Supporting information

For

A High-Throughput Platform for Efficient Exploration of Functional Polypeptides Chemical Space via Automation and Machine Learning

Table of Contents

Materials and Instrumentation

Materials

All chemicals were purchased from commercial sources and used as received unless otherwise specified. NADPH tetrasodium salt was purchased from Bide pharmatech. Triethylene glycol was purchased from Shanghai Hushi Co. Ltd. Glutathione reductase (GR) from baker's yeast was purchased from Sigma-Aldrich (product number: G3664, 196 units/mg). The aqueous solution of *tert*-Butyl hydroperoxide (TBHP, 70%) was purchased from Macklin Biochemical Co., Ltd. Phosphate buffer saline (PBS, 20 ×) was purchased from Sangon Biotech (Shanghai) Co., Ltd. Zeba Spin Desalting Columns Plates (7K MWCO) was purchased from ThermoFisher Scientific. PD-10 desalting columns was purchased from Cytiva. The high throughput synthesis and purification were carried out in Nunc™ 96-Well Polypropylene Storage Microplates (product number: 267245) equipped with Nunc™ 96-Well Cap Mats (product number: 276002). The GPx activity assay was performed in Corning® 96-well Clear Flat Bottom Polystyrene TC-treated Microplates (product number: 3599). The fluorescent intensity was recorded in Corning® 96-well Solid Black Flat Bottom Polystyrene TC-treated Microplates (product number: 3916).

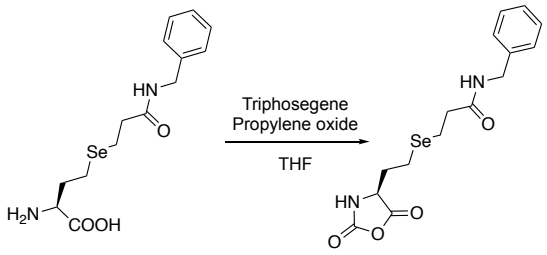
Instruments

^{1}H and ^{13}C NMR spectra were recorded on a 400 MHz Bruker ARX400 FT-NMR spectrometer. Fourier transform infrared spectra (FT-IR) were recorded on a Bruker ALPHA II with ATR. High-resolution mass spectrometry (HR-MS) analyses were performed on Q-Exactive Plus (Thermo Scientific). Size exclusion chromatography (SEC) experiments were performed on a system equipped with an isocratic pump (Model 1100, Agilent Technology, Santa Clara, CA), a DAWN HELEOS 9-angle laser light scattering detector (MALLS, Wyatt Technology, Santa Barbara, CA), and an Optilab rEX refractive index detector (Wyatt Technology, Santa Barbara, CA). The detection wavelength of MALLS was 658 nm and the temperature of both the refractive index and the MALLS detectors was 25 °C. Separations were performed using serially connected size exclusion columns (500, 10^3,10^4, and 10^5 Å Phenogel columns, 5 μm, 7.8 × 300 mm, Phenomenex, Torrance, CA) at a flow rate of 1.0 mL/min and 50 °C using DMF containing 0.1 M LiBr as the mobile phase. The dn/dc value was measured offline by using the internal calibration system (by the ASTRA V software version 5.1.7.3 provided by Wyatt Technology). Circular dichroism (CD) spectra (190 - 250 nm) were recorded on a circular dichroism spectrometer J-815 (JASCO) using a quartz cuvette of 0.1 cm path length. The spectra were reported as an average of 3 scans and in units of molar ellipticity [θ] (deg cm^2 $dmol^{-1}$). GPx-like activity was analyzed on an EnSpire multimode plate reader from PerkinElmer. The liquid handling was performed on a JANUS automated workstation by PerkinElmer. The plate for high throughput synthesis was incubated in Multitron Standard Incubator shaker (Infors HT).

Experimental Procedures

Synthesis of P(pAm-SeHC)

Synthesis of (*S*)-*N*-benzyl-3-((2-(2,5-dioxooxazolidin-4-yl)ethyl)selanyl)propenamide (pAm-SeHC NCA)



In a 75 mL seal tube, pAm-SeHC (1.0 g, 2.9 mmol) and triphosgene (0.4, 1.3 mmol) were suspended in THF (30 mL). To the system was added propylene oxide (~2.0 mL, ~29 mmol) and stirred at room temperature for 3 hours. The solvent was removed and the product was purified by silica gel chromatography (mobile phase: EA/PE). Pure pAm-SeHC NCA was recovered as a light-yellow oil (0.78g, yield: 72%). Caution: silica gel needs to be dried at 90 °C for at least 4 hours before use.
HR-ESI-MS $[M + H]^+$ Calcd. for $C_{15}H_{19}N_2O_4Se$, 371.0505, found 371.0506. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.96 (s, 1H), 7.32 – 7.16 (m, 5H), 6.95 (t, 1H), 4.41 – 4.25 (m, 3H), 2.83 – 2.50 (m, 6H), 2.16 – 1.89 (m, 2H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 172.32, 170.63, 152.47, 137.95, 128.69, 127.52, 57.18, 43.60, 37.18, 31.96, 19.21, 18.89, 18.70.

Ring opening polymerization of pAm-SeHC NCA

The polymerization was carried out under nitrogen flow as reported by Wooley *et al.*[1] In a 10 mL Schlenk flask equipped with a magnetic stir bar, pAm-SeHC NCA was dissolved in dry DMF (100 mg/mL). To the solution was quickly added the initiator benzylamine or the macroinitiator mPEG-NH_2 (M_n 5000) in dry DMF (0.1 M). The polymerization was stirred under nitrogen flow for 4 hours and monitored by FT-IR. After the full conversion of monomer, the polypeptide was precipitate and washed with ether twice. The M_n of the polypeptide was characterized by size exclusion chromatography. After dry in vacuum, The product P(pAm-SeHC) was obtained as white sticky gum (yield: ~60%).

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.53 – 7.88 (m, 2H), 7.23 (m, 5H), 4.48 – 3.96 (m, 3H), 2.93 – 2.43 (m, 8H, including signal of DMSO), 2.41 – 1.73 (m, 2H).

PEG-*b*-P(pAm-SeHC)
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.26 – 7.06 (m, 5H), 4.44 – 3.96 (m, 3H), 3.71 – 3.54 (m, 23H), 3.39 – 3.36 (m, 0.13H), 2.87 – 1.95 (m, 8H).

Selenoxide Elimination of the Polypeptides

Selenoxide elimination of P(pAm-SeHC)

p(pAm-SeHC) obtained from ROP of 680 mg pAm-SeHC NCA was dissolved in 50 mL THF and 100 mL chloroform. To the solution was added 70% TBHP (2760 μL, 12 eq of Se) and stirred at room temperature for 2 hours before TEA (800 μL, 3 eq of Se) and 75 mL $NaHCO_3$ solution (1M) were added. The system was stirred under 37 °C for 16 hours during which a clear phase separation was observed, which was used as an indication of the completion of the reaction. The aqueous phase was washed with 150 mL DCM, and dialyzed (MWCO 3500 Da) against 0.5 M NaCl for 12 hours, followed by water for 48 hours with changed twice every day. The remaining content was lyophilized to give the final product as a pale yellow powder (256 mg, yield: 64% from NCA).

^{1}H NMR (400 MHz, D_2O) δ 4.49 – 4.18 (m, 1H), 2.80 – 2.47 (m, 2H), 2.28 – 1.90 (m, 2H).
^{77}Se NMR (95 MHz, D_2O) δ 1188.03.

Selenoxide elimination of PEG-*b*-P(pAm-SeHC)

To avoid extensive emulsification, PEG-*b*-$PSeO_2Na$ was prepared using a slightly modified procedure, in which the oxidized product in chloroform was firstly precipitated with ether followed by the elimination in a basic aqueous solution. PEG-*b*-P(pAm-SeHC), prepared by mPEG-NH_2 (M_n 5000, 293 mg)-initiated ROP of pAm-SeHC NCA (870 mg), was dissolved in 200 mL THF and to the solution was added 70% TBHP (3025 μL, 23.5 mmol). The solution was stirred at room temperature for 2.5 hours and then TEA (990 μL, 7.0 mmol) was added. After another 14 hours, the solution was condensed by evaporation and precipitated with ether. The precipitate was re-dissolved in premixed H_2O/methanol (100 mL, v/v = 1:1) and pH was adjusted to ~8 by NaOH. To the solution was added 70% TBHP (3025 μL, 23.5 mmol). After 20 hours stirring at room temperature, the system was dialyzed (MWCO 3500 Da) against 0.5 M NaCl for 12 hours follow by water for 48 hours with water changed twice every day. After lyophilization, the product was obtained as a light-yellow powder (409 mg, yield: 50% from NCA).

^{1}H NMR (400 MHz, D_2O) δ 4.47 – 4.29 (m, 1H), 3.71 – 3.51 (m, 14H), 3.35 – 3.28 (m, 0.07H), 2.97 – 2.73 (m, 2H), 2.37 – 2.07 (m, 2H).

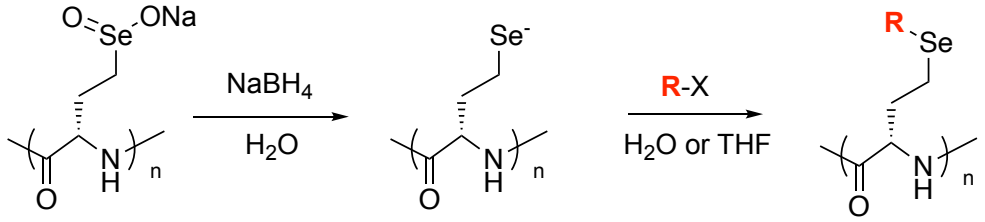
General procedure for fluorescence labeling

The fluorescent labeling can be performed before or after the selenoxide elimination. For the labeling before elimination, the system after polymerization was added with 5(6)-carboxyfluorescein *N*-succinimidyl ester (FAM-SE, 1eq to the initiator) and stirred overnight at room temperature. Then the selenoxide elimination was carried out by following the general procedure described above and the free fluorescein was removed by SEC (0.01 M NaOH and 0.15M NaCl, Superdex^{TM} 75, HiLoad^{TM} 16/600). The collected solution was dialyzed against deionized water for 24 hours and lyophilized to get the final product. For the labeling after elimination, the solution of $PSeO_2Na$ (or PEG-$PSeO_2Na$) was dialyzed against water for 24 hours and condensed to ~ 5 mL by rotatory evaporation. The 5(6)-carboxyfluorescein *N*-succinimidyl ester (1eq to the initiator) was dissolved in 2 mL DMSO and added to the system on ice. The solution was protected from light and stirred for 24 hours. Then it was dialyzed against water for 48 hours with water changed twice every day. The free fluorescein was

removed by SEC (0.01 M NaOH and 0.15M NaCl, Superdex^{TM} 75 pg, HiLoad^{TM} 16/600). The final product was recovered by lyophilization.

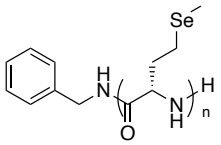
Post Polymerization Modification (PPM)

General procedure for post polymerization modification of $PSeO_2Na$

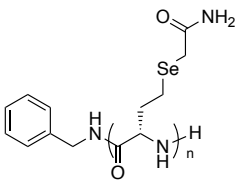


Under the protection of nitrogen, $PSeO_2Na$ was dissolved in water (5-10 mg/mL). To the solution was added $NaBH_4$ (5 mg $NaBH_4$ for 10 mg $PSeO_2Na$) followed by the addition of TFA (1.5 μL for 5 mg $NaBH_4$, **caution: gas emission**). Precipitation was observed in ~10 minutes. After 20 minutes, another portion of TFA (1.5 μL) was added. The system was stirred at room temperature for 30 minutes and the completion of the reduction was indicated by the re-dissolving of the precipitate. Then to the polymer solution was added the modifier in THF or water (specified below). The system was stirred under indicated temperature. The selenopolypeptide was purified by dialysis and/or SEC and recovered by lyophilization. The post polymerization modification of PEG-*b*-$PSeO_2Na$ was carried out similarly.

Post polymerization modification for homopolypeptides

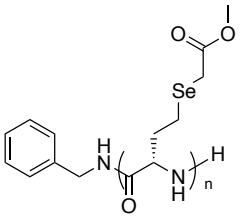


P1 (poly-selenomethionine)
P1 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. Iodomethane (19.0 mg, 0.135 mmol) was dissolved in 2 mL THF and added to the reduced polymer solution. The reaction time was set for 40 minutes to avoid over-modification. The polymer was obtained by centrifugation and washed with water. After lyophilization, the product was recovered as a white powder (6.4 mg, yield: 79%).
^{1}H NMR (400 MHz, TFA-*d*) δ 5.00 – 4.75 (m, 1H), 2.85 – 2.56 (m, 2H), 2.40 – 2.19 (m, 2H), 2.17 – 1.97 (m, 3H).

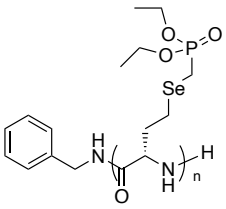


P2
P2 was prepared by following the general procedure for modification with 10 mg $PSeO_2Na$. Iodoacetamide (25.0 mg, 0.135 mmol) in 0.5 mL water was added to the reduced polymer

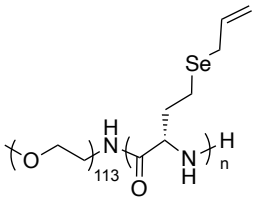
solution and stirred for 60 minutes. The polymer was obtained by centrifugation and washed with water. After lyophilization, the product was recovered as a white powder (7.5 mg, yield: 73%).
^{1}H NMR (400 MHz, TFA-*d*) δ 4.96 – 4.59 (m, 1H), 3.64 – 3.36 (m, 2H), 3.10 – 2.75 (m, 2H), 2.58 – 2.17 (m, 2H).



P3
P3 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. Methyl bromoacetate (20.5 mg, 0.135 mmol) in 2 mL THF was added to the reduced polymer solution and stirred for 6 hours. The polymer was purified by dialysis against water (MWCO 3500 Da) for 24 hours. After lyophilization, the product was recovered as a white powder (6.8 mg, yield: 62%).
^{1}H NMR (400 MHz, TFA-*d*) δ 5.02 – 4.76 (m, 1H), 4.09 – 3.77 (m, 3H), 3.54 – 3.21 (m, 2H), 3.07 – 2.76 (m, 2H), 2.55 – 2.15 (m, 2H).

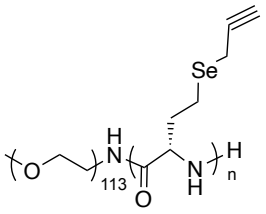


P4
P4 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. Diethyl iodomethylphosphonate (38 mg, 0.135 mmol) in 2 mL THF was added to the reduced polymer solution and stirred for 15 hours under 50 °C. The system was washed with 15 mL ether for twice and then dialyzed against water for 24 hours. After lyophilization, the product was recovered as a light yellow powder (7.0 mg, yield: 48%).
^{1}H NMR (400 MHz, TFA-*d*) δ 5.01 – 4.64 (m, 1H), 4.49 – 4.11 (m, 4H), 3.12 – 2.13 (m, 6H), 1.57 – 1.21 (m, 6H).



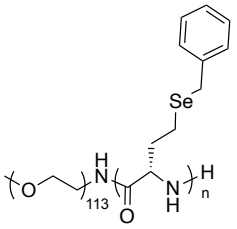
P5
P5 was prepared by following the general PPM procedure with 10 mg PEG-*b*-$PSeO_2Na$ (molecular weight of PEG = 5000 Da, DP of $PSeO_2Na$ = 40). Allyl bromide (9.8 mg, 0.081 mmol) in 2 mL THF was added to the reduced polymer solution and stirred for 3 hours. The polymer was purified by dialysis against water (MWCO 3500 Da) for 24 hours. After lyophilization, the product was recovered as a white powder (9.7 mg, yield: 97%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.95 – 5.79 (m, 1H), 5.14 – 4.86 (m, 2H), 4.24 – 3.92 (m, 1H), 3.87 – 3.44 (m, 18H), 3.39 – 3.37 (m, 0.11H), 3.30 – 3.06 (m, 2H), 2.83 – 2.46 (m, 2H), 2.44 – 2.03 (m, 2H).



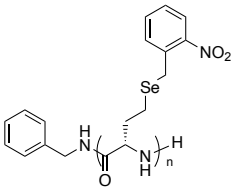
P6

P6 was prepared by following the general PPM procedure with 10 mg PEG-*b*-$PSeO_2Na$ (molecular weight of PEG = 5000 Da, DP of $PSeO_2Na$ = 40). Propargyl bromide (9.6 mg, 0.081 mmol) in 2 mL THF was added to the reduced polymer solution and stirred for 4 hours. The polymer was purified by dialysis against water (MWCO 3500 Da) for 24 hours. After lyophilization, the product was recovered as a white powder (9.4 mg, yield: 97%).
^{1}H NMR (400 MHz, $CDCl_3$) δ 4.22 – 4.02 (m, 1H), 3.87 – 3.44 (m, 22H), 3.38 – 3.38 (m, 0.12H), 3.33 – 3.13 (m, 2H), 3.06 – 2.77 (m, 2H), 2.55 – 2.11 (m, 3H).



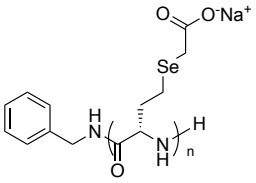
P7

P7 was prepared by following the general PPM procedure with 10 mg PEG-*b*-$PSeO_2Na$ (molecular weight of PEG = 5000 Da, DP of $PSeO_2Na$ = 40). Benzyl bromide (13.8 mg, 0.081 mmol) in 2 mL THF was added to the reduced polymer solution and stirred for 3 hours. The polymer was purified by dialysis against water (MWCO 3500 Da) for 24 hours. After lyophilization, the product was recovered as a white powder (8.2 mg, yield: 75%).
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.25 – 7.04 (m, 5H), 4.28 – 3.91 (m, 1H), 3.86 – 3.44 (m, 13H), 3.39 – 3.36 (m, 0.1H), 2.85 – 1.92 (m, 4H).



P8

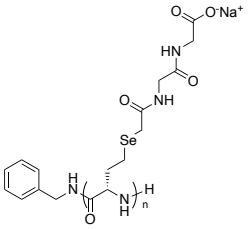
P8 was prepared by following the general PPM procedure 10 mg $PSeO_2Na$. 2-Nitrobenzyl bromide (33 mg, 0.135 mmol) in 2 mL THF was added to the reduced polymer solution and stirred for an hour. The polymer was obtained by centrifugation and washed with water. After lyophilization, the product was recovered as a yellow powder (10.8 mg, yield: 78%).
^{1}H NMR (400 MHz, $CDCl_3$/TFA-*d*) δ 8.11 – 7.78 (m, 1H), 7.62 – 7.29 (m, 3H), 4.53 – 4.21 (m, 1H), 4.19 – 3.88 (m, 2H), 2.92 – 2.50 (m, 2H), 2.47 – 2.05 (m, 2H).



P9

P9 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. Bromoacetic acid (19 mg, 0.135 mmol) in 0.1 mL water was added to the reduced polymer solution. The solution immediately became opaque. Then 27 μL 5 M NaOH solution was added, which turn the system back to a clear solution. The system was stirred for 4.5 hours. The polymer was purified by dialysis against water (MWCO 3500 Da) for 24 hours. After lyophilization, the product was recovered as a white powder (10.0 mg, yield: 89%).

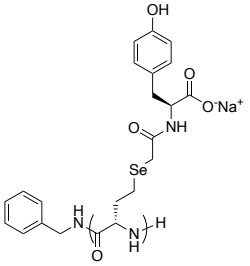
^{1}H NMR (400 MHz, D_2O/NaOD) δ 4.44 – 4.25 (m, 1H), 3.15 – 2.99 (m, 2H), 2.71 – 2.45 (m, 2H), 2.15 – 1.91 (m, 2H).



P10

P10 was prepared by following the general PPM procedure with 5 mg $PSeO_2Na$. **Hal-10** (5.7 mg, 0.027 mmol) was dissolved in 0.5 mL NaOH solution and added to the reduced polymer solution. The system was stirred for 5 hours under room temperature. The polymer was purified by PD-10. After lyophilization, the product was recovered as a fluffy powder (6.4 mg, yield: 78%).

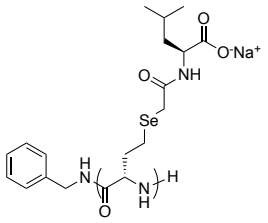
^{1}H NMR (400 MHz, D_2O) δ 4.46 – 4.26 (m, 1H), 3.98 – 3.78 (m, 2H), 3.75 – 3.63 (m, 2H), 3.33 – 3.14 (m, 2H), 2.81 – 2.49 (m, 2H), 2.23 – 1.89 (m, 2H).



P11

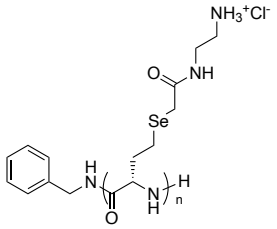
P11 was prepared by following the general PPM procedure with 5 mg $PSeO_2Na$. **Hal-11** (7.1 mg, 0.027 mmol) was dissolved in 0.5 mL water and added to the reduced polymer solution. The system was stirred for 5 hours under room temperature. The polymer was purified by PD-10. After lyophilization, the product was recovered as a fluffy powder (8.0 mg, yield: 86%).

^{1}H NMR (400 MHz, D_2O) δ 7.03 – 6.84 (m, 2H), 6.71 – 6.54 (m, 2H), 4.39 – 4.13 (m, 2H), 3.16 – 2.84 (m, 3H), 2.77 – 2.56 (m, 1H), 2.49 – 2.16 (m, 2H), 2.12 – 1.71 (m, 2H).



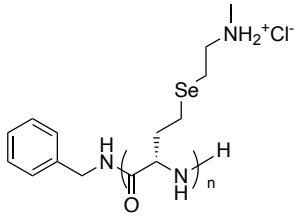
P12

P12 was prepared by following the general PPM procedure with 5 mg $PSeO_2Na$. **Hal-12** (5.7 mg, 0.027 mmol) in 0.5 mL water was added to the reduced polymer solution. The system was stirred for 2 hours under room temperature. The polymer was purified by PD-10. After lyophilization, the product was recovered as a fluffy powder (7.1 mg, yield: 87%).
^{1}H NMR (400 MHz, D_2O) δ 8.44 – 7.77 (m, 1H), 4.28 – 3.90 (m, 2H), 3.44 – 3.03 (m, 2H), 2.97 – 2.54 (m, 2H), 2.51 – 1.97 (m, 2H), 1.77 – 1.30 (m, 3H), 1.09 – 0.60 (m, 6H).



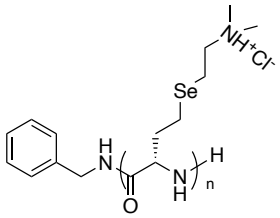
P13

P13 was prepared by following the general PPM procedure with 5 mg $PSeO_2Na$. **Hal-13** (4.8 mg, 0.027 mmol) in 0.5 mL water was added to the reduced polymer solution. The solution immediately became opaque and turned back to clear after 2 minutes. The system was stirred for 2 hours under room temperature. The polymer was purified by PD-10. After lyophilization, the product was recovered as a fluffy powder (5.0 mg, yield: 72%).
^{1}H NMR (400 MHz, D_2O) δ 4.42 – 4.19 (m, 1H), 3.45 – 3.32 (m, 2H), 3.29 – 3.13 (m, 2H), 3.08 – 2.91 (m, 2H), 2.86 – 2.54 (m, 2H), 2.32 – 1.91 (m, 2H).



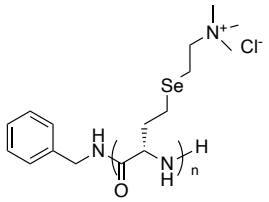
P14

P14 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. 2-Bromo-*N*-methylethanamine hydrobromide (30.0 mg, 0.135 mmol) in 0.5 mL water was added to the reduced polymer solution. The solution became opaque immediately and turned back to clear after an hour. The system was stirred for 9 hours under room temperature. The polymer was dialyzed against 0.5 M NaCl solution for 12 hours followed by water for 48 hours. After lyophilization, the product was recovered as a white powder (7.0 mg, yield: 60%).
^{1}H NMR (400 MHz, D_2O) δ 4.47 – 4.34 (m, 1H), 3.30 – 3.18 (m, 2H), 2.88 – 2.75 (m, 2H), 2.72 – 2.49 (m, 5H), 2.15 – 1.99 (m, 2H).



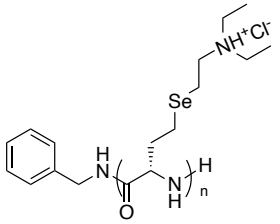
P15

P15 was prepared by following the general PPM procedure with 30 mg $PSeO_2Na$. 2-Bromo-*N,N*-dimethylethanamine hydrobromide (94.0 mg, 0.405 mmol) in 1.5 mL water was added to the reduced polymer solution (4.5 mL). The system became opaque immediately and turned back to clear after an hour. The system was stirred for 5 hours under room temperature. The polymer was dialyzed against 0.5 M NaCl solution for 12 hours followed by water for 48 hours. After lyophilization, the product was recovered as a white powder (27 mg, yield: 73%).
^{1}H NMR (400 MHz, D_2O) δ 4.44 – 4.33 (m, 1H), 3.43 – 3.32 (m, 2H), 2.90 – 2.78 (m, 8H), 2.75 – 2.53 (m, 2H), 2.18 – 1.99 (m, 2H).



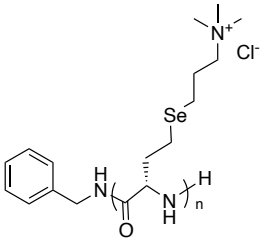
P16

P16 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. Bromocholine bromide (33 mg, 0.135 mmol) in 2 mL water was added to the reduced polymer solution and stirred for 9 hours under room temperature. The polymer was dialyzed against 0.5 M NaCl solution for 12 hours followed by water for 48 hours. After lyophilization, the product was recovered as a white powder (9.0 mg, yield: 68%).
^{1}H NMR (400 MHz, D_2O) δ 4.53 – 4.27 (m, 1H), 3.70 – 3.48 (m, 2H), 3.20 – 3.01 (m, 9H), 2.97 – 2.83 (m, 2H), 2.81 – 2.55 (m, 2H), 2.27 – 2.01 (m, 2H).



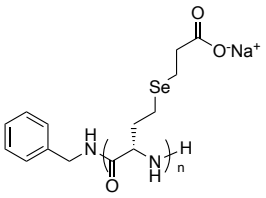
P17

P17 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. 2-Bromo-*N*,*N*-diethylethylamine hydrobromide (35.2 mg, 0.135 mmol) in 0.5 mL water was added to the reduced polymer solution and stirred for 9 hours under room temperature. The polymer was dialyzed against 0.5 M NaCl solution for 12 hours followed by water for 48 hours. After lyophilization, the product was recovered as a white powder (10.1 mg, yield: 77%).
^{1}H NMR (400 MHz, D_2O) δ 4.48 – 4.33 (m, 1H), 3.38 – 3.25 (m, 2H), 3.22 – 3.07 (m, 4H), 2.89 – 2.78 (m, 2H), 2.76 – 2.53 (m, 2H), 2.18 – 1.98 (m, 2H), 1.29 – 1.15 (m, 6H).



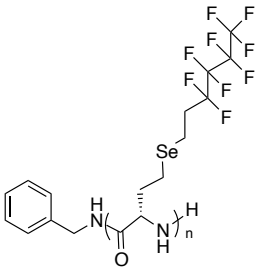
P18

P18 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. 3-Bromo-*N*,*N*,*N*-trimethylpropan-1-aminium bromide (35.3 mg, 0.135 mmol) in 2 mL water was added to the reduced polymer solution and stirred for 9 hours under room temperature. The polymer was dialyzed against 0.5 M NaCl solution for 12 hours followed by water for 48 hours. After lyophilization, the product was recovered as a white powder (11.4 mg, yield: 84%).
^{1}H NMR (400 MHz, D_2O) δ 4.48 – 4.35 (m, 1H), 3.45 – 3.28 (m, 2H), 3.15 – 2.96 (m, 9H), 2.74 – 2.47 (m, 4H), 2.23 – 1.93 (m, 4H).



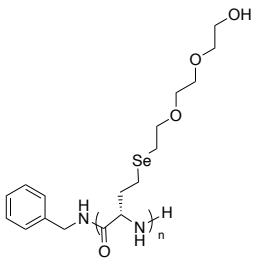
P19

P19 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. 3-Bromopropionic acid (20.5 mg, 0.135 mmol) in 0.5 mL water was added to the reduced polymer solution. The solution immediately became opaque. Then 10 μL 5M NaOH solution was added, which turn the system back to a clear solution The system was stirred for 9 hours under room temperature. The polymer was dialyzed against 0.5 M NaCl solution for 12 hours followed by water for 48 hours. After lyophilization, the product was re-dissolved in 0.01M NaOH solution and desalted by PD-10. After lyophilization, the product was recovered as a white powder (6.2 mg, yield: 53%).
^{1}H NMR (400 MHz, D_2O) δ 4.44 – 4.19 (m, 1H), 2.83 – 2.36 (m, 6H), 2.20 – 1.89 (m, 2H).



P20

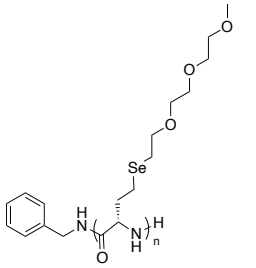
P20 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. 1,1,1,2,2,3,3,4,4-Nonafluoro-6-iodohexane (50.5 mg, 0.135 mmol) in 2.0 mL THF was added to the reduced polymer solution and stirred for 6 hours under room temperature. The polymer was dialyzed against water for 48 hours. After lyophilization, the product was recovered as a light yellow solid (8.0 mg, yield: 42%). Note: loss of the product is probably owing to the sticky nature of the polymer.
^{1}H NMR (400 MHz, TFA-*d*) δ 4.98 – 4.55 (m, 1H), 3.18 – 2.72 (m, 4H), 2.70 – 2.20 (m, 4H).



P21

P21 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. **Hal-21** (46 mg, 0.180 mmol) in 2 mL THF was added to the reduced polymer solution and stirred for 3 hours under room temperature. The polymer was purified by dialyzed against water (MWCO 3500 Da) for 48 hours. After lyophilization, the product was recovered as a light yellow gum (11.0 mg, yield: 80%).

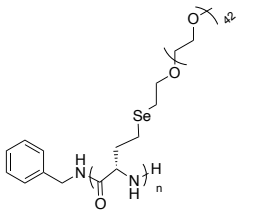
^{1}H NMR (400 MHz, D_2O) δ 4.32 – 4.00 (m, 1H), 3.84 – 3.48 (m, 10H), 2.98 – 2.51 (m, 4H), 2.51 – 1.96 (m, 2H).



P22

P22 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. **Hal-22** (49 mg, 0.180 mmol) in 2 mL THF was added to the reduced polymer solution and stirred for 3 hours under room temperature. The polymer was purified by dialyzed against water (MWCO 3500 Da) for 48 hours. After lyophilization, the product was recovered as a light yellow gum (9.1 mg, yield: 63%).

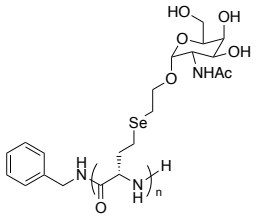
^{1}H NMR (400 MHz, $CDCl_3$) δ 4.17 – 3.95 (m, 1H), 3.81 – 3.50 (m, 10H), 3.46 – 3.32 (m, 3H), 2.91 – 2.55 (m, 4H), 2.44 – 2.10 (m, 2H).



P23

P23 was prepared by following the general PPM procedure with 10 mg $PSeO_2Na$. mPEG1900-I (360 mg, 0.180 mmol) in 2 mL water was added to the reduced polymer solution and stirred for 20 hours under room temperature. The polymer was purified by gel filtration (Superdex^{TM} 75 pg, HiLoad^{TM} 16/600). After lyophilization, the product was recovered as a white powder (61.0 mg, yield: 62%).

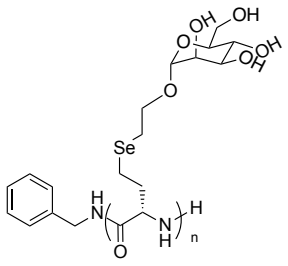
^{1}H NMR (400 MHz, D_2O) δ 3.84 – 3.37 (m, 165H), 3.33 – 3.24 (m, 3H), 2.89 – 2.39 (m, 4H), 2.35 – 1.73 (m, 2H).



P24

P24 was prepared by following the general PPM procedure with 5 mg $PSeO_2Na$. **Hal-24** (23 mg, 0.069 mmol) in 0.5 mL water was added to the reduced polymer solution and stirred for 6 hours under room temperature. The polymer was purified by PD-10. After lyophilization, the product was recovered as a white powder (8.0 mg, yield: 85 %).

^{1}H NMR (400 MHz, D_2O) δ 4.93 – 4.80 (m, 1H), 4.44 – 3.49 (m, 10H), 2.95 – 2.48 (m, 4H), 2.44 – 2.12 (m, 2H), 2.06 – 1.85 (m, 3H).



P25

P25 was prepared by following the general PPM procedure with 5 mg $PSeO_2Na$. 2-bromoethyl-mannose (20 mg, 0.069 mmol) in 0.5 mL water was added to the reduced polymer solution and stirred for 6 hours under room temperature. The polymer was purified by PD-10. After lyophilization, the product was recovered as a white powder (8.4 mg, yield: 100 %).

^{1}H NMR (400 MHz, D_2O) δ 4.90 – 4.76 (m, 1H), 4.44 – 3.26 (m, 9H), 3.12 – 1.91 (m, 6H).

Synthesis and quantitative assessment of RHP through PPM with two different organohalides

$PSeO_2Na$ (20 mg) was reduced with $NaBH_4$ in 2.0 mL water ([selenolate] = 45.9 mM) for subsequent PPM. Stock solutions of organohalides were prepared at 45.9 mM (equal to [selenolate]) for the activated organohalides **Hal-9**, **Hal-11** and **Hal-12**, or 55.1 mM (1.2 fold to [selenolate]) for the inactivated organohalide **Hal-16** and **Hal-22**, respectively. For each reaction, typically, the PSeNa solution (400 μL each) was added to 400 μL cocktail solutions prepared by premixing the stock solutions of two selected organohalides at different volume ratios ($V_{R2\text{-}X}/V_{total}$=12.5 %, 25.0 %, 50.0 %. 75.0 % and 87.5%, respectively). For the modification with only activated organohalides, the system was incubated at room temperature for 4 hours. For the modification with at least one inactivated organohalide, the system was incubated at 50 °C for 6 hours. The product was purified by desalting with PD-10. After lyophilization, the polymers were characterized with ^{1}H NMR spectroscopy.

High throughput synthesis, purification and GPx-like activity assay

The high throughput synthesis was performed with $PSeO_2Na$ (DP = 70) prepared from the same batch. The polypeptide was labeled with FAM-SE before the selenoxide elimination followed the aforementioned procedure. Under the protection of nitrogen, 30.0 mg $PSeO_2Na$ was dissolved in 2.85 mL water and to the solution was added freshly prepared $NaBH_4$ aqueous solution (100 mg/mL × 150 μL). TFA (2.0 μL) was added for 2 times after 20 minutes and 40 minutes and the system was then stirred for 30 minutes (**caution: gas emission**). Twenty minutes before the modification, to the reaction was added water (5.85 mL), NaCl aqueous solution (2.4 M, 1.0 mL) and $NaBH_4$ aqueous solution (100 mg/mL × 150 μL) to give a PSeNa solution ([selenolate] = 13.7 mM) in total 10.0 mL. During the reduction, stock solutions of organohalides prepared in advance were dispensed to 96-well plate by an automated workstation with indicated volume ratios. The total volume of the organohalides solution was 60 μL in each well and 50 μL of the freshly prepared PSeNa solution was added to the wells by pipetting. The plates were then sealed and incubated at 50 °C with gentle agitating (220 rpm) for 5.5 hours. Then to each well was added 2.5 μL 200 mM glutathione (GSH) solution (prepared by dissolving GSH in water with 1 equivalence of NaOH) to consume the unreacted organohalides. After an hour, the plates were centrifuged (4000 g) for 20 minutes and the supernatant was transferred to a desalting plate (MWCO 7000) by pipetting to switch the buffer to PBS and remove small molecular impurities. The resulting RHP solutions were directly used for GPx-like activity assay using an enzyme activity assay.

For the GPx-like activity assay, several stock solutions were prepared as follow:

GR working solution: the suspension of GR was diluted with PBS to make the final concentration at 254 units/mL.
NADPH working solution: NADPH was dissolved in water to make the final concentration at 100 mM.
GSH working solution: GSH was dissolved in water to make the final concentration at 75 mM.
TBHP working solution: 21.5 μL 70% TBHP was diluted with 4.0 mL water. The solution was shaken vigorously for complete mixing.

Assay solution (per plate):

NADPH working solution	100 μL
GSH working solution	200 μL
GR working solution	100 μL
PBS	7600 μL
Total	8000 μL

Sample solution: All polypeptides were dissolved in PBS with indicated concentration. PBS was used as blank group and homopolypeptides prepared from **Hal-21** (1.85 mg/mL in PBS) was used as the control. The blank group was repeated for 3 times on each plate. For the determination of the absolute activity of the RHP. The sample solution of the indicated RHP was prepared in PBS and the molar concentration of selenium was set at 5 mM.

The assay was carried out in a 96-well plate with clear flat bottom. 80 μL assay solution was added followed by the addition of 40 μL sample solution. After mixing, the system was stood at room temperature for 10 minutes and added with 20 μL TBHP working solution to trigger the cascade reaction. The absorbance in 340 nm was recorded every 30 seconds for 50 times under 27 °C. After the assay, 10 μL of assay solution was diluted with 100 μL PBS (with 10%

ethanol) and transferred to a black plate with flat bottom for the record of fluorescence (excitation: 488 nm, emission: 520 nm).

Data analysis of GPx-like activity assay

The GPx-like activity of the RHP is based on the kinetics of NADPH oxidation Change of absorption of a sample at *t* min was calculated as:

$$\Delta A(t)\ [\mathrm{AU}] = A(0) - A(t)$$

where $A(0)$ is the absorption at 0 min and $A(t)$ is the absorption at *t* min.

Linear regression was then performed on the 20 data points of the first 10 minutes:

$$\Delta A(t) = kt + b$$

The activity was calculated from the slope through normalization:

$$E\ [\mathrm{AU}] = \alpha \frac{k_{sample} - \bar{k}_{blank}}{\log_{10}(F_{sample} - \bar{F}_{background})}$$

where $\alpha = 100$ is an scaling parameter, k_{sample} is the slope of the sample, $\bar{k}_{blank}$ is the averaged slope of blank group and positive control, F_{sample} is the fluorescent intensity of the sample and $\bar{F}_{background}$ is the averaged fluorescent intensity of the background. We noticed that molecular composition will influence the fluorescent intensity of the resulting polymer. Meanwhile, no polymer was observed remaining on the desalting plate after purification. Thus logarithmic operation was performed on the fluorescent intensity. In this case, only significant fluctuation of the fluorescent intensity will put an influence on the output signal. To make the data from different plates comparable, the activity of the sample was normalized to the control:

$$E_r = \frac{E_{sample}}{E_{control}}$$

The E_r was then used for the following Bayesian optimization.

Unit definition of the GPx-like activity: One unit will catalyze the oxidation of 1.0 μmole NADPH per min under the presence of GSH and GR, TBHP in PBS at 27 °C.

The absolute catalytic activity normalized to the amount selenium of was calculated as:

$$E_{abs}[\mathrm{U/mmol}] = \frac{(k_{sample} - \bar{k}_{blank})[\mathrm{min^{-1}}]}{\varepsilon^{\mu M}[\mathrm{L\ \mu mol^{-1}\ cm^{-1}}] \times L[\mathrm{cm}]} \div \frac{c[\mathrm{mmol\ L^{-1}}]}{D}$$

$\varepsilon^{\mu M}$: molar extinction coefficient of NADPH, 0.00622 L·μmol^{-1} cm^{-1}
L: optical path length, 0.386 cm
c: concentration of selenium in sample solution, 5 mM
D: dilution factor, 3.5

Random generation of RHPs

The 83 RHPs for random search were generated as a 83 × 7 array of uniformly-sampled values between 0 and 1. Each row (relative abundances) was normalized by its sum to give the composition of RHP in terms of mole fractions. These mole fractions were multiplied by a factor of 60 μL to yield the volume of the organohalide solutions that are required for each well.

Bayesian optimization

The Bayesian optimization (BO) is implemented in Python using GPyTorch, BoTorch and Ax[2] (see the notebook for the source code). The Gaussian process was chosen as the surrogate model owing to its suitability for low-data learning and inherent ability to estimate uncertainty. After each iteration, all data was randomly split into 80/20 training/testing for surrogate model training. The program performs a hyperparameter optimization to select between 4 kernels (RBF, Matern-0.5, Matern-1.5, and Matern-2.5), 10 random seeds and 3 different learning rates (0.01, 0.02 and 0.2) with RMSprop. The model with the lowest lost function (negative marginal log likelihood) on test set was chosen. After training the surrogate model, 83 candidates for successive iteration were chosen by gen() which is implemented in the BoTorch and Ax. The method generates the candidates by optimizing an expected improvement (EI) acquisition function with multi-start optimization (number of starting points: 5, number of samples for initialization: 100) on the consecutive design space, subject to the constraint that total mole fractions equal 1. The program was run on a Lenovo Legion R9000K laptop with AMD Ryzen 9 5900HX CPU and NVDIA GeForce RTX 3080 Laptop GPU (16GB).

Some constraints were set during the optimization. In the second round, the maximum content of x_7 proposed by BO was set to 0.5 with the concern of unwanted precipitation. Then it was realized that excessive positive charge (x_6) will cause precipitation during characterization. Thus for the rest of the screening, the maximum content of x_6 was set to 0.5 while the maximum content of x_7 was set back to 1.0. It should be noted that even through these constraints was not imposed on the random generation, all randomly generated RHPs naturally fell into the space because program is not very likely to generate a vector with element higher than 0.5.

Supplementary Figures

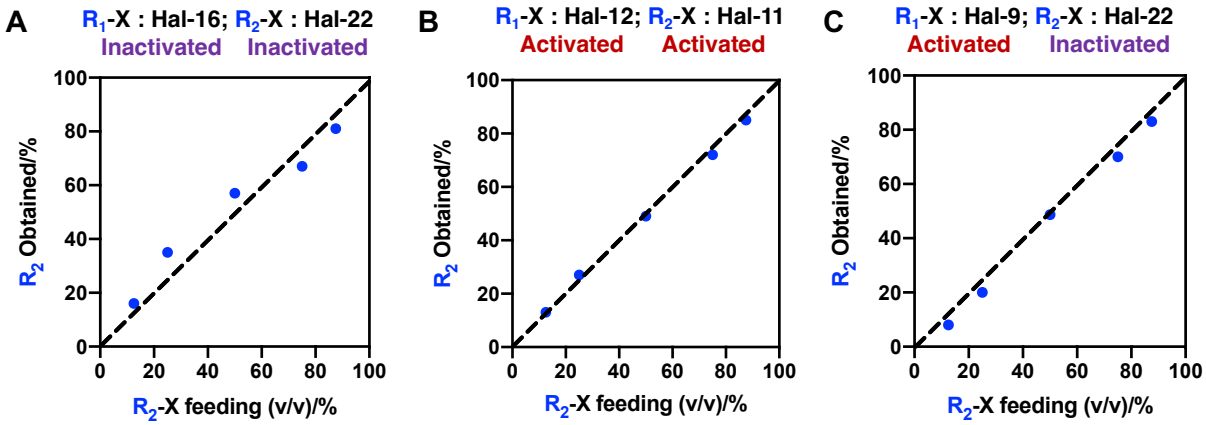


Figure S1: Control of molecular composition of RHP in binary organohalide systems. (A) Modification with two inactivated (**Hal-16** and **Hal-22**) organohalides. Concentration of stock solutions: [**Hal-16**] = [**Hal-22**] = [selenolate] ×1.2; condition: 50 °C for 6 hours. (B) Modification with two activated (**Hal-11** and **Hal-12**) organohalides. Concentration of stock solutions: [**Hal-11**] = [**Hal-12**] = [selenolate]; condition: room temperature for 4 hours. (C) Modification with an activated (**Hal-9**) and an inactivated (**Hal-22**) organohalides. Concentration of stock solutions: [**Hal-9**] = [selenolate], [**Hal-22**] = [selenolate] ×1.2; condition: incubation at 50 °C for 6 hours. The composition of the obtained RHP were determined by ^{1}H NMR spectroscopy after purification.

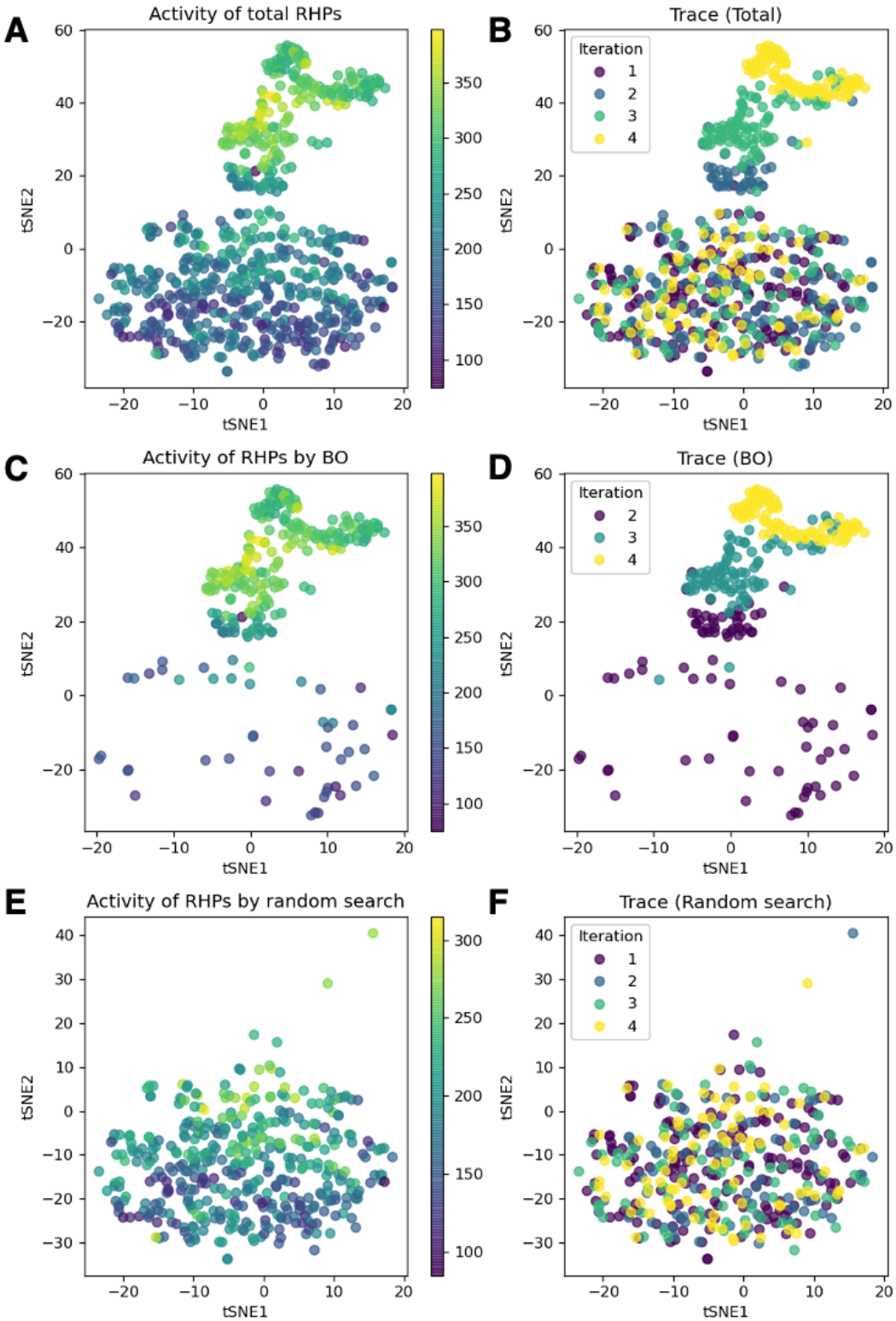


Figure S2: t-SNE projection of polymers selected for synthesis and evaluation, where each point is originally represented by the 7D vector of monomer mole fractions. (A, B) all RHPs acquired, (C, D) RHPs proposed by BO and (E, F) RHPs proposed by random search. Each point is annotated with (A, C, E) its activity (color bar) and (B, D, F) the iteration number during which it was acquired.

Reference

1 Zou, J. *et al.* A Facile Glovebox-Free Strategy To Significantly Accelerate the Syntheses of Well-Defined Polypeptides by N-Carboxyanhydride (NCA) Ring-Opening Polymerizations. *Macromolecules* **46**, 4223-4226 (2013). https://doi.org:10.1021/ma4007939

2 Balandat, M. *et al.* BoTorch: a framework for efficient Monte-Carlo Bayesian optimization. *Adv. Neural Inf. Process. Syst.* **33**, 21524-21538 (2020).