nature synthesis

Article

https://doi.org/10.1038/s44160-023-00294-7

A high-throughput platform for efficient exploration of functional polypeptide chemical space

Received: 3 August 2022

Accepted: 16 March 2023

Published online: 01 May 2023

Check for updates

Guangqi Wu[1,2,3], Haisen Zhou[1,2,3], Jun Zhang[4], Zi-You Tian[1,2,3], Xingyi Liu[1,2,3], Shuo Wang[1,2,3], Connor W. Coley[5] & Hua Lu[1,2,3]

Rapid and in-depth exploration of the chemical space of high-molecular-weight synthetic polypeptides via ring-opening polymerization of *N*-carboxyanhydride allows the discovery of protein mimics and functional biomaterials. The traditional synthetic workflow, however, is labour intensive and has limited throughput. Here we develop an approach for the high-throughput diversification of polypeptides based on a click-like reaction between selenolate and various electrophiles in aqueous solutions. Importantly, the platform is amenable to automation, which allows rapid generation of up to 1,200 homopolypeptides or random heteropolypeptides (RHPs) within one day. With the assistance of machine learning, iterative exploration of the RHP library identifies candidates with improved glutathione peroxidase-like activity from the complex chemical space of which we have little previous knowledge. This automated and high-throughput platform provides potential solutions to unmet challenges, such as the de novo design of artificial enzymes, biomacromolecule delivery and understanding of intrinsically disordered proteins.

Proteins are natural biopolymers with vast chemical space and sophisticated functions, such as binding, catalysis, transportation and signalling. For decades, an overarching goal of polymer science has been to create protein-like functional polymeric materials, not only for fundamental understanding of proteins but also for solving real-world challenges[1–5]. For instance, peptides made by solid-phase peptide synthesis (SPPS) have been widely explored. However, SPPS is generally limited by the small scale and short length of its products. Recent studies have shown that synthetic random heteropolymers with statistically controlled side chain compositions are able to exhibit protein-like functions even without the peptide backbone[6–9]. To this end, synthetic polypeptides prepared by the ring-opening polymerization (ROP) of *N*-carboxyanhydrides (NCAs) have emerged as promising protein mimics with the potential to combine the advantages of both peptides and synthetic polymers[10–15]. Specifically, polypeptides possess the same backbone and even biological functions as proteins and in the meantime can be produced efficiently at up to kilogram scales and a high number-averaged molecular weight (M_n)[16]. One vivid example is Copaxone—a random heteropolypeptide (RHP) made by the ROP of four different amino acid NCA monomers. Owing to its similar composition to myelin basic protein, Copaxone is used as an immunomodulatory drug to treat multiple sclerosis[17]. Nevertheless, similar to other polymers, RHPs are subjected to the curse of dimensionality, meaning the combination of just a few residues can lead to a chemical space that is too large to be fully explored[18,19]. To reach functional protein-mimicking polypeptides from the enormous chemical space, one needs to: (1) facilely produce polypeptides from the design space with high fidelity; and (2) establish

[1]Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing, People's Republic of China. [2]Key Laboratory of Polymer Chemistry and Physics, Ministry of Education, Peking University, Beijing, People's Republic of China. [3]Center for Soft Matter Science and Engineering, Peking University, Beijing, People's Republic of China. [4]Changping Laboratory, Beijing, People's Republic of China. [5]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. e-mail: chemhualu@pku.edu.cn

https://doi.org/10.1038/s44160-023-00294-7

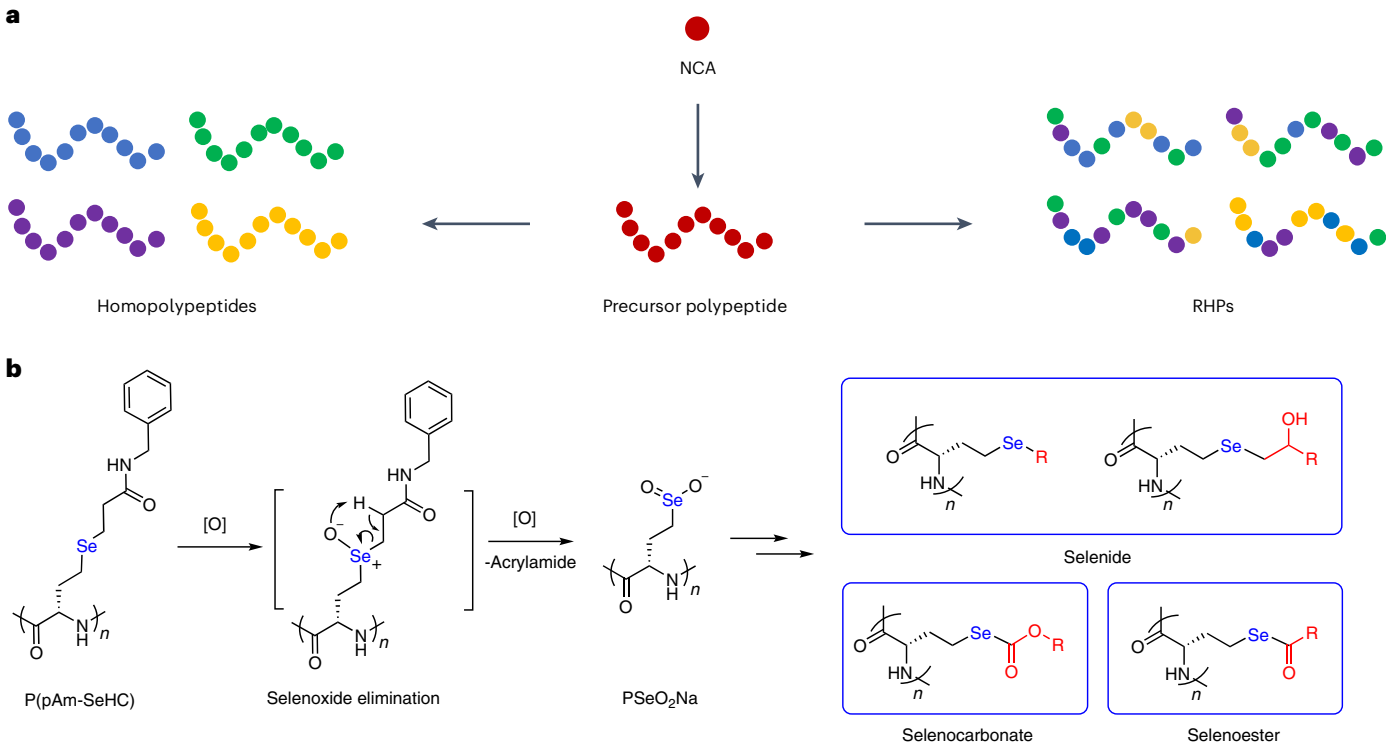


Fig. 1 | PPM of polypeptides through highly efficient selenium chemistry. a, Scheme of the PPM strategy for making homopolypeptides (left) and RHPs (right). **b**, The selenopolypeptide $PSeO_2Na$, derived from the selenoxide elimination of P(pAm-SeHC), is utilized to generate reactive selenium species for PPM.

an efficient strategy for effective exploration of the space at affordable labour and time costs.

The development of automated and high-throughput synthesis (HTS) has greatly facilitated small-molecule drug discovery and biomacromolecule evolution[20–24]. Furthermore, recent applications of machine learning in chemistry have further accelerated the function mining processes and the interpretation of the experimental results[25–31]. Although the employment of HTS and/or machine learning for making functional polymers can date back to the early 1990s[32–36], progress in this field has largely lagged behind compared with progress in small-molecule and biomacromolecule synthesis[19,37]. Currently, HTS of polymers is primarily accomplished by parallel step-growth polymerizations or chain-growth radical polymerizations[32,36,38]. Langer et al.[39,40] applied Michael addition to synthesize poly(β-amino ester)s for non-viral gene delivery. Boyer et al. used photo-induced electron transfer–reversible addition–fragmentation chain transfer for the HTS of poly(acryl amide)s to explore polymers with protein binding[41] and antimicrobial activity[42]. More recently, Leibfarth et al. and Gormley et al. applied controlled radical polymerization in combination with automated synthesis and machine learning to identify polymers for enhanced magnetic resonance signals[19] and protein preservation[43,44], respectively. As a complementary approach to the aforementioned parallel polymerization, post-polymerization modification (PPM)[45,46] is also common for library generation[47]. Compared with the parallel polymerization approach, the PPM strategy is advantageous in that all data can be generated from one shared precursor polymer synthesized from a single batch. For this, highly efficient Huisgen cycloaddition[48], activated ester–amine conjugation[49,50] and thiol-ene reactions[51] are among the most frequently employed reactions.

To date, attempts to incorporate the NCA and polypeptide chemistry into HTS workflows have been challenging and sparse. This is partially attributable to the high sensitivity to moisture of the ROP system, the poor aqueous solubility of polypeptides and laborious workup processes. In a pioneering work, Deming et al.[52] applied the parallel copolymerization strategy to generate RHPs. However, the synthesis procedure was time consuming, with ~500 RHPs produced in 2 weeks. PPM has also been utilized for the synthesis of both homopolypeptides and RHPs (Fig. 1a)[53–59], but again most of the early seminal works were performed in a low-throughput fashion. Moreover, many click reactions often introduce bulky and hydrophobic spacer moieties (for example, triazole) that might affect both the secondary structure and solubility of the resulting polypeptides[60,61].

One underlying challenge for machine learning-assisted polymer design is the limited availability of high-quality data[37]. While most studies exploit data from the literature and virtual experiments (for example, electronic structure calculations or simulations), an ideal platform is capable of performing new experiments to support model training. We envision that a powerful PPM chemistry would not only simplify the HTS procedure, but is also beneficial for the generation of standard data. In this Article, we report the development of a HTS platform in aqueous solutions for polypeptides based on a click-like reaction between selenolate and electrophiles. This quantitative chemistry gave accurate control of the molecular composition of RHPs and was amenable to automated synthesis, which allowed efficient generation and purification of over 1,200 polymers within one day. Compared with other click-type reactions, this chemistry only introduced a selenium atom as a miniature yet reactive linker, which poses minimum influence on the overall polymer structure and offers desirable responsiveness and functions to the materials. With the assistance of machine learning model-guided optimization, we were able to perform iterative exploration of the RHP chemical space for enzyme mimics and identified candidates with improved glutathione peroxidase (GPx)-like activity in a more efficient and effective way.

Results and discussion

Design and synthesis of the precursor for PPM

Inspired by studies on selenocysteine-based bioconjugation[62–67], we sought to introduce a selenolate—arguably one of the most nucleophilic

https://doi.org/10.1038/s44160-023-00294-7

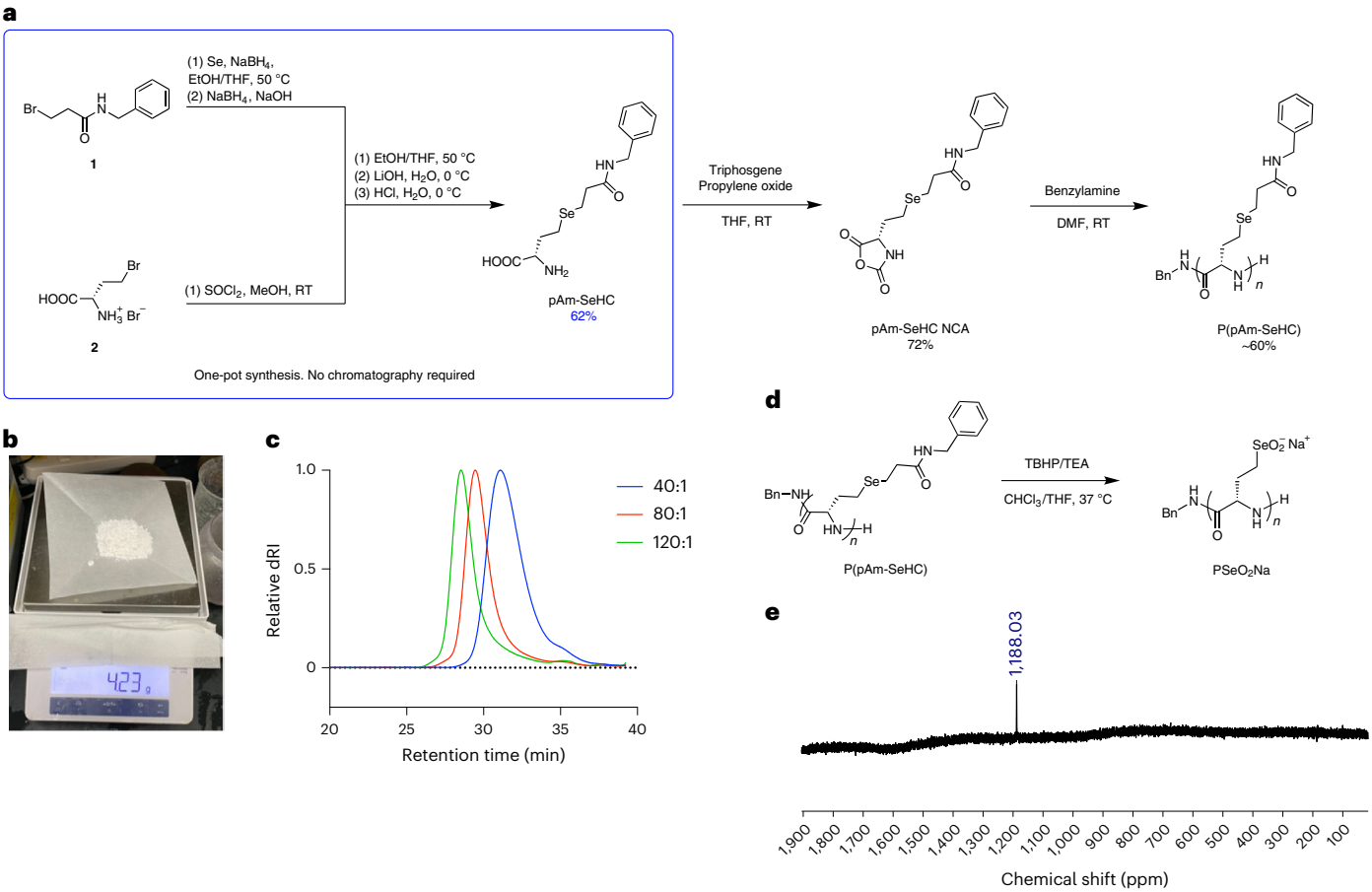


Fig. 2 | Synthesis of the precursor polypeptide $PSeO_2Na$. a, Scheme for the synthesis of P(pAm-SeHC). Bn, benzyl; DMF, *N*, *N*-dimethylformamide; RT, room temperature. **b**, Photograph of the amino acid pAm-SeHC at gram scale. **c**, SEC of P(pAm-SeHC) prepared at various [M]/[I] ratios. dRI, differential refractive index. **d**, Oxidative elimination of P(pAm-SeHC) in a biphasic system gave the precursor polypeptide $PSeO_2Na$. TEA, triethylamine. **e**, ^{77}Se NMR spectrum of $PSeO_2Na$.

functional groups in aqueous solutions—to the side chain for the PPM of polypeptides (Fig. 1b). Among various selenol-protecting strategies[68], we chose to generate the reactive selenium species after polymerization through selenoxide elimination (Fig. 1b), owing to its efficiency and mildness. This type of reaction was previously documented for the introduction of a double bond[69]. However, the resulting selenium species were normally considered as byproducts and discarded[70–73]. Notably, Chen et al.[74–76] reported that if a selenium and an amide carbonyl were placed at the δ and ζ positions of the side chain of amino acids, respectively, the elimination would exclusively take place at the Se_δ–C_ε bond instead of the C_β–Se_δ bond (Fig. 1b). Based on this discovery, we designed a selenopolypeptide, P(pAm-SeHC), whose pendant group is a latent selenolate. The seleno-amino acid (pAm-SeHC) was synthesized from L-methionine and elemental selenium in a one-pot manner with no need for chromatography, affording the pure product at multi-gram scale (Fig. 2a,b and Supplementary Figs. 1 and 2). The monomer, namely pAm-SeHC NCA, was then obtained using a moisture-tolerant method recently developed by our group (Supplementary Figs. 3 and 4)[77]. Benzyl amine-initiated ROP of pAm-SeHC NCA at various monomer-to-initiator concentration ([M]/[I]) ratios gave complete monomer conversion, generating the desired selenopolypeptide P(pAm-SeHC) with a controlled M_n of up to 35.8×10^3 g mol^{-1} and a dispersity of ~1.10 (Fig. 2c, Supplementary Fig. 5 and Supplementary Table 1). It is also worth mentioning that the whole process did not involve any volatile selenium species and thus avoided undesired stench, which has been a common safety concern for organoselenium compounds.

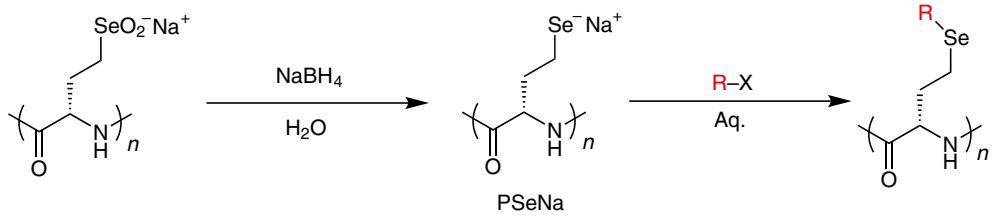
Next, we performed the selenoxide elimination of P(pAm-SeHC) with *tert*-butyl hydroperoxide (TBHP). However, initial attempts all failed in most common organic solvents, possibly owing to the poor organic solubility of both the oxidized intermediate and eliminated product. This hurdle was eventually overcome using a one-pot, two-step process in a biphasic system (Fig. 2d)[73]. Briefly, P(pAm-SeHC) was first oxidized in a mixed tetrahydrofuran (THF)/chloroform solution using TBHP, followed by the addition of sodium bicarbonate solution and trimethylamine, affording exclusively the expected $PSeO_2Na$ in the aqueous phase. The generation of $PSeO_2Na$ was confirmed by both ^{1}H and ^{77}Se NMR spectroscopy with an overall yield of ~60% calculated from NCA (Fig. 2e and Supplementary Fig. 6). Remarkably, $PSeO_2Na$ was stable under an ambient atmosphere for at least 3 months at −20 °C, making it an ideal intermediate for scaled-up synthesis and long-term storage. The block copolymer with poly(ethylene glycol) (PEG) (that is, PEG-*b*-$PSeO_2Na$) was also prepared from PEG-*b*-P(pAm-SeHC) with a modest yield (Supplementary Figs. 7 and 8).

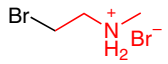
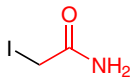
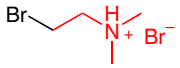
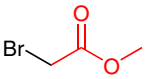
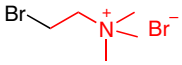
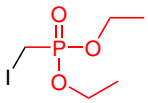
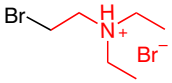
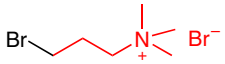
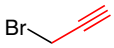
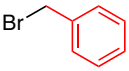
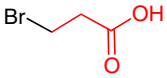
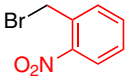
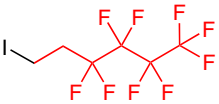
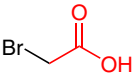
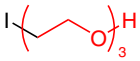
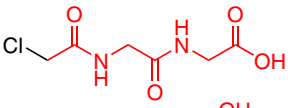
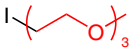
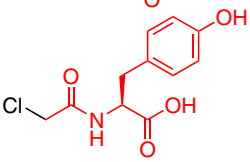
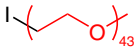
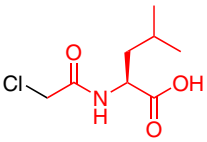
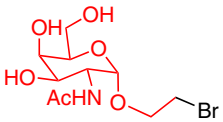
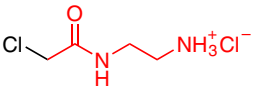
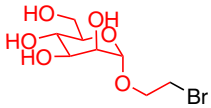
Synthesis of homopolypeptides in solution and solid phase

Next, $PSeO_2Na$ was reduced with $NaBH_4$ in water, affording the selenolate-bearing polypeptide (PSeNa) for further functionalization (Supplementary Fig. 9). This in situ-generated PSeNa was soluble in aqueous solutions and highly reactive for substitution reactions with various activated (**Hal-1**–**13** in Table 1) or inactivated (**Hal-14**–**25** in Table 1) electrophiles. The PPM exhibited almost quantitative side chain conversion, as indicated by ^{1}H NMR spectroscopy (Table 1, Fig. 3a and Supplementary Figs. 10–35) and was equally efficient with regard to PSeNa

https://doi.org/10.1038/s44160-023-00294-7

Table 1 | PPM of $PSeO_2Na$ with activated and inactivated halides



Activated R–X		Product (yield[a], GE[b])	Inactivated R–X		Product (yield[a], GE[b])
Hal-1		**P1** (79%, >95%)	**Hal-14**		**P14** (60%, >95%)
Hal-2		**P2** (73%, >95%)	**Hal-15**		**P15** (73%, >95%)
Hal-3		**P3** (62%, >95%)	**Hal-16**		**P16** (68%, >95%)
Hal-4		**P4** (48%, >95%)	**Hal-17**		**P17** (75%, >95%)
Hal-5		**P5** (99%, >95%)[c]	**Hal-18**		**P18** (84%, >95%)
Hal-6		**P6** (97%, >95%)[c]			
Hal-7		**P7** (75%, >95%)[c]	**Hal-19**		**P19** (53%, >95%)
Hal-8		**P8** (78%, >95%)	**Hal-20**		**P20** (42%, >95%)
Hal-9		**P9** (89%, >95%)	**Hal-21**		**P21** (81%, >95%)
Hal-10		**P10** (78%, >95%)	**Hal-22**		**P22** (63%, >95%)
Hal-11		**P11** (86%, >95%)	**Hal-23**		**P23** (62%, >95%)
Hal-12		**P12** (87%, >95%)	**Hal-24**		**P24** (85%, >95%)
Hal-13		**P13** (72%, >95%)	**Hal-25**		**P25** (100%, >95%)

[a]Purification yield. [b]Grafting efficiency based on the ^{1}H NMR spectra of the product. [c]Prepared from PEG-*b*-$PSeO_2Na$. Aq., aqueous solution.

with various degrees of polymerization (degree of polymerization from 45–130). The size exclusion chromatography (SEC) chromatogram of the modified polymer remained a unimodal peak after the modification, suggesting negligible backbone degradation, chain scission or crosslink during the process (Fig. 3b and Supplementary Fig. 36). Some of the obtained polypeptides (for example, **P21**, **P23** and **P24**; Fig. 3c and Supplementary Fig. 37) exhibited typical right-handed α-helical conformation in water, implying minimum racemization of the backbone under the PPM conditions. The preservation of the chirality could offer an opportunity for facile conformational regulation via the selection of starting amino acids with different configurations (for example, D- or DL-methionine). Side chain oxidation-induced helix-to-coil transition

https://doi.org/10.1038/s44160-023-00294-7

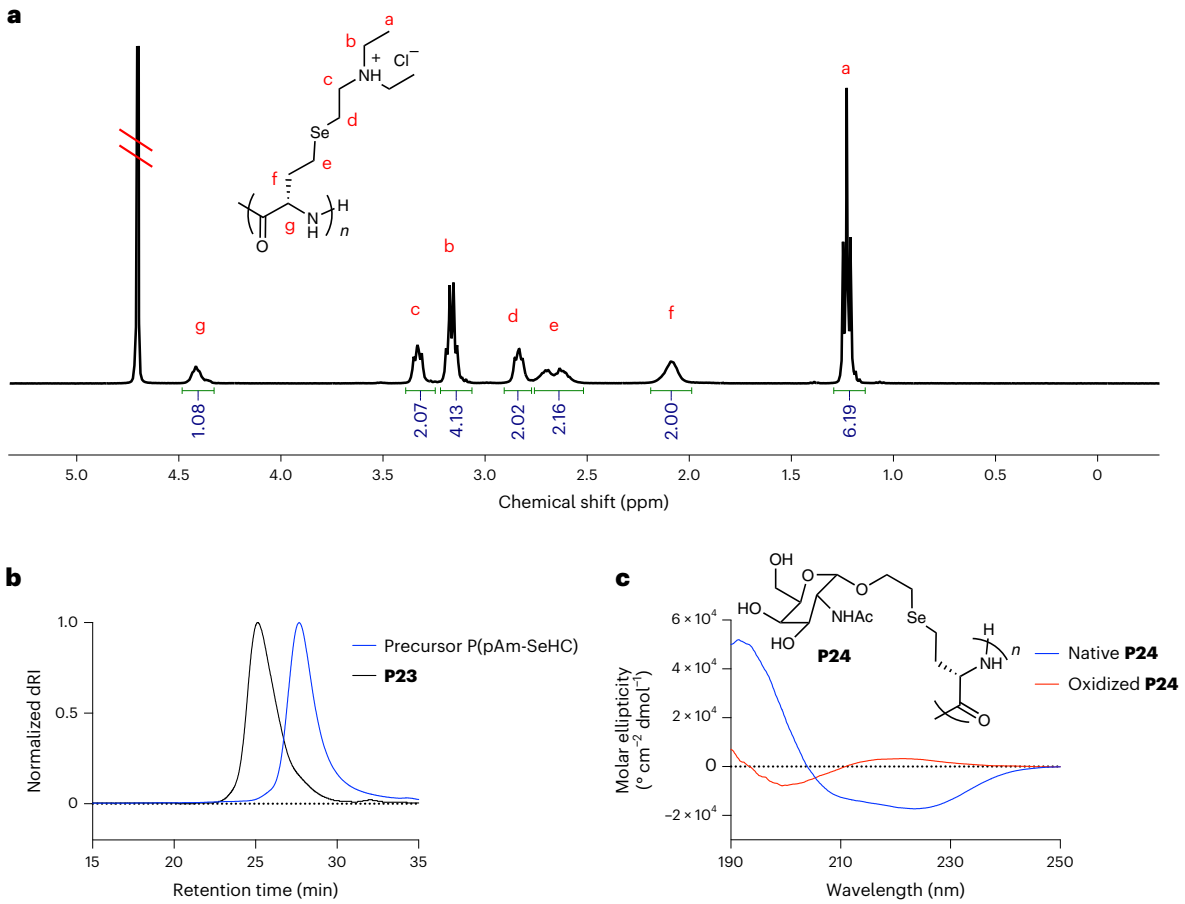


Fig. 3 | Representative characterizations of homopolypeptides prepared by PPM of PSeO$_2$Na. a, ^{1}H NMR spectrum of **P17** in D$_2$O. **b**, Overlay of the SEC traces of **P23** and its precursor P(pAm-SeHC). **c**, Circular dichroism spectra of the GalNAc-grafted selenopolypeptide (**P24**) and its oxidized form.

(Fig. 3c) was also observed, which was similar to the oligo(ethylene glycol)-grafted selenopolypeptides[78].

The PPM of PSeNa showed remarkable tolerance to various functionalities, including those that are difficult to introduce directly through the ROP of NCA. For instance, poly(selenomethionine) (**P1**) and the hydrofluorocarbon-bearing polymer (**P20**) were smoothly synthesized despite their poor solubility in common solvents. Reactive functional groups such as alkenes and alkynes were also facilely grafted to the side chain (**P5** and **P6**)[79]. The photo-labile *o*-nitro benzyl group was efficiently incorporated (**P8**) for potential light-responsive materials[80]. We also generated various anionic and cationic polyelectrolytes bearing diverse functionalities, such as carboxylic acids, primary, secondary and tertiary amines and quaternary ammonium (**P9**–**P19**). The applicability of the strategy was further demonstrated by the preparation of densely packed brush-like polymer (**P23**), which has long been challenging owing to steric hindrance[57]. The synthesis of glycopolypeptides used to be laborious; here, polypeptides modified with GalNAc (**P24**) and mannose (**P25**) were readily achieved in one step, holding great promise for biomedical applications, including lysosome-targeting chimeras[81,82] and immunotherapy[83]. Disappointingly, modification of PSeNa with inactivated secondary organohalides or primary organochlorides gave low grafting efficiency, probably due to their low reactivity. Of note, some of the synthesized selenopolypeptides had extremely poor solubility even in trifluoroacetic acid (TFA). To circumvent the problem, these polymers were prepared using the more soluble block copolymer PEG-*b*-PSeO$_2$Na as the precursor of PPM (**P5**, **P6** and **P7**).

Other electrophilic substrates besides organohalides were also applied for PSeNa modification (Table 2). For example, modification of PSeNa with epoxides smoothly generated polypeptides with β-hydroxyl selenide (**P26** and **P27**; Supplementary Figs. 38 and 39). We also successfully fabricated two polypeptides bearing selenoester (**P28** and **P29**; Supplementary Figs. 40 and 41)—a functional group incompatible with the ROP of NCA due to its vulnerability to nucleophiles[84,85]. Similarly, a polypeptide tethering selenocarbonate was prepared from chloroformate (**P30**; Supplementary Fig. 42). These polymers were all obtained with more than 80% separation yields and almost quantitative grafting efficiency.

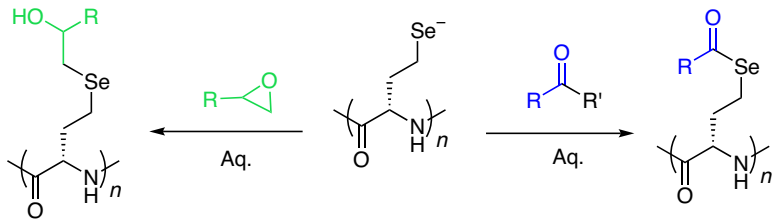
The ROP, oxidative elimination and subsequent PPM processes were also attempted on an amine-functionalized resin (Supplementary Fig. 43), to examine the feasibility of integrating this chemistry with the well-established SPPS. It was found that the product remained controlled on the surface of the resin, affording P(pAm-SeHC) with high M_n and a unimodal SEC peak after cleavage (Supplementary Fig. 43). Moreover, the oxidative elimination and subsequent PPM with **Hal-13** were also performed on the resin and found to be highly effective. After cleavage, pure product **P13** was conveniently obtained with a quantitative grafting efficiency (Supplementary Fig. 44).

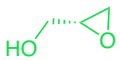
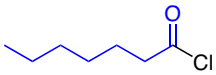
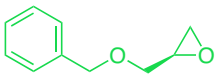
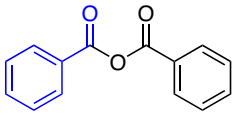
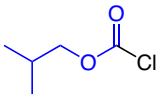
Synthesis of RHPs in solution

With the success in making homopolypeptides, we next exploited the chemistry for RHP synthesis by simultaneously treating the precursor with multiple organohalides. To facilitate future HTS and machine learning algorithm development, we sought to predictively control the molecular composition of the RHP through machine-readable input[37,86]. For this, we chose the feeding volume ratio of the organohalides, as it could be directly transformed to the command for automated

https://doi.org/10.1038/s44160-023-00294-7

Table 2 | PPM of $PSeO_2Na$ with epoxide, acyl chloride, anhydride and chloroformate



Epoxide		Product (yield[a], GE[b])	Acylation reagent		Product (yield[a], GE[b])
EP-1		**P26** (46%, >95%)	**AC-1**		**P28** (88%, >95%)[c]
EP-2		**P27** (79%, >95%)	**AC-2**		**P29** (83%, >95%)[c]
			AC-3		**P30** (81%, >95%)[c]

[a]Purification yield. [b]Grafting efficiency based on the ^{1}H NMR spectra of the product. [c]Prepared from PEG-*b*-$PSeO_2Na$.

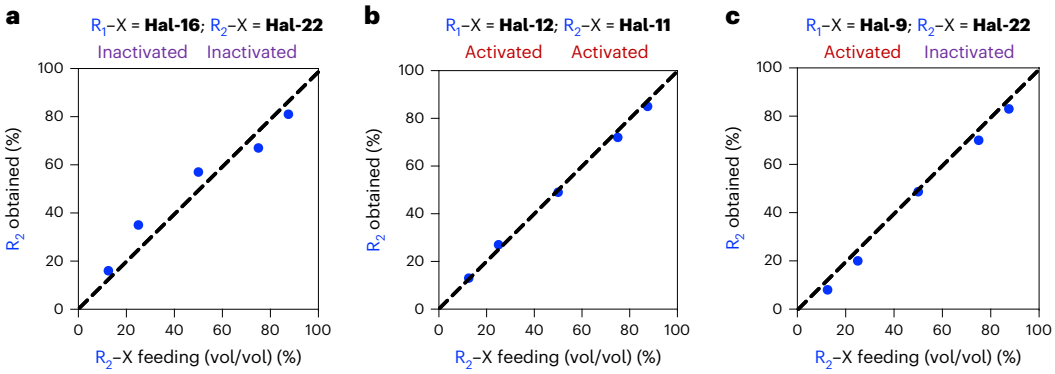


Fig. 4 | Control of the molecular composition of RHPs in binary organohalide systems. a, Modification with two inactivated organohalides (**Hal-16** and **Hal-22**). Concentrations of stock solutions: [**Hal-16**] = [**Hal-22**] = [selenolate] × 1.2. Conditions: 50 °C for 6 h. **b**, Modification with two activated organohalides (**Hal-11** and **Hal-12**). Concentrations of stock solutions: [**Hal-11**] = [**Hal-12**] = [selenolate]. Conditions: room temperature for 4 h. **c**, Modification with an activated organohalide (**Hal-9**) and an inactivated organohalide (**Hal-22**). Concentrations of stock solutions: [**Hal-9**] = [selenolate] and [**Hal-22**] = [selenolate] × 1.2. Conditions: incubation at 50 °C for 6 h. The compositions of the obtained RHPs were determined by ^{1}H NMR spectroscopy after purification.

liquid-handling workstations. To achieve this goal, the conditions (such as stock solution concentration, stoichiometry and temperature) for the RHP synthesis were carefully optimized on three binary systems: (1) both organohalides were inactivated (**Hal-16** and **Hal-22**); (2) both organohalides were activated (**Hal-11** and **Hal-12**); and (3) one organohalide was activated and the other was inactivated (**Hal-9** and **Hal-22**, respectively). After the trial-and-error attempts and considering the difference in relative reactivity, we fixed the concentrations of the stock solutions at 1.2 and 1.0 equivalent to those of the selenolate for the inactivated and activated organohalides, respectively. Meanwhile, the total volume of organohalides was set to be equal to the volume of the PSeNa solution. With these optimized conditions, a good match between the input volume ratio and the actual molecular composition of the RHP was obtained for all three scenarios (Fig. 4 and Supplementary Figs. 45–47). Thanks to the high reactivity of the selenolate, it was unnecessary (and sometimes even deteriorative) to use a large excess of the organohalides in the RHP synthesis (Supplementary Figs. 48 and 49). Overall, this capability to precisely control the composition of RHP laid a firm foundation for generating high-quality datasets for machine learning and paved the way for a quantitative structure–activity relationship (SAR).

Exploration of functional RHP with automation and machine learning

We then explored the feasibility of transferring the synthesis from flasks (5.0 mg ml^{-1} **$PSeO_2Na$**; 1.0–4.0 ml) to multi-well plates (1.0 mg ml^{-1} **$PSeO_2Na$**; 100–200 µl). In model studies within an NMR tube, the PPM was found to have completed after 6 h at 50 °C or 8 h at 37 °C when an inactivated organohalide **Hal-16** was used (Supplementary Fig. 50). For the activated organohalide **Hal-9**, the completed modification required

https://doi.org/10.1038/s44160-023-00294-7

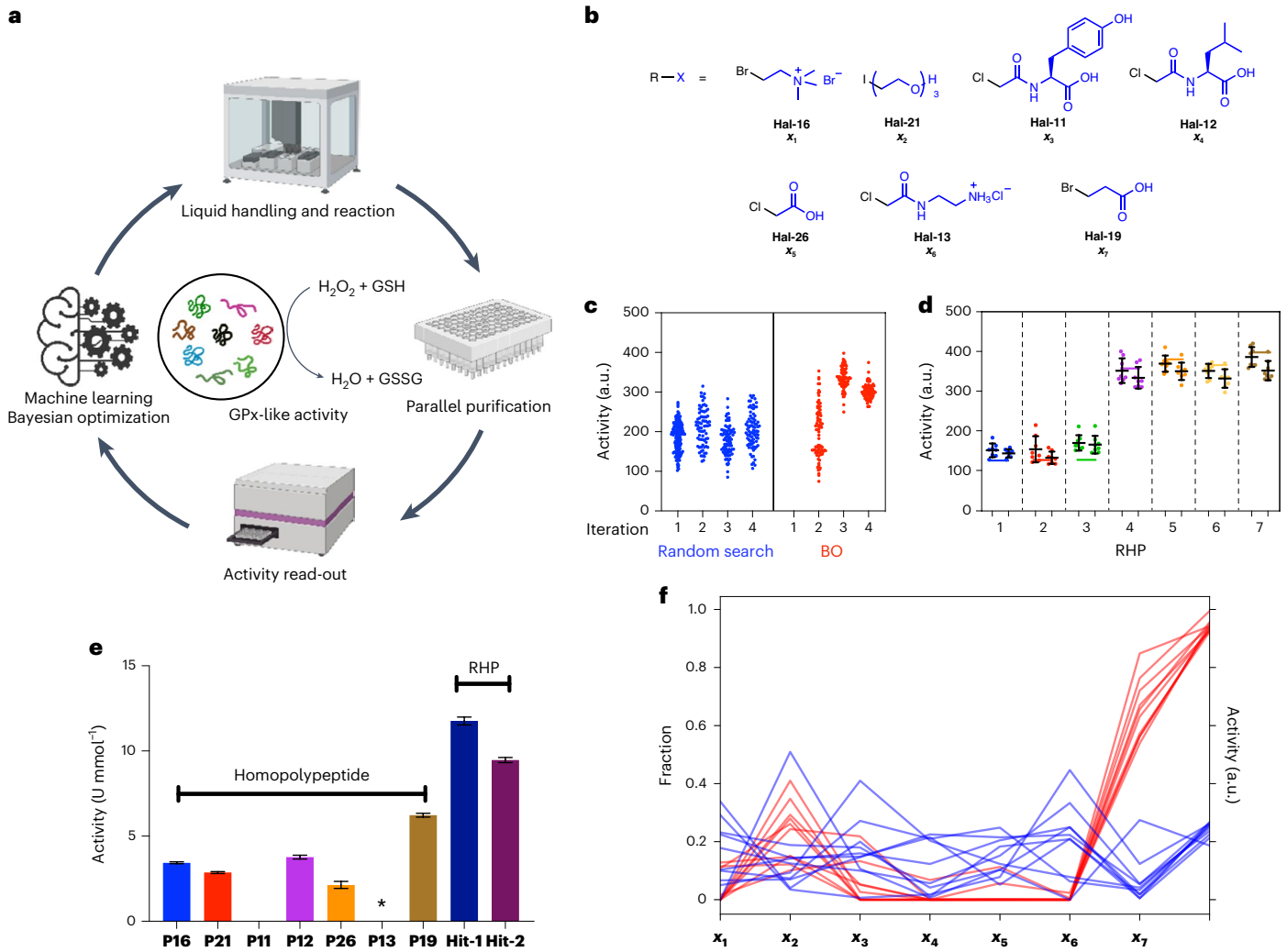


Fig. 5 | Closed-loop optimization of GPx activity of the RHPs via HTS and machine learning. a, Illustration of the closed-loop workflow containing four modules, including HTS, parallel purification, activity readout and Bayesian optimization. GSH, reduced glutathione; GSSG, oxidized glutathione. Created with BioRender.com. **b**, Structures of the seven selected organohalides for RHP library generation and aim of optimization. The molecular composition of RHPs is descripted as a seven-dimensional vector, $\mathbf{x} = (x_1, \ldots, x_7)$, where x_n ($n = 1$–7) is the relative volume ratio of the organohalide and $c_n > 0$ ($n = 1$–7) is the upper limit of x_n (see Supplementary Information for the detailed design space). The aim of optimization is to find a molecular composition $\mathbf{x}^*$ that maximizes the catalytic activity, subject to the constraint that the sum of x_n ($n = 1$–7) is equal to 1. **c**, GPx-like activity of RHPs in each iteration via random searching (blue) or Bayesian optimization (BO; red). **d**, Data validation within a plate ($n = 8$) and between two different plates. RHPs with low (lanes 1–3) and high (lanes 4–7) GPx-like activities from the database were selected for validation. The dots on the right and left side in each lane represent the results from different plates. The black central lines and error bars in each lane represent the mean and s.d. The coloured line in each lane is the original activity of the RHP from the database. **e**, Comparison of the GPx-like activities of the two RHP hits with the seven homopolypeptides each modified with one individual organohalide used in HTS ($n = 3$). **Hit-1**: $\mathbf{x} = (0.12, 0.12, 0, 0, 0, 0, 0.76)$ and **Hit-2**: $\mathbf{x} = (0, 0.24, 0.22, 0, 0, 0, 0.54)$. All polymers were synthesized in a flask and then purified for GPx activity. The data are presented as means ± s.d. Note that the activity of the homopolypeptide **P13** could not be measured properly owing to precipitation during testing (data replaced with an asterisk). **f**, Parallel coordinate plot describing the copolymer composition and performance of the best ten (red) and worst ten (blue) performing RHPs.

only 1 h at 50 °C (Supplementary Fig. 51). Based on this finding, the HTS of RHP was established with the assistance of a commercialized automated workstation for dispensing stock solutions of organohalides to plates (Fig. 5a). Then, the freshly reduced PSeNa solution was added through pipetting and the plate was sealed and incubated in an oven at 50 °C for 6–8 h. The resulting polymers were purified in parallel using a desalting plate[87]. This semi-automated workflow greatly boosted the synthesis capability and enabled the parallel preparation of ~400 RHPs (four plates) in one day. The throughput could be improved easily to ~1,200 RHPs (12 plates) per day if only activated organohalides were used for modification. Of note, precipitations were observed in our preliminary trials when: (1) the content of relatively hydrophobic modifiers was high (for example, >0.5); (2) the contents of negatively and positively charged modifiers were roughly equal to each other, which neutralized the net charge and led to insoluble polyplex; and (3) the conversion of the side chain was incomplete and the residual selenolate was gradually oxidized into diselenide and formed a crosslinked network. To avoid such undesired precipitation, several strategies were found to be effective: (1) adding a certain amount of organic solvents, such as *N*,*N*-dimethylformamide or dimethyl sulfoxide when a hydrophobic modifier was used with high ratios; (2) adjusting the pH of the reaction system to avoid the isoelectric point of the RHP; or (3) attaching PEG to the precursor polymers.

Next, the HTS system was coupled to functional analysis for protein-like activity. Many organoselenium compounds show activities similar to GPx enzymes[88,89]—a class of proteins that catalyse the

reduction of peroxide by glutathione. Since GPx plays important roles in retaining cellular redox homoeostasis[90], the development of GPx mimics may lead to antioxidative therapeutics for the treatment of stroke, reperfusion injury and neurodegenerative diseases[91–95]. We previously synthesized two selenopolypeptides P(EG$_x$-SeHC) (x = 3 or 4), but they only exhibited weak GPx-like activities[78]. Here we sought to improve the GPx-like activity of the RHP with the platform and provide information on SARs for future works.

Seven organohalides with good water solubility were selected for PPM (Fig. 5b). These organohalides offer different structural features, including charges, hydrogen bond donors and acceptors, neutral polar groups, alky groups and aromatic groups. Mole fractions of the components were combined as a vector representation of each RHP. A fluorescein was tethered to the amino terminus of the precursor polymer (degree of polymerization = 70) as an internal label to quantify the polymer content in each well. The GPx-like activity of each RHP was determined by normalizing the absolute readout from a modified NADPH-coupled assay[96] to the fluorescence intensity of each well (see the Supplementary Information section 'Data analysis of GPx-like activity assay' for details). Based on previous optimization studies (Fig. 4 and Supplementary Figs. 50 and 51), we built a workflow that executes the HTS, parallel purification and activity analysis of the RHPs within 10 h (83 experiments per plate with two plates in parallel; Supplementary Fig. 52). However, even with this throughput, the space of hypothetical RHPs was too large to be fully explored. This situation motivates the use of a model-guided optimization strategy to help to prioritize experiments and accelerate discovery. Bayesian optimization is a powerful tool for various design problems[97] and is receiving increasing attention in the chemistry community[44,98–100]. Because of the data efficiency of Bayesian optimization relative to brute force or random screening, it is especially useful for problems where evaluation is expensive. We applied a Bayesian optimization framework based on BoTorch and Ax[101] and established a closed-loop design–build–test–learn workflow (Fig. 5a).

Within 4 d, four iterations comprising a total of ~660 experiments were performed (Fig. 5c), with 166 experiments per iteration. Initially, 166 RHPs were randomly chosen and analysed from the designed space to train a Gaussian process regression model. Candidates for successive iterations were chosen by selecting compositions that optimized an expected improvement acquisition function, subject to the constraint that total mole fractions equalled 1. To avoid trapping into local minimums, random search and Bayesian optimization were performed simultaneously in each round. All of the data, including those from previous rounds, were used to retrain the surrogate Gaussian process before proposing 83 candidates for the next iteration. It was found that while random search consistently found candidates with activities near a range of 150–200, Bayesian optimization efficiently found RHPs with substantially higher GPx-like activity, particularly in the third and fourth iteration (Fig. 5c). T-distributed stochastic neighbour embeddings showed that Bayesian optimization quickly identified an area in the design space that achieved higher activity (Supplementary Fig. 53). Replicate experiments of seven randomly selected RHPs from the database were carried out in different plates and the results validated the reproducibility of the data (Fig. 5d and Supplementary Table 2).

To further validate the results, two hits from Bayesian optimization were synthesized in a flask (Supplementary Figs. 54 and 55) and their GPx-like activity was evaluated. **Hit-1** exhibited around twofold higher GPx-like activity than the homopolypeptide modified with **Hal-19**, the most active homopolypeptide in the design space (Fig. 5e), which meant that the activity of RHPs is not merely the normalized average of the activity of each component. Although the freshly prepared native GPx was substantially more reactive than **Hit-1**, the catalytic activity of native GPx was found to slump to zero during incubation at 37 °C for 12 h; in contrast, **Hit-1** fully retained its activity during incubation under the same conditions (Supplementary Fig. 56). This stability underscores one of the many advantages of RHPs in biological and biomedical applications. Analysis of the high-performing RHPs pointed out a correlation with high x_7 (Fig. 5f and Supplementary Tables 3 and 4), which was supported by the retrospective analysis of the full dataset with a linear regression surrogate model (Supplementary Fig. 57). Interestingly, while the linear model suggested maximized x_7 to be advantageous for activity, the homopolymer with x_7 = 1 was less active than **Hit-1** or **Hit-2**. Meanwhile, x_2 was also positively correlated with GPx activity. Altogether, it is likely that the functional groups introduced to the RHPs by **Hal-16**, **Hal-21** and **Hal-11** created a suitable microenvironment favouring the desired chemical transformations, which resembled the catalytic pocket of enzymes to some extent[88].

While the detailed SAR is currently under investigation, the above results illustrate that the application of a machine learning model or other optimization algorithms can facilitate materials discovery.

Conclusion

In summary, we report a robust, quantitative and divergent strategy for the rapid expansion of a polypeptide library based on a universal precursor selenopolypeptide. After controlled ROP and a regioselective elimination reaction, the in situ-generated selenolate on the side chain was readily modified with a school of electrophiles, creating homopolypeptides and RHPs with broad chemical diversity. This PPM strategy avoided the laborious efforts of making a variety of NCAs that are synthetically challenging. Compared with many other click reactions, the selenolate is miniature in size and does not create a bulky linkage moiety after reaction. Moreover, these polymers can be used to design materials with interesting properties and functions by harnessing the unique chemistry and properties of selenium[63,102–105].

The potential of this modification chemistry was further highlighted by the establishment of HTS and machine learning model-guided optimization of functional RHP. Enabled by the efficiency of the reaction, a map from the ratio of the feeding volume to the molecular composition was directly created. Because all polypeptides were derivatized from the same precursor, this strategy could be particularly useful in generating a standardized dataset. Moreover, the HTS was performed in aqueous solutions and open air, which allowed convenient transfer of the resulting polymers to subsequent biological assays. As a proof of concept, we demonstrated a concise workflow enabling the rapid identification of RHPs with promising GPx-like activity. The identified RHP exhibited GPx-like activity around twofold higher than the most active homopolypeptide. While the detailed SAR is still under investigation, these results underscore the power of HTS and machine learning in exploring systems of which people have little knowledge. Of note, under such synthesis capability, parallel characterization of the material properties may become the rate-limiting step in the closed-loop optimization, which could be accelerated by further implementation of automation to build a self-driving laboratory. Furthermore, by performing the reported ROP and modification chemistry on an automated solid-phase peptide synthesizer, the level of automation could be further augmented. Our preliminary results on solid-phase ROP and modification (Supplementary Figs. 43 and 44) support this notion. With this rich and robust chemistry, we envision that the potential of this platform is far beyond artificial enzymes and can be readily expanded to applications such as the discovery of antimicrobial agents, the understanding of protein phase separation and the development of intracellular delivery systems for therapeutic biomacromolecules.

Methods

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 this solution was added 70% TBHP (2,760 µl; 12 equiv. Se), which was then stirred at

room temperature for 2 h before trimethylamine (800 µl; 3 equiv. Se) and 75 ml $NaHCO_3$ solution (1 M) were added. The system was stirred at 37 °C for 16 h, 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 dichloromethane and dialysed (MWCO 3,500 Da) against 0.5 M NaCl for 12 h, and then against water (which was changed twice every day) for 48 h. The remaining content was lyophilized to give the final product as a pale yellow powder (256 mg; yield = 64% from NCA).

General procedure for PPM of $PSeO_2Na$

Under the protection of nitrogen, $PSeO_2Na$ was dissolved in water (5–10 mg ml^{-1}). 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 min. After 20 min, another portion of TFA (1.5 µl) was added. The system was stirred at room temperature for 30 min and the completion of the reduction was indicated by the re-dissolution of the precipitate. Then, to the polymer solution was added the modifier in THF or water (as specified in the Supplementary Information section 'Post polymerization modification for homopolypeptides'). The system was stirred at the indicated temperature. The selenopolypeptide was purified by dialysis and/or SEC and recovered by lyophilization. The PPM of PEG-*b*-$PSeO_2Na$ was carried out similarly.

Random generation of RHPs

The 83 RHPs for random search were generated as an 83 × 7 array of uniformly sampled values between 0 and 1. Each row (relative abundance) 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 volumes of the organohalide solutions required for each well.

Bayesian optimization

The Bayesian optimization was implemented in Python using GPyTorch, BoTorch and Ax[101] (see the Jupyter notebook, BO_GPx_Ax022.ipynb, in Supplementary Information 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 were randomly split into 80/20 training/testing for surrogate model training. The program performed a hyperparameter optimization to select between four kernels (RBF, Matern-0.5, Matern-1.5 and Matern-2.5), ten random seeds and three different learning rates (0.01, 0.02 and 0.20) with RMSprop. The model with the lowest lost function (negative marginal log likelihood) on the test set was chosen. After training the surrogate model, 83 candidates for successive iteration were chosen by gen(), which was implemented in BoTorch and Ax. This method generates the candidates by optimizing an expected improvement 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 the total mole fraction equals 1. Bayesian optimization was performed on a Lenovo Legion R9000K laptop with an AMD Ryzen 9 5900HX CPU and a NVIDIA GeForce RTX 3080 Laptop GPU (16 GB).

Some constraints were set during the optimization. For x_1–x_5, the upper limits were set to 1 throughout the optimization. In the second round, the upper limit of x_7 (c_7) proposed by Bayesian optimization 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 synthesis and characterization. Thus, for the rest of the screening, the c_6 was set to 0.5 while the c_7 was set back to 1.0. It should be noted that even through these constraints were not imposed on the random generation, all randomly generated RHPs naturally fell into the space because the program is not very likely to generate a vector with an element >0.5.

Additional methods are provided in Supplementary Information.

Data availability

All of the data generated or analysed in this study are included in this published article and its Supplementary Information files. Source data are provided with this paper.

Code availability

The code for the generation of random RHPs and Bayesian optimization in this study is included in Supplementary Code 1.

References

1. Cole, J. P., Hanlon, A. M., Rodriguez, K. J. & Berda, E. B. Protein-like structure and activity in synthetic polymers. *J. Polym. Sci. A Polym. Chem.* **55**, 191–206 (2017).
2. Rothfuss, H., Knofel, N. D., Roesky, P. W. & Barner-Kowollik, C. Single-chain nanoparticles as catalytic nanoreactors. *J. Am. Chem. Soc.* **140**, 5875–5881 (2018).
3. Bonduelle, C. Secondary structures of synthetic polypeptide polymers. *Polym. Chem.* **9**, 1517–1529 (2018).
4. Varanko, A. K., Su, J. C. & Chilkoti, A. Elastin-like polypeptides for biomedical applications. *Annu. Rev. Biomed. Eng.* **22**, 343–369 (2020).
5. Callmann, C. E., Thompson, M. P. & Gianneschi, N. C. Poly(peptide): synthesis, structure, and function of peptide–polymer amphiphiles and protein-like polymers. *Acc. Chem. Res.* **53**, 400–413 (2020).
6. Jiang, T. et al. Single-chain heteropolymers transport protons selectively and rapidly. *Nature* **577**, 216–220 (2020).
7. Panganiban, B. et al. Random heteropolymers preserve protein function in foreign environments. *Science* **359**, 1239–1243 (2018).
8. Hilburg, S. L., Ruan, Z. Y., Xu, T. & Alexander-Katz, A. Behavior of protein-inspired synthetic random heteropolymers. *Macromolecules* **53**, 9187–9199 (2020).
9. Han, Z., Hilburg, S. L. & Alexander-Katz, A. Forced unfolding of protein-inspired single-chain random heteropolymers. *Macromolecules* **55**, 1295–1309 (2022).
10. Song, Z. Y., Tan, Z. Z. & Cheng, J. J. Recent advances and future perspectives of synthetic polypeptides from *N*-carboxyanhydrides. *Macromolecules* **52**, 8521–8539 (2019).
11. Song, Z. Y. et al. Synthetic polypeptides: from polymer design to supramolecular assembly and biomedical application. *Chem. Soc. Rev.* **46**, 6570–6599 (2017).
12. Zhou, X. F. & Li, Z. B. Advances and biomedical applications of polypeptide hydrogels derived from α-amino acid *N*-carboxyanhydride (NCA) polymerizations. *Adv. Healthcare Mater.* **7**, e1800020 (2018).
13. Deng, C. et al. Functional polypeptide and hybrid materials: precision synthesis via α-amino acid *N*-carboxyanhydride polymerization and emerging biomedical applications. *Prog. Polym. Sci.* **39**, 330–364 (2014).
14. Hou, Y. Q. & Lu, H. Protein PEPylation: a new paradigm of protein–polymer conjugation. *Bioconjugate Chem.* **30**, 1604–1616 (2019).
15. Deming, T. J. Synthetic polypeptides for biomedical applications. *Prog. Polym. Sci.* **32**, 858–875 (2007).
16. Liu, Y., Li, D., Ding, J. X. & Chen, X. S. Controlled synthesis of polypeptides. *Chin. Chem. Lett.* **31**, 3001–3014 (2020).
17. Ruggieri, M., Avolio, C., Livrea, P. & Trojano, M. Glatiramer acetate in multiple sclerosis: a review. *CNS Drug Rev.* **13**, 178–191 (2007).
18. Taylor, S. V., Walter, K. U., Kast, P. & Hilvert, D. Searching sequence space for protein catalysts. *Proc. Natl Acad. Sci. USA* **98**, 10596–10601 (2001).
19. Reis, M. et al. Machine-learning-guided discovery of F-19 MRI agents enabled by automated copolymer synthesis. *J. Am. Chem. Soc.* **143**, 17677–17689 (2021).
20. Macarron, R. et al. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10**, 188–195 (2011).

21. Coley, C. W., Eyke, N. S. & Jensen, K. F. Autonomous discovery in the chemical sciences part I: progress. *Angew. Chem. Int. Ed.* **59**, 22858–22893 (2020).
22. Yang, L. L. et al. High-throughput methods in the discovery and study of biomaterials and materiobiology. *Chem. Rev.* **121**, 4561–4677 (2021).
23. Soheilmoghaddam, F., Rumble, M. & Cooper-White, J. High-throughput routes to biomaterials discovery. *Chem. Rev.* **121**, 10792–10864 (2021).
24. DeBenedictis, E. A. et al. Systematic molecular evolution enables robust biomolecule discovery. *Nat. Methods* **19**, 55–64 (2022).
25. Gromski, P. S., Granda, J. M. & Cronin, L. Universal chemical synthesis and discovery with 'the Chemputer'. *Trends Chem.* **2**, 4–12 (2020).
26. Pollice, R. et al. Data-driven strategies for accelerated materials design. *Acc. Chem. Res.* **54**, 849–860 (2021).
27. Correa-Baena, J. P. et al. Accelerating materials development via automation, machine learning, and high-performance computing. *Joule* **2**, 1410–1420 (2018).
28. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **16**, 687–694 (2019).
29. Vamathevan, J. et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).
30. Kumar, R. et al. Efficient polymer-mediated delivery of gene-editing ribonucleoprotein payloads through combinatorial design, parallelized experimentation, and machine learning. *ACS Nano.* **14**, 17626–17639 (2020).
31. Kumar, R., Le, N., Oviedo, F., Brown, M. E. & Reineke, T. M. Combinatorial polycation synthesis and causal machine learning reveal divergent polymer design rules for effective pDNA and ribonucleoprotein delivery. *JACS Au* **2**, 428–442 (2022).
32. Goldberg, M., Mahon, K. & Anderson, D. Combinatorial and rational approaches to polymer synthesis for medicine. *Adv. Drug. Deliv. Rev.* **60**, 971–978 (2008).
33. Baudis, S. & Behl, M. High-throughput and combinatorial approaches for the development of multifunctional polymers. *Macromol. Rapid Commun.* **43**, 2100400 (2022).
34. Holmes, P. F., Bohrer, M. & Kohn, J. Exploration of polymethacrylate structure–property correlations: advances towards combinatorial and high-throughput methods for biomaterials discovery. *Prog. Polym. Sci.* **33**, 787–796 (2008).
35. Gormley, A. J. & Webb, M. A. Machine learning in combinatorial polymer chemistry. *Nat. Rev. Mater.* **6**, 642–644 (2021).
36. Upadhya, R. et al. Automation and data-driven design of polymer therapeutics. *Adv. Drug. Deliv. Rev.* **171**, 1–28 (2021).
37. Patel, R. A., Borca, C. H. & Webb, M. A. Featurization strategies for polymer sequence or composition design by machine learning. *Mol. Syst. Des. Eng.* **7**, 661–676 (2022).
38. Oliver, S., Zhao, L., Gormley, A. J., Chapman, R. & Boyer, C. Living in the fast lane high throughput controlled/living radical polymerization. *Macromolecules* **52**, 3–23 (2019).
39. Lynn, D. M., Anderson, D. G., Putnam, D. & Langer, R. Accelerated discovery of synthetic transfection vectors: parallel synthesis and screening of a degradable polymer library. *J. Am. Chem. Soc.* **123**, 8155–8156 (2001).
40. Green, J. J., Langer, R. & Anderson, D. G. A combinatorial polymer library approach yields insight into nonviral gene delivery. *Acc. Chem. Res.* **41**, 749–759 (2008).
41. Gormley, A. J. et al. An oxygen-tolerant PET–RAFT polymerization for screening structure–activity relationships. *Angew. Chem. Int. Ed.* **57**, 1557–1562 (2018).
42. Judzewitsch, P. R. et al. High-throughput process for the discovery of antimicrobial polymers and their upscaled production via flow polymerization. *Macromolecules* **53**, 631–639 (2020).
43. Kosuri, S. et al. Machine-assisted discovery of chondroitinase ABC complexes toward sustained neural regeneration. *Adv. Healthcare Mater.* **11**, 2102101 (2022).
44. Tamasi, M. J. et al. Machine learning on a robotic platform for the design of polymer–protein hybrids. *Adv. Mater.* **34**, 2201809 (2022).
45. Gauthier, M. A., Gibson, M. I. & Klok, H. A. Synthesis of functional polymers by post-polymerization modification. *Angew. Chem. Int. Ed.* **48**, 48–58 (2009).
46. Gunay, K. A., Theato, P. & Klok, H. A. Standing on the shoulders of Hermann Staudinger: post-polymerization modification from past to present. *J. Polym. Sci. A Polym. Chem.* **51**, 1–28 (2013).
47. Zhong, Y. B., Zeberl, B. J., Wang, X. & Luo, J. T. Combinatorial approaches in post-polymerization modification for rational development of therapeutic delivery systems. *Acta Biomater.* **73**, 21–37 (2018).
48. Ladmiral, V. et al. Synthesis of neoglycopolymers by a combination of "click chemistry" and living radical polymerization. *J. Am. Chem. Soc.* **128**, 4823–4830 (2006).
49. Wong, S. Y., Sood, N. & Putnam, D. Combinatorial evaluation of cations, pH-sensitive and hydrophobic moieties for polymeric vector design. *Mol. Ther.* **17**, 480–490 (2009).
50. Pedone, E., Li, X. W., Koseva, N., Alpar, O. & Brocchini, S. An information rich biomedical polymer library. *J. Mater. Chem.* **13**, 2825–2837 (2003).
51. Yan, Y. F. et al. Functional polyesters enable selective siRNA delivery to lung cancer over matched normal cells. *Proc. Natl. Acad. Sci. USA* **113**, E5702–E5710 (2016).
52. Wyrsta, M. D., Cogen, A. L. & Deming, T. J. A parallel synthetic approach for the analysis of membrane interactive copolypeptides. *J. Am. Chem. Soc.* **123**, 12919–12920 (2001).
53. Deming, T. J. Synthesis of side-chain modified polypeptides. *Chem. Rev.* **116**, 786–808 (2016).
54. Deming, T. J. Functional modification of thioether groups in peptides, polypeptides, and proteins. *Bioconjugate Chem.* **28**, 691–700 (2017).
55. Lu, H. et al. Ring-opening polymerization of γ-(4-vinylbenzyl)-L-glutamate *N*-carboxyanhydride for the synthesis of functional polypeptides. *Macromolecules* **44**, 6237–6240 (2011).
56. Zhou, J. R. et al. A simple and versatile synthetic strategy to functional polypeptides via vinyl sulfone-substituted L-cysteine *N*-carboxyanhydride. *Macromolecules* **46**, 6723–6730 (2013).
57. Engler, A. C., Lee, H. I. & Hammond, P. T. Highly efficient "grafting onto" a polypeptide backbone using click chemistry. *Angew. Chem. Int. Ed.* **48**, 9334–9338 (2009).
58. Krannig, K. S. & Schlaad, H. pH-responsive bioactive glycopolypeptides with enhanced helicity and solubility in aqueous solution. *J. Am. Chem. Soc.* **134**, 18542–18545 (2012).
59. Cao, J. B. et al. Non-ionic water-soluble "clickable" α-helical polypeptides: synthesis, characterization and side chain modification. *Polym. Chem.* **6**, 1226–1229 (2015).
60. Xie, Y., Lopez-Silva, T. L. & Schneider, J. P. Hydrophilic azide-containing amino acid to enhance the solubility of peptides for SPAAC reactions. *Org. Lett.* **24**, 7378–7382 (2022).
61. Pickens, C. J., Johnson, S. N., Pressnall, M. M., Leon, M. A. & Berkland, C. J. Practical considerations, challenges, and limitations of bioconjugation via azide–alkyne cycloaddition. *Bioconjugate Chem.* **29**, 686–701 (2018).
62. Liu, J., Chen, Q. Q. & Rozovsky, S. Utilizing selenocysteine for expressed protein ligation and bioconjugations. *J. Am. Chem. Soc.* **139**, 3430–3437 (2017).
63. Zhao, Z. G., Shimon, D. & Metanis, N. Chemoselective copper-mediated modification of selenocysteines in peptides and proteins. *J. Am. Chem. Soc.* **143**, 12817–12824 (2021).
64. Quaderer, R., Sewing, A. & Hilvert, D. Selenocysteine-mediated native chemical ligation. *Helv. Chim. Acta* **84**, 1197–1206 (2001).

65. Li, X. et al. Stable and potent selenomab–drug conjugates. *Cell Chem. Biol.* **24**, 433–442 e436 (2017).
66. Li, X. L. et al. Site-specific dual antibody conjugation via engineered cysteine and selenocysteine residues. *Bioconjugate Chem.* **26**, 2243–2248 (2015).
67. Sayers, J. et al. Construction of challenging proline–proline junctions via diselenide–selenoester ligation chemistry. *J. Am. Chem. Soc.* **140**, 13327–13334 (2018).
68. Flemer, S. Jr. Selenol protecting groups in organic chemistry: special emphasis on selenocysteine Se-protection in solid phase peptide synthesis. *Molecules.* **16**, 3232–3251 (2011).
69. Sharpless, K. B., Lauer, R. F. & Teranishi, A. Y. Electrophilic and nucleophilic organoselenium reagents. New routes to α,β-unsaturated carbonyl compounds. *J. Am. Chem. Soc.* **95**, 6137–6139 (1973).
70. Wu, J. A. et al. The functionalization of poly(ε-caprolactone) as a versatile platform using ε-(α-phenylseleno) caprolactone as a monomer. *Polym. Chem.* **10**, 3851–3858 (2019).
71. Yu, L., Zhang, M., Du, F. S. & Li, Z. C. ROS-responsive poly(ε-caprolactone) with pendent thioether and selenide motifs. *Polym. Chem.* **9**, 3762–3773 (2018).
72. Wang, L. et al. ROS-triggered degradation of selenide-containing polymers based on selenoxide elimination. *Polym. Chem.* **10**, 2039–2046 (2019).
73. Reich, H. J., Wollowitz, S., Trend, J. E., Chow, F. & Wendelborn, D. F. Syn elimination of alkyl selenoxides. Side reactions involving selenenic acids. Structural and solvent effects on rates. *J. Org. Chem.* **43**, 1697–1705 (1978).
74. Yang, Y. et al. Genetically encoded releasable photo-cross-linking strategies for studying protein–protein interactions in living cells. *Nat. Protoc.* **12**, 2147–2168 (2017).
75. Yang, Y. et al. Genetically encoded protein photocrosslinker with a transferable mass spectrometry-identifiable label. *Nat. Commun.* **7**, 12299 (2016).
76. Lin, S. et al. Genetically encoded cleavable protein photo-cross-linker. *J. Am. Chem. Soc.* **136**, 11860–11863 (2014).
77. Tian, Z. Y., Zhang, Z. C., Wang, S. & Lu, H. A moisture-tolerant route to unprotected α/β-amino acid *N*-carboxyanhydrides and facile synthesis of hyperbranched polypeptides. *Nat. Commun.* **12**, 5810 (2021).
78. Wu, G. et al. Synthesis of water soluble and multi-responsive selenopolypeptides via ring-opening polymerization of *N*-carboxyanhydrides. *Chem. Commun.* **55**, 7860–7863 (2019).
79. Lin, Y. Y. A. et al. Rapid cross-metathesis for reversible protein modifications via chemical access to Se-allyl-selenocysteine in proteins. *J. Am. Chem. Soc.* **135**, 12156–12159 (2013).
80. Reddy, K. M. & Mugesh, G. Application of dehydroalanine as a building block for the synthesis of selenocysteine-containing peptides. *RSC Adv.* **9**, 34–43 (2019).
81. Banik, S. M. et al. Lysosome-targeting chimaeras for degradation of extracellular proteins. *Nature* **584**, 291–297 (2020).
82. Ahn, G. et al. LYTACs that engage the asialoglycoprotein receptor for targeted protein degradation. *Nat. Chem. Biol.* **17**, 937–946 (2021).
83. Delaveris, C. S., Chiu, S. H., Riley, N. M. & Bertozzi, C. R. Modulation of immune cell reactivity with *cis*-binding Siglec agonists. *Proc. Natl Acad. Sci. USA* **118**, e2012408118 (2021).
84. Du, J. J. et al. Glycopeptide ligation via direct aminolysis of selenoester. *Chin. Chem. Lett.* **29**, 1127–1130 (2018).
85. Temperini, A., Piazzolla, F., Minuti, L., Curini, M. & Siciliano, C. General, mild, and metal-free synthesis of phenyl selenoesters from anhydrides and their use in peptide synthesis. *J. Org. Chem.* **82**, 4588–4603 (2017).
86. Chen, L. H. et al. Polymer informatics: current status and critical next steps. *Mater. Sci. Eng. R Rep.* **144**, 100595 (2021).
87. Upadhya, R., Kanagala, M. J. & Gormley, A. J. Purifying low-volume combinatorial polymer libraries with gel filtration columns. *Macromol. Rapid Commun.* **40**, 1900528 (2019).
88. Barbosa, N. V. et al. Organoselenium compounds as mimics of selenoproteins and thiol modifier agents. *Metallomics* **9**, 1703–1734 (2017).
89. Huang, X., Liu, X. M., Luo, Q. A., Liu, J. Q. & Shen, J. C. Artificial selenoenzymes: designed and redesigned. *Chem. Soc. Rev.* **40**, 1171–1184 (2011).
90. Seibt, T. M., Proneth, B. & Conrad, M. Role of GPX4 in ferroptosis and its pharmacological implication. *Free Radic. Biol. Med.* **133**, 144–152 (2019).
91. Yant, L. J. et al. The selenoprotein GPX4 is essential for mouse development and protects from radiation and oxidative damage insults. *Free Radic. Biol. Med.* **34**, 496–502 (2003).
92. Xu, C. X. et al. The glutathione peroxidase Gpx4 prevents lipid peroxidation and ferroptosis to sustain Treg cell activation and suppression of antitumor immunity.*Cell Rep.* **35**, 109235 (2021).
93. Parnham, M. & Sies, H. Ebselen: prospective therapy for cerebral ischaemia. *Expert Opin. Investig. Drugs* **9**, 607–619 (2000).
94. Landgraf, A. D. et al. Neuroprotective and anti-neuroinflammatory properties of ebselen derivatives and their potential to inhibit neurodegeneration. *ACS Chem. Neurosci.* **11**, 3008–3016 (2020).
95. Yamagata, K., Ichinose, S., Miyashita, A. & Tagami, M. Protective effects of ebselen, a seleno-organic antioxidant on neurodegeneration induced by hypoxia and reperfusion in stroke-prone spontaneously hypertensive rat. *Neuroscience.* **153**, 428–435 (2008).
96. Paglia, D. E. & Valentine, W. N. Studies on the quantitative and qualitative characterization of erythrocyte glutathione peroxidase. *J. Lab. Clin. Med.* **70**, 158–169 (1967).
97. Shahriari, B., Swersky, K., Wang, Z. Y., Adams, R. P. & de Freitas, N. Taking the human out of the loop: a review of Bayesian optimization. *Proc. IEEE* **104**, 148–175 (2016).
98. Shields, B. J. et al. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).
99. Nambiar, A. M. K. et al. Bayesian optimization of computer-proposed multistep synthetic routes on an automated robotic flow platform. *ACS Cent. Sci.* **8**, 825–836 (2022).
100. Hase, F., Roch, L. M., Kreisbeck, C. & Aspuru-Guzik, A. Phoenics: a Bayesian optimizer for chemistry. *ACS Cent. Sci.* **4**, 1134–1145 (2018).
101. Balandat, M. et al. BoTorch: a framework for efficient Monte-Carlo Bayesian optimization. *Adv. Neural Inf. Process. Syst.* **33**, 21524–21538 (2020).
102. Shao, L. X., Li, Y. M., Lu, J. M. & Jiang, X. F. Recent progress in selenium-catalyzed organic reactions. *Org. Chem. Front.* **6**, 2999–3041 (2019).
103. Reich, H. J. & Hondal, R. J. Why nature chose selenium. *ACS Chem. Biol.* **11**, 821–841 (2016).
104. Xia, J. H., Li, T. Y., Lu, C. J. & Xu, H. P. Selenium-containing polymers: perspectives toward diverse applications in both adaptive and biomedical materials. *Macromolecules* **51**, 7435–7455 (2018).
105. Li, Q. L. et al. Organoselenium chemistry-based polymer synthesis. *Org. Chem. Front.* **7**, 2815–2841 (2020).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2019YFA0904203 to H.L.), National Natural Science Foundation of China (22125101 and 21975004 to H.L. and 22105008 to G.W.) and Beijing Natural Science Foundation (2220023 to H.L.), a fellowship of the China Postdoctoral Science Foundation (2020M680192 to G.W.) and the Li Ge-Zhao Ning Life Science Youth Research Foundation (LGZNQN202206 to H.L.).

G.W. thanks the Boya Postdoctoral Fellowship of Peking University for financial support. G.W. thanks S. Goldman (Department of Chemical Engineering, Massachusetts Institute of Technology) and Z. Wang (Department of Industrial Engineering, Tsinghua University) for the discussion and R. Mercado (Department of Chemical Engineering, Massachusetts Institute of Technology) for examining the code. Figure 5a and graphical abstract created with BioRender.com.

Author contributions

G.W. and H.L. designed and directed the research. G.W., Z.-Y.T., X.L. and S.W. performed the synthesis of monomers and the precursor polypeptide. G.W. and H.Z. built and validated the HTS and characterization platform and performed the screening of GPx mimics. G.W. and J.Z. wrote the code of Bayesian optimization. G.W., C.W.C. and H.L. wrote the original draft. All authors reviewed and accepted the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44160-023-00294-7.

Correspondence and requests for materials should be addressed to Hua Lu.

Peer review information *Nature Synthesis* thanks Adam Gormley, Michael Webb and Huaping Xu for their contribution to the peer review of this work. Primary Handling Editor: Peter Seavill, in collaboration with the *Nature Synthesis* team.