Supporting Information

Tertiary structure-based prediction of how ATRP initiators react with proteins

Sheiliza Carmali[†], Hironobu Murata[†], Erika Amemiya[†], Krzysztof Matyjaszewski[†], Alan J. Russell[†, *]

[†] Center for Polymer-Based Protein Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania, 15213 United States

Total Number of Pages: 15

Total Number of Figures: 11

Total Number of Tables: 5

Molecular Dynamics (MD) Simulation with Initiator 1

The initial structure for bromine-functionalized NHS-initiator 1 was built with Maestro built toolkit (Schrodinger). To remove any bias or constraints, the initiator structure was subjected to a simulated annealing (SA) protocol using Desmond (Maestro, Schrodinger)4 (

Table **S1**). The system for simulation was prepared using Desmond's system builder. The OPLS-2005 force field was used and SPC was chosen as a solvent model. An orthorhombic shape was chosen for the simulation box and its volume minimized with Desmond tool, no ions were added to neutralize the system.

Table S1 Three stage protocol used for the simulated annealing simulation

Stages	1	2	3
Duration (ps)	100	300	600
Temperature (K)	300-400	450-300	300

NVT ensemble and Berendsen thermostat method were used for temperature coupling with a relaxation time of 1 ps. A cutoff of 9 Å for van der Waals interactions was applied, and the particle mesh Ewald algorithm was used for Coulomb interactions with a switching distance of 9 Å. The total simulation time was 1 ns with recording interval energy 1.2 ps and recoding trajectory of 5 ps. The final structure obtained after SA was submitted for a 1 ns MD simulation.

For the trajectory energy values were recorded every 1.2 ps a structure every 4.8 ps. The simulation was conducted at 300 K with a time-step bonded of 2 fs. NPT ensemble was used and the default relaxation model applied. The 'Nose-Hoover chain' thermostat method and 'Martyna-Tobias-Klein' Barostat method with 2 ps relaxation time and isotropic coupling style were used.

A cutoff of 9 Å for van der Waals interactions was applied, and the particle mesh Ewald algorithm was used for Coulomb interactions with a switching distance of 9 Å. No ions were added to the solution.

The Desmond application within Maestro (GUI) provides a Simulation Event Analysis tool for trajectory analysis that was used to measure end-to-end distances. Distances were measured by the manual selection of two atoms (bromine and a methylene carbon from the N-hydroxysuccinimide group) and values were determined by the Simulation Event Analysis tool. Average end-to-end distance value over the 1 ns simulation period was calculated and found to be 8.4 Å.

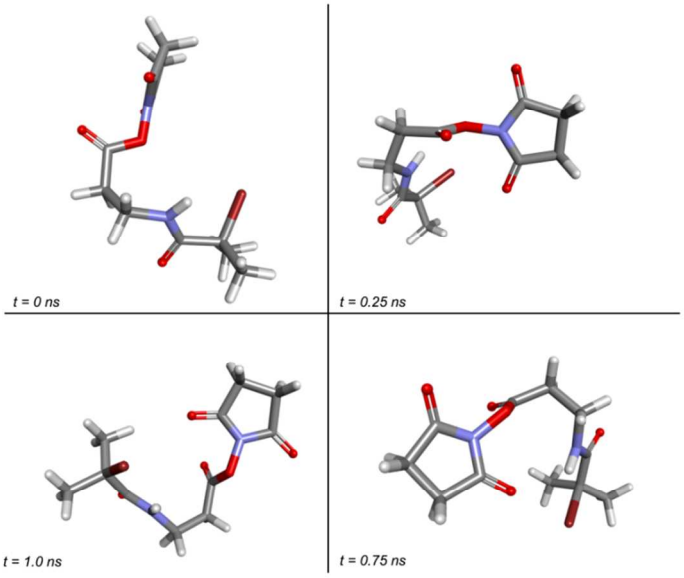


Figure S1 Structures of bromine-functionalized NHS initiator **1** obtained after a 1 ns MD simulation in water. Four time points are shown (t=0, 0.25, 0.75 and 1 ns) to provide an overview of the trajectory of the initiator.

MALDI-TOF Characterization for Initiator Modification of Lysozyme

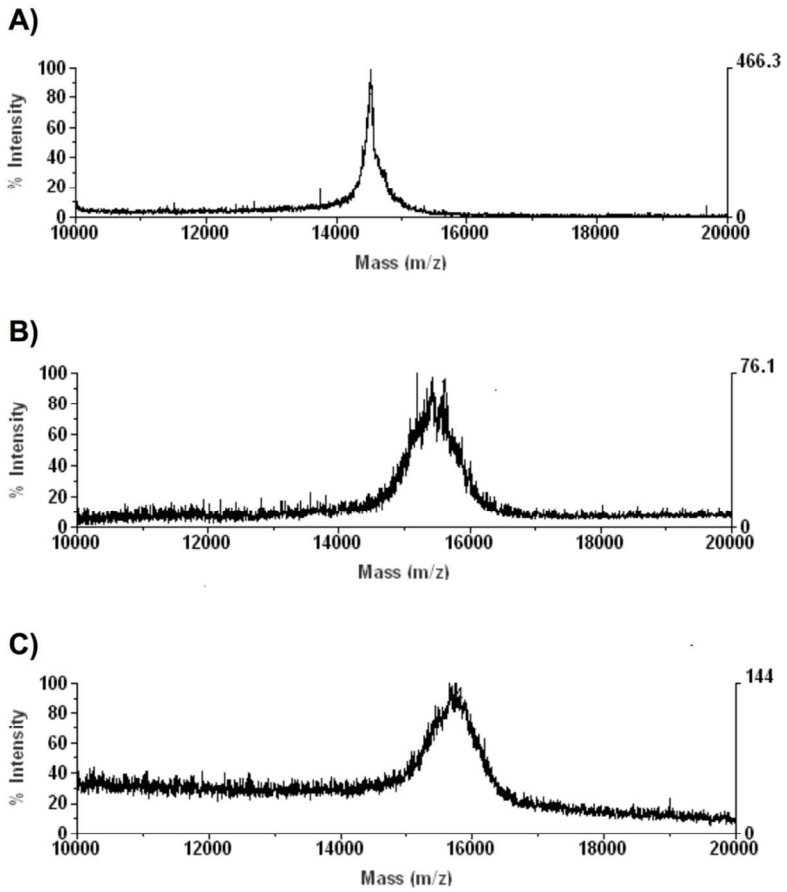


Figure S2 MALDI-TOF spectra for A) native Lysozyme, B) Initiator-modified Lysozyme with one ATRP initiator (Lyz-Br1) and C) Initiator-modified Lysozyme with five ATRP initiators (Lyz-Br5). The number of initiators in each macro-initiator complex was calculated by subtracting the Lyz-Br m/z values from native lysozyme m/z and dividing by the molecular weight of initiator **1** (Mw = 220.9 g/mol).

MALDI-TOF Characterization for Initiator Modification of Chymotrypsin

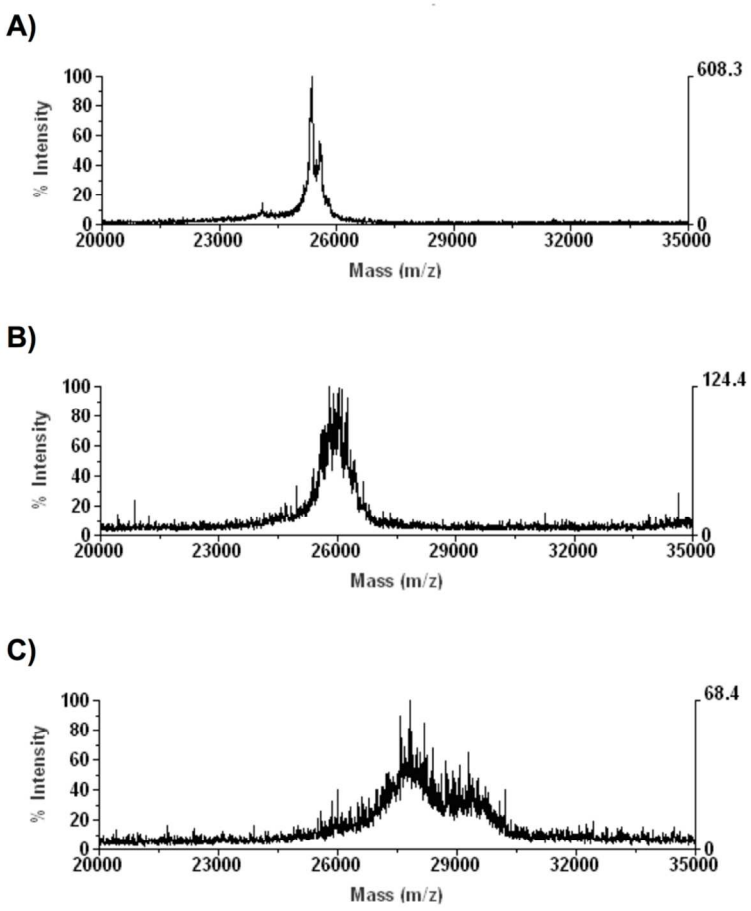


Figure S3 MALDI-TOF spectra for A) native Chymotrypsin, B) Initiator-modified Chymotrypsin with one ATRP initiator (CT-Br1) and C) Initiator-modified Chymotrypsin with eleven ATRP initiators (CT-Br11). The number of initiators in each macro-initiator complex was calculated by subtracting the CT-Br m/z values from native chymotrypsin m/z and dividing by the molecular weight of initiator **1** (Mw = 220.9 g/mol).

Trypsin and NTCB digestion studies with Lysozyme

Table S2 Summary of peptide fragments with theoretical and observed masses for native lysozyme after trypsin digestion

Peptide Fragment	Peptide Mass (Da)	Theoretical Mass	Observed mass
GYSLGNWVCAA**K**^{33}	1325.5	1330.37 $[M+H]^+$	1329.74
		696.77 $[M+3ACN+2H]^{2+}$	698.65
^{34}FESNFNTQATNR	1428.48	1433.35 $[M+NH4]^+$	1433.15
NLCNIPCSALLSSDITASVNCA**K**^{96}	2508.86	2513.73 $[M+NH4]^+$	2513.33
NLCNIPCSALLSSDITASVNCA**KK**^{97}	2655.05	2738.11 $[M+2ACN+H]^+$	2740.02
^{98}IVSDGNGMNAWVAWR	1675.88	1680.75 $[M+NH4]^+$	1681.17
		1634.76 $[M+ACN-Na]^+$	1633.51
^{117}GTDVQAWIR	1045.16	1050.03 $[M+NH4]^+$	1049.37
^{62}WWCNDGR	936.01	997.08 $[M+ACN+Na]^+$	997.66

Note: Fragments with cysteine residues were corrected to incorporate mass increases due to alkylation with iodoacetamide (+57.05 Da)

Table S3 Summary of peptide fragments with theoretical and observed masses for native lysozyme and initiator-modified lysozyme after NTCB digestion

Peptide Fragment	Peptide Mass (Da)	Theoretical Mass	Observed mass
CELAAAM^{13}**K**RHGLDNYRGYSLGNWV	2725.1	2808.16 $[M+2ACN+H]^+$	2816.90
CELAAAM^{13}K***IBr***RHGLDNYRGYSLGNWV	2946.0	3029.06 $[M+2ACN+H]^+$	3034.48
CAA^{33}**K**FESNFNTQATNRNTDGSTDYGILQINSRWW	3910.2	3993.26 $[M+2ACN+H]^+$	3990.07
CAA^{33}**K*IBr***FESNFNTQATNRNTDGSTDYGILQINSRWW	4131.1	4210.12 $[M+2ACN+H]^+$	4206.36
CA^{96}**K**^{97}**K**IVSDGNGMNAWVAWRNR	2376.73	2460.79 $[M+2ACN+H]^+$	2473.37
CA^{96}**K**^{97}**K(*IBr*)**IVSDGNGMNAWVAWRNR	2597.63	2681.69 $[M+2ACN+H]^+$	2690.69
C^{116}**K*IBr***GTDVQAWIRG	1554.41	1638.46 $[M+2ACN+H]^+$	1654.49

Note: Fragments containing ***IBr*** were corrected to include mass increases due to initiator modification (+220.9 Da)

Trypsin and NTCB digestion studies with Chymotrypsin

Table S4 Summary of peptide fragments with theoretical and observed masses for native chymotrypsin after trypsin digestion

Peptide Fragment	Peptide Mass (Da)	Theoretical Mass	Observed mass
^{1}CGVPAIQPVLSGLSR	1553.84	1512.72 $[M+ACN+2Na]^{+}$	1517.93
IVNGEEAVPGSWPWQVSLQDK^{36}	2339.59	802.85 $[M+3Na]^{3+}$	800.09
TGFHFCGGSLINENWVVTAAHCGVTTSDVVVAGEF DQGSSSE**K**^{79}	4558.93	4641.99 $[M+2ACN+H]^{+}$	4661.91
VF**K**^{90}	392.50	785.01 $[2M+H]^{+}$	785.66
YNSLTINNDITLL**K**^{107}	1621.85	1685.87 $[M+ACN+Na]^{+}$	1675.84
LQQASLPLLSNTNC**K**^{169}	1686.94	1645.82 $[M+ACN+Na]^{+}$	1648.90
LQQASLPLLSNTNC**K**^{169}**K**^{170}	1833.13	1856.12 $[M+Na]^{+}$	1861.23
YWGT**K**^{175} I**K**^{177}	913.09	1890.20 $[2M+ACN+Na]^{+}$	1888.11
DAMICAGASGVSSCMGDSGGPLVC**K**^{202}	2487.83	1285.95 $[M+2ACN+2H]^{2+}$	1287.30
DAMICAGASGVSSCMGDSGGPLVC**K**^{202} **K**^{203}	2634.02	1379.56 $[M+3ACN+2H]^{2+}$	1378.49

Note: Fragments with cysteine residues were corrected to incorporate mass increases due to alkylation with iodoacetamide (+57.05 Da)

Table S5 Summary of peptide fragments with theoretical and observed masses for native chymotrypsin after NTCB digestion

Peptide Fragment	Peptide Mass (Da)	Theoretical Mass	Observed mass
^{1}**C**GVPAIQPVLSGLSRIVNGEEAVPGSWPWQVSLQD^{36}**K**TGFHF	4433.01	4495.03 $[M+ACN+Na]^+$	4497.11
C**K**^{202}**K**^{203}NGAWTLVGIVSWGSST	2019.3	2064.27 $[M+2Na+H]^+$	2067.11

MALDI-TOF as a semi-quantitative technique

Native lysozyme solutions (0 – 2 mg/mL) were prepared in deionized water. For MALDI-TOF analysis, 10 µL of each lysozyme solution was mixed with 5 µL of trypsin standard solution (10 mg/mL, in deionized water) and 15 µL of the matrix solution. Sample solutions were vortexed to ensure the samples were well-mixed and the dried droplet method was used to prepare the sample target: 2 µL of analyte mixture was applied to the MALDI-TOF plate and air-dried. MALDI-TOF experiment was carried out as described in the experimental section and a calibration curve was plotted using Excel (Microsoft Office).

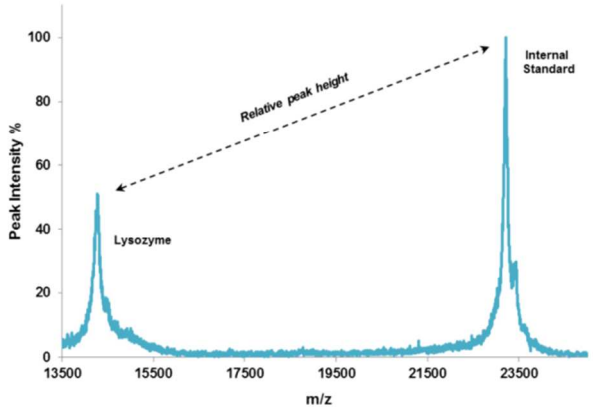


Figure S4 MALDI-TOF spectrum showing the relative quantification of lysozyme to trypsin using peak heights.

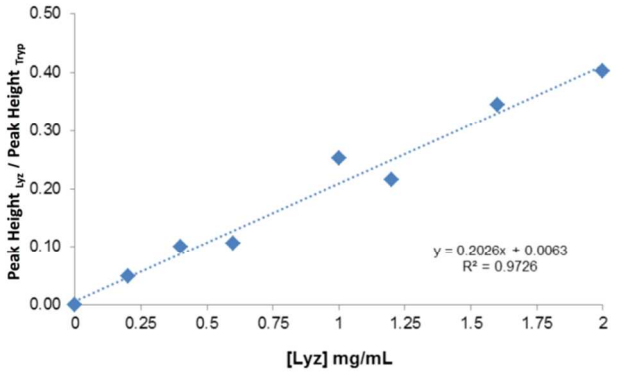


Figure S5 Calibration curve for native lysozyme using peak intensity measurements against trypsin standard, depicting the semi-quantitative nature of MALDI-TOF.

Aminolysis Rates for Lysozyme and Chymotrypsin

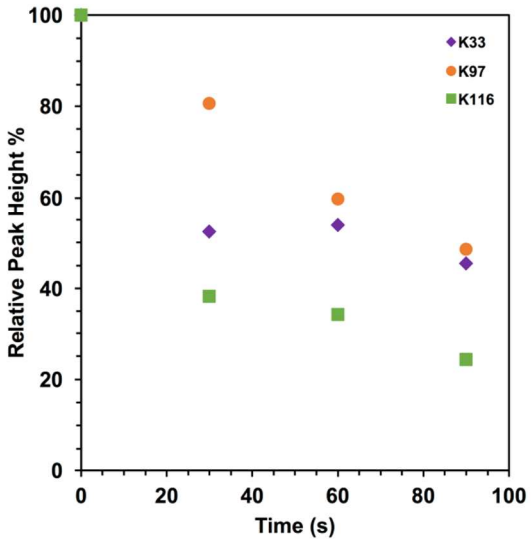


Figure S6 Aminolysis rates for lysine residues K33, K97 and K116 in lysozyme determined by tryptic digestion followed by MALDI-TOF analysis. The corresponding peptide fragments were identified and their relative peak height relative to a non-modified peptide fragment was monitored over time.

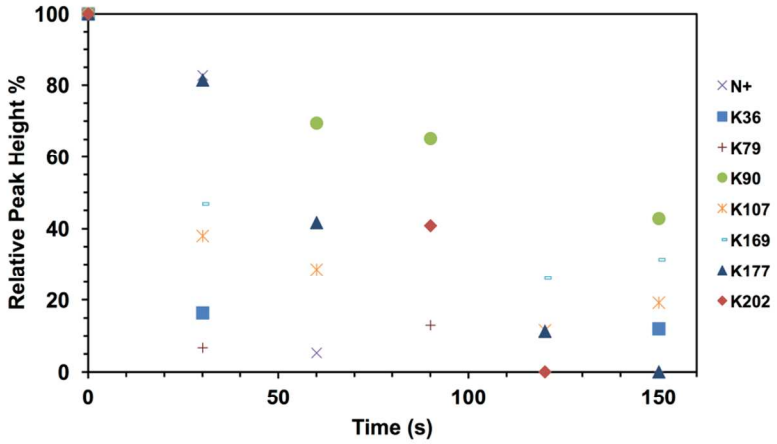


Figure S7 Aminolysis rates for the N-terminus and K36, K79, K90, K107, K169, K177 and K202 in CT were determined by tryptic digestion followed by MALDI-TOF analysis. The corresponding peptide fragments were identified and their relative peak height relative to a non-modified peptide fragment was monitored over time.

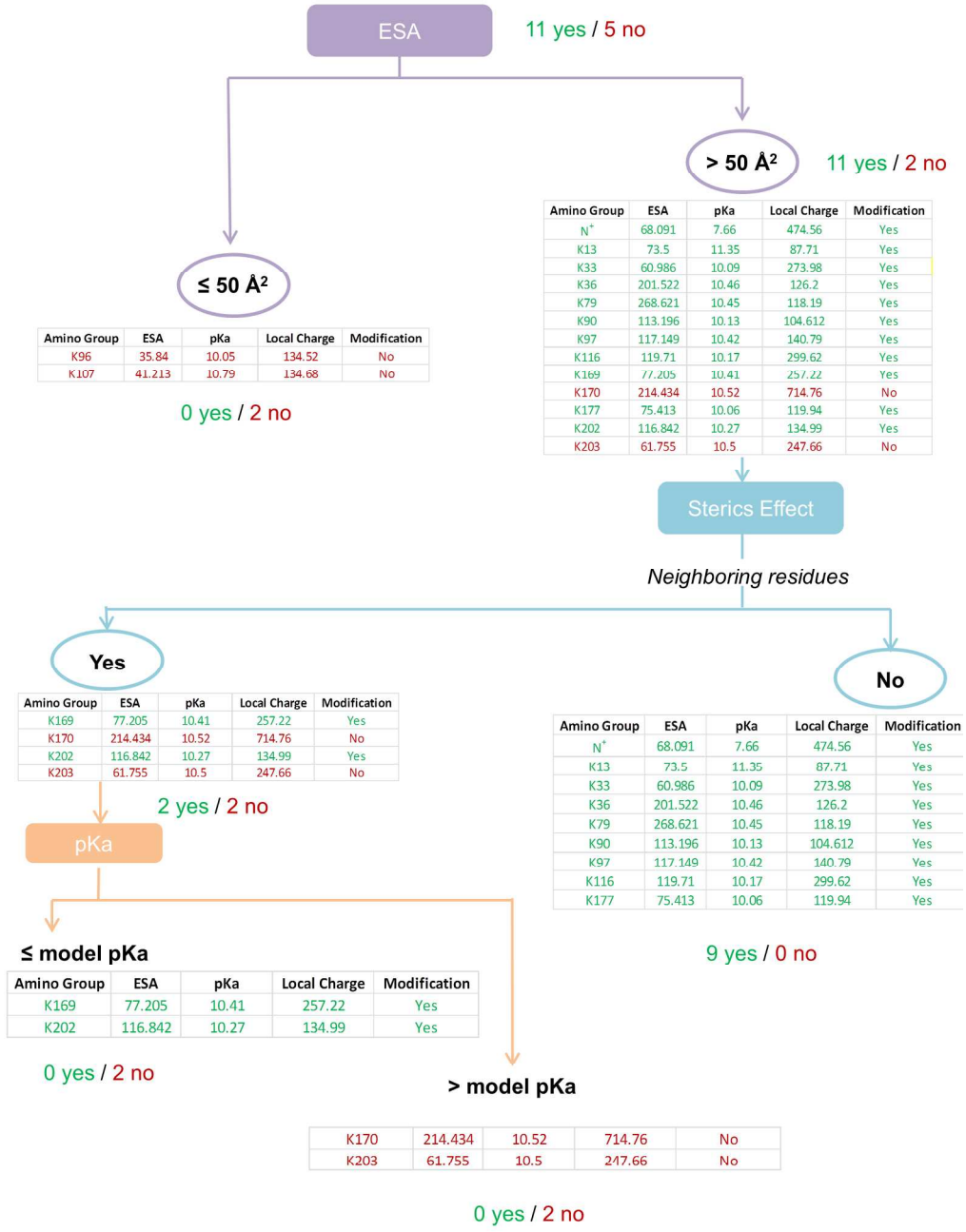
Decision Tree Development

To develop the decision tree to predict initiator-protein specificity, attributes (ESA, pKa, steric hindrance, secondary structure and the amino groups' microenvironment) of the amino groups from both lysozyme and chymotrypsin were used as a dataset. The training dataset was therefore compiled of a total of 15 amino groups (11 amino groups modified and 4 non-modified).

Amino Group	ESA	pKa	Local Charge	Modification
N^+	68.091	7.66	474.56	Yes
K13	73.5	11.35	87.71	Yes
K33	60.986	10.09	273.98	Yes
K36	201.522	10.46	126.2	Yes
K79	268.621	10.45	118.19	Yes
K90	113.196	10.13	104.612	Yes
K96	35.84	10.05	134.52	No
K97	117.149	10.42	140.79	Yes
K107	41.213	10.79	134.68	No
K116	119.71	10.17	299.62	Yes
K169	77.205	10.41	257.22	Yes
K177	75.413	10.06	119.94	Yes
K170	214.434	10.52	714.76	No
K202	116.842	10.27	134.99	Yes
K203	61.755	10.5	247.66	No

Figure S8 Complete dataset obtained from theoretical and experimental studies with lysozyme and chymotrypsin

Question 1: How to predict if an amino group is modified?



Amino Group	ESA	pKa	Local Charge	Modification
K96	35.84	10.05	134.52	No
K107	41.213	10.79	134.68	No

Amino Group	ESA	pKa	Local Charge	Modification
N^+	68.091	7.66	474.56	Yes
K13	73.5	11.35	87.71	Yes
K33	60.986	10.09	273.98	Yes
K36	201.522	10.46	126.2	Yes
K79	268.621	10.45	118.19	Yes
K90	113.196	10.13	104.612	Yes
K97	117.149	10.42	140.79	Yes
K116	119.71	10.17	299.62	Yes
K169	77.205	10.41	257.22	Yes
K170	214.434	10.52	714.76	No
K177	75.413	10.06	119.94	Yes
K202	116.842	10.27	134.99	Yes
K203	61.755	10.5	247.66	No

Amino Group	ESA	pKa	Local Charge	Modification
K169	77.205	10.41	257.22	Yes
K170	214.434	10.52	714.76	No
K202	116.842	10.27	134.99	Yes
K203	61.755	10.5	247.66	No

Amino Group	ESA	pKa	Local Charge	Modification
N^+	68.091	7.66	474.56	Yes
K13	73.5	11.35	87.71	Yes
K33	60.986	10.09	273.98	Yes
K36	201.522	10.46	126.2	Yes
K79	268.621	10.45	118.19	Yes
K90	113.196	10.13	104.612	Yes
K97	117.149	10.42	140.79	Yes
K116	119.71	10.17	299.62	Yes
K177	75.413	10.06	119.94	Yes

Amino Group	ESA	pKa	Local Charge	Modification
K169	77.205	10.41	257.22	Yes
K202	116.842	10.27	134.99	Yes

K170	214.434	10.52	714.76	No
K203	61.755	10.5	247.66	No

Figure S9 Prediction of modification of an amino group. The training dataset was split into subsets according to their attributes until pure subsets were obtained.

Question 2: How to predict the rate of amino modification?

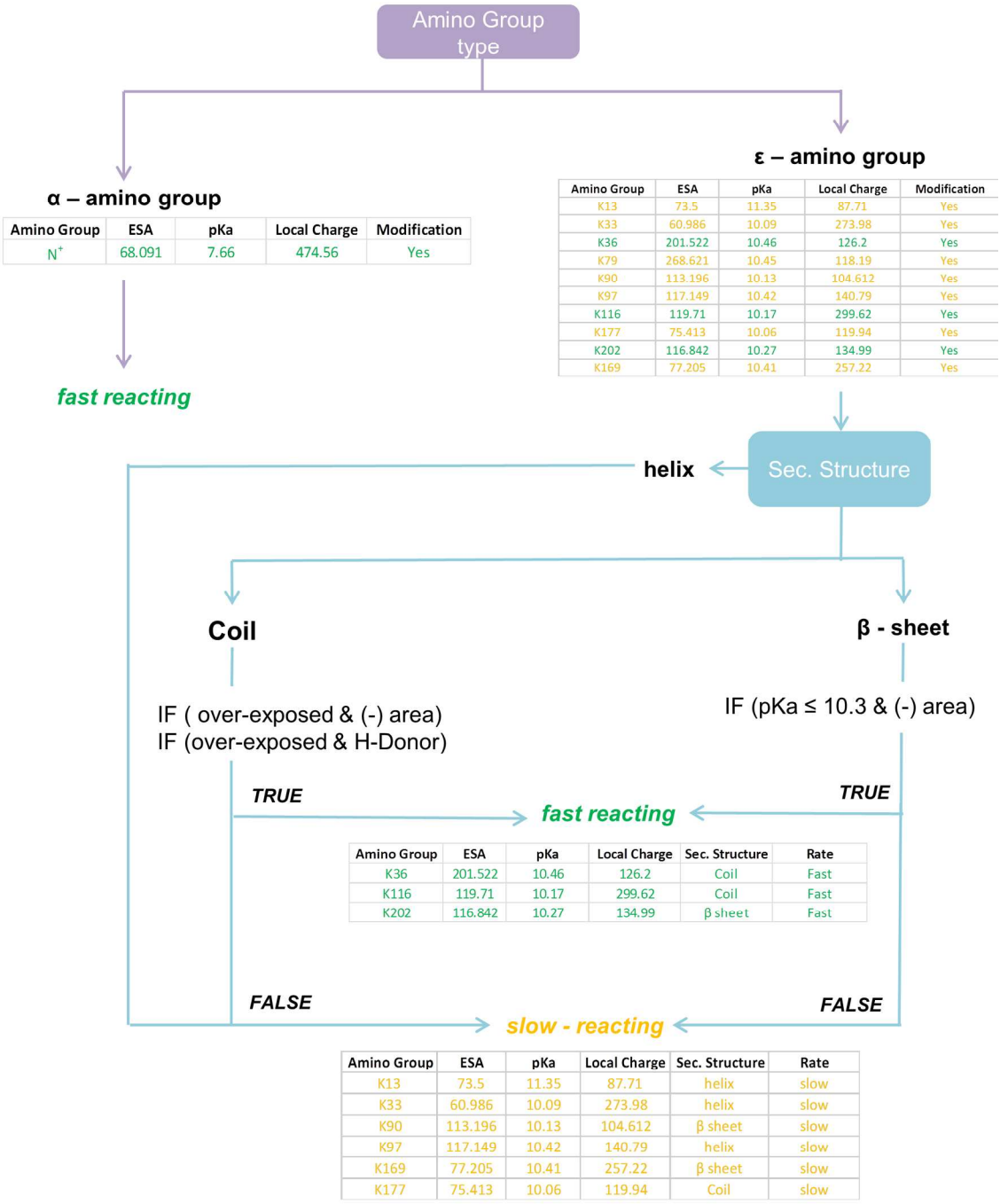


Figure S10 Prediction of the relative rate of modification of an amino group. The rate of reaction was found to be dependent on the type of amino group, the secondary structure and the amino groups' micro-environment.

Molecular Dynamics (MD) Simulation of Native CT

The initial structure for native chymotrypsin was obtained from the PDB Database (PDB ID: 4CHA). To remove any bias or constraints, the initiator structure was subjected to a simulated annealing (SA) protocol using Desmond (Maestro, Schrodinger)4 (

Table **S1**). The system for simulation was prepared using Desmond's system builder. The OPLS-2005 force field was used and SPC was chosen as a solvent model. An orthorhombic shape was chosen for the simulation box and its volume minimized with Desmond tool, no ions were added to neutralize the system.

NVT ensemble and Berendsen thermostat method were used for temperature coupling with a relaxation time of 1 ps. A cutoff of 9 Å for van der Waals interactions was applied, and the particle mesh Ewald algorithm was used for Coulomb interactions with a switching distance of 9 Å. The total simulation time was 1 ns with recording interval energy 1.2 ps and recoding trajectory of 5 ps. The final structure obtained after SA was submitted for a 20 ns MD simulation.

For the trajectory energy values were recorded every 1.2 ps a structure every 4.8 ps. The simulation was conducted at 300 K with a time-step bonded of 2 fs. NPT ensemble was used and the default relaxation model applied. The 'Nose-Hoover chain' thermostat method and 'Martyna-Tobias-Klein' Barostat method with 2 ps relaxation time and isotropic coupling style were used. A cutoff of 9 Å for van der Waals interactions was applied, and the particle mesh Ewald algorithm was used for Coulomb interactions with a switching distance of 9 Å. No ions were added to the solution.

The Desmond application within Maestro (GUI) provides a Simulation Event Analysis tool for trajectory analysis that was used to measure end-to-end distances. Distances were measured by

the manual selection of the ε-amino groups in lysine residues K90 and K93, K169 and K170 and K202 and K203. Values were determined by the Simulation Event Analysis tool and plotted using Microsoft Office Excel (Figure S11)

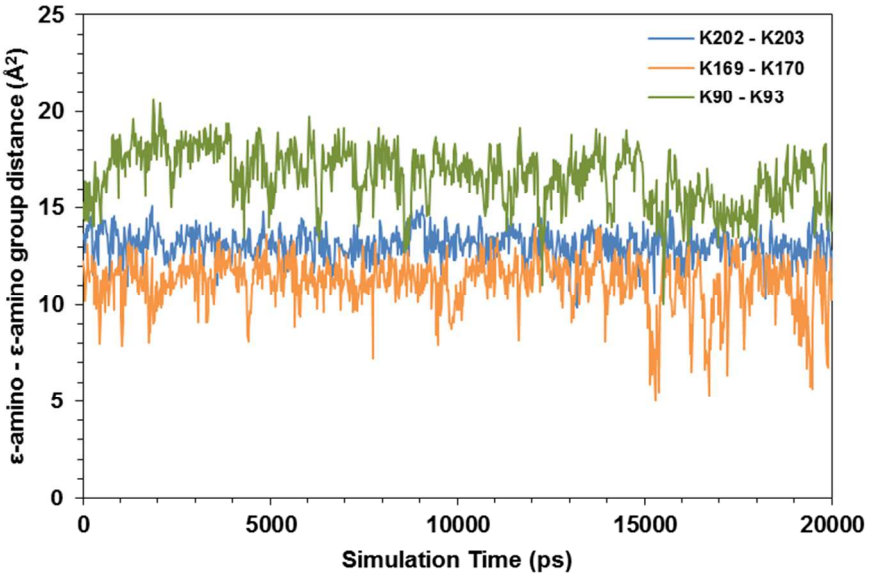


Figure S11 End-to-end distances from the ε-amino groups in lysine residues K90 and K93, K 169 and K170 and K202 and K203 determined after a 20 ns molecular dynamics simulation of native chymotrypsin in water. Distances were found to be closer for neighboring residues, which can contribute towards a steric hindrance effect.