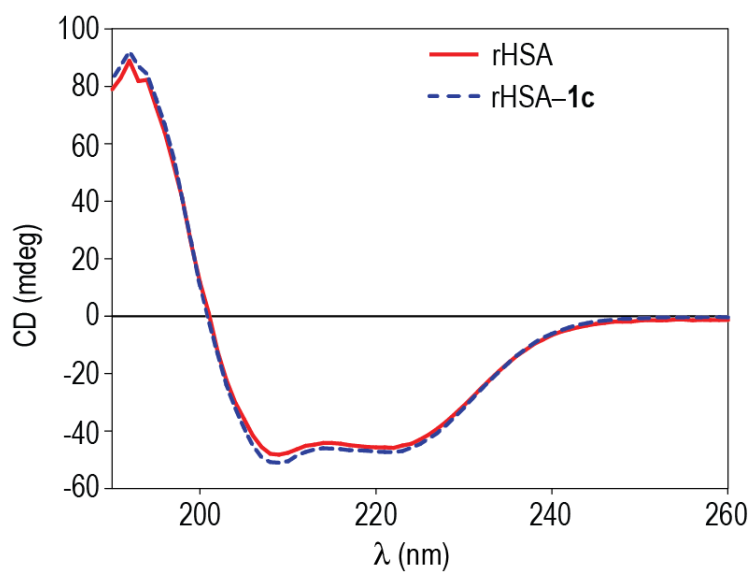
Supporting Figure 31. Structural analysis of rHSA and rHSA–**1c** by CD.

Circular dichroism of rHSA-K573P and rHSA-K573P–1c
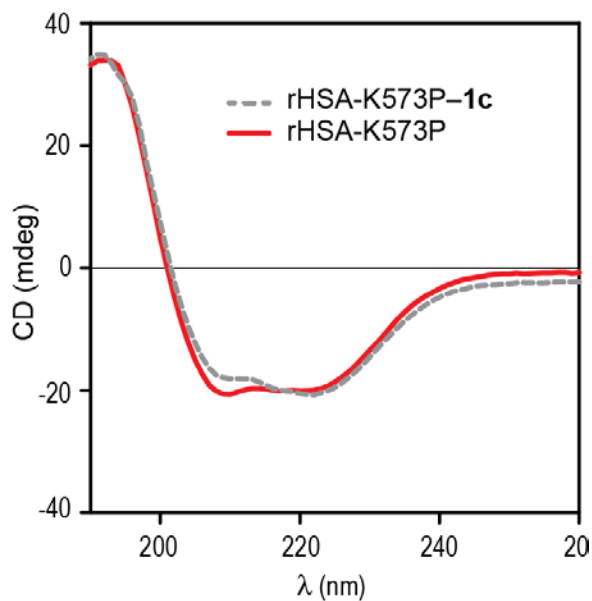
Supporting Figure 32. Structural analysis of rHSA-K573P and rHSA-K573P–**1c** by CD.

Surface plasmon resonance analysis of rHSA–1c and rHSA-K573P–1c.

SPR experiments were performed using a Biacore 3000 instrument (GE Healthcare). Flow cells of CM5 sensor chips were coupled with soluble human FcRn (1505 RU) using amine coupling chemistry as described in the protocol provided by the manufacturer (GE Healthcare). The coupling was performed by injecting 10 μg/mL of the protein in 10 mM sodium acetate pH 4.5 (GE healthcare). Phosphate buffer (25mM sodium acetate, 25 mM NaH_2PO_4, 150mM NaCl, 0.01% T-20, pH 5.5) was used as running buffer and dilution buffer. Regeneration of the surfaces were performed using injections of HBS-EP buffer (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% surfactant P20) at pH 7.4 (GE Healthcare). Post immobilisation, the chip was left to stabilise with a constant flow (5 μL/min) of running buffer. The chip surface was conditioned by injecting 3x injections of running buffer followed by 3x injections of regeneration buffer. Surfaces were checked for activity with unmodified albumin controls. For determination of binding kinetics, serial dilutions of non-modified and modified albumins (10-0 μM) were injected over immobilized receptor at a constant flow rate (30 μL/min) at 25 °C. In all experiments, data were zero adjusted and the reference cell subtracted. Data evaluations were performed using BIAevaluation 4.1 software (GE Healthcare).
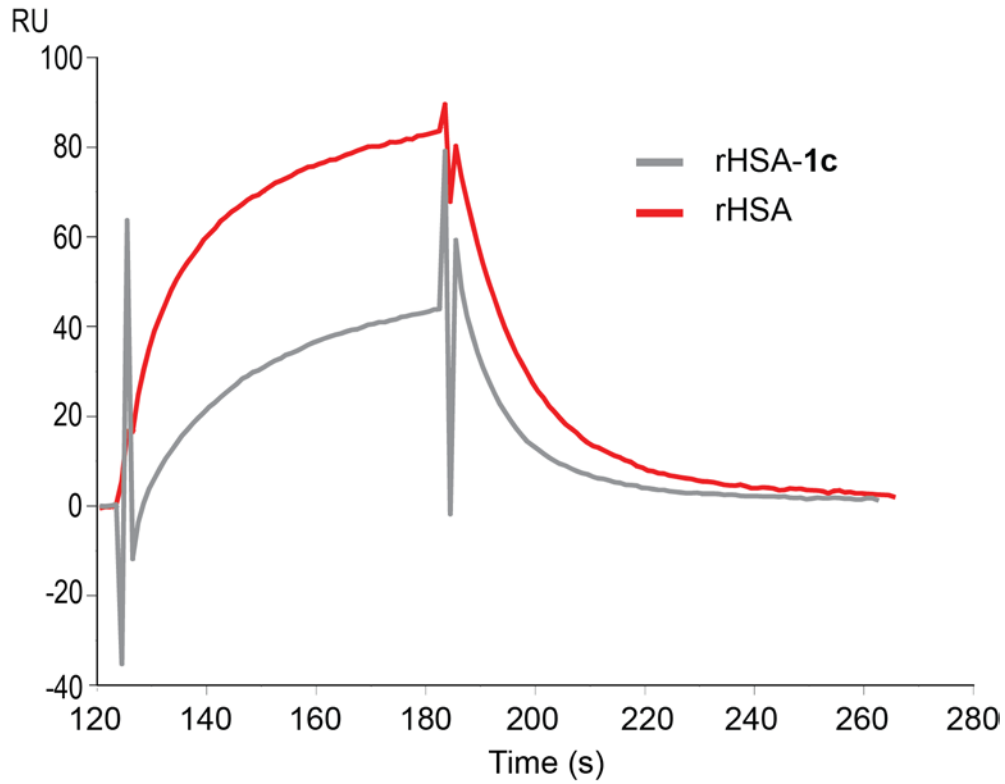
Supporting Figure 33. Biacore curves for human FcRn binding rHSA and rHSA–**1c**.
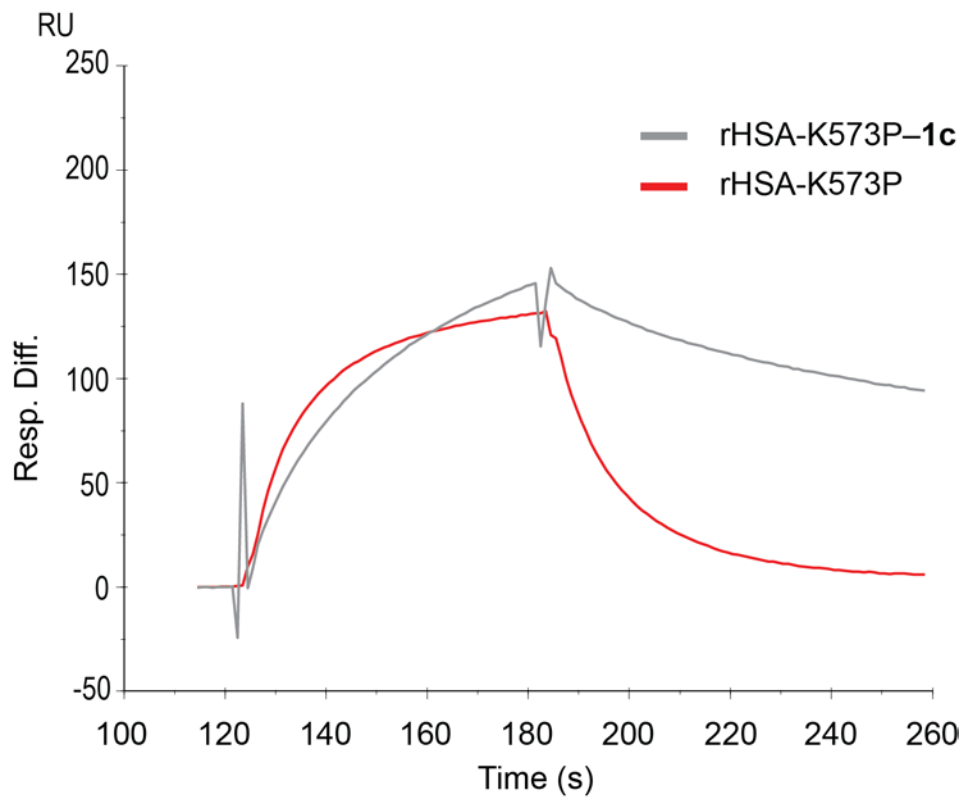
Supporting Figure 34. Biacore curves for human FcRn binding rHSA-K573P and rHSA-K573P–**1c**.

Supporting Table 4. Kinetic analysis at pH 5.5 of rHSA derivatives binding to human FcRn receptor.

Albumin	k_{on} (10^{-3}/Ms)	k_{off} (10^3/s)	k_D (µM)
rHSA	3.92	63.3	16.13
rHSA–**1c**	2.15	62.5	29.10
rHSA-K573P	9.75	4.88	0.501
rHSA-K573P–**1c**	7.92	4.65	0.587

Values are mean of duplicates. There are differences observed in FcRn binding kinetics between the albumins. For rHSA/rHSA–**1c**: FcRn binding is negatively impacted when **1c** was installed at position 573. The k_{on} for rHSA–**1c** is much slower, driving the k_D value up to 29.1 µM, which is 2x fold less than rHSA; for rHSA-K573P/ rHSA-K573P–**1c**: The k_{on} when position 4 is modified is slightly slower (7.925) compared to non-modified rHSA-K573P (9.755), pushing the affinity k_D for the human FcRn, up slightly.

Reaction of rHSA with MS(PEG)4 Methyl-PEG-NHS-Ester
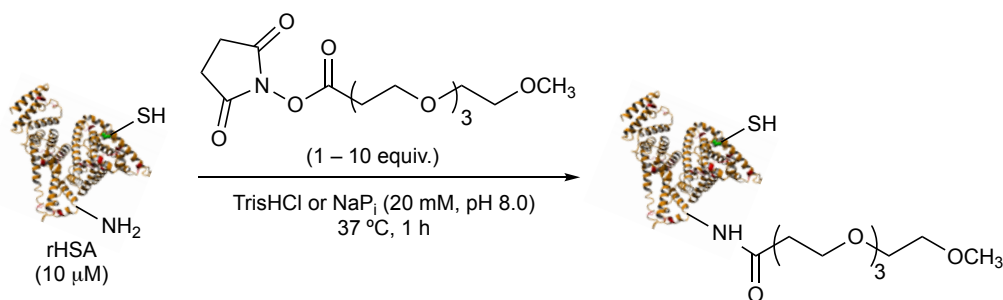
1 equiv. – incomplete conversion, heterogenous mixture of products with one and two modifications
10 equiv. – complete conversion, heterogenous mixture of products with three up to eight modifications

The reaction was performed according to the general procedure. To an eppendorf with 30,7 µL of TrisHCl (20 mM, pH 8.0) and 2.6 µL of DMF, a 5.3 µL aliquot of a stock solution of rHSA (150 µM) was added and the resulting mixture vortexed for 10 seconds. Afterwards, a 0.54 mM solution of MS(PEG)4 Methyl-PEG-NHS-Ester (1.4 µL, 1 equiv.) in DMF was added and the reaction mixed for 1 h at 37 ºC. A 10 µL aliquot was thenanalysed by LC–MS and indicating the presence of three species corresponding to the starting material (66472 Da) and conjugates bearing one (66656 Da) and two (66873 Da) modifications. Using 10 equiv. of MS(PEG)4 Methyl-PEG-NHS-Ester, a complex mixture of products with three up to eight modifications detected.
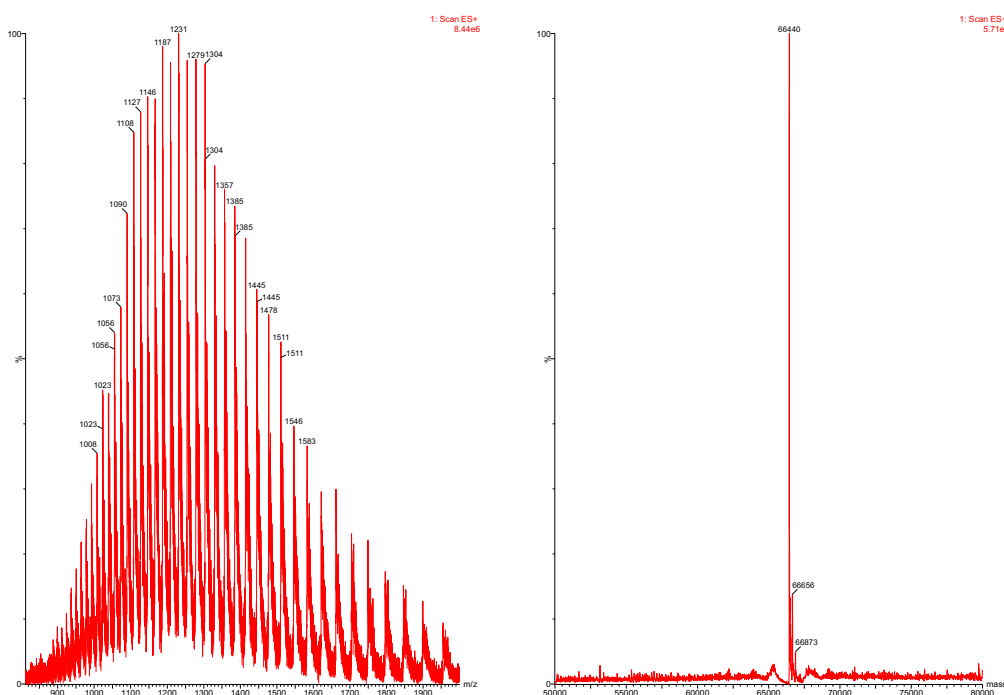
Supporting Figure 35. Combined ion series and deconvoluted mass spectrum of the reaction of rHSA (10 µM) with 1 equiv. of MS(PEG)4 Methyl-PEG-NHS-Ester after 1 h at 37 ºC. Identical data was obtained at 24 h.
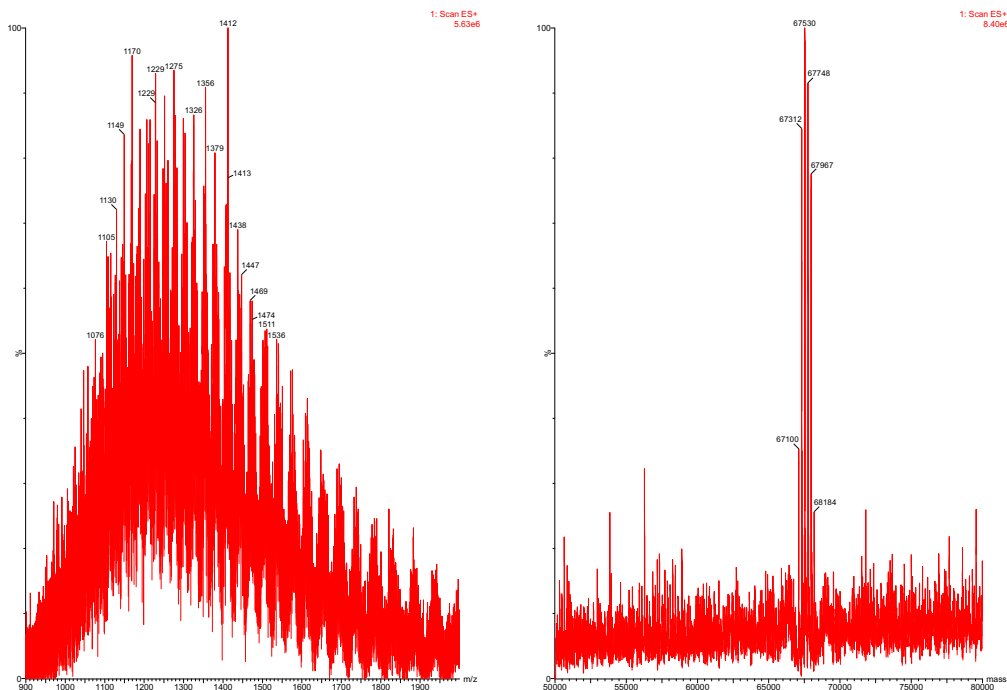
Supporting Figure 36. Combined ion series and deconvoluted mass spectrum of the reaction of rHSA (20 µM) with 10 equiv. of MS(PEG)4 Methyl-PEG-NHS-Ester after 1 h at 37 ºC.

6. Reaction of lysozyme with 1c and characterization
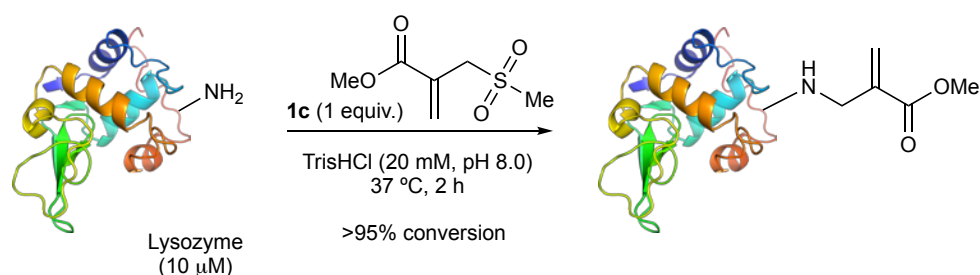
The reaction was performed according to the general procedure. To an eppendorf with 30.3 μL of TrisHCl (20 mM, pH 8.0) and 3.3 μL of DMF, was added a 5.7 μL aliquot of a stock solution of lysozyme (69.93 μM) and the resulting mixture was vortexed for 10 seconds. Afterwards, a 0.560 mM solution of methyl 2-(methylsulfonyl)acrylate **1c** (0.7 μL, 1 equiv.) in DMF was added and the reaction mixed for 2 and 24 h at 37 ºC. At each reaction time, a 10 μL aliquot was analysed by LC–MS and complete conversion to the expected product was observed (calculated mass, 14402 Da; observed mass, 14403 Da).
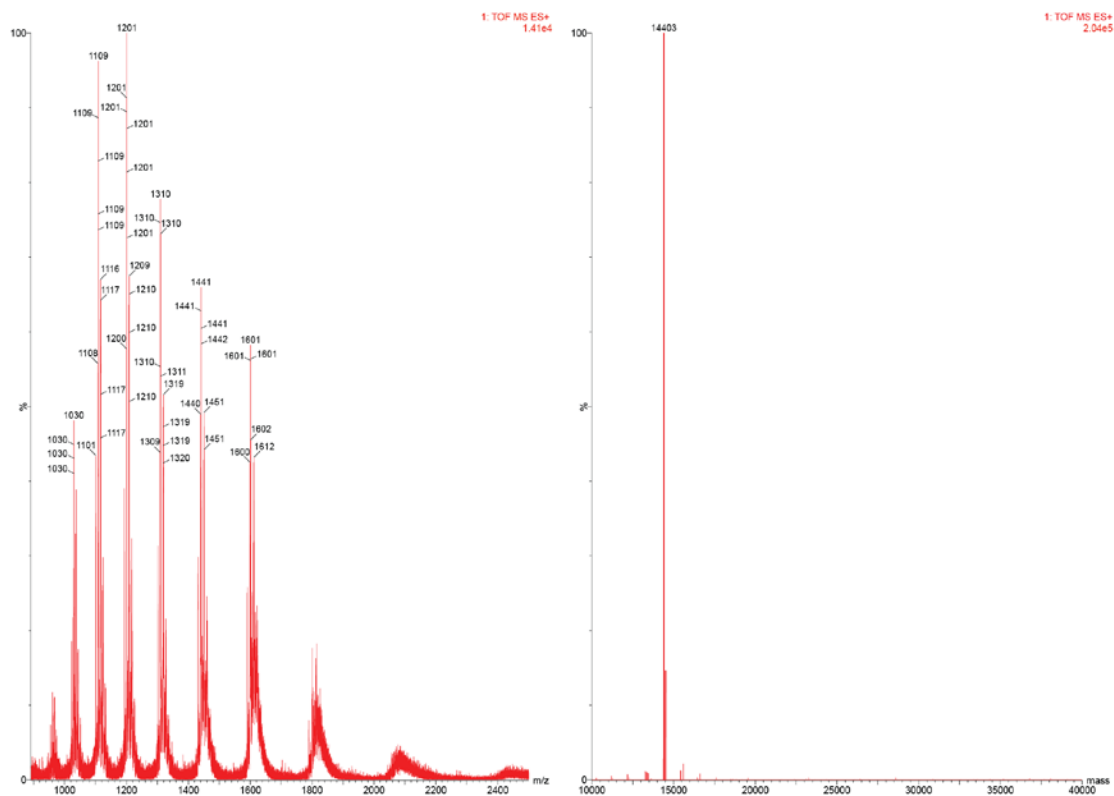
Supporting Figure 37. Combined ion series and deconvoluted mass spectrum of the reaction of lysozyme (10 µM) with 1 equiv. of **1c** after 2 h at 37 ºC. Identical data was obtained at 24 h.

7. Reactions and characterization with Annexin V-conjugates

Reaction of Annexin V with 1c
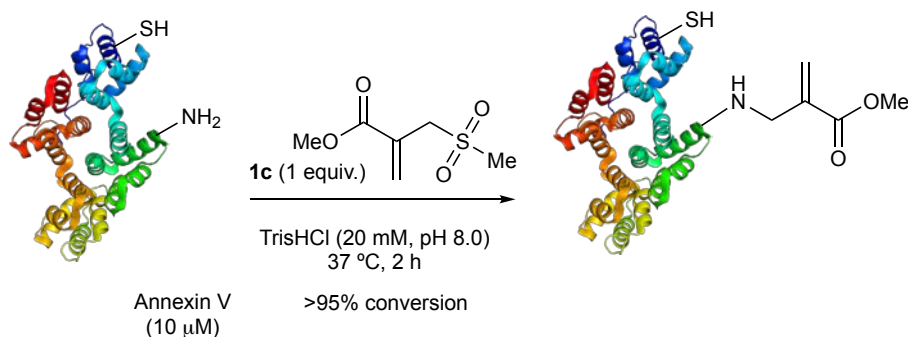
The reaction was performed according to the general procedure. To an eppendorf with 21.7 μL of TrisHCl (20 mM, pH 8.0) and 3.3 μL of DMF, was added a 14.3 μL aliquot of a stock solution of Annexin V (27.9 μM) and the resulting mixture was vortexed for 10 seconds. Afterwards, a 0.560 mM solution of **1c** (0.7 μL, 1 equiv.) in DMF was added and the reaction mixed for 2 h at 37 °C. At the end, a 10 μL aliquot was analysed by LC–MS and complete conversion to the product was observed (calculated mass, 35903 Da; observed mass, 35904 Da).
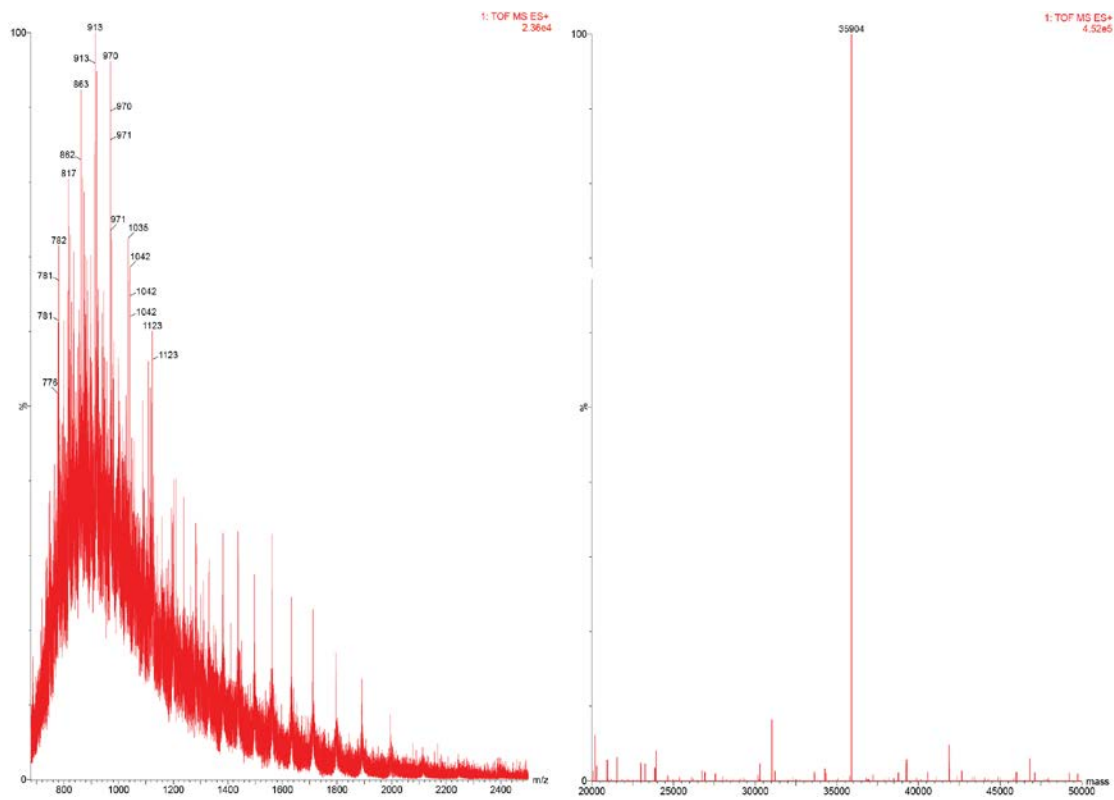
Supporting Figure 38. Combined ion series and deconvoluted mass spectrum of the reaction of Annexin V (10 μM) with 1 equiv. of **1c** after 2 h at 37 ºC. Identical data was obtained at 24 h.

Stability of Annexin V–1c in human plasma

A 20 μL aliquot of Annexin V–**1c** (10 μM) in TrisHCl (20 mM, pH 8.0) was thawed. 1 μL of reconstituted human plasma (Sigma-Aldrich) was added at room temperature and the resulting mixture vortexed for 10 seconds. The resulting reaction mixture was then mixed at 37 ºC. After 1 and 48 h, a 10 μL aliquot of each reaction mixture was analysed by LC–MS. No significant degradation of the adduct was observed at either time point.
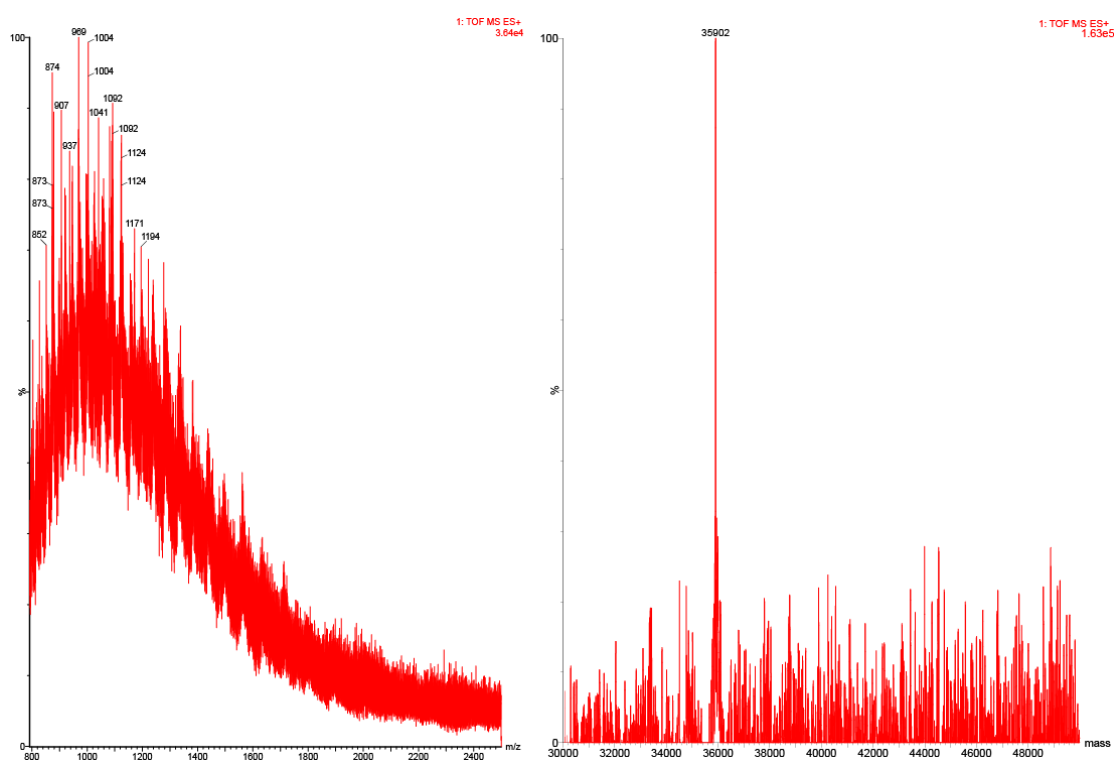
Supporting Figure 39. Combined ion series and deconvoluted mass spectrum of the reaction of Annexin V–**1c** after incubation in human plasma for 48 h at 37 ºC.

Stability of Annexin V–1c in the presence of GSH (1 mM)

A 20 μL aliquot of Annexin V–**1c** (10 μM) in TrisHCl (20 mM, pH 8.0) was thawed. 1 μL of a 20 mM glutathione solution (6 mg glutathione dissolved in 1 mL of TrisHCl (20 mM, pH 8.0) was added at room temperature and the resulting mixture vortexed for 10 seconds. The resulting reaction mixture was then mixed at 37 ºC. After 1 and 48 h, a 10 μL aliquot of each reaction mixture was analysed by LC–MS. No significant degradation of the adduct was observed at either time point.
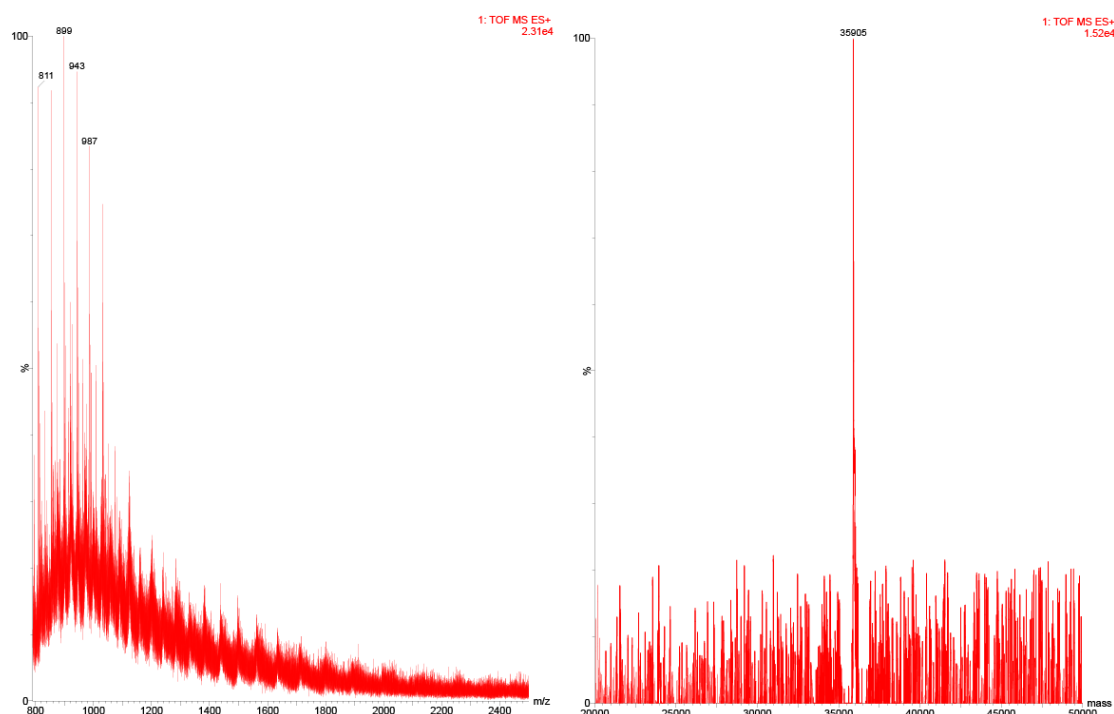
Supporting Figure 40. Combined ion series and deconvoluted mass spectrum of the reaction of Annexin V–**1c** after incubation with GSH for 48 h at 37 ºC.

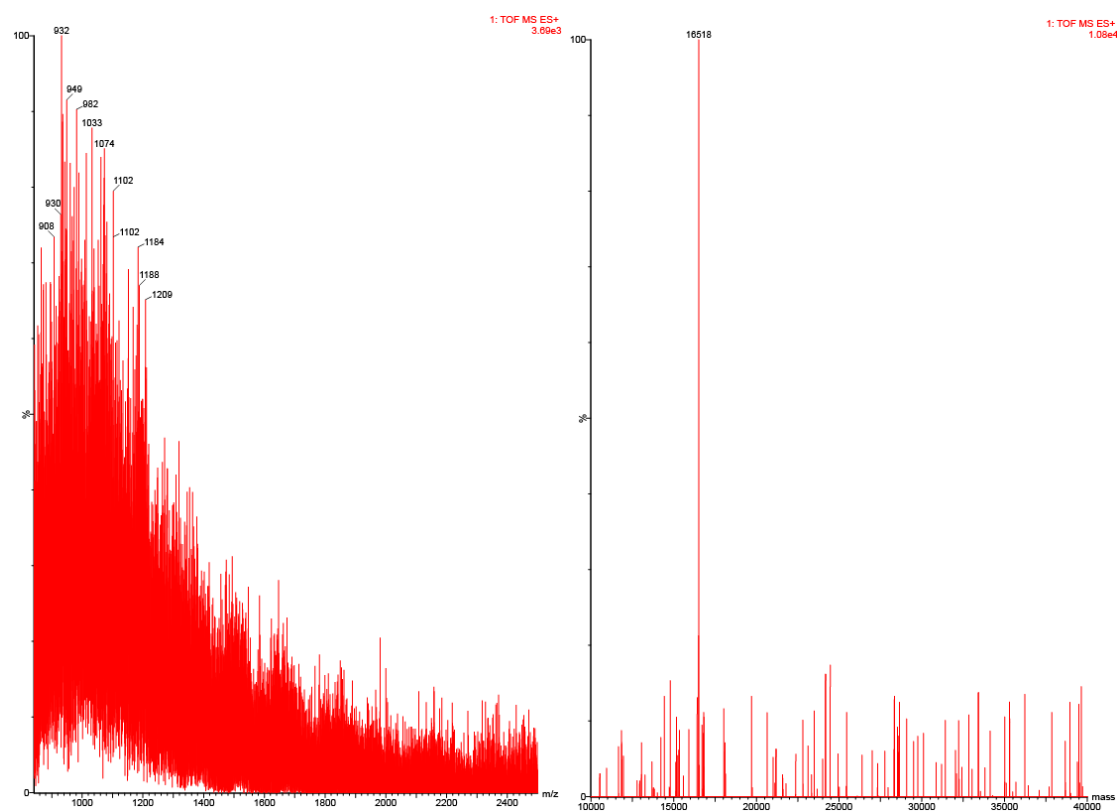## Reaction of Annexin V–1c with Ellman's reagent

A 40 μL aliquot of Annexin V–**1c** (10 μM) was transferred to a 0.5 mL eppendorf tube. An aliquot of 8.0 μL (500 equiv.) of a stock suspension of Ellman's reagent (50.5 μM) was added and the resulting mixture vortexed for 10 seconds. After 4 h of additional mixing, at 37 ºC, small molecules were removed from the reaction mixture by loading the sample into a Zeba Spin Desalting Column previously equilibrated with TrisHCl (20 mM, pH 8.0). The sample was eluted *via* centrifugation (2 min, 1000xg). A 10 μL aliquot was analysed by LC–MS and full conversion to the expected doubly modified protein was observed (calculated mass, 36103 Da; observed mass, 36094 Da). Note: in our experiments we noticed that the presence of a thiol source (GSH, ellman's) can lead to hydrolysis of the methyl ester.

The use of same conditions but performing the reactions in reverse order (i.e., Ellman's mixed disulfide formation followed by reaction with **1c** gave identical results.
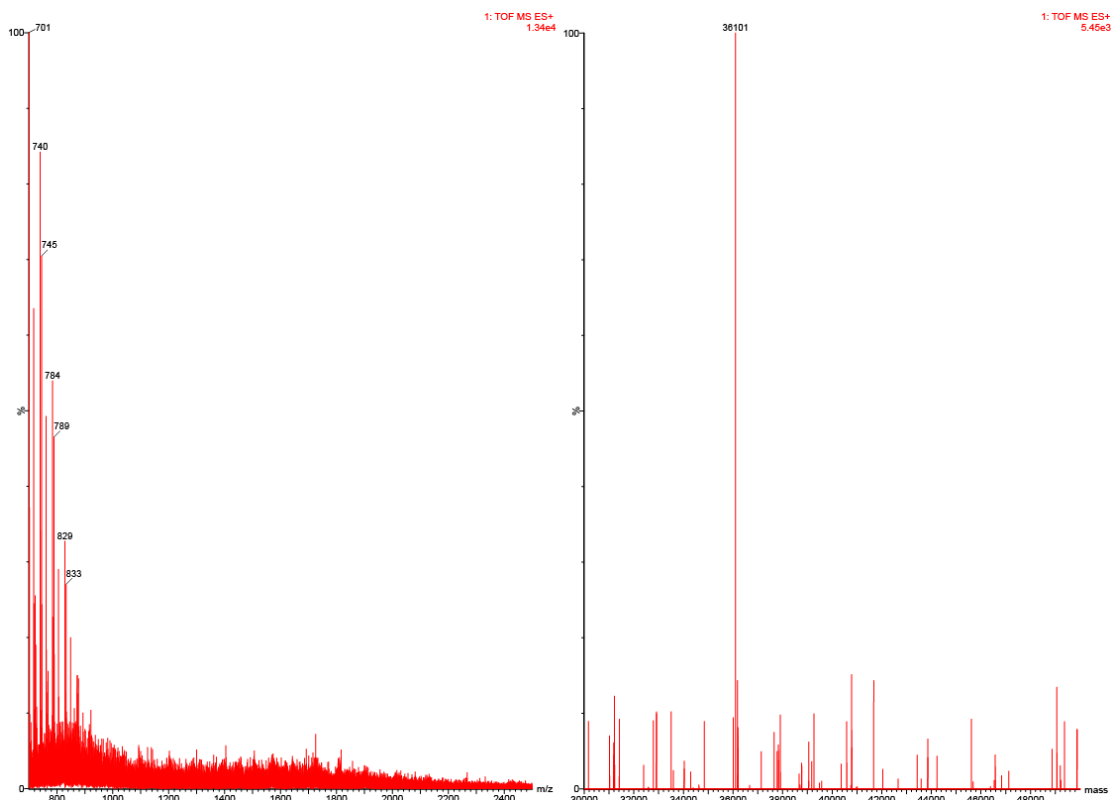
Supporting Figure 41. Combined ion series and deconvoluted mass spectrum of the reaction of Annexin V–**1c** with Ellman's reagent (500 equiv.) after 4 h at 37 ºC.
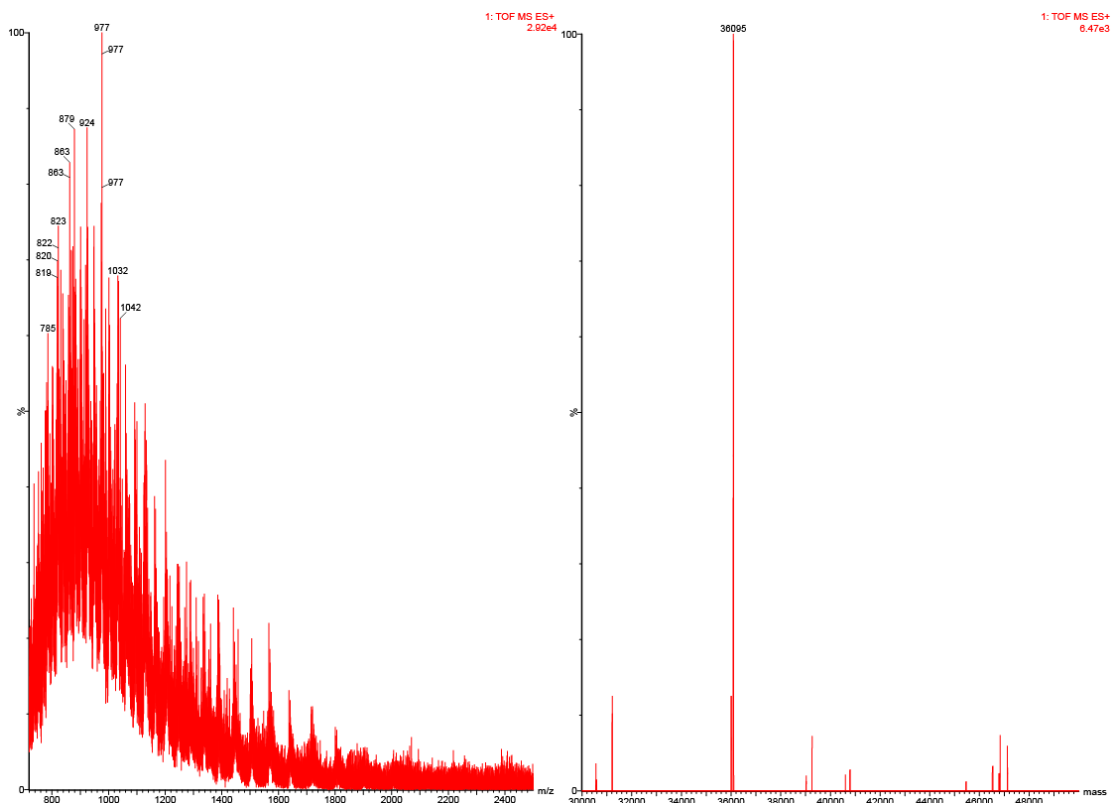
Supporting Figure 42. Combined ion series and deconvoluted mass spectrum of the reaction of Annexin V–**1c** with Ellman's reagent (500 equiv.) for 24 h at 37 ºC.

Reaction of Annexin V–1c with benzylamine
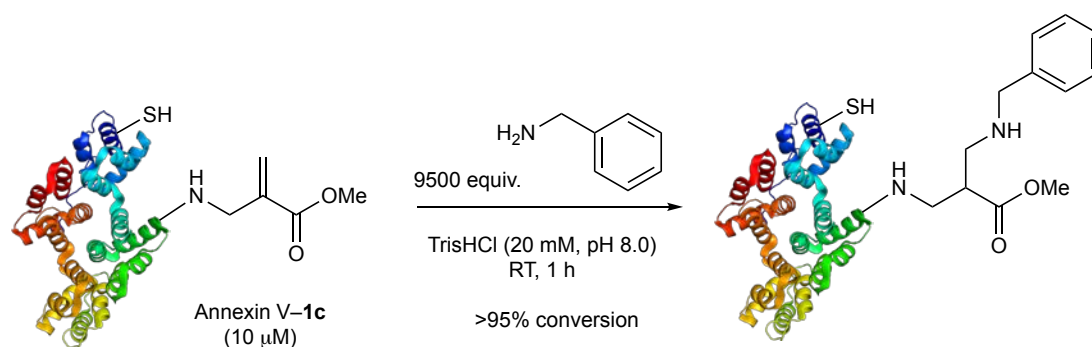
A 20 μL aliquot of Annexin V–**1c** (10 μM) in TrisHCl (20 mM, pH 8.0) was thawed. Benzylamine (9500 equiv.) was added and the resulting mixture was vortexed for 10 seconds. After 1 h of additional mixing at room temperature, a 10 μL aliquot was analysed by LC–MS and complete conversion to the expected product was observed (calculated mass, 36010 Da; observed mass, 36008 Da).
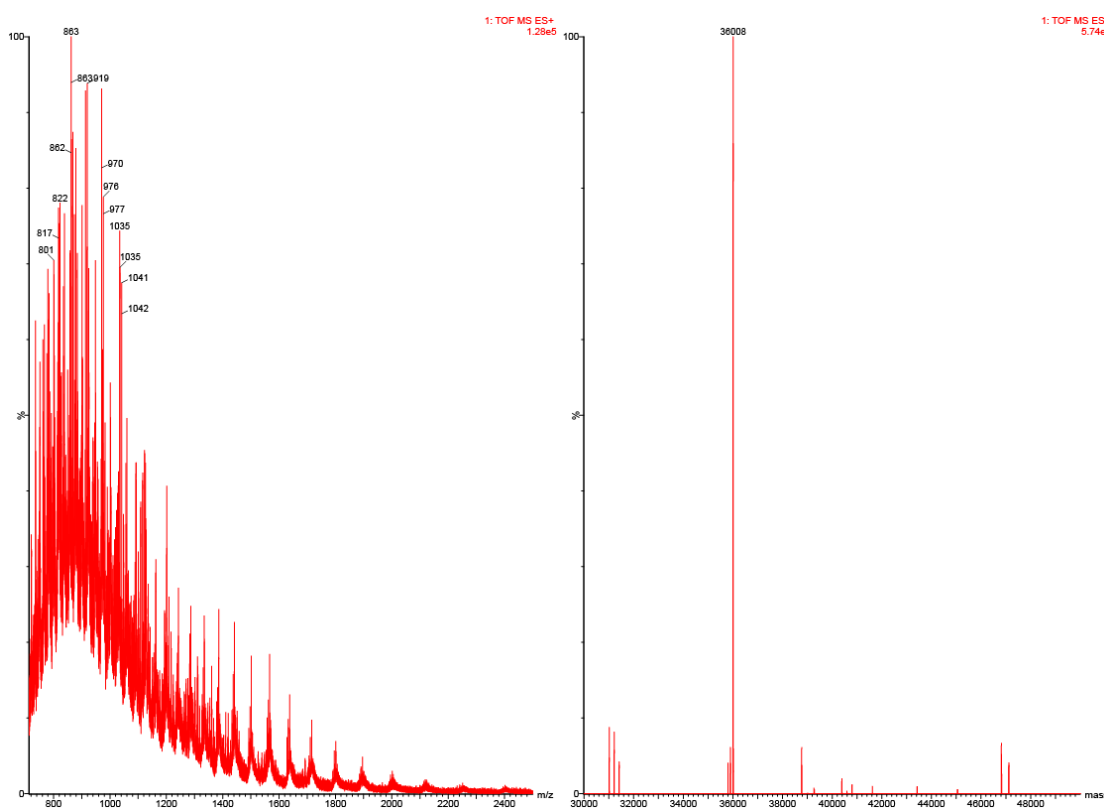
Supporting Figure 43. Combined ion series and deconvoluted mass spectrum of the reaction of Annexin V–**1c** with benzylamine (9500 equiv.) for 1 h at room temperature.

2-(2-(2-Methoxyethoxy)ethoxy)ethylamine addition to Annexin V–1c
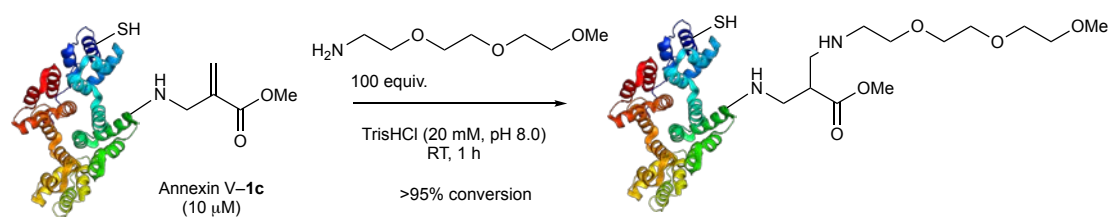
A 20 μL aliquot of Annexin V–**1c** (10 μM) in TrisHCl (20 mM, pH 8.0) was thawed. 2-(2-(2-Methoxyethoxy)ethoxy)ethylamine (100 equiv.) was added and the resulting mixture was vortexed for 10 seconds. After 1 h of additional mixing at room temperature, a 10 μL aliquot was analysed by LC–MS and complete conversion to the product was observed (calculated mass, 36066 Da; observed mass, 36060 Da).
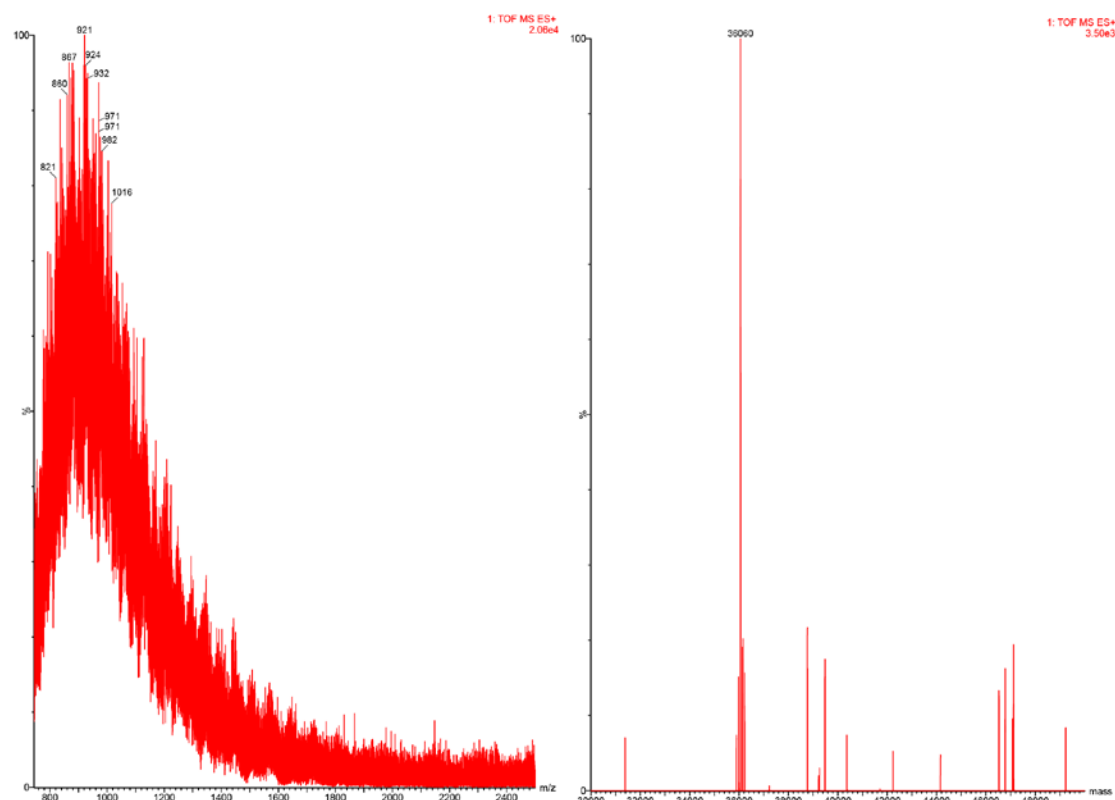
Supporting Figure 44. Combined ion series and deconvoluted mass spectrum of the reaction of Annexin V–**1c** with 2-(2-(2-methoxyethoxy)ethoxy)ethylamine (100 equiv.) for 1 h at room temperature.

8. Reactions and characterization with C2Am-conjugates

Reaction of C2Am with 1c
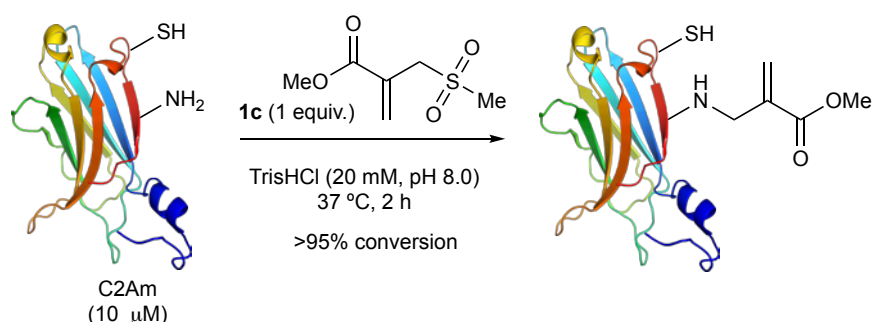
The reaction was performed according to the general procedure. To an eppendorf with 29.5 μL of TrisHCl (20 mM, pH 8.0) and 3.3 μL of DMF, a 6.5 μL aliquot of a stock solution of C2Am (46.2 μM) was added and the resulting mixture vortexed for 10 seconds. Afterwards, a 6.09 mM solution of **1c** (0.71 μL, 1 equiv.) in DMF was added and the reaction mixed for 2 and 24 h at 37 ºC. An aliquot of 1 μL (10 equiv.) of a stock solution of TCEP (4.0 mM) was added, for partial reduction of the dimer before the analysis of the sample. At each reaction time, a 10 μL aliquot was analysed by LC–MS and complete conversion to the product was observed (calculated mass, 16319 Da; observed mass, 16321 Da).
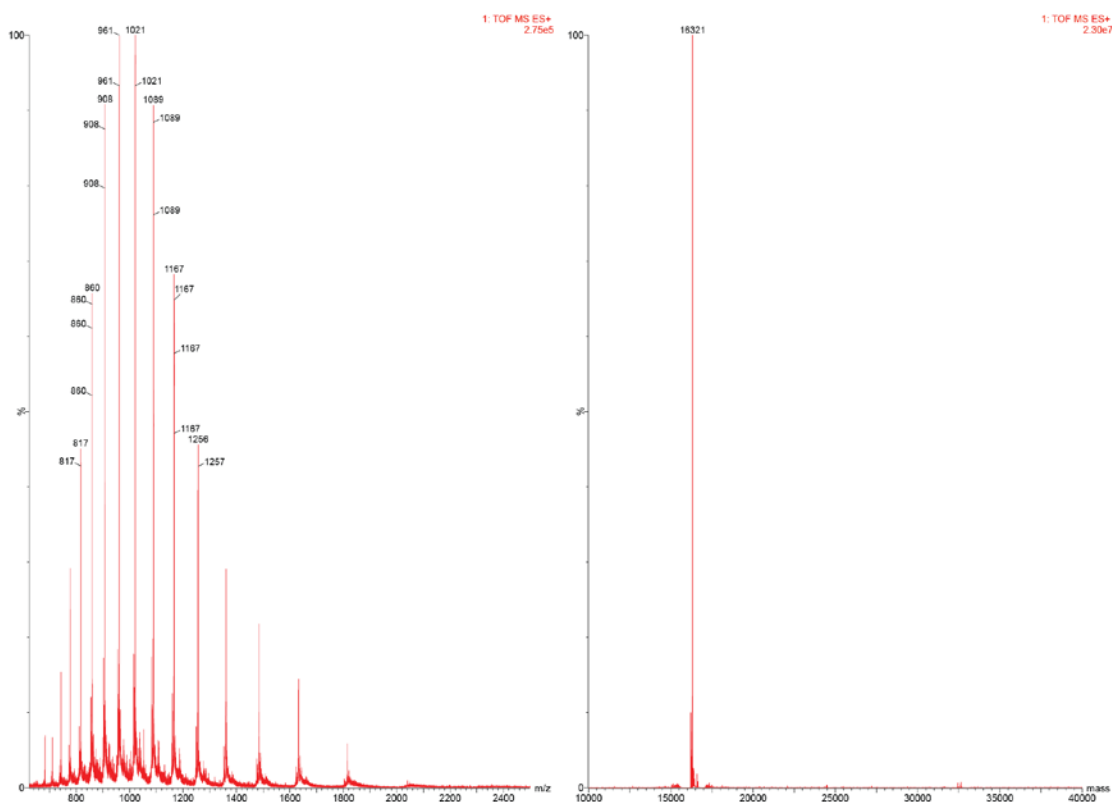
Supporting Figure 45. Combined ion series and deconvoluted mass spectrum of the reaction of C2Am (10 μM) with 1 equiv. of **1c** after 2 h at 37 ºC. Identical data was obtained at 24 h.

Reaction of C2Am–1c with Ellman's reagent
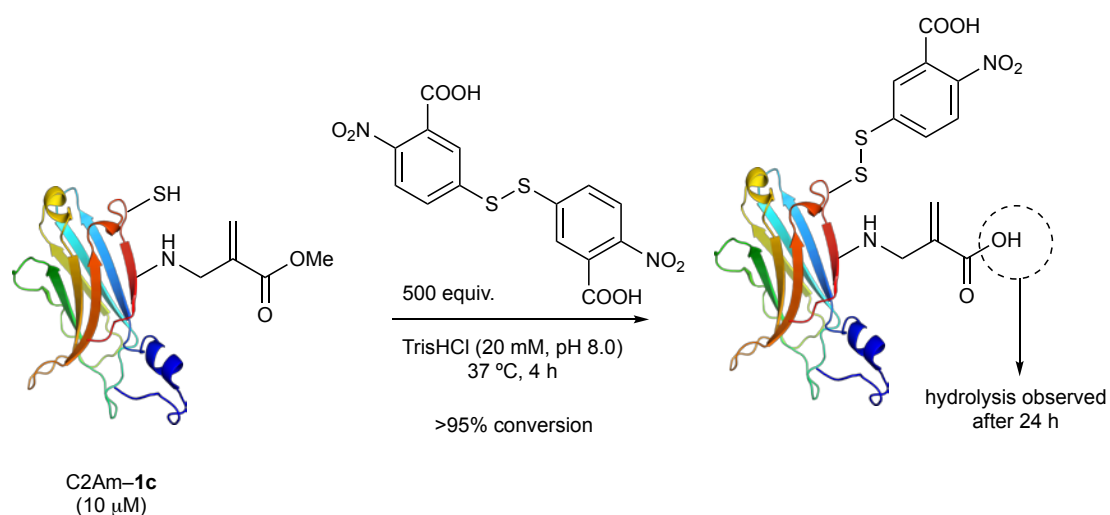
A 40 μL aliquot of C2Am–**1c** (10 μM) was transferred to a 0.5 mL eppendorf

tube. An aliquot of 8.0 μL (500 equiv.) of a stock suspension of Ellman's reagent

(50.5 μM) was added and the resulting mixture vortexed for 10 seconds. After

4 h of additional mixing, at 37 ºC, small molecules were removed from the reaction mixture by loading the sample into a Zeba Spin Desalting Column previously equilibrated with TrisHCl (20 mM, pH 8.0). The sample was eluted *via* centrifugation (2 min, 1000xg). A 10 µL aliquot was analysed by LC–MS and full conversion to the doubly modified protein was observed (calculated mass, 16517 Da; observed mass, 16503 Da). <u>Note:</u> in our experiments we noticed that the presence of a thiol source (GSH, Ellman's) can lead to hydrolysis of the methyl ester.

The use of same conditions but performing the reactions in reverse order (i.e., Ellman's mixed disulfide formation followed by reaction with **1c** gave identical results.

Supporting Figure 46. Combined ion series and deconvoluted mass spectrum of the reaction of C2Am–**1c** with Ellman's reagent (500 equiv.) after 4 h at 37 ºC.
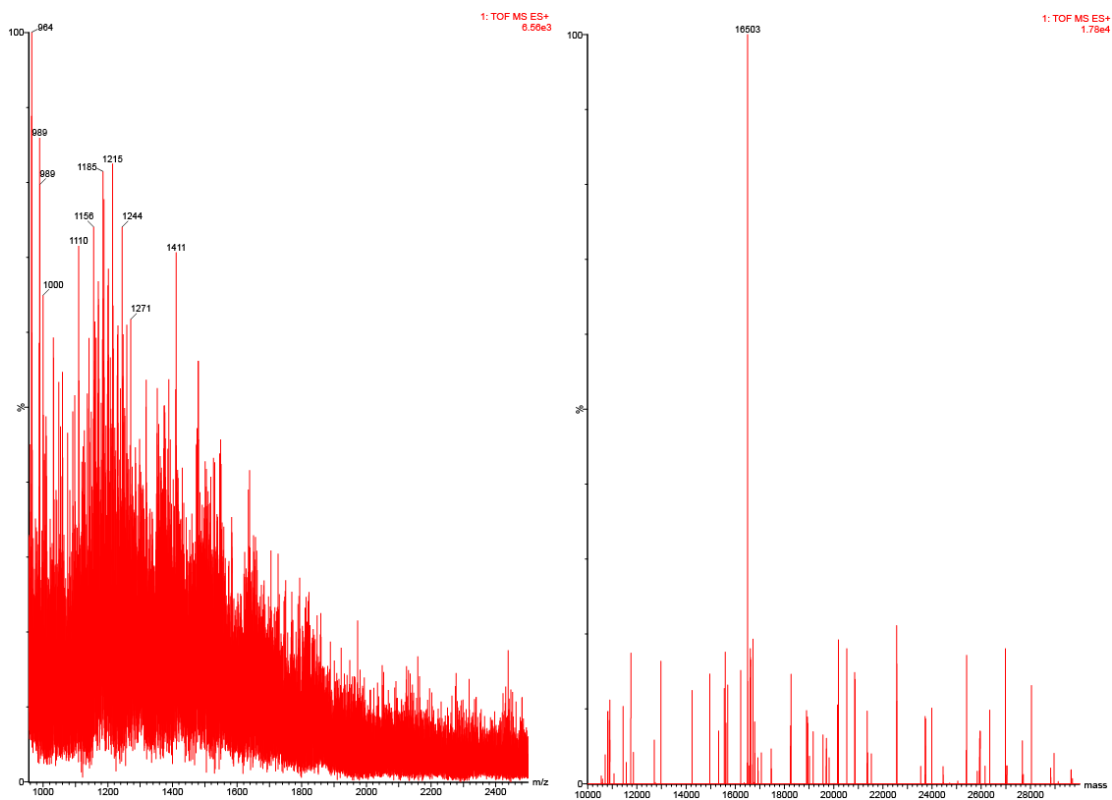
Supporting Figure 47. Combined ion series and deconvoluted mass spectrum of the reaction of C2Am–**1c** with Ellman's reagent (500 equiv.) after 24 h at 37 ºC.

FITC–PEG3NH$_2$ addition to C2Am
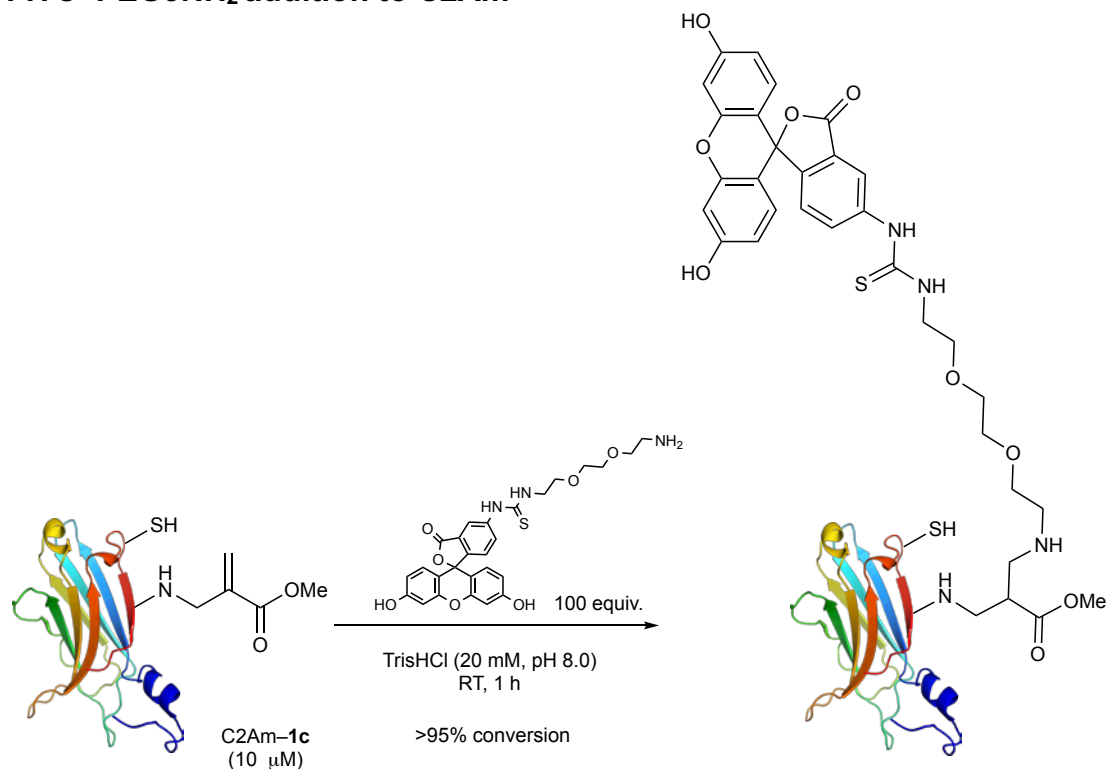
A 20 μL aliquot of C2Am–**1c** (10 μM) in TrisHCl (20 mM, pH 8.0) was thawed. FITC–PEG3NH$_2$ (100 equiv.) was added and the resulting mixture was vortexed for 10 seconds. After 1 h of additional mixing at room temperature, a 10 μL aliquot was analysed complete conversion to the expected product was observed by polyacrylamide gel electrophoresis.

Fluorescent

Coomassie
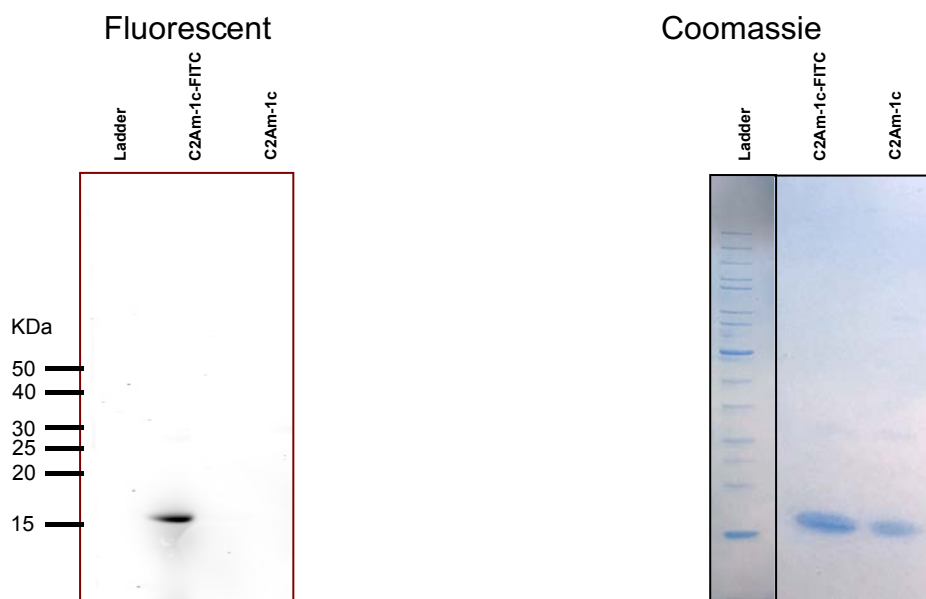
Supporting Figure 48. SDS-PAGE analysis of the reaction of C2Am–**1c** with FITC–PEG3NH$_2$.

CD of C2Am–1c–FITC
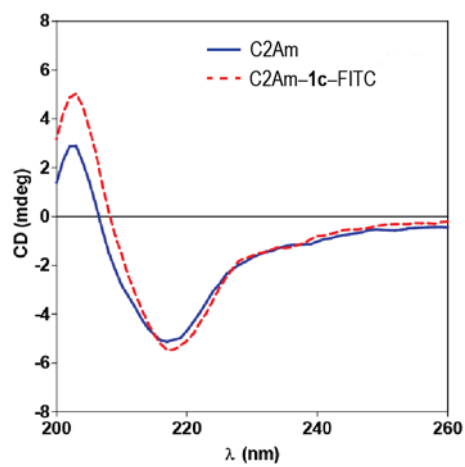
Supporting Figure 49. Structural analysis of C2Am and C2Am–**1c**–FITC by CD.

9. Image analysis of C2Am–1c–FITC binding to dying cells

Microscopy studies

Human embryonic kidney HEK293 cells were maintained at 37 ºC in a humidified atmosphere of 5% CO_2 in DMEM high glucose (Gibco) supplemented with 10% heat inactivated fetal bovine serum (HI-FBS, Gibco) and 1x antibiotic-antimycotic (Gibco). When cells have reached the appropriate density (70%–80% confluent), the medium was aspirated and cells harvested with 0.25% Trypsin-EDTA. Then, 200 µL of the cell suspension (~50 000 cells) was applied on top of 12 mm glass coverslips pre-coated with poly D-Lys (Corning™ BioCoat™) placed inside a 24-well plate. After 1–2 h of incubation to allow cells to adhere, 200 µL of additional media was added to flood the wells. Cells were then grown for more 8 h at 37 ºC before apoptosis was induced by treatment for 12 h with 2 µM of actinomycin D in fresh growth media. Untreated cells at the same density were included as a control. After apoptosis induction the media was removed, cells were washed with D-PBS (Gibco) containing 1% (m/v) of bovine serum albumin and then incubated at 37 ºC for 15 minutes with C2Am–**1c**–FITC at a concentration of 0.5 µM in 10 mM HEPES pH 7.4 supplemented with 140 mM NaCl and 2.5 mM $CaCl_2$ (Annexin binding buffer from Molecular Probes). At the same time blocking studies were performed by pre-incubating apoptotic cells with a 10x excess of non-fluorescent C2Am wild type (5 µM, 37 ºC for 15 minutes) before incubation with the fluorescent probe C2Am–**1c**–FITC for additional 15 minutes at 37 ºC. For fluorescent DNA nuclei staining, Hoechst 33342 (0.8 µg/mL, Sigma Aldrich, 15 minutes at 37 ºC) was used. After labelling, the cells were washed in 2x PBS pH 7.4. Finally, cells

were fixed with PBS pH 7.4 containing 4% (w/v) formaldehyde for 15 minutes at room temperature, further washed two times with Milli-Q water, and mounted on slides with Ibidi mounting medium. Fluorescence microscopy was performed using an inverted epifluorescent microscope (Olympus IX-71) connected to a F-view digital camera (Soft Imaging System). Images were acquired in the FITC and Hoechst channels and analysed using the software Cell-F. Identical image acquisition settings were used for the control, experimental and blocking data sets. Confocal images were acquiring on a Leica SP5 confocal microscope equipped with Leica Application Suite Advanced Fluorescence (LAS AF) software and an oil-immersion 63x objective of numerical aperture of 1.4. The FITC was detected upon excitation with the 488 nm laser line of an argon laser and detected on a 500 to 560 nm spectral bandwidth, while the excitation and emission wavelengths of Hoechst 33258 were 405 and 425–490 nm, respectively. The confocal stacks (10 sections 1024 x 1024) were analysed using ImageJ software.
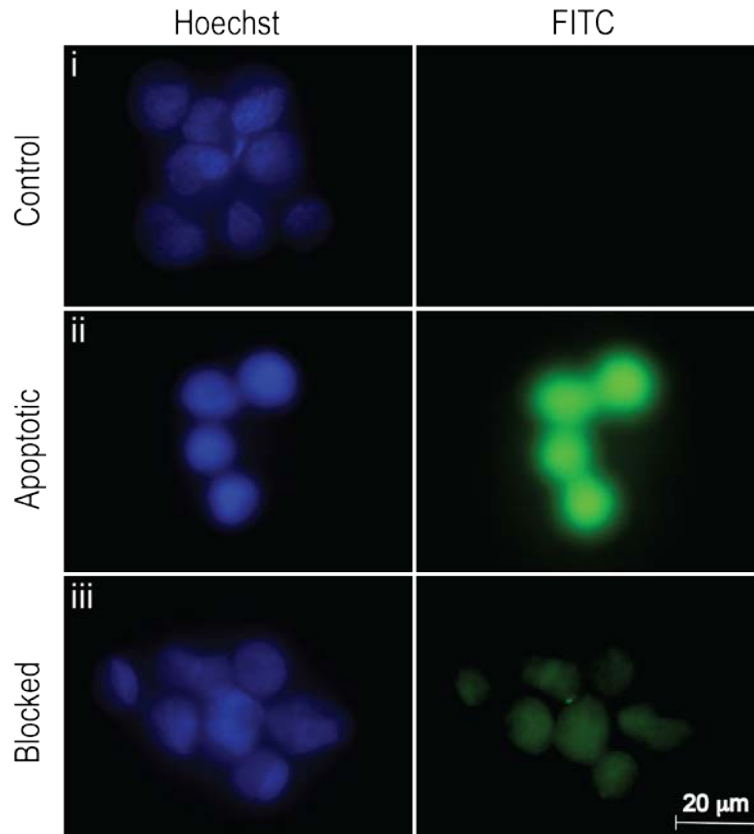
Supporting Figure 50. Images of non-apoptotic (control, **i**) and apoptotic HEK293 cells (**ii** and **iii**) after labeling with C2Am–**1c**–FITC. Apoptotic cells stained with C2Am–**1c**–FITC showed a significant increase in the fluorescence intensity (**ii**) when compared to non-apoptotic cells (**i**). In addition, apoptotic cells previously incubated with a 10x excess of non-fluorescent C2Am showed a significant decrease of binding of C2Am–**1c**–FITC (**iii**). These blocking studies confirm a specific binding mechanism. Apoptotic cells are shown green, while the nuclei counterstained with Hoechst 33342 are shown blue. Scale bar represents 20 μm.
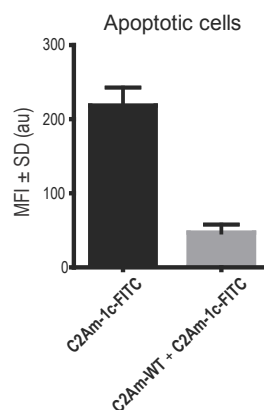
Supporting Figure 51. The mean fluorescent intensity (MFI) of apoptotic cells incubated with C2Am–**1c**–FITC or preincubated with C2Am–WT and then C2Am–**1c**–FITC.
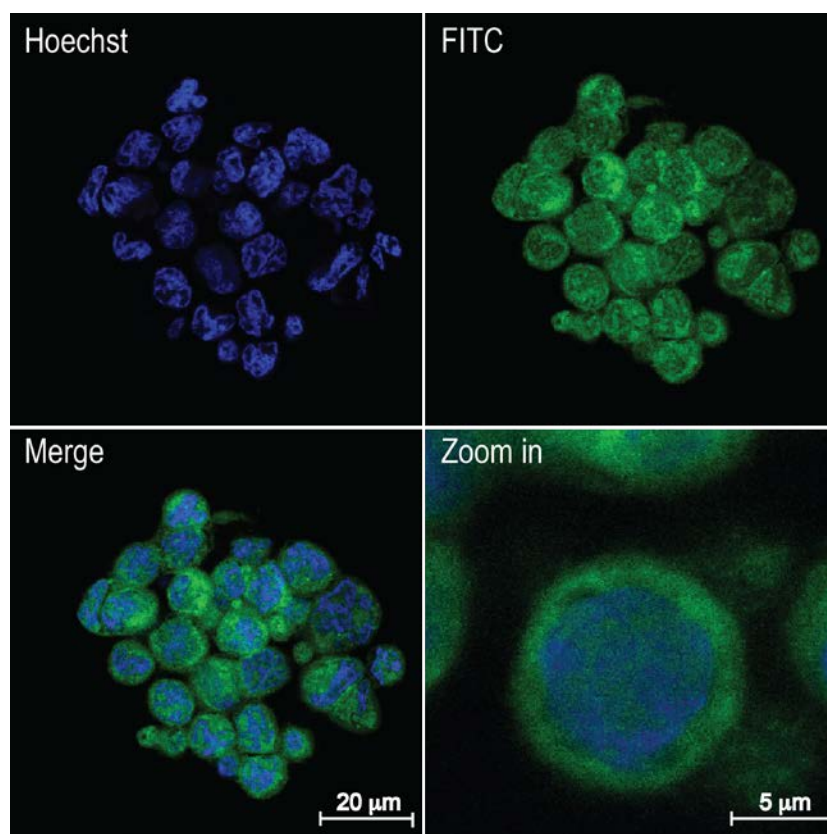
Supporting Figure 52. Confocal images of apoptotic cells treated with C2Am–**1c**–FITC. Zoom in and Z-stacking showing peripheral staining of the cellular membrane. Membranes of apoptotic cells are shown in green, while the nuclei in blue.

10. Reactions and characterization of Trastuzumab-conjugates

Reaction of 1c with Trastuzumab
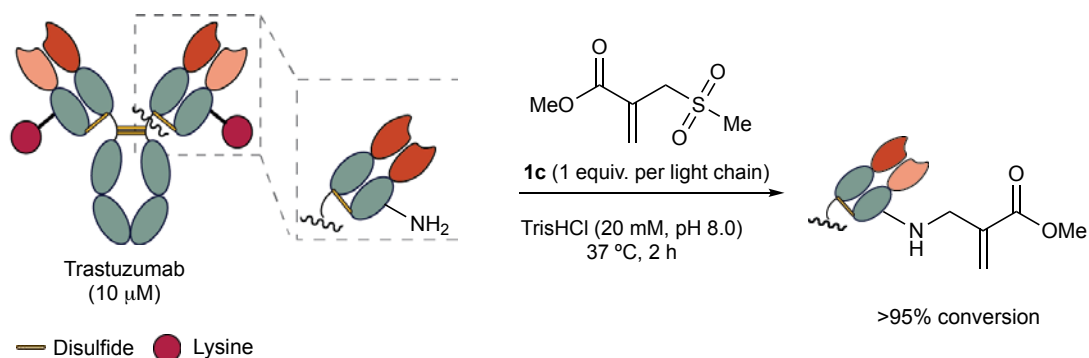
The reaction was performed according to the general procedure. To an eppendorf with 20.7 μL of TrisHCl (20 mM, pH 8.0) and 3.3 μL of DMF, was added a 15.3 μL aliquot of a stock solution of Trastuzumab (26.1 μM). Afterwards, a 0.56 mM solution of **1c** (0.7 μL, 1 equiv. per light-chain) in DMF was added and the reaction mixed for 2 and 24 h at 37 ºC. At each reaction time, a 10 μL aliquot, with 1 μL (10 equiv.) of a stock solution of TCEP (4 mM), was analysed by LC–MS and complete conversion to the product was observed (calculated mass, 23540 Da; observed mass, 23541 Da). The modification occurred exclusively at the light-chain and no mass alterations were found on the heavy-chain.
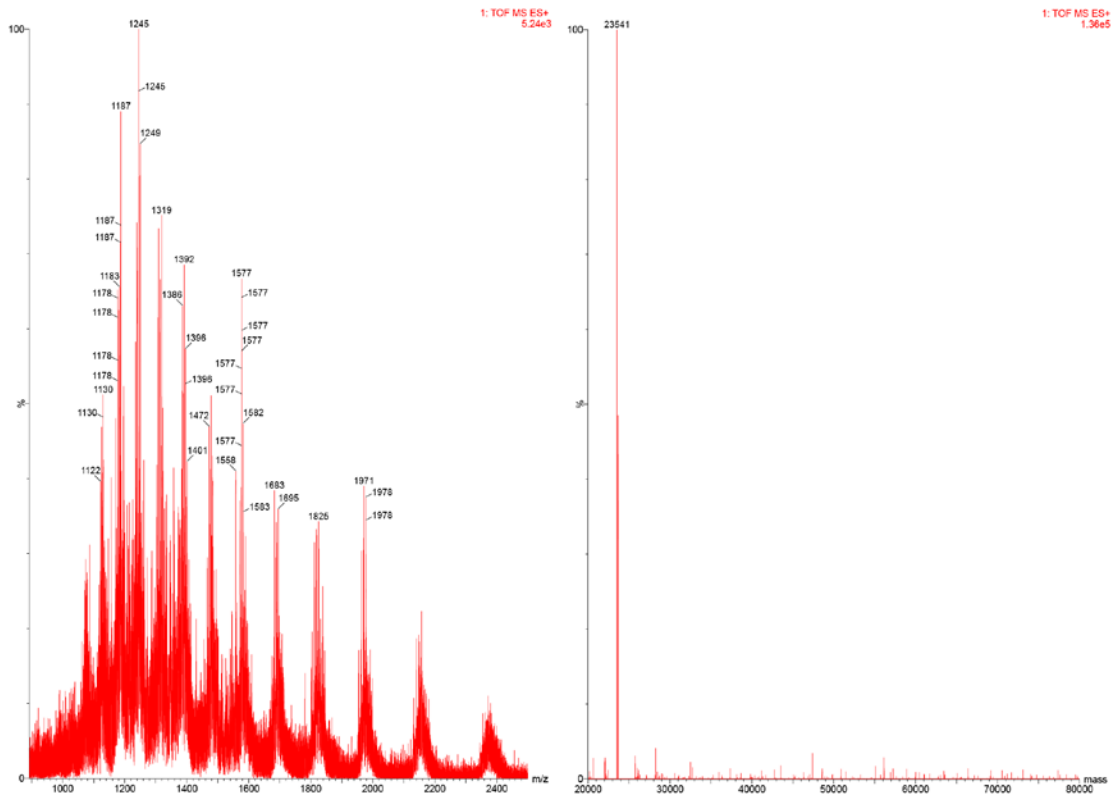
Supporting Figure 53. Combined ion series and deconvoluted mass spectrum of the light-chain of Trastuzumab after reaction with **1c** (1 equiv. per light-chain) for 2 h at 37 ºC.
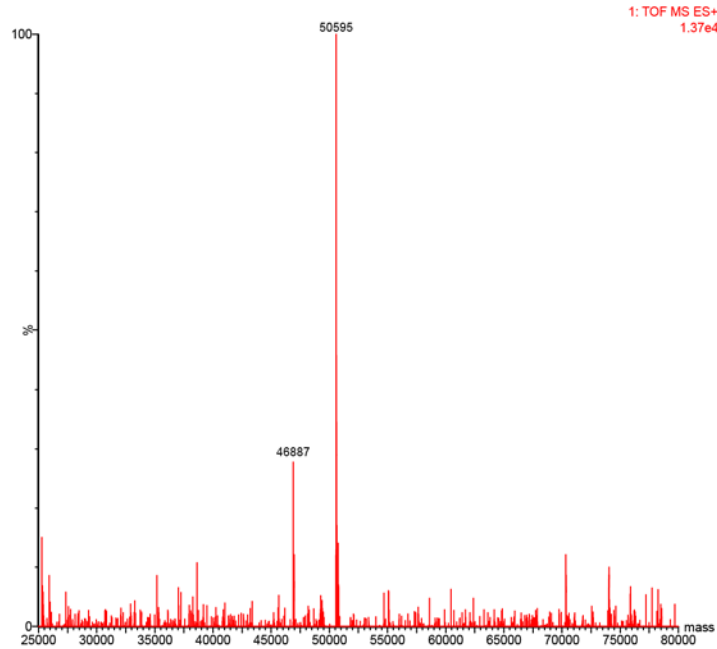
Supporting Figure 54. Deconvoluted mass spectrum of the heavy-chain of Trastuzumab after reaction with **1c** (1 equiv. per light-chain) for 2 h at 37 ºC.

FITC–PEG3NH$_2$ addition to Trastuzumab–1c
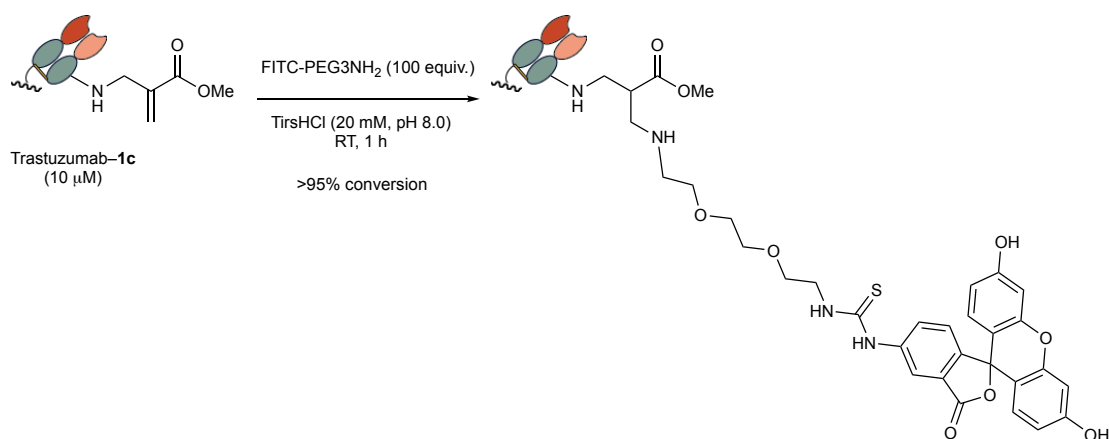
A 20 μL aliquot of Trastuzumab–**1c** (10 μM) in TrisHCl (20 mM, pH 8.0) was thawed. FITC–PEG3NH$_2$ (100 equiv.) was added and the resulting mixture was vortexed for 10 seconds. After 1 hour of additional mixing at room temperature, a 10 μL aliquot was analysed and complete conversion to the expected product was observed by polyacrylamide gel electrophoresis.
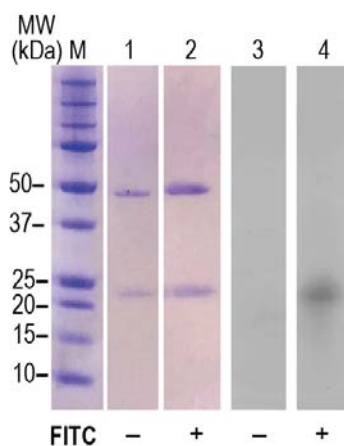
Supporting Figure 55. SDS-PAGE of the reaction of Trastuzumab–**1c** with FITC–PEG3NH$_2$.

Reaction of Trastuzumab–1c with crizotinib
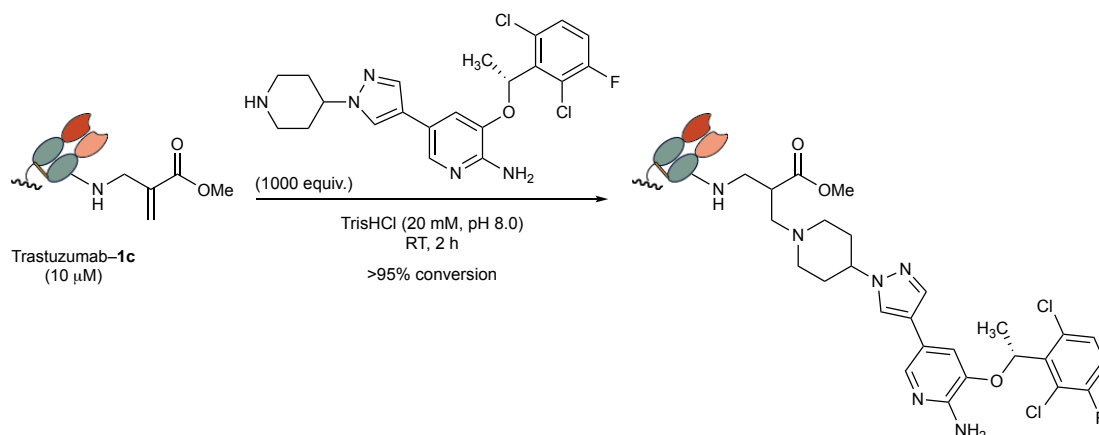
A 20 μL aliquot of Trastuzumab–**1c** (10 μM) in TrisHCl (20 mM, pH 8.0) was thawed. Crizotinib (10 mM, 1000 equiv.) was added and the resulting mixture was vortexed for 10 seconds. After 2 h of additional mixing at room temperature, a 10 μL aliquot was analysed by LC–MS and complete conversion to the expected product was observed (calculated mass of light chain, 23986 Da; observed mass, 23983 Da). The modification occurred exclusively at the light-chain and no mass alterations were found on the heavy-chain. When the reaction was scaled up for biophysical studies, an overnight purification by dialysis was performed.
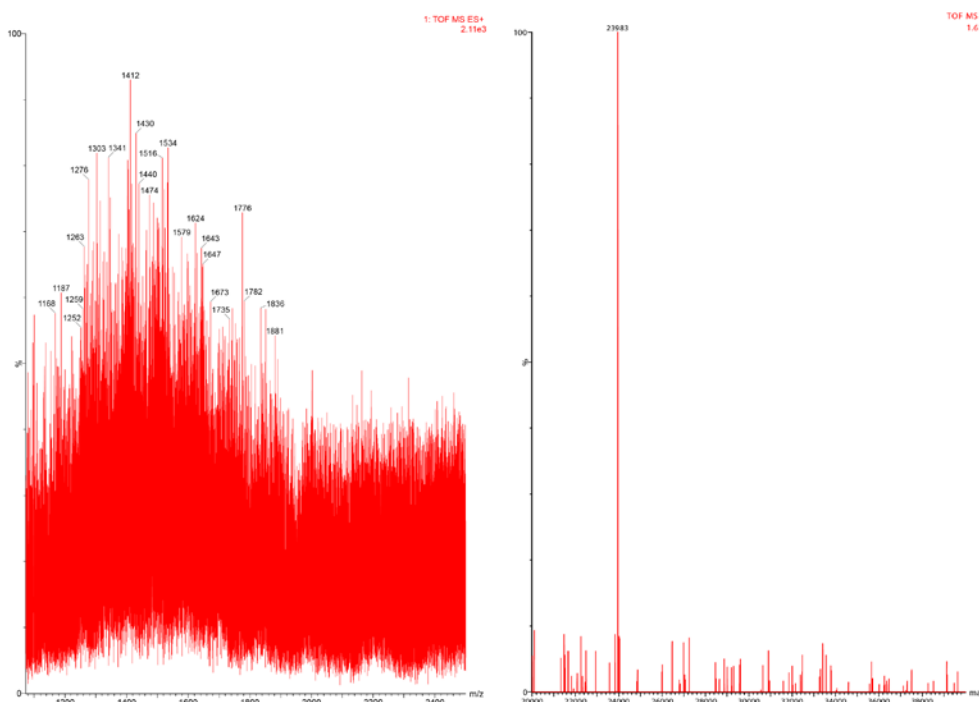
Supporting Figure 56. Combined ion series and deconvoluted mass spectrum of the reaction of Trastuzumab–**1c** with crizotinib (10 mM) for 2 h at room temperature.
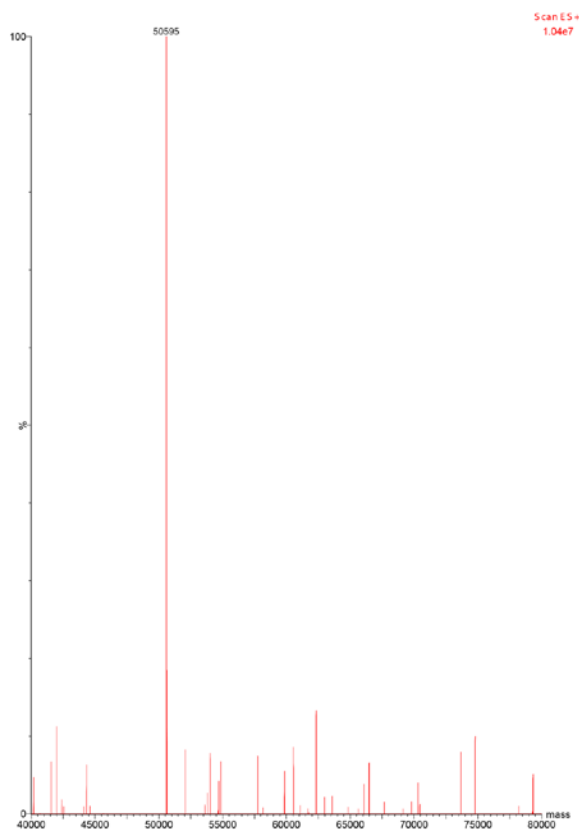
Supporting Figure 57. Deconvoluted mass spectrum of the heavy-chain of Trastuzumab–**1c**–crizotinib.
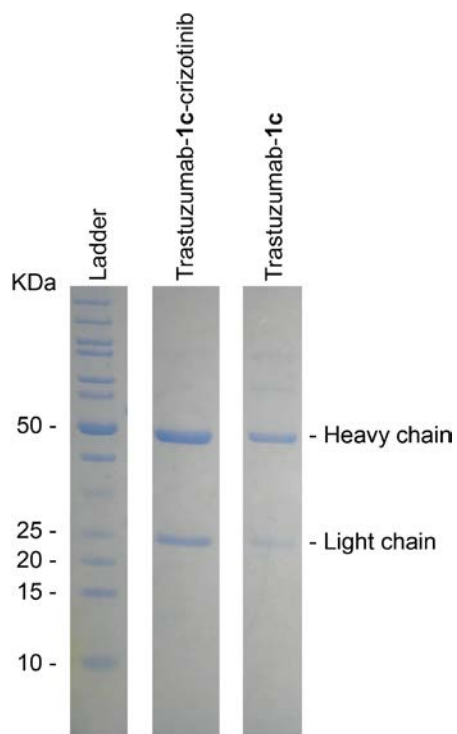
Supporting Figure 58. SDS-PAGE of the reaction of Trastuzumab–**1c** with crizotinib.

CD of Trastuzumab–1c–crizotinib
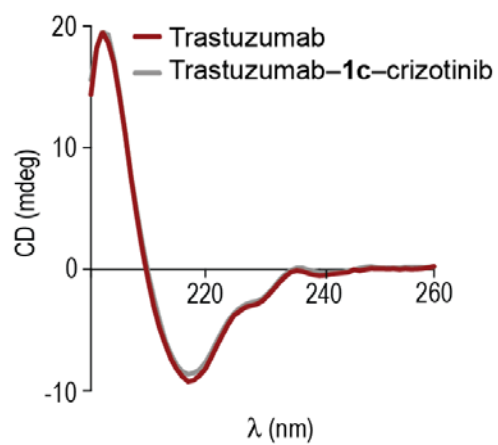
Supporting Figure 59. Structural analysis of Trastuzumab and Trastuzumab–**1c**–crizotinib by CD.

## 11. Determination of antibody-conjugates binding affinity

Biotinylation of antibodies

Non-modified Trastuzumab and Trastuzumab–**1c** were conjugated to a Q8 biotin linker using Biotin-(PEG)4-NHS (Thermofisher Scientific) to carry out BLI experiments using Streptavidin (SA) Biosensors. A solution of EZ-Link NHS-(PEG)4-Biotin (20 mL, 200 mM in PBS) was added to the corresponding protein (20 mL, 20 mM in PBS pH 7.4) and was left at room temperature for 30 minutes. The crude reaction mixture was buffer exchanged with PBS for three times to remove the excess of NHS-(PEG)4-Biotin, obtaining a biotin-to-antibody ratio around 1.6.

Bio-layer interferometry

Binding assays were performed on an Octet Red Instrument (fortéBIO). Ligand immobilization, binding reactions, regeneration and washes were conducted in wells of black polypropylene 96-well microplates. Trastuzumab and Trastuzumab–**1c** (20 nM) were immobilized on Streptavidin (SA) Biosensors in PBS with 0.1% BSA and 0.02% tween at 30 ºC. Binding analysis were carried out at 25 ºC, 1,000 r.p.m. in PBS pH 7.4 with 0.1% BSA and 0.02% tween, with a 600 s of association followed by a 2,200 s of dissociation, using different concentrations of recombinant Her2 receptor to obtain the association curve. Glycine pH 2.0 was used as a regeneration buffer. Data were analysed using Data Analysis (fortéBIO), with Savitzky-Golay filtering. Binding was fitted to a 2:1 Heterogeneous ligand model, steady state analysis was performed to obtain the binding kinetics constants (K_D).
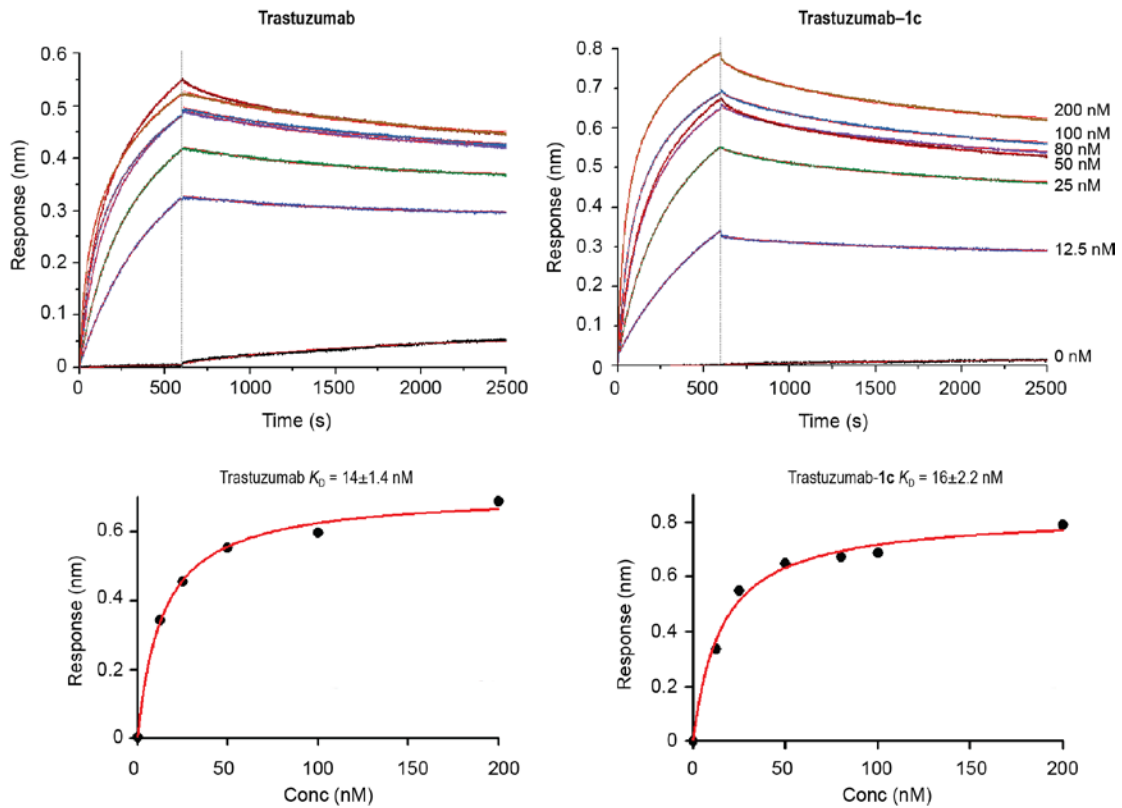
Supporting Figure 60. Bio-Layer Interferometry (BLI) curves (in blue) and fitting curves (in red) obtained for Trastuzumab and Trastuzumab–**1c** with Her2 receptor, together with the K_D constants derived from BLI experiments.
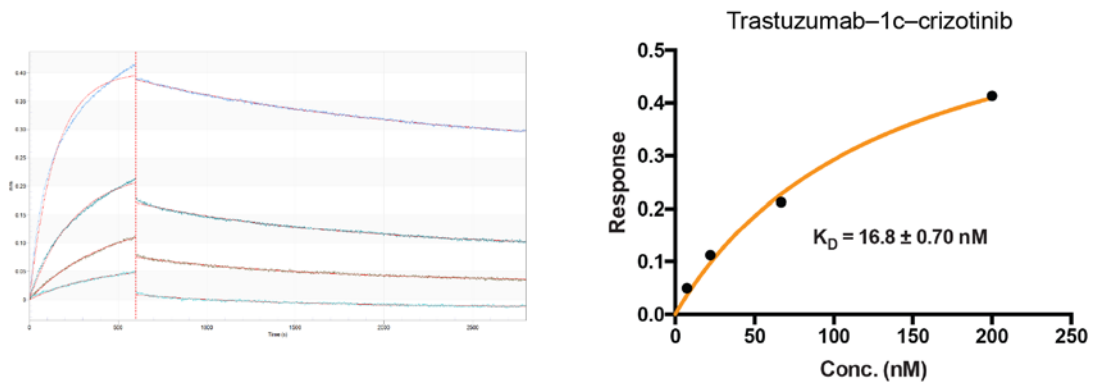
Supporting Figure 61. Bio-Layer Interferometry (BLI) curves obtained for Trastuzumab–**1c**–crizotinib with Her2 receptor, together with the K_D constant derived from BLI experiments.

12. Determination of antibody-conjugates specificity

Cell culture conditions

SKBR3 (ATCC HTB-30), HepG2 (ATCC HB-8065) and HEK293T (ATCC CRL-3216) cells were routinely grown in a humidified incubator at 37 ºC under 5% CO_2, and split before reaching confluence using TrypLE Express. Both cells lines were grown on DMEM medium supplemented with 10% heat-inactivated FBS, 2mM GlutaMAX, 10mM HEPES, Q9 1% NEAA, 1mM sodium pyruvate, 100 units mL^{-1} penicillin and 100 mg mL^{-1} streptomycin. All reagents were bought from Gibco, Life Technologies (USA), unless otherwise stated.

Flow cytometry analysis of Trastuzumab–1c–FITC

The binding affinity of the antibody Trastuzumab–**1c**–FITC (compared to Trastuzumab) was determined by Flow Cytometry analysis. For this purpose, SKRB3 cells (with high expression of Her2 receptor) and HepG2 cells (with low expression of Her2 receptor) were plated in 96-well plates (100,000 cells per well) and incubated with 10 µL of the antibodies at different concentrations (10, 50 and 150 µM) at room temperature, in the dark. After 1 h of incubation 100 µL of 10% FBS in PBS was added and the cells were centrifuged for 5 minutes at 400 G. The supernatant was then removed, the cells were re-suspended in 400 µL of 10% FBS in PBS and transferred to flow cytometry tubes. Acquisition was done using a BD LSR Fortessa set up with a 488 nm laser and a combination of a 505 nm long-pass and a 530/30 nm band-pass filter (combination used for FITC detection). Data analysis was done with FlowJo

(version 6.3.4, FlowJo) software. Data represents mean ± s.d. of 3 biological

replicates and only single-cell events are shown.
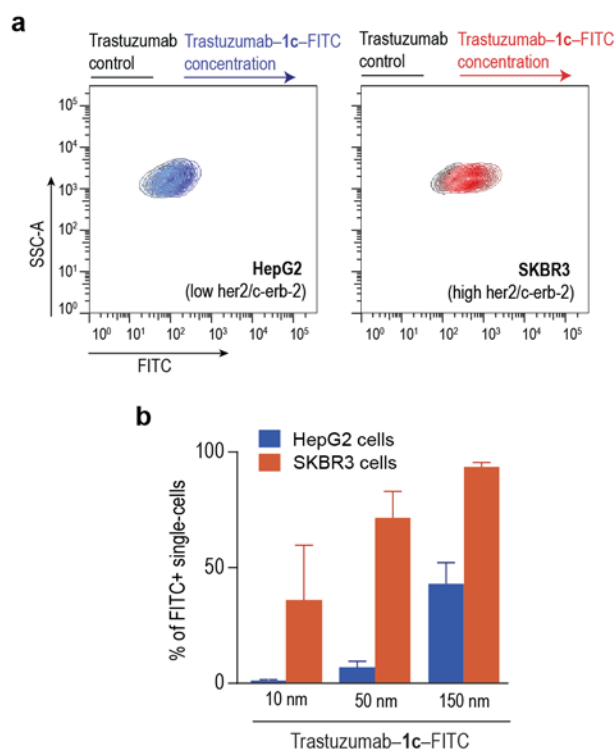
Supporting Figure 62. Analysis of specificity of Trastuzumab–**1c**–FITC towards Her2 by flow-cytometry. **a.** Superposition of contour plots of side-scatter detection *versus* FITC-equivalent fluorescence intensity, in HepG2 cells (blue, expressing low levels of her2/c-erb-2), and in SKBR3 cells (red, expressing high levels of her2/c-erb-2). Controls were treated with non-conjugated Trastuzumab while samples were treated with increasing concentrations of Trastuzumab–**1c**–FITC (10, 50 and 150 nM). **b.** Percentage of FITC-positive single cells, after treatment with fluorescently labelled or non-labelled Trastuzumab, both in HepG2 cells (blue) and SKBR3 cells (red).

Flow cytometry analysis of Trastuzumab–1c–crizotinib

The binding affinity of the antibody Trastuzumab–**1c**–crizotinib was determined

by Flow Cytometry analysis. For this purpose, SKRB3 cells (with high

expression of HER2 receptor) and HEK293T cells (HER2 low expressing cells)

were plated in 96 well round bottom plates (100.000 cells per well) and were

incubated with 30 µL of 10 µM Trastuzumab–**1c**–crizotinib at room temperature.

After 1 h of incubation 100 µL of 10%FBS in PBS was added and the cells were

centrifuged for 5 min at 400 G. Following this washing step, the cells were incubated with 30 µL/well of Goat anti-Human IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 647 (cat. No A21445, Life Technolonies) at 10 µg/mL, for 1 h, at room temperature, in the dark. After this incubation period, the cells were washed one time as previously described, re-suspended in 400 µL of 10% FBS in PBS and transferred to flow cytometry tubes. Acquisition was done using a BD LSR Fortessa set up with a 640 nm laser and a 670/14 nm band-pass filter (combination used for APC detection). Data analysis was done with FlowJo (version 6.3.4, FlowJo) software. Data represents mean ± s.d of 3 biological replicates and only single-cell events are shown.

13. Theoretical calculation of the most reactive lysine on three proteins and obtained LC–MS/MS analysis confirming the modified site

Lysozyme (6 Lys, no free Cys, 4 disulfides)
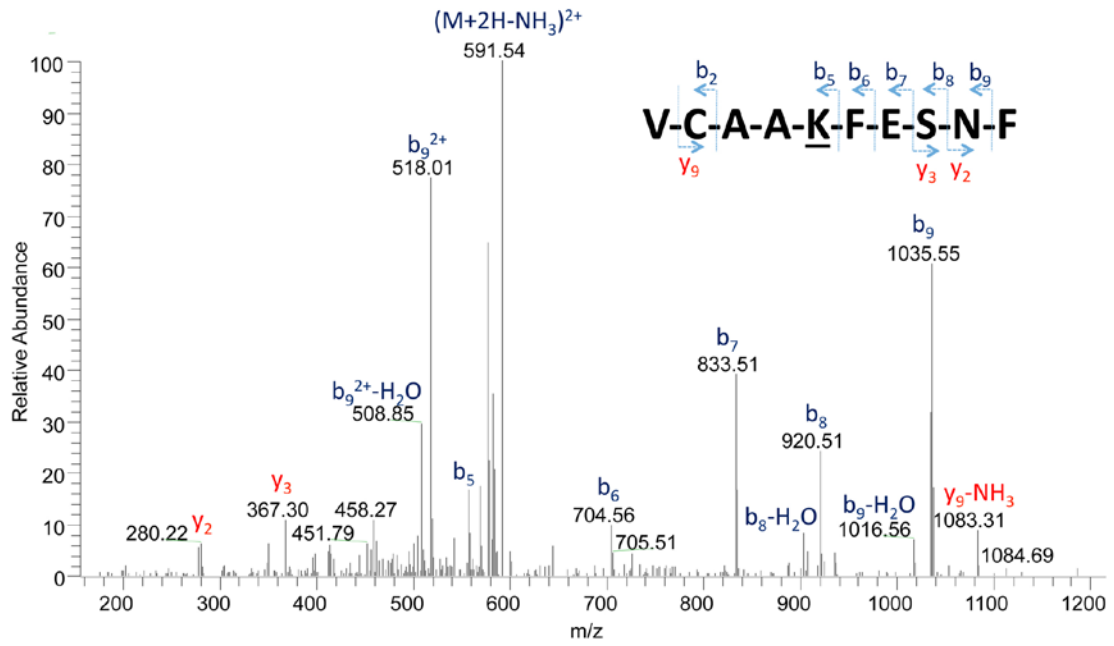
Lys 33 – determined computationally
Lys 33 – found experimentally
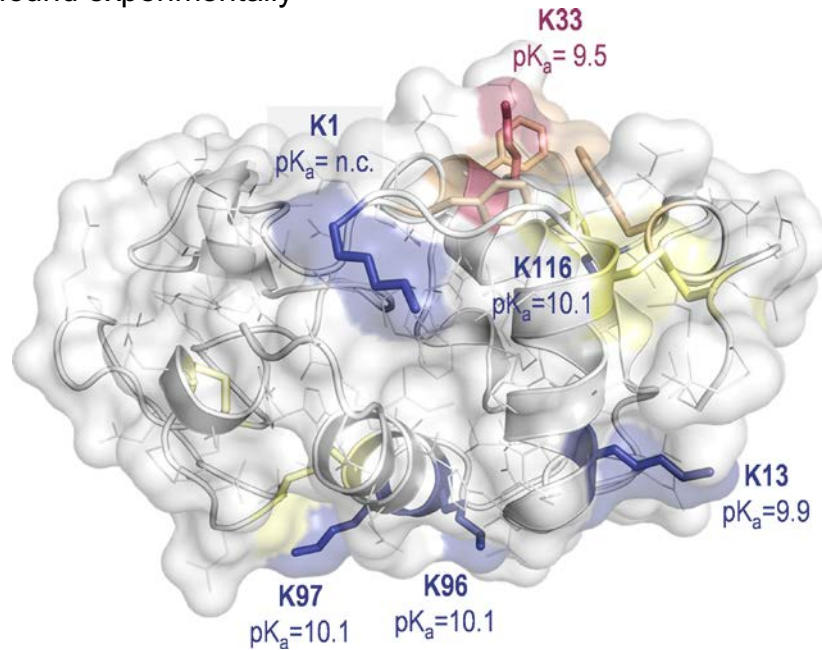
Sequence:

KVFGRCELAA AMKRHGLDNY RGYSLGNWVC AA**K**FESNFNT QATNRNTDGS   50
TDYGILQINS RWWCNDGRTP GSRNLCNIPC SALLSSDITA SVNCAKKIVS   100
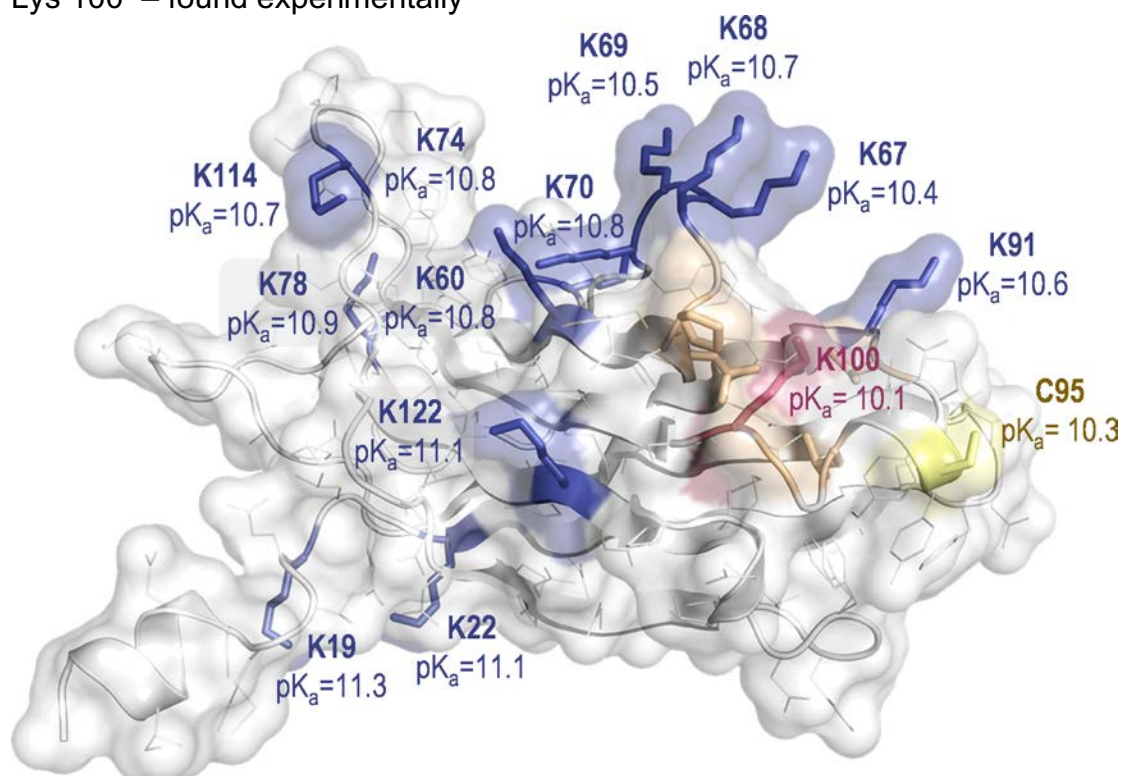DGNGMNAWVA WRNRCKGTDV QAWIRGCRL 129

Supporting Figure 62. MS/MS spectrum of the *m/z* 600.27 doubly charged ion of the lysine modified peptide VCAAKFESNF from chicken egg white lysozyme. The underscore relates to the modified amino acid.

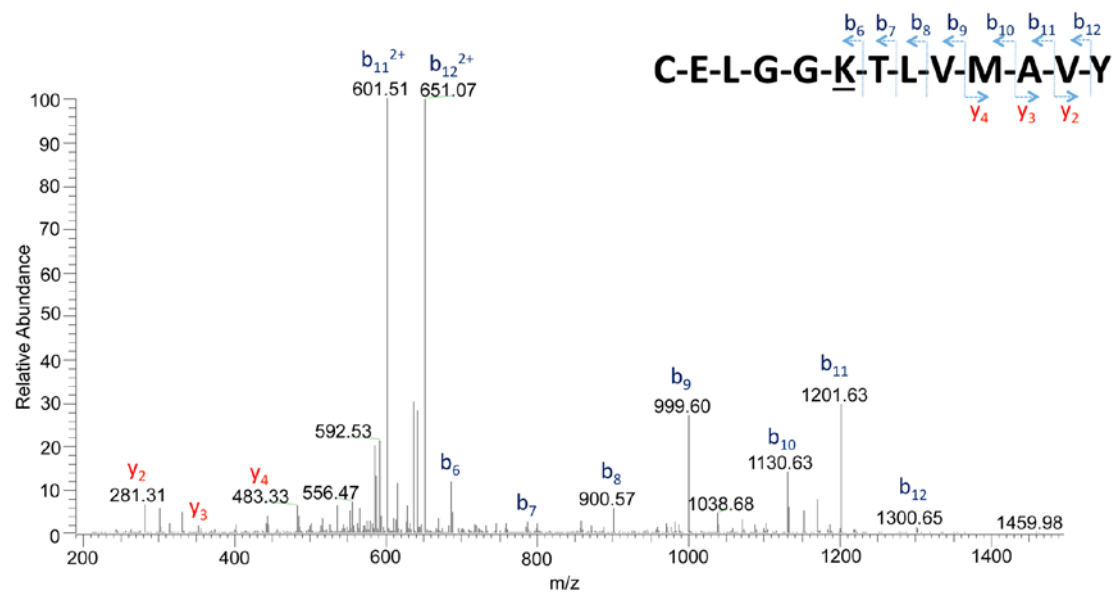**C2Am** (14 Lys, 1 free Cys, no disulfides)

Lys 100 – determined computationally
Lys 100 – found experimentally

Sequence:

GSPGISGGGG GILDSMVEKL GKLQYSLDYD FQNNQLLVGI IQAAELPALD   50
MGGTSDPYVK VFLLPDKKKK FETKVHRKTL NPVFNEQFTF KVPYCELGG**K** 100
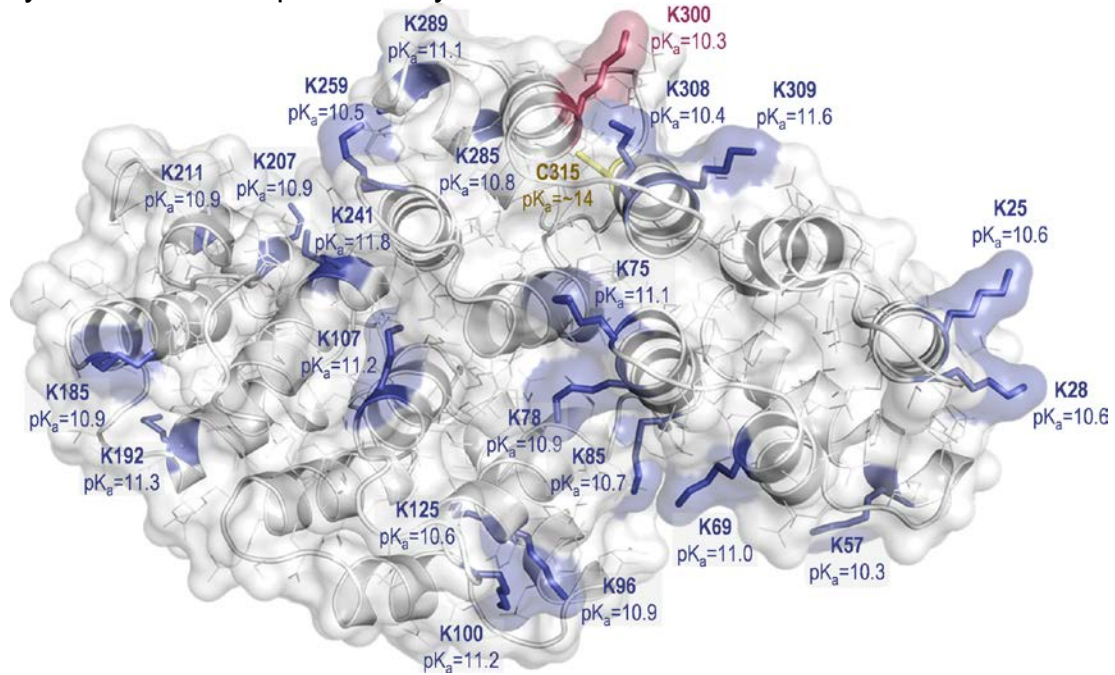TLVMAVYDFD RFSKHDIIGE FKVPMNTVDF GHVTEEWRDL QSAEK 145

Supporting Figure 63. MS/MS spectrum of the *m/z* 741.37 doubly charged ion of the lysine modified peptide CELGG**K**TLVMAVY from the C2Am. The underscore relates to the modified amino acid.

Annexin V (22 Lys, 1 free Cys, no disulfides)

Lys 300 – determined computationally
Lys 300 – found experimentally

Sequence:

```
AQVLRGTVTD  FPGFDERADA  ETLRKAMKGL  GTDEESILTL  LTSRSNAQRQ  50
EISAAFKTLF  GRDLLDDLKS  ELTGKFEKLI  VALMKPSRLY  DAYELKHALK   100
GAGTNEKVLT  EIIASRTPEE  LRAIKQVYEE  EYGSSLEDDV  VGDTSGYYQR  150
MLVVLLQANR  DPDAGIDEAQ  VEQDAQALFQ  AGELKWGTDE  EKFITIFGTR  200
SVSHLRKVFD  KYMTISGFQI  EETIDRETSG  NLEQLLLAVV  KSIRSIPAYL   250
AETLYYAMKG  AGTDDHTLIR  VMVSRSEIDL  FNIRKEFRKN  FATSLYSMI**K**  300
GDTSGDYKKA  LLLLCGEDD   319
```
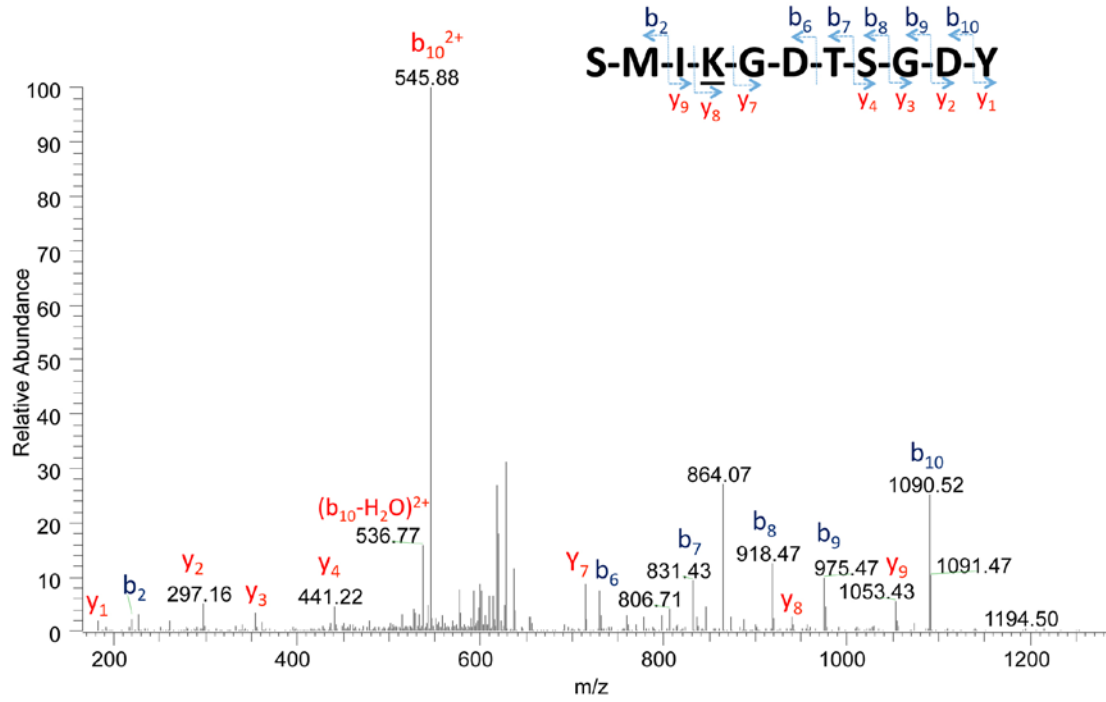
Supporting Figure 64. MS/MS spectrum of the *m/z* 636.27 doubly charged ion of the lysine modified peptide SMIKGDTSGDY from Annexin V. The underscore relates to the modified amino acid.

14. References

(1) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09* (Gaussian, Inc., Wallingford CT, 2009).

(2) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(3) Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110.

(4) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B.* **2011**, *115*, 14556–14562.

(5) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154–2161.

(6) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523–5527.

(7) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038–2048.

(8) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen ,H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. *AMBER 2016* (University of California, San Francisco, 2016).

(9) Tsui, V.; Case, D. A. *Biopolymers* **2000**, *56*, 275–291.

(10) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.

(11) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.

(12) Uberuaga, B. P.; Anghel, M.; Voter, A. F. *J. Chem. Phys.* **2004**, *120*, 6363–6374.

(13) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. *J. Chem. Theory Comput.* **2009**, *5*, 1624–1631.

(14) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(15) Vazquez, M. L.; Mueller, R. A.; Talley, J. J.; Getman, D. P.; DeCrescenzo, G. A.; Sun, E. T.; Google Patents: 1998.

(16) Davoust, M.; Briere, J.-F.; Metzner, P. *Org. Biomol. Chem.* **2006**, *4*, 3048–3051.

(17) Bertoldo, J. B.; Razzera, G.; Vernal, J.; Brod, F. C. A.; Arisi, A. C. M.; Terenzi, H. *Biochim. Biophys. Acta.* **2011**, *1814*, 1120–1126.

(18) Alam, I. S.; Neves, A. A.; Witney, T. H.; Boren, J.; Brindle, K. M. *Bioconjugate Chem.* **2010**, *21*, 884–891.

(19) Cal, P. M. S. D.; Sieglitz, F.; Santos, F.; Carvalho, C. P.; Guerreiro, A.; Bertoldo, J. B.; Pischel, U.; Gois, P.; Bernardes, G. J. L. *Chem. Commun.* **2016**, *53*, 368–371.

(20) Asano, S.; Patterson, J. T.; Gaj, T.; Barbas, C. F. *Angew. Chem. Int. Ed.* **2014**, *53*, 11783–11786.