UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Biodegradable, Biohybrid, and Biocheminformatically Designed Plastics

Permalink

https://escholarship.org/uc/item/34k5k1h1

Author

Jayapurna, Ivan Ferdo

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Biodegradable, Biohybrid, and Biocheminformatically Designed Plastics

By

Ivan Ferdo Jayapurna

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Applied Science & Technology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ting Xu, Chair

Professor Haiyan Huang

Professor Kevin Healy

Spring 2024

Copyright © 2024

By

Ivan F Jayapurna

Abstract

Biodegradable, Biohybrid, and Biocheminformatically Designed Plastics

By

Ivan Ferdo Jayapurna

Doctor of Philosophy in Applied Science & Technology

University of California, Berkeley

Professor Ting Xu, Chair

Water pollution, climate change, and human health are among many critical challenges that biodegradable plastics can help solve. The ideal plastic can be rapidly produced in large volumes with flexible dimensionality, as well as rapidly degraded to molecular products to avoid micro- and nano-plastic generation. Synthetic polymers offer desirable production time- and length-scales while biological polymers have desirable degradation time- and length-scales. By blending biotic and abiotic components a single biohybrid material with both synthetic scalability and biological sophistication can be realized. Pre-programming over which function will be dominant at each timescale can be achieved by designing across molecular-to-macroscopic length scales. By nanoscopically dispersing enzymes in a hydrophobic polymer matrix that is also the substrate for biocatalysis, we achieve control over the resulting hybrid plastic's degradation timescale. During production and use, the abiotic properties are dominant, until biotic functionality is controllably activated upon end-of-life by basic environmental triggers *e.g.,* water. To fully unlock the economic and bioremediative potential of biohybrid, biodegradable plastics a fundamental scientific understanding of molecular mechanisms and biotic-abiotic interactions is required. Synthetic random heteropolymers (RHPs) are rationally designable chaperones that can modulate biotic-abiotic interfaces during both the production and degradation of plastics. RHPs are also a platform technology that, beyond biohybrid materials, can also independently recapitulate protein-like functions.

Here, we elucidated the molecular-scale feedback loop between enzyme-catalyzed plastic degradation and molecular transport through an evolving, hierarchical, semi-crystalline plastic matrix. Modulating the enzyme-polymer interface resulted in polyester degrading enzymes that were thermostable in over 200 $^{\circ}$C polymer melts, enabling industrially scalable biohybrid plastic production. To accelerate the design and optimization of RHPs as either protein chaperones or mimics, we built the *RHPapp*: a RAFT kinetic heteropolymerization simulator that feeds into a biocheminformatic sequence analysis framework. *In silico* generation of accurate heteropolymer sequences in lieu of sequencing technologies allows us to leverage the advances made in bioinformatics and artificial intelligence to lay a foundation for the burgeoning field of macromolecular cheminformatics.

Biodegradable, Biohybrid, and Biocheminformatically Designed Plastics

Table of Contents

Acknowledgements iii

Introduction 1

Chapter 1: Biocatalysis under nanoconfinement: a molecular scale feedback loop

1.1 Plastic chemical transformation by nanoscopically embedded catalysts 4
1.2 Limited diffusion shuts down catalysis 6
1.3 Water transport modulates catalysis 8
1.4 Molecular mechanism: degradation drives diffusion and vice-versa 11
1.5 Implications and applications: real-world environments and upcycling 14
1.6 Conclusion 17
1.7 Methods and materials 18

Chapter 2: Enzyme degradation mechanisms on and in semi-crystalline substrates

2.1 Challenges with enzymatic degradation of semi-crystalline polymers 21
2.2 Hydrolase features predict promiscuity for polyester degradation 23
2.3 A medley of mechanisms can degrade semi-crystalline polyesters 25
2.4 Active site features and mechanism predict random heteropolymer interaction 32
2.5 Enzyme confinement and nano-confinement shift degradation mechanism 35
2.6 In silico degradation mechanism prediction via molecular simulations 41
2.7 Methods and materials 44

Chapter 3: Modulating protein stability via intermolecular protectants

3.1 Industrial bioactive plastic processing 47
3.2 Thermostable proteins in 200°C polymer melts 50
3.3 Universal rules governing protein thermostability in polymer melts and matrices 53
3.4 Salt-enzyme-polymer specific thermostabilizing interactions 57
3.5 Future work 59
3.6 Methods and materials 60

Chapter 4: Biocheminformatic sequence design of biomimetic heteropolymers

4.1 Heteropolymerization simulation: an orthogonal path to sequencing 61
4.2 *RHPapp 1.0* (Methods) 63
4.3 Simulation scale and fidelity 66
4.4 Sequence design of protein-interfacing random heteropolymers 69
4.5 Sequence design of random heteropolymer with biomimetic function 72

Chapter 5: Improving simulated heteropolymer sequence fidelity

5.1 Limitations of *RHPapp 1.0* 74
5.2 Determining reactivity ratios in RAFT heteropolymerization 75
5.3 Composition evolution during synthesis and purification 77
5.4 *RHPapp 2.0*: a RAFT heteropolymer kinetics and sequence simulator 83
5.5 Effect of RAFT polymerization kinetics on binary copolymer sequences 89
5.6 Effect of RAFT polymerization kinetics on heteropolymer sequences 94
5.7 Methods and materials 100

Chapter 6: Deep learning driven heteropolymer design

6.1 Learning the language of life, synthetically 103
6.2 A hybrid variational autoencoder for designing synthetic protein mimics 104
6.3 Transferability of protein large language models to synthetic mimics 109
6.4 Future work 115
6.5 Methods 116

References

- Introduction 118
- Chapter 1 121
- Chapter 2 125
- Chapter 3 130
- Chapter 4 133
- Chapter 5 138
- Chapter 6 144

Supporting Information

- Chapter 1 147
- Chapter 2 151
- Chapter 3 162
- Chapter 4 166
- Chapter 5 170
- Chapter 6 181

Acknowledgements

This dissertation would not have been possible without Prof. Ting Xu believing in me and encouraging me to pursue a PhD, despite it not being something I had ever even dreamt of. Throughout the past 5 years she has supported, mentored, and pushed me to improve as a scientist. Similarly, I'm very thankful for Aaron and Deepak, both of whom provided me invaluable mentorship and opportunities that led me to where I am today, as well as the unconditional support from my family, friends, and especially Soo. I'm also grateful for my many great lab mates and collaborators, whose significant contributions are acknowledged in the relevant sub-chapters: 5.2, 5.3, and 6.2. Finally, it has been the greatest honor and joy to have spent the last 5 years mentoring and working every day with my undergraduate research team: Meera, Prajna, Ariel, Spencer, Sanjana, Lani, Beatrice, Phillip, and Pierre.

Introduction

Plastic pollution is one of many sustainability problems that can be solved by polymers with advanced functionalities, with solutions including compostable, chemically recyclable, edible, and nanocomposite materials.[1–5] Driven by the advent of artificial intelligence, *in silico* modeling, and high throughput robotics, the long-term future of functional polymer development may be de novo design and machine-automated evolution.[6,7] However, at present, there is still a gap in efficacy between the spanning set of human-machine-designed solutions and those found in nature, which have evolved over eons. Natural polymer composite materials such as the lotus leaf, can be both biodegradable and superhydrophobic, seemingly antithetical functionalities. Water barrier properties exist for the exact timescale required before breaking down completely and enriching its surrounding environment, achieving full circularity.[4,8] This is the target level of design sophistication we should aspire towards - a single hybrid material system with pre-programmed timescale control that is achieved by designing across molecular to macroscopic length scales. The conventional material design paradigms of today often work to destroy nature. Instead, future materials should be co-designed with nature, leveraging ancient biological wisdom to improve functionality without compromising on sustainability or scalability.

The incompatibility between nature-designed biopolymers and human demand is a mismatch of length scales and timescales. For example, polyethylene (PE) is a versatile, synthetic material that can be manipulated into any desired dimensionality, and can be produced at a rate of ~3.5 metric tons per second.[9] In comparison, bamboo, one of the fastest growing biological materials, has an upper bound growth of 0.001 cm per second.[10] Human-engineered biodegradable plastics are a multibillion-dollar industry with 20-30% projected annual growth that have the potential to mitigate single-use plastic waste.[11] However, most of these plastics have degradation timescales that are incompatible with both end-of-life environments that are natural (*e.g.,* oceans and rivers) and specifically designed (*e.g.,* compost and landfill). This results in material waste with length scales that are non-circular and even harmful (*e.g.,* micro- and nano-plastics) whose dangers we are still in the infancy of understanding.[12–14] Generally, synthetic materials have desirable production time- and length-scales, while biological materials have desirable degradation time- and length-scales.

Materials with both synthetic scalability and biological sophistication can be achieved by blending biotic and abiotic components with desired properties into a single biohybrid material. However, modulation of which function amongst multiple would be dominant at the desired timescales is required. Achieving this requires a deep understanding to control design of all relevant length scales. In the example of bioactive plastics, a duet of synergistic functionalities is imparted by biotic and abiotic components. Often a polymer matrix provides mechanical strength, flexibility, and barrier properties, while embedded biological additives impart advanced functionality *e.g.,* catalysis, sensing and/or therapeutics.[15–17] Enzymes are naturally occurring, best-in-class catalysts. Many examples have shown enzymes (hydrolases and oxidases) from bacteria, fungi, and animals

to have sufficient promiscuity to degrade synthetic polymer chains.[18,19] When the host matrix of a biohybrid material is the substrate for enzymatic catalysis, the resulting material is a biodegradable plastic.[20–22] During the timescale of use, the abiotic material properties should be dominant until biotic material function is activated upon end-of-life. Embedding biocatalysts within plastics provides this portable, on-demand triggerable degradation (*i.e.,* time-scale control), as the enzymes in these bioactive plastics only require water and warm temperatures to activate biotic material function. This perfectly matches industrial compost conditions, thus unlocking accelerated, circular, polymer-to-monomer degradation.[23] Bioactive plastics are a particularly exciting solution as they are theoretically compatible with existing traditional plastics manufacturing infrastructure, in which biological components can be incorporated as simple additives.[15,24] Furthermore, for decades we have produced enzymes at scale for use in a plethora of commercial products.[25,26]

The key knowledge gap remaining is the lack of control of the interface between the biotic and abiotic phases. To design as nature does, we need an understanding of interfacial interactions, mechanisms and kinetics of catalysis and diffusion, and morphological evolution across the entire molecular-nano-micro-macro-scale hierarchy and the entire biohybrid material life cycle: from processing to use to end-of-life. In practice, there are two outstanding issues in incorporating proteins and other biotics into biohybrid materials. First, as discussed, time- and length- scales must be guided to avoid partially degraded macro-, micro-, and nano- plastics contaminating composts and natural degradation environments. Secondly, enzymes are fragile outside of their native environments and in general, not innately amenable to organic solvents nor high temperature, viscous polymer melts required for plastic processing.[27] Both issues can be addressed by a rationally designable interfacing agent that is able to chaperone biotic-abiotic interactions.

Random heteropolymers (RHPs) are composed of three or more monomers, with sequences that are statistically defined. RHPs have been shown to successfully increase protein thermostability and solvent stability in non-native environments.[27,28] RHPs were designed to chaperone biotic-abiotic interactions by mimicking intrinsically disordered and small heat shock proteins, with dynamic interactions of moderate strength.[29–32] In comparison to proteins, RHPs are synthetic, and polydisperse in molecular weight and composition. However, recent work has shown that RHPs alone can independently recapitulate protein-like function.[33–35] Heteropolymer design not only offers a rationally designable interfacing agent, but it also unlocks a whole suite of advanced functional materials, of which biohybrid materials are just a small subset, and biodegrade plastics an even smaller sub-subset.

To navigate through the combinatorially complex heteropolymer design space of tunable parameters, computation and automation must come into play. Historically, in biological polymers, DNA sequencing, bioinformatics and publicly collaborative databases of millions of sequences unlocked the explosion of rational design in genetics and protein engineering.[36] Transferring learnings, methods, and tools from bioinformatics can accelerate the development of RHPs as chaperone agents and synthetic protein mimics. However, to effectively communicate between the threeway nature-human-machine design team, a

shared language is required. Recent breakthroughs in artificial intelligence, namely large language models, can guide us towards learning the “language of life” (*i.e.,* the “grammar rules” of biology), that should hopefully provide wisdom in designing biocheminformatically inspired synthetic and hybrid systems.

In this work, I addressed the challenges and opportunities presented above by exploring the making and breaking of bioinspired, biodegradable, and biohybrid functional plastics. We designed functional biohybrid macromolecules that can be manufactured and degraded in time-scales compatible with societal demands. Specifically, we developed truly compostable plastics via catalysis under nanoconfinement in the solid-state and elucidated its degradation pathway from the molecular through to macroscopic length-scale. In the process we share practical takeaways for the general development and engineering of sustainable plastics and plastic degrading enzymes. Our analysis considered the often-neglected influences of real-world environments and semi-crystalline substrates. Fundamentals of enzyme stability in extreme conditions were investigated, unlocking thermostable proteins in >200^{o}C plastic melts, revealing new learnings in solid-state enzymology, and offering practical engineering applications for industrial bioactive plastic manufacturing. Finally, a first-of-its-kind heteropolymer biocheminformatics simulation, design and analysis framework and software (*RHPapp*) has been developed with demonstrated applications. The *RHPapp* includes a kinetic Monte Carlo RAFT polymerization simulator for heteropolymers to generate sequence information in lieu of sequencing technologies, bioinformatics inspired sequence analysis modules, and leverages large protein language models to generate biologically labeled and biologically rich latent spaces to guide advanced functional material analysis and design.

Chapter 1

Biocatalysis Under Nanoconfinement: A Molecular Scale Feedback Loop

1.1 Plastic chemical transformation by nanoscopically embedded catalysts

Nanoscopically embedded catalysts inside of polymeric matrices can maximize catalyst-substrate interfaces, modulate reaction pathways, program reaction latency, and ensure continued chemical modification upon plastic disintegration to microplastics. Nanoscopic reactions on the surrounding polymeric substrates can affect catalytic activity, alter reaction pathways, and create a molecular feedback loop. The resultant reaction rate, pathways and extent of reaction reflect the interplay among competing processes. Despite the system complexity associated with this interdependent relationship where host polymers serve as both the host and substrate, understanding catalyzed reactions under nanoconfinement will determine our ability to chemically transform plastics, should it be upcycling, recycling, circular usage, or degradation under real-world environments.

The rate and extent of catalyzed plastic chemical transformations depends on molecular-scale reactions, transport through the molecular to nano-micro-macroscopic structural hierarchy, and interplay among competing processes (**Fig 1.1**). Nanoscopically embedded catalysts can program polymer degradation to meet demands for compostable single-use plastics.[1–8] Degradation leads to nanoscale voids and their subsequent growth constantly changes the local microenvironment, molecular transport, and thus substrate and product concentrations.[9,10] These factors shift kinetic and thermodynamic equilibria and compromise our ability to modulate competition between different pathways. Once the void growth exceeds a percolation threshold, cohabitants in the external environment can catalyze, poison catalysts or act as competing substrates. The evolution of kinetics (time) and morphology (space) are thus intertwined.

Here, we investigated enzyme-catalyzed polyester hydrolysis under nanoconfinement and focused on interplay among reactant diffusion, byproduct removal, hierarchical structure evolution of the host matrix, and competing reactions from cohabitants. The system of choice is a 3-component model system (ePCL) consisting of: 99.16 wt% polyester host matrix (80 kDa polycaprolactone, PCL), 0.04 wt% hydrolase (lipase from *Burkholderia cepacia*, LIP), and 0.8 wt% random heteropolymer (RHP) as an enzyme protectant and dispersing agent.[11] ePCL films, 20-100 μm in thickness, were solution cast and thermally treated to control morphology and thermal history. Three degradation environments were tested: 40-60% wet basis compost soil, ≥99% relative humidity (99% RH) air to mimic hot, humid climates such as Southeast Asia (a hotspot for single-use plastics consumption and subsequent ocean pollution),[12] and pH 7.4 phosphate buffer solution (PBS). Unless otherwise stated, all ePCL films were incubated for degradation at 40°C, within range of compost operating temperatures. We showed that understanding the transport of substrates e.g., water, acidic small molecules, heavy elements, and nanoparticles is critical for modulating nanoconfined catalysis. We presented two methods to deactivate

nanoconfined biocatalysis: trapped product accumulation and heavy element cohabitant adsorption. Water availability e.g. storage humidity and compost moisture, can modulate the extent of degradation. Partially degraded plastics naturally transformed into sponge-like, hierarchically porous structures with high surface area containing reactive end-groups. This morphology leads to a kinetic competition between chain-end interactions with catalysts for polymer chain scission and side-reactions with cohabitants. Thus, unlocking opportunities for plastic waste upcycling to porous scaffolds for heterogeneous catalysis and environmental remediation of cohabiting pollutants.

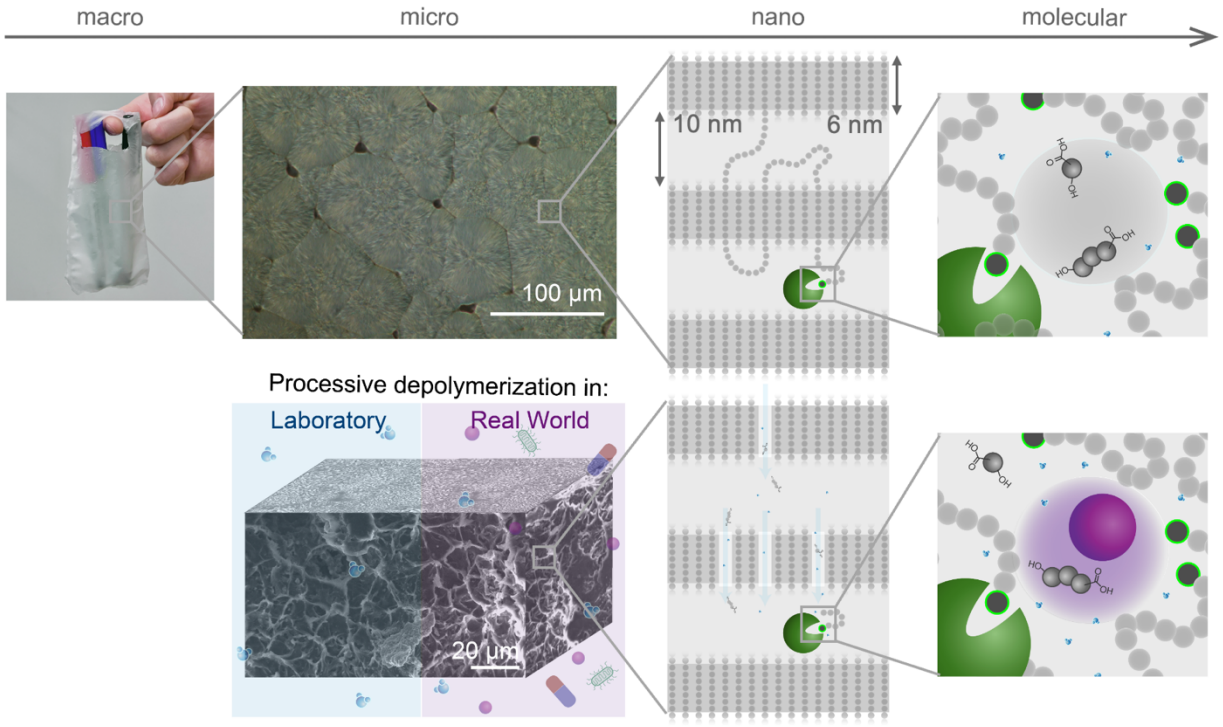


Figure 1.1 | Macro through molecular structure hierarchy of ePCL degradation. Macro: image of an enzyme-embedded polycaprolactone (ePCL) plastic bag. Micro: polarized optical microscopy image of ePCL surface and a scanning electron-microscopy derived schematic that illustrates a partially degraded plastic in real world conditions with cohabitants e.g., water, heavy metals, organic compounds, microbes. Nano: Enzymes in the amorphous regions of ePCL lamellae structure, a single polymer chain is depicted on the monomer scale, and its processive depolymerization creating an angstrom-scale tunnel for small molecule transport is shown. Molecular: species at local reaction sites compete for interactions that result in the end fate of plastics.

1.2 Limited diffusion shuts down catalysis

For catalyzed hydrolysis under confinement to occur, water and polymer chain substrates must be molecularly local to catalyst active sites and properly oriented for substrate docking. To assess enzyme activity and enzyme and/or PCL chain mobility post-processing, ePCL films were characterized before and after 24-hour incubation in 40°C 99% RH air. Macroscopically, no detectable mass loss was observed within the analytical balance sensitivity (±0.1 mg). Yet, quasi-static tensile tests revealed the Young's modulus of ePCL decreased by 26% from 463±73 (n=3) to 342±55 (n=3) MPa, suggesting polymer degradation. The Young's modulus of control PCL films (no enzyme) did not decrease (**Fig. S1.1**). ePCL was subsequently dissolved for gel permeation chromatography (GPC) analysis. There was a decrease in average PCL molecular weight by 1.5 kDa, corresponding to a chain-scission of ~12 repeat units (**Fig. 1.2a**). We hypothesized that accumulation of acidic hydrolysis products may change enzyme local microenvironment and inhibit further biocatalysis. To visually confirm enzymatic activity and thus the generation and local accumulation of acidic products, we co-embedded a pH-sensitive dye, bromocresol green (BG), into ePCL (ePCL-BG) and control (Control-BG) films. In ePCL-BG films, the BG co-localized with the enzyme. Water transported rapidly across the film surface and underwent observable micro-scale colour change at pH 4 within seconds (**Fig. S1.2**). After 40°C 99% RH air incubation, ePCL-BG changed colour from blue to green, signaling a decrease in pH to ≤4 (**Fig. 1.2b**). Small angle x-ray scattering (SAXS) done on an ePCL film before and after 99% RH air incubation resulted in an increase in slope at low scattering vector (q), with power law factor (P) increasing from ~3.2 to ~3.3 (**Fig. 1.2c**). The increasing low q slope suggests the formation of surface fractal-like features sized < 60 nm, further indicative of enzymatic activity. A control film under the same conditions reduced from P ~3.2 to ~3.1. Thus, in warm, humid air, enzymes are active and enzyme/PCL mobility is not the rate limiting factor for catalyzed hydrolysis. Further evidence of chain mobility is provided by the slight left-shift in both ePCL and control film SAXS lamellae structure factor peaks, indicating slight lamellae thickening and lamellae spacing increase, as well as **Figure S1.1**. Thus, insufficient reactive water and/or limited product diffusion caused the nanoconfined catalysis to shut down.

Water is more than a hydrolysis substrate. It is also an important environmental variable and plays additional roles in influencing the kinetics of polyester degradation. We proposed a 3-step kinetic cycle to decipher the roles of water in ePCL degradation (**Fig. 1.2d**). At the molecular level, water must be present near the catalytic site to facilitate a nucleophilic attack of the acyl-enzyme intermediate (E*) and release the PCL chain-end (P). Locally, water modulates the microenvironment for embedded enzymes (E), affecting active site accessibility.[9,13–15] Water may also assist the removal of acidic small molecule hydrolysis products (M).[16] A buildup of M could acidify the enzyme's surroundings, triggering enzyme conformational shift and catalyst deactivation ($R_{in} \rightarrow 0$). The increased concentration of M could also create a depolymerization-polymerization equilibrium ($R_p = R_{dp}$).

There are 3 water sources: external water from the environment, hydration water molecules in RHP and enzyme, and those within the host polymer matrix. Siparsky et al. showed that at 40°C 90% RH, water can diffuse through a ~52% crystalline PCL film with a

permeability coefficient (P_{PCL}) of ~4800 barrer, and that rate of diffusion was not rate limiting for hydrolysis until films reached cm-scale thickness.[17] Similarly, Ghanem et al. reported P_{PCL} ~3430 barrer for a ~72% crystalline film.[18] Schmitt et al. showed that similar polyesters e.g., poly(DL-lactide-co-glycolide) had hydrolysis rates independent of total water content.[19] Water in the matrix was found to be split into 2 populations: non-reactive, bulk water with free rotation that occupies space between chains, and reactive water with hindered rotation that are bound to oxygen of ester carbonyl groups. Our system differs by the inclusion of nanoconfined enzymes, shifting the role of water from attacking P to attacking E*. This difference may allow for free rotation water or waters bound to the RHP and enzyme to be reactive. We estimated that there are more internal water molecules in ePCL than hydrolysable ester bonds (**Methods**). Even if only PCL bound waters are reactive, internal water alone should be sufficient to degrade the polyester matrix if the enzymes are active, thus leaving limited product diffusion as the trigger for nanoconfined catalyst deactivation.

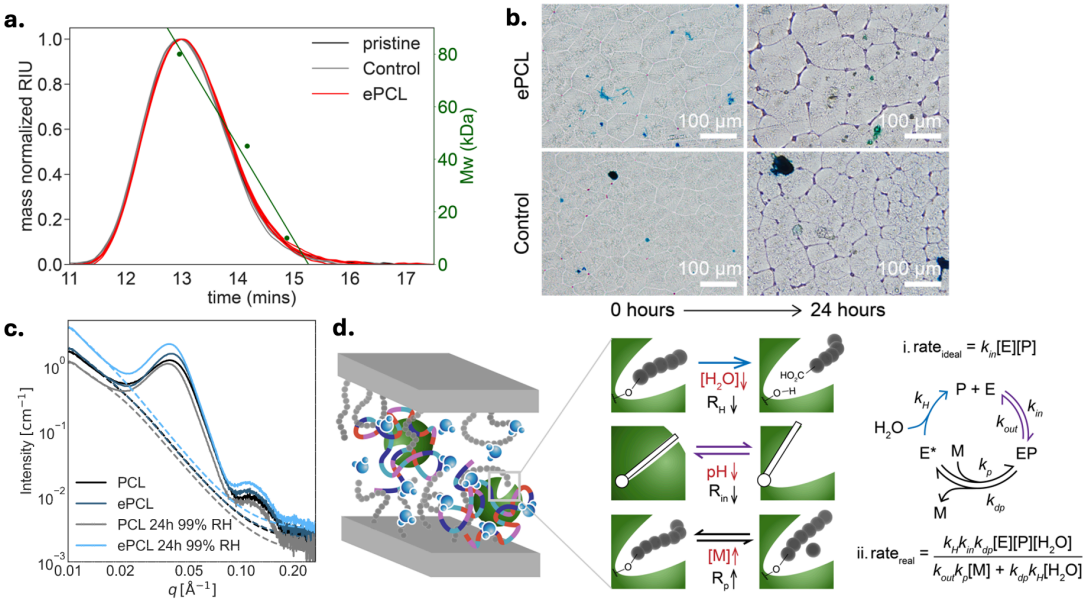


Figure 1.2 | Limited diffusion catalysis shutdown. ePCL and Control (PCL) films pre (0 hour, pristine) and post-24h incubation in 99% RH air. **a,** Mass normalized gel permeation chromatography (GPC) triplicates. The pristine and control curves almost entirely overlap. The calibration curve (green) was used to determine Mw and to calculate chain shortening. **b,** Optical microscopy images of films containing bromocresol green (BG) pH indicator. **c,** SAXS profile (solid lines) with power law fit to low q and extrapolated to high q (dashed lines). **d,** Proposed kinetic cycle for ePCL degradation. Under abundant water conditions, the rate equation simplifies from the "real" rate law (ii) to the ideal rate (i). Schematics show factors contributing to degradation halt in each step of the mechanism under low water conditions, including lack of product release (blue), pH-induced trapping of deactivated conformation (purple), and repolymerization of monomers accumulating in the local enzyme environment (black).

1.3 Water transport modulates catalysis

Kinetics depend on local concentrations of substrates, products, and catalysts. These are influenced by transport, which is dictated by the interior surface chemistry and morphology of the surrounding environment. Unlike solution-based chemistry, solid-state chemical transformations strongly depend on local structure. For catalyst-embedded semi-crystalline polymers, the degradation fronts are buried internally. Water and degradation products must traverse a tortuous path through the amorphous regions to get to and from catalytic sites.[17,20] In ePCL films "acidified" by incubation in warm, humid air or pH 4.0 phosphate buffer solution, a concentration gradient driven diffusion of acidic small molecules out of the PCL matrix was expected but not observed. This diffusion limitation is likely caused by specific interactions between acidic products and local catalytic environment that cannot be overcome in even 99% RH (saturated water vapor). However, acidified films could be "reactivated" by liquid water soaking, in which films completely degraded within 24 hours.

To investigate external liquid water, we used compost soil to modulate surrounding environment water availability. ePCL films were incubated in pH 7.4 phosphate buffer solution (PBS), composts with varied water content and 99% RH air for 72 hours (**Fig. 1.3a**). In the water-rich extreme, PBS, ePCL completely degraded within 6 hours. In 60% wet basis soil, complete degradation was observed within 72 hours. Films in 50% wet basis soil had 44±17% mass remaining after 24 hours and 40±6% mass remaining after 72 hours. Despite significant mass loss, the films maintained structural integrity without forming visible microplastics. For films in 40% wet basis soil, 91±5% remained after 24 hours and no further change was observed at 72 hours. After one week, no further mass loss was observed for samples in 40 and 50% wet basis soils. A rate law for degradation was derived for both liquid water-rich (R_{ideal}) and liquid water-poor (R_{real}) environments (**Fig. 1.2c**). With reduced [H_2O], reduced [E], and increased [M], the degradation shuts down ($R_{real} \rightarrow 0$). Under water-rich conditions ([H_2O] $\rightarrow \infty$, [M] = 0) the rate simplifies to a substrate-diffusion limited reaction that has been previously discussed.[1] To determine if the mechanism causing $R_{real} \rightarrow 0$ is the same for 99% RH and in ≤50% wet soil, GPC analysis was done on ePCL films incubated in 40% and 50% wet soil, as well as in PBS as a water-rich control (**Fig. 1.3b**). The average molecular weight films incubated in ≤50% wet soil decreased by 1-2 kDa, while films incubated in PBS increased by 1-2 kDa**.** The 2-4 kDa PCL molecular weight difference post-degradation in liquid water-poor compared to liquid water-rich conditions could thus be attributed to the premature halt of depolymerization caused by a local accumulation of trapped acidic small molecules. Liquid water modulates the local environment to overcome the entrapping interactions, facilitate product departure and prevent or reverse catalyst poisoning *i.e.,* lid closing. Thus, although surrounding internal water is sufficient for embedded enzymes to "activate" and degrade local polymer chains, external liquid water is needed to remove degradation products through the plastic matrix and provide a "clean" microenvironment for enzymes.

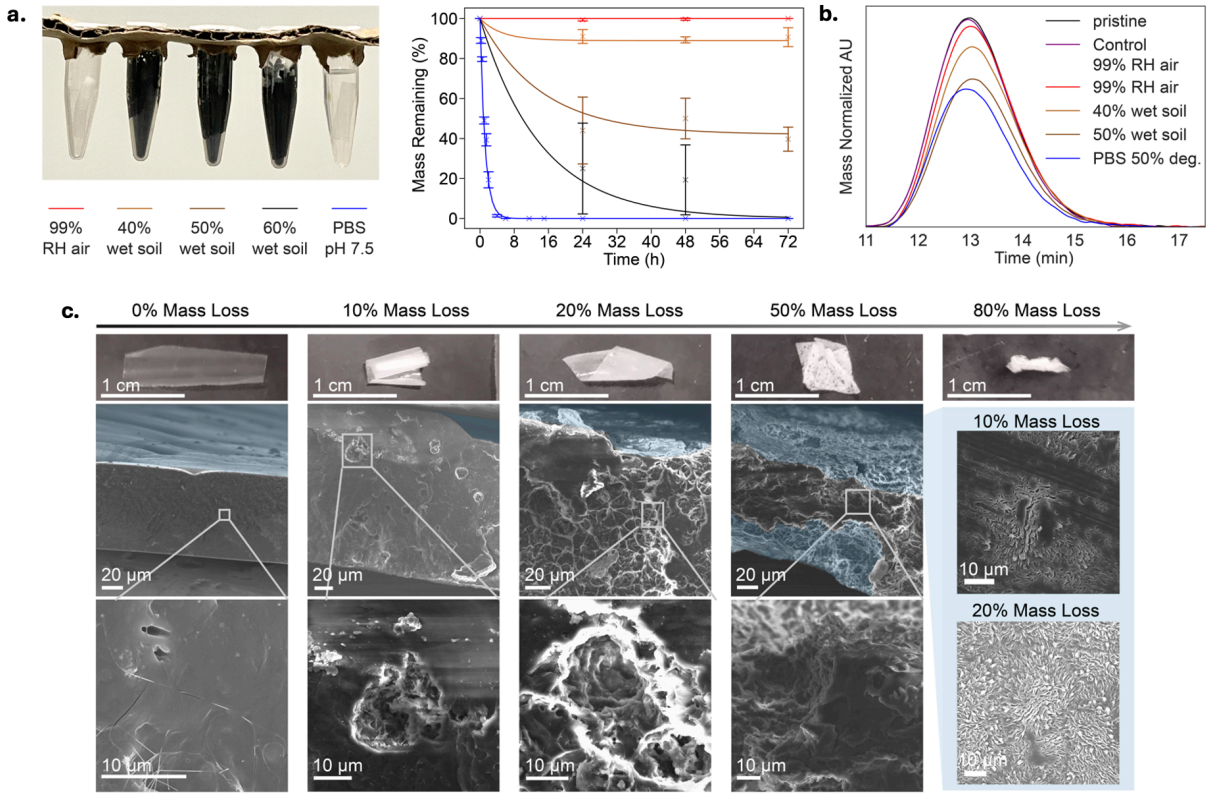


Figure 1.3 | Water transport modulates catalysis. a, Degradation profiles of ePCL films under different environmental water availabilities (n=3). Image shows experimental setup for different levels of external water, prior to soil mixing. **b,** Gel permeation chromatography (GPC) spectra of films pre and post-degradation normalized by initial film mass. Soil and air degradation samples reached degradation plateau and films degraded in PBS are of specified mass loss. **c,** Images showing the progression from a solid film to a sponge-like matrix at varying extents of degradation. The middle row is cross-sectional scanning electron microscopy (SEM) images. Exterior film surfaces are shaded blue. The bottom row SEM images are zoom-ins of features. Exterior surface SEM images of 10% and 20% mass loss films are shown in the blue box inset.

Because ePCL nanoconfined catalysis is water and acidic product diffusion controlled, the derived kinetic rate laws cannot be isolated from matrix morphology. To investigate the structural evolution during enzyme-mediated degradation of PCL, we characterized morphological changes as a function of degradation. ePCL films were incubated in PBS for varied durations to generate films of 0-80% mass loss (**Fig. S1.3**). PCL films were also exposed to enzymes (LIP) in solution (Control) to compare surface versus embedded enzyme-catalyzed degradation. Scanning electron microscopy (SEM) images were taken for the surfaces and cryofractured interior cross-sections of ePCL (**Fig. 1.3c**) and Control (**Fig. S1.4a**) films from 0-50% mass loss. Prior to degradation, ePCL and Control

films had smooth exposed and interior surfaces. As degradation proceeded, the exposed surfaces of both films developed rougher surfaces with dendrite-etched outlines in impinged spherulites of 50-100 μm, indicating amorphous domain preferential surface erosion. This suggests that enzymes at the surface of ePCL degrade by a similar mechanism to enzymes in solution.

ePCL film interior surfaces at 10% degradation exhibited sparse microscopic, spherical pores. No observable dendrite-etching suggested that embedded degradation is agnostic to domain crystallinity. These pores have no visible connection to the film surface. Thus, embedded catalysts indeed degrade PCL from within, and in water-rich environments, water and hydrolysis products can diffuse through the semi-crystalline matrix to and from embedded catalytic sites respectively, through nanoscopic channels with length scales below the resolution of SEM. As degradation continues, 20% mass loss ePCL forms sponge-like networks of impinging microscopic pores with direct-to-surface tunnels in some regions of the film. Microscopic voids imaged in 10% and 20% degraded ePCL consist of coalesced nanoscopic pores and features indicative of nanoscale roughness. By 50% degradation, ePCL films are up to 50% (50 μm) thinner and the delineation between features observed in the cross-section and surface becomes increasingly blurred. Spherulite-sized (50 μm diameter) holes are visible, their surfaces consist of impinging smaller porous spheres. In contrast, Control film interiors remained smooth up to 50% mass loss, revealing that in surface degradation LIP does not diffuse into the semi-crystalline matrix – the reaction fronts exclusively remain on exposed surfaces. Beyond 50% mass loss the mechanical strength of films sharply declined and were too fragile for SEM (**Fig. S1.3**). With 80% mass loss, films lost structural integrity, shriveling up when removed from solution.

Surface and interior SEM images were also taken for ePCL films incubated for 24 hours in 99% RH (**Fig. S1.4b**). These looked identical to images of non-degraded ePCL, suggesting that vacancies left by the removal of ~12 monomeric units do not form voids visible by SEM resolution. This result highlights the gap in what SEM can show us. As polymer-to-monomer degradation proceeds, small molecule removal leads to molecular voids that subsequently grow through the molecular-nanoscopic-microscopic-macroscopic structure hierarchy. Here we showed the growing pores eventually coalesce into percolation networks that facilitate internal transport, shaped by catalyst distribution and the original semi-crystalline matrix morphology. However, the molecular and nanoscale have not been definitively observed. Elucidating the molecular-scale degradation mechanism could provide an answer to the still open question of how liquid water assists in trapped catalysis product removal through the tortuous, hydrophobic semi-crystalline matrix in the initial stages of nanoconfined degradation.

1.4 Molecular mechanism: degradation drives diffusion and vice-versa

Nanoscopic reactions on polymer substrates result in matrix modification – molecular scale rearrangements that lead to reaction modification and create a molecular scale feedback loop. This feedback loop has trickle-up implications through the structural hierarchy and ultimately drives macroscopically observable behaviors, such as the end-fate of nanoscopically embedded enzyme plastic waste in real-world environments. To study mechanisms on the molecular-nanoscopic length scale, small angle x-ray scattering (SAXS) was done on ePCL films from 0 to 50% mass loss (**Fig 1.4a**). The low *q* regime increased in contrast with increasing extent of degradation. The increasing slope suggest the formation of features sized < 60 nm, thus providing evidence for a nano-scale reaction, as was seen in **Figure 1.2c**. Fitting a power law to the low *q* regime gives power law factors (*P*) increasing from 3.2 in non-degraded ePCL to 3.8 in the 50% degraded film. This range corresponds to surface fractal like behavior, suggesting the formation of nanoscopically rough surfaces within the films (nanoporosity) resulting from degradation. The semi-crystalline nature of PCL can be seen from the two structure factor peaks. The smaller, higher *q* shows the lamellae thickness of ~6 nm and the larger, lower *q* peak shows the lamellae spacing of ~10 nm. Analyzing the changes or lack thereof in these peaks can shed insight into the precise mechanism of semi-crystalline plastic degradation. To deconvolute the evolution of the lamellae structures from the effects from nanoscopic surface roughness and porosity, the low *q* power law (Porod) fit was extrapolated to the high *q* of each SAXS profile. Subtracting this fit model from the original data revealed that there was negligible change in peak shape and position up to 35% mass loss, as well as only minor change in peak magnitudes (**Fig. 1.4b**). This is a clear indicator that when up to 35% plastic mass has been removed the lamellae geometry does not change at all. Increasing from 35 to 50% mass loss results in broadening between the two structure factor peaks, indicating the beginning of loss of well-defined crystal structure. SAXS results support tensile test and SEM characterizations that showed 50% mass loss is near the inflection point of structural collapse (**Figs. 1.3c** and **S1.3**).

The first implication of structure factor peak stability over at least the initial 35% of degradation is that both amorphous and crystalline regions must be degrading in relatively balanced proportion. If only the crystalline region was removed the structure factors would disappear or significantly reduce. Conversely, if only the amorphous region was removed, the structure would certainly collapse and the lamellae spacing would reduce. **Figure 1.4c** shows the scattering intensity normalized (linearly scaled) by mass loss. As films lose mass the number of scatterers that can produce the lamellae structure factors equally reduces. Normalized peak magnitudes slightly increased with mass loss, suggesting slightly preferential amorphous region chain-scission. However, SAXS peak intensity increases was shown to be negligible by differential scanning calorimetry (DSC), that showed the bulk crystallinity of ePCL did not increase from 0 to 60% mass loss, remaining at 48±2% (**Fig. S1.5**). The SAXS and DSC profiles further provide nano-scale proof, as was hypothesized in prior works, that processive chain-scission is the dominant mechanism for ePCL plastic degradation.[1] If random chain-scission was the primary mechanism, we would expect to observe majority amorphous region mass loss, as well as reduction in lamellae thickness by local destruction of crystalline regions that would result in a broadening of the lamellae

spacing peak to lower *q* and the lamellae thickness peak to higher *q*. Neither of these are observed. Assuming the Flory switchboard model for polymer crystallization with intramolecular nucleation, an enzyme that processively cleaves PCL unit-by-unit without "releasing" the chain explains the lamellae structure preservation. DSC further showed negligible changes in the percent crystallinity 48±2% of partially degraded ePCL films across incubation environments (**Fig. S1.5**), suggesting that the dominant processive, chain-end scission degradation mechanism is environment-independent.

Single chain removal from lamellae crystals one-by-one while not destroying the structure, could lead to a "loosening" of the lamellae, leading to the formation of molecular-scale channels that serve as shortcuts to transport water and small molecules through, directly leading to and from the enzyme active site that created the tunnel in the first place. Catalytic sites are effectively creating their own transport channels to alleviate the diffusion limitations brought about by the semi-crystalline, hydrophobic matrix. These transport channels will naturally be transient as PCL is well above Tg, the chains will over time reorient and close the gaps. However, the rapid degradation and DSC curves show that in liquid water-rich environments, the rate of catalyzed degradation is much faster than the rate of lamellae thickening and chain-reorganization. Cheng et al. showed water vapor transports faster through molecular-scale hydrophobic pores than liquid water.[21] They attributed it to a lack of intermolecular interactions between water molecules in continuum flow that make them effectively larger than water vapor molecules following a free molecular flow regime. The transient, self-created transport channels may allow for liquid water flow through the hydrophobic matrix with a high enough flux to overpower entrapping specific interactions and carry the acidic degradation products out the same way the water came in.

ePCL degrades over time with a characteristic near-exponential decay (**Figs. 1.3a** and **S1.3**). This could be caused by any combination of active embedded-enzyme concentration reduction (e.g., leaching, deactivation), acidic product accumulation, and decrease in concentration of substrates (water, PCL chain-ends) local to enzymes. Hierarchical porosity can either increase or decrease the rate of substrate and product transport through a scaffold.[22–25] The hydrolysis of polyesters results in a large surface area functionalized with polar carboxylate and alcohol chain ends. Image analysis was done on SEM images of 3 regions of a 20% degraded film interior: a porous region, a non-porous region, and a magnified pore (**Fig. S1.6**). Regions of both high and low porosity likely exist due to uneven enzyme dispersion and impurity-driven inhomogeneous crystallization, leading to inconsistent local enzyme concentrations and matrix crystallinity. The resulting porosity distribution is hierarchical and trimodal, with the main peak at <1 μm, and peaks at 8 and 17 μm with lower magnitudes (**Fig. 1.4d**). Water static contact angles on ePCL films increased from 81.8±4.4 (n=3) when non-degraded to 89.3±5.4 (n=5) when 20% degraded, showing that increased hydrophobicity imparted by structural roughness outweighs the effects of increased surface polar groups (**Fig. 1.4e**). This suggests that internal water transport may be inhibited by increased porosity, coupling slowed degradation kinetics to morphological evolution. However, incubating 0-80% degraded ePCL in 4-nitrophenyl butyrate (4NPB) substrate solution showed that the rate of 4NPB hydrolysis by LIP was higher for porous structured films (15-50%) than non-porous (0%) and structurally collapsed (80%) films (**Fig.**

1.4f). The structural porosity enhanced small molecule transport through ePCL, despite increased surface hydrophobicity and confirmed that active embedded-enzyme concentration reduction is negligible. Thus, in buffered, aqueous environments where acidic products do not accumulate, PCL degradation slowdown must be due to a decrease in local substrate concentration, specifically the number of buried PCL chain-ends proximal to catalytic active sites.

Here we presented a more complete understanding of the entire polymer-to-monomer degradation process across the relevant length scales and time scales (**Fig. 1.1**). For nanoscopically embedded catalysis with the matrix as the substrate, local morphology dictates the immediate surrounding environment. SAXS elucidated the mechanism of processive chain-scission leading to the formation of molecular transport tunnels that subsequently grow through the molecular-nanoscopic-microscopic-macroscopic structure hierarchy. Hierarchical porosity changes the local concentrations of substrates and product surrounding the reaction fronts. The evolution of nanoscale reaction kinetics and morphology are thus intertwined in a molecular-scale feedback loop that cannot be considered in isolation of one another.

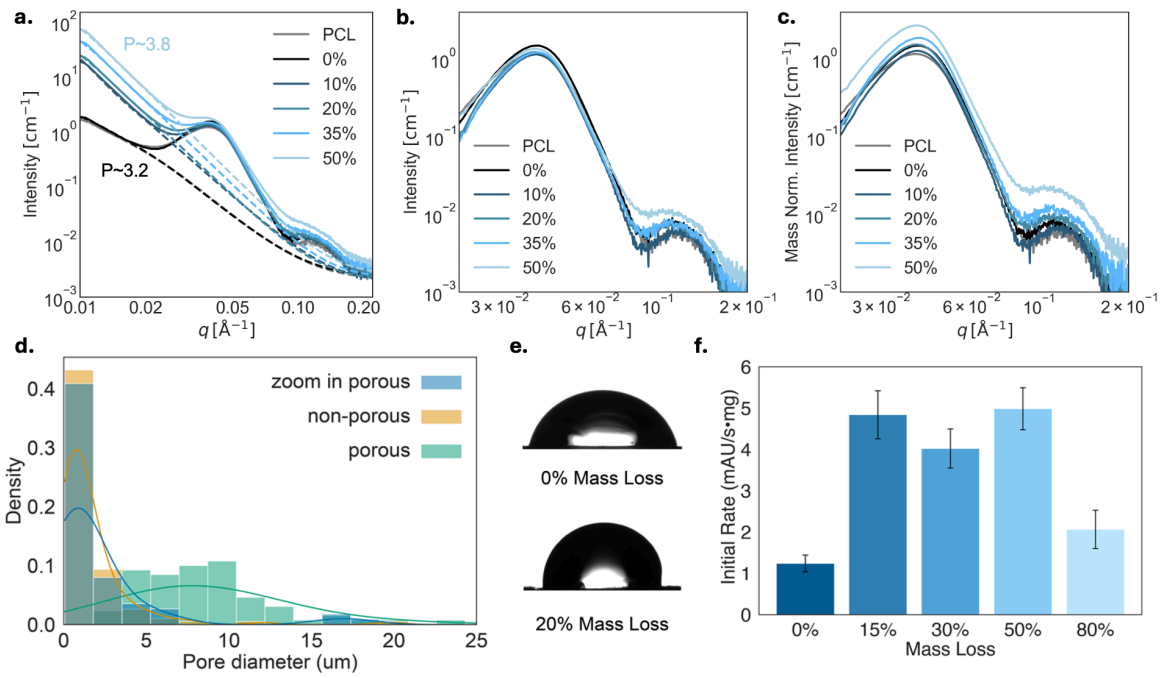


Figure 1.4 | Molecular mechanism. a, SAXS profile (solid lines) with power law fit to low q and extrapolated to high q (dashed lines). **b,** fit subtracted SAXS zoomed in on lamellae structure factor peaks. **c,** (b) mass normalized ($I_{norm} = I/f_{mass\ loss}$). **d,** Image analysis of pore sizes for SEM images of porous and non-porous regions at various magnifications of a 20% mass loss film. **e,** Static contact angle water drop images on 0% and 20% mass loss films. **f,** Initial rate of 4-nitrophenylbutyrate hydrolysis of ePCL films at varying extents of degradation, normalized by film mass.

1.5 Implications and applications: real-world environments and upcycling

To effectively combat the global pollution crisis, basic polymer development must consider real-world conditions.[26–29] Plastics have infiltrated the entire globe from deserts to the tropics to the arctic, and thus during-use and end-of-life could be exposed to the full range of global temperatures from -110°C to 80°C and relative humidity from 1% to 100%.[30–32] When plastics are composted or sent to landfill, they experience variations in pH, UV light exposure, soil water content, and more. The accepted moisture content for effective composting ranges between 40-65%.[33] Landfill cover soils have lower, less controlled moisture levels between 5-60%.[34,35] The results in **Figure 1.3a** showed the degradation profiles of ePCL films incubated in composts with varied water content and simulated humid air storage for 72 hours. The derived R_{real} and R_{ideal} correspond to experimental observations in compost with <60% water and ≥60% water respectively. Thus, water content in compost is the key parameter that determines the extent of plastic degradation and ePCL complete degradation within 3 days requires ~60% soil wet basis, within range for compost and landfill degradation.

Controlled plastic degradation offers a sustainable alternative to synthesize hierarchical porous materials with tunable structure and uniform surface chemistry, that are essential for separations, catalysis, tissue engineering and more.[23,25,24] On the flipside, these large, densely functionalized surface areas can lead to complications and uncertainties surrounding the fate of partially degraded plastic waste. In the real-world, they are surrounded by a plethora of cohabitants that can either adsorb and/or react to the exposed chemical groups. There are abundant microbes, heavy elements, and organic compounds in compost, landfill, and aquatic environments,[36–40] with several emerging examples showing that cohabitants may be capable of reacting with the moieties in increasingly chemically complex modern plastics.[41–43] The outcome of plastic waste management depends on how well we can manipulate the kinetics of plastic chemical transformations against various possible side-reactions. As proof of concept, we upcycled a partially degraded commodity polyester to synthesize catalytically active nanoparticles. In parallel, we also tested the fate of degrading plastic waste in simulated real-world end-of-life environments.

We selected ePCL with 20% mass loss for its optimal balance of porosity, thus high-surface area of exposed carboxylates, and mechanical strength. Upon incubation in Turkevich reaction solution at room temperature for 24 hours, cross-section SEM images confirmed 360±110nm(n=18) gold nanoparticle (AuNP) formation within the film (**Fig. 1.5a**), and macroscopically, the film visually changed from white to purple-metallic and gained 6.5±1.1% mass (**Fig. 1.5b**). The coverage of nanoparticles on the exposed surface is high, suggesting a high density of carboxylates. Visible AuNP adsorption present even in regions surrounded by no observable signs of degradation by SEM resolution suggest that gold cations (~1.5Å atomic radius) are transported through the film through the same nanoscopic channels as water and hydrolysis products, forming AuNP in situ within the film. Solution-formed AuNP of with average diameter of ~10nm would be too large, as schematically depicted to scale in **Figure 1.1.** To test this, partially degraded ePCL was soaked in pre-formed red AuNP from solution. Films turned a lighter pink-metallic and gained a lower

3.6±0.4% mass. The lower mass gain may further support the existence of ePCL in situ AuNP synthesis enabled by gold cation diffusion through nanoscopic water transport channels to regions of the films pre-formed AuNP are unable to reach. AuNP functionalized ePCL (AuNP-ePCL) are catalytically active and effectively degraded oxytetracycline, a ubiquitous water pollutant,[44] at room temperature (**Fig. S1.7**). Under the same conditions, pristine ePCL gained a metallic coating, 1±1% mass, and were only minimally catalytically active against the antibiotics. Once nanoparticles are adsorbed, AuNP-ePCL showed a slower degradation, but fully degraded after a month in 40°C PBS. Upon release, the nanoparticles aggregate and crash out of solution (**Fig. 1.5c**). Thus, AuNP-ePCL decomposed into monomers and gold for easy separation and subsequent recycling.

Real-world environments are full of organics and heavy elements, abundant with upper-bounded order of magnitude concentrations of 0.01 mM in aquatic bodies (oceans, rivers etc.) and 10 mM in landfills.[37–40] Thus, given that our results suggested that partially degraded polyesters may be effective sponges to adsorb real-world nanoparticle pollution, it is conceivable that the demonstrated catalyzed formation of metal nanoparticles and subsequent tetracycline degradation could spontaneously occur. Although heavy element and antibiotic environmental remediation would be beneficial, the formation and release of toxic nanoparticles into the environment upon eventual plastic degradation is a concern. To investigate nanoparticle adsorption and/or formation by degrading plastics in non-aqueous environments such as landfills and composts, we buried partially degraded ePCL in room temperature compost soil made to 60% wet basis with 15 mM $AuCl_3$ and 30 mM citrate. After 24 hours there was no evidence of AuNP adsorption, suggesting that despite higher concentrations, nanoparticle formation is kinetically slow in non-aqueous environments.

To test heavy element ion adsorption, partially degraded ePCL was incubated in room temperature 2mM $AuCl_3$ without the presence of citrates to prevent independent AuNP formation in solution. These films did not visibly change and only gained 1±1% mass. We then incubated pristine ePCL films for 24 hours in 40°C PBS with and without $AuCl_3$. $AuCl_3$ inhibited complete ePCL degradation to only 11±1% mass loss (**Fig. 1.5d**). 9±1% mass loss, 82% of the total degradation over 24 hours, occurred within the first 30 minutes, suggesting that enzymatic degradation of ePCL is initially rapid, but significantly slows once nanoscopic tunnels form allowing for water transport gold cations with it, to bind to both the plastic matrix and perhaps the enzymes (deactivating them), which outcompete ePCL degradation. This showed that partially degraded polyesters can adsorb heavy elements, but requires cohabitants *e.g.,* citrates to drive observable nanoparticle formation in situ. Plastic surface occupation by heavy metal ions and nanoparticles both slowed down the degradation. There is a competition at PCL chain-ends between enzyme-catalyzed hydrolysis and heavy element adsorption. In early degradation, chain-ends are buried near enzymes, so hydrolysis is faster, but adsorption becomes dominant as plastic porosity increases and transport to chain-ends becomes facile. This also further showed that initial degradation is reaction controlled not water diffusion controlled, as gold cations would diffuse in with water. Once the water and gold cations do diffuse in, degradation halts to a plateau.

Importantly, the concentrations of heavy elements in real-world waters are on average 2 orders of magnitude lower than our tested solutions. To test the outcome of this kinetic competition in real waters, ePCL was incubated at 40°C in ocean water collected from the Berkeley Marina and completely degraded within 24 hours. Though not conclusive, these results suggest that the risks for heavy element adsorption inhibiting ePCL degradation in real aquatic bodies, as well as hazardous nanoparticle formation catalyzed by degrading plastics in real-world environments are minimal, provided that plastic degradation is faster than side-reactions such as heavy metal adsorption.

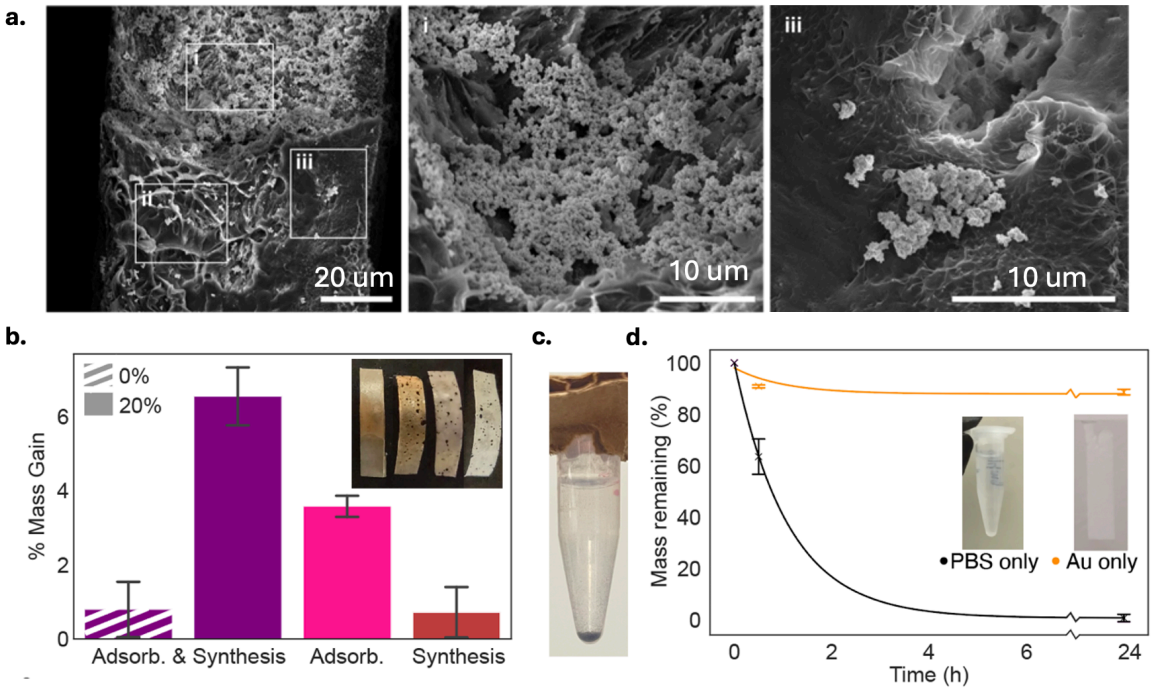


Figure 1.5| Implications and applications. a, Cross-sectional SEM of 20%-AuNP showing adsorption of AuNP on porous surfaces. Right image is a magnified subsection of the left image. **b,** Gains in mass for 0% and 20% mass loss ePCL films after 24 h in gold chloride and sodium citrate solution (Adsorb. & Synthesis), preformed AuNP solution (Adsorb.), and gold chloride solution (Synthesis). Insets show corresponding macroscopic images of films after each treatment from left to right. **c,** Image of fully degraded AuNP functionalized ePCL, AuNP crashed out of PBS buffer solution and aggregated at the bottom of the vial. **d,** Mass loss profile for pristine ePCL incubated at 40 ºC in PBS with and without $AuCl_3$.

1.6 Conclusion

Understanding the transport of substrates e.g., water, acidic small molecules, heavy elements, and nanoparticles is critical for modulating nanoconfined catalysis. Leveraging this, we presented 2 methods to deactivate nanoconfined biocatalysis: trapped product accumulation and heavy element cohabitant adsorption. We demonstrated intentional, controlled transformation of these issues unique to nanoconfined catalysis in a solid-state hydrophobic matrix into economic opportunities. The observed reaction kinetics and end fate of product e.g., plastic waste in real-world environments, are both manifestations of the feedback loop that exists between reaction kinetics and morphological evolution at the molecular scale.

Designing biodegradable plastics is complicated and must be done with real-world considerations in mind. Our results highlight that real-world plastic degradation is in kinetic competition with cohabitant interactions and that the fate of plastics is a matter of speed. Embedded enzymes are an effective strategy to tailor kinetics to match desired outcomes across environments. Opportunities are opened for plastic waste upcycling in controlled settings to highly functionalized materials with tunable hierarchical porosity. Plastics can be designed to match compost or landfill variables e.g., temperature, moisture, and pH, such that hydrolysis outcompetes side-reactions, to achieve complete polymer-to-monomer degradation. Here we focused on enzyme-catalyzed PCL degradation, however, even non-catalyzed plastics can evolve into structural materials with reactive surfaces.[28] Systematic studies of plastic waste interactions with cohabitants across all time- and length-scales are lacking. To avoid inadvertently creating worse pollution in search of sustainability,[8,45] present studies should be expanded upon by environmental specialists to cover a wider range of plastics, cohabitants, and environments. The knowledge presented herein can inform the design of future plastics “kinetically matched” to real-world environments with improved storage/use durability and end-of-life control to prevent unintended reactions with cohabitants, or even to intentionally react with co-pollutants for waste sequestration efforts.

1.7 Methods and materials

Film Preparation: Amano PS Lipase from *Burkholderia cepacia* (LIP) was purchased from Sigma-Aldrich and purified following previously established procedures.[46] Concentration of lipase stock solutions were determined using ultraviolet–visible (UV-vis) absorbance at 280 nm. Random heteropolymer (M_n = 44 kDa; polydispersity, PDI = 1.26) was synthesized as previously reported following a monomer molar composition of 50% methyl methacrylate, 20% ethyl hexyl methacrylate, 25% oligo ethylene glycol methacrylate (M_n = 500 g mol^{-1}), and 5% sulfopropyl methacrylate potassium salt.[11] RHP and purified enzyme solutions were mixed in a mass ratio of 20:1, flash-frozen at -80 ºC, and lyophilized overnight to form the RHP-enzyme complexes.

Polycaprolactone 80 kDa (PCL) was purchased from Sigma Aldrich. PCL was dissolved in toluene at 4 wt% concentration and stirred for at least 4 h to ensure complete dissolution. The dried RHP-LIP complexes were resuspended directly in the polymer solution at 0.04 wt% enzyme concentration (total polymer matrix mass = 99.16 wt%). Mixtures were vortexed for 5 min, cast on a glass slide and air-dried at room temperature to an average film thickness of 20 µm. For thicker films (average 100 µm), air-dried films were redissolved in toluene to 20 wt% PCL concentration. This mixture was then cast, spread using a doctor blade to achieve uniform thickness and air-dried at room temperature.

To ensure uniform starting crystalline lamellae thickness, dry films were melted at 90°C and recrystallized for 27 min during a gradual ramp to 35°C (**Fig. S1.8**). Dry films were then submerged in water and removed from glass slides via capillary peeling.

Film Degradation: Triplicates of ePCL films with an average 3 mg initial mass were degraded in three environment types enclosed within a 40°C water bath: 10 mM pH 7.4 phosphate buffer solution (PBS), 40-60% wet basis compost soil and ≥99% relative humidity air. Compost soil was massed as purchased and oven dried to quantify initial moisture content of 38% wet weight basis. PBS was added and mixed to a homogeneous soil mixture to make samples of approximately 40, 50 and 60% wet basis as per composting standards that on average range between 40-65%. Samples were buried in the compost soils. Humid air samples were placed in empty centrifuge tubes with needle-punctured lids to equilibrate to water bath humidity. Films were removed from degradation environments at designated time points, any residual buffer salts or soil were washed with Milli-Q water or removed by gentle air blowing, then lyophilized overnight, then weighed to determine mass loss.

Film Characterization: Thickness of air-dried films were determined using a digital caliper. Crystallinity of the films were determined using differential scanning calorimetry (DSC), with a temperature ramp from 20 to 75°C at a rate of 1.5 °C min^{-1}. To determine the percent crystallinity, the sample's integrated enthalpy of melting was normalized against the enthalpy of melting for 100% crystalline PCL, 151.7 J g^{-1}.[47]

Quasi-static tensile tests were performed to determine the mechanical properties of films using a Dynamic Mechanical Analyzer 850. Films were cut into 5x10x0.1 mm strips and tested at room temperature with a 0.5 mm min^{-1} ramp rate up to 1.5mm displacement.

Gel permeation chromatography (GPC) was used to monitor molecular weight distribution as a function of degradation. Partially degraded films and degradation product solution were lyophilized overnight and dissolved in tetrahydrofuran (THF). 20 µL of solution was injected into an Agilent PolyPore 7.5 × 300 mm column. GPC spectra were baseline corrected and area-under-the-curve normalized by initial film masses using a custom Python script. A calibration curve of PCL standards was used, and all characterizations were repeated at least in triplicates.

Static water contact angle measurements were obtained on triplicate samples with a Ramé-Hart goniometer. 10 µL of Milli-Q water was dispensed onto films. 100 measurements of both left and right contact angles were taken with 0.005 second delay between them to get a time average.

For SAXS studies, 20-30 µm-thick films were folded 4 times to reach 160-240µm-thick film samples and run at the Stanford Synchrotron Radiation Lightsource (SSRL) SLAC National Accelerator Laboratory beam line 1-5 using 15 keV X-rays, 5 minutes of exposure time, beam nominal spot size of 300x500 µm and a Pilatus 1M detector at 3 m distance. AgBe and glassy carbon were used as calibrants. Air background was subtracted from all processed data. Images were plotted as intensity (I) versus q, where $q = (4\pi/\lambda)\sin\theta$, where λ is the wavelength of the incident X-ray beam and 2θ is the scattering angle. The sector-average profiles of SAXS patterns were extracted using Igor Pro with the Nika package.

Local pH visualization: ePCL embedded with pH indicators were prepared to probe local enzyme activity. Bromocresol green (BG) purchased from Sigma Aldrich was dissolved in Milli-Q water and mixed together with RHP and enzyme in a 1:20:1 BG:RHP:enzyme mass ratio before flash-frozen and lyophilized overnight. To ensure co-localization of enzyme and indicator, enzyme was labeled with NHS-Fluorescein (5/6-carboxyfluorescein succinimidyl ester). Dried complexes were resuspended in PCL/toluene solution and cast as previously described. Films were placed in 99% RH air for 24 hours, massed, and examined with optical microscopy (OM). A U-MWBS3 mirror unit with 460-490 nm excitation wavelength was used to take the fluorescence microscopy images, which were overlaid with brightfield optical microscopy images in ImageJ software (**Supporting Fig. 2d**). Control films containing only BG and RHP were prepared at the same mass ratio (1:20) as above to determine BG sensitivity to pH change in the solid state. 0.2 M pH 4 and pH 8 citrate-phosphate buffer was dropped onto air-dried control films to calibrate expected pH induced color change from green to blue respectively, observed both macroscopically and under OM.

ePCL internal water estimation: The amount of water present in a 5 mg ePCL film stored at ambient conditions was estimated. Water saturated in PCL pellets and toluene were approximated to be 0.012 wt%[17] and 0.059 wt%,[48] respectively. Residual water attached to oligo ethylene glycol methacrylate monomers after the lyophilization process was approximated to be 112 water molecules per RHP based on polymer molar mass and monomer distribution.[49] Using these numbers, water content in 5g ePCL was approximated at 1.06e-4 mol. The number of ester bonds for a film of the above dimension was approximated to be 4.38e-5 bonds based on 80 kDa PCL.

Small molecule assay for solid-state enzyme activity: 4-Nitrophenyl butyrate (4NPB) was used as a small molecule ester substrate to test for enzyme activity. ePCL films at various stages of mass loss were submerged in 4NPB solution (1 mL, 0.5 mM) dissolved in PBS. A sample was taken every 20 min and monitored for absorbance at 410 nm using a Nanodrop 2000 Spectrophotometer over a period of 1.6 h. A regression was performed on the linear region of the enzyme activity plot to determine initial rates. All data points were performed in triplicate and mass normalized.

Cohabitant Interactions: Gold nanoparticles (AuNP) were synthesized following a modified room temperature Turkevich method.[50] Sodium citrate (8 mM) and gold (III) chloride (4 mM) were mixed in a 1:1 volume ratio and titrated to pH 5. The solution was stirred for 24 h at room temperature with partially degraded film in solution. For adsorption only samples, AuNP was first synthesized, then stirred with films for 24 h. Partially degraded films were also buried in room temperature compost soil made to 60% wet basis with 15 mM $AuCl_3$ and 30 mM $Na_3C_6H_5O_7$. To test heavy element adsorption, partially degraded films were incubated at room temperature in 2mM $AuCl_3$ for 24 hours. To explore the kinetic competition between degradation and adsorption, pristine films were incubated in 40°C PBS with and without 2mM $AuCl_3$. In all cases, films were rinsed and lyophilized overnight before masses were measured.

We soaked films in 200uL room temperature solutions of 37.5 μM oxytetracycline and 2.5 μM cumene hydroperoxide in pH 8 phosphate buffer solution: pristine ePCL, 20% degraded ePCL, AuNP synthesis pristine ePCL, AuNP synthesis 20% degraded ePCL, and pre-synthesized AuNP adsorbed 20% degraded ePCL. The films used ranged in mass between 3-5 mg and in thickness between 75-110 um. Films were wrapped around the sides of 96-well plate incubated at 20 C. 150 uL of 50 uM oxytetracyclene and 50 uL of 10 mM cumene hydroperoxide were added to each well. Each well was sampled for 320-400 nm absorbance spectra every 30 minutes for 16 hours to detect the reduction in the oxytetracycline substrate peak at 360 nm over time.

Morphology Imaging: Surface and cross-sectional scanning electron microscopy (SEM) images were performed with a Hitachi S-5000 microscope, except for gold nanoparticle functionalized polymer film imaging, which were performed on a FEI Quanta 3D FEG FIB-SEM. Films with thickness of 100 μm or more were washed with Milli-Q water and lyophilized prior to imaging. To obtain cross-sectional images, films were flash frozen in liquid nitrogen, cryo-fractured and mounted on sample holder with carbon tape. Samples were sputter coated with gold/palladium for 60 s before imaging.

SEM images were analyzed to determine the size of hierarchal porous structures. Visible pores were fit and marked with elliptical shapes, which were then analyzed using the particle analysis function of ImageJ software. Diameters were determined from the measured areas based on the assumption that all pores were circular. Histograms of pore size were normalized by total count of pores in the image.

Chapter 2

Enzyme Degradation Mechanisms On and In Semi-Crystalline Substrates

2.1 Challenges with enzymatic degradation of semi-crystalline polymers

In response to the global plastic waste crisis there has been rapid progress in discovering enzymes with plastic degrading activity. These enzymes originate from across the kingdoms of life, from fungi to bacteria to animals.[1–4] Some enzymes are promiscuous with similar native substrates, while others have recently evolved to adapt to environments with high density of plastic waste such as landfills and plastic recycling facilities.[5–7] The enzyme engineering community has had success using directed evolution and de novo design to significantly improve plastic degrading enzyme catalytic efficiency, in particular focusing on poly(ethylene terephthalate) (PET) as the model substrate.[8,9]

However, most of these studies only focus on amorphous substrates, as an outstanding problem at the interface of molecular biology and material science is that enzymes have a notoriously difficult time degrading crystalline material.[1,10–13] Crystalline chains are more closely packed with reduced chain mobility and accessibility for both enzyme and water, a key co-substrate for enzymatic hydrolysis. Samples are partially degraded with the amorphous regions preferentially removed. This is an issue because over half of the world's plastics are semi-crystalline, including most PET bottles, significantly limiting the applicability of catalyzed plastic biodegradation.[12,14] Concerningly, amorphous region preferential partial degradation can also release into the environment, highly crystalline micro- and nano-plastics that are longer bound by amorphous domains.[15]

The ability to design enzymes that degrade polymer crystalline regions with at least equal preference to amorphous regions is thus critical for both actual commercial utility and mitigation of hazardous micro- and nano-plastic generation.[16,17] Solutions have been proposed that include engineering thermostable enzymes to degrade at high temperatures where polymer chains are more mobile.[8] As well as nanoscopically embedding enzymes within the polymer matrix to bias degradation towards a processive, chain-end scission mechanism that degrades both amorphous and crystalline regions equally.[18–21] **Chapter 1** explored this mechanism in great depth, providing evidence for enzymatic "pulling" of single chains that are threaded through multiple crystalline lamellae. However, the driving force for enzymatic lamellae chain pulling is still inconclusive. We can look to nature for clues, as there are many processive mechanisms that are critical for life.[22,23] The most relevant example is the enzymatic hydrolysis of crystalline cellulose, where endoglucanases break down cellulose chains at random positions and exoglucanases (*i.e.,* cellobiohydrolases) processively cleave off cellobiose from chain ends. The catalytic domains of cellulase 48F (C48F) and cellobiohydrolase I (CBHI), both cellobiohydrolases, have long tunnel-like active sites lined with aromatic residues. They stabilize cellulose chains with hydrophobic and π-π stacking interactions while hydrolysis occurs, with a cleft at the tunnel end for product exit. The secondary structure of the tunnels of C48F and CBHI differ significantly, being an

$(\alpha\alpha)_6$ barrel and a β-sandwich structure respectively.[23,24] The commonality across processive mechanisms in nature seems to be deep, narrow active site geometries that are relatively rigid and populated with chemistries that promotes favorable substrate binding interactions. Although the total sum of intermolecular attraction between neighboring chains in a polymer lamella is very high, Beckham et al showed that for cellulose the decrystallization free energy was 6.7 and 5.4 kcal/mol/cellobiose unit for chains in the middle and edge of microfibrils respectively.[25,26] A greater attractive force provided by a well-designed enzyme binding pocket may thus provide a driving force for individual molecular bonds to locally peel off one at a time and "slide" into the binding pocket.[27,28] Simulations by Knott et al suggested that cellulose chains during processive chain slide in the active site of another cellobiohydrolase, Cel7A, increased in stability by 8 kcal/mol.[29] Compared to cellulose crystals with abundant hydrogen bonding, the decrystallization free energy for a synthetic polymer with only van der Waals interactions will be far lower.[30]

Importantly, DelRe et al showed that the most common product of the processive degradation of poly(caprolactone) was a trimer not a monomer, meaning despite processive translocation some ester bonds were skipped by the active site.[18] Similarly, proteases can be processive but still selectively cleave certain residues only and disaggregases processively translocate without any bond scission or formation at all.[31,32] Thus, a chain-end, processive degradation mechanism may yield undesirably long oligomeric products (single chain nanoparticles). A counterintuitive best-case scenario might also be possible where an enzyme can grab and initiate degradation from any random location along a chain, after which it processively slides and continues to degrade without dissociation. The slowest step is substrate binding, so expanding the pool of substrates from chain ends only to any chain segment could lead to an incredible boost in degradation rate. This might be a programmable feature when designing biocatalysts.

In this work our goal was to identify enzyme features predictive of polyester hydrolysis mechanism and degradation product distribution *e.g.,* microplastics, oligomers, monomers. Informed by examples in nature, our guiding hypothesis was that enzymes with ideal processive degradation mechanisms would have deep, narrow, and rigid, tunnel-like binding pockets lined with hydrophobic surfaces to strongly attract hydrophobic polymer chains. Our results will help guide the direction of future plastic degrading enzyme discovery and evolution towards searching not just for rapid degradation rates on amorphous substrates but to also optimize for semi-crystalline substrates and monomeric product distributions that can be subsequently recycled (repolymerized) or mineralized in compost.[33,34]

2.2 Hydrolase features predict promiscuity for polyester degradation

We chose poly(caprolactone) (PCL) as a model plastic system for investigation as it is semicrystalline polyester that broken down by hydrolysis – a very important degradation pathway for the multi-billion-dollar biodegradable plastics industry.[35] PCL can be easily solvent cast into films, and there is a decently sized library of commercially available, wild type hydrolases with well characterized crystal structures that have been screened for promiscuous activity on PCL substrates.[36–42] At least one enzyme, a lipase from *Burkholderia cepacia* was shown to degrade PCL via a processive mechanism, albeit only when nanoscopically confined.[18,19] Importantly, enzymes that did not degrade PCL have also been reported, providing important negative controls. There has been no systematic reporting of the enzyme degradation mechanism and degradation product molecular weight distributions. Ultimately, 6 hydrolases with reported PCL degradation and 5 negative controls were selected for experimental investigation (**Table S2.1**). 4 additional hydrolases with reported activity on PCL (2 positive, 2 negative) that were not commercially available were computationally analyzed as proof of concept for mechanism prediction purely in silico.

To remove bias in binding pocket residue selection we wrote an algorithm to automate binding pocket residue identification and showed with universal hyperparameters shared across all 14 enzymes, it had great similarity with our manual labeling using PyMOL. The script automated measurement of key enzyme surface and binding pocket features from a crystal structure, either experimentally measured or generated in silico *i.e.,* via AlphaFold2 (AF2) (**Methods and materials**).[43] A caveat on AF2 generated structures is discussed in **Methods and materials**. The key surface and binding pocket chemistry, geometry, and flexibility (secondary structure) features for 15 hydrolases were calculated and visualized on ternary plots (**Fig. 2.1**). Enzymes were classified by literature-reported PCL degrading activities, which should be treated with some skepticism as reports span multiple papers that used varying degradation conditions (*e.g.,* temperature, buffer solution) and substrate PCL (e.g., molecular weight, % crystallinity, aspect ratio).

We found that although the hydrolase surface chemistries were all in a similar range, there was a slight skew towards PCL degrading enzymes having more hydrophobic and less polar surfaces. This might be useful for substrate binding to highly hydrophobic PCL. There is no clear trend with surface charge ratios. More spherical enzymes are seemingly preferred for PCL degradation, most likely because they correlate with deeper binding pockets. Overly large enzymes also appear less amenable to PCL degradation. Flexibility wise, although there are outliers, generally having more α-helices promotes PCL degradation. Binding pocket analysis produces a much broader spread of features across the 15 hydrolases that echo entire surface features but with more nuance. For example, it appears there is a goldilocks zone for PCL degradation of 15-35% polarity and 50-70% hydrophobicity. Deeper pockets that we hypothesized would be essential for processive translocation were also seemingly predictive of enzyme promiscuity towards PCL degrading activity independent of mechanism. Although protease from *Streptomyces griseus* (P-SG) is an outlier having an extremely deep pocket but no PCL degrading activity, it does have reported activity on PLA and is likely precluded from PCL degradation due to its chemistry and secondary structure

features. Astonishingly, a threshold of ~38% of active site residues being part of α-helices was able to linearly separate enzymes with reported PCL degrading activity (>38%) and negative controls. Alternatively, it could also be that random coils are a better predictor, where all the PCL degraders have <~60% random coil pockets, and the 2 exceptions of proteins that have reported non-activity on PCL are proteases with reported PLA degrading activity. To use these features to predict if PCL degrading enzymes operate via a processive or distributive (random chain scission) mechanism, systematic experimental characterization is required.

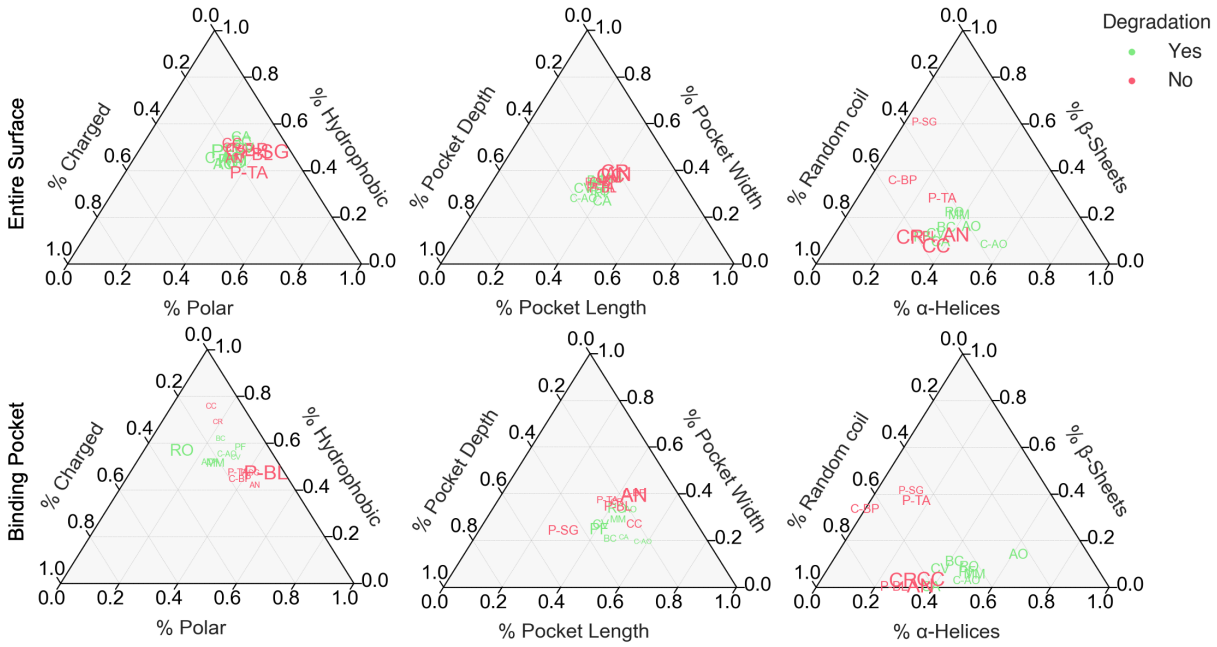


Figure 2.1 | Ternary plots showing key surface and binding pocket chemistry, geometry, and flexibility (secondary structure) features for 15 hydrolases. The size of label markers for chemistry, geometry and flexibility plots represent positive : negative charge ratio, volume, and molecular weight respectively. The % width, % depth and % length were calculated as % depth = depth / (length + width + depth), so a perfectly spherical enzyme or hemispherical binding pocket would have coordinates [0.33,0.33,0.33]. The depth, width and length for entire protein were chosen to match the axes used to define the binding pocket depth, width, and length. Enzymes are colored by literature reported poly(caprolactone) (PCL) degrading activity.

2.3 A medley of mechanisms can degrade semi-crystalline polyesters

Commercially available enzymes were used to enhance the practical applicability of these results. However, the downside to affordable and scalable production is low enzyme purity. Non-protein impurities and target enzyme concentrations were quantified by UV-Vis (**Fig. S2.1**) and SDS-PAGE (**Fig. S2.2**). AO, purified BC, C-BP, P-BL, P-SG and P-TA have low nucleic acid contamination and low protein aggregation, with A320 corrected A260/A280 ~0.6. The rest of the enzymes are far less pure, CR and PF were extremely turbid. SDS-PAGE revealed that 100% of the proteins in BC and CA are the target enzyme, however, their concentrations are extremely low meaning most of the mass is non-protein. We quantified in a prior work that the impurity in BC is mostly diatomaceous earth.[18,44] MM, RO, PF, and AO all have clear bands where target enzyme molecular weights (Mw) are expected, however there are also other strong bands present, in particular in PF, where the darkest band is a higher Mw protein. Both the impurities and the true enzyme concentrations in each commercial protein powder are varied and uncertain. As our goal is not to compare exact degradation rates but rather degradation mechanism, we decided that for PCL degradation rather than keeping the commercial powder mass constant, enzyme solution concentrations should be normalized by activity on an ester-bond containing small molecule substrate, 4-nitrophenyl butyrate (4NPB) (**Fig. S2.3, Table S2.2**). There was no clear correlation between enzyme purity and ester-bond cleaving activity. BC and CR that were characterized as extremely impure by SDS-PAGE and A260/A280 respectively had very high 4NPB hydrolysis rates per unit mass. To test the difference enzyme purity will make on mechanism we purified BC (BCp). The improvement in purity was evident across SDS-PAGE, UV-Vis and 4NPB hydrolysis. Purification was attempted for other enzymes (PF, RO) but became cost prohibitive.

The mass loss of PCL films after 24 hours in 40°C enzyme solutions match predictions from literature on which hydrolases had promiscuity for PCL degradation (**Fig. 2.2a**). Given the importance of substrate binding affinity, enzyme solution concentrations normalized by 4NPB hydrolysis rates unsurprisingly had vastly varying PCL hydrolysis rates (**Table S2.2**). CA had the slowest 4NPB hydrolysis rates but one of the fastest PCL hydrolysis rates, being one of the only 2 enzymes that were able to degrade the PCL film nearly completely within 24 hours (>98%). Several visible microplastics remained in solution. This was not the case for AO, which also fully degraded the film but, unlike CA, generated no human-eye visible microplastics. BC, BCp, and MM all degraded between 70-80% of the original PCL film on average. PF and RO were the slowest enzymes that showed PCL degrading activity, achieving 43±8% and 10±2% mass loss respectively. Varied degradation mechanisms resulted in macroscopically distinguishable features in the post-degraded, remaining films (**Fig. S2.14a**). BC and BCp degraded films retained their shape but got thinner and smaller with some whitening indicative of internal porosity formation. Mild whitening was also observed in PF, as well as the formation of macroscopic holes that were unique to PF. MM degraded to similar mass loss as BC but did not retain its shape, shrinking to half the length of the starting film. Given the much slower degradation rate of RO, we chose not to proceed with characterizing its mechanism any further. All enzymes assumed to be negative controls were confirmed and were also no longer characterized.

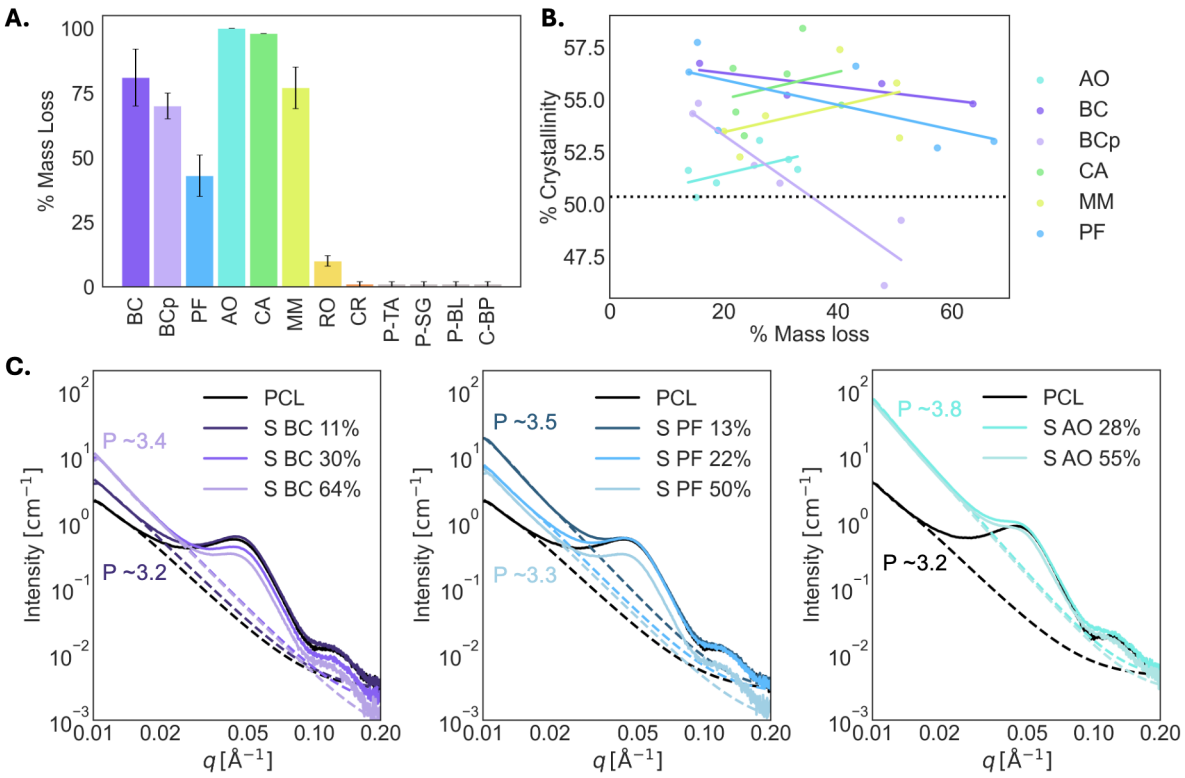


Figure 2.2 | Enzymatic solution (surface) degradation of Poly(caprolactone) (PCL). **A.** Average mass loss of PCL films after 24h in 40°C phosphate buffer solution 1x pH 7.4 containing enzymes of varying concentrations normalized by catalytic activity on an ester-bond containing small molecule (**Fig. S2.3**). Error bars show standard deviation of triplicates. **B.** Differential scanning calorimetry (DSC) measured crystallinity of PCL films at varying % mass loss resulting from the activity of various enzymes. Linear slope fit to scatter data, dashed black line is the starting PCL crystallinity prior to degradation. **C.** Transmission small angle x-ray scattering (SAXS) profiles (solid lines) on PCL films degraded by enzymes (left to right): BC, PF, AO with power law fit to low q and extrapolated to high q (dashed lines).

PCL films were incubated in 40˚C solutions of each of the five remaining enzymes for varying time points between 0.5 to 24 hours to generate films of varying mass loss. These snapshots at different points of degradation were characterized to shed insight into how degradation mechanisms may vary between enzymes. Differential scanning calorimetry (DSC) and transmission small angle x-ray scattering (SAXS) were used to characterize remaining solid films between 0-65% mass loss to probe the initial degradation mechanism. All non-degraded PCL films were 50% crystalline as characterized by DSC (**Fig. 2.2b**). The first ~15% of mass loss by all enzymes except AO increased rapidly by between 3-6%. It could be that the initial degradation was all amorphous region, which would make sense. However, the immediate crystallinity increase is more likely attributable to the effects observed in

Figure S1.1 and **S1.5**, in which despite no mass loss occurring, incubating a pure PCL film at 40°C increased crystallinity by 6% due to chain rearrangement/lamellae thickening. A control experiment is currently being run to conclusively demystify this by leaving films in buffer solution with no enzyme between 0 to 30 minutes and then running DSC on them. From ~18-65% mass loss, the data became a lot clearer as AO, CA, and MM all increased in crystallinity with increasing mass loss, while BC, BCp, and PF decreased in crystallinity with increasing mass loss.

SAXS data suggested that during surface degradation by all five enzymes, most of the lamellae structures are untouched, even up to ~60% mass loss. When PCL is degraded by BC, the power law factor (P) of the slope at low scattering vector (q) increases from 3.2 at 0% to 3.4 at 64% mass loss (**Fig. 2.2c**). This increase is not that steep compared to what was observed for embedded degradation in **Chapter 1**, showing the contrast with air is only minimally changing. There was also a steady reduction in the intensity of the lamellae contrast peaks, which indicated a reduction in crystal structure. The P at 30% and 64% degraded films are identical but the lamellae peak intensities are significantly lower at 64%, which suggests that this is not a thickness loss normalization effect, as that would lead to intensity reduction at all q. In PCL degradation by PF, P similarly increases by ~0.2, but then over time it starts to reduce. The lamellae peak is constant from 13 to 22%, but then, like BC, drops at 50% mass loss. SAXS results for BC and PF matched DSC results, as initially % crystallinity increases before then slowly reducing over time. For PCL degraded by BC and PF, the variance in P across % mass loss snapshots sampled indicate shifting nanoscopic features that result in varying degrees of contrast between PCL and air. This coupled with the non-conventional ability to preferentially degrade crystalline regions suggested that enzymes may be processively degrading PCL chains, and the unthreading of chains from crystalline lamellae resulted in formation of nanoscopic pores from the surface. Indeed, macroscopic pores were visible in both films, in particular PF (**Fig. S2.14a**).

In comparison, PCL films degraded by AO, CA, and MM had a sharper increase in P of ~0.3-0.6 by as early as 9% mass loss, and unlike BC and PF, up to 61% mass loss (except for one outlier) P did not change any further (**Figs. 2.2c** and **S2.5a**). The lamellae contrast peaks of all AO, CA, and MM degraded films (except for one outlier) also negligibly deviate from those of non-degraded pure PCL. As significant mass loss necessarily results in significant reduction in the number of scatterers, this matches DSC results where PCL degraded by AO, CA, and MM increased in crystallinity with mass loss. AO, CA, and MM degraded PCL, immediately reach a steady state of nanoscopic roughness (*i.e.,* surface fractal patterning, contrast with air). They do not appear to alter the lamellae structure of PCL at all, even up to 60% mass loss. This can be understood if degradation mechanism from the surface is distributive, etching nanoscopic surface fractal features that randomly chip away and thin the film over time. Without adding any additional roughness internally, the same surface roughness just migrates down the film thickness, or as observed in MM macroscopically, down the film length (**Fig. S2.14a**).

Despite this characterized distinction between “processive” enzymes (BC, PF) and “distributive” enzymes (AO, CA, MM), both groups can seemingly degrade amorphous and

crystalline domains. DSC showed that 50% crystalline PCL films could lose up to 50% of its mass while at most increasing in crystallinity by 7%. The caveat to this is that the PCL film spherulite sizes were on the sub-micron/nanoscale (**Fig. S2.4a**). The distributive mechanism enzymes may only be degrading amorphous regions around crystalline spherulites that then detach and release into solution as nanoplastics. Measuring the mass loss of remaining films by mass balance would not capture this. Nor would these nanoplastics be optically visible, captured by DSC, or SAXS that were done on the remaining intact films only. However, preliminary have shown that AO, CA, MM, and BC are all able to degrade thermally annealed PCL with average spherulite diameter of 50-100 microns by a similar mechanism as non-annealed PCL (**Fig. S2.5b**).

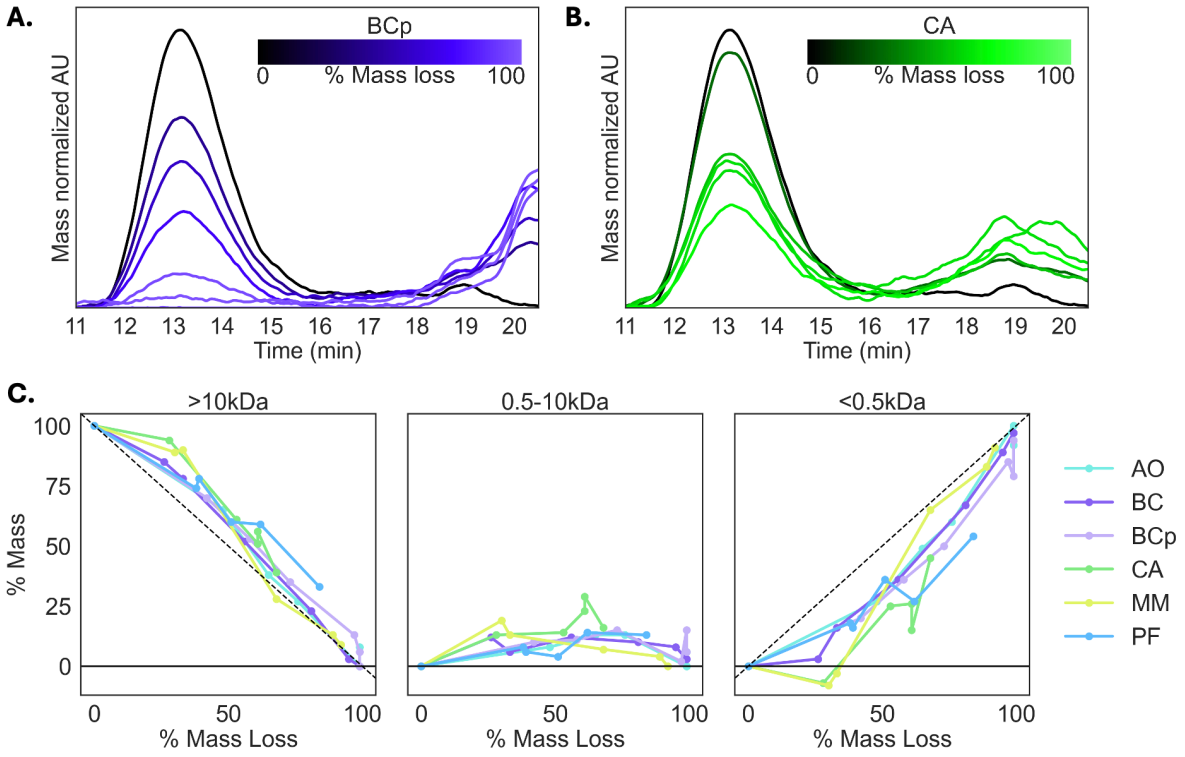


Figure 2.3 | Enzymatic solution (surface) degradation product molecular weight distribution of poly(caprolactone) (PCL), measured by gel permeation chromatography (GPC). GPC spectra mass normalized by initial mass of films pre-degradation for **A.** BCp, closest to processive mechanism and **B.** CA, closest to distributive mechanism. **C.** Evolution of PCL mass distribution with increasing mass loss split into 3 categories: long chains, (>10 kDa) short chains and oligomers (0.5-10 kDa), small molecules (<500 Da, 4-mer or less), as measured by integrating area-under-curve (AUC).

Gel permeation chromatography (GPC) was used to characterize the full molecular weight distribution of degradation products. Both the leftover film and products in solution were lyophilized and redissolved in organic solvent to be run as a single sample. Although nano- and micro- plastics in solution are not explicitly quantified, their existence can be indirectly quantified as shown in **Figure 2.3**. **Figures 2.3a** and **2.3b** respectively show the molecular weight distributions of PCL degraded by BCp and CA with increasing mass loss, normalized by initial film mass prior to degradation. This normalization is important as it allows for quantitative integration of the area under the curve (AUC). GPC spectra for BC, AO, PF and MM are shown in **Figure S2.6.** For each enzyme, separate integrations were done for 11-15 minutes and 15-20.5 minutes that based on a calibration curve correspond to chains >10 kDa and between 0.5 – 10 kDa respectively. 0.5 kDa was the lower bound detection limit of our GPC. These AUCs divided by the AUC from 11-20.5 minutes for 0% degraded PCL yield the % long chain, % short chain and oligomers, and by subtraction, also yields the "off the plot" distribution corresponding to % small molecule products that are 4-mers or smaller (**Fig. 2.3c**). The results show that despite the different mechanisms suggested by DSC and SAXS, the resulting product distributions are more similar than they are different. All enzymes deviate from a perfect processive mechanism with 100% small molecule products. From 0 to 50% there is more remaining mass > 0.5 kDa measured by GPC than remaining mass intact on the main film, confirming as expected that plastic pieces are falling off into solution. However, as degradation continues, at least for BC, BCp, AO and MM, the macro-, micro- and/or nano-plastics that chip off can continue degrading as at 100% film mass loss, there is near zero deviation from the ideal. AO-degraded PCL films that are 100% degraded have 95±4% (n=3) conversion to 4-mers or smaller, 100% degraded PCL by BC and BCp are 97% (n=1) and 94% (n=1) ≤4-mers, and 92% degraded PCL by MM is 91% (n=1) ≤4-mers. BCp and CA are specifically compared in **Figure 2.3** because they represent one of the most ideal and one of the least ideal degradation mechanisms. The distributions of BCp, BC, and AO are nearly identical, where the main peak reduces, the primary products are ≤4-mers and the only oligomers that remain have elution peak at 20.5 mins (**Fig. S2.6**). In comparison, the main CA peak displayed right-tail broadening as it decreased with products formed at all molecular weights with a distinct peak at ~19 min elution – a larger main oligomeric product length. PCL degraded by MM behaved similarly to CA with early stage right-tail broadening and primary oligomer peak at 19 mins elution. PCL degraded by PF reduced without peak broadening like BC and AO, however, the extent of decrease of the GPC peak is less than the mass loss, implying long chains exist as chipped away smaller plastics in solution. The main oligomer product peak of PF is also more like those of MM and CA. There is likely a link between the worse product distribution of PF and its slower PCL degradation rate compared to the other 4 enzymes.

Our GPC data did not follow the neat binary mechanism groupings suggested by DSC and SAXS. From it, we first conclude that processive chain-pulling is not a prerequisite for polymer-to-≤4-mer degradation *i.e.,* microplastic elimination, at least for these 5 hydrolases on PCL substrate. Ongoing SAXS experiments characterizing the degradation of thermally annealed PCL with spherulite diameters of 50-100 microns may change this conclusion and add further mechanistic insight. We can also conclude that the purified and non-purified BC

had no mechanistic difference. More data is needed to make a general claim that impurities do not affect enzymatic polymer degradation mechanism.

To investigate if PCL degradation mechanism can be predicted by enzyme features, we did a simplified grouping of BC and PF as having processive and AO, CA, MM having distributive mechanisms (**Fig. 2.4**). No entire surface enzyme feature correlated with mechanism. Within the PCL-degrading binding pocket chemistry goldilocks zone (50-70% hydrophobicity, 15-35% polarity), there may be an even smaller goldilocks zone for processive mechanisms: 58-62% hydrophobicity. BC and PF also have deeper pockets (>30%) than AO, CA, and MM. These results match our nature-inspired hypothesis that a processive mechanism would require active sites with deeper binding pockets lined with hydrophobic surfaces to strongly attract hydrophobic polymer chains. CV and C-AO are enzymes with demonstrated PCL degradation activity from literature. Using our results to predict their degradation mechanism suggests that CV is very likely to degrade by a processively as it has similar binding pocket chemistry and geometry to BC and PF. Interestingly, comparing the origins of all the PCL degrading enzymes, only the processive and potentially processive BC, PF, and CV originate from bacteria, with all distributive mechanism enzymes originating from fungi (**Fig. S2.7a**). C-AO shares similar chemistry to the processive enzymes, however its binding pocket is shallower, so prediction is difficult.

However, the differences in the degradation rate and product distributions of BC and PF are significant. As is the outstanding speed and % ≤4-mer product of AO despite its lack of processive mechanism as characterized by DSC and SAXS. Ultimately, this suggests that there is a medley of possible mechanisms that can successfully degrade crystalline lamellae into small molecule products. More data is needed to define these more rigorously beyond just our trinary classification of processive, distributive and does not degrade. Despite that, the insights from the work herein can be used to accelerate plastic degrading enzyme discovery and evolution by screening sequences (and thus crystal structures) for promising hits. Predicting if an enzyme will demonstrate promiscuity for PCL can be done with high confidence, overcoming the first practical bottleneck completely in silico. Our methods of featurization and characterization can be broadly applied towards other plastics such as poly(L-lactic acid) (PLLA). P-TA and P-SG that were negative controls for PCL were reported PLLA degraders. These enzymes also have rigid pockets low in % random coil. However, rather than α-helices, β-sheets are preferred. They serve as a starting point for a similar analysis where a different set of design rules can similarly be extracted.

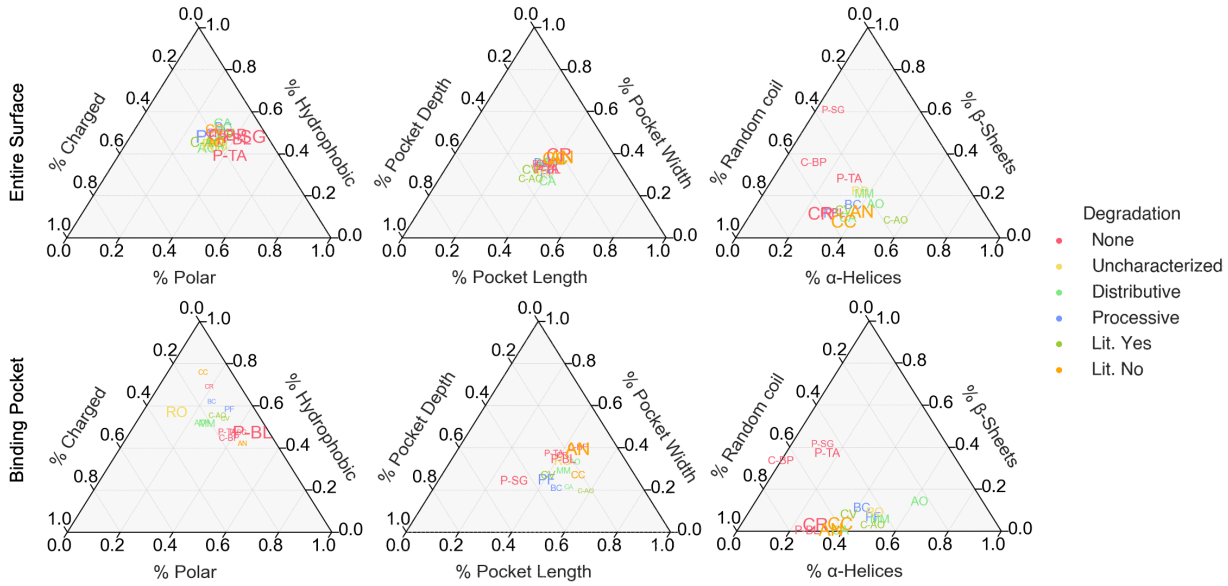


Figure 2.4 | Ternary plots to predict mechanism, showing key surface and binding pocket chemistry, geometry, and flexibility (secondary structure) features for 15 hydrolases. The size of label markers for chemistry, geometry and flexibility plots represent positive : negative charge ratio, volume, and molecular weight respectively. The % width, % depth and % length were calculated as % depth = depth / (length + width + depth), so a perfectly spherical enzyme or hemispherical binding pocket would have coordinates [0.33,0.33,0.33]. The depth, width and length for entire protein were chosen to match the axes used to define the binding pocket depth, width, and length. Enzymes are colored by experimental results reported herein on poly(caprolactone) (PCL) degradation mechanism.

2.4 Active site features and mechanism predict random heteropolymer interactions

DelRe et al showed that the degradation mechanism of BC shifted when nanoscopically embedded within a polymer matrix to achieve processivity.[18] Here we showed that even BC degrading PCL from the surface in solution operates by a processive mechanism. There is a gap in understanding how exactly enzyme mechanism may be changed by confinement in the host plastic matrix. **Chapter 1** explored how the processive mechanism of BC, when nanoscopically embedded with random heteropolymer (RHP) in PCL, was critical to establishing a molecular feedback loop between degradation reaction kinetics and transport to and from embedded enzymes. Here we aimed to delineate the differences in polymer chain translocation, scission mechanisms, and degradation product distributions between enzymes embedded in solution, those confined within the plastic matrix, and those nanoscopically dispersed via RHPs (nano-confined). Several prior works have embedded aggregated enzymes within polymer matrices with demonstrated degradation activity.[15,38,45,46] We want to fully demystify if nanoscopic dispersion just results in faster rate due to higher degradation surface area, or if it fundamentally changes mechanism compared to the same enzyme acting from solution or embedded without nanoscopic dispersion.

Six enzymes from the previous section that were shown to have PCL substrate compatibility and characterized surface degradation mechanism were embedded in PCL with 1wt% RHP. BCp was embedded at 0.05wt% loading while all commercial enzymes were embedded with 0.5wt% loading to have sufficient enzyme, accounting for significant impurities. **Figure 2.5a** shows the mass loss of enzyme-RHP embedded films after 1 day and 1 week incubation in 40°C PBS. BCp degraded faster than it did in solution with the entire film degrading within 24 hours and no visible microplastics remaining after 1 week, as was reported in prior works.[18,19,44] Embedded enzyme masses were not normalized by exact active enzyme mass (impossible due to impurities) nor by activity on 4NPB hydrolysis (**Fig. S2.3**). 4NPB hydrolysis normalized concentration ratios were not used for embedded films as the % loading of CA and RO, and to a lesser extent PF, MM, and BC, that would be required to match the 4NPB hydrolysis activity of BCp and AO would be unreasonably high. This would certainly affect matrix morphology and act as a confounding factor in analysis. Differential scanning calorimetry (DSC) confirmed that embedding enzyme-RHP with unquantified and varied impurities at <1wt% into films negligibly impacted the total crystalline lamellae fraction of embedded films (**Fig. S2.8c**). Polarized optical microscopy confirmed that the microscopic distribution of crystallinity *i.e.,* spherulite diameter distribution was also unchanged with the addition of <1wt% enzyme-RHP (**Figs. S2.4b, 2.8a,** and **2.8b**).

Although the mass losses shown in **Figure 2.5a** cannot generally be compared across enzymes, 24-hour PCL hydrolysis rates normalized by enzyme mass were calculated for comparison between surface and embedded enzymes (**Table S2.2**). BCp, BC, and PF all had higher PCL hydrolysis rates embedded compared to in solution while AO, CA, and MM had much lower hydrolysis rates embedded compared to in solution. This difference could be because of the effect confinement has on enzymes of different chain scission mechanisms. BC and PF were shown to degrade PCL by a processive mechanism while AO, CA, and MM

operated by a distributive mechanism. However, an additional variable introduced when embedding enzymes is the RHP chaperone. The role of the RHP was to stabilize enzymes during organic solvent processing and to nanoscopically disperse enzymes within the PCL matrix.[18,47] Prior works have also shown that RHP composition optimization is necessary for compatibility with certain enzymes, although no concrete design rules have been provided.[48,49] To differentiate activity loss due to confinement versus incompatible RHP binding, 4NPB hydrolysis assays were done using enzyme solutions lyophilized with and without RHP. **Figure 2.5b** shows that AO, CA, MM, and RO had significant reductions in activity when lyophilized with RHP. These were the exact enzymes that had a reduction in embedded activity, suggesting that the reduction in PCL mass was caused by RHP blocking the enzyme active site. It is reasonable that the same features that make an enzyme distributive (shallower depth, smaller volume, less hydrophobic and more charged), would also make pockets more easily blocked by RHP (**Fig. S2.7b**). Thus, the same features that predict processivity can also predict RHP binding compatibility. The functional reversal of this is that processivity can be predicted by a simple RHP binding proxy assay.

AO is one of the most promising enzymes due to its complete degradation to mostly small molecule products in 0.75wt% enzyme solution in under 24 hours. However, AO suffered the largest activity decrease when bound to RHP (**Fig. 2.5a** and **2.5b**). We hypothesized that this was due to the distributive mechanism enzymes' binding pockets being more charged. In particular, the AO active site is highly positively charged and shallow and can thus interact very strongly with negatively charged RHP. We tested AO binding with a more optimized RHP design with no charge, to interact with the active site more transiently. This yielded significant improvement with activity increasing from 29.3±6.0% to 83.8±0.7% (**Fig. 2.5c**). However, embedded AO with no RHP degraded even faster, leading to 100% degradation within 24 hours. Embedding CA and MM with no RHP also yielded significant increases in PCL degradation within 24 hours from 5.0±1.9% and 5.0±1.5% to 33.3±1.9% and 74.8±2.4% respectively. Although many enzymes require RHP protectants to be processed in organic solvents and plastics, these lipases are highly resistant to misfolding in organic environments.[50,51] We demonstrated the practicality of AO embedded PCL films by making a 6.35 x 2.75 cm PCL bag embedded with 0.5wt% AO lipase and no RHP. Although white speckling is visible indicating enzyme aggregation within the film and lack of nanoscopic dispersion, the film completely degraded in 24 hours when incubated in 750mL of 40°C PBS (**Fig. 2.5d**). Thus, RHP is necessary to nanoscopically disperse enzymes within the solid matrix to minimize aggregation and maintain optical clarity, here we demonstrated that for some enzymes like AO, the RHP is not necessary for rapid plastic degradation catalyzed by < 0.5wt% embedded enzyme.

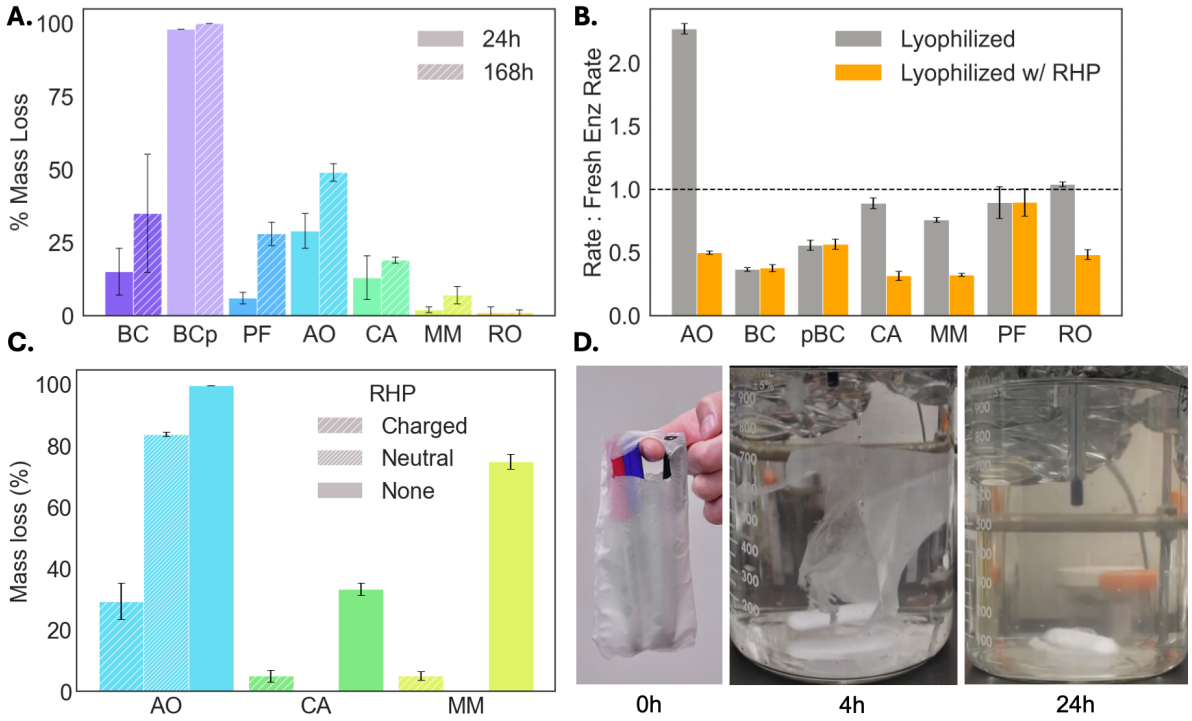


Figure 2.5 | Degradation of poly(caprolactone) (PCL) by embedded enzymes. All bar plots are averages of triplicates and error bars show standard deviation. The standard random heteropolymer (RHP) is a 4-monomer (M_n = 32 kDa; polydispersity, PDI = 1.44) of 50% methyl methacrylate, 20% ethyl hexyl methacrylate, 25% oligo ethylene glycol methacrylate (M_n = 500 g mol^{-1}), and 5% sulfopropyl methacrylate. **A.** Average mass loss of PCL films with embedded enzymes (0.5wt%, except BCp at 0.05wt%) with 1wt% random heteropolymer (RHP) after 24h and 168h in 40°C phosphate buffer solution 1x pH 7.4 (PBS). **B.** 4NPB hydrolysis rate of enzymes lyophilized with and without RHP. All rates are normalized against the rates of enzymes without RHP prior to lyophilization (black dotted line). **C.** Average mass loss of PCL films with embedded enzymes (0.5wt%, except BCp at 0.05wt%) after 24h and 168h in 40°C phosphate buffer solution 1x pH 7.4. Charged refers to the standard RHP, Neutral refers to a 3-monomer RHP without sulfopropyl methacrylate. **D.** 6.35 x 2.75cm PCL bag with 0.5wt% AO degraded in PBS with images taken at 4h and 24h (video of degradation: www.youtube.com/watch?v=aqZHU56oxbc).

2.5 Enzyme confinement and nano-confinement shift degradation mechanism

Embedded AO was able to completely degrade PCL films to products invisible to the human eye, however, its mechanism remains a mystery. AO in solution degraded with a different pathway from the well characterized BC. DSC and SAXS showed AO chain scission in solution operated with a distributive mechanism with preferential amorphous region degradation and no film nano-porosity introduced (**Fig. 2.2**). Yet GPC showed that the products of PCL degradation by AO were mostly small molecules < 500 Da, nearly identical to the product molecular weight distribution generated by the processive enzyme BC (**Figs. 2.3** and **S2.6**). We characterized embedded enzyme degraded films by DSC, SAXS, and GPC to investigate the effect that enzyme confinement within the solid-state substrate matrix has on degradation mechanism. To understand the role that RHP plays, if any, we characterized PCL degraded by embedded BC and AO both with and without RHP. This offered a comparison between the effects RHP may have on enzymes with different surface degradation mechanism but that both still yield complete, rapid PCL degradation. The embedded mechanisms of CA and MM, both without RHP, and PF with RHP were also characterized to identify predictive relationships between enzyme binding pocket features and mechanism shift from surface to confined degradation.

Variations in embedded degradation mechanism resulted in macroscopically observable features in remaining films despite uniform starting film appearance. All films whiten, indicating internal porosity formation (**Fig. S2.14b**). PCL films significantly degraded by embedded BC-RHP retain original shape in solution but shrivel once removed, due to loss of structural rigidity. This was indicative of a high degree of porosity. Embedded PF degradation uniquely generates abundant macroscopic pores on PCL films. All pre-degradation embedded films were ~50% crystalline (**Fig. S2.8c**). As with surface degradation, crystallinity has an initial jump in crystallinity of in 3-6% within the first ~15% of mass loss, after which up to 65% mass loss. PCL degraded by embedded AO and CA increased in crystallinity with increasing mass loss and PCL degraded by embedded BC-RHP and PF-RHP decreased in crystallinity with increasing mass loss (**Fig. 2.6a**). These trends match those observed for PCL degradation by enzyme solution (**Fig. 2.2b**). However, BC without RHP, AO-RHP, and MM, all demonstrated a mechanism shift upon embedding to having no preference for amorphous nor crystalline region degradation.

SAXS revealed that when PCL is degraded by embedded BC-RHP, the power law factor (P) of the slope at low scattering vector (q) increased from 3.2 at 0% to 3.9 at 47% mass loss (**Fig. 2.6b**). A steady broadening and intensity reduction of lamellae contrast peaks with increasing mass loss also occurred. These results suggested nanoporosity formation increasing polymer-air contrast at low q and a steady decrease in film crystallinity respectively. Together, they suggest a processive mechanism that was similarly observed in the SAXS profiles of thermally annealed PCL degraded by BC-RHP (**Fig. 1.4a**). PCL degraded by BC solution had significantly less increase in P with PCL mass loss and the SAXS intensity across all q decreased with increasing mass loss (**Fig. 2.2c**). This indicated that nanoscopic dispersion of BC-RHP greatly increased the surface area of matrix confined catalytic reaction fronts resulting in nanoscopic porosity formation rather than film thinning from the

surface. The SAXS profiles of PCL degraded by embedded BC without RHP is more similar to PCL degradation by BC solution than embedded BC-RHP (**Fig. 2.6c**). P only increased from 3.2 to 3.4 between 0 to 49% PCL mass loss, and the decreasing intensity across all *q* indicates film thinning. The main unique characteristic of PCL degradation by BC was that in the initial 9% mass loss PCL significantly lost crystalline structure with the two lamellae contrast peaks broadening significantly and merging into a single broad hump, before regaining clear lamellae structure factor by 20% PCL mass loss. The combination of these features suggested that BC without RHP initially destroyed its surrounding crystal structure, but between 10-20% mass loss, sufficiently large (>5 nm) transport channels were created allowing embedded enzymes to leach into solution and proceed to degrade by surface mechanism resulting in film thinning. The sharp peaks that appear at q~0.11$Å^{-1}$ superimposed on top of lamellae thickness contrast peaks across PCL samples degraded by various embedded enzyme such as BC-RHP could be indicative of crystalline rearrangement and lamellae thickening in competition with internal degradation.

SAXS data was collected on PCL films degraded by embedded 0.05 wt% AO + 1 wt% RHP, 0.5 wt% AO + 1 wt% RHP, and AO with no RHP (**Figs. 2.6d,e**,**f**). SAXS profiles for all three embedded systems had minimally deviating P from 10-53% mass loss. The higher RHP:AO ratio of 20:1 resulted in an increase in P of ~1.0, in comparison to a lower increase in P of ~0.4-0.5 that was observed when the RHP:enzyme ratio was 2:1 and when no RHP was used at all. This was similarly observed in PCL degraded by AO solution (**Fig. 2.2c**). In AO solution degradation, intensity at all *q* slightly decreases from 0 to 61% mass loss due to film thinning. This is also observed in PCL degraded by embedded AO. However, for AO-RHP degraded films, intensity at low *q* slightly increases with % mass loss indicating the formation of some internal nanoscopic porosity. Comparing the lamellae structure factors of PCL degraded by AO and BC, AO destroyed the lamellae structure more significantly than BC within the first 10% of PCL degradation. BC retained lamellae contrast peaks with higher intensity until very high % mass loss. Lamellae recrystallization was observed in the PCL sample degraded by AO-RHP to 53% mass loss. Lamellae spacing shifted to lower *q* increasing from ~13 nm to ~15 nm and a new smaller, sharper lamellae thickness peak appearing, still of ~5-6 nm average thickness. Taken together, DSC and SAXS data suggested that the dominant mechanism of PCL degradation by embedded AO with and without RHP is still distributive chain scission that destroyed lamellae structures local to embedded enzymes. Additionally, a sufficiently high ratio of RHP:AO resulted in an improvement in nanodispersion, forming a larger number of initial pores that are still formed by a distributive chain-scission.

After the formation of sufficiently large transport channels within PCL, embedded AO and BC without RHP may be leaching out and degrading from the surface. Embedding enzymes with RHP incentivizes the formation of porosity within PCL films by mediating the enzyme-matrix interface. Despite different PCL morphological evolutions resulting from BC and AO embedded degradation, for both enzymes, the RHP serves as a dynamic/transient adhesive/surfactant to prevent enzyme leaching from the matrix surface. It is unclear how the molecular feedback loop that governs small molecule transport and matrix morphology evolution that was elucidated in **Chapter 1** for PCL embedded with BC-RHP, differs for PCL embedded with AO-RHP. If AO does not processively degrade PCL chains, then an alternate

mechanism for the formation of nanoscopic water and acidic product transport channels must exist to achieve rapid initial degradation kinetics. Perhaps degradation starts from AO embedded on PCL film surface and the resulting cavities they form allow transport to and from enzymes embedded deeper in the interior.

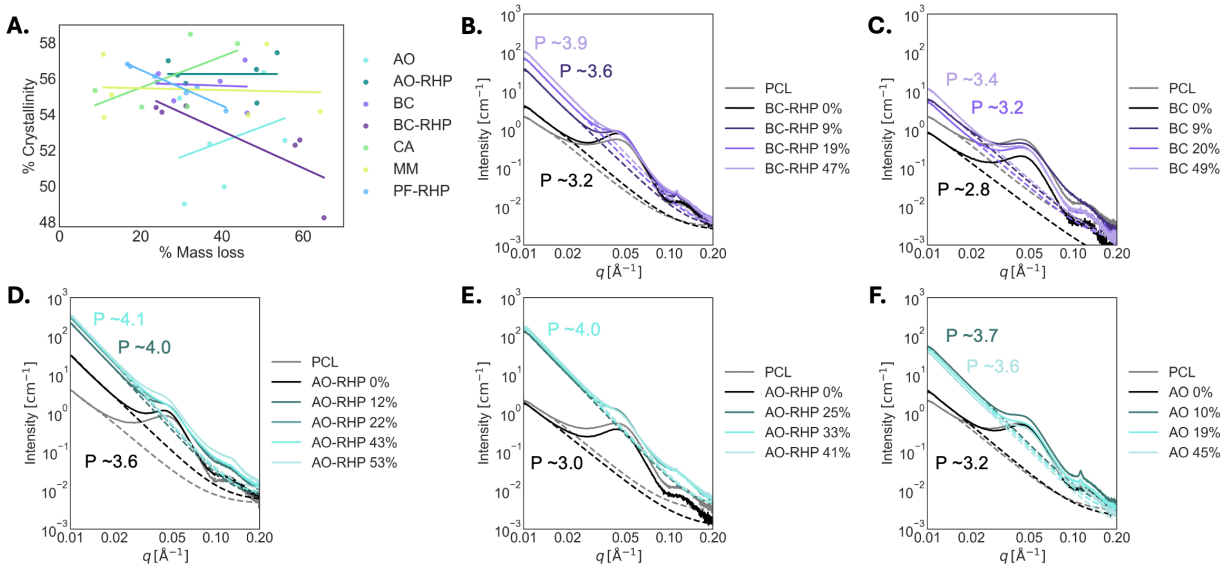


Figure 2.6 | **Embedded enzyme degradation** of Poly(caprolactone) (PCL) by enzymes with and without 1 wt% random heteropolymer (RHP). **A.** Differential scanning calorimetry (DSC) measured crystallinity of PCL films at varying % mass loss resulting from the activity of various enzymes with and without RHP. Linear slope fit to scatter data**. B**. 0.05wt% BC with RHP **C.** 0.05 wt% BC, **D.** 0.5wt% AO with RHP, **E.** 0.05wt% AO with RHP, and **F.** 0.5wt% AO embedded in PCL films at varying % mass loss characterized by transmission small angle x-ray scattering (SAXS) profiles (solid lines) with power law fit to low q and extrapolated to high q (dashed lines).

The minimum RHP:enzyme ratio threshold (>2:1) required for the nanoscopic dispersion of AO in PCL was not universally observed across enzymes. SAXS profiles of PCL degraded by embedded PF-RHP demonstrated that a 2:1 RHP:enzyme mass ratio was sufficient to increase P by ~0.4-0.7 with increasing mass loss from 0 to 52%, compared to only a ~0.2 increase to P in PCL degraded to the same mass losses by PF solution (**Figs. 2.2c** and **S2.9c**). We hypothesized that a far greater ratio of impurities in PF might result in a much higher real RHP:enzyme ratio. Additionally, dispersion at 2:1 RHP:enzyme mass ratio may be suboptimal but the inherent processive nature of PF chain scission allowed non-maximally dispersed PF to pull and remove chains that switchboard outwards in all directions from enzyme clusters, creating a vast network of nanoporous tunnels, as was observed in BC. However, despite the similarities between the mechanisms of PCL degradation by BC and PF observed by DSC and SAXS, there are also many subtle and important differences in the

SAXS profiles that indicate their mechanisms of chain scission are not the same. These may explain the vastly slower normalized hydrolysis rate of PF compared to BC both in solution and when embedded. Similarly, although the DSC and SAXS characterized mechanisms of PCL degradation by solutions of MM, CA, and AO were similar, when embedded, all three enzymes deviated in characterized features indicative of uniquely different mechanisms (**Figs. 2.2c, S2.5a,** and **S2.9a,b**). CA had the most unchanged mechanism from surface to embedded whereas MM appeared to exhibit features more associated with processive chain scission. We hypothesized that this may be due to the larger binding pocket volume of MM compared to CA and AO (**Fig. 2.4**). Confinement appeared to accentuate the subtle differences in enzymes that result in varying embedded PCL degradation mechanisms. Nanoconfinement by RHP did not change the intrinsic chain scission mechanism that enzymes would exhibit otherwise. The RHP instead serves two key roles as (1) a nanoscopic dispersion agent to increase the surface area of reaction fronts and (2) a dynamic adhesive/surfactant to strengthen the enzyme-PCL interface. Perhaps RHP composition and molecular weight can be tailored to create hybrid binding pockets to shift intrinsic enzyme mechanism, however, no evidence to support this was observed herein.

GPC showed that in the degradation of PCL by enzyme solutions, despite the varying degradation mechanisms suggested by DSC and SAXS, the resulting product molecular weight (Mw) distributions were very similar, mostly degrading to small molecule products but with clear deviation from a perfect polymer-to-monomer depolymerization (**Fig. 2.3c**). GPC of enzyme embedded films and their degradation products showed a significant shift in the degradation product distributions resulting from enzyme confinement (**Fig. 2.7a**). Overall, embedded degradation shifts the product distribution of all enzymes towards smaller molecules rather than microplastics, nanoplastics, and oligomers. However, there are varying degrees of change. CA has a very clear right shift in high Mw PCL peak indicative of random chain scission and still produced significant oligomeric products with a 19 min elution peak similar to its solution degradation behavior (**Fig. S2.9e**). Unlike CA, MM does not exhibit high Mw film peak right shift (**Fig. S2.9d**). Yet, it still deviates from a perfect polymer-to-monomer depolymerization, particularly when mass loss is < 70% where leached long chains and oligomers (in the form of micro- or nano-plastics) make up ~20% of the total mass. Despite SAXS and DSC suggesting PF operates by processive chain scission, between 35 to 75% PCL mass loss by embedded PF-RHP, 10-15% of the total PCL mass consistently existed as leached long chains and oligomers (**Fig. S2.9f**). We hypothesized that although PF has the appropriate chemistry and binding pocket depth to processively pull/translocate PCL chains, because its pocket volume is too large (*i.e.,* too wide and non-tunnel-like), chains can thread through and skip multiple ester bonds without being hydrolyzed (**Fig. 2.4**). Finally, BC-RHP, AO-RHP, and AO had near 100% < 4-mer product distributions at all % mass losses (**Figs. 2.7b** and **S2.10**). Averaged across all measured % mass losses the total mass that existed as leached long chains and oligomers were -0.2±2.6%, 3.0±5.1%, and 3.3±2.4% respectively. However, BC with no RHP at 100% film mass loss had 22±4% of total PCL mass as leached chains > 500 Da. This supports hypothesis that RHP is necessary to prevent enzyme leaching from the polymer matrix. AO no RHP does not suffer this issue likely because an order of magnitude more AO was

embedded than BC so there is a higher concentration of enzymes that have leached into solution once bulk films structurally collapse. Moreover, we showed that AO surface degradation is much faster than BC surface degradation when normalized by either of 4NPB hydrolysis rate or enzyme solution concentration (**Fig. 2.2** and **Table S2.2**). It is likely that given more time the PCL micro- and nano-plastics will be degraded by BC in solution.

Control experiments were done degrading ~2.5 mg films in ~50 mL water (0.00025 mg enzyme / mL buffer solution) at 40°C, with and without RHP to test our embedded enzyme leaching without RHP hypothesis. This experiment further provided an isolated embedded enzyme degradation product distribution (*i.e.,* minimized impact of degradation by leached enzyme in solution). AO without RHP can reach 100% film mass loss within 24h in ~50 mL water. GPC of the resulting film and lyophilized products in solution showed that 3.0±3.6% total mass remained as long chains >10kDa, comparable to 3.0±1.7% for 1 mL PBS degradation (**Fig. 2.7c**). However, in ~50mL water degradation a large, broad oligomer peak consisting of 13.7±9.1% total mass is observed at ~19 min elution peak with a long, left tail. This is the isolated, non-negligible contribution of purely surface degradation. In 1 mL PBS degradation, leached enzymes can continue degrading from solution, resulting in 0±0% oligomers in the 0.5-10 kDa range. Control triplicate films of pure PCL in ~50 mL water showed no mass loss after 24 hours in 0.00025 mg/mL AO solution. In comparison, AO-RHP embedded degradation at ~66% PCL mass loss in 1mL PBS and ~50mL water both had ~30-40%, ~2-5%, and 55-65% of total mass as >10 kDa, 0.5-10 kDa, and <500 Da respectively. In other words, only 2.0±4.8% of total mass is >500 Da when the intact film has lost ~66% mass. This supports the hypothesis that RHP prevents enzyme leaching and that maintaining embedded, non-leached degradation can skew product distribution to smaller molecules.

PCL degraded by embedded BC without RHP to 100% film mass loss within 24 hours in ~50 mL water was shown by GPC to have 20.7±6.8% of total PCL mass as >10 kDa leached chains in solution. This is similar to degradation in 1mL PBS where 22±4% of total PCL mass was > 10 kDa. However, like with AO, surface degradation in the small volume environment resulted in 0±0% oligomers in the 0.5-10 kDa range. Whereas in the larger degradation volume where leached enzymes are too low in concentration (0.000025 mg/mL), a total mass of 15.0±13.1% remained as 0.5-10 kDa oligomers in solution. Thus, we have shown that despite AO and BC degrading PCL with vastly different mechanisms, the effect of RHP is the same. Without RHP you have leaching. Although no additional >10kDa polymer chains micro-/nano-plastics are released into solution, additional oligomers between 0.5 – 10 kDa are released comprising up to 30% of the PCL film's initial total mass.

Ultimately, we have shown that there are multiple possible pathways an enzyme can follow to degrade semi-crystalline plastic. All 5 investigated enzymes were able to degrade both crystalline and amorphous regions of PCL plastic films. A binary categorization separating plastic degradation mechanisms into processive and distributive (random) chain scission is too simplistic. Although that simplified model offers some utility for plastic hydrolysis by enzyme solution, it offers minimal value for analyzing embedded enzyme degradation. SAXS and DSC showed that 5 lipases each degraded embedded PCL with a unique mechanism, and GPC showed that those mechanisms manifested in unique product

molecular weight distributions. Future work should investigate promising features of AO in relation to hydrophobic polyester degradation. From an engineering perspective it serves as a good target for RHP optimization. The non-charged 3-monomer RHP showed improved activity and provided evidence that optimization is feasible, however the design goal should be to nanoscopically disperse AO with leaching prevention while maintaining or improving degradation rate compared to AO embedded without RHP. There is also abundant foundational knowledge that can be extracted from deeply understanding the mechanism of AO degradation as was done for BC in **Chapter 1**. This would provide an alternative enzyme design paradigm for sustainable plastic biodegradation.

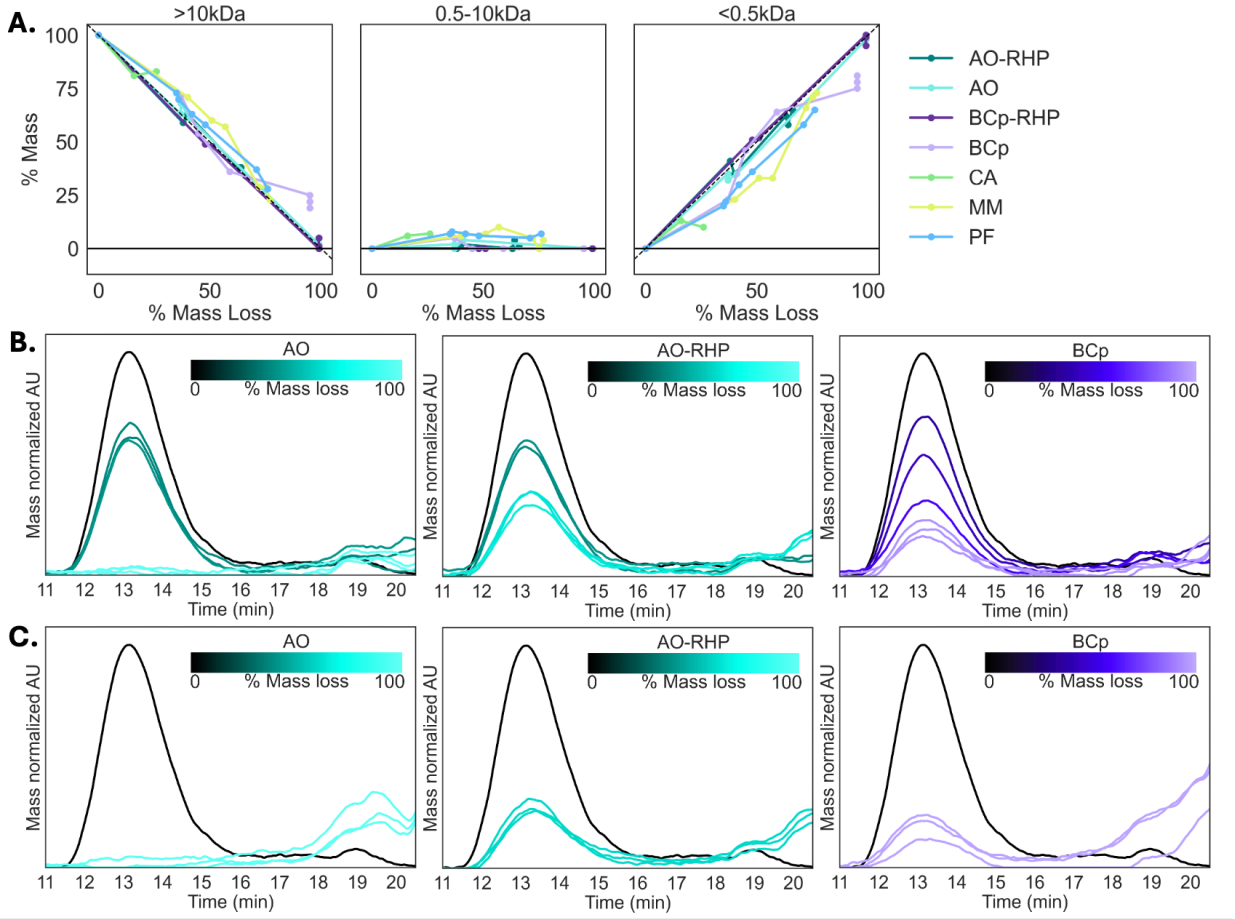


Figure 2.7 | Embedded enzyme degradation product molecular weight distribution of poly(caprolactone) (PCL), measured by gel permeation chromatography (GPC). All spectra are mass normalized by initial mass of films pre-degradation. **A.** Evolution of PCL mass distribution with increasing mass loss split into 3 categories: long chains, (>10 kDa) short chains and oligomers (0.5-10 kDa), small molecules (<500 Da, 4-mer or less), as measured by integrating area-under-curve (AUC) from mass normalized GPC spectra **B.** 0.5wt% AO, AO-RHP (0.5-1 wt%), and purified BC (0.05wt%) embedded films degraded in 1 mL PBS, **C.** and 50 mL Milli-Q water.

2.6 In silico degradation mechanism prediction via molecular simulations

We were unable to predict embedded polyester degradation mechanism in silico due to the medley of mechanisms unveiled by characterization across the investigated enzymes. More sophisticated featurization and experimental data are required. Featurization could be improved by developing a more precise definition of geometry and chemistry by constructing a feature vector that captures 3-dimensional spatial distributions of surface chemistry and pocket depth. For example, we hypothesized that for BC, the bulk of the binding pocket is hydrophobic to attract hydrophobic PCL chains to stably dock in position for hydrolysis to occur. However, at the exiting end of the pocket is a polar patch to attract the carboxylic acid end group that is generated post-hydrolysis to processively pull the chain through the pocket and position the next ester bond at the catalytic serine (**Fig. 2.8a**).[18] To visualize this processive chain slide, we performed molecular simulations of PCL oligomers docking onto enzyme binding pockets and metadynamics simulations of any possible chain slides through the binding pocket. This method may allow for the high throughput generation of mechanism characterization data in silico. Which, when coupled with advanced featurization, may lead to an improved understanding of how to predict and design enzymes that can depolymerize crystalline lamellae.

Semi-flexible molecular (Glide) docking simulations were performed docking a PCL dimer ligand on the binding pockets of all enzymes that were experimentally tested. We rationalized that a semi-flexible system may mimic an embedded enzyme where PCL chain ends have mobility, but the enzyme is largely conformationally trapped due to confinement. The top 20 docked poses were generated and analyzed for evidence of a potential processive chain slide pathway through the enzyme binding pocket. Our criteria for what constituted evidence for potential processivity were that within the top 20 docked poses there existed (1) poses with favorable binding affinity (*i.e.,* negative docking scores) that (2) exist in pre-slide and mid/post-slide configurations that (3) have an incentivized motion trajectory (*i.e.,* poses move from higher to lower docking score). PCL docked on the lid-open binding pocket of BC exhibited the potential for a clear processive chain slide pathway (**Fig. 2.8a**). Surprisingly, molecular docking indicated that AO might have a processive chain slide pathway as it fits all our criteria (**Fig. 2.8b**). However, it is evident that the pocket is significantly shallower, and docking is less stable, as indicated by the significantly lower docking scores (-0.1 to -0.7 kcal/mol) in comparison to those of the PCL dimer docked on BC (-4.5 to -5.8 kcal/mol). This could explain the distributive chain-scission mechanism suggested by DSC and SAXS, as the PCL chains are far more likely to dissociate off the AO active site rather than sliding through it processively. Moreover, there is no large polar patch on the opposite side of a hydrophobic across the AO pocket to create clear entry and exit pathways for the chain to follow.

Disappointingly, none of the other enzymes showed any signs of processive chain slide that could unlock a crystalline lamella (**Fig. S2.11**). This includes PF, that DSC and SAXS experimentally suggested could processively pull PCL chains. However, the structure of PF was generated via AlphaFold 2 and as a result there may be some caveats to this result (**Methods and materials**). Further complications to acknowledge are that PCL dimers are only an approximation and long chain PCL may dock completely differently. Also, it is

unlikely that pockets remain rigid during substrate binding. Lipases have characteristic lids that can open and close and are important to catalytic activity.[52,53] For BC the lid was fixed to an open position, for AO the lid was just manually removed (**Fig. 2.12**). These approximations are likely to lead to further deviations between simulated docking and reality.

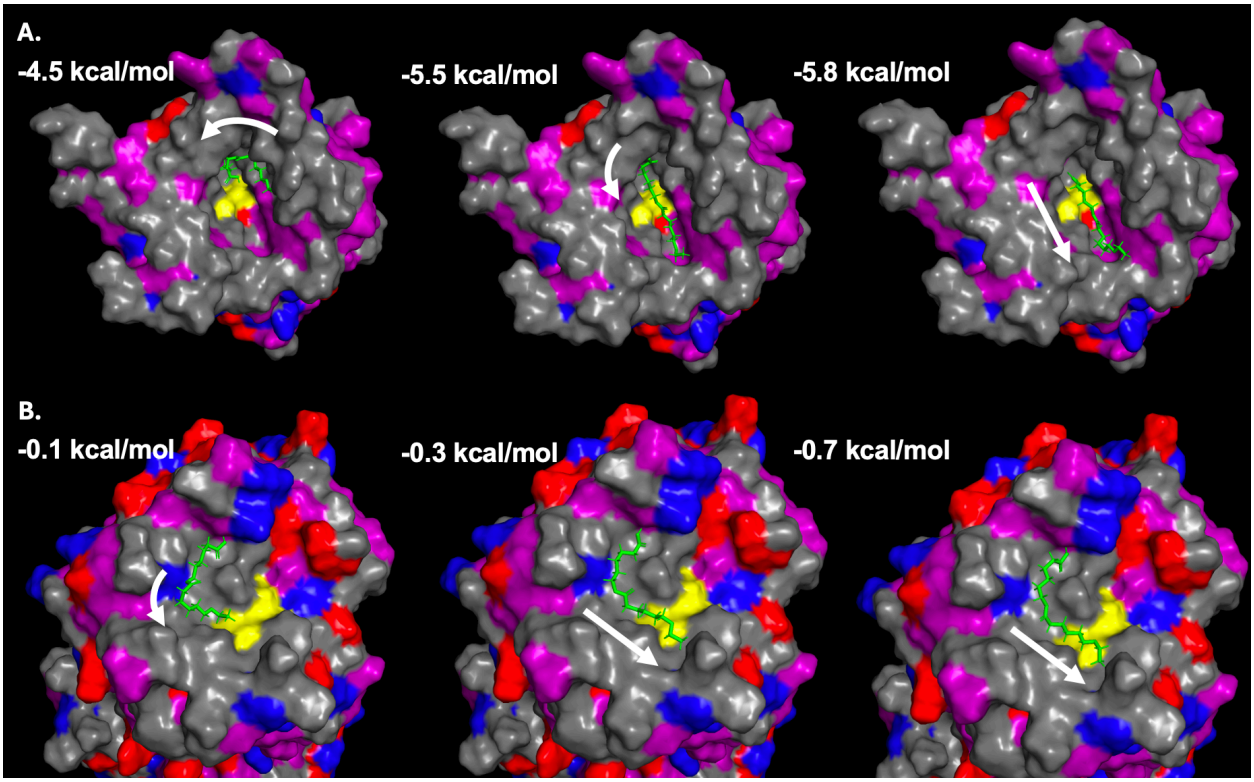


Figure 2.8 | Molecular docking of polycaprolactone (PCL) dimers (green) onto crystal structures of **A.** BC in lid open conformation **B.** AO with lid manually removed. The top 3 poses and their docking energy scores are shown, with illustrative arrows suggesting a possible processive chain slide pathway from least to most stable docked structures, passing across the catalytic triad (yellow). Hydrophobic, polar, positive, and negative residues are colored grey, purple, red, and blue respectively.

Thus, we can see that there are too many issues to make docking a viable prediction strategy. A full molecular dynamics simulation would solve several key issues: binding pockets would be flexible (lids can open and close), longer chain substrates could be accommodated, and instead of interpolation between static substrate poses, a fluid motion trajectory can be observed. However, substrate sliding is a relatively slow event in the timescale of molecular motion. Once a substrate has stably docked (starting pose) with stabilizing interactions between its head group (that has just undergone hydrolysis) and the catalytic serine there is a large energetic barrier for its release. Metadynamics was used to speed up simulation time. Optimization of simulation parameters was done for a PCL 4-mer

initially docked with head group at the catalytic serine in a 315K box of water or octanol to mimic surface and embedded degradation environments respectively. Processive chain slide motion was observed through the binding pocket of BC within 2 ns (**Fig. 2.9a**). In comparison, using the same simulation parameters, the PCL 4-mer does not chain slide but rather flies off CA showing weak binding interactions that may suggest a distributive (random) chain scission mechanism. Simulations were repeated, and screenings were done across varying temperatures and gaussian heights (**Fig. S2.13**).

Ultimately, it would be ideal to develop a fully in silico pipeline to discover or evolve enzymes with mechanisms that can (1) unlock crystalline lamellae depolymerization and (2) have monomeric products rather than micro-/nano-plastics or oligomers. The pipeline can be generalized for use across target substrate plastics (*e.g.,* PLLA, PBAT) that are sustainably sourced but do not have existing recycling infrastructure and do not biodegrade in short enough timescales. The pipeline would work synergistically with high throughput sequence screening, accurate sequence-to-structure prediction, and AI-driven acceleration. We hope this work has demonstrated the importance for enzyme engineering for plastic biodegradation to not only focus on catalysis rate but also on the ability for enzymes to degrade semi-crystalline substrates to recyclable or mineralizable final products.

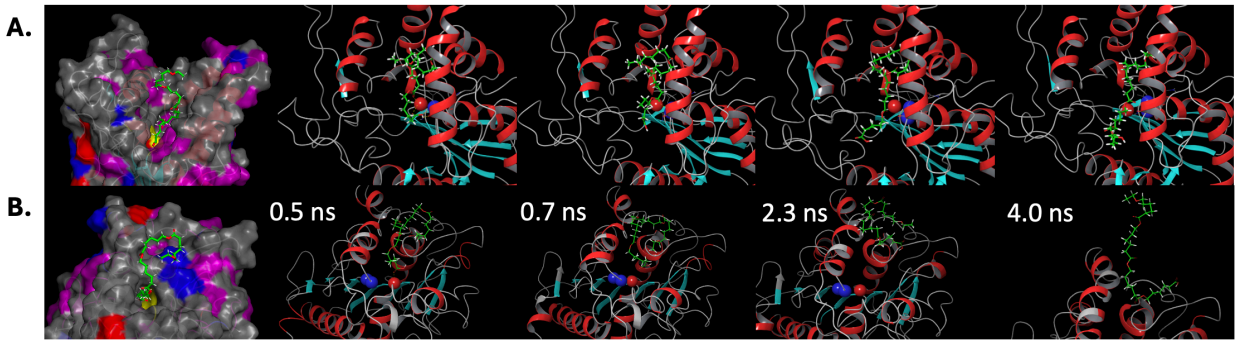


Figure 2.9 | Metadynamics simulation of processive chain slide. Snapshots taken from metadynamics simulations of PCL 4-mer (green) docked to the binding pockets of **A.** BC and subsequent successful chain **B.** CA and subsequent unsuccessful chain slide. Full video of BC processive chain slide: www.youtube.com/watch?v=OViXEAkaRm0

2.7 Methods and materials

Enzyme binding pocket identification and featurization algorithm: is available and described at https://github.com/ivanjayapurna. A caveat on AlphaFold2 (AF2) generated structures: originally, we wrote this algorithm intending it to couple it with AF2 to do high throughput in silico screenings for enzyme sequences with ideal PCL degrading binding pockets. However, we found that many AF2 generated structures were unreliable, possessing long tails with extremely low confidence scores that are likely intrinsically disordered regions that the model is unable to accurately predict. We compared sample hydrolases with known crystallographic structures to AF2 generated structures and the catalytic triads had significant differences in the surrounding residues that made up the active sites. We leave this in silico screening pipeline on non-structurally characterized sequences as future work.

Enzymes: were all purchased from Sigma-Aldrich. Lipase from *Burkholderia cepacia* (BC) was purified following previously established procedures.[54] All other enzymes were used as purchased. Concentration of enzyme stock solutions were made to 1 mg/mL of enzyme in Milli-Q (MQ) water and quantified using ultraviolet–visible (UV-vis) absorbance at 280 nm. To get A280 signal, BC was 10x concentrated, PF was 5x concentrated, RO was 3x concentrated and CA was 2x concentrated. Full details for each abbreviated enzyme in **Table S1**.

Random heteropolymers (RHP): (M_n = 31 kDa; polydispersity, PDI = 1.44) was synthesized as previously reported following a monomer molar composition of 50% methyl methacrylate, 20% ethyl hexyl methacrylate, 25% oligo ethylene glycol methacrylate (M_n = 500 g mol^{-1}), and 5% sulfopropyl methacrylate potassium salt.[55] Non-charged (Neutral), 3-monomer RHP (M_n = 31 kDa; polydispersity, PDI = 1.36) was similarly synthesized following a monomer molar composition of 51% methyl methacrylate, 22% ethyl hexyl methacrylate, 27% oligo ethylene glycol methacrylate (M_n = 500 g mol^{-1}).

4-Nitrophenyl butyrate (4NBP) hydrolysis assay: 4NPB was used as a small molecule ester substrate to test enzyme activity with and without RHP. Enzyme solution was either first mixed with RHP solution in a mass ratio of 2:1, or directly flash-frozen at -80 ºC, and lyophilized overnight to form RHP-enzyme complexes. Lyophilized solids were then resuspended in 50 uL 1x PBS pH 7.4 and transferred to 96-well plate. 200 µL of 0.5 mM 4NPB in 1x PBS pH 7.4 was added. Absorbance at 410 nm was measured for kinetic trials at 30 second intervals using an infinite M nano plate reader. A regression was performed on the linear region of the enzyme activity plot to determine initial rates. All data points were performed in triplicate and mass normalized.

Film preparation and degradation: Polycaprolactone 80 kDa (PCL) was purchased from Sigma Aldrich. PCL was dissolved in toluene at 4 wt% concentration and stirred for at least 4 h at 55 °C to ensure complete dissolution. For degradation by enzyme solution, PCL-toluene solution was stirred for ~30 min before directly casted on a glass slide and air dried overnight at room temperature to an average film thickness of 20 µm. For embedded enzyme degradation, enzyme solution in water was either first mixed with RHP solution in water in a mass ratio of 2:1, or directly flash-frozen at -80 ºC, and lyophilized overnight to form RHP-

enzyme complexes. The dried solids were resuspended directly in the polymer solution at either 0.05 wt% (20:1 RHP:enzyme ratio) for purified enzymes or 0.5 wt% (2:1 RHP:enzyme blend ratio) for commercial enzymes. Mixtures were vortexed for 5 min, cast on a glass slide and air-dried overnight at room temperature to an average film thickness of 20 μm.

Triplicates of rectangular films with an average 2.5 mg initial mass were degraded in 1x pH 7.4 phosphate buffer solution (PBS) in a 40°C water bath. For enzyme solution degradation, 4NPB hydrolysis activity-normalized quantities of aqueous enzyme were pipetted into the buffer solution prior to 40°C incubation. Most trials were done in 1 mL PBS, large volume trials were done in ~50 mL Milli-Q water. Films were removed from at designated time points ranging between 0.5 to 168 hours. Any residual buffer salts or soil were washed with Milli-Q water or removed by gentle air blowing, then lyophilized overnight, then weighed to determine mass loss.

Film characterization: Crystallinity of the films were determined using differential scanning calorimetry (DSC), with a temperature ramp from 20 to 75°C at a rate of 1.5 °C min^{-1}. To determine the percent crystallinity, the sample's integrated enthalpy of melting was normalized against the enthalpy of melting for 100% crystalline PCL, 151.7 J g^{-1}.[56]

Gel permeation chromatography (GPC) was used to monitor molecular weight distribution as a function of degradation. Partially degraded films and degradation product solution were lyophilized overnight and dissolved in tetrahydrofuran (THF). 20 μL of solution was injected into an Agilent PolyPore 7.5 × 300 mm column. GPC spectra were baseline corrected and area-under-the-curve normalized by initial film masses using a custom Python script. A calibration curve of PCL standards was used.

For SAXS studies, 20-30 μm-thick films were folded 4 times to reach 160-240μm-thick film samples and run at the Stanford Synchrotron Radiation Lightsource (SSRL) SLAC National Accelerator Laboratory beam line 1-5 using 15 keV X-rays, 5 minutes of exposure time, beam nominal spot size of 300x500 μm and a Pilatus 1M detector at 3 m distance. AgBe and glassy carbon were used as calibrants. Air background was subtracted from all processed data. Images were plotted as intensity (I) versus q, where $q = (4\pi/\lambda)\sin\theta$, where λ is the wavelength of the incident X-ray beam and 2θ is the scattering angle. The sector-average profiles of SAXS patterns were extracted using Igor Pro with the Nika package.

Black and white, reflection mode, polarized optical microscopy (POM) images taken of plastic films on a heating stage covered by a box to eliminate external light, wind, and dust interference. As-cast images captured and focused on. Samples were then heated to melt at 90°C at a ramp rate of 30°C/min. Films were held for 2 minutes at 90C then temperature was cooled to 49°C at ramp rate of 2°C /min. A 24-hour time lapse with 5-minute imaging intervals was started to capture spherulite growth at 49°C.

Molecular simulations: were performed on Schrodinger Maestro. Protein PDB files were processed by the protein preparation wizard to delete extraneous ligands, add missing hydrogens, correct hydrogen bonding network, and minimize. Enzymes with closed lid structures were manually opened. The ligand, a 2-mer or 4-mer of PCL with -COOH end

group to mimic the product of hydrolysis, was manually drawn and optimized using Ligand Prep using default settings. Cubic docking grids were generated of length 20Å and centered about the catalytic serine. Glide docking was done considering the enzymes as rigid and the top 20 poses were generated.

Metadynamics simulations were performed using a PCL 4-mer docked in a hypothesized pre-slide position on the enzyme binding pocket in a water or octanol box as the starting state. Water was chosen to mimic enzyme solution degradation and octanol was chosen to mimic the hydrophobic environment of embedded enzyme degradation. Collective variable was set between the catalytic serine oxygen and the center of mass of the 7 atoms of the PCL head group. Gaussian hill heights were varied between 0 – 0.05 kcal/mol, temperature was varied between 285 - 315K, and simulations were run for 15-30 nanoseconds.

Chapter 3

Modulating protein stability via intermolecular protectants

3.1 Industrial bioactive plastic processing

Bioactive plastics *e.g.,* plastics with embedded biotic species such as proteins, are versatile functional materials that leverage nature-optimized designs in human-engineered systems. At present they are mostly pharmaceuticals and smart food packaging with embedded antimicrobial agents.[1,2] However, bioactive plastics are also excellent for heterogeneous catalysis applications such as water purification, as unlike solution catalysis, solid films can be easily transported and placed in varying environments in situ without facing post-reaction separation issues. The range of synthesizable bioactive plastics was significantly broadened with the development of random heteropolymers (RHPs), amphiphilic copolymers that can stabilize proteins in non-native environments such as organic solvents.[3] RHPs can make ordinarily fragile proteins compatible with solvent-casted plastic processing by preventing their misfolding and aggregation en route to solid-state encapsulation.[4] External small molecules aside, the substrate for heterogeneous catalysis could also be the plastic matrix itself. Nanoscopically embedding enzymes into biodegradable plastics is a promising approach to modulating degradation latency, eliminating single-use plastic waste, and mitigating microplastic formation.[5–10]

Unfortunately solvent processing using volatile organic compounds is not environmentally sustainable and thus is often legally restricted.[11] For industrial relevance at scale, proteins must be compatible with traditional melt extrusion and additive manufacturing plastic processing, in which plastic pellets are heated at high temperatures to melt, mixed with additives for a residence time, extruded, then molded or air blown to shape.[2,12] Proteins must retain folded structure after incubation in high temperature polymer melts for the duration of the residence time to achieve commercially viable bioactive plastics. Residence times are highly dependent on the polymer, extruder type and operating parameters, and are usually minimized because polymers at high temperatures can also thermally or hydrolytically degrade. Poly(lactic acid) (PLA) is a commercially relevant platform for bioactive plastic processing exploration. In 2023, PLA represented 31% of the 2.18 million metric tons of global bioplastics produced annually.[13] Reported PLA residence times vary from as low as 20 seconds to as high as 8 minutes, with most on average being under 5 minutes.[14,15] Thus, enzymes that degrade PLA must retain folded structure after ~5 minutes of residence time in viscoelastic polymer melt above the maximum plastic melt temperature (T_m) *i.e.,* 150-200°C for PLLA.[14] This temperature is common for most commercial plastics that range in T_m from 100-200°C. After extrusion, the enzymes must remain catalytically active at 40-65°C to match compost conditions.[16]

Most enzymes denature in aqueous solutions as temperatures approach 60-80°C, and even the most hyperthermophilic proteins are only thermostable up to 117°C.[17] Above 110°C, molecules such as amino acids and metabolites become highly unstable and

hydrophobic interactions significantly weaken. Moreover, these hyperthermophiles evolved in extreme environments, so their optimal operating temperatures are also inconveniently high. The >100°C thermostability gap to $T_{m,PLLA}$ is unlikely to be bridged by enzyme engineering,[18] but rather by the design of protective chaperones. Proteinase K (PK) is the most well demonstrated PLLA degrading enzyme, however, it loses all catalytic activity when heated in solution is 5-10 minutes at 75°C.[19] Recent works have immobilized PK onto polyacrylamide microparticles to retain 2% activity after 5 minutes at 200°C,[19] and onto ZIF-8 to retain 61.8% activity after 10 minutes at 160°C.[20] However, no existing result is amenable for practical use, which requires engineering stable PLLA degrading enzymes such as PK for a minimum of 5 minutes of residence time at up to 200°C in PLLA melt.

RHPs were designed for solution stabilization and have been shown to increase protein thermostability in solution.[21] It was shown that RHP does not actually prevent misfolding but rather chaperones enzyme refolding back to the native structure during the cooling process. However, in solution interactions, hydrophobic effects dominate, it is an open question if that remains true in polymer melts. There are three fundamental ways in which protein stability may differ in polymer melts compared to biological solutions: polarity, water deficiency, and viscosity. Although plastics vary in polarity, in general compared to aqueous solutions, most plastics are hydrophobic. In solutions the hydrophobic effect is dominant. Losing this in polymer melts could cause protein misfolding. However, the large body of research on enzyme stability in organic solutions showed that it might help enzyme rigidification due to the retention of the core hydrophobic shell – good for stability, bad for activity (as a certain degree of flexibility is required for function).[22,23] Hydrophobic plastic matrices at 200°C will be almost if not entirely water deficient. The lack of water will destabilize hydrophobic and hydrogen bonding stabilizing interactions. Finally, considering viscosity, the density of amorphous PLA is 1248 g l^{-1},[24] an order of magnitude higher than the estimated total concentration of macromolecules in the cytoplasm, that ranges from 50 to 400 g l^{-1}.[25] Comparing to enzymes in dilute aqueous solutions, the contrast in molecular crowding is even more stark. High viscosity helps some proteins remain stable and causes others to misfold depending on interactions. Although Nature has evolved proteins to not interact with locally surrounding macromolecules to avoid misfolding, it has done no such fine-tuning for synthetic polymer interactions. Thus, we expect high viscosity to hurt rather than help protein stability. However, this effect may be counterbalanced or exceeded by an increase in enzyme structural stability afforded by immobilization in a viscoelastic state. If dynamics are low, perhaps proteins will not have the flexibility to misfold even in a liquid polymer melt.

Here we demonstrated through a series of systematic experiments, fundamental differences between protein thermostability in polymer matrices and melts, in comparison to aqueous and organic solutions. We showed that multiple plastic-degrading serine hydrolases are thermostable up to 145-155°C in plastics of varying polarity, both biodegradable and traditional polyolefins, as well as melts of varying viscosity. Excitingly, we exceeded the engineering goal of stabilizing proteins that catalyze PLLA degradation in PLLA melt, by demonstrating that 2 hydrolases: proteinase K from *Tritirachium album* and

protease from *Streptomyces griseus*, retain activity after >8 minutes of residence time in 200°C PLLA melt. We achieved this by simply co-embedding the enzymes with $MgSO_4$, and subsequently demonstrated that films containing 2wt% enzyme and 10wt% $MgSO_4$ catalyze the rapid degradation of their host PLLA films up to 50% mass loss within 24 hours at 40°C. This offers a cheap, simple, and scalable enzyme-salt additive for bioactive and truly compostable plastic melt processing. We further showed that the deactivation temperature is universal across the tested hydrolases that span a diverse range of secondary structures, and that the stabilization effects of $MgSO_4$ are universal. However, control experiments showed that these results are not purely due to local hydration modulation. Specific structural change and/or chaperoned refolding must also be occurring. However, several fundamental questions remain and thus we conclude this chapter by laying out a proposed roadmap for subsequent investigation.

3.2 Thermostable proteins in 200°C polymer melt

The primary engineering goal of this study is to enhance protein stability during industrial polymer melt extrusion. However, even micro-extruders on a lab scale produce films of at least gram-scale, making melt extrusion inefficient for the repetitive screening and assaying of biological additives due to material wastage, slow processing, and the confounding effects of mixing gradients and mechanical shear. Although mechanical impacts are relevant to protein thermostability in melt extrusion, they will be left as a topic for future exploration. Our current focus is on how the temperature, chemistry, and viscosity of stationary polymer melts affect protein thermostability. To this end, we developed a proxy thermal processing method (**Fig. S3.1, Methods and materials**). Polymer, enzyme, and any intermolecular chaperones were mixed in organic solvent, solution cast, air dried into films, and then heat treated for a simulated extruder residence time of 5 minutes. After heat treatment, the films were rapidly cooled and assayed for enzymatic activity using a kinetic heterogeneous catalysis colorimetric assay. Enzyme embedded films are soaked in solution of 4-nitrophenyl butyrate (4NPB), an ester bond containing small molecule whose hydrolysis is catalyzed. Product absorbance at 410nm is measured over time, and initial enzyme activity rates are calculated via an initial, linear slope fitting. Activity rates of heat-treated films were normalized by the activity rates of pre-heat treated (as-cast, AC) films of the exact composition, yielding % retained activity (% RA) values that are plotted. Normalization is important as it allows for fair comparison across experiments that used different enzyme batches, enzyme loading %, or even different enzymes, plastics, or small molecules entirely. Varying each of these will result in different extents of activity loss due to misfolding during organic solvent processing, AC catalysis rates, and diffusion pathways to and from solid-state enzymes embedded in hydrophobic films. Thus, normalization removes all the confounding variable for comparison of only thermostability in the polymer matrix/melt.

Applying this method, we found that 2wt% proteinase K (PK) is outstandingly thermostable in 155°C 30kDa poly(L-lactic acid) (PLLA30), maintaining 59±10% RA at 155 °C (**Fig 3.1a**). However, PK is fully deactivated by 185°C, and so would not survive melt extrusion in PLLA30. PK stability at 155°C was surprising considering without any optimization, confinement in PLLA30 resulted in thermostability ~80°C higher than T_m in solution. Recently published literature also observed similar results, and may suggest that the molecular crowding effect of densely entangled polymer chains effectively immobilize the enzyme, removing the flexibility required to misfold.[19,20] Indeed, differential scanning calorimetry (DSC) showed that at 155°C PLLA30 is above T_g but below the onset T_m (**Fig 3.1b, Table S3.1**). Incredibly, we discovered that the inclusion of just 10wt% $MgSO_4$ enables the retention of 44±11% RA at 185 °C, 15±2% RA at 200 °C, and 6±2% RA at 215 °C, before fully deactivating by 230°C. DSC results further showed that including $MgSO_4$ in films resulted in a minor 4.1±4.8% (n=6) decrease in film crystallinity calculated from the total integrated area of cold-crystallization and melting peaks (**Fig. 3.1b**). An exact % crystallinity of films was not calculated as there are enthalpic relaxation peaks at lower temperatures as PLLA30 heats past its T_g that were not measured. Embedding 10wt% $MgSO_4$ resulted in the PLLA30 melt temperature mode (T_m) only increasing from 173 to 174°C, and the maximum ($T_{m,max}$) remained constant at 178°C. Thus, it is safe to conclude that by 185 °C all polymer chains in

all films with or without salt are fully melted, and the salt effect improves enzyme stability rather than plastic matrix stability.

To test if our proxy assay for retained activity could translate to post heat treated plastic degrading activity, we heat treated PLLA30 films for 8 minutes at 185 °C to ensure that all films were fully melted. Films were air cooled then incubated for 24 hours in 40°C pH 8.0 tris buffer solution. Post-incubation films were lyophilized to remove adsorbed water. Films with only 2wt% PK barely decreased in mass by 0.8±0.2%, whereas films with co-embedded 2wt% PK and 10wt% $MgSO_4$ decreased in mass by 43.4±10.4% (**Fig. 3.1c**). Macroscopically PK only films gained a brown tint, indicative of misfolded and aggregated proteins, whereas the PK+$MgSO_4$ films turned completely opaque white and became brittle, indicative of the formation of internal porosity. However, films embedded with $MgSO_4$ lost 3.9±0.2% mass, suggesting that ~25% of total salt mass leached out of the film. This was expected as prior literature has intentionally solution-cast poly(L-lactic acid) with embedded salt and then let the salt leach to create microporous foams.[24] Given that internal degradation creates nanoporous small molecule transport channels (**Chapter 1**), a higher rate of salt leaching is expected in partially degraded films. Yet, even if we assume 100% of $MgSO_4$ leached out in the PK+$MgSO_4$ films and subtract that from the mass loss, lower bounded PK degradation resulted in 32% PLLA30 mass loss. Additionally, buffer solution pH was measured after removing films with PK+$MgSO_4$ embedded, and it dropped to pH 5. This suggested significant lactic acid production and further confirmed PK was made thermostable after 8 minutes at 185 °C by co-embedding $MgSO_4$ and suggest that degradation rate might have been higher had a larger volume or a continuous flow of buffer solution been used to maintain the optimal pH 8.

In this work the mechanical properties of plastic films were largely neglected. Depending on the engineering application, foam-like microporous plastics could be desired, however, for plastic packaging where barrier properties are required, this is definitely not the case. We screened the thermostability of PK in PLLA30 with varying $MgSO_4$ loading and found the optimum to be within the 2-20 wt% range (**Fig. 3.1d**). In this work 10wt% was used to optimize for enzyme activity retention, however in practice the trade-off with film mechanical properties can be considered when picking a desired salt loading. We leave further plastic engineering optimization as future work, separate from the upcoming fundamental scientific exploration presented herein. Ultimately, we successfully demonstrated thermostabilization of a functional enzyme in a mass-produced biopolymer using a common salt ($MgSO_4$). These results may pave the way for industrially viable, truly compostable and/or bioactive plastics.

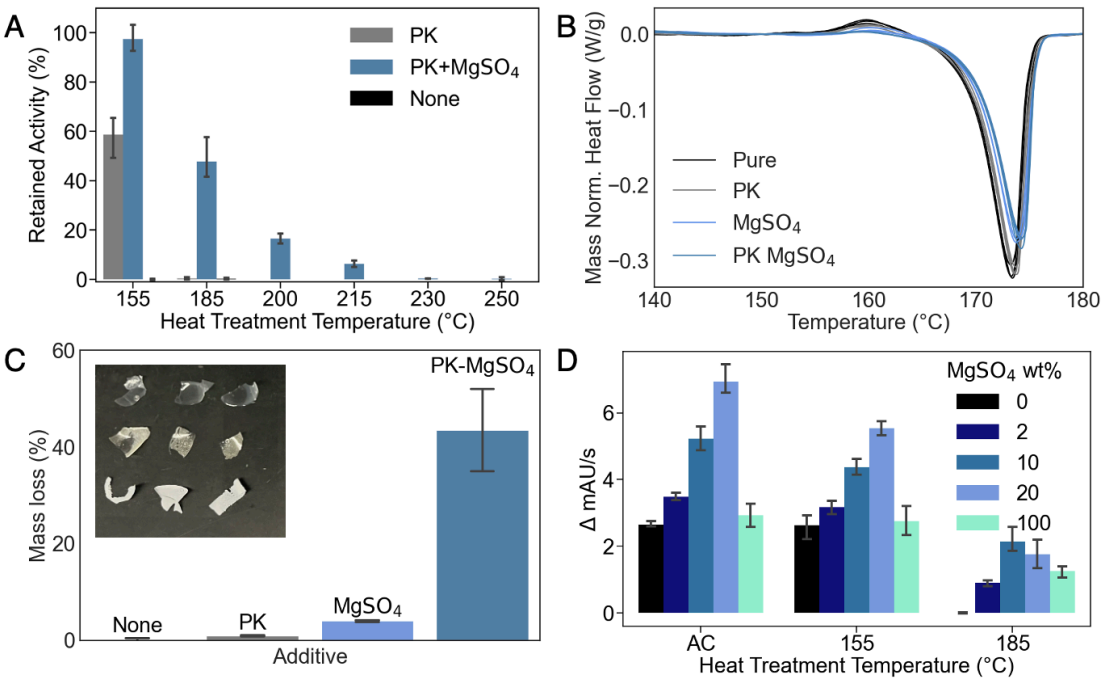


Fig 3.1 | $MgSO_4$ thermostabilizes Proteinase K (PK) in 30kDa Poly(L-lactic acid) (PLLA30). For all bar plots triplicates were averaged and standard deviations were plotted as error bars. **A. %** Retained activity of PK, PK and $MgSO_4$, and control films with no additive after heat treatment for 5 minutes at temperatures ranging from 155-250°C. **B.** Triplicates of differential scanning calorimetry (DSC) spectra of films containing no additive (Pure), 2wt% PK, 10wt% $MgSO_4$ and both together. **C.** % Mass loss of 8 minute 185°C heat treated films with no additive, 2wt% PK, 10wt% $MgSO_4$, and both after 24 hours in 40°C pH 8.0 tris buffer. **D.** Activity rate of 2wt% PK co-embedded with varying ratios of $MgSO_4$ as-cast (AC) and heat treated for 5 minutes at 155 and 185 °C.

3.3 Universal rules governing protein thermostability in polymer melts and matrices

We next investigated if the remarkable observed protein thermostability in plastic melt that is further enhanced by co-embedding $MgSO_4$ is universal or system specific to PK in PLLA30. We systematically screened thermostability across various proteins-polymer pairings. The thermostabilities of 5 hydrolases: PK, AO, BC, BL, SG and 1 fluorescent reporter protein: GFP (**Methods and materials**), were investigated across 6 plastics: PLLA30, PLLA85, PLLA100, PMMA, PCL, PS (**Methods and materials**). The enzymes are relatively pure A260/A280 ~0.6, with only SG slightly less pure at ~0.65 (**Fig. S3.3**). Turbidity of all enzyme solutions was also low, quantified by absorbance at 320nm. Enzyme surfaces features were computationally extracted using the pipeline developed in **Chapter 2** (**Fig. S3.4**). 3 enzymes originate from bacteria, 2 from fungi. 2 of the hydrolases are lipases and 3 are proteases. 2 enzymes contain Ca^{2+}, 1 has both Ca^{2+} and Na^{+}, and 2 have no metal ions. Chemical surfaces of proteins are relatively similar overall, with the largest discrepancy being AO having the most charged surface with ~25% charged residues that are more negative, while SG has the least charged surface at only ~10% coverage, of which more are positive. All enzymes are extremely spherical, and notably is SG uniquely beta-sheet dominant (~60% beta-sheet, ~35% random coil), while all other enzymes are <35% beta-sheet, with a more equal distribution (35-45% alpha-helix, 40-60% random coil). Plastics investigated ranged in polarity and in molecular weight (Mn) from ~30-500kDa. The three PLA films and PCL are semi-crystalline, while PMMA and PS are amorphous. These features were hypothesized to be predictive of embedded enzyme thermostability and were summarized in **Table S3.1**.

Thermostability screens of PK, AO, BC, and BL across six plastics showed that all 24 combinations had 0% RA after heat treatment by 155-165 °C. Up to 145 °C, AO and PK were generally more thermostable across plastics in comparison to BC and BL. There are 2 notable enzyme-plastic pairing exceptions to this that stand out as more thermostable than the rest: BC-PCL and BL-PMMA. Interestingly, the % RA of PK, BC and BL seems to decrease from 135-165 °C with a smooth exponential decay, however, AO experiences a sharp step-function drop in % RA between 145 and 155°C. This could be due to AO glycosylation that was observed in **Chapter 2**. Despite these variations in enzyme thermostability after heat treatment in the 135-145 °C range, the overwhelming result remains that none of the protein nor plastic features are significant when pushing towards $T_{m,PLA}$. Importantly, although 165°C happens to be the onset of PLA melting, this temperature marks no significant chain conformation or dynamics transition for any of the other 3 plastics. This screening was also repeated with enzyme loading increased from 1-2wt% to 10wt% (**Fig. S3.6**). The same outcome was observed, demonstrating that thermostability cannot be increased by increasing enzyme loading. These results point to some universal mechanism of deactivation that is independent of surface chemistry interactions and chain density confinement effects.

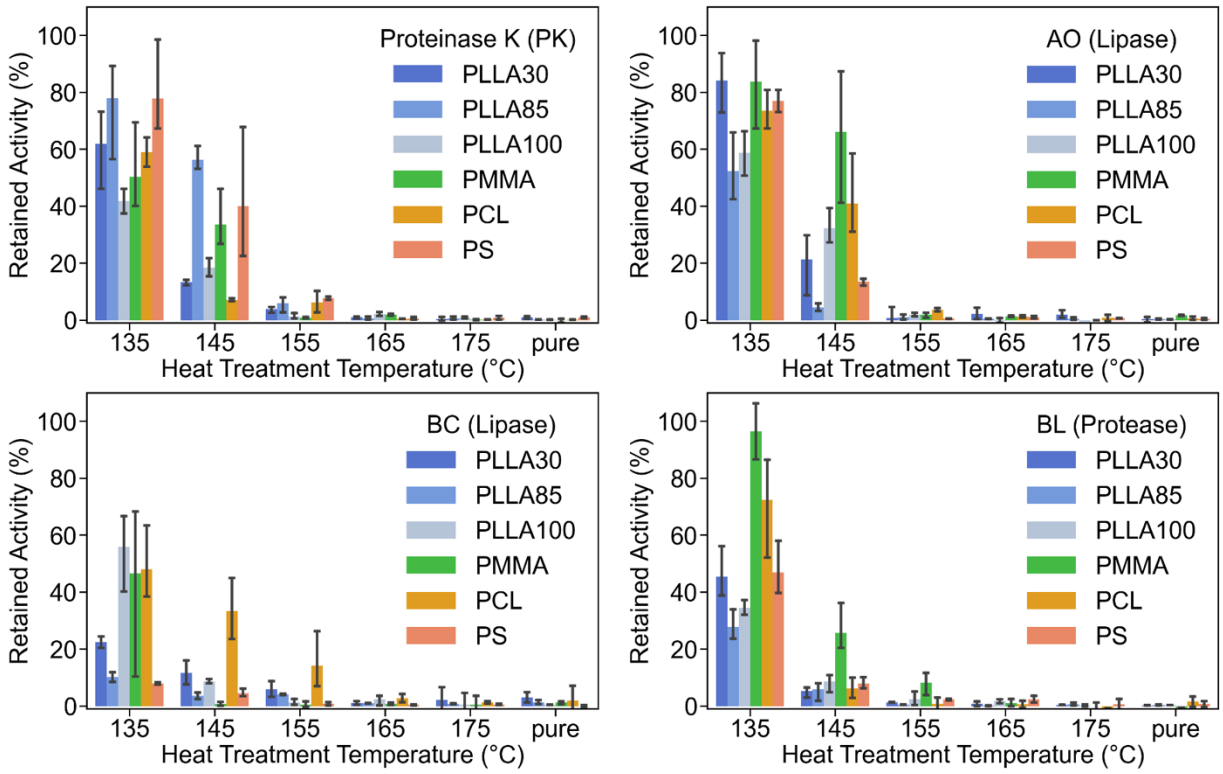


Fig 3.2 | 155°C universal deactivation across 24 protein-plastic pairings. The retained activity % of 4 Hydrolases (labeled in the legend titles) were screened across 6 plastics from 135 to 175 °C.

Water is the remaining universal factor. We hypothesized that universal misfolding across all protein-plastic pairings was due to evaporation of solid-state water bound to enzyme (hydration shell layer), RHP and oxygen groups in polyesters. This temperature would be relatively agnostic to variations in surrounding plastic, enzyme chemistry and matrix/melt morphology. If this were true, the role of $MgSO_4$ could therefore be to modulate the water molecules local to embedded enzymes. $MgSO_4$ is a well-known dehydrating agent, can form various highly hydrated states, and thus absorb or release water depending on environmental humidity.[26–28] We screened various protein-plastic pairings with and without $MgSO_4$ to test the universality of its thermostabilizing effect. In short, $MgSO_4$ boosted the thermostability of every pairing tested. When plastic is varied from the more polar PLLA30 to the most hydrophobic PS, PK was still thermostabilized by $MgSO_4$, retaining 34±3% RA after 185 °C heat treatment (**Fig 3.3a**). PK only control films with no salt under the same conditions were completely deactivated (~0.1% RA). We further showed that in addition to PK, all of AO, BC, and SG can be thermostabilized in PLLA by $MgSO_4$ up to heat treatment of at least 185°C, demonstrating universality across hydrolases even with significantly varying secondary structures, species of origin, metal ions and native substrates (**Fig. 3.3b**). An important caveat is that SG was shipped with calcium acetate (CaAc) for solubility and was used as purchased without desalting. SG-CaAc had 53±9% and 5±1% RA after 185 °C and 200°C heat treatments respectively. When $MgSO_4$ is co-embedded with SG and CaAc, both increased to 140±3% and 99±15% RA after 185 °C and 200°C heat treatments respectively. >100% RA indicates that $MgSO_4$ also stabilizes SG-CaAc in organic solvent during solution casting. It is very likely that this combination could be pushed farther than the 230 °C max temperature observed for PK- $MgSO_4$ in PLLA.

Prior literature demonstrated SG had PLLA degrading activity,[29] thus we performed a similar degradation experiment as was done with PK in **Figure 3.1c**, with SG (**Fig. 3.3c**). Similar to PK, post-heat treatment PLLA degradation results were in line with the results from post-heat treatment proxy small molecule degradation assays. SG-CaAc lost 3.9±0.2% mass, while co-embedding $MgSO_4$ boosted this to 22.2±8.5% mass loss. Assuming 100% of $MgSO_4$ leached, the lower bound for PLLA mass loss is ~7%, a slight increase from the SG-CaAc film with no $MgSO_4$. Control films with no enzyme only changed in mass by 0.1±0.4%. The significant reduction in PLLA mass loss over 24 hours by SG compared to PK is likely not an issue of thermostability but rather one of enzyme-substrate promiscuity (**Chapter 2**).

Finally, to test if the remarkable observed protein thermostability in plastic melts generalizes beyond hydrolases, we followed the same processing and heat treatment procedure to embed GFP in PLLA. Instead of a colorimetric kinetic assay, fluorescence of films was simply measured (**Methods and materials**). **Figure 3.3d** shows that all 5 hydrolases we screened are unusually stable in solvent-casting processing, as GFP films has near-zero fluorescence even as-cast (prior to heat treatment). GFP films co-embedded with random heteropolymers (RHPs) to chaperone the solvent processing, showed significant fluorescence as-cast, but almost entirely deactivates by 145 °C. The key region for GFP is a buried chromophore, compared to the surface-exposed binding pockets of the 5 hydrolases we tested. It remains an open question why this behavior is different, and if $MgSO_4$ as well as CaAc are universal beyond hydrolases to stabilize all proteins in al plastics.

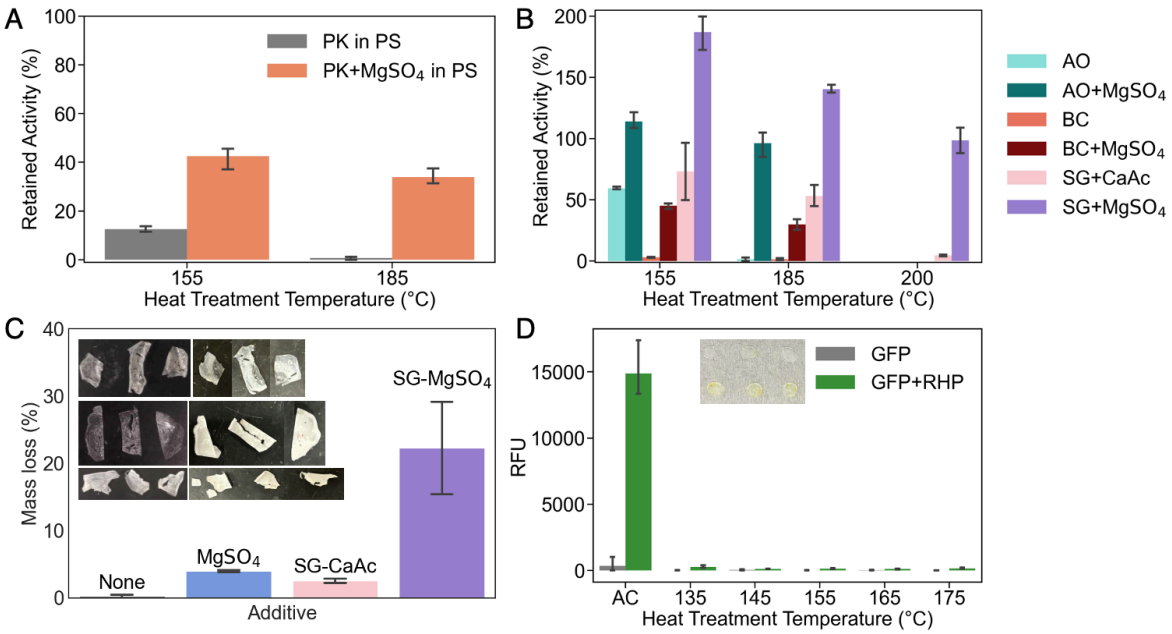


Fig 3.3 | $MgSO_4$ universal thermostabilization across all hydrolase-plastic pairings. For all bar plots triplicates were averaged and standard deviations were plotted as error bars. **A. %** Retained activity of PK, PK and $MgSO_4$, after heat treatment for 5 minutes in PS. **B. %** Retained activity of AO, BC and SG-CaAc with and without $MgSO_4$, after heat treatment for 5 minutes in PLLA80. **C.** % Mass loss of PLLA30 films with no additive, 2wt% SG-CaAc, 10wt% $MgSO_4$, and both, all heat treated for 8 minutes at 185°C, after 24 hours in 40°C pH 8.0 tris buffer. **D.** Fluorescence of GFP embedded PLLA80 films with and without a 4-monomer random heteropolymer (RHP) as cast, and heat treated up to 5 minutes in 185 °C.

3.4 Salt-enzyme-polymer specific thermostabilizing interactions

To test the bound hydration water evaporation hypothesis, we did a modified heat treatment procedure where instead of an oven (dry environment) we heat treated in an environment saturated with ≥99% relative humidity steam (**Figure S3.1**). In such an environment we minimize the driving force for evaporation. PK, AO, and BC in PLLA were steam heat treated, with results matching those done in the standard, dry heat treatment setup, as all 3 enzymes fully deactivate by 155°C (**Fig. 3.4a**). To further probe our hydration hypothesis, we screened PK thermostability in PLLA with several salts that can form highly hydrated states (**Fig. 3.4b**). Although lanthanides are extremely hygroscopic, all 3 tested (nitrates of Eu, Nd and Dy) deactivated enzymes completely prior to heat treatment. This suggests there was likely a binding interaction that was too strong, causing PK to misfold. Like $MgSO_4$, $MgCl_2$ and $Mg(NO_3)_2$ are common hygroscopic salts and share the same cation that was able to significantly improve the thermostability of all tested hydrolases in all tested plastics.[26] However, even these common salts fully deactivated PK in PLLA as cast. Future work should be done to determine if this deactivation is occurring during aqueous mixing, lyophilization, organic solvent mixing or when dried into solid state. Na_2SO_4 and $ZnSO_4$ are common hygroscopic salts that share the same anion as $MgSO_4$. Co-embedding these salts with PK in PLLA did not immediately kill the enzyme during solvent processing, however, they also did not improve the thermostability of PK without salt. Only cobalt acetate (CoAc) successfully managed to retain PK activity after heat treatment at 185 °C. Perhaps CoAc stabilizes PK by the mechanism by which CaAc stabilizes SG. However, it should be noted that the CoAc mixture was extremely insoluble in organic solvent, so it's likely most of the salt, and a large portion of salt-bound enzyme, did not end up in the plastic films. Together, these results showed that the universal deactivation of all tested hydrolases in all tested plastics at 155-165 °C as well as the thermostabilization mechanism of $MgSO_4$ are not purely water/hydration driven phenomena.

Our next hypothesis was that ionic salts are binding to hydrolases, resulting in specific structural changes and/or are enabling chaperoned refolding during the cooling process after heat-treatment. These specific structural changes/chaperoned-refolding must come from interactions between protein-cation, protein-anion, or both. The Na_2SO_4 and $ZnSO_4$ results show that $MgSO_4$ does not thermostabiliize enzymes by a specific sulfate anion effect alone. However, we cannot as easily conclude the same for the Mg^{2+} cation, as both $MgCl_2$ and $Mg(NO_3)_2$ completely killed embedded enzymes as-cast. Enzymes likely aggregated and/or misfolded during the solvent-processing, which is likely a separate mechanism, not correlated to thermostability in plastics. Given that $MgSO_4$ does not deactivate PK during solution processing, we next investigated if the chloride and nitrate anions were responsible. We chose salts with ammonium and potassium cations that would likely not strongly interact with PK detrimentally. Thermostability assay results showed that nitrates indeed almost entirely deactivated as-cast PK in PLLA (**Fig. 3.4c**). However, surprisingly NH_4Cl and KCl not only did not deactivate PK in solution, but KCl even increased PK stability in PLLA up to 5 minutes at 175 °C. Despite not being as effective as $MgSO_4$, this is still a significant increase.

KCl having moderate success further supports the specific ionic interactions hypothesis over the initial hydration modulation hypothesis, as it is not hygroscopic and does not form highly hydrated states. As with $MgSO_4$, we then tested the universality of KCl as a stabilizer across varying hydrolases and plastics. Without KCl, BL is fully deactivated by heat treatment at 155°C in both PLLA and PMMA (**Fig. 3.4d**). We showed that BL was stabilized by KCl in PLLA and retained 34.9±5.1% RA after heat treatment at 195 °C. However, this extreme stabilization increase by KCl was not observed in PMMA, where retained activity only increased to 5.2±1.1% and 3.4±1.2% after heat treatment at 155 and 175 °C respectively. Thus, unlike how $MgSO_4$ was able to successfully stabilize all hydrolase-plastic pairings tested up to at least 5 minutes at 185°C, KCl could only demonstrate a similar effect for BL-PLLA, and not BL-PMMA nor PK-PLLA. Interestingly, this shows that there must be a specific salt-enzyme-plastic tri-interaction. KCl also was shown to have a much higher optimal salt loading compared to $MgSO_4$, with an optimum loading to thermostabilize 1wt% BL in PLLA being somewhere between 10-100 wt%. This further demonstrates that although $MgSO_4$ and KCl are both ionic salts, they may be thermostabilizing enzymes with entirely different mechanisms, one that is universal and one that is salt-enzyme-plastic specific.

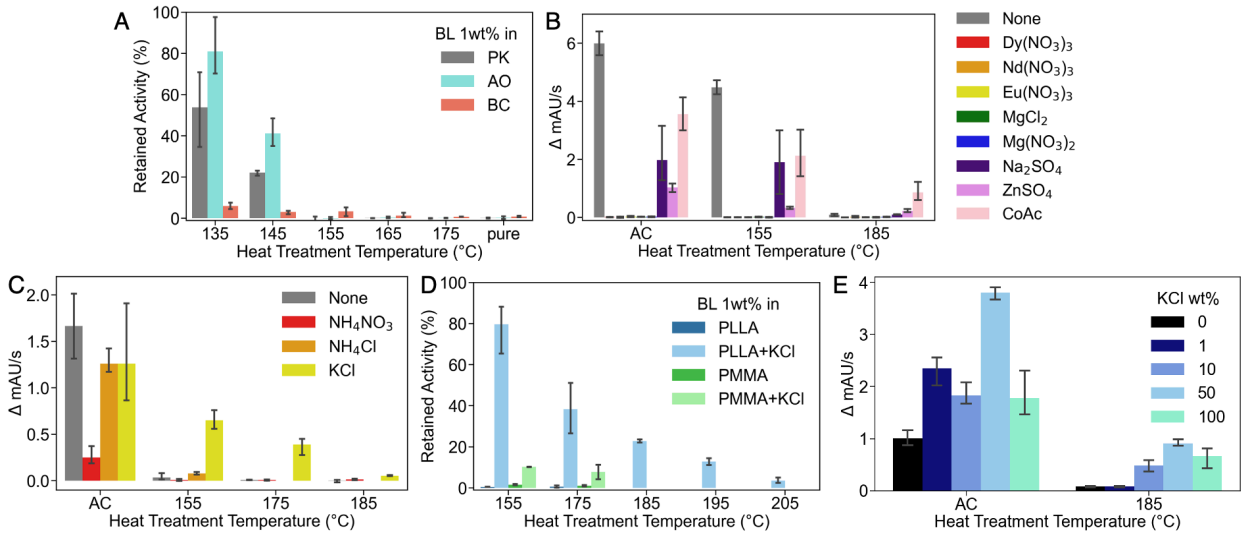


Fig 3.4 | Thermostability screen with varying salts, plastics, and enzymes. For all bar plots heat treatment was done for 5 minutes, triplicates were averaged, and standard deviations were plotted as error bars. **A. %** Retained activity of PK, AO, and BC in PLLA80, after heat treatment in ≥99% RH air environment. **B.** Activity rate of 2wt% PK in PLLA80 co-embedded with varying hygroscopic and efflorescent salts. **C.** Activity rate of 2wt% PK in PLLA80 co-embedded with salts with chloride and nitrate anions. **D.** % Retained activity of BL in PLLA and PMMA with and without KCl. **E.** Activity rate of 1wt% BL co-embedded in PLLA with varying ratios of KCl as-cast (AC).

3.5 Future work

This work lays the foundation for understanding protein stability in polymer melts. In successfully demonstrating extreme protein thermostabilization, we have raised more questions that provide fertile ground for future investigations. The key gap is a molecular understanding of what is happening to proteins during film processing and heat treatment. Developing reliable methods to probe the evolution of protein structures in polymer melts and matrices before, during and after heat treatment could lead to an understanding of the mechanisms by which $MgSO_4$, KCl and to an extent CoAc and CaAc, are able to stabilize proteins. We propose a 2-phased future investigation.

First, we investigate if protein structure is changing upon confinement with and without salt. This can be done by comparing the secondary structure distributions of proteins with and without co-lyophilized salts in aqueous circular dichroism (CD) and post-embedding into plastic films via solid-state CD. A good starting set of salts to screen would include a seemingly universally thermostabilizing salt ($MgSO_4$), and chemically similar salts that fully deactivate enzyme prior to heat treatment ($MgCl_2$) and that neither increases nor decreases thermostability (Na_2SO_4). KCl would also be interesting as a second successful thermostabilizer that may operate via a different mechanism.

In Phase 2 we utilize the same characterization platform to probe protein structural change with and without salt before and after heat treatment. We ideally want to specify if proteins are misfolding and aggregating at high temperatures or during the cooling process. Similarly, are salts buffer protein structural changes at high temperatures, or if they are chaperoning their refolding during the cooling process? This could also be investigated by increasing the thermal annealing time to find the maximum at 185°C and contrasted with multiple short cycles of heating and cooling. As a side-benefit this would yield practical engineering knowledge to inform bioactive plastic melt extrusion processing guidelines.

Better mechanical properties could be achieved by using intermolecular chaperones that are more miscible in organic solvents and plastics than the inorganic salts used herein. Organic salts or random heteropolymers (RHPs) would achieve far superior nanoscopic dispersion. Perhaps charges could be incorporated onto RHPs to mimic the thermostabilizing effects of ionic salts such as $MgSO_4$ or KCl. Alternatively, 4-component systems of plastic-enzyme-salt-RHP could be developed to achieve thermostability, nanoscopic dispersion, and minimize leaching. GFP is the prime candidate to begin this exploration as if GFP can be thermostabilized in a PLLA-GFP-Mg_2SO_4-RHP system, that would demonstrate the broad thermostabilizing efficacy of Mg_2SO_4 beyond hydrolases and show that ionic salts and RHPs can have synergistic effects. Going down this route would be a combinatorially complex search space and would greatly benefit from in silico and robotic high throughput assays to accelerate exploration. Computational tools to understand how RHPs may interface with protein surfaces would also be extremely valuable to unlock more fine-tuned, rational design. Thus, we enter the second half of this dissertation, in which we focus on developing in silico tools to rationally design RHPs.

3.6 Methods and materials

Film Preparation: Proteinase K from *Tritirachium album* (PK), lipase from *Burkholderia cepacia* (BC), lipase from *Aspergiullus oryzae* (AO), and protease from Bacilus licheniformis (BL) were purchased from Sigma-Aldrich and purified following previously established procedures.[30] Protease from *Streptomyces griseus* (SG) was purchased from Sigma-Aldrich and used as purchased. Green fluorescent protein (GFP) was gifted from the Wenjun Zhang lab at the University of California, Berkeley. Concentration of enzyme stock solutions were made ranging between 0.5-1.5 mg/mL of enzyme in Milli-Q (MQ) water and quantified via absorbance at 280 nm. $MgSO_4$, KCl, NH_4NO_3, NH_4Cl, $MgCl_2$, $Mg(NO_3)_2$, Na_2SO_4, $ZnSO_4$, cobalt (II) acetate, dysprosium (III) nitrate, neodymium nitrate and europium (III)nNitrate were purchased from Sigma-Aldrich. Salt stock solutions were made in MQ water ranging between 10-100 mg/mL. lyophilized Salt and enzyme solutions were mixed at mass ratios between 1:1 and 100:1, frozen at -80 ºC, and lyophilized for a minimum of 8 hours.

Polycaprolactone (PCL), Polystyrene (PS) and Polymethylmethacrylate (PMMA) were purchased from Sigma Aldrich. Polylactic acid (PLLA) polymers were purchased from PolySciTech. Full plastic details can be found in **Table S3.1**. PCL, PLLA and PMMA was dissolved in dichloromethane (DCM) at 4 wt% concentration and stirred for 4 hours to ensure complete dissolution. PS was dissolved in DCM at 6 wt%. The co-lyophilized enzyme and salt complexes were resuspended directly in the polymer solution at 0.1-2 wt% enzyme loading. Mixtures were vortexed for 5 min before 20 µL drops were cast on glass sides to obtain approximately 1 mg films for thermostability assay trials. For larger degradation trials, 20 mg films were cast. All films were air-dried at room temperature. Unless otherwise specified, ~85 kDa PLLA was used for thermostability experiments and ~30 kDa PLLA was used for degradation experiments.

Thermostability Assays: Quadruplicates of 1 mg plastic films were used for each temperature trial and were thermally annealed in a Fischer Scientific Isotemp oven for 5-8 minutes. A thermocouple was also placed in the oven right above the films to get the most temperature (**Fig. S3.1**). For 99% RH steam heat treatment, a hotplate was used with 3x 30 mL beakers filled with distilled water and covered with a glass cover to create a saturated humid air environment. Oven to hotplate temperature calibration was done. 4-Nitrophenyl butyrate (4NPB) was used as a small molecule ester substrate to test enzyme activity for all hydrolase except SG, where N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide (N-Suc) was used instead. Each film was peeled off glass using a razor blade and placed in individual wells in a 96-well plate filled with 200 µL of 0.5 mM 4NPB or 0.7mM N-Suc solution in 1x PBS. Absorbance at 410 nm was measured at 30 second intervals for each film for 30 minutes using an infinite M nano plate reader. A regression was performed on the linear region of the enzyme activity plot to determine initial rates.

PLLA Degradation: ~20 mg films triplicates were thermally annealed in the isotherm oven for 12 minutes, ~6 mg strips were cut and placed in 7 mL glass vials filled with 5 mL of 0.01x Tris pH8.0 buffer. Films were removed from a 40°C water bath after 24 hours, any residual buffer salts were washed with MQ water, lyophilized overnight, and then weighed to determine mass loss.

Chapter 4

Biocheminformatic Design of Biomimetic Heteropolymers

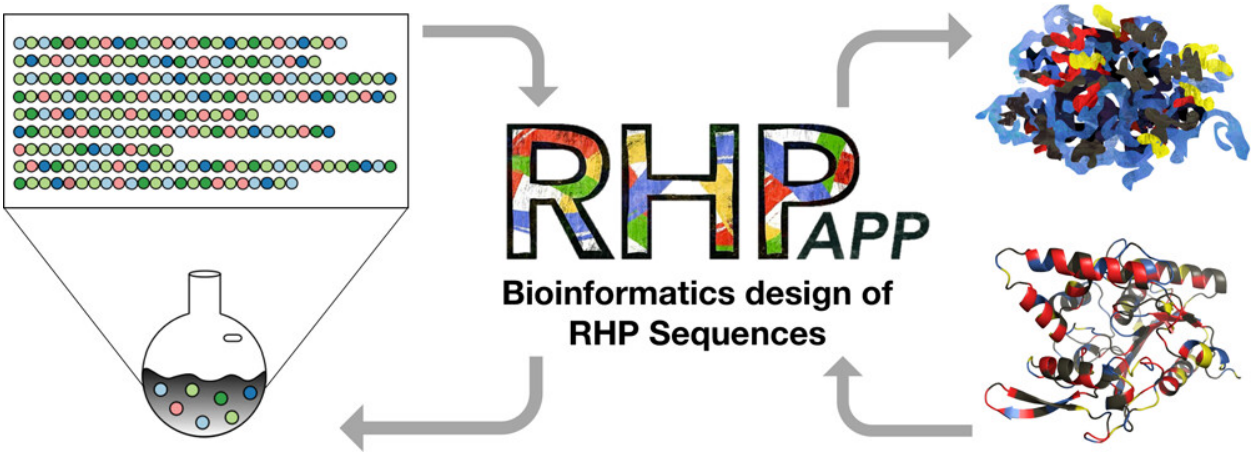
4.1 Heteropolymerization simulation: an orthogonal path to sequencing

Utilizing and mimicking protein function is a key approach to unlocking advanced, robust, cheap, and scalable functional materials. Heteropolymers are routinely used for surfactants,[1–5] hydrogels,[6,7] polyelectrolytes,[8–10] gene delivery,[11–13] and more.[14–17] Chemistry diversification through monomer increments, side-chain modifications, or block copolymerization have been unsystematically explored as the primary design criteria for material functionalization. However, a more general chemical heterogeneity framework for rational design of protein-like heteropolymers is still lacking. Random heteropolymers (RHPs) are composed of more than two monomers, with sequences that are statistically defined. In comparison to proteins, RHPs are synthetic, and polydisperse in molecular weight and composition. RHPs can have batch-to-batch variations but cannot yet be sequenced with monomeric specificity. Despite key differences, several computational and experimental studies have demonstrated the ability for RHPs to recapitulate protein-like behaviors.[18–30] Unlike sequence-specific heteropolymers,[31–34] the lack of RHP sequence information significantly hampers our ability to further leverage the full potential of this unique class of polymers for precisely tailored functionality. Synthetic breakthroughs in reversible-deactivation radical polymerization (RDRP) have made it possible to synthesize heteropolymers with improved reproducibility and control over the probability of each monomer along the polymer chain.[35–43] This narrows the gap between theoretically ideal polymerization and synthesized heteropolymers.

Since the 1940s there have been numerous efforts to simulate heteropolymers using experimental inputs such as monomer concentrations and reactivity ratios.[44–54] Among those, Compositional Drift, a Monte Carlo method RDRP simulator, has been developed to simulate RHP sequences based on the Mayo–Lewis model.[55] RDRP synthesis input parameters provide direct handles to tune key batch level characteristics such as monomer composition and average degree of polymerization. Synthetic control of batch level properties can be experimentally verified using common instrumentation such as nuclear magnetic resonance spectroscopy (NMR) and gel permeation chromatography (GPC). Matching these experimentally measurable, batch level, key characteristics in reality and theory allows for simulated outputs to be useful once abstracted from the level of monomeric precision to an analysis on the batch level of statistical distributions and sequence patterns. Prior works have shown the use of RDRP synthetic parameters to design RHPs with statistically controlled sequences as protein mimics with a wide variety of promising applications inclusive of enzyme stabilization, biodegradable plastics, and selective ion transport.[4,6,10,25,26,55,56]

Sequence analysis is routinely performed for proteins to evaluate statistical distributions of residues along a chain, identify key motifs, and assess similarity across

proteins.[57–62] Applying similar analysis to RHPs will advance our ability to design functional polymeric materials. Here we demonstrate a bioinformatics-inspired sequence analysis on batches of simulated RHP sequences. Specifically, we present the RHPapp (**Scheme 4.1**), a more comprehensive version of the Compositional Drift simulation software, integrated into a suite of analytical modules for RHP design and analysis. Through the RHPapp, the synthesis of batches of RHPs with varying experimentally measurable, batch level characteristics are simulated. Common methods such as binarizing sequences by hydrophobicity and plotting hydropathy are applied to the simulated RHP sequences. To fully understand the heterogeneity of RHPs, we highlighted the importance of analyzing simulated outputs at multiple levels of abstraction: single chain segments (segment level), across sequences (sequence level), and across batches (batch level). Current results revealed that given appropriate evaluation metric selection and segment length definition, sequence analysis on simulated RHP sequences can help to rationalize experimental findings, guide subsequent experimental design in an iterative fashion, and realize designed function without the need for full sequence specificity or sequencing technology.



Scheme 4.1 | Schematic introducing the RHPapp: a polymer sequence simulation and analytics software that offers a design framework for functional heteropolymers e.g., to mimic protein function.

4.2 RHPapp 1.0 (Methods)

We simulated RHP polymerization over ranges of values for each input parameter to RHPapp and compared batches across various modular metrics for evaluation and 3 levels of statistical heterogeneity: (1) single chain segments (segment level), (2) across sequences (sequence level), and (3) across batches (batch level). Target % conversion and polydispersity (PDI) of batches of simulated RHP sequences are tunable parameters. However, for all RHP sequences in this work conversion is fixed to 50% and polydispersity (PDI) is kept low (below 1.2) to reflect previous experimental results. Reactivity ratios used in this work are presented in **Table S4.1**. All oligomers (sequences of degree of polymerization (DP) < 15) are neglected, as they are removed experimentally in the purification process of RHP synthesis, as is described in prior works.[25] Currently, 4 evaluation metrics have been implemented as Python modules, with which RHP sequences are analyzed. These metrics are modular; thus, the addition of new metrics as future work is straightforward.

Chemical heterogeneity: analysis of how monomers along the sequence vary. Subsets of sequences from simulated batches are visualized for segment level comparison along a single chain (**Fig. 4.1a**). On the sequence level, kernel density estimate (KDE) plots show the distribution of monomers across all chains in a single batch. A curve fitted to a histogram is plotted for each unique monomer to show the distribution of monomer fractions on each chain in a single batch of simulated polymer (**Fig. 4.2a**). For each batch, the full width at half-maximum (FWHM) of the peaks in KDE plots is calculated, normalized by each monomer feeding fraction (nFWHM), and visualized on a single scatter-line plot (**Fig. 4.2b**) The *x*-axis of nFWHM batch-to-batch level comparison plots can be varied to probe trends across different input parameters such as number of sequences and average degree of polymerization.

Segmental hydrophobicity: sequences are binarized into hydrophobic or hydrophilic monomers and grouped into segments. The hydrophile–lipophile balance (HLB) value was used to evaluate the solubility of monomer side-chains through group contribution theory. Using the equation $HLB = 7 + \sum_i n_i HLB_i$, where n_i is the number of the ith chemical group in the molecule with corresponding value HLB_i. The HLB value for each monomer side chain used in this work was estimated as HLB[methyl methacrylate (MMA)] = 8.45, HLB[2-ethylhexyl methacrylate (EHMA)] = 5.12, HLB[poly(ethylene glycol) average M_n 500 (OEGMA)] = 11.4, HLB[3-sulfopropyl methacrylate potassium salt (SPMA)] = 18.5, and HLB[styrene (STY)] = 4.865. Lower HLB values indicate higher hydrophobicity, and a higher value means greater hydrophilicity. A hydrophilic–hydrophobic cutoff value (HLB-threshold) of 9 was set to distinguish hydrophobic and hydrophilic monomers. A hyrophobic (or hydrophilic) segment is a contiguous run of hyrophobic (or hydrophilic) monomers. These segments can be visualized on sampled subsets of binarized sequences (**Fig. 4.1b**). Sequence level heterogeneity is analyzed by counting the hydrophobic segments of each block length on each chain. The number of hydrophobic segments is then averaged per chain. For visualization, the average frequencies per chain for select segment lengths (i.e., 1, 3, 5, and 10 monomers long), which is the same as the block length for this analysis, are plotted (**Fig.**

4.2c). Segment length and count frequencies on each chain can also be summed (rather than averaged) across all chains in a batch for a batch level heterogeneity comparison of total segment distributions (**Fig. 4.3b**).

Sliding window analysis: sliding window analysis is routinely used for protein sequence analysis to reduce random noise and obtain coarse-grained but more obvious characteristics at the segment level. We thus applied a level of convolution to RHP sequences prior to segmental hydrophobicity analysis. Average segmental HLB values are continuously calculated for a window sliding from the alpha to the omega ends of the simulated RHP chains. The window is advanced by one monomer each time. We used a span containing odd numbers of monomers and assigned the average HLB value of that span to its middle monomer. Various window sizes of 5, 9, and 15 were adopted from previous works to study the effects of small, medium, and large numbers of neighbor monomers, respectively.[57–60] Hydropathy plots were generated to visualize randomly sampled sequences for each RHP composition and window size (**Fig. S4.1**). The hydrophobic/hydrophilic segments were identified using the same definition from Metric 2, except that the HLB value for each monomer was replaced with window-averaged values. At the sequence level, hydropathy plots were averaged across all chains within a batch and distribution statistics were plotted (**Fig. S4.2**).

Specific segment search: distributions of specific segments of interest are analyzed. This metric is similar to Metric 2 in that sequence level analysis averages chain distributions and batch level comparisons sum the distributions. However, instead of hydrophobic/hydrophilic segments, segments are specifically defined by a desired chemical (monomer) pattern. An example module has been implemented that searches for hydrophobic segments containing 1 embedded OEGMA monomer that is 2 or more monomers away from the end of the segment. For sequence level analysis, manually selected segment lengths of 5, 8, 10, and 13 are plotted. For batch level analysis, kernel density distribution of the specific segments for the given batch are plotted (**Fig. 4.5c**).

Any combination of evaluation metrics may be used for a given sequence analysis. To generate reaction schematics that physically realize this desired sequence distribution, the corresponding input parameters and additional information about the reaction scale, monomer molecular weight, monomer density, initiator, chain-transfer agent, and solvent, are solved in a system of equations. Accordingly, the required volumes and masses of each reagent are output.

Protein sequences can be analyzed through the same workflow for direct comparison to the RHP sequences. A dictionary is created mapping amino acids into groups of roughly corresponding RHP monomers by hydrophobicity/hydrophilicity/charge. The dictionary is applied to convert protein primary sequences into RHP-monomer equivalent sequences – a dimensionality reduction from an alphabet size of 20 (AA residues) to between 2 and 5 (synthetic monomers). A specific protein sequence is then split into overlapping segments governed by a segment length = 100 and offset (spacing between segment start monomers) = 10. An example is shown in **Figure 4.1c**. These segments are passed in as simulated RHP-sequence equivalents into the workflow for direct comparison.

All code used for calculation and visualization in this work are provided as an open-source repository (https://github.com/ivanjayapurna/RHPapp), to serve as a tool for community use.

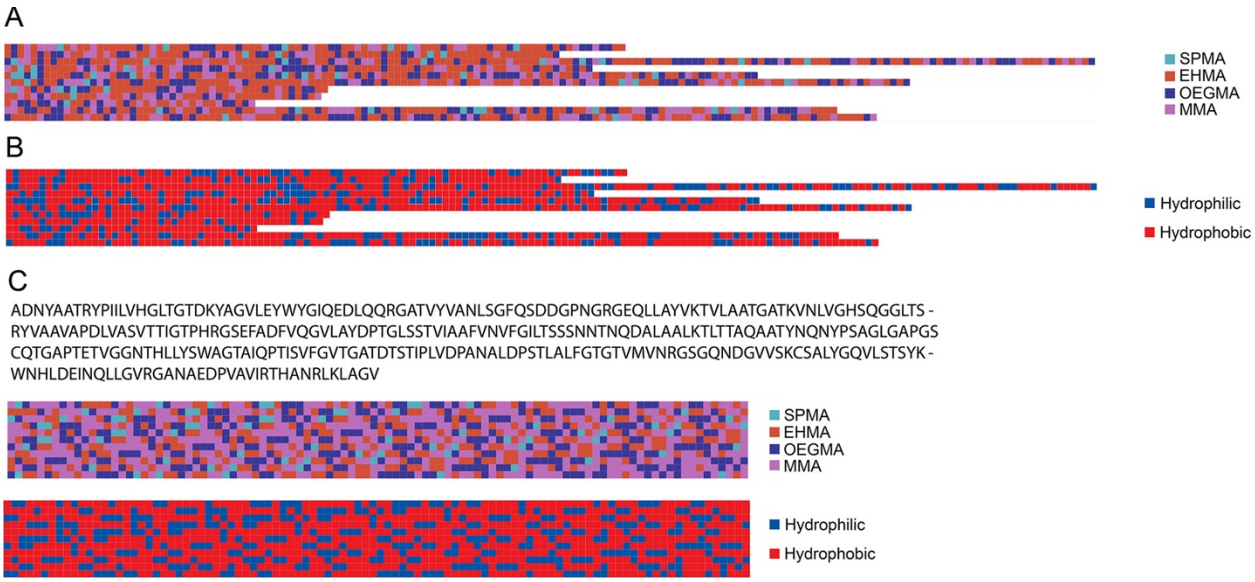


Figure 4.1 | Random heteropolymer (RHP) and protein sequence comparison. (**A**) Random sequences sampled from a simulated 4 monomer RHP of 50% methyl methacrylate (MMA), 25% poly(ethylene glycol) average M_n 500 (OEGMA500), 20% 2-ethylhexyl methacrylate (EHMA), and 5% 3-sulfopropyl methacrylate potassium salt (SPMA) (**B**) RHP sequences binarized to hydrophobic and hydrophilic. (**C**) Full sequence for *Burkholderia cepacia* lipase (BC-Lip), segmented and translated to RHP sequence space and then binarized to hydrophobic and hydrophilic units.

4.3 Simulation scale and fidelity

The entire premise of RHP design by simulation assumes a controlled link between synthetic design and actualized statistical monomer distribution. RHP sequence simulation can only be insightful when a polymerization retains its livingness such that synthesis is predictable. To experimentally verify synthetic control, we conduct routine characterizations on synthesized materials by nuclear magnetic resonance spectroscopy (NMR) and gel permeation chromatography (GPC). With these two common characterization techniques we confirm in our samples (1) bounded polydispersity that confirms good RDRP control, (2) reaction conversion percentage that confirms no compositional drift, (3) achieved targeted molecular weight, and (4) approximate composition percentages (**Fig. S4.3**). However, an important caveat is that both in-lab and in-silico experiments only probe a tiny subpopulation of statistically possible RHP sequences. To illustrate the scale of our materials in number of polymer chains, let us assume the synthesis of 1 g of a hypothetical methacrylate based RHP, with an average monomer molecular weight of 100 g per mole and an average degree of polymerization (DP) of 100. This would yield on the order of 10^{18} polymer chains synthesized. GPC or NMR will use on the order of 1 mg of sample, which is on the order of 10^{15} chains. Thus, when characterizing with GPC or NMR, we assume that an approximately 0.1% sample is representative of the total batch synthesized. A similar approximation will need to be made computationally.

A key parameter in stochastic (Monte Carlo) simulators is the minimum sample size required for simulated results to converge to a stable value. For the RHPapp, the key parameter is the number of chains simulated (NC). Computational performance limits our ability to simulate 10^{18} chains, an approximate magnitude of a real synthesis of a batch of RHP. A similar subsampling approximation as was made for experimental characterization (roughly 3 orders of magnitude lower) must be done for in-silico characterization. In a good stochastic simulation of polymerization, results should converge to the same, stable value regardless of how many sequences are simulated. The optimal NC is the minimum required to converge to a stable distribution of results to maximize simulation accuracy and minimize computational time. As an example, to illustrate finding an NC minimum for a batch of 4-monomer methacrylate-based RHPs of DP of 100, we simulated multiple batches with varied NC while keeping all other simulation parameters fixed.

When using chemical heterogeneity as a metric, there is visible lack of smoothness in the KDE fitting when only 100 sequences are simulated, when compared to 15000 and 100000. However, the important high-level features across all 3 batches of varying NC such as peak location, height, and width remain similar (**Fig. 4.2a**). The normalized full width at half-maxima (nFWHM) of all peaks from the KDE plot were estimated (**Fig. 4.2b**). For an average DP of 100, although the differences in chemical heterogeneity are minor between the number of sequences simulated, the monomer of lowest feed ratio (SPMA) has a convergence of initial nFWHM oscillation at around NC = 15000. This minimum threshold is approximately 14 orders of magnitude lower than a real experimental RHP synthesis and is acceptable as it requires minimal compute power to simulate at NC ≥ 15000. The minimum NC is a parameter intrinsically linked to co-input parameters and may vary significantly when

other parameters change, such as molar feed ratios or number of monomers. However, in some cases, such as increasing or decreasing average DP of our example system and keeping all else constant, the minimum NC is similar. Although the actual nFWHM values for average DP 50 and 300 differ significantly from those of DP 100, all initial oscillations stabilize at a similar NC threshold.

The minimum NC can also be estimated using sequence hydrophobic/philic segmental distributions, as seen in the binarized sequences in **Figure 4.2b**. Comparing different NC simulations at the sequence level within batches of DP 100 RHPs, there is no shift in the primary peak mean or heights in the average frequency per chain distributions of hydrophobic and hydrophilic segments. However, there is a noticeable gain in smoothness of fits and disappearance of misleading minor peaks as NC increases (**Fig. 4.2c**). As the increase in fit smoothness is negligible between 15000 and 100000 NC, the segmental hydrophobicity metric at the sequence level suggests NC = 15000 is sufficient for stable, accurate simulation distributions. Analysis at the batch level shows a negligible difference in batch level frequency and distribution of segments normalized by number of sequences (**Fig. S4.4**). Thus, NC was set to 15000 for all 3 subsequent example use-cases using the same 4 monomer methacrylate RHP presented in this work.

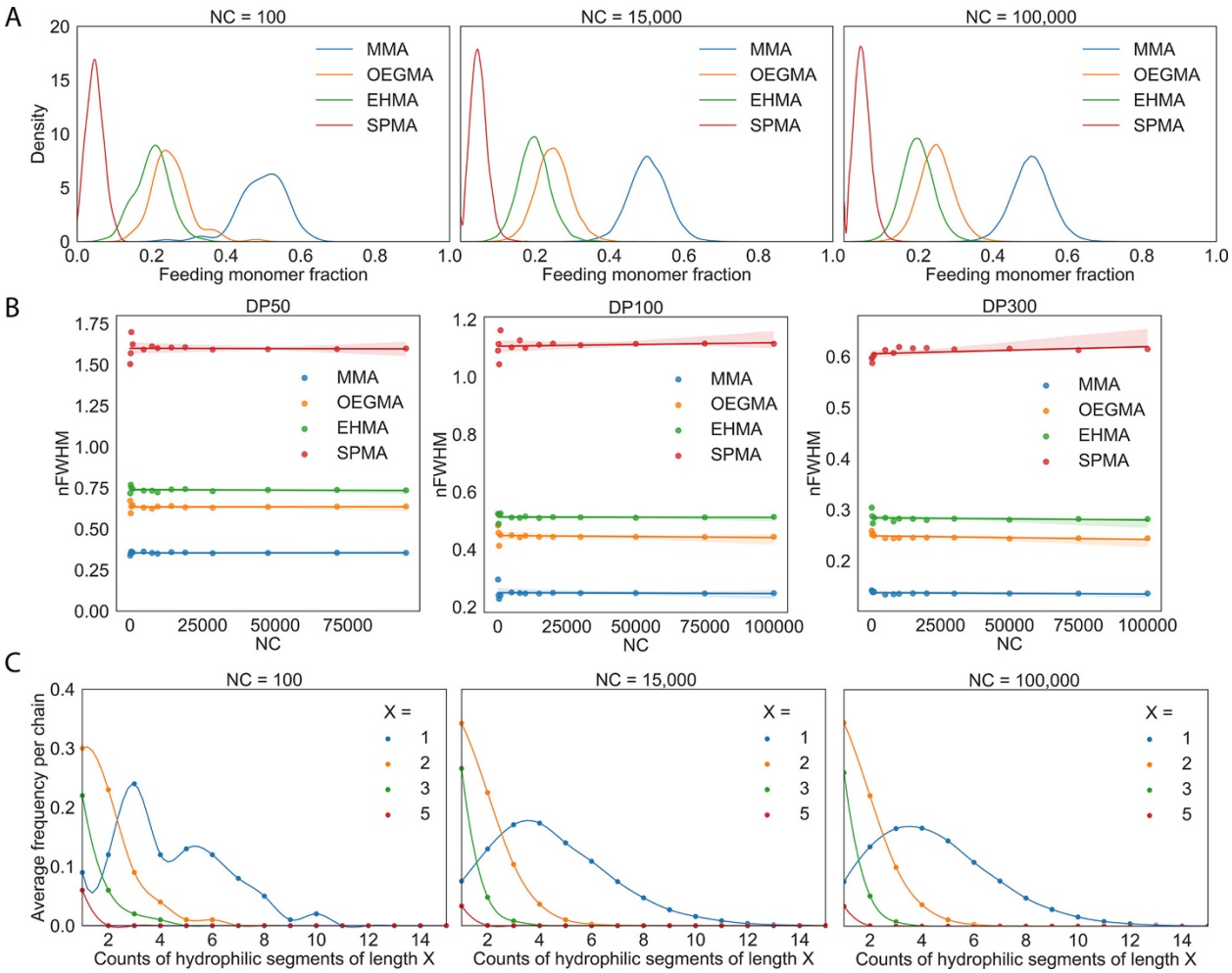


Figure 4.3 | Varying the number of chains (NC) simulated, for a simulated 4 monomer RHP of 50% methyl methacrylate (MMA), 25% poly(ethylene glycol) average M_n 500 (OEGMA), 20% 2-ethylhexyl methacrylate (EHMA), and 5% 3-sulfopropyl methacrylate potassium salt (SPMA). (A) Sequence level monomer distributions for batches of NC = 100, 15000, and 100000 (left to right). (B) Normalized full width at half-maximum (nFWHM) plots of batch level monomer feed ratio distributions for increasing degree of polymerization (DP) of 50, 100, and 300 (left to right). (C) Sequence level hydrophobic segment distributions, highlighting segments of lengths 1, 2, 3, and 5 for batches of NC = 100, 15000, and 100000 (left to right).

4.4 Sequence design of protein-interfacing random heteropolymers

Panganiban et al. proposed that 4 monomer RHPs can stabilize proteins in aqueous and organic solutions when they have both (1) chemical heterogeneity and (2) hydrophobic/hydrophilic block length and count distributions that mimic those of intrinsically disordered proteins.[25] We used simulated RHP sequences to decouple these two hypotheses and assist in the design of RHP sequences of a varying number of unique monomers that still retain the same hydrophobic/hydrophilic segment distributions as the original 4 monomer RHPs. The two fixed monomers are MMA and OEGMA, with EHMA added for monomer 3, SPMA for monomer 4, and STY for monomer 5. Styrene was chosen to represent a common non-methacrylate monomer to demonstrate the potential for monomeric diversity in this design framework. Initial guesses of appropriate monomer feeding ratios for each RHP were made to approximate similar hydrophobic/hydrophilic segment length and count distributions. **Figure 4.3a** shows that the sequence level chemical heterogeneities of RHP batches of varying monomers are vastly different. However, despite the disparity in chain chemistry, at the binarized hydrophobic/hydrophilic level, the segmental distributions for the 2 and 4 monomer RHP batches are nearly identical, suggesting no need to alter initial guess monomer feeding ratios (**Fig. 4.3b**). This evaluation metric was used to further fine-tune the designs of the 3 and 5 monomer RHPs for a more precise distribution match. An iterative design approach on our initial guesses of the 3 and 5 monomer RHPs enabled precise monomer feeding ratio alterations to preserve the targeted chemical heterogeneity (**Fig. 4.3d**) while minimizing the disparity in theoretically predicted hydrophobic/hydrophilic segment length distributions between each RHP batch (**Fig. 4.3c**). Thus, when designing an RHP synthesis for a materials application, analysis of simulated RHP sequences can serve as an in-silico prescreening, to inform and accelerate the rational design of compositions to experimentally characterize.

DelRe et al. demonstrated that RHPs can nanoencapsulate and preserve the activity of enzymes in solid polymeric matrices. RHP composition was shown to regulate substrate binding and active site availability.[4] However, only a few RHP compositions were tested, due to the lack of high throughput material synthesis and characterization. Simulated sequence analysis can assist in data deficient modeling, analysis and serve as a useful tool to suggest explanations for experimental findings. The 4 monomer RHP used has 2 hydrophobic and 2 hydrophilic monomers, giving several handles to tune. The first we chose to modulate is the MMA:OEGMA ratio. MMA is our proxy for segmental hydrophobicity and OEGMA for segmental hydrophilicity. Tuning this handle yields clear differences in segment length distributions (**Fig. 4.4a**) that could be compared to differences in enzyme nanoencapsulation behavior based on enzyme surface hydrophilic and hydrophobic patch distribution patterns.

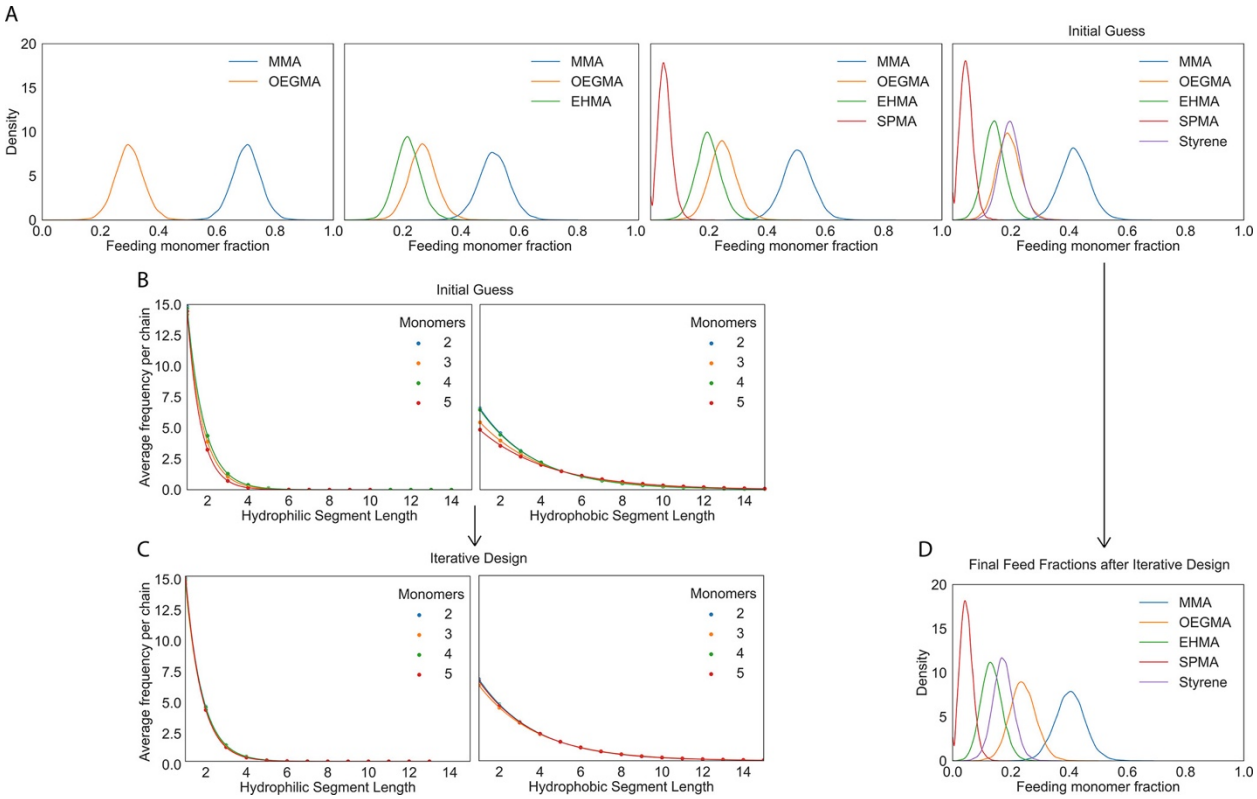


Figure 4.3 | Varying the number of unique monomers. For simulated RHP batches of average degree of polymerization 100 with varying number of unique monomers from 2 to 5 of methyl methacrylate (MMA), poly(ethylene glycol) average M_n 500 (OEGMA), 2-ethylhexyl methacrylate (EHMA), and 3-sulfopropyl methacrylate potassium salt (SPMA). (A) Sequence level monomer distributions for each simulated batch with initial guess monomer feeding ratios of 70:30, 51:27:22, 50:25:20:5, and 45:20:15:5:15 (left to right). (B) Batch level hydrophilic and hydrophobic segment distribution heterogeneities of initial guess monomer feeding ratios. (C) Batch level hydrophilic and hydrophobic segment distribution heterogeneities after iterative adjustment of monomer feeding ratios to 70:30, 50:29:21, 50:25:20:5, and 43:25:14:5:13 (left to right). (D) Sequence level monomer distributions for iteratively designed 5 monomer RHP with 43:25:14:5:13 feeding ratio.

However, polymer–protein interactions are sensitive and complex. Rather than leading to improved performance, too drastic of a change in RHP composition could overshoot the scale of differentiation between enzyme chemical distributions resulting in worse chaperone performance or even polymer gelation issues as not all RHP compositions can be synthesized.[29] To tune with higher sensitivity the MMA:EHMA ratio can be varied. Although both monomers are considered hydrophobic by the HLB threshold parameter of 9, this binarization is just an artifact of analysis. A lower HLB threshold that would split EHMA and MMA could be chosen that would yield different results. In the current analytical setup, adjusting the MMA:EHMA ratio has no apparent effect (**Fig. 4.4b**), contrary to experimental results. To more subtly fine-tune using a method that is more robust to threshold parameter

selection, we redefine what it means to be a hydrophobic/hydrophilic segment. A level of sliding window convolution prior to binarizing into contiguous hydrophobic/hydrophilic segments adds an abstraction layer from monomer specificity, which is inherently stochastic and noisy. Sliding window analysis allows us to loosen the rigid prior definition of what is considered a segment. The results of the analysis suggested that, within a batch, window average HLB distributions are invariant to central monomer position along the chain, except for increased variance at the omega end of simulated chains, where due to polydispersity there are fewer data points to average and converge to the expected statistical distribution (**Fig. S4.2**). Thus, sequences can be further averaged across positions along the chain to make cleaner batch-average segmental (window) hydrophobicity comparisons. Differences can be observed in batch level segmental distribution statistics, where segments are now defined by window average HLB values (**Fig. 4.4c**). The mean values and trend of increasing window average HLB as the MMA:EHMA ratio increases are consistent across varying window sizes. Although variance reduces with increasing window size, here we have demonstrated that applications where we primarily consider the resulting trends in average values, such as this RHP–enzyme interaction analysis, are invariant and thus robust to the range of window sizes selected.

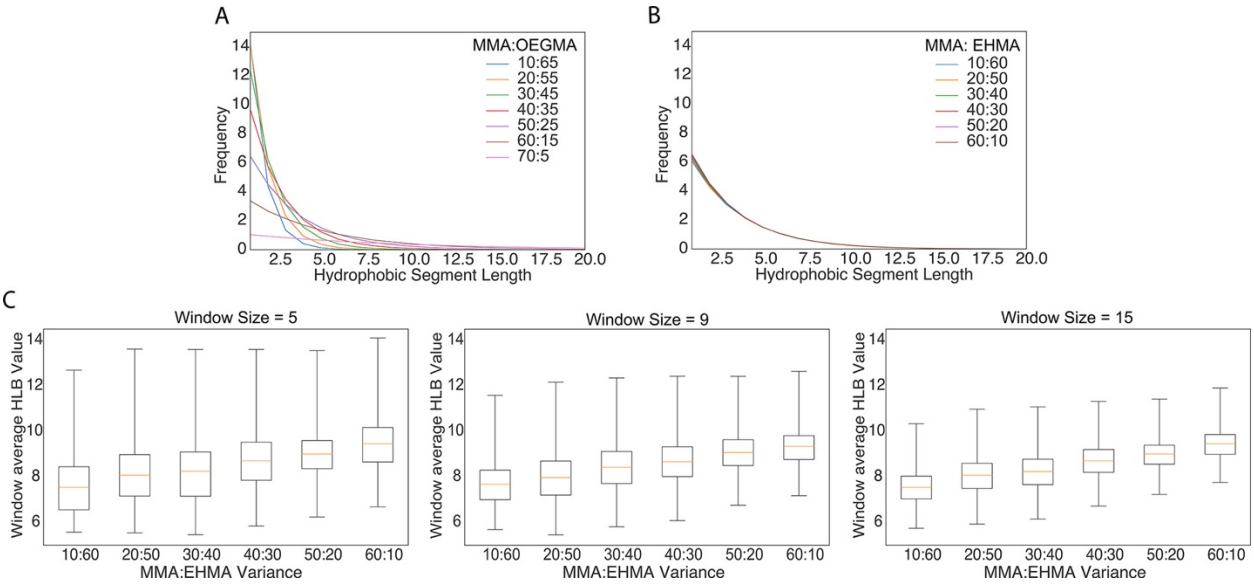


Figure 4.4 | Varying the monomer feed ratio. Batch level hydrophilic segment distribution heterogeneities for a simulated 4 monomer RHP of methyl methacrylate (MMA), poly(ethylene glycol) average M_n 500 (OEGMA), 2-ethylhexyl methacrylate (EHMA), and 3-sulfopropyl methacrylate potassium salt (SPMA) of degree of polymerization 100: (A) varying MMA:OEGMA feed ratios and fixed 20% EHMA and 5% SPMA; (B) varying MMA:EHMA feed ratios and fixed 25% OEGMA and 5% SPMA. (C) Batch level statistics after sliding windows of sizes 5, 9, and 15 (left to right) are applied, and the resulting sequence level segment information is averaged by sequence position.

4.5 Sequence design of heteropolymers with biomimetic function

Protein sequences can be convolved into corresponding RHP sequence space, segmented to form a batch, and then analyzed through the RHPapp workflow for direct comparison to RHP sequences. Six sample proteins (3 enzyme hydrolases, 2 transmembrane proteins, and 1 β barrel structure protein) were analyzed through the RHPapp to demonstrate facile comparison to a 4 monomer methacrylate-based RHP (**Fig. S4.6**). On the sequence level, proteins each have characteristic chemical heterogeneities despite functional and evolutionary similarity. Membrane proteins AquaporinZ and PepTSo have different chemical heterogeneities despite similar function (**Fig. 4.5a**). Despite the chemical heterogeneity differences, all 6 proteins display similar hydrophobic segment distributions on the sequence level (**Fig. 4.5b**). This may suggest a degree of generality in the design of RHPs, as was demonstrated by Panganiban et al. where a single RHP design stabilized various proteins in solutions. Chemical heterogeneity differences may explain why DelRe et al. observed that different RHPs were required for optimal nanoencapsulation of different hydrolases (**Fig. S4.5**).

In addition to interfacing with proteins as binders, Jiang et al. showed RHPs can independently mimic membrane protein function to undergo rapid and selective proton transport across lipid bilayers at a rate similar to those of natural proton channels.[26] Specific RHP segments of critical importance to recapitulating transport function are hydrophobic segments containing 1 embedded OEGMA monomer that is 2 or more monomers away from the end of the segment. All functional chains contained this pattern within their random sequence. This functional RHP of DP 100 has a different chemical heterogeneity profile compared to both membrane proteins Aquaporin Z and PepTSo (**Fig. 4.5a**), suggesting global chemistry is nonessential to mimicking protein function. A more local analysis of the proton transport specific segment pattern was done comparing the RHP to 6 proteins at the batch level (**Fig. 4.5c**). Of the sampled proteins, only the 2 membrane proteins display a similar distribution to the functional RHP. Specifically, the distributions suggestive of selective proton transport function display a lower density peak at specific segment length 9 and a longer tail with sizable population between 20 and 30 segment length. This example demonstrates the comparison of statistical distributions of specific segment motifs on RHPs to those on proteins to explain experimentally the observed protein-like function.

Here we demonstrated a viable path toward guiding the rational design of RHPs as synthetic protein mimics through the combination of RDRP polymerization simulation and bioinformatics-inspired sequence analysis. The RHPapp, more than an open-source toolkit, is a design and analysis approach that can be applied to a diverse range of impactful projects well beyond the methacrylate backbone monomers presented herein. More complex random heteropolymers with monomers of varied reactivity ratios could further enhance control by incorporating partial blocky features into the statistical randomness to create more complex repeating motifs similar to those found in natural proteins. Mirroring the nature of RHPs, the RHP design approach and software presented are general in principle for broad applicability, but also modular and easily fine-tuned to suit projects with exact specificity. Our vision is for the RHPapp to take as input a protein sequence and desired

function as a starting point, from which an ideal sequence distribution to target can be designed and translated into controllable RDRP synthetic parameters to be experimentally realized. An example demonstrating this inverse design paradigm for RHP sequences is shown in **Chapter 6**. This analysis framework for simulated heteropolymer sequences couples powerfully with advances in high throughput synthesis and characterization. Just as bioinformatics, the inspiration for our work, has trended toward the realms of big data and more sophisticated statistical modeling techniques, analysis, and machine learning, we propose that this emergent field of macromolecular cheminformatics is ripe to follow suit.

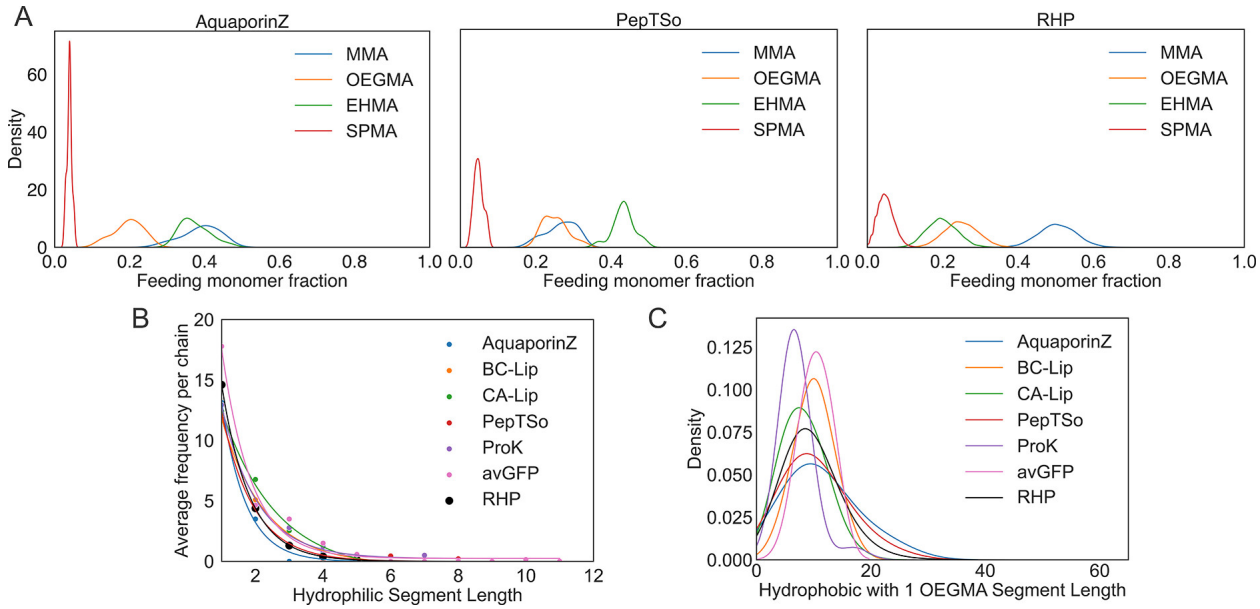


Figure 5 | Protein and RHP comparison. 6 proteins were convolved into RHP sequence space, segmented to form a batch, and analyzed through the RHPapp. Proteinase K (ProK), lipase from *Burkholderia 73ictoria* (BC-Lip), lipase from *Candida antarctica* (CA-Lip), Aquaporin Z (AquaporinZ), POT family transporter (PepTSo), and green fluorescent protein from Aequorea 73ictoria (avGFP) were compared to a simulated 4 monomer RHP of 50% methyl methacrylate (MMA), 25% poly(ethylene glycol) average M_n 500 (OEGMA), 20% 2-ethylhexyl methacrylate (EHMA), and 5% 3-sulfopropyl methacrylate potassium salt (SPMA). All sequences are of average degree of polymerization 100. (A) Sequence level monomer distributions for AquaporinZ, PepTSo, and RHP (left to right). (B) Batch level hydrophilic segment distributions. (C) Batch level kernel density estimate of hydrophobic segments with 1 OEGMA per chain.

Chapter 5

Improving Heteropolymer Sequence Fidelity

5.1 Limitations of RHPapp 1.0

Recent works have shown that 4 methacrylate monomer RHPs of mostly similar compositions and molecular weights have been able to stabilize enzymes, assist in protein folding, unlock processive polyester degradation and even directly mimic protein functions and mixtures.[1–6] However, more complex applications and functionalities of RHPs will require more complex RHPs with increased monomer variety. To rationally design more advanced synthetic polymers, we need more advanced design tools. For example, in biological polymers, DNA sequencing, bioinformatics and publicly collaborative databases of millions of sequences unlocked the explosion of rational design in genetics and protein engineering.[7] Unfortunately, like most synthetic polymers, we cannot yet sequence RHPs.

In the previous chapter we showed that orthogonal to sequencing, owing to synthetic breakthroughs in reversible-deactivation radical polymerization (RDRP),[8–11] we can instead use a simulator to generate polymer sequences *in silico*. Many RDRP simulators both deterministic and stochastic exist, either focusing on modeling accurate kinetics or sequence composition.[12–22] However, most published simulators are not open-source, and of these simulators, to the best of our knowledge the Mayo-Lewis equation based *CompDrift* that has been incorporated into the *RHPapp*, are the only simulators that generate explicit sequences for heteropolymers.[23–25] These simulators minimally consider kinetics, taking as input only monomer feeding ratios, reactivity ratios, target conversion and average degree of polymerization. Despite this, the *RHPapp* has shown wide utility in experimental data analysis, RHP design, and generating input data for various downstream computational tasks ranging from bioinformatics sequence analysis to hidden Markov modelling to molecular dynamics to deep learning.[5,25–29] However, the results of these analyses are only as accurate as the generated sequences used as input. As heteropolymer complexity increases, a sequence simulator with more complexity would provide heightened trust in sequencing fidelity.

In this chapter we highlight three of many remaining areas in which *in silico* sequence simulation can increase in fidelity to physical reality. Specifically, we build the *RHPapp2.0*, a more accurate heteropolymer sequence simulator that incorporates more accurate experimentally measured reactivity ratios, factors in composition evolution during synthesis and purification, and models RDRP polymerization kinetics during sequence generation.

5.2 Determining reactivity ratios in RAFT heteropolymerization[1]

In heteropolymerization, the intrinsic reactivities and local compositions of each monomer (n) determine the arrangement of monomers along a heteropolymer chain. Leveraging the intrinsic reactivity amongst comonomers allows for tuning segmental sequences and biasing the flanking sequences of specific residues. However, determining the reactivity ratios of comonomers becomes challenging when the number of monomers exceeds three, due to the complex, interdependent reaction kinetics and the mathematical models involved.[30–32] In *RHPapp 1.0* and other prior works, reactivity ratios obtained from binary copolymerization pairs have been commonly used.[33–35] However, neglecting competitive monomer addition in heteropolymer (n>2) systems can lead to substantial deviations between experimental results and theoretical predictions.[36,37] Thus, accurate determination of reactivity ratios is essential for accurate simulation of heteropolymer sequences and their computationally guided design and analysis.

A 4-monomer system was used as a platform for investigation, consisting of: methyl methacrylate (MMA), 2-ethylhexyl methacrylate (EHMA), oligo(ethylene glycol) methyl ether methacrylate (OEGMA; average $M_n \sim 500$ Da), and methacrylic acid *N*-hydroxysuccinimide ester (NHSMA) (**Fig. 5.1a**). MMA, EHMA, and OEGMA have been routinely used to mimic amino acids in prior RHP applications.[5,25] 3-Sulfopropyl methacrylate potassium salt (SPMA) was previously used as the fourth monomer. However, the chemical shifts of SPMA and EHMA largely overlap in the NMR spectra. Thus, NHSMA was chosen to replace SPMA as the fourth monomer to facilitate facile characterization. NHSMA has a pendant imide group and may exhibit different reactivity compared with the other three monomers. This can potentially boost composition drift during copolymerization, presenting supplementary measures to tune the monomer distribution in RHPs.

To account for competitive monomer addition in 4-monomer heteropolymerization, the Jaacks method was used to determine all 12 reactivity ratios (**Fig. 5.1b**).[38] This method applies integrated Mayo–Lewis equations to heteropolymerization (n>2) systems, and allows for the simultaneous determination of multiple reactivity ratios in a single 4-monomer heteropolymerization experiment while accounting for the influence of competitive monomer addition. We performed 4-monomer RAFT polymerizations with one monomer in large excess (molar ratio: 20:1:1:1) in each experiment. To account for the influence of RAFT equilibrium on the propagation kinetics, we conducted these experiments under the same conditions as those commonly used in RHP RAFT polymerization {using 4-cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl]pentanoic acid as CTA and azobis(isobutyronitrile) (AIBN) as initiator at 80 °C in DMF}.

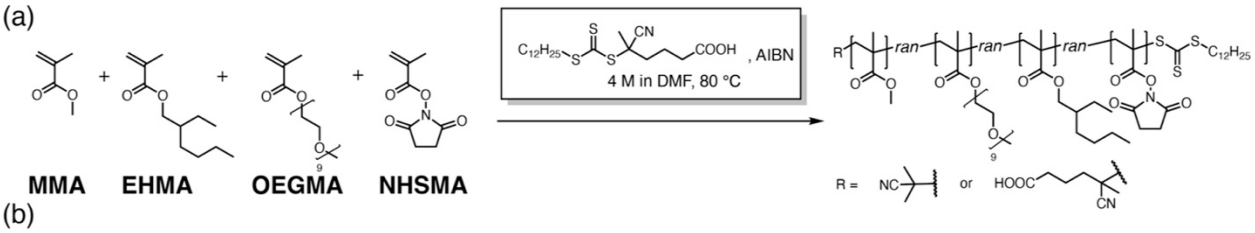
The four selected monomers exhibit distinct reactivity ratios, with values ranging between 0.4 and 2.0. In the four-monomer RAFT copolymerization, cross-propagation of the growing chain end toward MMA is more favorable than self-propagation ($r_{\text{non-MMA monomer/MMA}} <$

[1] The experimental data reported in Chapter 5.2 was collected by Dr. Hao Yu and Ruilin Yin

1), whereas self-propagation is generally preferred over cross-propagation toward OEGMA ($r_{EHMA/OEGMA}$ and $r_{NHSMA/OEGMA}$ > 1). These results suggest the important influence of the side-chain steric effects, with MMA having the smallest side group and OEGMA featuring a bulky side-chain. Notably, NHSMA can lead to more significant composition drifts. Polymer chains with terminal NHSMA group undergo self-propagation approximately two times faster than cross-propagation toward EHMA ($r_{NHSMA/EHMA}$ = 2.0), but 30% slower toward MMA ($r_{NHSMA/MMA}$ = 0.7). Therefore, in RHP copolymerization, a higher prevalence of NHSMA-MMA segments can be generated compared to NHSMA-EHMA or NHSMA–NHSMA segments. Prior work has demonstrated that leveraging intrinsic monomer reactivity offers a facile method for governing segmental sequence.[39,40] Additionally, employing a sequential addition approach can afford a more precise level of control over monomer distribution.[41,42] We envision that the investigation on monomer reactivity ratios here lays a foundation for effectively regulating comonomer distribution in RHPs, with adaptable applicability to various heteropolymer (n>2) systems.

In *RHPapp 1.0,* reactivity ratios were taken from binary systems, which for methacrylate based RHPs come out to be ~1 for almost all combinations. Here we showed that was an inaccurate assumption, at least for this specific system, that will significantly impact sequencing fidelity. For accurate sequence generation, reactivity ratios must be measured for the exact system under exact conditions rather than taken from literature.

(a)



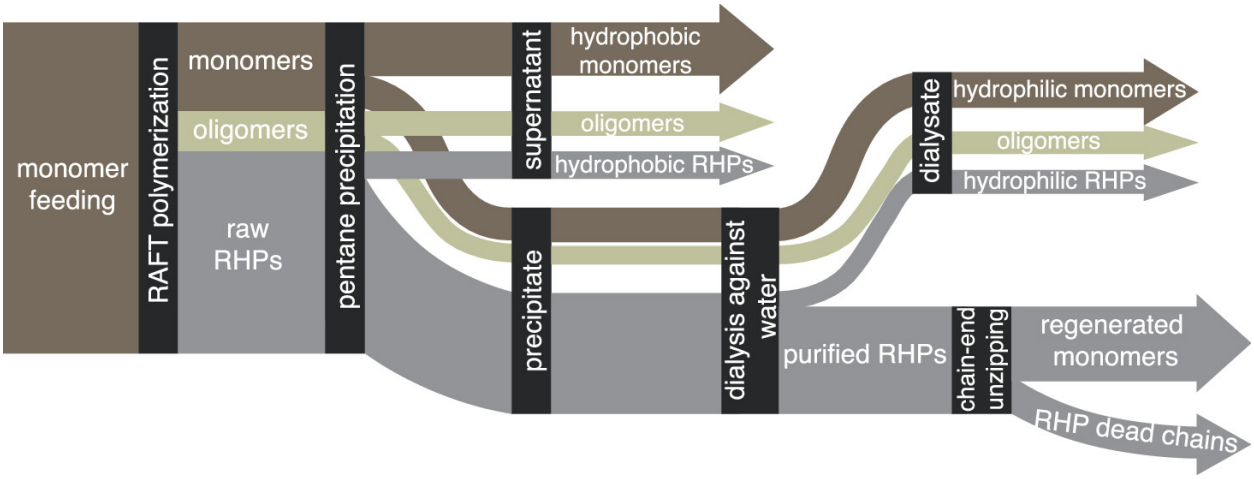
(b)

#	MMA	EHMA	OEGMA	NHSMA
MMA	-	1.3 ± 0.1	0.9 ± 0.02	0.5 ± 0.01
EHMA	0.4 ± 0.1	-	1.3 ± 0.1	0.6 ± 0.02
OEGMA	0.7 ± 0.2[a]	0.4 ± 0.1[a]	-	0.6 ± 0.1
NHSMA	0.7 ± 0.1	2.0 ± 0.3	1.6 ± 0.2	-

Figure 5.1 | A. Synthesis scheme of four-monomer RHPs. **B.** Reactivity ratios determined in this work ($r_{row/Column} = k_{rowrow}/k_{rowcolumn}$). [a]$r_{OEGMA/MMA}$ and $r_{OEGMA/EHMA}$ were determined separately by two-monomer copolymerization experiments to minimize analysis error due to peak overlap in the ^{1}H NMR spectra.

5.3 Composition evolution during synthesis and purification[2]

Heteropolymers are polydisperse in chain length, molecular weight, monomer composition, and sequence. *RHPapp 1.0* simulates an ideal simulation process with no accurate molecular weight distribution modeling, resulting in non-accurately distributed chain lengths. Additionally, the simulator simply takes a starting pool of monomers and presents final polymeric products, without factoring in non-ideal reactor conditions and post-reaction purification to remove excess reagent and reaction solvent. Both factors may bias the subpopulation of sequences that end up in the final characterized and functionally applied polymer products. Shea et al attempted an affinity chromatography approach to separate chains from a pool of heteropolymers.[43] They demonstrated that a subpopulation of the as-synthesized heteropolymers can exhibit more than 1000-fold higher affinity for a target protein than the average affinity.[44] Thus, a systematic investigation of the entire heteropolymer preparation process is required. We tracked the monomer composition evolution throughout the entire preparation process and investigated how the synthesis and purification manage the heterogeneities inherent in random heteropolymer (RHP) ensembles. Specifically, we quantitatively mapped out the evolutionary trajectory of monomer compositions across: (i) raw RHP products; (ii) purified RHPs; and (iii) the living chain subpopulations [chain terminated by a chain transfer agent (CTA)] (**Scheme 5.1**)



Scheme 5.1 | Composition Evolution of RHPs over the Course of the Entire Preparation Process. Due to composition drift and chain removal during purification, the composition of the final RHP products may differ from the initial feeding composition.

[2] [2] The experimental data reported in Chapter 5.3 was collected by Dr. Hao Yu and Ruilin Yin

Using experimentally determined monomer reactivity ratios, we designed four model RHPs with the ratio of hydrophobic monomers (MMA and EHMA) increasing progressively from **RHP1** to **RHP4** (**Fig. 5.2a**). This formulation leads to a progressive increase in the average chain hydrophobicity from **RHP1** to **RHP4**. We designed the monomer compositions based on previously reported RHPs consisting of MMA, EHMA, OEGMA, and SPMA,[5] along with our empirical experience on finding an optimal balance in solubility. For example, maintaining a ratio of EHMA or NHSMA below 50% is needed to avoid polymer precipitation in DMF before reaching the desired monomer conversion (>60%) during polymerization. In contrast, a ratio of OEGMA above 10% is necessary to ensure the water solubility of the final products. In radical polymerization, there is a trade-off between achieving higher monomer conversion and ensuring control over the reaction. This is due to monomer depletion and potential loss of end-group fidelity at higher conversions. However, attaining high monomer conversion reduces purification cost and minimizes waste generation. Additionally, the extent of compositional drift during polymerization is contingent on overall monomer conversion.[23] Therefore, we targeted **RHP1–4** for different overall monomer conversions as part of our study of optimal RHP production strategies.

(a)

entry	MMA (%)	EHMA (%)	OEGMA (%)	NHSMA (%)	targeted monomer conversion (%)
RHP1	10	20	45	25	~ 90
RHP2	40	10	40	10	~ 90
RHP3	50	20	20	10	~ 60
RHP4	25	40	25	10	~ 60

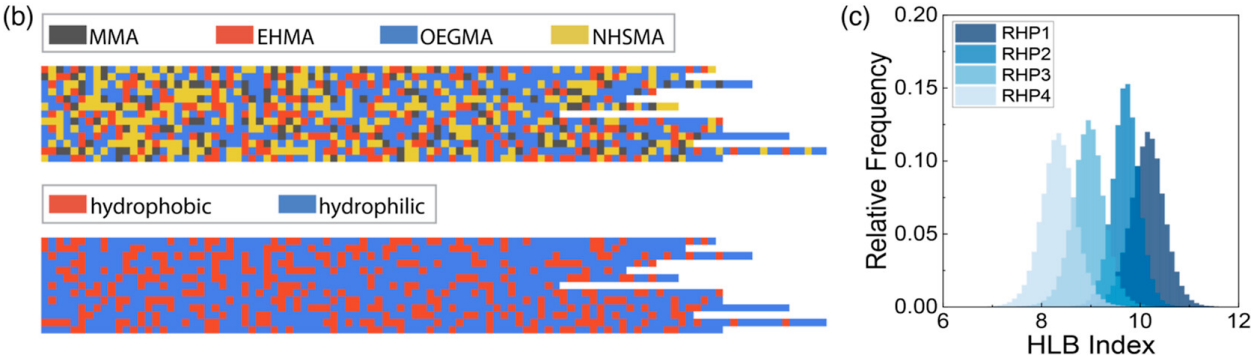


Figure 5.2 | **A.** Design of RHPs with increased hydrophobicity by increasing the compositions of hydrophobic monomers (MMA and EHMA) from **RHP1** to **RHP4**. **B.** Random sequences sampled from simulated **RHP1** using measured reactivity ratios and targeted monomer conversion. The RHP sequences shown at the bottom were binarized into hydrophobic (MMA and EHMA) and hydrophilic (OEGMA and NHSMA) units to display the distributions of hydrophobic/hydrophilic segments. **C.** Distribution of chain hydrophobicity in in silico-designed **RHP1–4**. Chain hydrophobicity was estimated using previously reported HLB values (MMA: 8.45, EHMA: 5.125, OEGMA: 11.42, and NHSMA: 12.775).[5]

To facilitate our design, we simulated RHP sequences using the *RHPapp*.[24,25] A total of 100,000 RHP sequences were generated for each model RHP batch for statistical analysis. The four designed RHPs exhibit sequence heterogeneity within local segments and between different chains (**Fig. 5.2b** and **Figs. S5.1-5.3**). There is no discernible monomer gradient observed along the RHP chains. To evaluate the RHP design, we estimated the average chain hydrophilicity using hydrophilic–lipophilic balance (HLB) values.[45] The lower HLB values indicate greater hydrophobicity, whereas the higher values reflect increased hydrophilicity. The results indicate a consistent decrease in hydrophilicity from **RHP1** to **RHP4**, aligning with our intended design (**Fig. 5.2c**). Collectively, this set of four designed model RHPs spans the representative range of hydrophobicity that is commonly encountered in the RHP design, benchmarked with poly(MMA) (HLB value: 8.45) and poly(OEGMA) (HLB value: 11.42). We anticipate that these four model polymers will serve as benchmarks, improving the prediction of the solution behavior in heteropolymers for future designs.

Recent studies on polymer diffusion kinetics during dialysis have revealed crucial influence of chemical composition on polymer retention.[46,47] Given the inherent sequence heterogeneity of RHPs, it naturally leads to the question of whether certain RHP chains are removed during the purification process (**Scheme 5.1**) Thus, we monitored the composition evolution of **RHP1–4** at every stage of the preparation process (**Fig. 5.3a**). The raw RHPs were purified by drop precipitation into pentane and subsequent dialysis against water (MWCO 3000 Da). The raw products after polymerization, precipitation supernatants, dialysate solutions, and purified RHPs were collected and analyzed using NMR and GPC (**Figs. 5.3b, 5.3c, S5.4-5.10**) Minimal batch-to-batch variations were observed among triplicate batches.

We first investigated the composition drift during RHP RAFT polymerization. The compositions of raw RHP products after RAFT polymerization were determined by quantifying the conversion of each monomer using ^{1}H NMR from the initial feedstock and post-polymerization aliquots, assuming that all consumed monomers participated in the polymerization (**Table S5.1**). Upon analysis, we noticed a decrease in the total peak integral when we compared the ^{1}H NMR spectra of the initial feedstocks with the post-polymerization aliquots. This decrease indicated a loss of monomers during the freeze–pump–thaw degassing process, which we attributed to the evaporation of volatile MMA under vacuum. A loss of approximately 20% in MMA was identified when contrasting the NMR spectra of the initial feedstock and the post-polymerization aliquots. To accurately estimate the composition of the raw products, this loss of MMA was factored into our calculations in determining the raw RHP compositions. With this correction, the experimental findings closely match the simulated results using experimental inputs, with a deviation of less than 4% (**Fig. 5.4, Table S5.2**) We conjecture that these deviations are likely due to the formation of dead chains during RAFT polymerization, which is currently not considered in the *RHPapp*.[25] Despite this, these results reinforce the validity of in silico RHP design by using experimentally measured and Jaacks method calculated reactivity ratios determined from 4-monomer heteropolymerization.

(a)

entry	Conversion (%)[a]	M_n^{theo} (kDa)	M_n^{exp} (kDa)[b]	DP_n^b	$Đ^b$
RHP1	90	29.4	22.8	70	1.39
RHP2	88	25.3	20.0	70	1.33
RHP3	64	14.3	13.3	60	1.31
RHP4	67	16.6	16.7	64	1.26

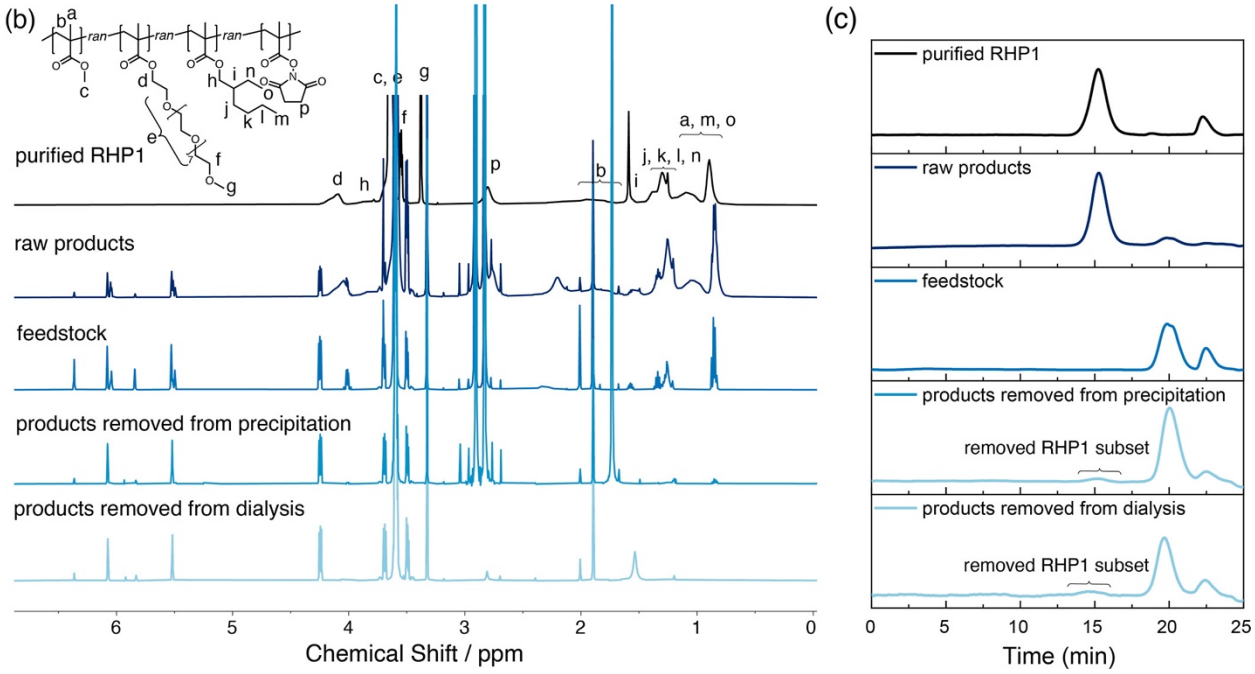


Figure 5.3 | **A.** Summary of the characterization results (averaged from three independent synthesis batches) of **RHP1–4**. [a]Overall monomer conversion determined from pre- and post-polymerization aliquots [b]determined using THF GPC with PMMA standards. DP_n was calculated using an averaged molecular weight based on the anticipated monomer composition. **B.** Representative ^{1}H NMR spectra ($CDCl_3$) and (c) GPC traces (THF) of **RHP1** and intermediates during synthesis. Residual GPC solvent peak (THF) is shown at 22.5 min.

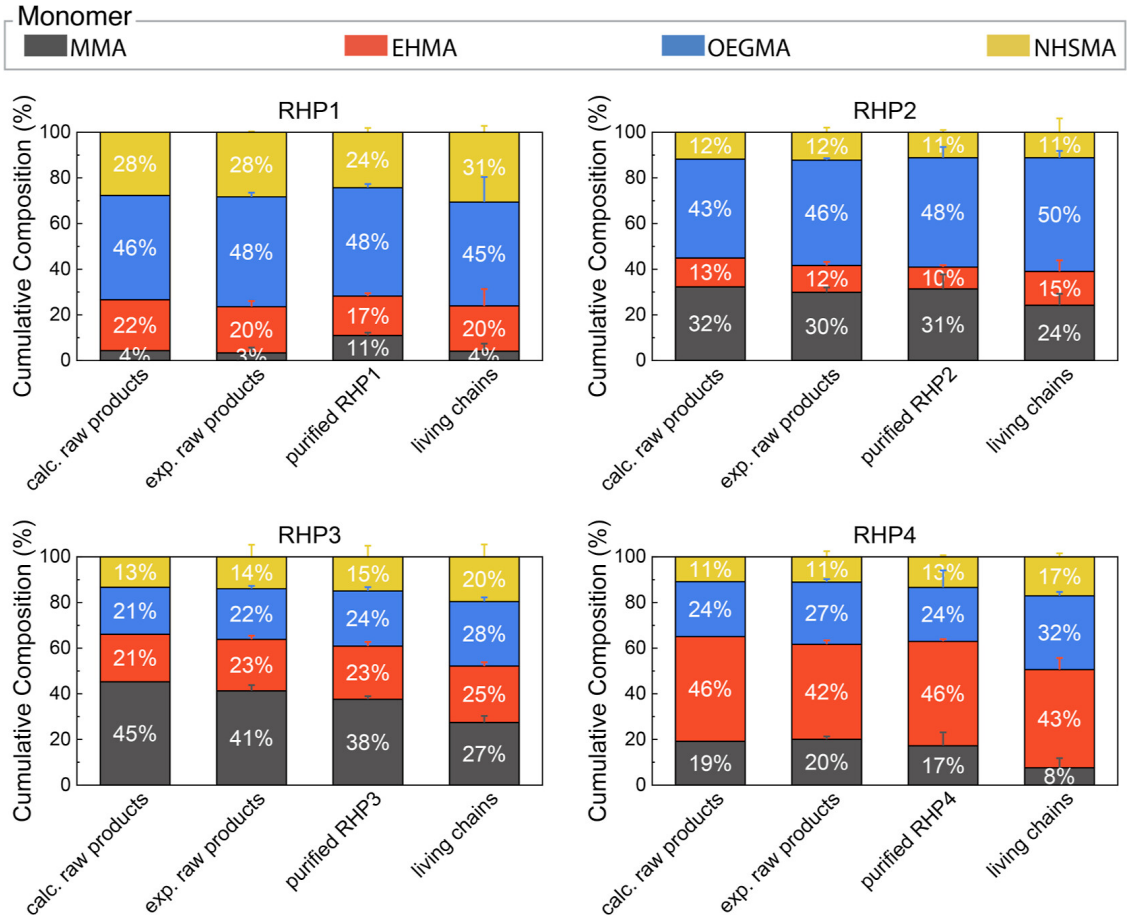


Figure 5.4 | RHP composition during RAFT polymerization and purifications. The percentage represents molar percent composition, and the error bars represent deviation from three independent RHP synthesis batches. The simulation results represent averaged composition of 100,000 simulated individual chains. The compositions of living chains were calculated from at least nine depolymerization experiments.

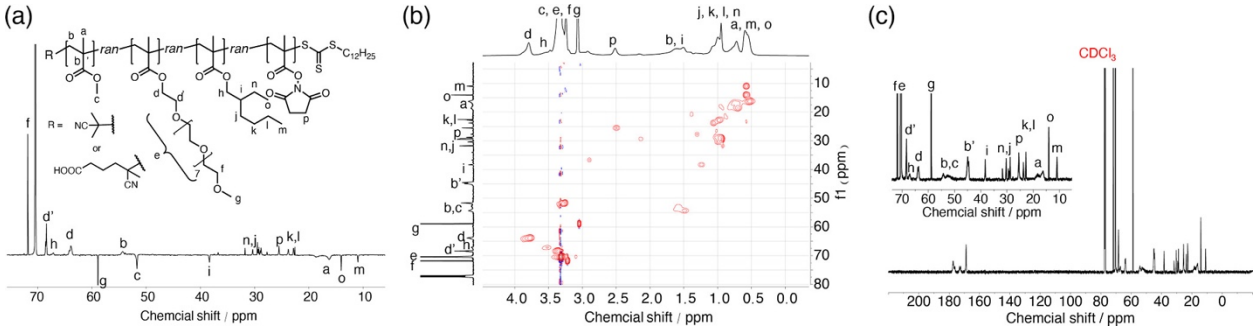


Figure 5.5 | DEPT-135 ^{13}C, **A.** ^{1}H–^{13}C HSQC, **B.** and quantitative ^{13}C NMR, **C.** of purified **RHP1** ($CDCl_3$).

We observed distinct impacts of pentane precipitation and dialysis on the **RHP1–4** ensembles. GPC analysis shows that pentane precipitation eliminated chains from **RHP1**, **3**, and **4** (**Figs. 5.3c**, **S5.9, S5.10**). Conversely, dialysis against water removed chains from more hydrophilic **RHP1–2** (**Figs. S5.7, S5.8**), with minimal observable effects on less hydrophilic **RHP3–4**. These observations support our hypothesis that the post-polymerization purification processes can eliminate RHP chains, contingent on their apparent hydrophobicity.

Quantifying the compositions of eliminated RHP chains in precipitation supernatants and dialysate solutions was challenging because their signals were generally low in the NMR spectra and obscured by the residual monomers (**Figs. 5.3b, S5.4-5.6**). To better quantify the impacts of purification process, we compared the compositions of raw and purified RHPs. Considering the peak broadening of polymers in ^{1}H NMR, we used DEPT-135 ^{13}C and ^{1}H–^{13}C HSQC NMR experiments to identify the characteristic ^{13}C signals of each monomer in purified RHPs, followed by studies using quantitative ^{13}C NMR (**Figs. 5.5**, **Table S5.3**). Comparing raw and purified RHPs, there was a minimal shift in composition (<4%), with the MMA fraction of **RHP1** exhibiting the largest shift (7%) (**Fig. 5.4**). These observations were consistent with GPC analysis, which showed that only a minor fraction of RHP chains were removed during purification. Within the systematic errors in quantitative NMR analysis,[48] these findings imply that despite the removal of partial RHP chains during the purification, the overall composition and sequence heterogeneity remain largely unaffected.

To examine the statistical control obtained in the four-monomer RHP RAFT copolymerization, a further assessment of the composition of RHP living chains was done (reported in detail by Yu et al).[49] We found that the molar percent of MMA in RHP living chains was 7–9% lower compared to the corresponding purified RHPs (**Fig. 5.4**, **Table S5.4**). These results indicate that the dead chains within **RHP1–4** are more MMA-rich compared with the living chains. In a recent study, Zhang revealed that a larger fraction of dead chains was generated at the early lower conversion stages of RAFT polymerization due to the higher abundance of primary free radicals at the initial stages of the reaction.[50,51] Likewise, as monomer conversion and chain length increase, a significant reduction in the termination rate coefficient, exceeding 3 orders of magnitude, was observed in the radical polymerization of MMA.[52,53] Furthermore, prior work has reported that the termination rate coefficient of MMA in radical polymerization is 1 order of magnitude higher than that of methacrylates with longer side chains such as butyl and dodecyl methacrylate.[53,54] In light of these results, the higher fraction of MMA in RHP dead chains can be ascribed to (1) the majority of RHP dead chains being formed during the early stages of polymerization, (2) during which RHP chains with a higher fraction of MMA are generated as suggested by stochastic simulation (**Fig. S5.11**), due to their high reactivity ($r_{\text{non-MMA monomer/MMA}} < 1$, **Fig. 5.1b**), and (3) the susceptibility of MMA-terminated chains to termination. Collectively, these findings represent an initial proof that the living and dead chains in RHPs may exhibit different composition distributions. For future efforts in synthesizing block heteropolymers through chain-end extension, we advocate for a thorough composition analysis that considers the compositional differences between the living and dead chains when analyzing the final products.

5.4 *RHPapp 2.0*: a RAFT heteropolymer kinetics and sequence simulator

The preceeding section demonstrated the importance of molecular weight distribution of chains, including the molecular weight and composition split between living-dead population species. It also highlighted discrepancies between in silico modeling and experimental results determined by NMR, in which the *RHPapp* predicted monomer conversion within a 10% error margin.[49] This is despite using improved reactivity ratio measurements and considering the subpopulation selecting effects of monomer evaporation and post-synthesis purification. We hypothesized that the prediction error originated from the lack of modeling of RDRP polymerization kinetics in sequence simulation, in particular termination reactions and living versus dead chain distribution differentiation. Although there have been studies showing that copolymer sequence distribution can be controlled by varying temperature, solvent, ionic strength, and locally available monomer concentration i.e., emulsions,[55–58] a more fundamental understanding of the effects of RDRP polymerization kinetics on copolymerization is not well known. Most published RHP synthesis has been done by reversible addition–fragmentation chain transfer (RAFT) polymerization.[1,59–61] However, RAFT kinetics are still not fully understood. Several competing models agree on the main RAFT mechanism, however, they differ in explaining the characteristic slowing of propagation rate observed when increasing the concentration of the chain transfer agent. Rate retardation caused by intermediate radical termination with oligomers (IRTO) is a currently popular theory and offers a simple set of starting assumptions for kinetic modeling.[62–64]

We hypothesized that incorporating RAFT kinetics into sequence simulation would increase sequence pattern accuracy. Here we developed the first simulator to give both accurate RAFT polymerization kinetics with full molecular weight distribution and explicit heteropolymer sequence prediction – the *RHPapp 2.0* (**Fig. 5.7**). The IRTO model is assumed but the simulator is designed to allow for facile future model modification. Here we demonstrated the *RHPapp 2.0* on binary copolymer and 4-monomer RHP systems, showing its utility as a tool for the fundamental study of how RAFT kinetics affects copolymer and heteropolymer sequences. Specifically, we showed relationships between rates (propagation, RAFT adduct fragmentation, termination) and sequence distributions to be system dependent. A binary styrene-glycidyl methacrylate system exhibited high sensitivity in sequence to propagation and termination constants, with low sensitivity to RAFT adduct fragmentation. A 4-monomer methacrylate-only RHP was much more stable to variance in termination parameters. A direct comparison between *RHPapp1.0* and *RHPapp2.0* generated RHP sequences showed negligible differences, suggesting that reactivity ratios (propagation) alone dominated all other kinetic effects and the *RHPapp 1.0* is a good approximation for the specifically explored 4-monomer methacrylate system. However, as shown by the binary system this does not generalize to all copolymer and heteropolymer systems, so the *RHPapp2.0* is a prerequisite step to benchmark the sensitivity of a given polymeric system, prior to using the simpler “kinetic-free” approximation. Open-source, user-friendly code is provided for community use for reaction planning and sequence analysis, leveraging compatibility with the *RHPapp 1.0* analytical frameworks.

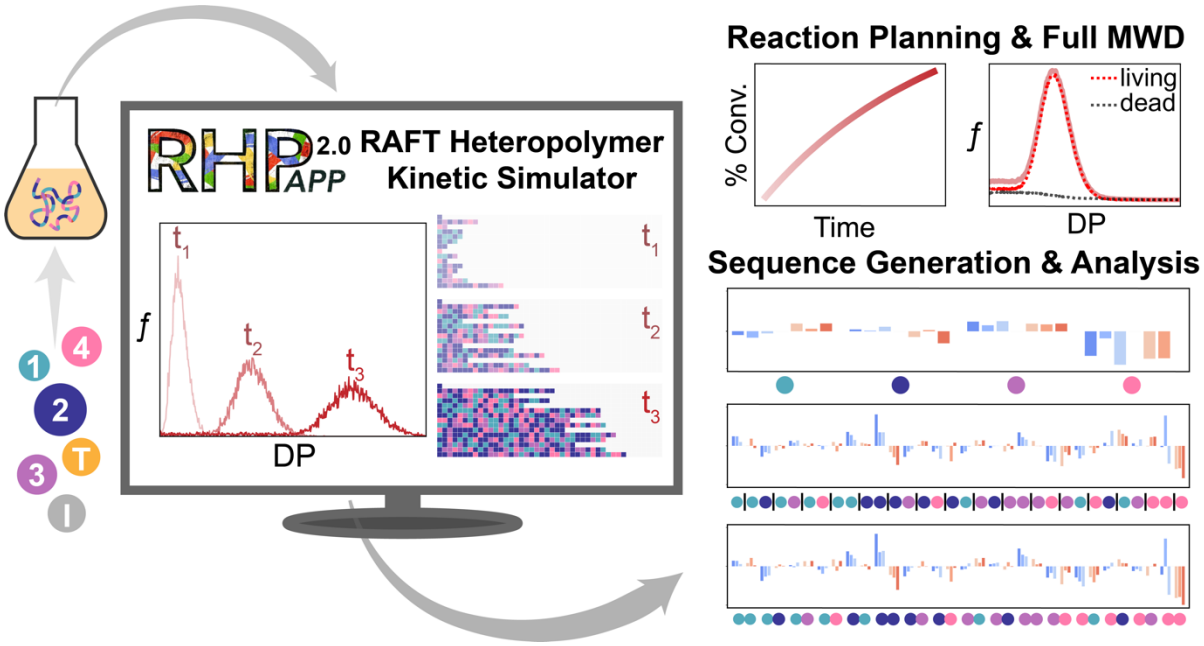
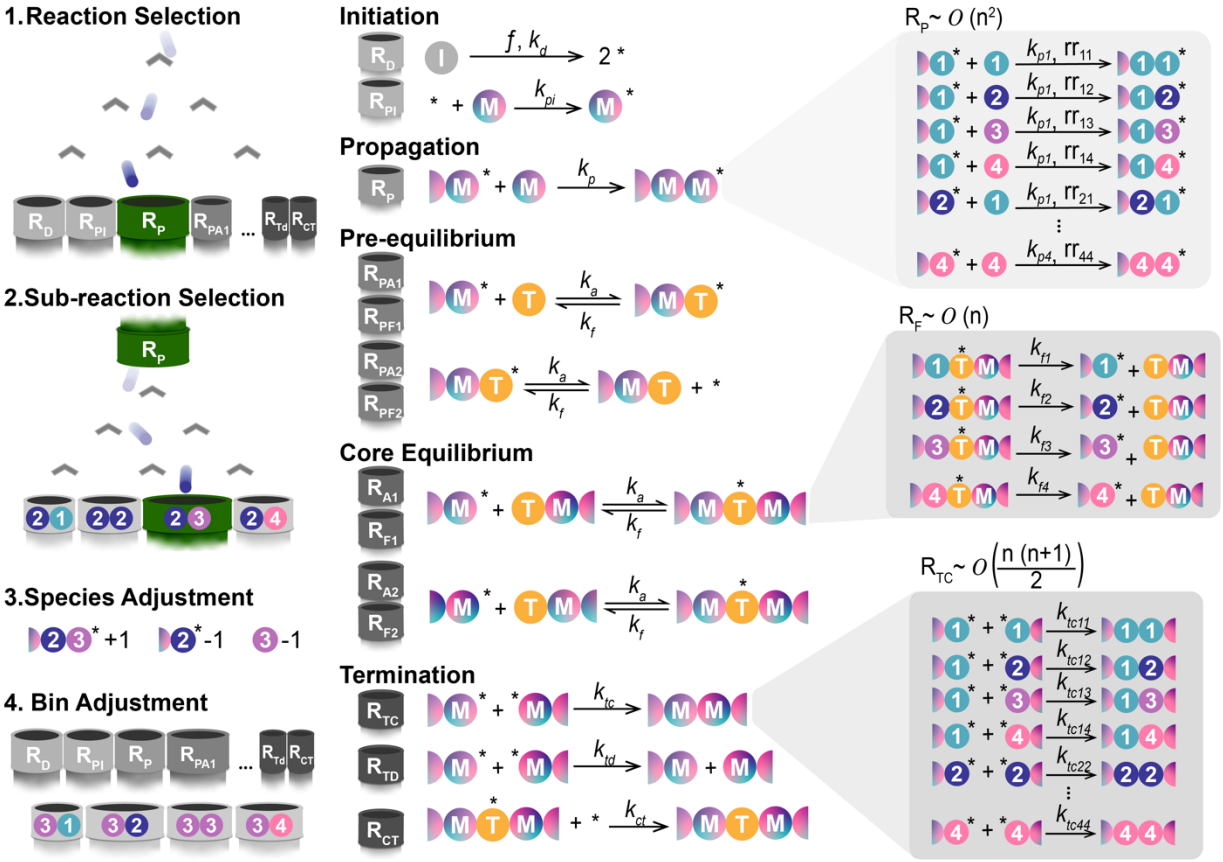


Figure 5.7 | Schematic flow of *RHPapp 2.0*.

The *RHPapp 2.0* is a Monte Carlo kinetic simulator that follows the exact method developed by Gillespie (**Scheme 5.2a**).[65] Substrate molecules, product molecules, a control volume, and the set of possible chemical transformations are explicitly initiated. At each time step the reaction rates for each unique chemical transformation are updated based on the concentration of substrates and user input kinetic constants. Time progresses iteratively with a single reaction occurring at each step, chosen by a reaction rate probability distribution weighted sampling, after which the reacting species counts, reaction rates, and the total estimated real-world elapsed reaction time are all updated. A detailed description of the core RAFT homopolymerization kinetic simulator algorithm as well as the choice of Julia as an ideal programming language were described by Pintos et al.[16] In order to model explicit heteropolymer sequences, the system of reactions, algorithms and data structures presented by Pintos et al required radical expansion and overhauling.

To accommodate the different kinetics of each unique monomer in a heteropolymerization, when a reaction e.g., propagation is selected, a sub-reaction selection step was added. This creates complexity as 14 unique homopolymerization reactions expand to over 100 branching reactions for a 5-monomer heteropolymerization (**Scheme 5.2b**). As a result, iterative reaction rate probability distribution updating and sampling become more time expensive. The number of required kinetic constants also expands and reactivity ratios are required. Our goal was to build a user-friendly, accurate simulator with low memory requirement that can complete an *in silico* polymerization in a reasonable timescale i.e., executable on most polymer scientists' personal computers in under 24 hours. Thus, designing for low memory requirement was the priority.



Scheme 5.2 | **A.** LHS showing 4-step Gillespie algorithm, **B.** RHS showing the full set of homopolymerization reactions modeled as well as the relevant kinetic parameters, and then zoom in on 3 sample reactions: propagation, core equilibrium fragmentation 1, and termination by combination, to show how the number of reactions expand either quadratic, linear or n(n+1)/2 for heteropolymerization with n monomers.

System scale was the primary computational memory bottleneck for simulation. Viable stochastic simulation scale is lower bounded by the minimum number of polymer chains required for both molecular weight and sequence distributions to converge to the same stable values that match experimentally observed ground truth. When simulation scale is too small, low concentration species *i.e.,* radical species that are explicitly represented (and thus integer valued) may go to zero. This would inhibit any further propagation, resulting in early termination and inaccurate kinetics. The minimum viable scale will vary depending on the specific polymerization simulated. Approximate polymerization and simulation scales are shown in **Figure 5.8a**. Most existing stochastic homopolymer simulators reduce the required computational memory by representing growing polymer chains implicitly *i.e.,* a single array of integers where the value of each array position *i* represents the count of chains of length *i*. This allows simulators to achieve accuracy afforded by higher system scales. However, this compressed representation

cannot be used for heteropolymers, as equal length chains will have non-identical sequence patterns. Moreover, the explicit sequence patterns of all chains are a key desired output (**Fig. 5.8b**). Thus, explicit sequencing is prohibitively memory expensive without further algorithm and data structure optimizations. Jayapurna et al. showed that when using the *RHPapp 1.0* to simulate a 4-monomer methacrylate heteropolymer system, approximately 10,000 chains alone were sufficient to yield accurate sequence distributions, far less than the millions of chains required for accurate kinetic modeling (**Fig. 5.8a**).[25] Leveraging this, the key algorithm-data-structure innovation for the *RHPapp 2.0* was to implement a 2-tiered system scale selection. Simulate millions of growing heteropolymer chains that are implicitly tracked for accurate kinetics, but only explicitly track the molecular sequences for a couple thousand. **Figure 5.8c** shows an arbitrary example where the kinetics of the smaller explicitly tracked subset is representative of the total simulated chains. Complete algorithm, data structure design selections, optimizations and data analysis tools are described in our GitHub repository (https://github.com/ivanjayapurna). *RHPapp 1.0* sequences used for comparison were generated as per methods previously described.[25]

We used 2 model systems for demonstration. First, a binary copolymerization of styrene (S) with Glycidyl Methacrylate (G) was used to simplify data analysis by having less input variables and output sequence data complexity. This allowed us to better ascertain trends between RAFT kinetic parameters and resulting sequence patterns. The S-G system was specifically chosen as Tenorio-Lopez et al published a compiled a list of RAFT kinetic parameters, their own simulated results and experimental kinetic data for us to validate our simulated kinetics against and verify accuracy.[19] Second, a 4-monomer random heteropolymer (M-O-E-N) system of methyl methacrylate (M), oligo(ethylene glycol) methyl ether methacrylate (O; average M_n ~ 500 Da), 2-ethylhexyl methacrylate (E), and methacrylic acid N-hydroxysuccinimide ester (N) from Yu et al was used to analyze trends in more complex heteropolymers and compare the results of RHPapp 2.0 versus RHPapp 1.0.[49] Yu et al provided the reactivity ratios and experimental system inputs. All other kinetic constants were estimated from literature.[54,66–72] Tables of complete simulation inputs for all systems used can be found in **Tables S5.5** and **S5.6**.

The accuracy of simulated sequences and kinetics will only be as accurate as the input kinetic parameters. Trustworthy kinetic values are obtainable from literature when using common reagents such as azobisisobutyronitrile (AIBN) or methyl methacrylate. However, accurate values will often not exist for exact system conditions (temperature, solvent, co-monomers, etc.) for less common reagents. These kinetic parameters can be challenging to experimentally measure or predict *in silico* e.g., RAFT addition and fragmentation rate constants. In these cases, the *RHPapp 2.0* can be a useful tool to screen *in silico*, the effects of modulating each input parameter on resulting kinetics and sequences. An estimate of the sensitivity for each parameter of given polymerization systems can thus be obtained, allowing the polymer chemist to determine which parameters are crucial to accurately measure or predict, and which can be treated as pseudo-constants across monomers and/or reaction environments – reducing system input complexity. *RHPapp 2.0* allows for sensitivity testing of all kinetic parameters. Here to demonstrate, we focused on screening the kinetic constants suggested by literature to be most likely to significantly

impact sequence distributions: monomer propagation (k_p), RAFT equilibrium chain fragmentation (k_f), and termination (k_t).[49,54,66–72]

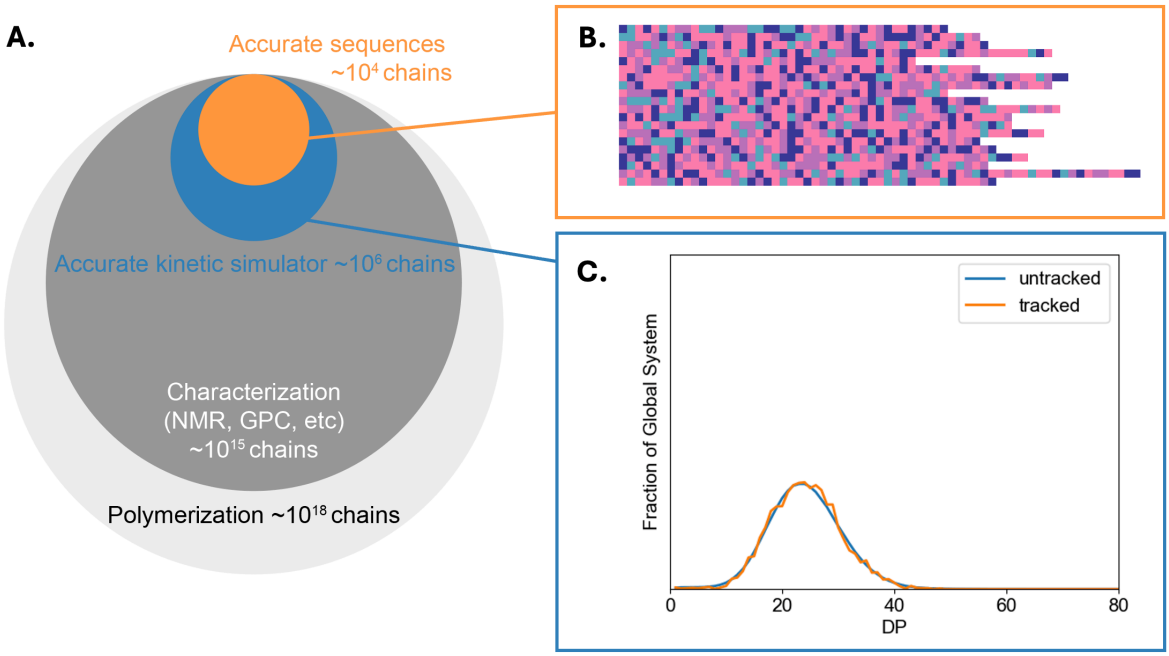


Figure 5.8 | A. Comparing the scale of experimental and simulated polymerization. Scale for experimental polymerization estimated based on the synthesis of 1 gram of polymer with average degree of polymerization (DP) of 1000 from monomers with average monomer molecular weight of 100 Da. Routine characterization to infer polymer molecular weight distributions or global sequence compositions requires on the order of 1 mg of polymer. Number of chains required for accurate simulated kinetics and sequence approximated from previous works and the results herein.[1,25,16] **B.** 20 explicitly simulated arbitrary 4-monomer heteropolymer sequences generated using the *RHPapp 2.0*, where rows represent polymer chains and each colored square represents a unique monomer. **C.** Snapshot of chain-length (DP) distribution of an arbitrary 4-monomer heteropolymer comparing ~10^6 implicitly untracked sequences and ~10^4 explicitly tracked sequences.

For both the S-G and M-O-E-N systems, a series of test simulations were performed with varying simulation scales to find the minimum system size required for accurate kinetic and sequence distributions. The number of simulated initiators controls simulation scale. Kinetic convergence was determined by the average degree of polymerization at a chosen monomer conversion. Different systems required different simulation scales to achieve statistical kinetic convergence, with S-G and M-O-E-N converging at 5e6 and 5e5 simulated initiators respectively (**Fig. 5.9a**). Simulated S-G polymerization kinetics closely followed the experimental results of Tenorio-lopez et al up to 120 minutes and conversions of 20-80% depending on the STY-GMA ratio (**Fig. 5.9b**).[19] This confirmed that our simulator had good kinetic fidelity with real experimental data. Sequence convergence for both systems were determined by each monomer composition at a given global composition. Again, different systems required different simulation scales to achieve statistical sequencing convergence, with S-G and M-O-E-N converging at 1e6 and 5e4 simulated initiators respectively (**Figs. 5.9c** and **5.9d**). Based on these two example systems, larger scale is required to reach statistical kinetic convergence than sequence convergence. Thus, 5e6 and 5e5 initiators, corresponding to 5e9 and 5e8 total molecules were respectively used for all subsequent simulations of S-G and M-O-E-N systems.

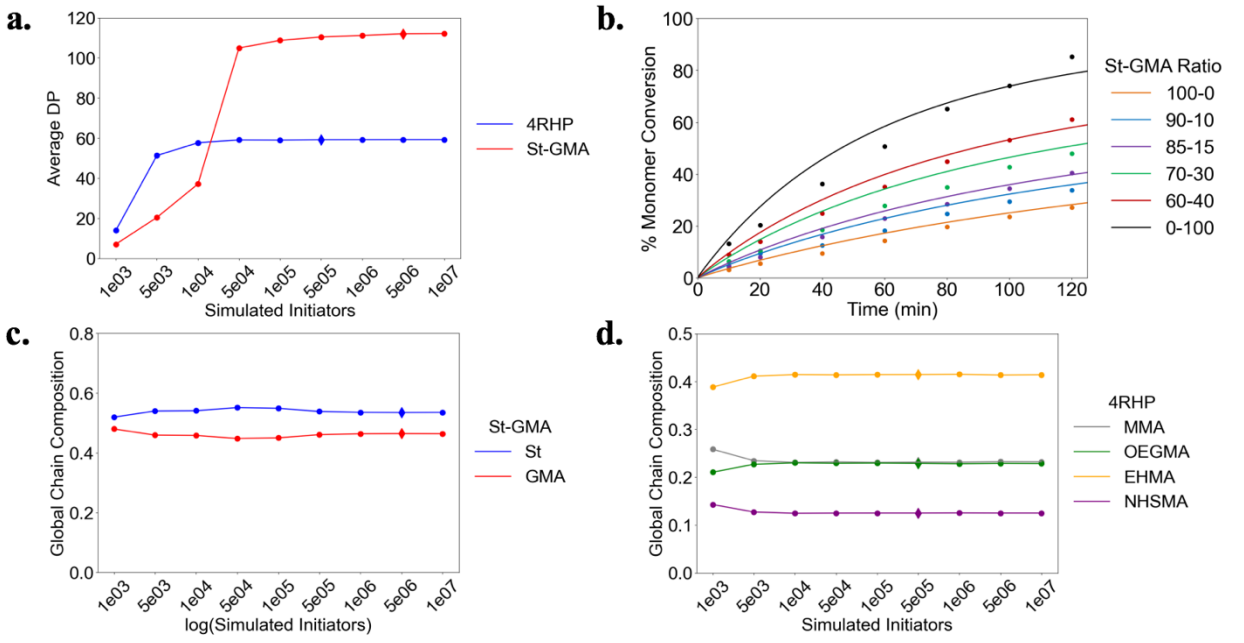


Figure 5.9 | A. Average degree of polymerization (DP) of STY-GMA and 4RHP systems as a function of the number of initiators simulated at fixed % conversion. Diamond indicates the chosen scale used for simulation at which statistical convergence is reached. **B.** Line curves of simulated STY-GMA global monomer conversion over time over scatter points of experimental data from Tenorio-Lopez et al,[19] for varying feeding ratios of STY-GMA. Global chain composition of (**C.**) STY-GMA and (**D.**) 4RHP as a function of the number of initiators at fixed % conversion. Diamonds indicate the chosen scale used for simulation.

5.5 Effect of RAFT polymerization kinetics on binary copolymer sequences

Two main visualizations of simulated data were used to study the effects of RAFT polymerization kinetics on resulting chain length and sequence distributions. To evaluate the sensitivity of stable polymerization kinetics to a kinetic parameter, we compare full chain length distributions at fixed conversions. Three batch-average plots were used to assess the impact varying a kinetic parameter will have on the resulting polymer sequence. Across the scaling range with which a kinetic parameter was varied we compared: monomer frequency, conditional probability, and dyad frequency. The conditional probability was determined by counting monomer frequency and dyad frequency, and then calculating according to Bayes' law: $P[M_{n+1}=Y, M_n=X] = P[M\ M_{n+1}=Y \mid M_n=X]\ P[M_n=X]$. Here, $P[M_{n+1}=Y, M_n=X]$ is the frequency of dyad XY, $P[M_n=X]$ is the overall frequency of monomer X in the monomer pool, and $P[M_{n+1}=Y \mid M_n=X]$ is the probability that monomer Y will follow monomer X. The y-axes of each of these plots show percent change relative to the standard kinetic parameters estimated from literature.

RAFT control is lost when the rate constant of propagation for S-terminal chains (k_p^S) increases, signified by the loss of a Gaussian chain length distribution (**Fig 5.10**). RAFT control is similarly lost when k_p^S and k_p^G are both increased, with the chain length distribution shifting from a Gaussian to an exponential decay. This is because propagation is too fast compared to RAFT equilibrium steps that are necessary to slow down and control propagation. When k_p^S decreases, overall propagation is slower and eventually the system does not reach the targeted % conversion due to all radicals terminating or near-infinitely looping in RAFT equilibrium steps. When both k_p^S and k_p^G are decreased RAFT control is also reduced with extra short chain length population observed, as when total rate of propagation is too slow compared to termination, the population of terminated short chains increases.

Increasing k_p^S resulted in sequences with more S and less G, which was the expected result. The relative rate of homopropagation of S (S|S) and G (G|G) both increased while heteropropagation decreased (S|G and G|S). Because the reactivity ratios naturally prefer homopropagation over heteropropagation for both monomers, increasing k_p^S just magnifies this effect. This interestingly suggests that when the rate of propagation is decreased relative to the rates of initiation, RAFT equilibrium, and termination, differences in co-monomer reactivity ratios can be "buffered" towards a more random distribution. To amplify differences in reactivity ratios, systems can be designed where k_p is maximized and all other reactions are minimized. This can be done by using less chain transfer agent or less initiator, among other strategies. Each strategy will have its corresponding downside *e.g.,* to reach the same target average chain length with a lower ratio of chain transfer agent relative to monomer will require lower monomer conversion, which is more material wasteful. This result suggests that the experimentally measured reactivity ratios for RAFT systems is likely dependent on the concentration/ratio of RAFT chain transfer agents relative to monomers and other factors. All the tunable inputs for RHP design are convoluted together and designing these convolved tradeoffs can be done with the *RHPapp* 1.0 synthesis planner and analytics modules.[25] In comparison, there is negligible change to sequence when both k_p^S and k_p^G are increased or decreased up to 10x and 0.1x, which makes sense as what is

important for sequence is relative rates (R_p^S relative to R_p^G), an effect which is likely captured by the experimentally measured reactivity ratios.

When the rate constant of fragmentation of living S-terminal chains from intermediate two-arm adducts (k_f^S) increased, RAFT control (Gaussian chain length distribution) was maintained up to 10x but was lost at 100x. When k_f^S decreased, the gaussian was rapidly is lost. Two-armed radicals are extremely short lived at the starting k_f values, so as fragmentation slows down, two-armed radicals become more stable and begin to affect kinetics. Thus, fragmentation has near-zero effect on kinetics when fragmentation is fast and when k_f for both monomers are within ~1.5 orders of magnitude. Interestingly, when both k_f^S and k_f^G increased or decreased even up to 100x and 0.01x, there was no change in chain length distribution. The effect of relative fragmentation on DP distributions can be understood when viewing RAFT-capped chains as dormancy traps. When a radical chain attacks a dormant RAFT-capped chain to form a two-armed radical, the probability distribution of which arm will then break to reform the active radical chain and dormant RAFT-capped chain is dictated by the relative fragmentation rate. If the T-M_1 bond is much more likely to break than the T-M_2 bond, then radical chains with terminal monomer M_2 will get trapped in RAFT-capped dormancy. This collapses the pseudo-Gaussian that is associated with RAFT control, into the exponential distribution that is associated with free radical polymerization and poor kinetic control.

Surprisingly, neither increasing or decreasing k_f^S nor increasing or decreasing both k_f^S and k_f^G up to 100x and 0.01x had any impact on sequence distributions. This is likely because even when chains are trapped in dormancy, it does not shift their reactivity ratios *i.e.,* once chains eventually come out of RAFT-capped dormancy they will propagate with the same probability distribution and thus result in the same statistical distribution, regardless of the long dormancy period. This result is useful because k_f values are extremely difficult to measure experimentally. Because k_f exhibits extremely high stability to perturbations, for future simulations with similar systems for sequence generation purposes, we can use DFT simulated/estimated values that are not system specific, without too much concern that it will have an outsized impact on sequence fidelity.[70]

Increasing the rate constant of termination by combination for S-terminal chains (k_{tc}^S) did not change the shape, just shifted the Gaussian to slightly shorter chain lengths. When the total rate of termination (R_t) increases, the approximate steady state concentration of radicals decreases, so the overall rate of monomer consumption also slows down. Therefore, it takes longer to reach the target monomer conversion %, resulting in more time for initiators to fragment, thus producing more chains overall, and thus a lower average degree of polymerization. When k_{tc}^S is pushed to its upper limits, the system is terminated before it can reach higher conversions. However, the Gaussians do not collapse because termination does not directly interfere with the core RAFT equilibrium. The observation of decreasing (k_{tc}^S) shifted the gaussian chain length distribution to slightly longer chains can be understood by similar reasoning.

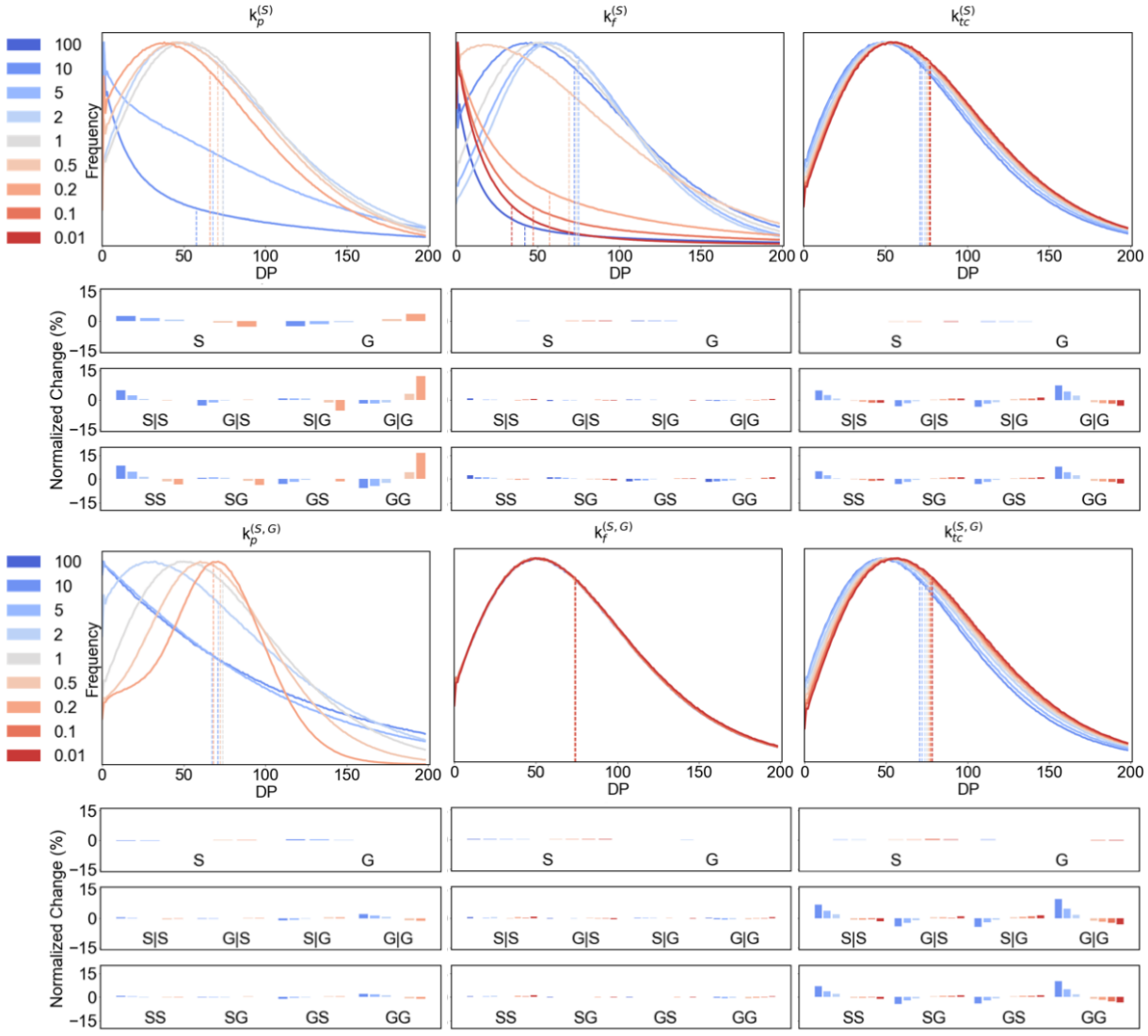


Figure 5.10 | S-G kinetics (top) and sequence (bottom) distributions resulting from screenings of varying kp, kf, and ktc (left to right). Top pair of kinetic and sequence distribution plots varied the kinetic parameters of only S. Bottom pair varied the kinetic parameters of both S and G. All data taken at 33% conversion. For kinetics plots: average degree of polymerization (DP) indicated by vertical lines and the maximum frequency of each distribution is set to 1. For sequence plots top sub-plot is monomer composition, middle sub-plot is conditional probability *i.e.,* S|G is the probability that a S follows a G along the chain. Bottom sub-plot is dyad distribution. Legend for all plots is scaling factor that is multiplied to the kinetic parameter(s) being varied. Not all scaling factors run for all kinetic parameter screenings, omissions either due to too early termination or slow propagation resulting in insufficient monomer conversion.

The impact that varying k_{tc}^S has on resulting sequence is perplexing. There is no change in overall monomer composition when scaled by up to 10x and 0.01x (100x increase in k_{tc}^S terminated too early to reach 33% conversion). However, despite the constant overall monomer composition, the conditional probability and dyad distributions shift with higher k_{tc}^S to prefer homopropagation (S|S and G|G) rather than heteropropagation. This result is far more complicated than the impact of varying k_p^S, in which the conditional probability shift can be explained as accompanying the shift in batch-average monomer composition. Varying both k_{tc}^S and k_{tc}^G had the exact same effect on both chain length and sequence distributions as was observed from just varying k_{tc}^S. Termination is diffusion controlled and the unscaled S-termination (1.16e12) term dominates G-termination (7.02e11). However, when k_{tc}^S is scaled down by 0.1x or 0.01x it should no longer be dominant. This suggests that increasing k_{tc}^S does not shift conditional probability towards more homopropagation due to an increase in the number of "homo-terminating" (SS and/or GG) chain-ends. If increased homopropagation is caused by increased homo-terminations, we would expect the change to occur in the dead chain distribution only. At 33%, dead chains make up to between 3-20% of the total chain population, where a larger k_{tc} leads to a larger dead chain % (**Fig. S5.12**). When we compared how the living and dead chains distributions varied with k_{tc}^S, we found that although the dead chain distributions are noisier due to a lower number of chains to average over, similar shifts in conditional probability occurred in both distributions (**Fig. 5.11**). Based on this result, we believe that k_{tc}^S is indirectly affecting the in-simulation kinetics by affecting the steady state concentration of radicals. Identical screenings were done varying the rate constant of termination by disproportionation for S-terminal chains (k_{td}^S) as well as varying both k_{td}^S and k_{td}^G and the impact on sequence distribution was near identical to the impact of varying k_{tc}^S. Although we have no definitive explanation, a possible hypothesis on how the rate constants of termination (k_t) affect overall kinetics is that R_t is slow relative to the rates of RAFT equilibrium steps and propagation. We empirically found that between any two termination steps there are ~1000 propagation steps and ~1000 equilibrium cycles. Thus, the set of living chains that are available for S+S, S+G, and G+G termination is completely different between termination instances. A live chain with active S-terminal (available for S+S and S+G termination) at a previous or future time point may have either S-terminal or G-terminal, based on some probability distribution that is a function of the steady state S-/G-terminal living radical ratio.

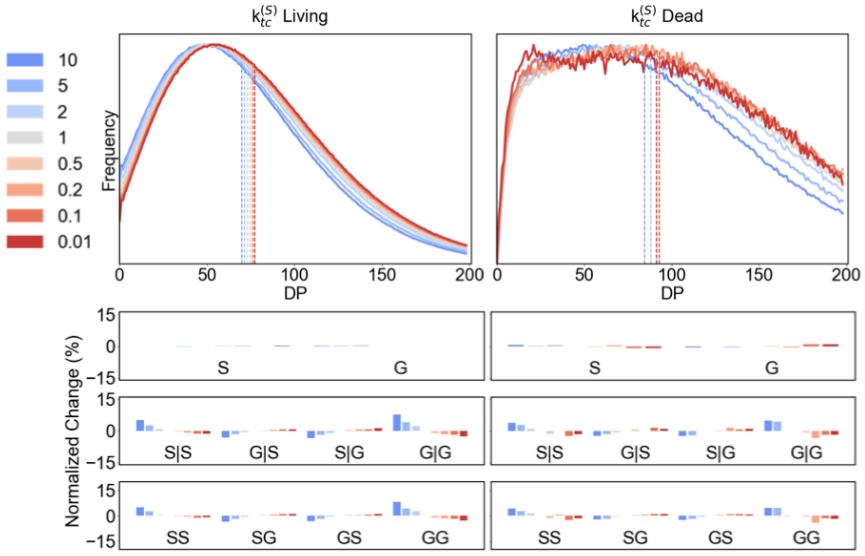


Figure 5.11 | S-G kinetics (top) and sequence (bottom) distributions split between living and dead chain populations resulting from varying k_{tc} of S. All data taken at 33% conversion. For kinetics plots: average degree of polymerization (DP) indicated by vertical lines and the maximum frequency of each distribution is set to 1**.** For sequence plots top sub-plot is monomer composition, middle sub-plot is conditional probability *i.e.,* S|G is the probability that S follows G along the chain. Bottom sub-plot is dyad distribution. Legend for all plots is scaling factor that is multiplied with k_{tc}

5.6 Effect of RAFT polymerization kinetics on heteropolymer sequences

Insights from the simpler binary copolymerization system can be applied to analyze a more complex 4-monomer random heteropolymer system (M-O-E-N). The fragmentation rate constants (k_f) in the S-G binary system showed remarkable stability against perturbations, not affecting sequence distributions. Consequently, in the M-O-E-N system, we prioritized examining the rate constants of propagation (k_p) and termination by combination (k_{tc}). Among the four monomers, we specifically varied the kinetics of OEGMA, because the kinetic parameters for MMA and EHMA are well-documented in the literature, whereas approximations were necessary for OEGMA and NHSMA. We selected OEGMA due to its higher initial concentration ratio (25%) compared to NHSMA (10%), hypothesizing that changes in its kinetic parameters would have a more significant impact. We conducted a screening where the kinetic parameters for both OEGMA and EHMA were adjusted. The combined feeding ratio of OEGMA and EHMA (approximately 67%) closely matches that of STY (60%) in the S-G system, allowing for comparisons of relative rate effects in a similar proportion of the system. Finally, we varied the kinetic parameters for all four monomers simultaneously to assess the impact of changes in total reaction rates (**Fig. 5.12**).

The relative value of k_p is crucial. As k_p increases, the chain length distribution slightly broadened on both ends. When both k_p^O and k_p^E were increased, the broadening was more pronounced, and when all monomers were increased, the system experienced the most significant broadening on both ends. Increasing propagation relative to other rates reduces RAFT control, leading to greater polydispersity. The less control exerted by RAFT on the monomers, the worse the overall distribution becomes. Conversely, reducing k_p does not affect the distribution's broadening or shrinking, but it does increase the formation of short chains (0 to 20 DP) due to relatively more dominant termination events.

Analyzing the sequence distribution, we observe that as the propagation constant for O-terminal chains (k_p^O) increased, the overall batch composition of O increased while other monomers decreased. This is expected because O-terminal chains propagate faster. The homopropagation of O (O|O) significantly increased, primarily at the expense of E|O and N|O. This suggests a significantly reduced probability of E or N following O. However, M|O is not as significantly affected. The reactivity ratio of M-M to M-O ($RR_{M,O}$) is 0.9, indicating a preference for M to add to O. Both $RR_{E,O}$ and $RR_{N,O}$ are > 1 so there is a preference to add to themselves. Thus, just as in the binary system, an increase in k_p amplified the existing reactivity ratios. In the binary S-G system, sequencing effects were negligible when the relative scale between monomers remained unchanged. However, in the M-O-E-N system, although overall monomer frequency is unchanged when all four k_p are varied, conditional probabilities are non-negligibly altered. With increasing total rate there is a significant increase to the homopropagation of O and N, at the cost of heteropropagation between the same two monomers. This could be another instance of reactivity ratio amplification that resulted in blockier sequence segments for 2 of the 4 monomers at higher k_p.

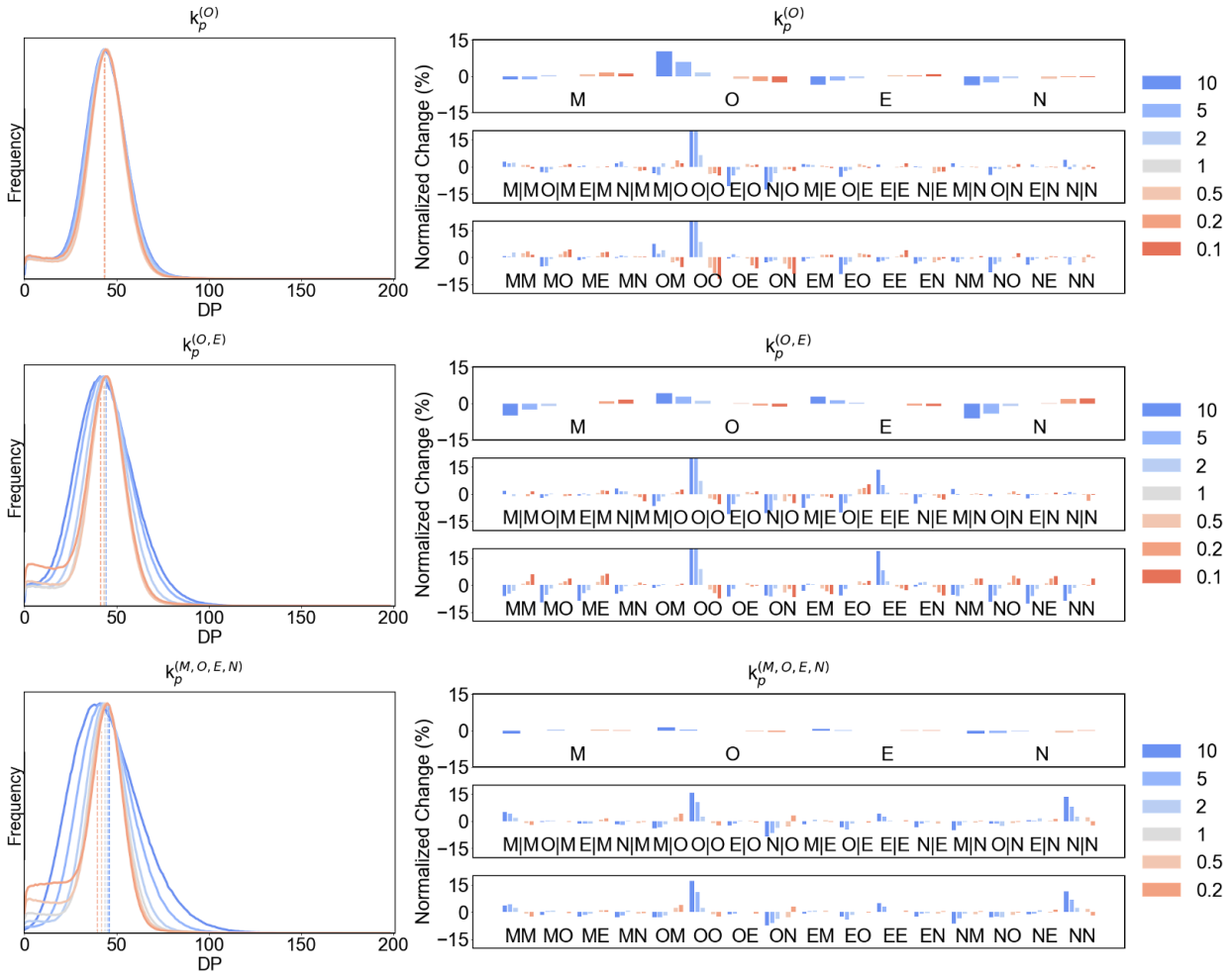


Figure 5.12 | M-O-E-N kinetics (left) and sequence (right) distributions resulting from screenings varying k_p. Kinetic and sequence distribution plots varied k_p of only O, both O+E, and all 4 monomers (from top to bottom). All data taken at 50% conversion. For kinetics plots: average degree of polymerization (DP) indicated by vertical lines and the maximum frequency of each distribution is set to 1**.** For sequence plots top sub-plot is monomer composition, middle sub-plot is conditional probability *i.e.,* O|M is the probability that O follows M along the chain. Bottom sub-plot is dyad distribution. Legend for all plots is scaling factor that is multiplied with k_p. Not all scaling factors run for all kinetic parameter screenings, omissions either due to too early termination or slow propagation resulting in insufficient monomer conversion.

Altering the termination constant for O-terminal chains (k_{tc}^{O}) for O had no significant effect on the kinetics, as it is diffusion-controlled and k_{tc}^{M} is the dominant factor, meaning most termination events will involve MMA. k_{tc}^{E} and k_{tc}^{M} are 2x and 20x larger than k_{tc}^{O} respectively. As the number of monomers with varied k_{tc} increases, we start to see variations in the presence of short terminated dead chains within the 0-20 DP range, which constitutes the only noticeable change in kinetics. This pattern mirrors observations in the binary system, where termination does not disrupt the core RAFT equilibrium. The overall impact of termination on kinetics aligns with findings from the binary system.

The macro trend observed for the impact of the kinetic parameter on sequencing in the M-O-E-N system mirrors that in the binary system, where changes in the monomer frequency across all k_{tc} screenings are negligible. Minor fluctuations in overall monomer frequency are insignificant compared to changes in conditional probabilities. Examining these conditional probabilities, homopropagation increases with the total termination rate (R_t), though patterns within these probabilities are difficult to distinguish. The probability of N|O paradoxically increases both when k_{tc}^{O} is raised and lowered, suggesting more random noise than a discernible pattern. The ambiguity implies that a larger system scale might be required. The presence of four monomers makes interactions to convoluted for systematic analysis, it was a challenge to fully understand the effect of k_t on sequencing even in the simpler binary system. Ultimately we can conclude that the 4-monomer RHP system (M-O-E-N) used in **Chapter 5.2** and **5.3** is quite robust to changes in k_t.

We (Yu et al) hypothesized that the lack of RAFT kinetic modeling, in particular termination kinetics may have been key to the ~10% error margin between experimentally measured and in silicon simulated sequences.[49] **Figure 5.14** compares batch-average monomer frequencies, conditional probabilities, and dyad distributions of RHP sequences simulated by *RHPapp 1.0* and *RHPapp 2.0*. The distribution for *RHPapp 2.0* is split into living and dead sequence populations, a feature unique to the newer model. Ultimately, although there is some disparity between the dead sequence population, the *RHPapp 2.0* living sequences and *RHPapp 1.0* sequences are near identical. At least for this system, the *RHPapp 1.0* is a very good approximation using feeding ratios and reactivity ratios (Mayo-Lewis) alone without any real kinetic modeling. This work inadvertently further supports the validity of the RHP sequence analyses published by the Xu Group thus far.

Our model was able to accurately simulate the kinetic and global composition experimental data presented in Tenorio-Lopez et al and Yu et al respectively. Yu et al provided an experimental upper bound for RHP-4 dead chain population of 18%. Here we provide a theoretical ideal lower bound 15% dead chains, so the bound is reasonably tight. This suggests that radical loss via chain termination during RHP depolymerization was minimal and therefore, the method used was accurate for determining living chain global monomer compositions.

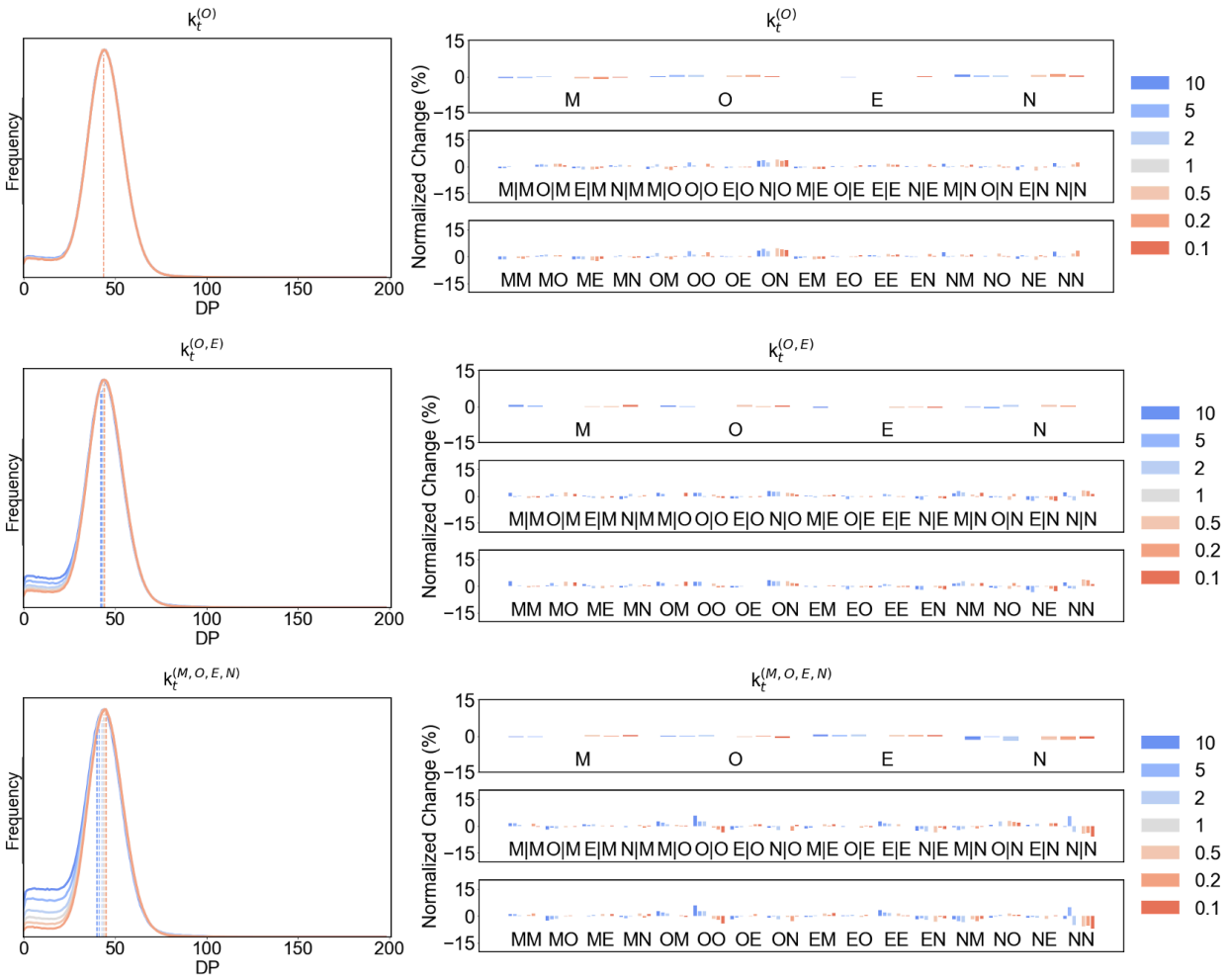


Figure 5.13 | M-O-E-N kinetics (left) and sequence (right) distributions resulting from screenings varying k_t. Kinetic and sequence distribution plots varied k_t of only O, both O+E, and all 4 monomers (from top to bottom). All data taken at 50% conversion. For kinetics plots: average degree of polymerization (DP) indicated by vertical lines and the maximum frequency of each distribution is set to 1**.** For sequence plots top sub-plot is monomer composition, middle sub-plot is conditional probability *i.e.,* O|M is the probability that O follows M along the chain. Bottom sub-plot is dyad distribution. Legend for all plots is scaling factor that is multiplied with k_t. Not all scaling factors run for all kinetic parameter screenings, omissions either due to too early termination or slow propagation resulting in insufficient monomer conversion.

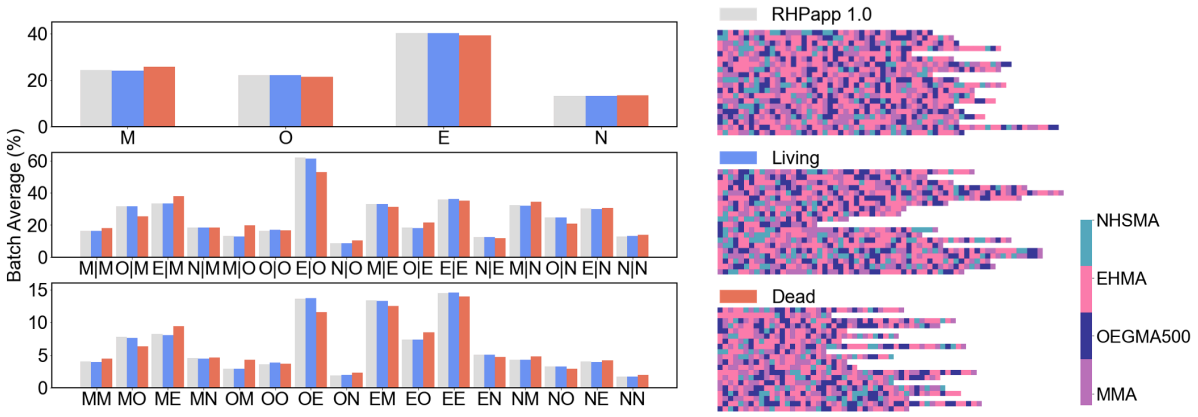


Figure 5.14 | M-O-E-N *RHPapp* 1.0 versus *RHPapp* 2.0 sequence comparison. Bar plots comparing the batch average monomer compositions (top), conditional probabilities (middle), and dyads (bottom) for the same batch of RHP simulated by *RHPapp* 1.0 and *RHPapp* 2.0. For *RHPapp* 2.0 living and dead populations were split. Legend includes 20 example monomer sequences of each.

Most sequence analysis of copolymers today only consider feeding ratios and global compositional drift. The *RHPapp 1.0* (Compositional Drift simulator module) used the simple Mayo-Lewis equation to generate sequences for downstream sequence pattern analysis. Yu et al (Chapters 5.2 and 5.3) showed the importance of experimentally measured reactivity ratios and accounting for composition evolution throughout reaction preparation and purification. To generate more accurate sequences in this work we built and presented a simulator that incorporates RAFT kinetics, the *RHPapp 2.0*.

In the process of demonstrating our unique dual kinetics and explicit sequence generating aspects of our simulator, we further explored its utility as a platform to shed fundamental insight into how RAFT kinetics do and do not affect copolymer and heteropolymer sequencing. We explored the impacts of the rate constants of propagation, two-arm adduct fragmentation, and termination on heteropolymer kinetics (molecular weight distributions at target % conversion) and sequence distributions. By exploring the sensitivity/stability of kinetic parameters to perturbations, we found that fragmentation rate constants are extremely stable and thus, these difficult to measure parameters can be approximated from DFT simulated/estimated values that are not system specific without suffering an outsized impact on sequence fidelity.

We also demonstrated that when the rate of propagation is decreased relative to the rates of initiation, RAFT equilibrium, and termination, differences in co-monomer reactivity ratios can be "buffered" towards a more random distribution. To amplify differences in reactivity ratios, systems can be designed to maximize propagation. As was shown herein, this is a viable strategy to achieve selectively blockier sequence segments for 2 of 4 monomers in a heteropolymer system. However, these strategies will leverage tunable RAFT synthesis inputs that are convoluted together, which is where the *RHPapp* can shine as an aide for reaction design. Future work could use the *RHPapp 2.0* platform to investigate the effect and stability of initiator dissociation and two-arm adduct addition rates, among other relevant parameters, on heteropolymer sequences.

We validated that for the 4-monomer methacrylate system used in prior works, Mayo-Lewis had sufficient complexity to be accurate. However, this result may not be generalizable to all copolymer and heteropolymer systems, as perhaps methacrylates are an edge case due to the high backbone chemical similarity and thus kinetic parameters of all monomers. Indeed, the binary copolymer system investigated herein showed that the stability of the Mayo-Lewis approximation can be system dependent. Moreover, the effects of termination rate constants remain a mystery to be solved in future work. We acknowledge that there are several abstractions our simulator made - real effects that our model doesn't capture – that if implemented, should further improve sequence fidelity to reality. These include considering a penultimate model, chain-length dependent kinetics, and non-ideal reactors. We open-sourced our code to encourage the community to co-implement these additions and more in the next generation simulator to come, perhaps the *RHPapp 3.0*.

5.7 Methods and materials

Materials: All reagents were purchased from commercial suppliers and used as received unless otherwise stated. Azobisisobutyronitrile (AIBN) was recrystallized in methanol prior to use. Prior to polymerization, methyl methacrylate (MMA), 2-ethylhexyl methacrylate (EHMA), and ethylene glycol methyl ether methacrylate (OEGMA) were passed over a column of neutral alumina to remove inhibitors.

Nuclear Magnetic Resonance (NMR) Spectroscopy: 1H and 13C NMR spectra (500 MHz and 150 MHz) were recorded at room temperature (298 K). Chemical shifts are reported in δ (ppm) referenced on residual solvent peaks. Coupling constants (J) are expressed in Hertz (Hz). Splitting patterns are designated as: s (singlet), d (doublet), t (triplet), and m (multiplet).

Determining Composition of Raw RHP Products: The consumption of each monomer during RAFT copolymerization was estimated by comparing the NMR spectra of feedstock and post-polymerization aliquots (using DMF peak as internal standard). The initial feedstock NMR aliquot was taken before the degassing process. The analysis revealed that a decrease in the total peak integral between the feedstock and post- polymerization NMR spectra, indicating the loss of monomers during the freeze-pump-thaw degassing procedure. This can be attributed to the loss of volatile MMA during the degassing process. Therefore, we estimated the loss of MMA by comparing the NMR spectra of initial feedstock and post-polymerization aliquots. We factored this loss into our calculations when determining the composition of the raw RHP. This calibration allows us to obtain an accurate estimation of the raw product composition.

Determining Composition of Purified RHP Products: Quantitative 13C NMR was performed using an inverse-gated decoupling sequence ("zgig" on a Bruker Avance III instrument equipped with a 5 mm 1H/BB Prodigy cryo-probe) with a 60s relaxation delay (150 MHz, CDCl3). RHPs monomer composition ratios were determined by integrating the peaks labeled with c (MMA), d (OEGMA), h (EHMA), p (NHSMA) and calculating the fraction for each monomer.

Gel Permeation Chromatography (GPC): THF GPC data were collected via an instrument equipped with an isocratic pump (1260 Infinity II, Agilent) and a differential refractometer (dRI) detector. Size exclusion columns were maintained at a temperature of 40 °C using THF as the mobile phase. All sample solutions were filtered through a 0.45 μm PTFE filter before injection. The molecular weight of each polymer was determined based on a PMMA calibration curve. DMF GPC data were collected using a Malvern OmniSEC equipped with refractive index, light scattering, and intrinsic viscosity detectors calibrated with a single PMMA standard. Analysis was performed in dimethylformamide containing 0.2% lithium bromide running at 0.7 mL/min and 55 °C. All sample solutions were prepared at a concentration of 2 mg/ml and filtered through a 0.45 μm PTFE filter before injection.

Spin Dialysis and Lyophilizing: Spin dialysis was performed using MilliporeSigma Amicon Ultra-15 Centrifugal Filter Units (Part number: UFC900324), diameter: 29.7 mm, MWCO: 3,000 d, volume: 15 mL. After pentane precipitation, RHPs were dissolved in 10 mL DI water

and transferred to the filter unit. The mixture was centrifuged at 6000 rpm for 30 min until the remaining solution was around 2 mL. 10 mL DI water was then added to the remaining mixture and the centrifuge separation was repeated. A total of 30 mL dialysate was collected for each RHP, and no residual monomers were observed in the NMR analysis of the final products. After dialysis, the purified products were dissolved in water to make a concentrated solution and transferred into vial for lyophilizing (using a 8L LABCONCO FreeZone Freeze Dryers). The purified RHPs were then characterized by NMR and GPC and used in depolymerization experiments.

High Performance Liquid Chromatography (HPLC): The monomer composition ratio after depolymerization were determined using a RP-HPLC (Agilent) on a C4 column (Viva 5 µm, 200 mm x 4.6 mm). The flow rate was 1 ml/min. Elution was monitored with a diode array detector at wavelengths of 210 nm. When measuring the UV- vis spectra of the monomers (MMA, EHMA, OEGMA, and NHSMA), no significant absorbance was observed above 220 nm. The selection of a 210 nm detection wavelength was based on its high sensitivity to detect ester compounds.1 Elution gradient: 20 min 20%IPA/80%water plus 0.1% (v/v) TFA then a 30 min linear AB gradient from 0% B to 50% B, where solvent A consisted of 20%IPA/80%water plus 0.1% (v/v) TFA and solvent B consisted of acetonitrile plus 0.1% (v/v) TFA.

Composition and Sequence Simulation: Composition and sequence simulations of the synthesized RHPs were performed using the *RHPapp*.[25] The simulation inputs (including the monomer composition, RAFT to initiator ratio, conversion fraction, and average degree of polymerization) were based on experimental design. A total of 100,000 RHP sequences were generated for each model RHP batch to ensure that there are sufficient chains to yield composition distributions for statistical analysis.

Hydrophobicity Calculation and Statistical Analysis: The RHP chain hydrophobicity was calculated based on previously reported hydrophilic- lipophile balance (HLB) scale estimated by considering the functional groups on the monomer side chains.3 The four monomers used in this experiment and simulation were each assigned a HLB index, where MMA was assigned a value of 8.45, EHMA a value of 5.125, OEGMA a value of 11.42, and NHSMA a value of 12.775.2,4,5 The hydrophobicity of each chain of the RHP and the distribution of that HLB value were calculated based on value assignment above.

General Procedures of RAFT Polymerization: A 20 mL Schlenk tube was charged with MMA, OEGMA, EHMA, NHSMA, and 2.5 mL DMF.

4-Cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl]pentanoic acid, and AIBN were added, and the reaction mixture was degassed by three freeze-pump-thaw cycles. After stirring at 80 °C for a given time (**RHP1-2**: 2.5 hr; **RHP3-4**: 1.5 hr), the reaction mixture was cooled in liquid nitrogen and the polymer was precipitated by dropwise addition of the mixture into stirring pentane. The yellow precipitate was then dissolved in water and transferred to a 3000 MWCO dialysis bag. The polymer was dialyzed against water and lyophilized to yield the final product. The purified polymer was subsequently characterized using GPC, 1H-NMR, and 13C- NMR.

RHP-1: Yellow gel. Prepared using MMA (0.100 g, 1.0 mmol), OEGMA (2,236 g, 4.5 mmol), EHMA (0.396 g, 2.0 mmol), NHSMA (0.457 g, 2.5 mmol), 4-Cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl]pentanoic acid (40 mg, 0.1 mmol), AIBN (3 mg, 0.02 mmol), and 2.5 mL DMF following the general RAFT polymerization procedures. Yield: 78 %.

RHP-2: Yellow gel. Prepared using MMA (0.402 g, 4.0 mmol), OEGMA (1.987 g, 4.0 mmol), EHMA (0.198 g, 1.0 mmol), NHSMA (0.183 g, 1.0 mmol), 4-Cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl]pentanoic acid (40 mg, 0.1 mmol), AIBN (3 mg, 0.02 mmol), and 2.5 mL DMF following the general RAFT polymerization procedures. Yield: 62 %.

RHP-3: Yellow gel. Prepared using MMA (0.502 g, 5.0 mmol), OEGMA (0.994 g, 2.0 mmol), EHMA (0.396 g, 2.0 mmol), NHSMA (0.183 g, 1.0 mmol), 4-Cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl]pentanoic acid (40 mg, 0.1 mmol), AIBN (3 mg, 0.02 mmol), and 2.0 mL DMF following the general RAFT polymerization procedures. Yield: 47 %.

RHP-4: Yellow gel. Prepared using MMA (0.251 g, 2.5 mmol), OEGMA (1.242 g, 2.5 mmol), EHMA (0.791 g, 4.0 mmol), NHSMA (0.183 g, 1.0 mmol), 4-Cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl]pentanoic acid (40 mg, 0.1 mmol), AIBN (3 mg, 0.02 mmol), and 2.0 mL DMF following the general RAFT polymerization procedures. Yield: 45 %.

Chapter 6

Deep learning driven heteropolymer design

6.1 Learning the language of life, synthetically

Proteins are essential for life. They perform many key functions from catalyzing crucial biochemical reactions, to transmitting chemical information, to providing nanoscale structure and more.[1] Proteins are natural polymers – physical strings of amino acids. Computationally, protein sequences can be viewed as sentences with a word dictionary size of 20 amino acids. The "grammar" of proteins affects how they fold into structures and thus how they function. It can thus be said that if we learn the "language" of protein sequences we will have learned the language of life.[2–7]

There has been growing interest in harnessing the power of proteins and recapitulating their function in engineered materials. Synthetic protein mimics lie on a sliding scale from most mimetic to most synthetic. An example of the most mimetic are proteins engineered via directed mutagenesis.[8] In the middle of this spectrum are the de novo design of proteins and synthesized polypeptides with specific sequences.[9] To be truly synthetic amino acids can be replaced as the polymeric building blocks with synthetic chemical monomer equivalents. These are synthetic heteropolymers and can range in statistical definition from fully sequence defined (as natural proteins are),[10] to only statistically defined.[11–15] Generally, the more synthetic a protein mimic is the more stable and compatible it is in commercial applications, better matching the timescales and length scales that humans demand.

In the process of considering and designing systems of synthetic protein mimics we can ask some important biological questions: What is the smallest alphabet size required to synthetically recapitulate protein function? Nature landed on 20 natural amino acids as its alphabet size after billions of years of evolution. Why 20? Is this fully necessary or is there redundancy in the alphabet of life? Performing dimensionality reduction to reduce system complexity would be ideal to improve system designability and interpretability. Just as English language sentences can be clustered into lexemes (*e.g.,* adjectives, nouns, verbs, and conjugations), protein sequences of amino acids can be clustered by groupings ranging from chemical properties (*i.e.,* lipophilic, hydrophilic, negative, positive, and aromatic) to hydrophobicity to evolutionary relationships.[16] Ideally, rather than using 20 building blocks a synthetic protein sequence would operate purely at a level of reduced dimensionality. In this chapter we use state-of-the-art deep learning techniques (*i.e.,* deep learning and large language models) to investigate synthetic protein mimics.

6.2 A hybrid variational autoencoder for designing synthetic protein mimics[3]

As it remains a challenge to synthesize read-write-able, sequence-specific polymers, progress has been made on developing biological mimics using statistically defined sequences only.[11–15] Designing statistically random heteropolymers (RHP) is challenging because both the exact monomeric sequences and conformations of synthetic RHP chains are nondeterministic. Traditional protein design methods rely heavily on high-throughput sequencing data and 3D structures.[8,9] Without exact sequences and structures, we are lacking in rational design principles for creating functional RHP chains.

We believe the work presented, and literature cited in **Chapters 4** and **5** cover the most up-to-date understanding of computational methods to generate and design RHP sequences. Moving forwards, just as the *RHPapp* took inspiration from bioinformatics to develop sequence analysis frameworks, we should leverage the state-of-the-art sequence analysis tools used in biology today - deep learning models. To our knowledge there is no prior literature using deep learning methods to model RHP sequences. The closest example used a hidden Markov model to characterize the functionality of proton transporting RHPs.[17] Looking to deep learning work on proteins for inspiration, variational autoencoder (VAE)-based architectures were some of the first model classes used to identify latent representations for biological sequences, and are useful in downstream tasks like identifying mutation effects and designing novel functional proteins.[18–21] However, a key difference is unlike how protein models take single chains as input, we need to instead model and analyze RHPs as an ensemble that is comprised of individual sequences.[13,14] Individual RHP chains are statistically, sequence-defined. They have different monomeric sequences but are statistically similar at the segment level, making them an ideal platform to materialize the segmental information extracted from primary protein sequences.[12–14]

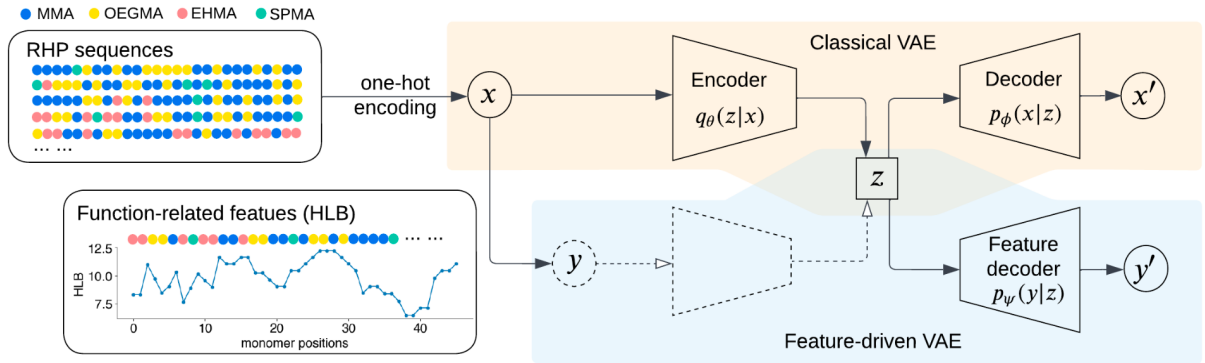


Figure 6.1 | *DeepRHP* model architecture consisting of a classical VAE equipped with an additional feature-based VAE.

[3] Ch. 6.2 was done in collaboration with Shuni Li, Zhiyuan Ruan, and Prof. Haiyan Huang.

Here we developed *DeepRHP*, a modified variational autoencoder trained under a semi-supervised framework, for modeling general RHP sequence data and discovering RHP compositions for protein function (**Fig. 6.1**). *DeepRHP* learns interpretable, low-dimensional, latent representations for RHP sequences and provides a platform to parameterize and visualize both chemical characteristics and sequential arrangement at the segmental level using the autoencoder model and establish design rules to modulate RHPs' similarity to proteins as an ensemble (**Fig. 6.2**).

DeepRHP was trained on protein sequences collected from the UniprotKB database.[22] Each sequence was first reduced into four types of pseudo-residue: hydrophilic, hydrophobic, very hydrophobic and charged, on the basis of the respective side-chain hydrophobicity of amino acids (**Table S6.2**).[23] Initial sequence analysis was also tested using two or eight pseudo-residues (**Tables S6.3** and **S6.4**). For each protein, the reassigned protein sequence was truncated into a collection of 50-mer segments for analysis (**Fig. S6.1**). The 50-residue interval was chosen because it captured most short- and long-range residue–residue contacts in native protein conformation. The residue order and local hydrophobicity of each 50-mer were extracted and represented by a low-dimensional vector, which was projected onto a space comprising the first two principal components (PC1 and PC2) from a Principal Component Analysis (PCA).[24,25] To a first approximation, the PC1 correlates with the apparent hydrophobicity of the 50-mer; the PC2 correlates with the sequential arrangement of different blocks within the 50-mer. A library of hypothetical chains containing the same four blocks arranged in different orders was analyzed for illustration and can be clearly differentiated based on the 2D sequence analysis (**Fig. 6.2a**). The PC1's dependence on the PC2 further highlights the importance of considering the block sequence along a chain and its potential to modulate interchain interactions.

The PCA analysis results effectively capture key sequence characteristics for known protein families (**Fig. 6.2b**). Membrane proteins have a subpopulation shifted along the PC1 axis compared to that of globular proteins. Thus, the model captures the distinction between the two protein families through their 50-mer segmental hydrophobicity. The large spatial overlap between the two protein families reflects that common protein motifs are present in both protein families. The two protein families were indistinguishable along the PC2 axis, and both are more concentrated in the center region of PC2. It is reasonable to speculate that proteins with alternating short hydrophobic and hydrophilic segments are less prone to aggregation and are preferable through evolution.

A library of RHP ensembles was designed to match the pseudo-residues used in the initial protein sequence analysis, using 2–4 out of six selected monomers including methyl methacrylate (MMA), 2-ethylhexyl methacrylate (2-EHMA), 3-sulfopropyl methacrylate potassium salt (3-SPMA), 2-(dimethylamino) ethyl methacrylate (DMAEMA) and oligo (ethylene glycol) methacrylate (OEGMA) with number-average molecular weight (M_n) of 300 or 500 Da, respectively (**Table S6.1**). The MMA:EHMA molar ratio was varied in 5 or 10% increments to modulate the distribution of segmental hydrophobicity. For each RHP ensemble, 2,000 simulated RHP sequences were generated using the *RHPapp 1.0*.[14] This

sample size is representative of the ensemble, because the occupied PCA space converged at roughly 1,500 chains (**Fig. S6.2**). The segment distribution from proteins and RHP1–7 was projected onto the same PCA space for similarity comparison (**Fig. 6.2c**). The occupied PCA space shifts along the PC1 axis, with some overlapping regions, as the monomer composition varies. RHP ensembles populating the left of the PCA space (lower PC1 value) have more hydrophobic segments than those to the right (**Fig. S6.3**). The PC2 distribution of RHP ensembles are similar to each other and comparable to those of both protein families. RHP ensembles can be designed to match the segmental diversity of both protein families. There is a sizeable overlap between four RHP ensembles, RHP1–4, with globular proteins; RHP1–7 have some overlap with membrane proteins.

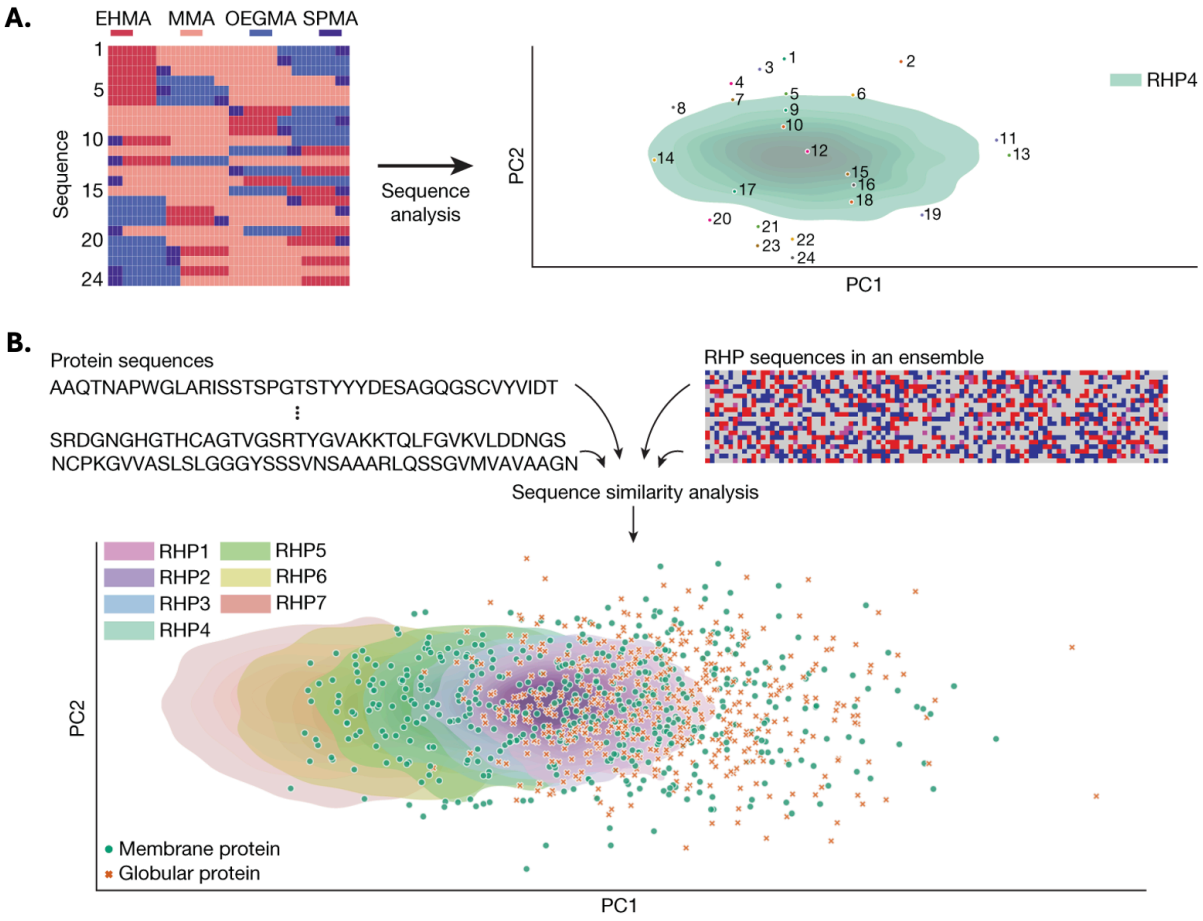


Figure 6.2 | A. Left: permutated sequences in descending order of PC2 value from top to bottom. Right: projection of 24 permutated sequences with RHP4 composition ratio and RHP4 ensemble onto the PCA space. **B.** 2D sequence analysis of proteins and heteropolymers to compare similarities in their chemical characteristics at the segmental level. 100 membrane protein sequences (green dot) and 100 globular protein sequences (orange cross) (both randomly selected) are shown for clarity. Each shaded area is a univariate distribution of an ensemble of 2,000 RHP chains using kernel density estimation.

Next, we demonstrated how *DeepRHP* can be used to accelerate RHP design by identifying promising monomer compositions. Aquaporins (Aqp) are membrane channel proteins that facilitate water transport between cells. Membrane proteins are unstable and prone to aggregation even under mild experimental conditions. Panganiban et al. successfully stabilized Aquaporin Z (AqpZ) and preserved its function in non-native environments with the presence of RHPs.[11] The RHP system used had 70% hydrophobic monomers and 30% hydrophilic monomers, selected based on a crude protein surface analysis on four protein sequences. We first validated this distribution of monomer hydrophobicities using our model. The latent factors of the 2-monomer RHPs and natural proteins are projected onto a two-dimensional space using PCA (**Fig. 6.3a**). All 2-monomer RHPs are composed of one hydrophobic monomer (EHMA) and one hydrophilic monomer (OEGMA). The compositions of *RHP A* through *RHP E* were selected to sufficiently reflect this hydrophobicity range (**Table S6.1**). We observed that PC1 correlates with hydrophobicity as RHPs span left to right, least hydrophobic to being most hydrophobic respectively. Most membrane and globular proteins overlapped with *RHP B* and *RHP C*, suggesting these two RHP compositions are most similar to natural proteins. Most hydrophobic membrane proteins overlapped with *RHP B* (70% hydrophobic, 30% hydrophilic), confirming that 30:70 is a good balance for the 2-monomer system.

We then fine-tuned the performance of the 30:70 distribution of hydrophilic and hydrophobic monomers by increasing the number of monomers from two to four (**Fig. 6.3b**). A library of 4-monomer-based RHPs was designed by varying the MMA:EHMA ratio (**Table S6.1**). Each of *RHP 1* through *RHP 7* is still composed of 30% hydrophilic monomers (OEGMA, SPMA) and 70% hydrophobic monomers (MMA, EHMA). Panganiban et al. did not rationalize the choice of four monomers for their design of protein-like RHPs. Our approach explains why the two-monomer alphabet size is insufficient. In **Figure 6.3b**, each of the RHP ensembles can be considered as a subset of *RHP B* and occupied a much more localized natural protein sequence space with smaller variance. We then projected AqpZ onto the 2-monomer (**Fig. 6.3c**) and 4-monomer PCA spaces (**Fig. 6.3d**). In the 2-monomer setting, the *RHP B* space is much larger than the span of AqpZ. In the 4-monomer setting, the AqpZ projections cover the *RHP 4* and *RHP 5* spaces almost entirely. Therefore, we believe the 2-monomer sequence space is too broad with respect to proteins while the 4-monomer sequence space is more localized.

In addition to providing heuristics regarding the number of monomers, *DeepRHP* sheds light on the choice of monomer compositions. In **Figure 6.3d** there is a large overlap between the projected proteins and the *RHP 4* and *RHP 5* contours. Panganiban et al. demonstrated that the optimal RHP has the same monomer ratio as that of *RHP 4* and is capable of stabilizing AqpZ. Thus, the overlap between RHPs and AqpZ in the PCA space can modulate their sequence correlation and molecular interactions in the aqueous solution. This indicates that the latent embeddings discovered by *DeepRHP* are chemically meaningful and can play a key role in discovering functional RHPs.

Overall, *DeepRHP* holds great promise for the future of integrating deep learning techniques, specifically VAEs, into RHP design. The hybrid VAE architecture of *DeepRHP* is

flexible and can be trained on any sequence family with variable sequence lengths and no multiple sequence alignment is needed. By exploring the low dimensional heteropolymer sequence representations learned by a semi-supervised framework, we provided fundamental insight on if the 'language of life' learned by models is transferable to synthetic protein mimics. We showed that a binary system was insufficient to recapitulate a protein-like latent space representation, but that 4 monomers was sufficient.

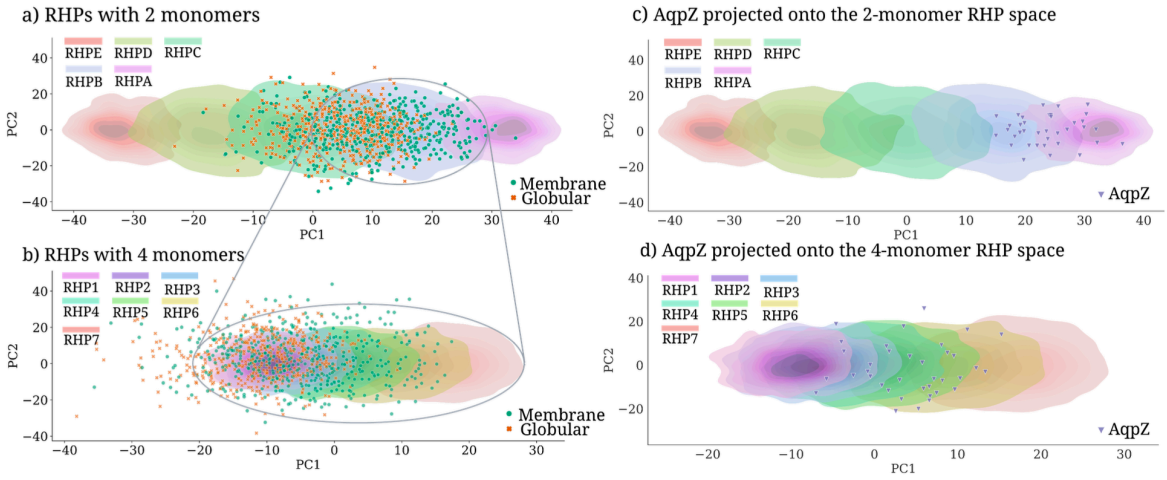


Figure 6.3 | PCA projections of RHP and protein latent factors. A. and **B.** project membrane and globular proteins onto two and four-monomer RHP space, respectively. **C.** and **D.** project AqpZ onto the same two RHP spaces.

6.3 Transferability of protein language models to synthetic mimics

The preceding section delved into the design of synthetic protein mimics through the utilization of a reduced alphabet set, shrinking from 20 amino acids to 8, 4, and 2. However, the approach employed by *DeepRHP* involved a semi-supervised learning framework, potentially leading to a bias towards overemphasizing the significance of hydrophobicity. In contrast, Ieremie et al. developed a fully unsupervised three-layer Bidirectional Long Short-Term Memory (BiLSTM) protein language model (PLM).[16] They trained this model on protein sequences that underwent conversion into a series of reduced amino acid alphabets (*e.g.,* 18, 17, 14, 12, 7, 5, and 4). Their investigation focused on elucidating the impact of diminishing protein sequence information on learned representations and downstream task performance. As anticipated, PLMs trained on reduced alphabets generally exhibited inferior performance compared to those trained on sequences encompassing the full 20 amino acids. Nevertheless, among the downstream prediction tasks (enzyme classification, homology detection, and three protein engineering datasets), certain PLMs trained on reduced alphabet sets demonstrated comparable performance to their full alphabet-trained counterparts. Promisingly, these findings suggesting that employing alphabet sets with reduced dimensionality could prove effective for protein mimics. However, unlike the outcomes observed in *DeepRHP*, PLMs trained on a reduced dataset of 4 representative amino acids exhibited exceedingly poor performance across all downstream tasks. A minimum alphabet size of 12 was necessary to attain moderately comparable performance to PLMs trained on the complete 20-amino acid alphabet. Furthermore, they highlighted that not all clustering schemes yield equivalent results. Simply clustering amino acids by hydrophobicity to 14 resulted in inferior downstream task performance compared to a more refined clustering scheme that had a smaller alphabet size of 12.

In this study, we aimed to bridge the methods and results of previous studies. Both *DeepRHP* and the models utilized by Ieremie et al. are considered small by contemporary standards and are now outdated compared to the latest advancements, particularly transformers that leverage attention.[26] Today, large-scale models are trained on billions of uncharacterized protein sequences. Using a fully unsupervised, state-of-the-art, large language transformer model should maximize the advance in our understanding of the applicability and transferability of protein language models to synthetic protein mimics. One of the most notable models in this domain is AlphaFold2.[27] However, AlphaFold2 and similar models utilize evolutionary information from sequence alignments and genetic databases to augment and accelerate learning. This is undesirable for transferability to synthetic protein mimics, which lack evolutionary histories and databases. Consequently, we chose to build upon the work of ProtTrans,[6] which provided six popular transformer architectures (BERT, Albert, Electra, T5, Transformer-XL, & XLNet), all pretrained solely on billions of raw protein sequences, without additional evolutionary data. Among the six ProtTrans models, we selected the T5 architecture as our starting point for experimentation as it had the best performance on every metric reported.

We used the pre-trained ProtT5 model to encode curated adversarial, random polypeptide (RPP) sequences that were designed to test the robustness and transferability

of PLMs to synthetic protein mimics such as random heteropolymers (RHPs). Then similar to the method of the preceding section, we visualized the RPP latent space coordinates against those of known reference proteins from the SCOPe database.[28] However, the top 2 principal components resulting from a Principal Component Analysis (PCA) could only explain 7% and 6% respectively. This result supported our initial hypothesis that semi-supervised learning with hydrophobicity as an input may be overemphasizing its learned importance. Thus, rather than PCA, t-SNE was used for visualization.[29]

RPPs sampled from a complete, uniform distribution of 20-amino acids (AAs) were barely in distribution (**Fig. 6.4**). This indicates the model can distinguish discriminate the random sequences from those of natural proteins - suggesting successfully learned protein grammar. **Figure 6.4** further showed that when the number of AAs is incrementally reduced the RPP sequences shift further out of the natural protein distribution. These initial results are promising, as they provide preliminary validation for the use of t-SNE + SCOPe as a visualization platform and discriminatory evaluation metric. The fact that a uniform distribution does not fool our evaluation platform allows for more complex analysis exploring sequence distributions.

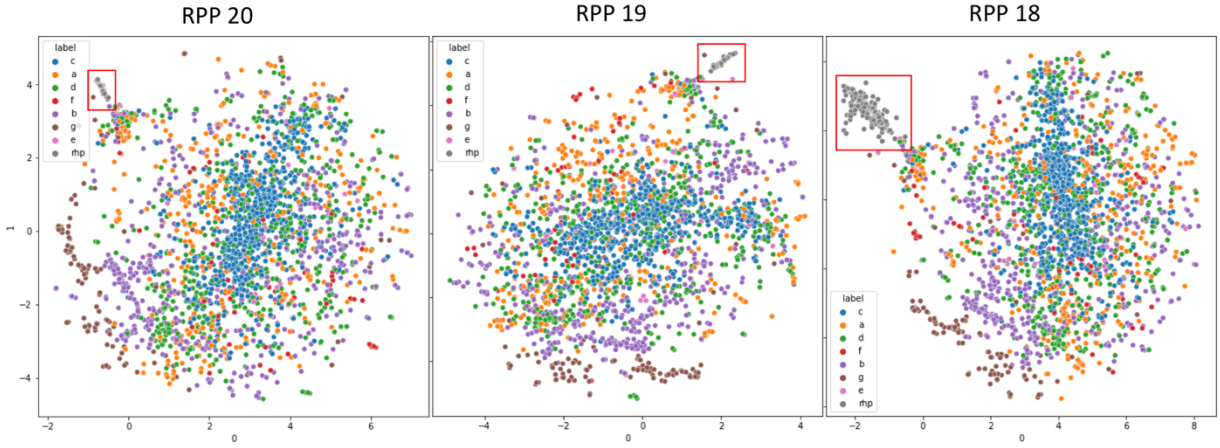


Figure 6.4 | t-SNE of 1024-dim encodings of SCOPe proteins and length-100 RPPs sampled from uniform distributions of 20, 19 and 18 unique amino acids of all possible combinations.

Shapes in t-SNE clusters are usually significant.[29] The long lines seen in RPP 20, 19 and 18 show that the absence or overabundance of some amino acids matters more others. Instead of using a non-biological uniform distribution, RPPs can be sampled from complete and reduced natural protein distributions of AAs. In other words, AAs that are more abundant in nature will be sampled at a higher frequency in RPPs. RPPs sampled from natural protein distributions blend in more with SCOPe proteins (**Fig. 6.5**). This trend continues as you reduce in AAs. Although the distribution of RPPs "peels off" more and more as number of AAs reduces, unlike in the uniform distribution case, the natural distribution RPP 18 are also

able to pull several natural proteins into its cluster. We see this pulling effect continue but slowly dwindle until RPP 10 (**Fig. S6.5**). By RPP 7 there is extremely clear t-SNE cluster separation between RPPs and SCOPe proteins, with 1 RPP sequence exception that prefers to cluster with SCOPe proteins (**Fig. 6.5**). There are no longer any outliers or overlapping regions between RPP and SCOPe clusters when reducing to less than 7 AAs. To determine which subsets of AAs are managing to successfully fool the transformer, the most naive approach is to take the top N most common AAs for each RPP N. Unfortunately, this approach did not yield convergence between RPP and SCOPe sequences (**Fig. S6.6**).

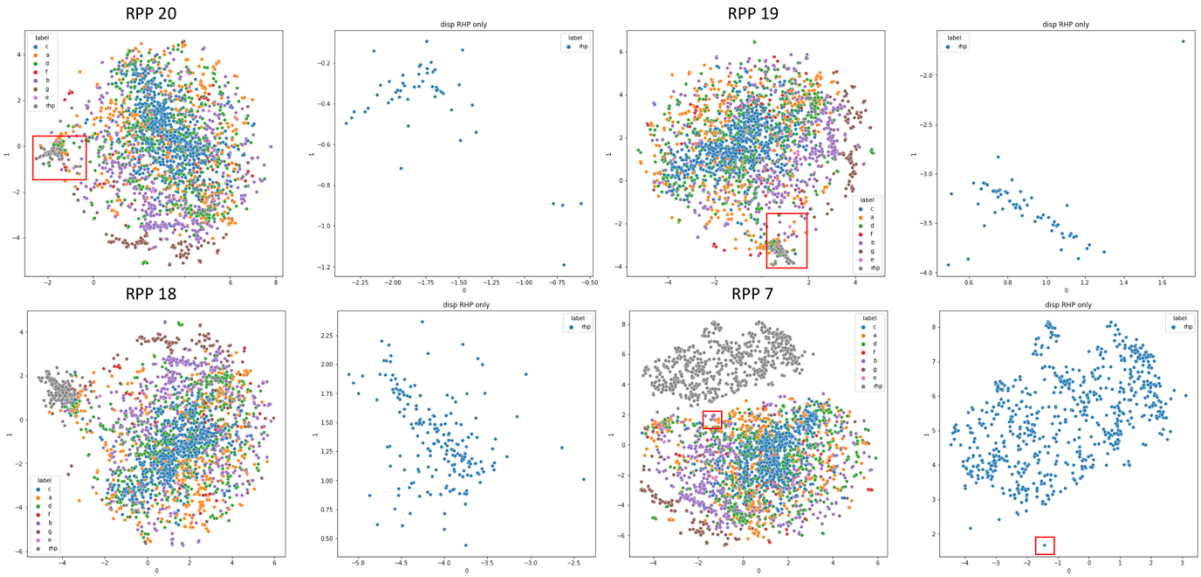


Figure 6.5 | t-SNE of 1024-dim encodings of SCOPe proteins and length-100 RPPs sampled from natural protein distributions of 20, 19,18 and 7 unique amino acids of all possible combinations. The left plots highlight the overall t-SNE, the right plots zoom in on only RPP sequences. For RPP7, a random subset of 700 sequences as sampled for visualization.

A noteworthy, exploitable feature of the SCOPe distribution is that it is sub-divided into specific subsets by protein structure. For example, the subset "c" that lies prominently in the center of SCOPe t-SNE visualizations consists of alpha/beta proteins (a/b) that are composed of alternating alpha helices and beta sheets, in which the beta sheets are mostly parallel to one another. In comparison, the subset "g" that clusters well amongst itself separate from the rest of SCOPe proteins is composed of "small proteins". We can plot and compare amino acid occurrence frequencies by sub-division (**Fig. S6.7**) and see that the a/b proteins (the center of SCOPe distribution on t-SNE) and our natural protein reference distribution from literature are very similar. However, the small proteins are both much shorter in average length (56 compared to 200) and have markedly different amino acid distribution. Small proteins are significantly richer in disulfide bond forming cysteine. It is no wonder the model can accurately group small proteins in latent space more similarly to one

another, separately to the other SCOPe proteins. It was a plausible hypothesis that sampling RPP sequences from the small protein distribution would result in proximal locations in latent space and thus proximity in t-SNE visualization, however, the resulting shift in t-SNE clustering is negligible (**Fig. S6.8**).

Instead of treating AAs as individual, independent alphabets, it would be more appropriate to think of them as words that share similarities to other words (AAs). We can reduce the alphabet size of RPPs not by random AA removal but by clustering AAs into several groups and then abstracting from the individual AA to the clustered-group level (**Fig. 6.6a**). 7 groups were hand-curated based on chemical similarity and a representative AA was chosen for each group based on abundance in natural proteins. The number of groups was chosen to match the lowest-alphabet size RPP for which we had a sequence that clustered with SCOPe, 7 AAs. RPPs can then be sampled from a reduced 7 AA dictionary with probabilities corresponding to the group probabilities in **Figure 6.6a**. **Figure 6.6b** shows a very extreme separation with no overlap. Even when adding 3 more amino acids to increase complexity and chemical diversity the separation from SCOPe is more extreme than when randomly removing AAs from the natural distribution (**Fig. 6.6c**). However, it can be seen in **Figure 6.6c** that there are some rare edge cases where outlier RPPs will be grouped with natural proteins when we increase the number of sequences generated and visualized. These results indicate that the wrong AAs were selected to represent each group or that the hand-curated groupings are detrimental.

Although it may visually appear that, performant outliers omitted, RPPs sampled from this group-distribution method results in strictly less-protein-like RPPs than those sampled from natural distributions and even uniform distributions, it should be noted that this is likely to be a t-SNE artifact. Spacing in t-SNE is often finicky.[29] A likely explanation for a more extreme spacing rather than less-protein-like RPPs, is increased similarity between RPPs generated by sampling from the group-distribution. More similar sequences will cluster better with each other and yield a very isolated cluster on the t-SNE. With the natural distribution, due to the high variance between RPPs due to randomly sampling from randomly cut subsets of AAs, there will be a gradient of similarity to natural proteins that results in a smaller spacing. We can highlight a limitation of using t-SNE as our evaluation metric in full effect by comparing a single sequence of an “optimized” RPP sequence sampled from a natural distribution of 10 amino acids (RPP 10), to a homo- polypeptide of all “C”s (homoPP-C). The resulting t-SNE + SCOPe visualizations are negligibly different with respect to distance in flattened latent space between RPP/homoPP-C and SCOPe proteins (**Fig. 6.7**). This shows we cannot use t-SNE + SCOPe as an evaluation metric in the case of a single RPP sequence.

Given the limitations and skepticism in going any deeper into latent space analysis using t-SNE + SCOPe as the sole evaluation metric, a viable route to assessing the applicability of language models trained on protein sequences to RPPs and RHPs is to probe direct applicability to improving downstream prediction tasks.

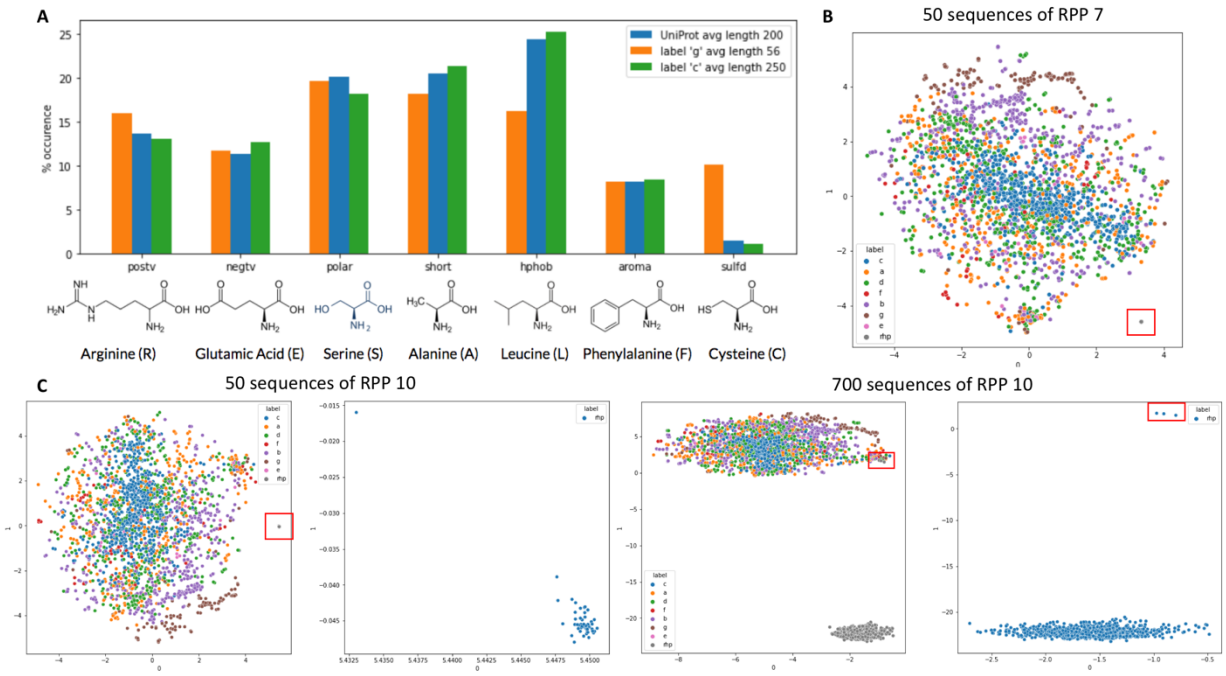


Figure 6.6 | A. Distributions of amino acids in naturally occurring proteins and sub-classes of proteins by secondary structure, grouped into 7 hand-curated groups based on chemical similarity: positive, negative, polar, short, hydrophobic, aromatic and sulfide (left to right). Representative amino acids for each group is displayed below the bar, chosen by highest abundance in natural proteins. **B.** t-SNE of 1024-dim encodings of SCOPe proteins and length-100 RPPs sampled from distribution of 7 amino acids from (a). **C.** same as (b) but with the 2nd most abundant amino acid added from the polar, short and hydrophobic groups for RPP 10, with same group distribution.

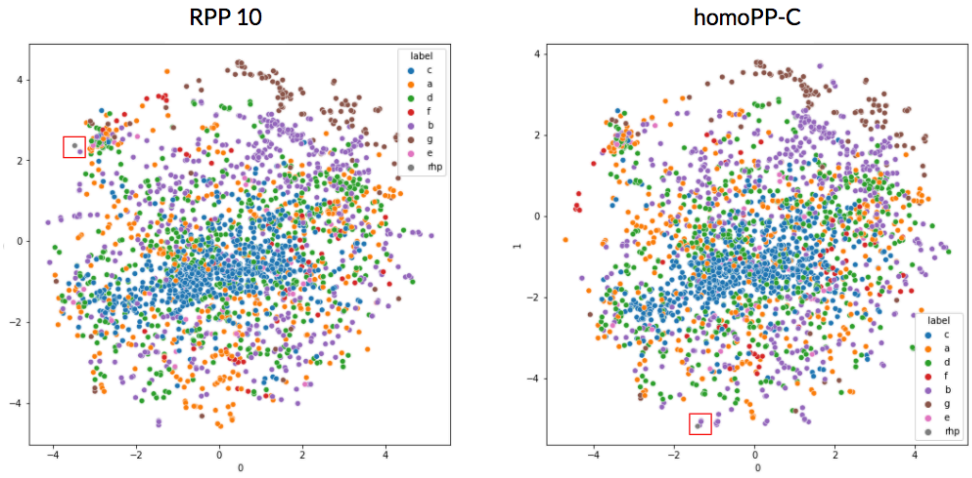


Figure 6.7 | t-SNE of 1024-dim encodings of SCOPe proteins and a single length-100 (left) RPP sampled from natural distribution of 10 unique amino acids and (right) homo–polypeptide of all cysteine (C).

Tamasi et al did active learning as part of a high-throughput robotic design platform to optimize RHPs as chaperones to increase the thermal stability of proteins. Their work provided a small experimental dataset mapping RHP batch characteristics to protein thermostability. We used this dataset as a ground truth to validate our supervised top model performance to predict protein stability. Prediction tasks using synthetic macromolecular sequence information as input is largely novel, with only very few mentions in published literature.[30]

We used ProtT5 outputted 1024-dim encodings of RHP-equivalent RPPs for inputs to a simple supervised regressor top model (**Scheme S6.1**). The results of varying input featurizations and model selections on 3 regression tasks are summarized in **Table 6.1**. No fine tuning or hyperparameter optimization was done for all input featurizations and models. The use of biologically rich sequence encodings over a simple batch-average characteristic feature vector as input slightly reduced RMSE for ridge regression - the best overall model for all 3 protein prediction tasks - and significantly improved Gaussian process regression (GPR) RMSE. The improvement in GPR is largely due to extreme outlier removal (**Fig. S6.9**). Finally, an ablation studies confirmed that the unsupervised-learned, biologically rich sequence information encoded in the 1024-dim input vector is what improved performance, and that it is not just an artifact of increasing the number of input features.

In conclusion, here we probed the extent to which a state-of-the-art transformer model can learn the language of proteins, by evaluating the ability for the model to discriminate hand-crafted adversarial examples of synthetic sequences of amino acids, defined as random polypeptides (RPPs). We then used latent space encodings as input to improve the predictive performance of a supervised, simple top model regressor on 3 protein-synthetic protein mimic hybrid function prediction tasks.

Table 6.1 | RMSE results on top model protein stability tasks evaluated over 50 trials. Best results are in bold.

Regression Task	Random Sampling	Mean Prediction	Batch-Ave + RidgeCV	Batch-Ave + GPR	RPP-seq encodings + RidgeCV	RPP-seq encodings + GPR
HRP	1144 ± 42	577	19.5 ± 1.7	246 ± 158	**17.8 ± 1.8**	18.9 ± 2.1
GOx	347 ± 12	174	11.0 ± 0.8	80 ± 77	**9.9 ± 0.9**	10.2 ± 1.1
Lip	1050 ± 47	522	15.7 ± 1.1	164 ± 133	**9.3 ± 1.0**	9.7 ± 1.3

6.4 Future work

The main bottleneck in the unsupervised latent space analysis portion of this work is the finicky and qualitative nature of tSNE as an evaluation tool. A more quantitative metric for evaluating how successfully "protein-like" a generated RPP sequence would be highly desirable. For example, a downstream top model classifier can be used to quantify the success of RPP adversarial attack. A similar quantitative downstream classifier would also improve *DeepRHP* as a synthetic protein mimic design tool.

One of the main takeaways from this study is that despite rigorous analysis into amino acid distributions of natural proteins, the ProtT5 transformer model was sophisticated enough to not be fooled by synthetic RPPs that have the same distributions. This clearly suggests that much like in the English language, if you randomly string together words by frequency in daily use, you will most likely still yield a nonsensical sentence. Perhaps like in sentences, where key word patterns are imperative to natural sentence structure, there are key amino acid motifs in proteins that must be present to fool the model. Future work should identify these key motifs and embed them into RPPs, such that RPPs are no longer truly random. Connecting this to synthetic random heteropolymers, a similar non-random sampling (*i.e.,* intentional sequence motif patterning) can be achieved by using chemical monomers with non-unitary reactivity ratios as was discussed in **Chapter 5**. The goal of this project was centered around interpretability rather than performance, however, simple hyperparameter tuning could drastically improve top model performance for all 3 model protein regression tasks. In particular, beyond regression hyperparameters, the conversion between synthetic monomers and biological amino acids was done based on chemical intuition. The amino acids selected to represent each monomer could be thoroughly explored and optimized - this would not only yield improved regression performance but also shed more light onto model interpretability and understanding how transformer models that claim to have learned the language of life can be used to better design synthetic protein mimics.

6.5 Methods

***DeepRHP* data:** The *RHPapp 1.0* was used to generate 2,000-15,000 chains for each RHP ensemble (DP = 50, **Table S6.1**).[14] To train the sequence autoencoder, 30,000 membrane protein sequences and 30,000 globular protein sequences with 50% identity threshold were collected from the UniProt database.[22] Sequences with uncommon amino acids were discarded. Each protein was reduced into four monomer codes (**Tables S6.2-6.4**). For each protein sequence, a set of consecutive protein motifs were collected by moving a 50-residue long window over each sequence with 15-residue long step size, resulting in a total of 1,046,845 training sequence motifs.

***DeepRHP* model:** an in-house latent variable model was developed to perform sequence dimensionality reduction and to learn protein/RHP sequence representations. The model was implemented based on a typical autoencoder architecture with an extra regression module to force the latent space to learn sequence hydrophilic-lipophilic balance (HLB) distributions (**Fig. 6.1**). Each protein sequence motif (in its RHP equivalent form) was first one-hot encoded and then passed through the encoder. The encoder embedded each sequence into a 16-dimensional latent vector, z, which was then fed into two parallel branches: the decoder and the regressor. The decoder intended to reconstruct the input sequence from z, whereas the regressor predicted the HLB values of an input sequence from z. The loss function was designed to be a weighted sum of the departure (for example, cross entropy) of the original and reconstructed sequence and the mean squared error of predicted and true HLB values. A meaningful low-dimensional latent space that captures both sequential features and HLB distributions was obtained by optimizing the reconstruction and regression loss together.

Both the encoder and the decoder were implemented with simple multilayer perceptrons. Each had three fully connected layers with 256, 128 and 64 hidden units. The regression module had two fully connected layers with 16 hidden units. Rectified linear unit non-linear functions were used throughout the network, except a Sigmoid activation that was used in the output layer of the decoder. The model was trained using the ADAM optimizer with a learning rate of 0.001. A learning rate scheduler was used to reduce the learning rate when the validation loss stopped improving. All model hyperparameters were optimized with Weights and Biases.[31] An LSTM-based autoencoder was also implemented, which demonstrated similar performance as the simpler autoencoder variant model.

***ProtT5* encoding pipeline:** *ProtT5* model has both encoder and decoder, for feature extraction we only load and use the encoder. Tokenizer and pre-trained model were loaded from Hugging Face.[32] Raw amino acid sequences were run through a batch encoder and the output 1024-dim encodings were saved (www,github.com/ivanjayapurna/ProtTrans_RHP).

Random polypeptide (RPP) sequence generation: RPP sequences were generated by mimicking random heteropolymer synthesis, where the next monomer in a growing sequence is randomly sampled based on a distribution. Here, we randomly sampled from 6 amino acid distributions and assumed all amino acids have uniform reactivity ratios (probability of subsequent residue is independent of previous position residue): (1)

Complete 20-amino acid, uniform distribution, (2) Reduced, uniform distributions, (3) Complete 20-amino acid, natural protein distribution,[33] (4) Reduced, natural protein distributions, (5) Complete 20-amino acid, specific natural protein subset distributions, and (6) Synthetic random heteropolymer reduced 4-amino acid distributions.

Latent space exploration of model discrimination ability via tSNE: Proteins were downloaded and formatted from the SCOPe dataset (10,000 sequences) to use for visualization.[28] SCOPe proteins were input to the ProtT5 encoding pipeline to generate 1024-dim encodings. Outputs were visualized on PCA (2D and 3D) no good separation was found. The first 2 principal components only can explain 7% and 6% respectively. Outputs were thus visualized on t-SNE. Perplexity and iteration counts were tuned.[29] SCOPe protein encodings and RPPs were co-visualized via t-SNE. RPP sequences were considered within natural protein distribution (*i.e.,* model has been fooled into thinking it is a real protein) if it is proximal in latent space on the flattened 2D visualization to the SCOPe protein sequences.

Protein-polymer hybrid top model training: Tamasi et al provided 624 random heteropolymer (RHP) batch-average characteristics (degree of polymerization and feeding monomer ratio) with labels for 3 different protein stabilization (prediction) tasks.[34] 3 regression models were trained to predict protein-RHP (synthetic protein mimic) stability for 3 different proteins with varying RHPs. Random sampling and mean prediction were used as trivial performance baselines. The same input features used in Tamasi et al. were used as a simple control model. Two models, ridge regression with cross validation (RidgeCV) and Gaussian process regression (GPR) were considered. RidgeCV is simple and often effective, and GPR is what was used in Tamasi et al. RHP input feature sets of batch-average characteristics were used to generate RPP sequences by a manually crafted conversion between synthetic chemistry monomers and natural amino acids (**Fig. S6.4**). RHP-representative RPP sequences were passed through ProtT5 to generate 1024-dim encodings to correspond as input to each output label. RidgeCV and GPR were used as models. Predictive performance is compared across models by averaged RMSE over 50 random trials.

Introduction

(1) Qin, B.; Liu, S.; Huang, Z.; Zeng, L.; Xu, J.-F.; Zhang, X. Closed-Loop Chemical Recycling of Cross-Linked Polymeric Materials Based on Reversible Amidation Chemistry. *Nat. Commun.* **2022**, *13* (1), 7595. https://doi.org/10.1038/s41467-022-35365-4.
(2) Abel, B. A.; Snyder, R. L.; Coates, G. W. Chemically Recyclable Thermoplastics from Reversible-Deactivation Polymerization of Cyclic Acetals. *Science* **2021**, *373* (6556), 783–789. https://doi.org/10.1126/science.abh0626.
(3) Christensen, P. R.; Scheuermann, A. M.; Loeffler, K. E.; Helms, B. A. Closed-Loop Recycling of Plastics Enabled by Dynamic Covalent Diketoenamine Bonds. *Nat. Chem.* **2019**, *11* (5), 442–448. https://doi.org/10.1038/s41557-019-0249-2.
(4) Milionis, A.; Sharma, C. S.; Hopf, R.; Uggowitzer, M.; Bayer, I. S.; Poulikakos, D. Engineering Fully Organic and Biodegradable Superhydrophobic Materials. *Adv. Mater. Interfaces* **2019**, *6* (1), 1801202. https://doi.org/10.1002/admi.201801202.
(5) Kim, D.-H.; Viventi, J.; Amsden, J. J.; Xiao, J.; Vigeland, L.; Kim, Y.-S.; Blanco, J. A.; Panilaitis, B.; Frechette, E. S.; Contreras, D.; Kaplan, D. L.; Omenetto, F. G.; Huang, Y.; Hwang, K.-C.; Zakin, M. R.; Litt, B.; Rogers, J. A. Dissolvable Films of Silk Fibroin for Ultrathin Conformal Bio-Integrated Electronics. *Nat. Mater.* **2010**, *9* (6), 511–517. https://doi.org/10.1038/nmat2745.
(6) De Pablo, J. J.; Jackson, N. E.; Webb, M. A.; Chen, L.-Q.; Moore, J. E.; Morgan, D.; Jacobs, R.; Pollock, T.; Schlom, D. G.; Toberer, E. S.; Analytis, J.; Dabo, I.; DeLongchamp, D. M.; Fiete, G. A.; Grason, G. M.; Hautier, G.; Mo, Y.; Rajan, K.; Reed, E. J.; Rodriguez, E.; Stevanovic, V.; Suntivich, J.; Thornton, K.; Zhao, J.-C. New Frontiers for the Materials Genome Initiative. *Npj Comput. Mater.* **2019**, *5* (1), 41. https://doi.org/10.1038/s41524-019-0173-4.
(7) Gormley, A. J.; Webb, M. A. Machine Learning in Combinatorial Polymer Chemistry. *Nat. Rev. Mater.* **2021**, *6* (8), 642–644. https://doi.org/10.1038/s41578-021-00282-3.
(8) Ensikat, H. J.; Ditsche-Kuru, P.; Neinhuis, C.; Barthlott, W. Superhydrophobicity in Perfection: The Outstanding Properties of the Lotus Leaf. *Beilstein J. Nanotechnol.* **2011**, *2*, 152–161. https://doi.org/10.3762/bjnano.2.19.
(9) Richardson, J. *Your complete and updated outlook for global polyethylene in 2023*. Asian Chemical Connections. https://www.icis.com/asian-chemical-connections/2023/01/your-complete-and-updated-outlook-for-global-polyethylene-in-2023/ (accessed 2024-04-29).
(10) Babu, K. M.; Chandrasekhara, S. M. *Bamboo Fibres: Processing, Properties and Applications*; Woodhead Publishing, 2022.
(11) EUBIO_Admin. *Market*. European Bioplastics e.V. https://www.european-bioplastics.org/market/ (accessed 2024-04-29).
(12) Vethaak, A. D.; Legler, J. Microplastics and Human Health. *Science* **2021**, *371* (6530), 672–674. https://doi.org/10.1126/science.abe5041.
(13) Mitrano, D. M.; Wick, P.; Nowack, B. Placing Nanoplastics in the Context of Global Plastic Pollution. *Nat. Nanotechnol.* **2021**, *16* (5), 491–500. https://doi.org/10.1038/s41565-021-00888-2.

(14) Gigault, J.; El Hadri, H.; Nguyen, B.; Grassl, B.; Rowenczyk, L.; Tufenkji, N.; Feng, S.; Wiesner, M. Nanoplastics Are Neither Microplastics nor Engineered Nanoparticles. *Nat. Nanotechnol.* **2021**, *16* (5), 501–507. https://doi.org/10.1038/s41565-021-00886-4.
(15) Breitenbach, J. Melt Extrusion: From Process to Drug Delivery Technology. *Eur. J. Pharm. Biopharm.* **2002**, *54* (2), 107–117. https://doi.org/10.1016/S0939-6411(02)00061-9.
(16) DelRe, C.; Huang, C.; Li, T.; Dennis, P.; Drockenmuller, E.; Xu, T. Reusable Enzymatic Fiber Mats for Neurotoxin Remediation in Water. *ACS Appl. Mater. Interfaces* **2018**, *10* (51), 44216–44220. https://doi.org/10.1021/acsami.8b18484.
(17) Bahrami, A.; Delshadi, R.; Assadpour, E.; Jafari, S. M.; Williams, L. Antimicrobial-Loaded Nanocarriers for Food Packaging Applications. *Adv. Colloid Interface Sci.* **2020**, *278*, 102140. https://doi.org/10.1016/j.cis.2020.102140.
(18) Sanluis-Verdes, A.; Colomer-Vidal, P.; Rodriguez-Ventura, F.; Bello-Villarino, M.; Spinola-Amilibia, M.; Ruiz-Lopez, E.; Illanes-Vicioso, R.; Castroviejo, P.; Aiese Cigliano, R.; Montoya, M.; Falabella, P.; Pesquera, C.; Gonzalez-Legarreta, L.; Arias-Palomo, E.; Solà, M.; Torroba, T.; Arias, C. F.; Bertocchini, F. Wax Worm Saliva and the Enzymes Therein Are the Key to Polyethylene Degradation by Galleria Mellonella. *Nat. Commun.* **2022**, *13* (1), 5568. https://doi.org/10.1038/s41467-022-33127-w.
(19) Tournier, V.; Duquesne, S.; Guillamot, F.; Cramail, H.; Taton, D.; Marty, A.; André, I. Enzymes' Power for Plastics Degradation. *Chem. Rev.* **2023**, *123* (9), 5612–5701. https://doi.org/10.1021/acs.chemrev.2c00644.
(20) DelRe, C.; Jiang, Y.; Kang, P.; Kwon, J.; Hall, A.; Jayapurna, I.; Ruan, Z.; Ma, L.; Zolkin, K.; Li, T.; Scown, C. D.; Ritchie, R. O.; Russell, T. P.; Xu, T. Near-Complete Depolymerization of Polyesters with Nano-Dispersed Enzymes. *Nature* **2021**, *592* (7855), 558–563. https://doi.org/10.1038/s41586-021-03408-3.
(21) DelRe, C.; Chang, B.; Jayapurna, I.; Hall, A.; Wang, A.; Zolkin, K.; Xu, T. Synergistic Enzyme Mixtures to Realize Near-Complete Depolymerization in Biodegradable Polymer/Additive Blends. *Adv. Mater.* **2021**, *33* (49), 2105707. https://doi.org/10.1002/adma.202105707.
(22) Kwon, J.; DelRe, C.; Kang, P.; Hall, A.; Arnold, D.; Jayapurna, I.; Ma, L.; Michalek, M.; Ritchie, R. O.; Xu, T. Conductive Ink with Circular Life Cycle for Printed Electronics. *Adv. Mater.* **2022**, *34* (30), 2202177. https://doi.org/10.1002/adma.202202177.
(23) Yeh, C. K.; Lin, C.; Shen, H. C.; Cheruiyot, N. K.; Camarillo, M. E.; Wang, C. L. Optimizing Food Waste Composting Parameters and Evaluating Heat Generation. *Appl. Sci.* **2020**, *10* (7), 2284. https://doi.org/10.3390/app10072284.
(24) Rauwendaal, C. *Polymer Extrusion*, 5th edition.; Hanser Publications ; Hanser Publication: Munich : Cincinnati, 2014.
(25) *Enzymes in Industry: Production and Applications*, 3rd, completely rev. ed ed.; Aehle, W., Ed.; Wiley-VCH: Weinheim, 2007.
(26) Headon, D. R.; Walsh, G. The Industrial Production of Enzymes. *Biotechnol. Adv.* **1994**, *12* (4), 635–646. https://doi.org/10.1016/0734-9750(94)90004-3.
(27) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera De La Cruz, M.; Xu, T. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018**, *359* (6381), 1239–1243. https://doi.org/10.1126/science.aao0335.

(28) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer–Protein Hybrids. *Adv. Mater.* **2022**, *34* (30), 2201809. https://doi.org/10.1002/adma.202201809.
(29) Kim, K. K.; Kim, R.; Kim, S.-H. Crystal Structure of a Small Heat-Shock Protein. *Nature* **1998**, *394* (6693), 595–599. https://doi.org/10.1038/29106.
(30) Van Montfort, R. L. M.; Basha, E.; Friedrich, K. L.; Slingsby, C.; Vierling, E. Crystal Structure and Assembly of a Eukaryotic Small Heat Shock Protein. *Nat. Struct. Biol.* **2001**, *8* (12), 1025–1030. https://doi.org/10.1038/nsb722.
(31) Wang, J.; Waltmann, C.; Umana-Kossio, H.; Olvera de la Cruz, M.; Torkelson, J. M. Heterogeneous Charged Complexes of Random Copolymers for the Segregation of Organic Molecules. *ACS Cent. Sci.* **2021**, *7* (5), 882–891. https://doi.org/10.1021/acscentsci.1c00119.
(32) Shakhnovich, E. I.; Gutin, A. M. Frozen States of a Disordered Globular Heteropolymer. *J. Phys. Math. Gen.* **1989**, *22* (10), 1647–1659. https://doi.org/10.1088/0305-4470/22/10/019.
(33) Jiang, T.; Hall, A.; Eres, M.; Hemmatian, Z.; Qiao, B.; Zhou, Y.; Ruan, Z.; Couse, A. D.; Heller, W. T.; Huang, H.; De La Cruz, M. O.; Rolandi, M.; Xu, T. Single-Chain Heteropolymers Transport Protons Selectively and Rapidly. *Nature* **2020**, *577* (7789), 216–220. https://doi.org/10.1038/s41586-019-1881-0.
(34) Ruan, Z.; Li, S.; Grigoropoulos, A.; Amiri, H.; Hilburg, S. L.; Chen, H.; Jayapurna, I.; Jiang, T.; Gu, Z.; Alexander-Katz, A.; Bustamante, C.; Huang, H.; Xu, T. Population-Based Heteropolymer Design to Mimic Protein Mixtures. *Nature* **2023**, *615* (7951), 251–258. https://doi.org/10.1038/s41586-022-05675-0.
(35) Hilburg, S. L.; Ruan, Z.; Xu, T.; Alexander-Katz, A. Behavior of Protein-Inspired Synthetic Random Heteropolymers. *Macromolecules* **2020**, *53* (21), 9187–9199. https://doi.org/10.1021/acs.macromol.0c01886.
(36) Gauthier, J.; Vincent, A. T.; Charette, S. J.; Derome, N. A Brief History of Bioinformatics. *Brief. Bioinform.* **2019**, *20* (6), 1981–1996. https://doi.org/10.1093/bib/bby063.

Ch 1. Biocatalysis under confinement: a molecular scale feedback loop

(1) DelRe, C.; Jiang, Y.; Kang, P.; Kwon, J.; Hall, A.; Jayapurna, I.; Ruan, Z.; Ma, L.; Zolkin, K.; Li, T.; Scown, C. D.; Ritchie, R. O.; Russell, T. P.; Xu, T. Near-Complete Depolymerization of Polyesters with Nano-Dispersed Enzymes. *Nature* **2021**, *592* (7855), 558–563. https://doi.org/10.1038/s41586-021-03408-3.
(2) DelRe, C.; Chang, B.; Jayapurna, I.; Hall, A.; Wang, A.; Zolkin, K.; Xu, T. Synergistic Enzyme Mixtures to Realize Near-Complete Depolymerization in Biodegradable Polymer/Additive Blends. *Adv. Mater.* **2021**, *33* (49), 2105707. https://doi.org/10.1002/adma.202105707.
(3) Kwon, J.; DelRe, C.; Kang, P.; Hall, A.; Arnold, D.; Jayapurna, I.; Ma, L.; Michalek, M.; Ritchie, R. O.; Xu, T. Conductive Ink with Circular Life Cycle for Printed Electronics. *Adv. Mater.* **2022**, *34* (30), 2202177. https://doi.org/10.1002/adma.202202177.
(4) Kalita, N. K.; Hakkarainen, M. Triggering Degradation of Cellulose Acetate by Embedded Enzymes: Accelerated Enzymatic Degradation and Biodegradation under Simulated Composting Conditions. *Biomacromolecules* **2023**, *24* (7), 3290–3303. https://doi.org/10.1021/acs.biomac.3c00337.
(5) Warner, M. J.; Lassa, J. P.; Narcross, H.; Commisso, A.; Ghosh, K.; Romero, M.; Schwartz, J. M.; Engler, A. C.; Kohl, P. A.; Leguizamon, S. C.; Jones, B. H. Chemical Recycling of Polybutadiene Rubber with Tailored Depolymerization Enabled by Microencapsulated Metathesis Catalysts. *ACS Sustain. Chem. Eng.* **2023**, *11* (39), 14538–14548. https://doi.org/10.1021/acssuschemeng.3c03907.
(6) Cooke, T.; Pomeroy, R. S. The Bioplastics Market: History, Commercialization Trends, and the New Eco-Consumer. In *Rethinking Polyester Polyurethanes*; Elsevier, 2023; pp 261–280. https://doi.org/10.1016/B978-0-323-99982-3.00003-1.
(7) Ghosh, K.; Jones, B. H. Roadmap to Biodegradable Plastics—Current State and Research Needs. *ACS Sustain. Chem. Eng.* **2021**, *9* (18), 6170–6187. https://doi.org/10.1021/acssuschemeng.1c00801.
(8) Nandakumar, A.; Chuah, J.-A.; Sudesh, K. Bioplastics: A Boon or Bane? *Renew. Sustain. Energy Rev.* **2021**, *147*, 111237. https://doi.org/10.1016/j.rser.2021.111237.
(9) Poulsen, K. R.; Snabe, T.; Petersen, E. I.; Fojan, P.; Neves-Petersen, M. T.; Wimmer, R.; Petersen, S. B. Quantization of pH: Evidence for Acidic Activity of Triglyceride Lipases. *Biochemistry* **2005**, *44* (34), 11574–11580. https://doi.org/10.1021/bi050634o.
(10) Grommet, A. B.; Feller, M.; Klajn, R. Chemical Reactivity under Nanoconfinement. *Nat. Nanotechnol.* **2020**, *15* (4), 256–271. https://doi.org/10.1038/s41565-020-0652-2.
(11) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera De La Cruz, M.; Xu, T. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018**, *359* (6381), 1239–1243. https://doi.org/10.1126/science.aao0335.
(12) Ng, C. H.; Mistoh, M. A.; Teo, S. H.; Galassi, A.; Ibrahim, A.; Sipaut, C. S.; Foo, J.; Seay, J.; Taufiq-Yap, Y. H.; Janaun, J. Plastic Waste and Microplastic Issues in Southeast Asia. *Front. Environ. Sci.* **2023**, *11*, 1142071. https://doi.org/10.3389/fenvs.2023.1142071.

(13) James, J. J.; Lakshmi, B. S.; Raviprasad, V.; Ananth, M. J.; Kangueane, P.; Gautam, P. Insights from Molecular Dynamics Simulations into pH-Dependent Enantioselective Hydrolysis of Ibuprofen Esters by Candida Rugosa Lipase. *Protein Eng. Des. Sel.* **2003**, *16* (12), 1017–1024. https://doi.org/10.1093/protein/gzg135.
(14) Trodler, P.; Schmid, R. D.; Pleiss, J. Modeling of Solvent-Dependent Conformational Transitions in Burkholderia Cepacia Lipase. *BMC Struct. Biol.* **2009**, *9* (1), 38. https://doi.org/10.1186/1472-6807-9-38.
(15) Peters, G. H.; Toxvaerd, S.; Olsen, O. H.; Svendsen, A. Computational Studies of the Activation of Lipases and the Effect of a Hydrophobic Environment. *Protein Eng. Des. Sel.* **1997**, *10* (2), 137–147. https://doi.org/10.1093/protein/10.2.137.
(16) Brogan, A. P. S.; Sharma, K. P.; Perriman, A. W.; Mann, S. Enzyme Activity in Liquid Lipase Melts as a Step towards Solvent-Free Biology at 150 °C. *Nat. Commun.* **2014**, *5* (1), 5058. https://doi.org/10.1038/ncomms6058.
(17) Siparsky, G. L.; Voorhees, K. J.; Dorgan, J. R.; Schilling, K. Water Transport in Polylactic Acid (PLA), PLA/ Polycaprolactone Copolymers, and PLA/Polyethylene Glycol Blends. *J. Environ. Polym. Degrad.* **1997**, *5* (3), 125–136. https://doi.org/10.1007/BF02763656.
(18) Ghanem, A. F.; Yassin, M. A.; Cosquer, R.; Gouanvé, F.; Espuche, E.; Abdel Rehim, M. H. Polycaprolactone Composite Films Infused with Hyperbranched Polyester/Reduced Graphene Oxide: Influence on Biodegradability, Gas/Water Transport and Antimicrobial Properties for Sustainable Packaging. *RSC Adv.* **2024**, *14* (9), 5740–5753. https://doi.org/10.1039/D3RA08948G.
(19) Schmitt, E. A.; Flanagan, D. R.; Linhardt, R. J. Importance of Distinct Water Environments in the Hydrolysis of Poly(DL-Lactide-Co-Glycolide). *Macromolecules* **1994**, *27* (3), 743–748. https://doi.org/10.1021/ma00081a019.
(20) Duan, Z.; Thomas, N. L. Water Vapour Permeability of Poly(Lactic Acid): Crystallinity and the Tortuous Path Model. *J. Appl. Phys.* **2014**, *115* (6), 064903. https://doi.org/10.1063/1.4865168.
(21) Cheng, P.; Fornasiero, F.; Jue, M. L.; Ko, W.; Li, A.-P.; Idrobo, J. C.; Boutilier, M. S. H.; Kidambi, P. R. Differences in Water and Vapor Transport through Angstrom-Scale Pores in Atomically Thin Membranes. *Nat. Commun.* **2022**, *13* (1), 6709. https://doi.org/10.1038/s41467-022-34172-1.
(22) Liu, M.; Wang, S.; Jiang, L. Nature-Inspired Superwettability Systems. *Nat. Rev. Mater.* **2017**, *2* (7), 17036. https://doi.org/10.1038/natrevmats.2017.36.
(23) Parlett, C. M. A.; Wilson, K.; Lee, A. F. Hierarchical Porous Materials: Catalytic Applications. *Chem Soc Rev* **2013**, *42* (9), 3876–3893. https://doi.org/10.1039/C2CS35378D.
(24) Jones, J. R.; Lee, P. D.; Hench, L. L. Hierarchical Porous Materials for Tissue Engineering. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2006**, *364* (1838), 263–281. https://doi.org/10.1098/rsta.2005.1689.
(25) Yang, X.-Y.; Chen, L.-H.; Li, Y.; Rooke, J. C.; Sanchez, C.; Su, B.-L. Hierarchically Porous Materials: Synthesis Strategies and Structure Design. *Chem. Soc. Rev.* **2017**, *46* (2), 481–558. https://doi.org/10.1039/C6CS00829A.

(26) Fotopoulou, K. N.; Karapanagioti, H. K. Degradation of Various Plastics in the Environment. In *Hazardous Chemicals Associated with Plastics in the Marine Environment*; Takada, H., Karapanagioti, H. K., Eds.; The Handbook of Environmental Chemistry; Springer International Publishing: Cham, 2017; Vol. 78, pp 71–92. https://doi.org/10.1007/698_2017_11.
(27) Chamas, A.; Moon, H.; Zheng, J.; Qiu, Y.; Tabassum, T.; Jang, J. H.; Abu-Omar, M.; Scott, S. L.; Suh, S. Degradation Rates of Plastics in the Environment. *ACS Sustain. Chem. Eng.* **2020**, *8* (9), 3494–3511. https://doi.org/10.1021/acssuschemeng.9b06635.
(28) Binda, G.; Costa, M.; Supraha, L.; Spanu, D.; Vogelsang, C.; Leu, E.; Nizzetto, L. Untangling the Role of Biotic and Abiotic Ageing of Various Environmental Plastics toward the Sorption of Metals. *Sci. Total Environ.* **2023**, *893*, 164807. https://doi.org/10.1016/j.scitotenv.2023.164807.
(29) Law, K. L.; Narayan, R. Reducing Environmental Plastic Pollution by Designing Polymer Materials for Managed End-of-Life. *Nat. Rev. Mater.* **2021**, *7* (2), 104–116. https://doi.org/10.1038/s41578-021-00382-0.
(30) Bergmann, M.; Collard, F.; Fabres, J.; Gabrielsen, G. W.; Provencher, J. F.; Rochman, C. M.; Van Sebille, E.; Tekman, M. B. Plastic Pollution in the Arctic. *Nat. Rev. Earth Environ.* **2022**, *3* (5), 323–337. https://doi.org/10.1038/s43017-022-00279-8.
(31) Windsor, F. M.; Durance, I.; Horton, A. A.; Thompson, R. C.; Tyler, C. R.; Ormerod, S. J. A Catchment-scale Perspective of Plastic Pollution. *Glob. Change Biol.* **2019**, *25* (4), 1207–1221. https://doi.org/10.1111/gcb.14572.
(32) Zhao, Y.; Norouzi, H.; Azarderakhsh, M.; AghaKouchak, A. Global Patterns of Hottest, Coldest, and Extreme Diurnal Variability on Earth. *Bull. Am. Meteorol. Soc.* **2021**, *102* (9), E1672–E1681. https://doi.org/10.1175/BAMS-D-20-0325.1.
(33) Yeh, C. K.; Lin, C.; Shen, H. C.; Cheruiyot, N. K.; Camarillo, M. E.; Wang, C. L. Optimizing Food Waste Composting Parameters and Evaluating Heat Generation. *Appl. Sci.* **2020**, *10* (7), 2284. https://doi.org/10.3390/app10072284.
(34) Afrin, S.; Uddin, Md. K.; Rahman, Md. M. Microplastics Contamination in the Soil from Urban Landfill Site, Dhaka, Bangladesh. *Heliyon* **2020**, *6* (11), e05572. https://doi.org/10.1016/j.heliyon.2020.e05572.
(35) Maphuhla, N. G.; Lewu, F. B.; Oyedeji, O. O. The Effects of Physicochemical Parameters on Analysed Soil Enzyme Activity from Alice Landfill Site. *Int. J. Environ. Res. Public. Health* **2020**, *18* (1), 221. https://doi.org/10.3390/ijerph18010221.
(36) Eggen, T.; Moeder, M.; Arukwe, A. Municipal Landfill Leachates: A Significant Source for New and Emerging Pollutants. *Sci. Total Environ.* **2010**, *408* (21), 5147–5157. https://doi.org/10.1016/j.scitotenv.2010.07.049.
(37) Thongyuan, S.; Khantamoon, T.; Aendo, P.; Binot, A.; Tulayakul, P. Ecological and Health Risk Assessment, Carcinogenic and Non-Carcinogenic Effects of Heavy Metals Contamination in the Soil from Municipal Solid Waste Landfill in Central, Thailand. *Hum. Ecol. Risk Assess. Int. J.* **2021**, *27* (4), 876–897. https://doi.org/10.1080/10807039.2020.1786666.
(38) Nannoni, F.; Mazzeo, R.; Santolini, R.; Protano, G. Multi-Matrix Environmental Monitoring to Assess Heavy Element Distribution around a Municipal Solid Waste Landfill

in Italy. *Int. J. Environ. Sci. Technol.* **2017**, *14* (12), 2591–2602. https://doi.org/10.1007/s13762-017-1342-y.
(39) Zhou, Q.; Yang, N.; Li, Y.; Ren, B.; Ding, X.; Bian, H.; Yao, X. Total Concentrations and Sources of Heavy Metal Pollution in Global River and Lake Water Bodies from 1972 to 2017. *Glob. Ecol. Conserv.* **2020**, *22*, e00925. https://doi.org/10.1016/j.gecco.2020.e00925.
(40) Smail, E. A.; Webb, E. A.; Franks, R. P.; Bruland, K. W.; Sañudo-Wilhelmy, S. A. Status of Metal Contamination in Surface Waters of the Coastal Ocean off Los Angeles, California since the Implementation of the Clean Water Act. *Environ. Sci. Technol.* **2012**, *46* (8), 4304–4311. https://doi.org/10.1021/es2023913.
(41) Tournier, V.; Duquesne, S.; Guillamot, F.; Cramail, H.; Taton, D.; Marty, A.; André, I. Enzymes' Power for Plastics Degradation. *Chem. Rev.* **2023**, *123* (9), 5612–5701. https://doi.org/10.1021/acs.chemrev.2c00644.
(42) Sanluis-Verdes, A.; Colomer-Vidal, P.; Rodriguez-Ventura, F.; Bello-Villarino, M.; Spinola-Amilibia, M.; Ruiz-Lopez, E.; Illanes-Vicioso, R.; Castroviejo, P.; Aiese Cigliano, R.; Montoya, M.; Falabella, P.; Pesquera, C.; Gonzalez-Legarreta, L.; Arias-Palomo, E.; Solà, M.; Torroba, T.; Arias, C. F.; Bertocchini, F. Wax Worm Saliva and the Enzymes Therein Are the Key to Polyethylene Degradation by Galleria Mellonella. *Nat. Commun.* **2022**, *13* (1), 5568. https://doi.org/10.1038/s41467-022-33127-w.
(43) James, B.; Ward, C.; Hahn, M.; Thorpe, S.; Reddy, C. *Minimizing the Environmental Impacts of Plastic through Eco-Design*; preprint; Chemistry, 2023. https://doi.org/10.26434/chemrxiv-2023-pxlxb.
(44) Li, Z.; Qi, W.; Feng, Y.; Liu, Y.; Ebrahim, S.; Long, J. Degradation Mechanisms of Oxytetracycline in the Environment. *J. Integr. Agric.* **2019**, *18* (9), 1953–1960. https://doi.org/10.1016/S2095-3119(18)62121-5.
(45) Ansink, E.; Wijk, L.; Zuidmeer, F. No Clue about Bioplastics. *Ecol. Econ.* **2022**, *191*, 107245. https://doi.org/10.1016/j.ecolecon.2021.107245.
(46) Bornscheuer, U.; Reif, O.; Lausch, R.; Freitag, R.; Scheper, T.; Kolisis, F.; Menge, U. Lipase of Pseudomonas Cepacia for Biotechnological Purposes: Purification, Crystallization and Characterization. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1994**, *1201* (1), 55–60. https://doi.org/10.1016/0304-4165(94)90151-1.
(47) Wurm, A.; Zhuravlev, E.; Eckstein, K.; Jehnichen, D.; Pospiech, D.; Androsch, R.; Wunderlich, B.; Schick, C. Crystallization and Homogeneous Nucleation Kinetics of Poly(ε-Caprolactone) (PCL) with Different Molar Masses. *Macromolecules* **2012**, *45* (9), 3816–3828. https://doi.org/10.1021/ma300363b.
(48) Chandler, K.; Eason, B.; Liotta, C. L.; Eckert, C. A. Phase Equilibria for Binary Aqueous Systems from a Near-Critical Water Reaction Apparatus. *Ind. Eng. Chem. Res.* **1998**, *37* (8), 3515–3518. https://doi.org/10.1021/ie970882j.
(49) Kitano, H.; Ichikawa, K.; Ide, M.; Fukuda, M.; Mizuno, W. Fourier Transform Infrared Study on the State of Water Sorbed to Poly(Ethylene Glycol) Films. *Langmuir* **2001**, *17* (6), 1889–1895. https://doi.org/10.1021/la0008986.
(50) Tyagi, H.; Kushwaha, A.; Kumar, A.; Aslam, M. A Facile pH Controlled Citrate-Based Reduction Method for Gold Nanoparticle Synthesis at Room Temperature. *Nanoscale Res. Lett.* **2016**, *11* (1), 362. https://doi.org/10.1186/s11671-016-1576-5.

Ch 2. Enzyme degradation mechanisms on and in semi-crystalline substrates

(1) Tournier, V.; Duquesne, S.; Guillamot, F.; Cramail, H.; Taton, D.; Marty, A.; André, I. Enzymes' Power for Plastics Degradation. *Chem. Rev.* **2023**, *123* (9), 5612–5701. https://doi.org/10.1021/acs.chemrev.2c00644.
(2) Sanluis-Verdes, A.; Colomer-Vidal, P.; Rodriguez-Ventura, F.; Bello-Villarino, M.; Spinola-Amilibia, M.; Ruiz-Lopez, E.; Illanes-Vicioso, R.; Castroviejo, P.; Aiese Cigliano, R.; Montoya, M.; Falabella, P.; Pesquera, C.; Gonzalez-Legarreta, L.; Arias-Palomo, E.; Solà, M.; Torroba, T.; Arias, C. F.; Bertocchini, F. Wax Worm Saliva and the Enzymes Therein Are the Key to Polyethylene Degradation by Galleria Mellonella. *Nat. Commun.* **2022**, *13* (1), 5568. https://doi.org/10.1038/s41467-022-33127-w.
(3) Ru, J.; Huo, Y.; Yang, Y. Microbial Degradation and Valorization of Plastic Wastes. *Front. Microbiol.* **2020**, *11*, 442. https://doi.org/10.3389/fmicb.2020.00442.
(4) Buchholz, P. C. F.; Feuerriegel, G.; Zhang, H.; Perez-Garcia, P.; Nover, L.; Chow, J.; Streit, W. R.; Pleiss, J. Plastics Degradation by Hydrolytic Enzymes: The PLASTICS-ACTIVE Enzymes Database— PAZy. *Proteins Struct. Funct. Bioinforma.* **2022**, *90* (7), 1443–1456. https://doi.org/10.1002/prot.26325.
(5) Din, S. U.; Kalsoom; Satti, S. M.; Uddin, S.; Mankar, S. V.; Ceylan, E.; Hasan, F.; Khan, S.; Badshah, M.; Beldüz, A. O.; Çanakçi, S.; Zhang, B.; Linares-Pastén, J. A.; Shah, A. A. The Purification and Characterization of a Cutinase-like Enzyme with Activity on Polyethylene Terephthalate (PET) from a Newly Isolated Bacterium Stenotrophomonas Maltophilia PRS8 at a Mesophilic Temperature. *Appl. Sci.* **2023**, *13* (6), 3686. https://doi.org/10.3390/app13063686.
(6) Lin, X.; Zhang, S.; Yang, S.; Zhang, R.; Shi, X.; Song, L. A Landfill Serves as a Critical Source of Microplastic Pollution and Harbors Diverse Plastic Biodegradation Microbial Species and Enzymes: Study in Large-Scale Landfills, China. *J. Hazard. Mater.* **2023**, *457*, 131676. https://doi.org/10.1016/j.jhazmat.2023.131676.
(7) Yoshida, S.; Hiraga, K.; Takehana, T.; Taniguchi, I.; Yamaji, H.; Maeda, Y.; Toyohara, K.; Miyamoto, K.; Kimura, Y.; Oda, K. A Bacterium That Degrades and Assimilates Poly(Ethylene Terephthalate). *Science* **2016**, *351* (6278), 1196–1199. https://doi.org/10.1126/science.aad6359.
(8) Guo, R.-T.; Li, X.; Yang, Y.; Huang, J.-W.; Shen, P.; Liew, R. K.; Chen, C.-C. Natural and Engineered Enzymes for Polyester Degradation: A Review. *Environ. Chem. Lett.* **2024**, *22* (3), 1275–1296. https://doi.org/10.1007/s10311-024-01714-6.
(9) Holst, L. H.; Madsen, N. G.; Toftgård, F. T.; Rønne, F.; Moise, I.-M.; Petersen, E. I.; Fojan, P. *De Novo* Design of a Polycarbonate Hydrolase. *Protein Eng. Des. Sel.* **2023**, *36*, gzad022. https://doi.org/10.1093/protein/gzad022.
(10) Li, S.; McCarthy, S. Influence of Crystallinity and Stereochemistry on the Enzymatic Degradation of Poly(Lactide)s. *Macromolecules* **1999**, *32* (13), 4454–4456. https://doi.org/10.1021/ma990117b.
(11) Tokiwa, Y.; Calabia, B.; Ugwu, C.; Aiba, S. Biodegradability of Plastics. *Int. J. Mol. Sci.* **2009**, *10* (9), 3722–3742. https://doi.org/10.3390/ijms10093722.

(12) Wallace, N. E.; Adams, M. C.; Chafin, A. C.; Jones, D. D.; Tsui, C. L.; Gruber, T. D. The Highly Crystalline PET Found in Plastic Water Bottles Does Not Support the Growth of the PETase -producing Bacterium *Ideonella Sakaiensis*. *Environ. Microbiol. Rep.* **2020**, *12* (5), 578–582. https://doi.org/10.1111/1758-2229.12878.
(13) Thomsen, T. B.; Almdal, K.; Meyer, A. S. Significance of Poly(Ethylene Terephthalate) (PET) Substrate Crystallinity on Enzymatic Degradation. *New Biotechnol.* **2023**, *78*, 162–172. https://doi.org/10.1016/j.nbt.2023.11.001.
(14) Mileva, D.; Tranchida, D.; Gahleitner, M. Designing Polymer Crystallinity: An Industrial Perspective. *Polym. Cryst.* **2018**, *1* (2). https://doi.org/10.1002/pcr2.10009.
(15) Wei, X.-F.; Capezza, A. J.; Cui, Y.; Li, L.; Hakonen, A.; Liu, B.; Hedenqvist, M. S. Millions of Microplastics Released from a Biodegradable Polymer during Biodegradation/Enzymatic Hydrolysis. *Water Res.* **2022**, *211*, 118068. https://doi.org/10.1016/j.watres.2022.118068.
(16) Vethaak, A. D.; Legler, J. Microplastics and Human Health. *Science* **2021**, *371* (6530), 672–674. https://doi.org/10.1126/science.abe5041.
(17) Mitrano, D. M.; Wick, P.; Nowack, B. Placing Nanoplastics in the Context of Global Plastic Pollution. *Nat. Nanotechnol.* **2021**, *16* (5), 491–500. https://doi.org/10.1038/s41565-021-00888-2.
(18) DelRe, C.; Jiang, Y.; Kang, P.; Kwon, J.; Hall, A.; Jayapurna, I.; Ruan, Z.; Ma, L.; Zolkin, K.; Li, T.; Scown, C. D.; Ritchie, R. O.; Russell, T. P.; Xu, T. Near-Complete Depolymerization of Polyesters with Nano-Dispersed Enzymes. *Nature* **2021**, *592* (7855), 558–563. https://doi.org/10.1038/s41586-021-03408-3.
(19) DelRe, C.; Chang, B.; Jayapurna, I.; Hall, A.; Wang, A.; Zolkin, K.; Xu, T. Synergistic Enzyme Mixtures to Realize Near-Complete Depolymerization in Biodegradable Polymer/Additive Blends. *Adv. Mater.* **2021**, *33* (49), 2105707. https://doi.org/10.1002/adma.202105707.
(20) Szabo; Horvath; Schad; Murvai; Tantos; Kalmar; Chemes; Han; Tompa. Intrinsically Disordered Linkers Impart Processivity on Enzymes by Spatial Confinement of Binding Domains. *Int. J. Mol. Sci.* **2019**, *20* (9), 2119. https://doi.org/10.3390/ijms20092119.
(21) Tennakoon, A.; Wu, X.; Paterson, A. L.; Patnaik, S.; Pei, Y.; LaPointe, A. M.; Ammal, S. C.; Hackler, R. A.; Heyden, A.; Slowing, I. I.; Coates, G. W.; Delferro, M.; Peters, B.; Huang, W.; Sadow, A. D.; Perras, F. A. Catalytic Upcycling of High-Density Polyethylene via a Processive Mechanism. *Nat. Catal.* **2020**, *3* (11), 893–901. https://doi.org/10.1038/s41929-020-00519-4.
(22) Matthews, B. W.; Breyer, W. A. Structure of Escherichia Coli Exonuclease I Suggests How Processivity Is Achieved. *Nat. Struct. Biol.* **2000**, *7* (12), 1125–1128. https://doi.org/10.1038/81978.
(23) Breyer, W. A.; Matthews, B. W. A Structural Basis for Processivity. *Protein Sci.* **2001**, *10* (9), 1699–1711. https://doi.org/10.1110/ps.10301.
(24) Divne, C.; Ståhlberg, J.; Teeri, T. T.; Jones, T. A. High-Resolution Crystal Structures Reveal How a Cellulose Chain Is Bound in the 50 Å Long Tunnel of Cellobiohydrolase I from *Trichoderma Reesei*1. *J. Mol. Biol.* **1998**, *275* (2), 309–325. https://doi.org/10.1006/jmbi.1997.1437.

(25) Beckham, G. T.; Matthews, J. F.; Peters, B.; Bomble, Y. J.; Himmel, M. E.; Crowley, M. F. Molecular-Level Origins of Biomass Recalcitrance: Decrystallization Free Energies for Four Common Cellulose Polymorphs. *J. Phys. Chem. B* **2011**, *115* (14), 4118–4127. https://doi.org/10.1021/jp1106394.
(26) Payne, C. M.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. Decrystallization of Oligosaccharides from the Cellulose Iβ Surface with Molecular Simulation. *J. Phys. Chem. Lett.* **2011**, *2* (13), 1546–1550. https://doi.org/10.1021/jz2005122.
(27) Li, Y.; Irwin, D. C.; Wilson, D. B. Processivity, Substrate Binding, and Mechanism of Cellulose Hydrolysis by Thermobifida Fusca Cel9A. *Appl. Environ. Microbiol.* **2007**, *73* (10), 3165–3172. https://doi.org/10.1128/AEM.02960-06.
(28) Horn, S. J.; Sikorski, P.; Cederkvist, J. B.; Vaaje-Kolstad, G.; Sørlie, M.; Synstad, B.; Vriend, G.; Vårum, K. M.; Eijsink, V. G. H. Costs and Benefits of Processivity in Enzymatic Degradation of Recalcitrant Polysaccharides. *Proc. Natl. Acad. Sci.* **2006**, *103* (48), 18089–18094. https://doi.org/10.1073/pnas.0608909103.
(29) Knott, B. C.; Crowley, M. F.; Himmel, M. E.; Ståhlberg, J.; Beckham, G. T. Carbohydrate–Protein Interactions That Drive Processive Polysaccharide Translocation in Enzymes Revealed from a Computational Study of Cellobiohydrolase Processivity. *J. Am. Chem. Soc.* **2014**, *136* (24), 8810–8819. https://doi.org/10.1021/ja504074g.
(30) Roberts, J. D.; Caserio, M. C. *Basic Principles of Organic Chemistry*, 2d ed.; W. A. Benjamin: Menlo Park, Calif, 1977.
(31) Mikita, N.; Cheng, I.; Fishovitz, J.; Huang, J.; Lee, I. Processive Degradation of Unstructured Protein by Escherichia Coli Lon Occurs via the Slow, Sequential Delivery of Multiple Scissile Sites Followed by Rapid and Synchronized Peptide Bond Cleavage Events. *Biochemistry* **2013**, *52* (33), 5629–5644. https://doi.org/10.1021/bi4008319.
(32) Gates, S. N.; Yokom, A. L.; Lin, J.; Jackrel, M. E.; Rizo, A. N.; Kendsersky, N. M.; Buell, C. E.; Sweeny, E. A.; Mack, K. L.; Chuang, E.; Torrente, M. P.; Su, M.; Shorter, J.; Southworth, D. R. Ratchet-like Polypeptide Translocation Mechanism of the AAA+ Disaggregase Hsp104. *Science* **2017**, *357* (6348), 273–279. https://doi.org/10.1126/science.aan1052.
(33) Satchwell, A. J.; Scown, C. D.; Smith, S. J.; Amirebrahimi, J.; Jin, L.; Kirchstetter, T. W.; Brown, N. J.; Preble, C. V. Accelerating the Deployment of Anaerobic Digestion to Meet Zero Waste Goals. *Environ. Sci. Technol.* **2018**, *52* (23), 13663–13669. https://doi.org/10.1021/acs.est.8b04481.
(34) Hobbs, S. R.; Parameswaran, P.; Astmann, B.; Devkota, J. P.; Landis, A. E. Anaerobic Codigestion of Food Waste and Polylactic Acid: Effect of Pretreatment on Methane Yield and Solid Reduction. *Adv. Mater. Sci. Eng.* **2019**, *2019*, 1–6. https://doi.org/10.1155/2019/4715904.
(35) Cooke, T.; Pomeroy, R. S. The Bioplastics Market: History, Commercialization Trends, and the New Eco-Consumer. In *Rethinking Polyester Polyurethanes*; Elsevier, 2023; pp 261–280. https://doi.org/10.1016/B978-0-323-99982-3.00003-1.
(36) Liu, Z.; Gosser, Y.; Baker, P. J.; Ravee, Y.; Lu, Z.; Alemu, G.; Li, H.; Butterfoss, G. L.; Kong, X.-P.; Gross, R.; Montclare, J. K. Structural and Functional Studies of Aspergillus Oryzae Cutinase: Enhanced Thermostability and Hydrolytic Activity of Synthetic Ester and Polyester Degradation. *J. Am. Chem. Soc.* **2009**, *131* (43), 15711–15716. https://doi.org/10.1021/ja9046697.

(37) Pastorino, L.; Pioli, F.; Zilli, M.; Converti, A.; Nicolini, C. Lipase-Catalyzed Degradation of Poly(ε-Caprolactone). *Enzyme Microb. Technol.* **2004**, *35* (4), 321–326. https://doi.org/10.1016/j.enzmictec.2004.05.005.
(38) Huang, Q.; Kimura, S.; Iwata, T. Development of Self-Degradable Aliphatic Polyesters by Embedding Lipases via Melt Extrusion. *Polym. Degrad. Stab.* **2021**, *190*, 109647. https://doi.org/10.1016/j.polymdegradstab.2021.109647.
(39) Gan, Z.; Liang, Q.; Zhang, J.; Jing, X. Enzymatic Degradation of Poly(ε-Caprolactone) Film in Phosphate Buffer Solution Containing Lipases. *Polym. Degrad. Stab.* **1997**, *56* (2), 209–213. https://doi.org/10.1016/S0141-3910(96)00208-X.
(40) Hermanová, S.; Omelková, J.; Voběrková, S.; Bálková, R.; Richtera, L.; Mravcová, L.; Jančář, J. The Effect of Processing of Polycaprolactone Films on Degradation Process Initiated by Aspergillus Oryzae Lipase. *Int. J. Polym. Anal. Charact.* **2012**, *17* (6), 465–475. https://doi.org/10.1080/1023666X.2012.696402.
(41) Hoshino, A.; Isono, Y. Degradation of Aliphatic Polyester Films by Commercially Available Lipases with Special Reference to Rapid and Complete Degradation of Poly(L-Lactide) Film by Lipase PL Derived from Alcaligenes Sp. *Biodegradation* **2002**, *13* (2), 141–147. https://doi.org/10.1023/A:1020450326301.
(42) Masaki, K.; Kamini, N. R.; Ikeda, H.; Iefuji, H. Cutinase-Like Enzyme from the Yeast Cryptococcus Sp. Strain S-2 Hydrolyzes Polylactic Acid and Other Biodegradable Plastics. *Appl. Environ. Microbiol.* **2005**, *71* (11), 7548–7550. https://doi.org/10.1128/AEM.71.11.7548-7550.2005.
(43) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
(44) Kwon, J.; DelRe, C.; Kang, P.; Hall, A.; Arnold, D.; Jayapurna, I.; Ma, L.; Michalek, M.; Ritchie, R. O.; Xu, T. Conductive Ink with Circular Life Cycle for Printed Electronics. *Adv. Mater.* **2022**, *34* (30), 2202177. https://doi.org/10.1002/adma.202202177.
(45) Huang, Q.; Hiyama, M.; Kabe, T.; Kimura, S.; Iwata, T. Enzymatic Self-Biodegradation of Poly(L -Lactic Acid) Films by Embedded Heat-Treated and Immobilized Proteinase K. *Biomacromolecules* **2020**, *21* (8), 3301–3307. https://doi.org/10.1021/acs.biomac.0c00759.
(46) Maraveas, C.; Kotzabasaki, M. I.; Bartzanas, T. Intelligent Technologies, Enzyme-Embedded and Microbial Degradation of Agricultural Plastics. *AgriEngineering* **2023**, *5* (1), 85–111. https://doi.org/10.3390/agriengineering5010006.
(47) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera De La Cruz, M.; Xu, T. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018**, *359* (6381), 1239–1243. https://doi.org/10.1126/science.aao0335.
(48) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of

Polymer–Protein Hybrids. *Adv. Mater.* **2022**, *34* (30), 2201809. https://doi.org/10.1002/adma.202201809.
(49) Jayapurna, I.; Ruan, Z.; Eres, M.; Jalagam, P.; Jenkins, S.; Xu, T. Sequence Design of Random Heteropolymers as Protein Mimics. *Biomacromolecules* **2023**, *24* (2), 652–660. https://doi.org/10.1021/acs.biomac.2c01036.
(50) Clark, D. S. Characteristics of Nearly Dry Enzymes in Organic Solvents: Implications for Biocatalysis in the Absence of Water. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2004**, *359* (1448), 1299–1307. https://doi.org/10.1098/rstb.2004.1506.
(51) Zaks, A.; Klibanov, A. M. The Effect of Water on Enzyme Action in Organic Media. *J. Biol. Chem.* **1988**, *263* (17), 8017–8021. https://doi.org/10.1016/S0021-9258(18)68435-2.
(52) Pleiss, J.; Fischer, M.; Schmid, R. D. Anatomy of Lipase Binding Sites: The Scissile Fatty Acid Binding Site. *Chem. Phys. Lipids* **1998**, *93* (1–2), 67–80. https://doi.org/10.1016/S0009-3084(98)00030-9.
(53) Verger, R. 'Interfacial Activation' of Lipases: Facts and Artifacts. *Trends Biotechnol.* **1997**, *15* (1), 32–38. https://doi.org/10.1016/S0167-7799(96)10064-0.
(54) Bornscheuer, U.; Reif, O.; Lausch, R.; Freitag, R.; Scheper, T.; Kolisis, F.; Menge, U. Lipase of Pseudomonas Cepacia for Biotechnological Purposes: Purification, Crystallization and Characterization. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1994**, *1201* (1), 55–60. https://doi.org/10.1016/0304-4165(94)90151-1.
(55) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera De La Cruz, M.; Xu, T. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018**, *359* (6381), 1239–1243. https://doi.org/10.1126/science.aao0335.
(56) Wurm, A.; Zhuravlev, E.; Eckstein, K.; Jehnichen, D.; Pospiech, D.; Androsch, R.; Wunderlich, B.; Schick, C. Crystallization and Homogeneous Nucleation Kinetics of Poly(ε-Caprolactone) (PCL) with Different Molar Masses. *Macromolecules* **2012**, *45* (9), 3816–3828. https://doi.org/10.1021/ma300363b.

Ch 3. Modulating protein stability and activity via intermolecular protectants

(1) Bahrami, A.; Delshadi, R.; Assadpour, E.; Jafari, S. M.; Williams, L. Antimicrobial-Loaded Nanocarriers for Food Packaging Applications. *Adv. Colloid Interface Sci.* **2020**, *278*, 102140. https://doi.org/10.1016/j.cis.2020.102140.
(2) Breitenbach, J. Melt Extrusion: From Process to Drug Delivery Technology. *Eur. J. Pharm. Biopharm.* **2002**, *54* (2), 107–117. https://doi.org/10.1016/S0939-6411(02)00061-9.
(3) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera de la Cruz, M.; Xu, T. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018**, *359* (6381), 1239–1243. https://doi.org/10.1126/science.aao0335.
(4) DelRe, C.; Huang, C.; Li, T.; Dennis, P.; Drockenmuller, E.; Xu, T. Reusable Enzymatic Fiber Mats for Neurotoxin Remediation in Water. *ACS Appl. Mater. Interfaces* **2018**, *10* (51), 44216–44220. https://doi.org/10.1021/acsami.8b18484.
(5) DelRe, C.; Jiang, Y.; Kang, P.; Kwon, J.; Hall, A.; Jayapurna, I.; Ruan, Z.; Ma, L.; Zolkin, K.; Li, T.; Scown, C. D.; Ritchie, R. O.; Russell, T. P.; Xu, T. Near-Complete Depolymerization of Polyesters with Nano-Dispersed Enzymes. *Nature* **2021**, *592* (7855), 558–563. https://doi.org/10.1038/s41586-021-03408-3.
(6) DelRe, C.; Chang, B.; Jayapurna, I.; Hall, A.; Wang, A.; Zolkin, K.; Xu, T. Synergistic Enzyme Mixtures to Realize Near-Complete Depolymerization in Biodegradable Polymer/Additive Blends. *Adv. Mater.* **2021**, *33* (49), 2105707. https://doi.org/10.1002/adma.202105707.
(7) Kwon, J.; DelRe, C.; Kang, P.; Hall, A.; Arnold, D.; Jayapurna, I.; Ma, L.; Michalek, M.; Ritchie, R. O.; Xu, T. Conductive Ink with Circular Life Cycle for Printed Electronics. *Adv. Mater.* **2022**, *34* (30), 2202177. https://doi.org/10.1002/adma.202202177.
(8) Kalita, N. K.; Hakkarainen, M. Triggering Degradation of Cellulose Acetate by Embedded Enzymes: Accelerated Enzymatic Degradation and Biodegradation under Simulated Composting Conditions. *Biomacromolecules* **2023**, *24* (7), 3290–3303. https://doi.org/10.1021/acs.biomac.3c00337.
(9) Warner, M. J.; Lassa, J. P.; Narcross, H.; Commisso, A.; Ghosh, K.; Romero, M.; Schwartz, J. M.; Engler, A. C.; Kohl, P. A.; Leguizamon, S. C.; Jones, B. H. Chemical Recycling of Polybutadiene Rubber with Tailored Depolymerization Enabled by Microencapsulated Metathesis Catalysts. *ACS Sustain. Chem. Eng.* **2023**, *11* (39), 14538–14548. https://doi.org/10.1021/acssuschemeng.3c03907.
(10) Maraveas, C.; Kotzabasaki, M. I.; Bartzanas, T. Intelligent Technologies, Enzyme-Embedded and Microbial Degradation of Agricultural Plastics. *AgriEngineering* **2023**, *5* (1), 85–111. https://doi.org/10.3390/agriengineering5010006.
(11) Siemann, U. Solvent Cast Technology – a Versatile Tool for Thin Film Production. In *Scattering Methods and the Properties of Polymer Materials*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 1–14. https://doi.org/10.1007/b107336.
(12) Rauwendaal, C. *Polymer Extrusion*, 5th edition.; Hanser Publications ; Hanser Publication: Munich : Cincinnati, 2014.

(13) EUBIO_Admin. *Market*. European Bioplastics e.V. https://www.european-bioplastics.org/market/ (accessed 2024-04-29).
(14) Velghe, I.; Buffel, B.; Vandeginste, V.; Thielemans, W.; Desplentere, F. Review on the Degradation of Poly(Lactic Acid) during Melt Processing. *Polymers* **2023**, *15* (9), 2047. https://doi.org/10.3390/polym15092047.
(15) Aldhafeeri, T.; Alotaibi, M.; Barry, C. F. Impact of Melt Processing Conditions on the Degradation of Polylactic Acid. *Polymers* **2022**, *14* (14), 2790. https://doi.org/10.3390/polym14142790.
(16) Yeh, C. K.; Lin, C.; Shen, H. C.; Cheruiyot, N. K.; Camarillo, M. E.; Wang, C. L. Optimizing Food Waste Composting Parameters and Evaluating Heat Generation. *Appl. Sci.* **2020**, *10* (7), 2284. https://doi.org/10.3390/app10072284.
(17) Vieille, C.; Zeikus, G. J. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65* (1), 1–43. https://doi.org/10.1128/MMBR.65.1.1-43.2001.
(18) Modarres, H. P.; Mofrad, M. R.; Sanati-Nezhad, A. Protein Thermostability Engineering. *RSC Adv.* **2016**, *6* (116), 115252–115270. https://doi.org/10.1039/C6RA16992A.
(19) Xu, C.; Battig, A.; Schartel, B.; Siegel, R.; Senker, J.; Von Der Forst, I.; Unverzagt, C.; Agarwal, S.; Möglich, A.; Greiner, A. Investigation of the Thermal Stability of Proteinase K for the Melt Processing of Poly(L -Lactide). *Biomacromolecules* **2022**, *23* (11), 4841–4850. https://doi.org/10.1021/acs.biomac.2c01008.
(20) Weng, Y.; Dunn, C. B.; Qiang, Z.; Ren, J. Immobilization of Protease K with ZIF-8 for Enhanced Stability in Polylactic Acid Melt Processing and Catalytic Degradation. *ACS Appl. Mater. Interfaces* **2023**, acsami.3c11979. https://doi.org/10.1021/acsami.3c11979.
(21) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer–Protein Hybrids. *Adv. Mater.* **2022**, *34* (30), 2201809. https://doi.org/10.1002/adma.202201809.
(22) Zaks, A.; Klibanov, A. M. The Effect of Water on Enzyme Action in Organic Media. *J. Biol. Chem.* **1988**, *263* (17), 8017–8021. https://doi.org/10.1016/S0021-9258(18)68435-2.
(23) Clark, D. S. Characteristics of Nearly Dry Enzymes in Organic Solvents: Implications for Biocatalysis in the Absence of Water. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2004**, *359* (1448), 1299–1307. https://doi.org/10.1098/rstb.2004.1506.
(24) Mikos, A. G.; Thorsen, A. J.; Czerwonka, L. A.; Bao, Y.; Langer, R.; Winslow, D. N.; Vacanti, J. P. Preparation and Characterization of Poly(l-Lactic Acid) Foams. *Polymer* **1994**, *35* (5), 1068–1077. https://doi.org/10.1016/0032-3861(94)90953-9.
(25) Shahid, S.; Hassan, Md. I.; Islam, A.; Ahmad, F. Size-Dependent Studies of Macromolecular Crowding on the Thermodynamic Stability, Structure and Functional Activity of Proteins: In Vitro and in Silico Approaches. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2017**, *1861* (2), 178–197. https://doi.org/10.1016/j.bbagen.2016.11.014.
(26) Barlas, N. T.; Bahamonde, H. A.; Pimentel, C.; Domínguez-Huidobro, P.; Pina, C. M.; Fernández, V. Evaluating Leaf Wettability and Salt Hygroscopicity as Drivers for Foliar Absorption. *Plants* **2023**, *12* (12), 2357. https://doi.org/10.3390/plants12122357.

(27) Ashraf, H.; Guo, Y.; Wang, N.; Pang, S.; Zhang, Y.-H. Hygroscopicity of Hofmeister Salts and Glycine Aerosols–Salt Specific Interactions. *J. Phys. Chem. A* **2021**, *125* (7), 1589–1597. https://doi.org/10.1021/acs.jpca.0c10710.
(28) Choi, M. Y.; Chan, C. K. Investigation of Efflorescence of Inorganic Aerosols Using Fluorescence Spectroscopy. *J. Phys. Chem. A* **2005**, *109* (6), 1042–1048. https://doi.org/10.1021/jp048103h.
(29) Williams, D. F. Enzymic Hydrolysis of Polylactic Acid. *Eng. Med.* **1981**, *10* (1), 5–7. https://doi.org/10.1243/EMED_JOUR_1981_010_004_02.
(30) Bornscheuer, U.; Reif, O.; Lausch, R.; Freitag, R.; Scheper, T.; Kolisis, F.; Menge, U. Lipase of Pseudomonas Cepacia for Biotechnological Purposes: Purification, Crystallization and Characterization. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1994**, *1201* (1), 55–60. https://doi.org/10.1016/0304-4165(94)90151-1.

Ch 4. Biocheminformatic Design of Biomimetic Heteropolymers

(1) Popot, J.-L.; Berry, E. A.; Charvolin, D.; Creuzenet, C.; Ebel, C.; Engelman, D. M.; Flötenmeyer, M.; Giusti, F.; Gohon, Y.; Hervé, P.; Hong, Q.; Lakey, J. H.; Leonard, K.; Shuman, H. A.; Timmins, P.; Warschawski, D. E.; Zito, F.; Zoonens, M.; Pucci, B.; Tribet, C. Amphipols: Polymeric Surfactants for Membrane Biology Research. *Cell. Mol. Life Sci. CMLS* **2003**, *60* (8), 1559–1574. https://doi.org/10.1007/s00018-003-3169-6.
(2) DelRe, C.; Chang, B.; Jayapurna, I.; Hall, A.; Wang, A.; Zolkin, K.; Xu, T. Synergistic Enzyme Mixtures to Realize Near-Complete Depolymerization in Biodegradable Polymer/Additive Blends. *Adv. Mater.* **2021**, *33* (49), 2105707. https://doi.org/10.1002/adma.202105707.
(3) DelRe, C.; Huang, C.; Li, T.; Dennis, P.; Drockenmuller, E.; Xu, T. Reusable Enzymatic Fiber Mats for Neurotoxin Remediation in Water. *ACS Appl. Mater. Interfaces* **2018**, *10* (51), 44216–44220. https://doi.org/10.1021/acsami.8b18484.
(4) DelRe, C.; Jiang, Y.; Kang, P.; Kwon, J.; Hall, A.; Jayapurna, I.; Ruan, Z.; Ma, L.; Zolkin, K.; Li, T.; Scown, C. D.; Ritchie, R. O.; Russell, T. P.; Xu, T. Near-Complete Depolymerization of Polyesters with Nano-Dispersed Enzymes. *Nature* **2021**, *592* (7855), 558–563. https://doi.org/10.1038/s41586-021-03408-3.
(5) Kwon, J.; DelRe, C.; Kang, P.; Hall, A.; Arnold, D.; Jayapurna, I.; Ma, L.; Michalek, M.; Ritchie, R. O.; Xu, T. Conductive Ink with Circular Life Cycle for Printed Electronics. *Adv. Mater.* **2022**, *34* (30), 2202177. https://doi.org/10.1002/adma.202202177.
(6) Smith, A. A. A.; Maikawa, C. L.; Lopez Hernandez, H.; Appel, E. A. Controlling Properties of Thermogels by Tuning Critical Solution Behaviour of Ternary Copolymers. *Polym. Chem.* **2021**, *12* (13), 1918–1923. https://doi.org/10.1039/D0PY01696A.
(7) Cardellini, A.; Jiménez-Ángeles, F.; Asinari, P.; Olvera de la Cruz, M. A Modeling-Based Design to Engineering Protein Hydrogels with Random Copolymers. *ACS Nano* **2021**, *15* (10), 16139–16148. https://doi.org/10.1021/acsnano.1c04955.
(8) Chang, L.-W.; Lytle, T. K.; Radhakrishna, M.; Madinya, J. J.; Vélez, J.; Sing, C. E.; Perry, S. L. Sequence and Entropy-Based Control of Complex Coacervates. *Nat. Commun.* **2017**, *8* (1), 1273. https://doi.org/10.1038/s41467-017-01249-1.
(9) Lytle, T. K.; Chang, L.-W.; Markiewicz, N.; Perry, S. L.; Sing, C. E. Designing Electrostatic Interactions via Polyelectrolyte Monomer Sequence. *ACS Cent. Sci.* **2019**, *5* (4), 709–718. https://doi.org/10.1021/acscentsci.9b00087.
(10) Yu, B.; Rumyantsev, A. M.; Jackson, N. E.; Liang, H.; Ting, J. M.; Meng, S.; Tirrell, M. V.; de Pablo, J. J. Complex Coacervation of Statistical Polyelectrolytes: Role of Monomer Sequences and Formation of Inhomogeneous Coacervates. *Mol. Syst. Des. Eng.* **2021**, *6* (10), 790–804. https://doi.org/10.1039/D1ME00076D.
(11) deRonde, B. M.; Tew, G. N. Development of Protein Mimics for Intracellular Delivery: Protein Mimics for Intracellular Delivery. *Biopolymers* **2015**, *104* (4), 265–280. https://doi.org/10.1002/bip.22658.
(12) Kumar, R.; Le, N.; Tan, Z.; Brown, M. E.; Jiang, S.; Reineke, T. M. Efficient Polymer-Mediated Delivery of Gene-Editing Ribonucleoprotein Payloads through Combinatorial

Design, Parallelized Experimentation, and Machine Learning. *ACS Nano* **2020**, *14* (12), 17626–17639. https://doi.org/10.1021/acsnano.0c08549.
(13) 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* **2022**, *2* (2), 428–442. https://doi.org/10.1021/jacsau.1c00467.
(14) Kuroda, K.; DeGrado, W. F. Amphiphilic Polymethacrylate Derivatives as Antimicrobial Agents. *J. Am. Chem. Soc.* **2005**, *127* (12), 4128–4129. https://doi.org/10.1021/ja044205+.
(15) Mowery, B. P.; Lee, S. E.; Kissounko, D. A.; Epand, R. F.; Epand, R. M.; Weisblum, B.; Stahl, S. S.; Gellman, S. H. Mimicry of Antimicrobial Host-Defense Peptides by Random Copolymers. *J. Am. Chem. Soc.* **2007**, *129* (50), 15474–15476. https://doi.org/10.1021/ja077288d.
(16) Pasparakis, G.; Krasnogor, N.; Cronin, L.; Davis, B. G.; Alexander, C. Controlled Polymer Synthesis—from Biomimicry towards Synthetic Biology. *Chem Soc Rev* **2010**, *39* (1), 286–300. https://doi.org/10.1039/B809333B.
(17) Biggs, C. I.; Bailey, T. L.; Ben Graham; Stubbs, C.; Fayter, A.; Gibson, M. I. Polymer Mimics of Biomacromolecular Antifreezes. *Nat. Commun.* **2017**, *8* (1), 1546. https://doi.org/10.1038/s41467-017-01421-7.
(18) Shakhnovich, E. I.; Gutin, A. M. Frozen States of a Disordered Globular Heteropolymer. *J. Phys. Math. Gen.* **1989**, *22* (10), 1647–1659. https://doi.org/10.1088/0305-4470/22/10/019.
(19) Fredrickson, G. H.; Milner, S. T.; Leibler, L. Multicritical Phenomena and Microphase Ordering in Random Block Copolymers Melts. *Macromolecules* **1992**, *25* (23), 6341–6354. https://doi.org/10.1021/ma00049a034.
(20) Golumbfskie, A. J.; Pande, V. S.; Chakraborty, A. K. Simulation of Biomimetic Recognition between Polymers and Surfaces. *Proc. Natl. Acad. Sci.* **1999**, *96* (21), 11707–11712. https://doi.org/10.1073/pnas.96.21.11707.
(21) Geissler, P. L.; Shakhnovich, E. I. Mechanical Response of Random Heteropolymers. *Macromolecules* **2002**, *35* (11), 4429–4436. https://doi.org/10.1021/ma012008e.
(22) Mao, S.; MacPherson, Q. J.; He, S. S.; Coletta, E.; Spakowitz, A. J. Impact of Conformational and Chemical Correlations on Microphase Segregation in Random Copolymers. *Macromolecules* **2016**, *49* (11), 4358–4368. https://doi.org/10.1021/acs.macromol.5b02639.
(23) Cardelli, C.; Bianco, V.; Rovigatti, L.; Nerattini, F.; Tubiana, L.; Dellago, C.; Coluzza, I. The Role of Directional Interactions in the Designability of Generalized Heteropolymers. *Sci. Rep.* **2017**, *7* (1), 4986. https://doi.org/10.1038/s41598-017-04720-7.
(24) Nguyen, T. D.; Qiao, B.; Olvera de la Cruz, M. Efficient Encapsulation of Proteins with Random Copolymers. *Proc. Natl. Acad. Sci.* **2018**, *115* (26), 6578–6583. https://doi.org/10.1073/pnas.1806207115.
(25) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera De La Cruz, M.; Xu, T. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018**, *359* (6381), 1239–1243. https://doi.org/10.1126/science.aao0335.

(26) Jiang, T.; Hall, A.; Eres, M.; Hemmatian, Z.; Qiao, B.; Zhou, Y.; Ruan, Z.; Couse, A. D.; Heller, W. T.; Huang, H.; De La Cruz, M. O.; Rolandi, M.; Xu, T. Single-Chain Heteropolymers Transport Protons Selectively and Rapidly. *Nature* **2020**, *577* (7789), 216–220. https://doi.org/10.1038/s41586-019-1881-0.
(27) Hilburg, S. L.; Ruan, Z.; Xu, T.; Alexander-Katz, A. Behavior of Protein-Inspired Synthetic Random Heteropolymers. *Macromolecules* **2020**, *53* (21), 9187–9199. https://doi.org/10.1021/acs.macromol.0c01886.
(28) Wang, J.; Waltmann, C.; Umana-Kossio, H.; Olvera de la Cruz, M.; Torkelson, J. M. Heterogeneous Charged Complexes of Random Copolymers for the Segregation of Organic Molecules. *ACS Cent. Sci.* **2021**, *7* (5), 882–891. https://doi.org/10.1021/acscentsci.1c00119.
(29) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer–Protein Hybrids. *Adv. Mater.* **2022**, *34* (30), 2201809. https://doi.org/10.1002/adma.202201809.
(30) Han, Z.; Hilburg, S. L.; Alexander-Katz, A. Forced Unfolding of Protein-Inspired Single-Chain Random Heteropolymers. *Macromolecules* **2022**, *55* (4), 1295–1309. https://doi.org/10.1021/acs.macromol.1c02411.
(31) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341* (6146), 1238149. https://doi.org/10.1126/science.1238149.
(32) Mutlu, H.; Lutz, J.-F. Reading Polymers: Sequencing of Natural and Synthetic Macromolecules. *Angew. Chem. Int. Ed.* **2014**, *53* (48), 13010–13019. https://doi.org/10.1002/anie.201406766.
(33) Lutz, J.-F. Defining the Field of Sequence-Controlled Polymers. *Macromol. Rapid Commun.* **2017**, *38* (24), 1700582. https://doi.org/10.1002/marc.201700582.
(34) Szymański, J. K.; Abul-Haija, Y. M.; Cronin, L. Exploring Strategies To Bias Sequence in Natural and Synthetic Oligomers and Polymers. *Acc. Chem. Res.* **2018**, *51* (3), 649–658. https://doi.org/10.1021/acs.accounts.7b00495.
(35) Perrier, S. *50th Anniversary Perspective* : RAFT Polymerization—A User Guide. *Macromolecules* **2017**, *50* (19), 7433–7447. https://doi.org/10.1021/acs.macromol.7b00767.
(36) Chiefari, J.; Chong, Y. K. (Bill); Ercole, F.; Krstina, J.; Jeffery, J.; Le, T. P. T.; Mayadunne, R. T. A.; Meijs, G. F.; Moad, C. L.; Moad, G.; Rizzardo, E.; Thang, S. H. Living Free-Radical Polymerization by Reversible Addition−Fragmentation Chain Transfer: The RAFT Process. *Macromolecules* **1998**, *31* (16), 5559–5562. https://doi.org/10.1021/ma9804951.
(37) Corrigan, N.; Jung, K.; Moad, G.; Hawker, C. J.; Matyjaszewski, K.; Boyer, C. Reversible-Deactivation Radical Polymerization (Controlled/Living Radical Polymerization): From Discovery to Materials Design and Applications. *Prog. Polym. Sci.* **2020**, *111*, 101311. https://doi.org/10.1016/j.progpolymsci.2020.101311.
(38) Szwarc, M. 'Living' Polymers. *Nature* **1956**, *178* (4543), 1168–1169. https://doi.org/10.1038/1781168a0.
(39) Matyjaszewski, K.; Gaynor, S.; Greszta, D.; Mardare, D.; Shigemoto, T. 'Living' and Controlled Radical Polymerization. *J. Phys. Org. Chem.* **1995**, *8* (4), 306–315. https://doi.org/10.1002/poc.610080414.

(40) Hawker, C. J.; Bosman, A. W.; Harth, E. New Polymer Synthesis by Nitroxide Mediated Living Radical Polymerizations. *Chem. Rev.* **2001**, *101* (12), 3661–3688. https://doi.org/10.1021/cr990119u.
(41) Matyjaszewski, K. Atom Transfer Radical Polymerization (ATRP): Current Status and Future Perspectives. *Macromolecules* **2012**, *45* (10), 4015–4039. https://doi.org/10.1021/ma3001719.
(42) Gauthier, M. A.; Gibson, M. I.; Klok, H.-A. Synthesis of Functional Polymers by Post-Polymerization Modification. *Angew. Chem. Int. Ed.* **2009**, *48* (1), 48–58. https://doi.org/10.1002/anie.200801951.
(43) Pelegri-O'Day, E. M.; Maynard, H. D. Controlled Radical Polymerization as an Enabling Approach for the Next Generation of Protein–Polymer Conjugates. *Acc. Chem. Res.* **2016**, *49* (9), 1777–1785. https://doi.org/10.1021/acs.accounts.6b00258.
(44) Mayo, F. R.; Lewis, F. M. Copolymerization. I. A Basis for Comparing the Behavior of Monomers in Copolymerization; The Copolymerization of Styrene and Methyl Methacrylate. *J. Am. Chem. Soc.* **1944**, *66* (9), 1594–1601. https://doi.org/10.1021/ja01237a052.
(45) Bruns, W.; Motoc, I.; O'Driscoll, K. F. *Monte Carlo Applications in Polymer Science*; Berthier, G., Dewar, M. J. S., Fischer, H., Fukui, K., Hall, G. G., Hartmann, H., Jaffé, H. H., Jortner, J., Kutzelnigg, W., Ruedenberg, K., Scrocco, E., Eds.; Lecture Notes in Chemistry; Springer Berlin Heidelberg: Berlin, Heidelberg, 1981; Vol. 27. https://doi.org/10.1007/978-3-642-93195-6.
(46) Klumperman, B.; O'Driscoll, K. F. Interpreting the Copolymerization of Styrene with Maleic Anhydride and with Methyl Methacrylate in Terms of the Bootstrap Model. *Polymer* **1993**, *34* (5), 1032–1037. https://doi.org/10.1016/0032-3861(93)90226-Z.
(47) Wulkow, M. Computer Aided Modeling of Polymer Reaction Engineering—The Status of Predici, I-Simulation. *Macromol. React. Eng.* **2008**, *2* (6), 461–494. https://doi.org/10.1002/mren.200800024.
(48) Zapata-González, I.; Saldívar-Guerra, E.; Ortiz-Cisneros, J. Full Molecular Weight Distribution in RAFT Polymerization. New Mechanistic Insight by Direct Integration of the Equations: Full Molecular Weight Distribution in RAFT Polymerization. .... *Macromol. Theory Simul.* **2011**, *20* (6), 370–388. https://doi.org/10.1002/mats.201100023.
(49) Drache, M.; Drache, G. Simulating Controlled Radical Polymerizations with mcPolymer—A Monte Carlo Approach. *Polymers* **2012**, *4* (3), 1416–1442. https://doi.org/10.3390/polym4031416.
(50) Ting, J. M.; Navale, T. S.; Bates, F. S.; Reineke, T. M. Precise Compositional Control and Systematic Preparation of Multimonomeric Statistical Copolymers. *ACS Macro Lett.* **2013**, *2* (9), 770–774. https://doi.org/10.1021/mz4003112.
(51) Zapata-González, I.; Hutchinson, R. A.; Matyjaszewski, K.; Saldívar-Guerra, E.; Ortiz-Cisneros, J. Copolymer Composition Deviations from Mayo-Lewis Conventional Free Radical Behavior in Nitroxide Mediated Copolymerization: Copolymer Composition Deviations from Mayo-Lewis Conventional Free Radical Behavior .... *Macromol. Theory Simul.* **2014**, *23* (4), 245–265. https://doi.org/10.1002/mats.201300137.
(52) Fortunatti, C.; Sarmoria, C.; Brandolin, A.; Asteasuain, M. Modeling of RAFT Polymerization Using Probability Generating Functions. Detailed Prediction of Full Molecular Weight Distributions and Sensitivity Analysis: Modeling of RAFT Polymerization

Using Probability Generating Functions …. *Macromol. React. Eng.* **2014**, *8* (12), 781–795. https://doi.org/10.1002/mren.201400020.
(53) Mastan, E.; Li, X.; Zhu, S. Modeling and Theoretical Development in Controlled Radical Polymerization. *Prog. Polym. Sci.* **2015**, *45*, 71–101. https://doi.org/10.1016/j.progpolymsci.2014.12.003.
(54) Pintos, E.; Sarmoria, C.; Brandolin, A.; Asteasuain, M. Modeling of RAFT Polymerization Processes Using an Efficient Monte Carlo Algorithm in Julia. *Ind. Eng. Chem. Res.* **2016**, *55* (31), 8534–8547. https://doi.org/10.1021/acs.iecr.6b01639.
(55) Smith, A. A. A.; Hall, A.; Wu, V.; Xu, T. Practical Prediction of Heteropolymer Composition and Drift. *ACS Macro Lett.* **2019**, *8* (1), 36–40. https://doi.org/10.1021/acsmacrolett.8b00813.
(56) Steube, M.; Johann, T.; Barent, R. D.; Müller, A. H. E.; Frey, H. Rational Design of Tapered Multiblock Copolymers for Thermoplastic Elastomers. *Prog. Polym. Sci.* **2022**, *124*, 101488. https://doi.org/10.1016/j.progpolymsci.2021.101488.
(57) White, S. H.; Jacobs, R. E. Statistical Distribution of Hydrophobic Residues along the Length of Protein Chains. Implications for Protein Folding and Evolution. *Biophys. J.* **1990**, *57* (4), 911–921. https://doi.org/10.1016/S0006-3495(90)82611-4.
(58) Rao, N.; Lei, X.; Guo, J.; Huang, H.; Ren, Z. An Efficient Sliding Window Strategy for Accurate Location of Eukaryotic Protein Coding Regions. *Comput. Biol. Med.* **2009**, *39* (4), 392–395. https://doi.org/10.1016/j.compbiomed.2009.01.010.
(59) Fares, M. A.; Elena, S. F.; Ortiz, J.; Moya, A.; Barrio, E. A Sliding Window-Based Method to Detect Selective Constraints in Protein-Coding Genes and Its Application to RNA Viruses. *J. Mol. Evol.* **2002**, *55* (5), 509–521. https://doi.org/10.1007/s00239-002-2346-9.
(60) Chen, K.; Kurgan, L.; Ruan, J. Optimization of the Sliding Window Size for Protein Structure Prediction. In *2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*; 2006; pp 1–7. https://doi.org/10.1109/CIBCB.2006.330959.
(61) Brendel, V.; Bucher, P.; Nourbakhsh, I. R.; Blaisdell, B. E.; Karlin, S. Methods and Algorithms for Statistical Analysis of Protein Sequences. *Proc. Natl. Acad. Sci.* **1992**, *89* (6), 2002–2006. https://doi.org/10.1073/pnas.89.6.2002.
(62) Karlin, S.; Brendel, V. Chance and Statistical Significance in Protein and DNA Sequence Analysis. *Science* **1992**, *257* (5066), 39–49. https://doi.org/10.1126/science.1621093.

Ch 5. Improving heteropolymer sequence fidelity

(1) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera De La Cruz, M.; Xu, T. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018**, *359* (6381), 1239–1243. https://doi.org/10.1126/science.aao0335.
(2) Jiang, T.; Hall, A.; Eres, M.; Hemmatian, Z.; Qiao, B.; Zhou, Y.; Ruan, Z.; Couse, A. D.; Heller, W. T.; Huang, H.; De La Cruz, M. O.; Rolandi, M.; Xu, T. Single-Chain Heteropolymers Transport Protons Selectively and Rapidly. *Nature* **2020**, *577* (7789), 216–220. https://doi.org/10.1038/s41586-019-1881-0.
(3) DelRe, C.; Jiang, Y.; Kang, P.; Kwon, J.; Hall, A.; Jayapurna, I.; Ruan, Z.; Ma, L.; Zolkin, K.; Li, T.; Scown, C. D.; Ritchie, R. O.; Russell, T. P.; Xu, T. Near-Complete Depolymerization of Polyesters with Nano-Dispersed Enzymes. *Nature* **2021**, *592* (7855), 558–563. https://doi.org/10.1038/s41586-021-03408-3.
(4) DelRe, C.; Chang, B.; Jayapurna, I.; Hall, A.; Wang, A.; Zolkin, K.; Xu, T. Synergistic Enzyme Mixtures to Realize Near-Complete Depolymerization in Biodegradable Polymer/Additive Blends. *Adv. Mater.* **2021**, *33* (49), 2105707. https://doi.org/10.1002/adma.202105707.
(5) Ruan, Z.; Li, S.; Grigoropoulos, A.; Amiri, H.; Hilburg, S. L.; Chen, H.; Jayapurna, I.; Jiang, T.; Gu, Z.; Alexander-Katz, A.; Bustamante, C.; Huang, H.; Xu, T. Population-Based Heteropolymer Design to Mimic Protein Mixtures. *Nature* **2023**, *615* (7951), 251–258. https://doi.org/10.1038/s41586-022-05675-0.
(6) Upadhya, R.; Di Mare, E.; Tamasi, M. J.; Kosuri, S.; Murthy, N. S.; Gormley, A. J. Examining POLYMER-PROTEIN Biophysical Interactions with Small-angle X-ray Scattering and Quartz Crystal Microbalance with Dissipation. *J. Biomed. Mater. Res. A* **2023**, *111* (4), 440–450. https://doi.org/10.1002/jbm.a.37479.
(7) Gauthier, J.; Vincent, A. T.; Charette, S. J.; Derome, N. A Brief History of Bioinformatics. *Brief. Bioinform.* **2019**, *20* (6), 1981–1996. https://doi.org/10.1093/bib/bby063.
(8) Chiefari, J.; Chong, Y. K. (Bill); Ercole, F.; Krstina, J.; Jeffery, J.; Le, T. P. T.; Mayadunne, R. T. A.; Meijs, G. F.; Moad, C. L.; Moad, G.; Rizzardo, E.; Thang, S. H. Living Free-Radical Polymerization by Reversible Addition−Fragmentation Chain Transfer: The RAFT Process. *Macromolecules* **1998**, *31* (16), 5559–5562. https://doi.org/10.1021/ma9804951.
(9) Szwarc, M. 'Living' Polymers. *Nature* **1956**, *178* (4543), 1168–1169. https://doi.org/10.1038/1781168a0.
(10) Matyjaszewski, K.; Gaynor, S.; Greszta, D.; Mardare, D.; Shigemoto, T. 'Living' and Controlled Radical Polymerization. *J. Phys. Org. Chem.* **1995**, *8* (4), 306–315. https://doi.org/10.1002/poc.610080414.
(11) Corrigan, N.; Jung, K.; Moad, G.; Hawker, C. J.; Matyjaszewski, K.; Boyer, C. Reversible-Deactivation Radical Polymerization (Controlled/Living Radical Polymerization): From Discovery to Materials Design and Applications. *Prog. Polym. Sci.* **2020**, *111*, 101311. https://doi.org/10.1016/j.progpolymsci.2020.101311.

(12) Mohammadi, Y.; Najafi, M.; Haddadi-Asl, V. Comprehensive Study of Free Radical Copolymerization Using a Monte Carlo Simulation Method, 1. *Macromol. Theory Simul.* **2005**, *14* (5), 325–336. https://doi.org/10.1002/mats.200400082.
(13) Regatte, V. R.; Gao, H.; Konstantinov, I. A.; Arturo, S. G.; Broadbelt, L. J. Design of Copolymers Based on Sequence Distribution for a Targeted Molecular Weight and Conversion: Design of Copolymers Based on Sequence …. *Macromol. Theory Simul.* **2014**, *23* (9), 564–574. https://doi.org/10.1002/mats.201400037.
(14) Gao, H.; Broadbelt, L. J.; Konstantinov, I. A.; Arturo, S. G. Acceleration of Kinetic Monte Carlo Simulations of Free Radical Copolymerization: A Hybrid Approach with Scaling. *AIChE J.* **2017**, *63* (9), 4013–4021. https://doi.org/10.1002/aic.15751.
(15) Zapata-González, I.; Saldívar-Guerra, E.; Ortiz-Cisneros, J. Full Molecular Weight Distribution in RAFT Polymerization. New Mechanistic Insight by Direct Integration of the Equations. *Macromol. Theory Simul.* **2011**, *20* (6), 370–388. https://doi.org/10.1002/mats.201100023.
(16) Pintos, E.; Sarmoria, C.; Brandolin, A.; Asteasuain, M. Modeling of RAFT Polymerization Processes Using an Efficient Monte Carlo Algorithm in Julia. *Ind. Eng. Chem. Res.* **2016**, *55* (31), 8534–8547. https://doi.org/10.1021/acs.iecr.6b01639.
(17) Wulkow, M. Computer Aided Modeling of Polymer Reaction Engineering—The Status of Predici, I-Simulation. *Macromol. React. Eng.* **2008**, *2* (6), 461–494. https://doi.org/10.1002/mren.200800024.
(18) Drache, M.; Drache, G. Simulating Controlled Radical Polymerizations with mcPolymer—A Monte Carlo Approach. *Polymers* **2012**, *4* (3), 1416–1442. https://doi.org/10.3390/polym4031416.
(19) Tenorio-López, J. A.; Benvenuta-Tapia, J. J.; García-Navarro, N.; Vivaldo-Lima, E.; Champagne, P.; Saldívar-Guerra, E. Mathematical Description of the RAFT Copolymerization of Styrene and Glycidyl Methacrylate Using the Terminal Model. *Polymers* **2022**, *14* (7), 1448. https://doi.org/10.3390/polym14071448.
(20) Jiang, J.; Wang, W.; Li, B.; Zhu, S. Modeling and Experimentation of RAFT Solution Copolymerization of Styrene and Butyl Acrylate, Effect of Chain Transfer Reactions on Polymer Molecular Weight Distribution. *Macromol. React. Eng.* **2017**, *11* (6), 1700029. https://doi.org/10.1002/mren.201700029.
(21) Ting, J. M.; Navale, T. S.; Bates, F. S.; Reineke, T. M. Precise Compositional Control and Systematic Preparation of Multimonomeric Statistical Copolymers. *ACS Macro Lett.* **2013**, *2* (9), 770–774. https://doi.org/10.1021/mz4003112.
(22) Wang, L.; Broadbelt, L. J. Explicit Sequence of Styrene/Methyl Methacrylate Gradient Copolymers Synthesized by Forced Gradient Copolymerization with Nitroxide-Mediated Controlled Radical Polymerization. *Macromolecules* **2009**, *42* (20), 7961–7968. https://doi.org/10.1021/ma901298h.
(23) Mayo, F. R.; Lewis, F. M. Copolymerization. I. A Basis for Comparing the Behavior of Monomers in Copolymerization; The Copolymerization of Styrene and Methyl Methacrylate. *J. Am. Chem. Soc.* **1944**, *66* (9), 1594–1601. https://doi.org/10.1021/ja01237a052.
(24) Smith, A. A. A.; Hall, A.; Wu, V.; Xu, T. Practical Prediction of Heteropolymer Composition and Drift. *ACS Macro Lett.* **2019**, *8* (1), 36–40. https://doi.org/10.1021/acsmacrolett.8b00813.

(25) Jayapurna, I.; Ruan, Z.; Eres, M.; Jalagam, P.; Jenkins, S.; Xu, T. Sequence Design of Random Heteropolymers as Protein Mimics. *Biomacromolecules* **2023**, *24* (2), 652–660. https://doi.org/10.1021/acs.biomac.2c01036.
(26) Hilburg, S. L.; Ruan, Z.; Xu, T.; Alexander-Katz, A. Behavior of Protein-Inspired Synthetic Random Heteropolymers. *Macromolecules* **2020**, *53* (21), 9187–9199. https://doi.org/10.1021/acs.macromol.0c01886.
(27) Jin, T.; Hilburg, S. L.; Alexander-Katz, A. Glass Transition of Random Heteropolymers: A Molecular Dynamics Simulation Study in Melt, in Water, and in Vacuum. *Polymer* **2023**, *265*, 125503. https://doi.org/10.1016/j.polymer.2022.125503.
(28) Jin, T.; Coley, C. W.; Alexander-Katz, A. Adsorption of Biomimetic Amphiphilic Heteropolymers onto Graphene and Its Derivatives. *Macromolecules* **2023**, *56* (5), 1798–1809. https://doi.org/10.1021/acs.macromol.2c02413.
(29) Zhou, Y.; Gong, B.; Jiang, T.; Xu, T.; Huang, H. Stochastic Variational Methods in Generalized Hidden Semi-Markov Models to Characterize Functionality in Random Heteropolymers. **2022**. https://doi.org/10.48550/ARXIV.2207.01813.
(30) Lynd, N. A.; Ferrier, R. C.; Beckingham, B. S. Recommendation for Accurate Experimental Determination of Reactivity Ratios in Chain Copolymerization. *Macromolecules* **2019**, *52* (6), 2277–2285. https://doi.org/10.1021/acs.macromol.8b01752.
(31) Wittmer, P. Kinetics of Copolymerization. *Makromol. Chem.* **1979**, *3* (S19791), 129–156. https://doi.org/10.1002/macp.1979.020031979107.
(32) Kazemi, N.; Duever, T. A.; Penlidis, A. Demystifying the Estimation of Reactivity Ratios for Terpolymerization Systems. *AIChE J.* **2014**, *60* (5), 1752–1766. https://doi.org/10.1002/aic.14439.
(33) Beauchemin, R.-C.; Dubé, M. A. Bulk Terpolymer Composition Prediction from Copolymer Reactivity Ratios. *Polym. React. Eng.* **1999**, *7* (4), 485–499. https://doi.org/10.1080/10543414.1999.10744527.
(34) Schoonbrood, H. A. S.; Van Eijnatten, R. C. P. M.; Van Den Reijen, B.; Van Herk, A. M.; German, A. L. Emulsion Co- and Terpolymerization of Styrene, Methyl Methacrylate, and Methyl Acrylate. I. Experimental Determination and Model Prediction of Composition Drift and Microstructure in Batch Reactions. *J. Polym. Sci. Part Polym. Chem.* **1996**, *34* (6), 935–947. https://doi.org/10.1002/(SICI)1099-0518(19960430)34:6<935::AID-POLA2>3.0.CO;2-H.
(35) Ren, S.; Zhang, L.; Dubé, M. A. Free-radical Terpolymerization of *n* -butyl Acrylate/Butyl Methacrylate/ D -limonene. *J. Appl. Polym. Sci.* **2015**, *132* (47), app.42821. https://doi.org/10.1002/app.42821.
(36) Scott, A. J.; Penlidis, A. Binary vs. Ternary Reactivity Ratios: Appropriate Estimation Procedures with Terpolymerization Data. *Eur. Polym. J.* **2018**, *105*, 442–450. https://doi.org/10.1016/j.eurpolymj.2018.06.021.
(37) Kressler, J.; Bieger, W.; Horvath, B.; Schmidt-naake, G. Experimental Investigation of Ternary Azeotropy in the Copolymerization of Acrylonitrile, Styrene, and Maleic Anhydride. *J. Macromol. Sci. Part - Chem.* **1987**, *24* (6), 681–687. https://doi.org/10.1080/00222338708058522.

(38) Jaacks, V. A Novel Method of Determination of Reactivity Ratios in Binary and Ternary Copolymerizations. *Makromol. Chem.* **1972**, *161* (1), 161–172. https://doi.org/10.1002/macp.1972.021610110.
(39) Yin, R.; Zhao, Y.; Gorczyński, A.; Szczepaniak, G.; Sun, M.; Fu, L.; Kim, K.; Wu, H.; Bockstaller, M. R.; Matyjaszewski, K. Alternating Methyl Methacrylate/ *n* -Butyl Acrylate Copolymer Prepared by Atom Transfer Radical Polymerization. *ACS Macro Lett.* **2022**, *11* (10), 1217–1223. https://doi.org/10.1021/acsmacrolett.2c00517.
(40) Ishizaka, S.; Nakagawa, S.; Yoshie, N. Computer-Aided Design of Copolymers with Controlled Comonomer Distributions. *Macromolecules* **2023**, *56* (18), 7312–7319. https://doi.org/10.1021/acs.macromol.3c00409.
(41) Lutz, J.-F. Writing on Polymer Chains. *Acc. Chem. Res.* **2013**, *46* (11), 2696–2705. https://doi.org/10.1021/ar400097a.
(42) Pfeifer, S.; Lutz, J.-F. A Facile Procedure for Controlling Monomer Sequence Distribution in Radical Chain Polymerizations. *J. Am. Chem. Soc.* **2007**, *129* (31), 9542–9543. https://doi.org/10.1021/ja0717616.
(43) Hoshino, Y.; Haberaecker, W. W.; Kodama, T.; Zeng, Z.; Okahata, Y.; Shea, K. J. Affinity Purification of Multifunctional Polymer Nanoparticles. *J. Am. Chem. Soc.* **2010**, *132* (39), 13648–13650. https://doi.org/10.1021/ja1058982.
(44) Zeng, Z.; Patel, J.; Lee, S.-H.; McCallum, M.; Tyagi, A.; Yan, M.; Shea, K. J. Synthetic Polymer Nanoparticle–Polysaccharide Interactions: A Systematic Study. *J. Am. Chem. Soc.* **2012**, *134* (5), 2681–2690. https://doi.org/10.1021/ja209959t.
(45) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157* (1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.
(46) Schuett, T.; Anufriev, I.; Endres, P.; Stumpf, S.; Nischang, I.; Hoeppener, S.; Zechel, S.; Schubert, U. S.; Geitner, R. A User-Guide for Polymer Purification Using Dialysis. *Polym. Chem.* **2023**, *14* (1), 92–101. https://doi.org/10.1039/D2PY00972B.
(47) Schuett, T.; Geitner, R.; Zechel, S.; Schubert, U. S. Dialysis Diffusion Kinetics in Polymer Purification. *Macromolecules* **2021**, *54* (20), 9410–9417. https://doi.org/10.1021/acs.macromol.1c01241.
(48) Malz, F.; Jancke, H. Validation of Quantitative NMR. *J. Pharm. Biomed. Anal.* **2005**, *38* (5), 813–823. https://doi.org/10.1016/j.jpba.2005.01.043.
(49) Yu, H.; Liu, L.; Yin, R.; Jayapurna, I.; Wang, R.; Xu, T. Mapping Composition Evolution through Synthesis, Purification, and Depolymerization of Random Heteropolymers. *J. Am. Chem. Soc.* **2024**, *146* (9), 6178–6188. https://doi.org/10.1021/jacs.3c13909.
(50) Guo, R.; Shi, Z.; Wang, X.; Dong, A.; Zhang, J. Separation and Quantification of Dead Species in Styrene RAFT Polymerization by Gradient Polymer Elution Chromatography. *Polym. Chem.* **2012**, *3* (5), 1314. https://doi.org/10.1039/c2py20102j.
(51) Russell, G. T.; Gilbert, R. G.; Napper, D. H. Chain-Length-Dependent Termination Rate Processes in Free-Radical Polymerizations. 1. Theory. *Macromolecules* **1992**, *25* (9), 2459–2469. https://doi.org/10.1021/ma00035a026.
(52) Buback, M.; Egorov, M.; Gilbert, R. G.; Kaminsky, V.; Olaj, O. F.; Russell, G. T.; Vana, P.; Zifferer, G. Critically Evaluated Termination Rate Coefficients for Free-Radical

Polymerization, 1. The Current Situation. *Macromol. Chem. Phys.* **2002**, *203* (18), 2570–2582. https://doi.org/10.1002/macp.200290041.
(53) Ballard, M. J.; Napper, D. H.; Gilbert, R. G.; Sangster, D. F. Termination-rate Coefficients in Methyl Methacrylate Polymerizations. *J. Polym. Sci. Part Polym. Chem.* **1986**, *24* (5), 1027–1041. https://doi.org/10.1002/pola.1986.080240518.
(54) Buback, M.; Kowollik, C. Termination Kinetics in Free-Radical Bulk Copolymerization: The Systems Dodecyl Acrylate−Dodecyl Methacrylate and Dodecyl Acrylate−Methyl Acrylate. *Macromolecules* **1999**, *32* (5), 1445–1452. https://doi.org/10.1021/ma9814806.
(55) Riahinezhad, M.; McManus, N.; Penlidis, A. Effect of Monomer Concentration and pH on Reaction Kinetics and Copolymer Microstructure of Acrylamide/Acrylic Acid Copolymer. *Macromol. React. Eng.* **2015**, *9* (2), 100–113. https://doi.org/10.1002/mren.201400053.
(56) Steube, M.; Johann, T.; Plank, M.; Tjaberings, S.; Gröschel, A. H.; Gallei, M.; Frey, H.; Müller, A. H. E. Kinetics of Anionic Living Copolymerization of Isoprene and Styrene Using *in Situ* NIR Spectroscopy: Temperature Effects on Monomer Sequence and Morphology. *Macromolecules* **2019**, *52* (23), 9299–9310. https://doi.org/10.1021/acs.macromol.9b01790.
(57) Semler, J. J.; Jhon, Y. K.; Tonelli, A.; Beevers, M.; Krishnamoorti, R.; Genzer, J. Facile Method of Controlling Monomer Sequence Distributions in Random Copolymers. *Adv. Mater.* **2007**, *19* (19), 2877–2883. https://doi.org/10.1002/adma.200602359.
(58) Khan, M.; Guimarães, T. R.; Choong, K.; Moad, G.; Perrier, S.; Zetterlund, P. B. RAFT Emulsion Polymerization for (Multi)Block Copolymer Synthesis: Overcoming the Constraints of Monomer Order. *Macromolecules* **2021**, *54* (2), 736–746. https://doi.org/10.1021/acs.macromol.0c02415.
(59) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer–Protein Hybrids. *Adv. Mater.* **2022**, *34* (30), 2201809. https://doi.org/10.1002/adma.202201809.
(60) Lee, J.; Mulay, P.; Tamasi, M. J.; Yeow, J.; Stevens, M. M.; Gormley, A. J. A Fully Automated Platform for Photoinitiated RAFT Polymerization. *Digit. Discov.* **2023**, *2* (1), 219–233. https://doi.org/10.1039/D2DD00100D.
(61) Smith, A. A. A.; Maikawa, C. L.; Lopez Hernandez, H.; Appel, E. A. Controlling Properties of Thermogels by Tuning Critical Solution Behaviour of Ternary Copolymers. *Polym. Chem.* **2021**, *12* (13), 1918–1923. https://doi.org/10.1039/D0PY01696A.
(62) Buback, M. Kinetics and Mechanism of RAFT Polymerizations. In *RAFT Polymerization*; Moad, G., Rizzardo, E., Eds.; Wiley, 2021; pp 59–93. https://doi.org/10.1002/9783527821358.ch4.
(63) Zapata-Gonzalez, I.; Hutchinson, R. A.; Buback, M.; Rivera-Magallanes, A. Kinetic Importance of the Missing Step in Dithiobenzoate-Mediated RAFT Polymerizations of Acrylates. *Chem. Eng. J.* **2021**, *415*, 128970. https://doi.org/10.1016/j.cej.2021.128970.
(64) Konkolewicz, D.; Hawkett, B. S.; Gray-Weale, A.; Perrier, S. RAFT Polymerization Kinetics: Combination of Apparently Conflicting Models. *Macromolecules* **2008**, *41* (17), 6400–6412. https://doi.org/10.1021/ma800388c.

(65) Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81* (25), 2340–2361.
(66) Moad, G. A Critical Assessment of the Kinetics and Mechanism of Initiation of Radical Polymerization with Commercially Available Dialkyldiazene Initiators. *Prog. Polym. Sci.* **2019**, *88*, 130–188. https://doi.org/10.1016/j.progpolymsci.2018.08.003.
(67) Beuermann, S.; Harrisson, S.; Hutchinson, R. A.; Junkers, T.; Russell, G. T. Update and Critical Reanalysis of IUPAC Benchmark Propagation Rate Coefficient Data. *Polym. Chem.* **2022**, *13* (13), 1891–1900. https://doi.org/10.1039/D2PY00147K.
(68) Smolne, S.; Weber, S.; Buback, M. Propagation and Termination Kinetics of Poly(Ethylene Glycol) Methyl Ether Methacrylate in Aqueous Solution. *Macromol. Chem. Phys.* **2016**, *217* (21), 2391–2401. https://doi.org/10.1002/macp.201600302.
(69) Yorimitsu, H.; Nakamura, T.; Shinokubo, H.; Oshima, K.; Omoto, K.; Fujimoto, H. Powerful Solvent Effect of Water in Radical Reaction: Triethylborane-Induced Atom-Transfer Radical Cyclization in Water. *J. Am. Chem. Soc.* **2000**, *122* (45), 11041–11047. https://doi.org/10.1021/ja0014281.
(70) Coote, M. L.; Krenske, E. H.; Izgorodina, E. I. Computational Studies of RAFT Polymerization–Mechanistic Insights and Practical Applications. *Macromol. Rapid Commun.* **2006**, *27* (7), 473–497. https://doi.org/10.1002/marc.200500832.
(71) Johnston-Hall, G.; Theis, A.; Monteiro, M. J.; Davis, T. P.; Stenzel, M. H.; Barner-Kowollik, C. Accessing Chain Length Dependent Termination Rate Coefficients of Methyl Methacrylate (MMA) via the Reversible Addition Fragmentation Chain Transfer (RAFT) Process. *Macromol. Chem. Phys.* **2005**, *206* (20), 2047–2053. https://doi.org/10.1002/macp.200500354.
(72) Barner-Kowollik, C.; Buback, M.; Charleux, B.; Coote, M. L.; Drache, M.; Fukuda, T.; Goto, A.; Klumperman, B.; Lowe, A. B.; Mcleary, J. B.; Moad, G.; Monteiro, M. J.; Sanderson, R. D.; Tonge, M. P.; Vana, P. Mechanism and Kinetics of Dithiobenzoate-mediated RAFT Polymerization. I. The Current Situation. *J. Polym. Sci. Part Polym. Chem.* **2006**, *44* (20), 5809–5831. https://doi.org/10.1002/pola.21589.

Ch 6. Deep learning driven heteropolymer design

(1) Whitford, D. *Proteins: Structure and Function*; J. Wiley & Sons: Hoboken, NJ, 2005.
(2) Bepler, T.; Berger, B. Learning the Protein Language: Evolution, Structure, and Function. *Cell Syst.* **2021**, *12* (6), 654-669.e3. https://doi.org/10.1016/j.cels.2021.05.017.
(3) Searls, D. B. The Language of Genes. *Nature* **2002**, *420* (6912), 211–217. https://doi.org/10.1038/nature01255.
(4) Scaiewicz, A.; Levitt, M. The Language of the Protein Universe. *Curr. Opin. Genet. Dev.* **2015**, *35*, 50–56. https://doi.org/10.1016/j.gde.2015.08.010.
(5) Shahzad, K.; Mittenthal, J. E.; Caetano-Anollés, G. The Organization of Domains in Proteins Obeys Menzerath-Altmann's Law of Language. *BMC Syst. Biol.* **2015**, *9* (1), 44. https://doi.org/10.1186/s12918-015-0192-9.
(6) Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; Bhowmik, D.; Rost, B. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44* (10), 7112–7127. https://doi.org/10.1109/TPAMI.2021.3095381.
(7) Ofer, D.; Brandes, N.; Linial, M. The Language of Proteins: NLP, Machine Learning & Protein Sequences. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1750–1758. https://doi.org/10.1016/j.csbj.2021.03.022.
(8) Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (12), 866–876. https://doi.org/10.1038/nrm2805.
(9) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537* (7620), 320–327. https://doi.org/10.1038/nature19946.
(10) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341* (6146), 1238149. https://doi.org/10.1126/science.1238149.
(11) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmuller, E.; Olvera De La Cruz, M.; Xu, T. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science* **2018**, *359* (6381), 1239–1243. https://doi.org/10.1126/science.aao0335.
(12) Jiang, T.; Hall, A.; Eres, M.; Hemmatian, Z.; Qiao, B.; Zhou, Y.; Ruan, Z.; Couse, A. D.; Heller, W. T.; Huang, H.; De La Cruz, M. O.; Rolandi, M.; Xu, T. Single-Chain Heteropolymers Transport Protons Selectively and Rapidly. *Nature* **2020**, *577* (7789), 216–220. https://doi.org/10.1038/s41586-019-1881-0.
(13) Ruan, Z.; Li, S.; Grigoropoulos, A.; Amiri, H.; Hilburg, S. L.; Chen, H.; Jayapurna, I.; Jiang, T.; Gu, Z.; Alexander-Katz, A.; Bustamante, C.; Huang, H.; Xu, T. Population-Based Heteropolymer Design to Mimic Protein Mixtures. *Nature* **2023**, *615* (7951), 251–258. https://doi.org/10.1038/s41586-022-05675-0.
(14) Jayapurna, I.; Ruan, Z.; Eres, M.; Jalagam, P.; Jenkins, S.; Xu, T. Sequence Design of Random Heteropolymers as Protein Mimics. *Biomacromolecules* **2023**, *24* (2), 652–660. https://doi.org/10.1021/acs.biomac.2c01036.

(15) Hilburg, S. L.; Ruan, Z.; Xu, T.; Alexander-Katz, A. Behavior of Protein-Inspired Synthetic Random Heteropolymers. *Macromolecules* **2020**, *53* (21), 9187–9199. https://doi.org/10.1021/acs.macromol.0c01886.
(16) Ieremie, I.; Ewing, R. M.; Niranjan, M. Protein Language Models Meet Reduced Amino Acid Alphabets. *Bioinformatics* **2024**, *40* (2), btae061. https://doi.org/10.1093/bioinformatics/btae061.
(17) Zhou, Y.; Gong, B.; Jiang, T.; Xu, T.; Huang, H. Stochastic Variational Methods in Generalized Hidden Semi-Markov Models to Characterize Functionality in Random Heteropolymers. **2022**. https://doi.org/10.48550/ARXIV.2207.01813.
(18) Sinai, S.; Kelsic, E.; Church, G. M.; Nowak, M. A. Variational Auto-Encoding of Protein Sequences. **2017**. https://doi.org/10.48550/ARXIV.1712.03346.
(19) Riesselman, A. J.; Ingraham, J. B.; Marks, D. S. Deep Generative Models of Genetic Variation Capture the Effects of Mutations. *Nat. Methods* **2018**, *15* (10), 816–822. https://doi.org/10.1038/s41592-018-0138-4.
(20) Greener, J. G.; Moffat, L.; Jones, D. T. Design of Metalloproteins and Novel Protein Folds Using Variational Autoencoders. *Sci. Rep.* **2018**, *8* (1), 16189. https://doi.org/10.1038/s41598-018-34533-1.
(21) Costello, Z.; Martin, H. G. How to Hallucinate Functional Proteins. **2019**. https://doi.org/10.48550/ARXIV.1903.00458.
(22) The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49* (D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100.
(23) Monera, O. D.; Sereda, T. J.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. Relationship of Sidechain Hydrophobicity and α-Helical Propensity on the Stability of the Single-Stranded Amphipathic α-Helix. *J. Pept. Sci.* **1995**, *1* (5), 319–329. https://doi.org/10.1002/psc.310010507.
(24) Kramer, M. A. Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. *AIChE J.* **1991**, *37* (2), 233–243. https://doi.org/10.1002/aic.690370209.
(25) Vidal, R.; Ma, Y.; Sastry, S. S. Principal Component Analysis. In *Generalized Principal Component Analysis*; Vidal, R., Ma, Y., Sastry, S. S., Eds.; Springer: New York, NY, 2016; pp 25–62. https://doi.org/10.1007/978-0-387-87811-9_2.
(26) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł. ukasz; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2017; Vol. 30.
(27) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
(28) Chandonia, J.-M.; Guan, L.; Lin, S.; Yu, C.; Fox, N. K.; Brenner, S. E. SCOPe: Improvements to the Structural Classification of Proteins – Extended Database to Facilitate Variant Interpretation and Machine Learning. *Nucleic Acids Res.* **2022**, *50* (D1), D553–D559. https://doi.org/10.1093/nar/gkab1054.

(29) Wattenberg, M.; Viégas, F.; Johnson, I. How to Use T-SNE Effectively. *Distill* **2016**, *1* (10), 10.23915/distill.00002. https://doi.org/10.23915/distill.00002.
(30) Patel, R. A.; Borca, C. H.; Webb, M. A. Featurization Strategies for Polymer Sequence or Composition Design by Machine Learning. *Mol. Syst. Des. Eng.* **2022**, *7* (6), 661–676. https://doi.org/10.1039/D1ME00160D.
(31) *Weights & Biases: The AI Developer Platform*. Weights & Biases. https://wandb.ai/site (accessed 2024-05-06).
(32) Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Davison, J.; Shleifer, S.; von Platen, P.; Ma, C.; Jernite, Y.; Plu, J.; Xu, C.; Scao, T. L.; Gugger, S.; Drame, M.; Lhoest, Q.; Rush, A. M. HuggingFace's Transformers: State-of-the-Art Natural Language Processing. **2019**. https://doi.org/10.48550/ARXIV.1910.03771.
(33) Carugo, O. Amino Acid Composition and Protein Dimension. *Protein Sci.* **2008**, *17* (12), 2187–2191. https://doi.org/10.1110/ps.037762.108.
(34) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer–Protein Hybrids. *Adv. Mater.* **2022**, *34* (30), 2201809. https://doi.org/10.1002/adma.202201809.

Ch 1. Biocatalysis under confinement: a molecular scale feedback loop

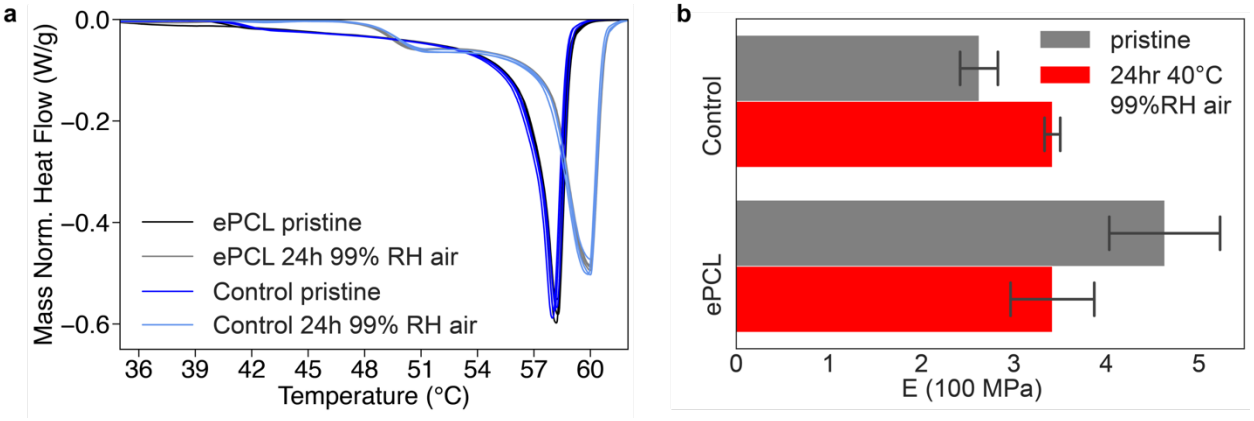


Figure S1.1 | Differential scanning calorimetry (DSC) (**a**) and Young's modulus from quasi-static tensile testing (**b**) of ePCL and no enzyme control films pre and post 24 h incubation in 40 ºC humid air, indicating that lamellae thickening is independent of degradation.

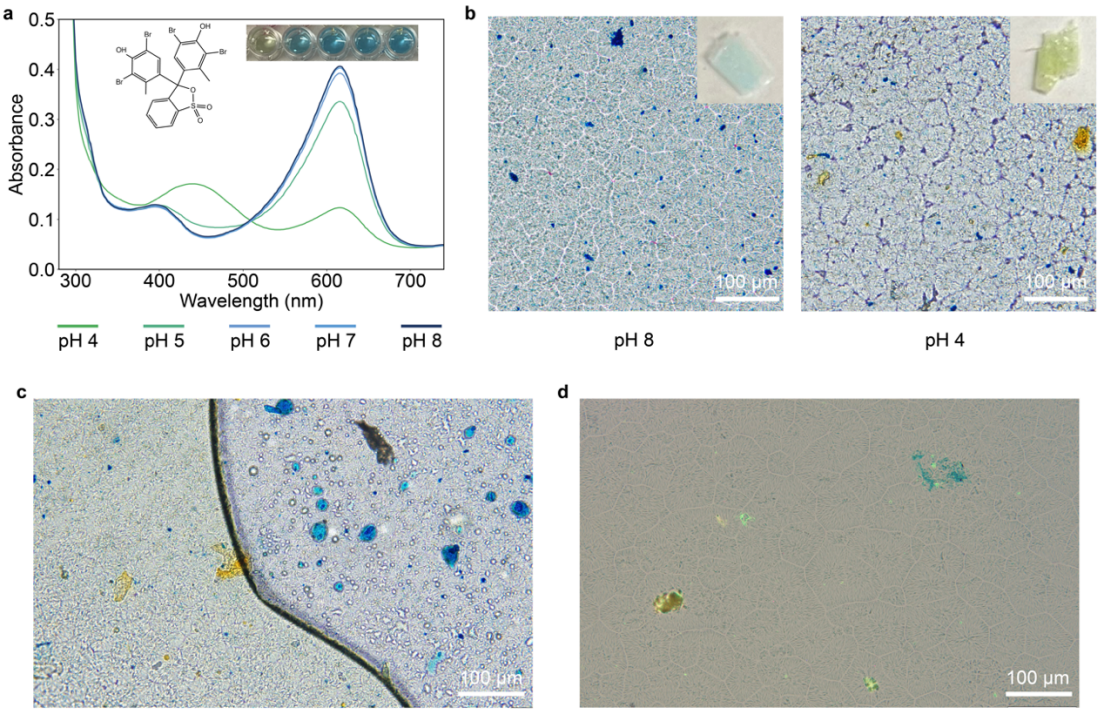


Figure S1.2 | Bromocresol green color change can be used to detect internal change of pH. **a**, UV-vis spectra, and visual color change of Bromocresol green at different pH. **b**, Optical microscope (OM) and macroscopic image of PCL-BG-RHP films at different pH. **c,** image showing pH change from 7 to 4 under the microscope as pH 4 buffer spreads across the

film. **d**, OM overlayed with fluorescent microscopy image of tagged lipase enzymes shows co-localization of lipase and pH indicator.

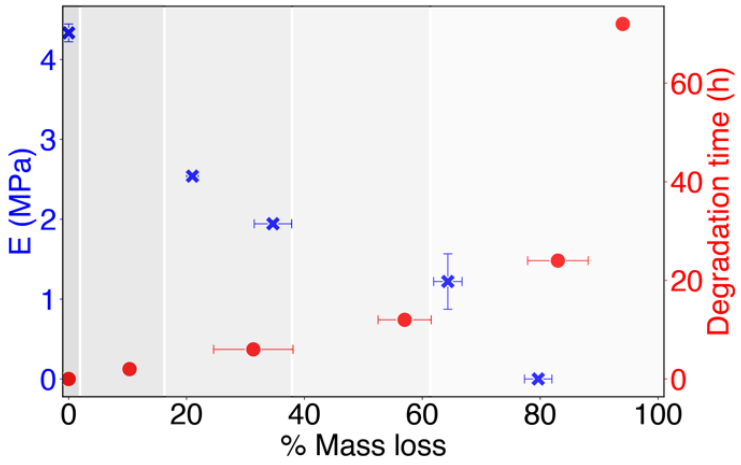


Figure S1.3 | Degradation time taken to reach ePCL films of varying percentage mass loss and the Young's modulus films at these points as measured by quasi-static tensile testing. Ranges in mass loss with similar morphological features are grouped together by grey bars.

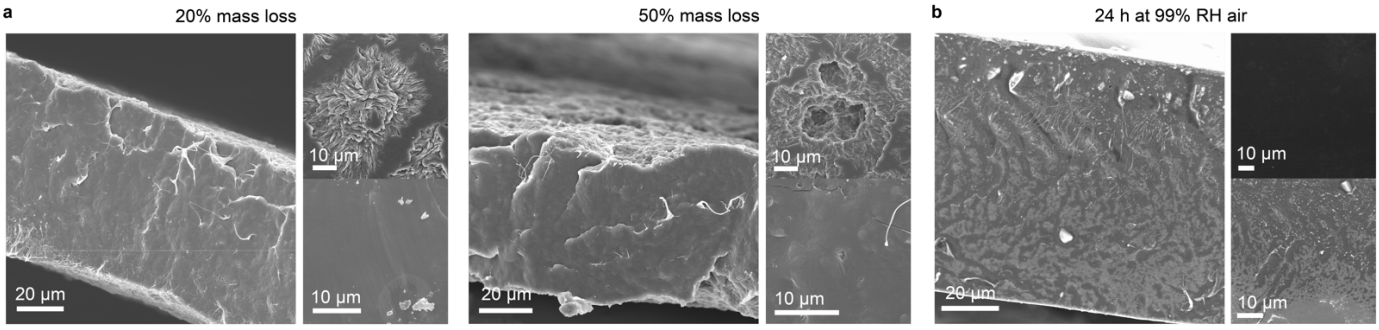


Figure S1.4 | Scanning electron microscopy (SEM) images of cryofractured corss-sections of films. Smaller images show details of the surface (top) and magnified cross sections (bottom) of films. **a,** Surface degraded PCL films at 20% and 50% mass loss. **b,** ePCL films after 24h incubation at 99% RH.

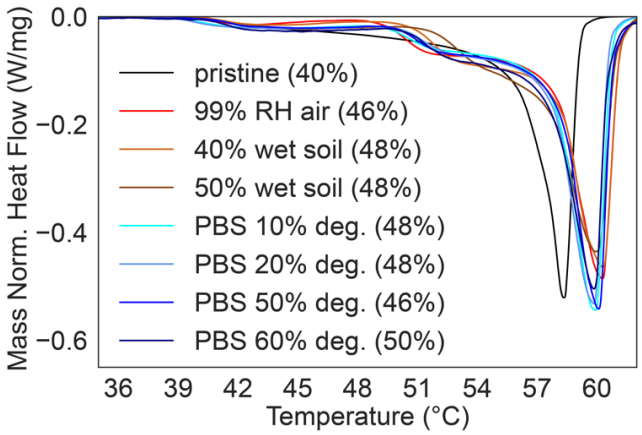


Figure S1.5 | Differential Scanning Calorimetry (DSC) profiles for films pre and post degradation. Data for soil and air are of 24h incubated samples and films degraded in PBS are of specified mass loss.

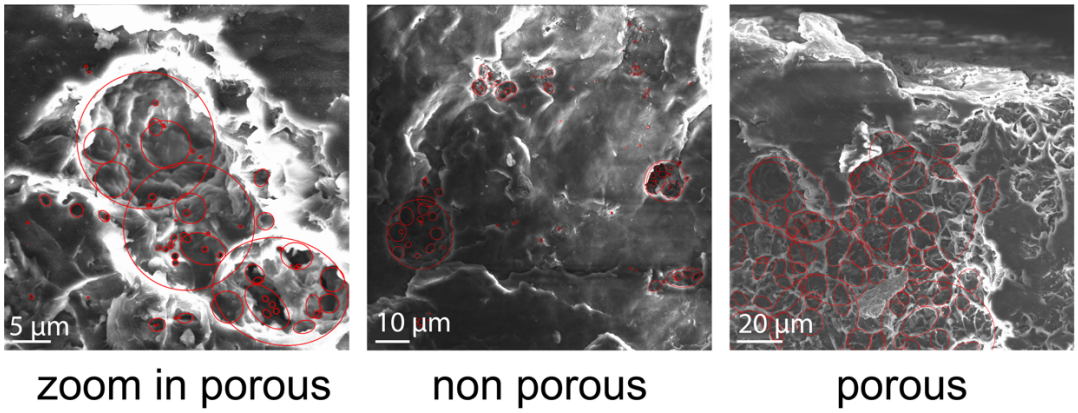


Figure S1.6 | Image analysis of scanning electron microscopy (SEM) images of a 20% porous film. Pores (assumed to be elliptical) are manually determined and labeled in red using Adobe Illustrator in the above images. Ellipses were filled with color and exported as images, then processed using the analyze particle function of ImageJ to obtain the distribution of area. A distribution of pore diameters was determined from the calculated area and plotted in **Figure 1.4d**.

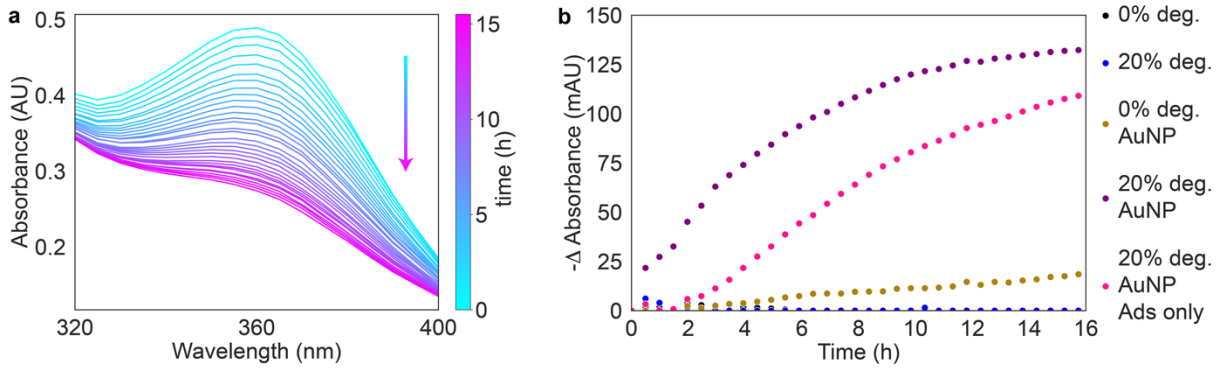


Figure S1.7 | **a**, UV-vis monitoring of oxytetracycline (360 nm) oxidation by functionalized films at 15 min intervals. **b,** rates of oxytetracycline oxidation for films characterized through deterioration of UV-vis signal.

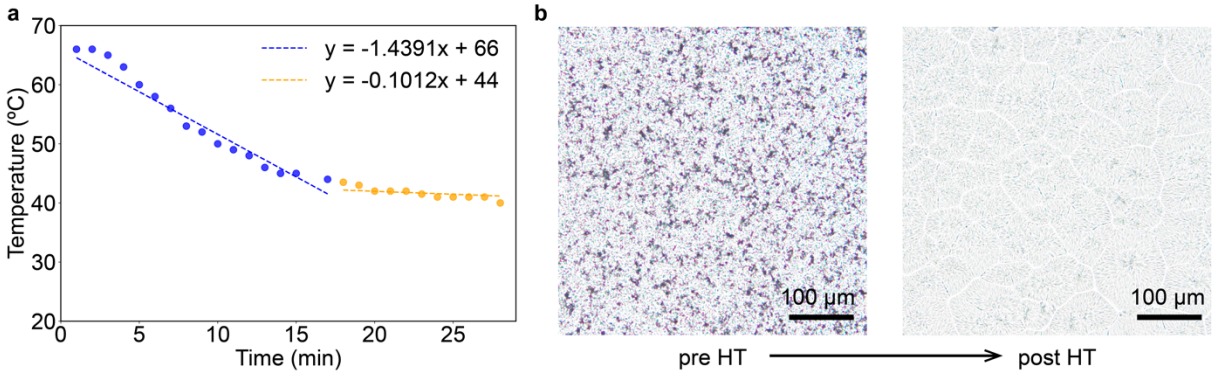


Figure S1.8 | Heat treatment of PCL films. **a**, temperature profile of PCL film undergoing heat treatment. **b**, Optical microscopy images of films pre and post heat treatment.

Ch 2. Enzyme degradation mechanisms on and in semi-crystalline substrates

Table S2.1 | Key information on the 15 hydrolases used. Crystal structures for enzymes without PDB codes were generated via AlphaFold2.

Code	Protein	Organism	Kingdom	PDB / UniProt	Molecular Weight	Metal Ion
BC	Lipase	Burkholderia cepacia	Bacteria	3LIP	33.13	Ca^{2+}
PF	Lipase	Pseudomonas fluorescens	Bacteria	Q7WZT7	32.37	None
AO	Lipase	Aspergillus oryzae	Fungi	5XK2	31.36	None
CA	Lipase	Candida antarctica	Fungi	1TCA	33.02	None
MM	Lipase	Rhizomucor (Mucor) miehei	Fungi	1LGY	29.59	None
RO	Lipase	Rhizopus oryzae	Fungi	1LGY	29.59	None
CR	Lipase	Candida rugosa	Fungi	1CRL	57.1	None
P-TA	Protease (Proteinase K)	Tritirachium album	Fungi	2PRK	28.91	Ca^{2+} (2)
P-BL	Protease	Bacillus Licheniformis	Bacteria	1SCA	27.29	Ca^{2+} , Na^{+}
P-SG	Protease	Streptomyces griseus	Bacteria	4SGA	18.01	None
C-BP	Chymotrypsin	Bovine Pancreas	Animal	1ACB	25.67	None
C-AO	Cutinase	Aspergillus oryzae	Fungi	3GBS	20.64	None
CV	Lipase	Chromobacterium viscosum	Bacteria	1CVL	33.16	None
AN	Lipase	Aspergillus niger	Fungi	1UKC	56.89	None
CC	Lipase	Candida cylindracea	Fungi	1CLE	57.3	None

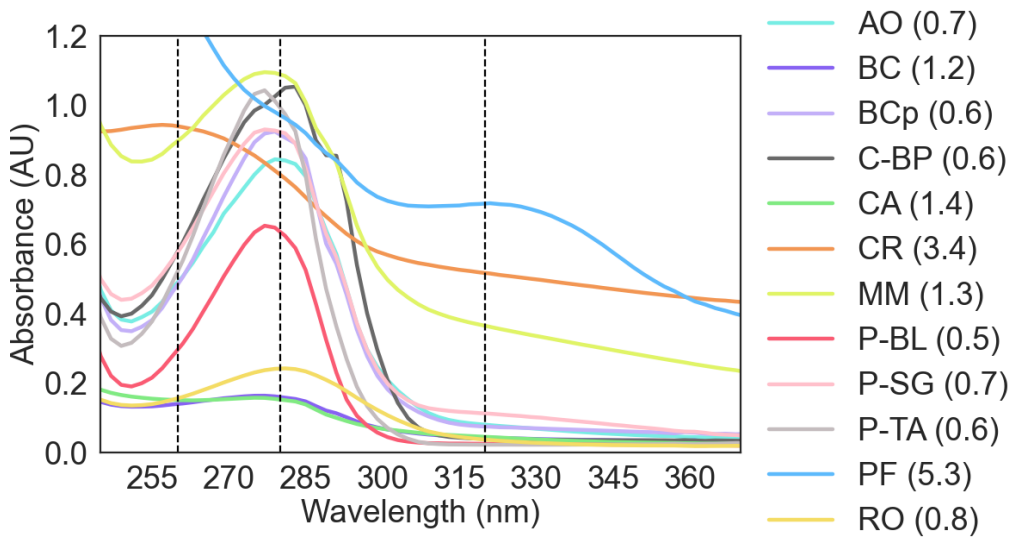


Figure S2.1 | UV-Vis absorbance to assess purity of commercially purchased lyophilized protein powder. Proteins made to 1 mg/mL solutions in water. To get A280 signal, BC was 10x concentrated, PF was 5x concentrated, RO was 3x concentrated and CA was 2x concentrated. Full details for each abbreviated enzyme in **Table S1**, and the A320 baseline corrected A260/A280 are displayed in the legend in brackets for each enzyme.

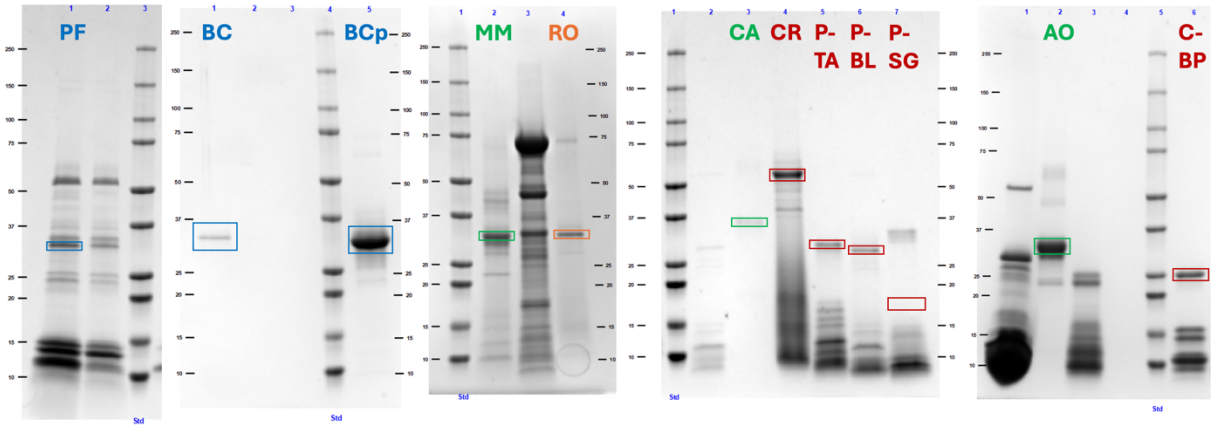


Figure S2.2 | SDS-PAGE of experimentally used proteins. Boxes indicate molecular weight corresponding to the protein crystal structure that was featurized and assumed to be representative of the active plastic degrading enzyme.

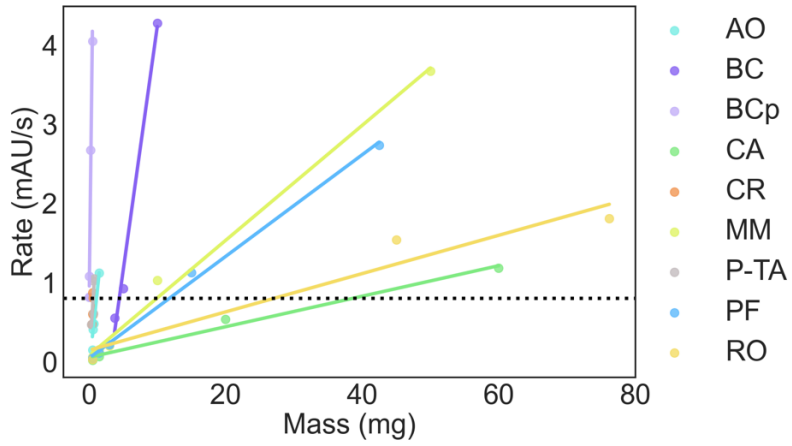


Figure S2.3 | Enzyme catalyzed 4-Nitrophenyl butyrate hydrolysis. Kinetic assay measuring 410nm product absorbance and taking the linear fit over initial slope as a proxy for enzyme activity rate on ester bond hydrolysis. 3-5 masses were screened for each enzyme and linear line of best fit is plotted. Interpolated mass required for rate of 0.8 mAU/s was used for as a scaling factor for enzyme solution concentration for plastic degradation, to account for impurities in commercial enzyme blends.

Table S2.2 | Normalized activity and masses for experimentally tested hydrolases with PCL degrading activity. *AO and CA surface hydrolysis rates and BCp embedded went to 100% degradation in under 24 hours so hydrolysis rate can only be provided as lower bounds. **MM and RO embedded mass losses were <2% and <1% respectively and are within experimental error so are likely innacurate.

Code	4NPB Normalized Activity ratio	Surface Degradation Enzyme (mg)	Embed Degradation Enzyme (mg)	24h Surface Hydrolysis Rate (g_{PCL} s^{-1} g_{enz}^{-1})	24h Embed Hydrolysis Rate (g_{PCL} s^{-1} g_{enz}^{-1})
BCp	1	0.00125	0.00125	121,000	169,000*
BC	115	0.144	0.0125	1,220	2,590
PF	278	0.348	0.0125	267	1,040
AO	15	0.0188	0.0125	11,500*	5,010
CA	840	1.05	0.0125	202*	2,250
MM	195	0.244	0.0125	682	356**
RO	600	0.750	0.0125	29	0**

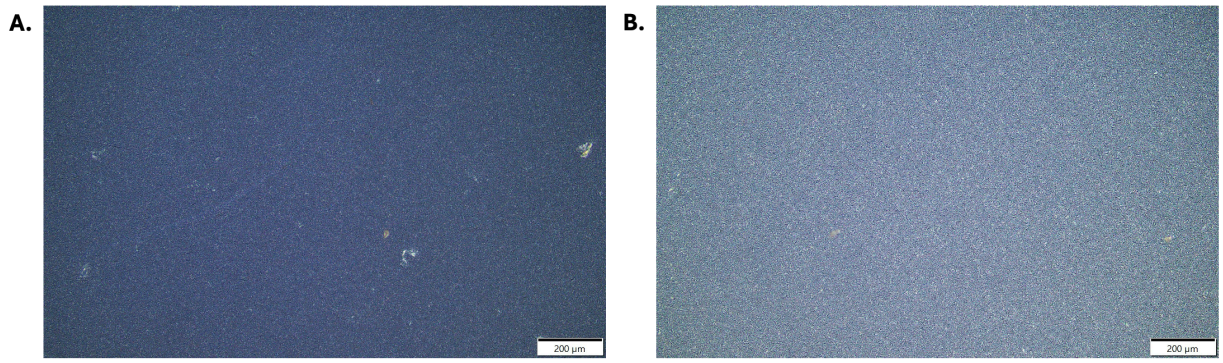


Figure S2.4 | Polarized optical microscopy (POM) images of A. pure PCL and **B.** PCL with embedded 0.05wt% BC lipase and 1wt% random heteropolymer.

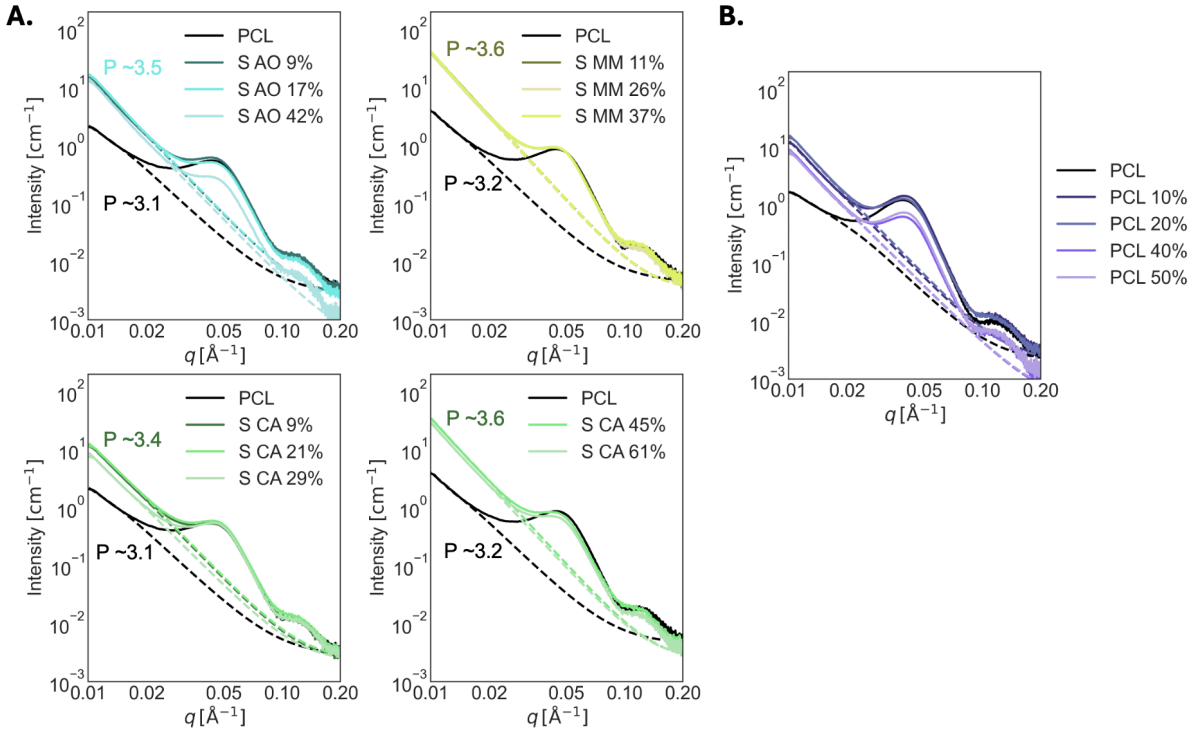


Figure S2.5 | Transmission small angle x-ray scattering (SAXS) profiles (solid lines) with power law fit to low q and extrapolated to high q (dashed lines). **A.** As-cast PCL films degraded by enzymes (top-left to bottom-right): AO Round 1, MM, CA Round 1, CA Round 2 Round 1 and Round 2 SAXS data were collected 4 months apart with different beamline condition. **B.** Thermally annealed PCL films degraded by BCp.

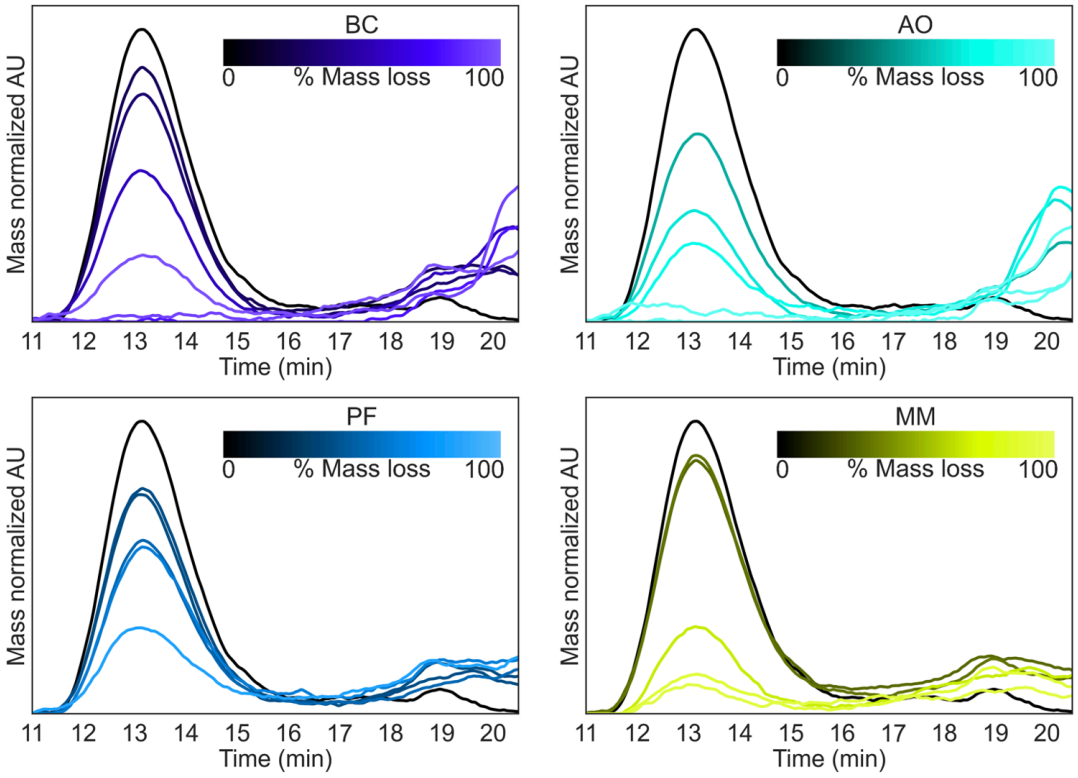


Figure S2.6 | Enzymatic solution (surface) degradation product molecular weight distribution of Poly(caprolactone) (PCL), measured by gel permeation chromatography (GPC). All spectra are mass normalized by initial mass of films pre-degradation.

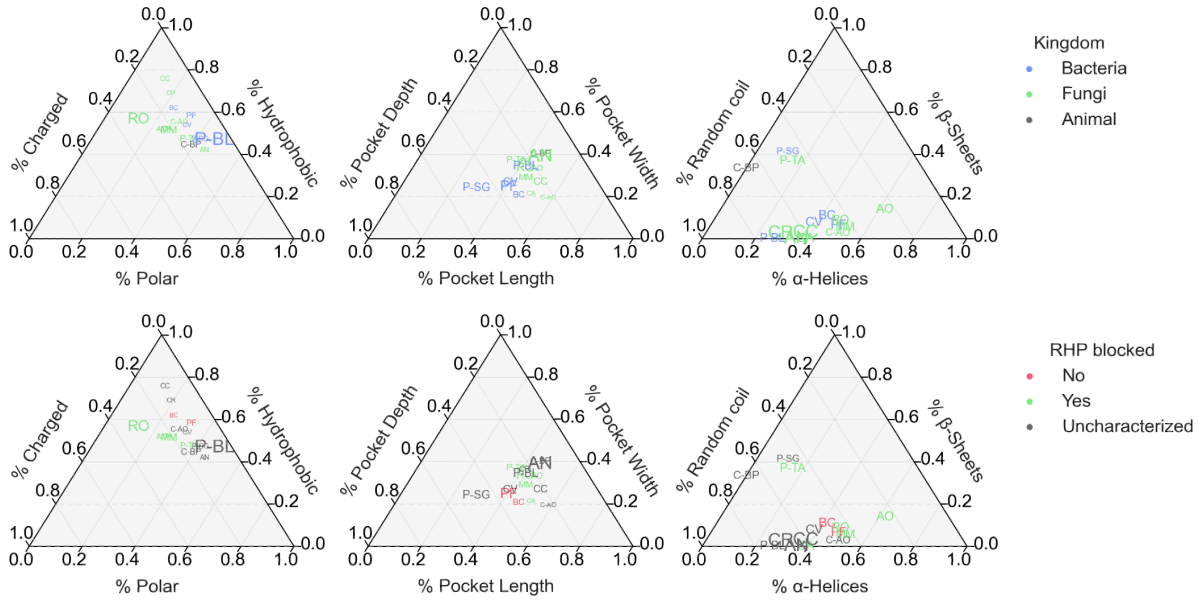


Figure S2.7 | Ternary plots showing key binding pocket chemistry, geometry, and flexibility (secondary structure) features for 15 hydrolases. The size of label markers for chemistry, geometry and flexibility plots represent positive : negative charge ratio, volume, and molecular weight respectively. The % width, % depth and % length were calculated as % depth = depth / (length + width + depth), so a perfectly spherical enzyme or hemispherical binding pocket would have coordinates [0.33,0.33,0.33]. The depth, width and length for entire protein were chosen to match the axes used to define the binding pocket depth, width, and length. Enzymes are colored by **A.** Kingdom of source organism **B.** If random heteropolymer binding blocks the enzyme binding pocket or not.

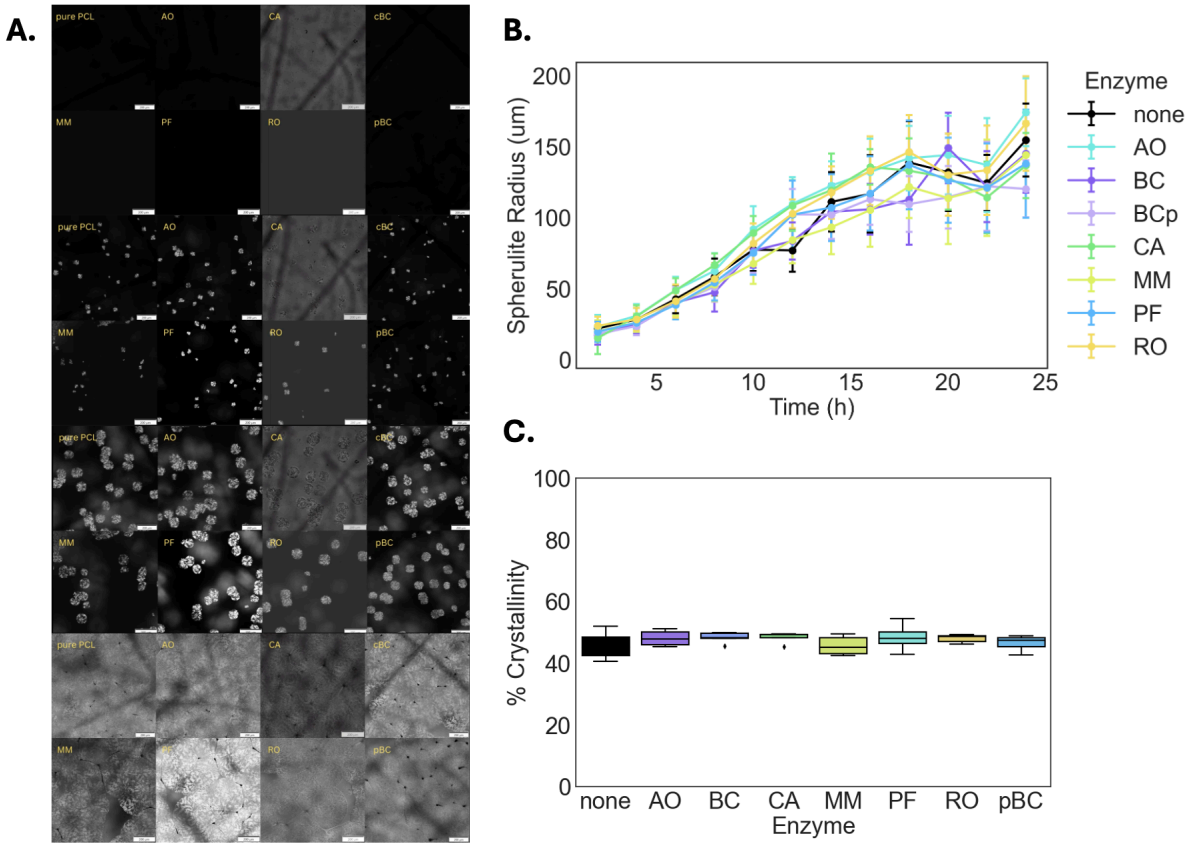


Figure S2.8 | Crystallinity of embedded films is negligibly affected by embedding nanoscopically dispersed commercial enzyme and RHP (<1wt%). A. Screenshots taken from full timelapse of PCL films with embedded enzymes (0.5wt%) and RHP (1wt%) starting from fully melted and recrystallized at 49°C for 24 hours. Full video can be found at: www.youtube.com/watch?v=KHf3uTg6Ups **B.** Manually measured spherulite radius distribution at various timepoints from (**A**). **C.** Starting crystallinity of as-cast films used in degradation trials for each enzyme with RHP.

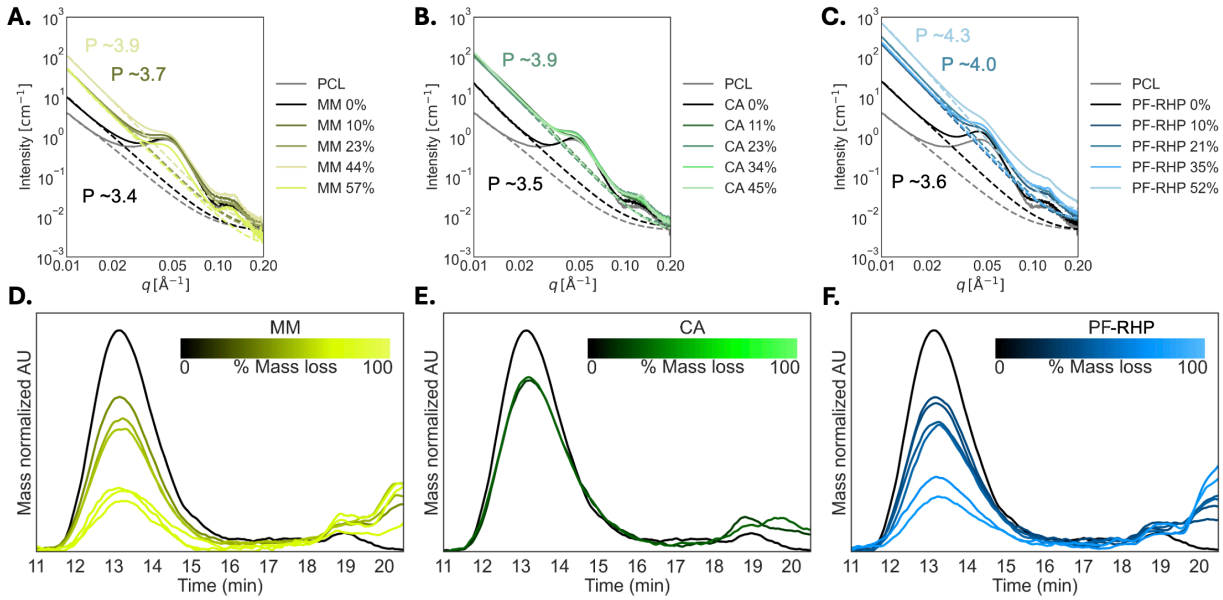


Figure S2.9 | Embedded enzyme degradation of Poly(caprolactone) (PCL) by 0.5wt% MM, 0.5wt% CA, and 0.5wt% PF and 1wt% random heteropolymer (RHP) characterized by transmission small angle x-ray scattering (SAXS) profiles (solid lines) with power law fit to low q and extrapolated to high q (dashed lines) (**A., B., C.**) and gel permeation chromatography (**D., E., F.**). All spectra are mass normalized by initial mass of films pre-degradation.

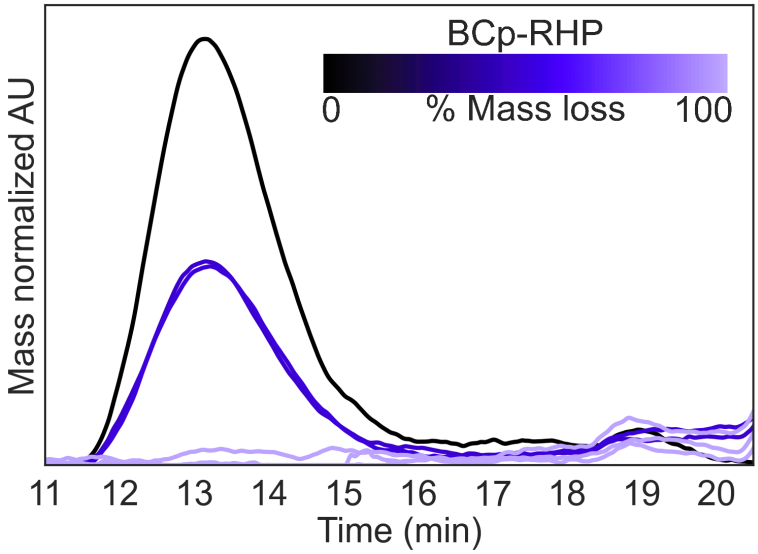


Figure S2.10 | BCp-RHP Embedded degradation product molecular weight distribution of Poly(caprolactone) (PCL), measured by gel permeation chromatography (GPC). All spectra are mass normalized by initial mass of films pre-degradation. BCp 0.05wt% and RHP 1wt% loading.

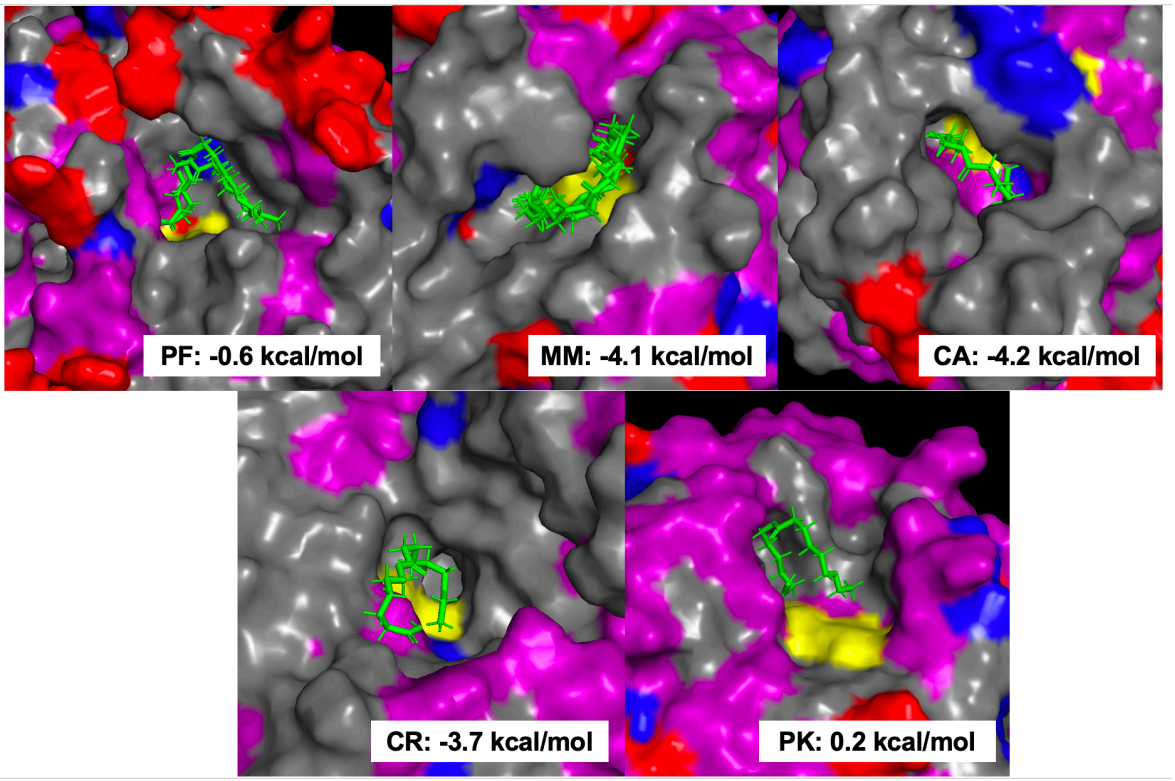


Figure S2.11 | Molecular docking of polycaprolactone (PCL) dimers (green) onto crystal structures of **PF, MM, CA, CR and PK**. The top 3 poses are shown superimposed, and the top docking energy score is labeled. Catalytic triad, hydrophobic, polar, positive, and negative residues are colored yellow, grey, purple, red, and blue respectively.

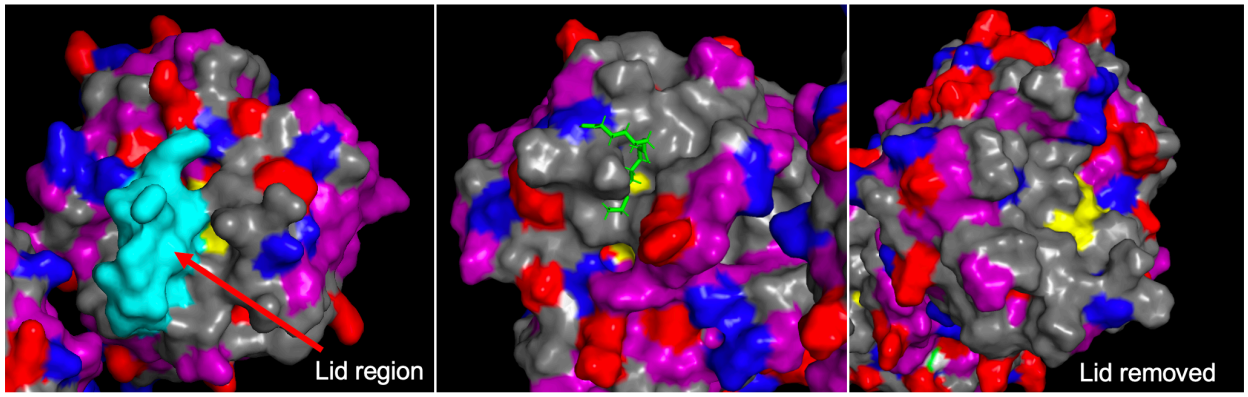


Figure S2.12 | Crystal structure of AO A. Original PDB file with lid (light blue) closed. **B.** Example of polycaprolactone (PCL) dimers (green) docked onto crystal structures from (A.) **C.** Manually removed lid structure used for docking in **Figure 2.8b**. Catalytic triad, hydrophobic, polar, positive, and negative residues are colored yellow, grey, purple, red, and blue respectively.

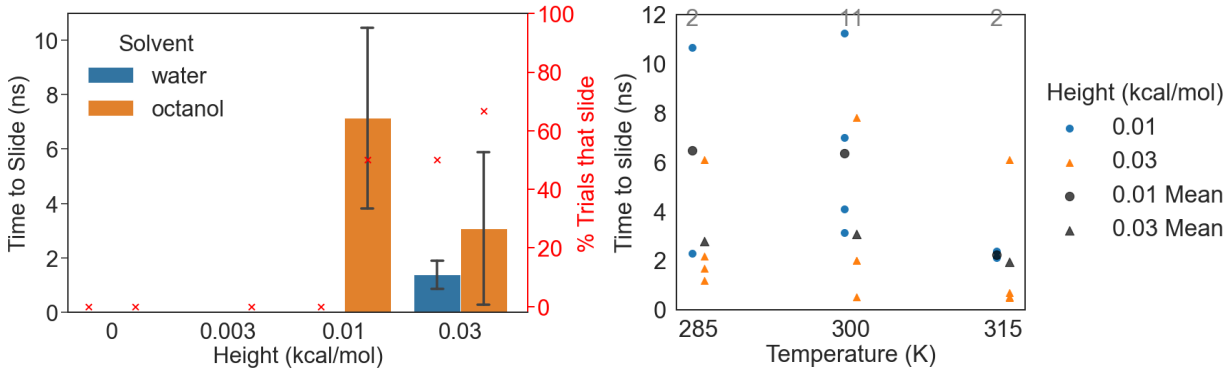


Figure S2.13 | Screening hyperparameters of a metadynamics simulation of PCL 4-mer chain slide through BC binding pocket. Optimization trials screening: **A.** Varying gaussian heights in water and octanol solvent boxes. Time to slide indicates the simulated time required to see binding interaction move from the head group of the 4-mer to an ester bond further down the chain. % Trials that slide is important as not all trials processively slid as there is a degree of stochasticity between each simulation. Trials that did not slide were not averaged for the time to slide bar plots. **B.** Varying simulated temperatures and gaussian hill heights in an octanol box. Grey points indicate mean time to slide, blue squares and orange triangles indicate individual simulated points, grey numbers at the top of the plot indicated the number of simulated trials in which no chain slide occurred.

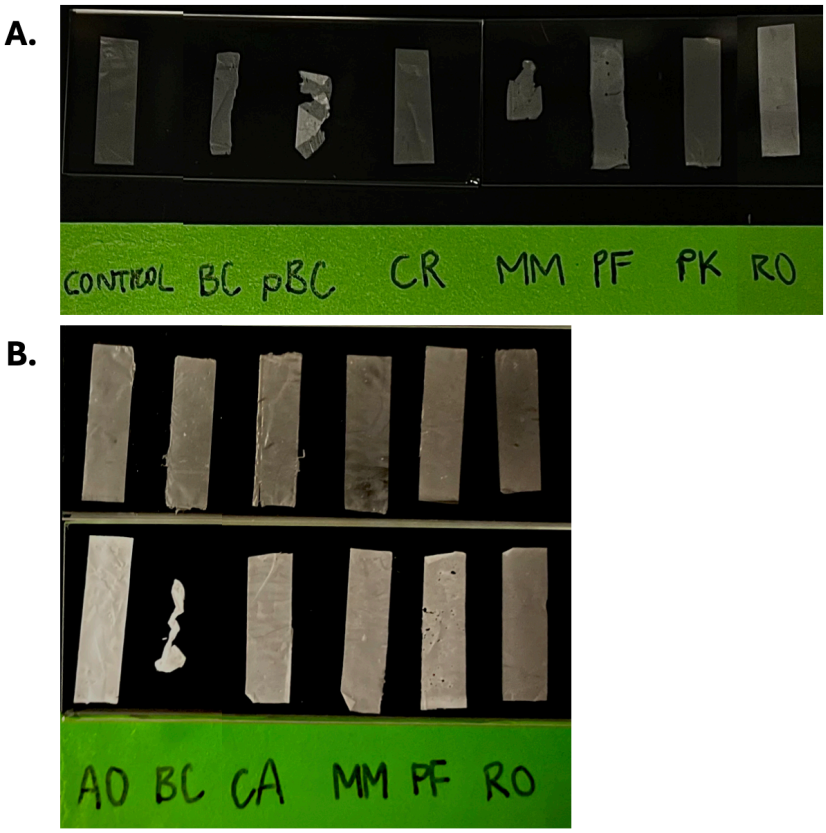


Figure S2.14 | Images of films before and after enzymatic degradation: **A.** Solution (control = before) AO and CA were not imaged as it fully degraded, all negative controls not shown look like control, CR and PF. **B.** Embedded (0.5wt%) degradation with RHP (1wt%), top row is pre-, and bottom row is post-24 hours in 40°C PBS. BCp fully degraded so not shown.

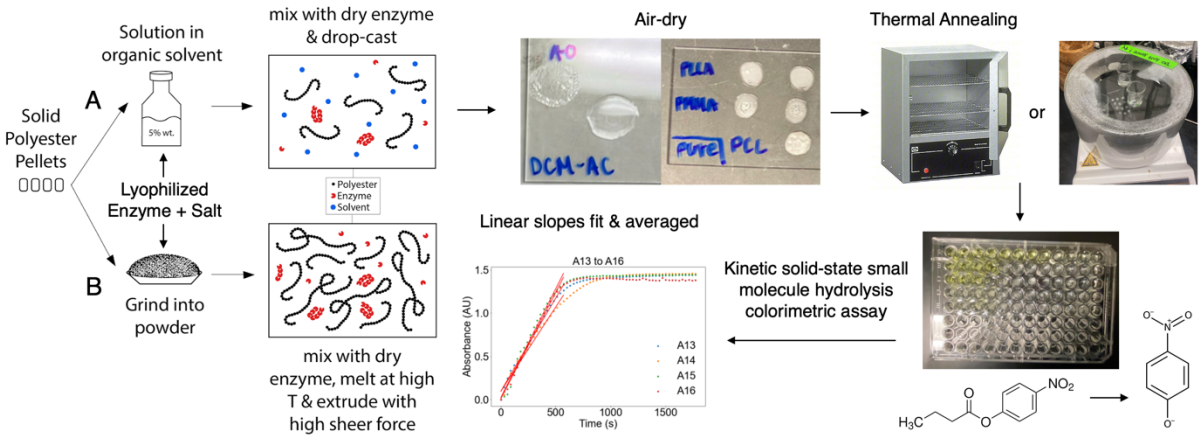


Fig S3.1 | Thermostability in polymer melt experimental setup. Solid polymer pellets either go through **A.** Our experimental assay that is trying to mimic **B.** industrial melt extrusion process that does not involve organic solvents.

Table S3.1 | Key information on the 6 plastics screened. Values in regular font were provided by the supplier, values in italics were taken from literature, values in bold were experimentally measured.

	PLLA30	**PLLA85**	**PLLA100**	**PMMA**	**PCl**	**PS**
Poly (...)	L-lactic acid	L-lactic acid	L-lactic acid	Methyl methacrylate	Caprolactone	Styrene
~Mn (kDa)	32	85	86	500	80	260
Crystallinity	**Semi**	**Semi**	**Semi**	**Amorphous**	**Semi**	**Amorphous**
~T_g (°C)	**58**			**105**	*-60*	**126**
~T_m (°C)	**163-177**			-	50-63	-
Polarity	*Most*	*Most*	*Most*	*More*	*Less*	*Least*

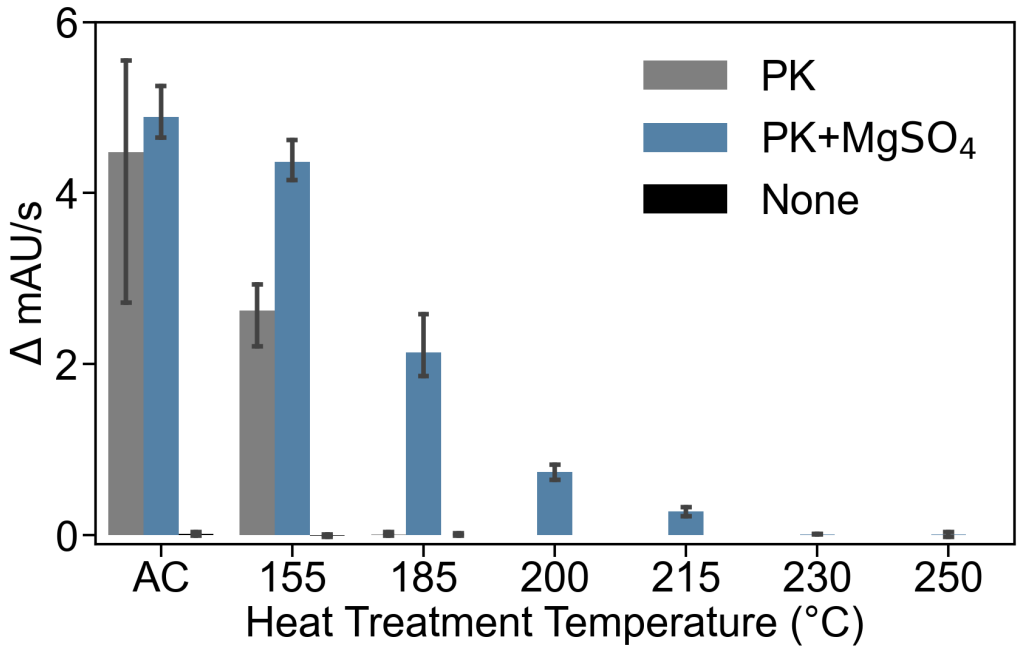


Fig S3.2 | $MgSO_4$ thermostabilized Proteinase K (PK). Non-As Cast (AC) normalized data for of **Figure 3.1a**.

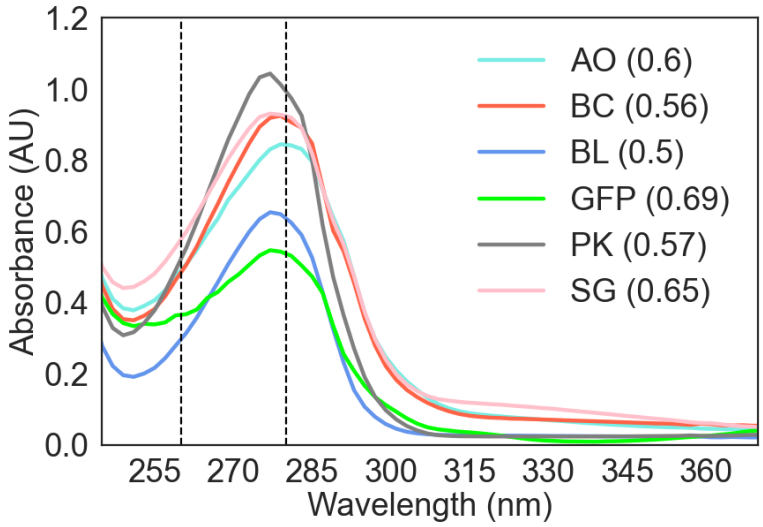


Fig S3.3 | UV-Vis of proteins. Representative spectra of triplicates of each protein shown. A260/A280 shown in brackets indicating very high purity for AO, BC, BL and PK, and reasonable purity for GFP and SG. **A.** PK **B.** BC **C.** AO **D.** BL. **E.** SG **F.** GFP.

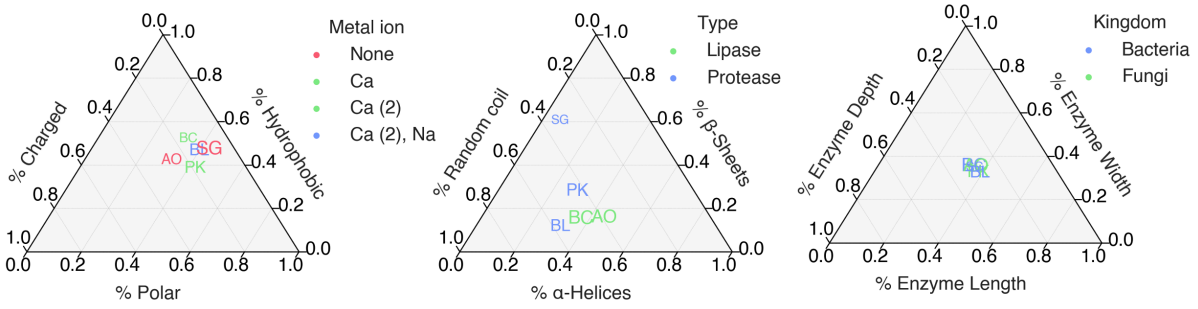


Fig S3.4 | Enzyme surface featurization. Ternary plots that compare enzyme surface: **A.** chemistry, where size corresponds to the ratio of positive : negative surface charges and color corresponds to metal ion presence. **B.** secondary structure, where size corresponds to molecular weight (Mw) and color corresponds to the type of enzyme. **C.** geometry, where each axis is the % of total dimension for example a perfectly spherical enzyme would be at coordinates [0.33,0.33,0.33], size corresponds to Mw and legend is what kingdom the enzyme's original organism is from.

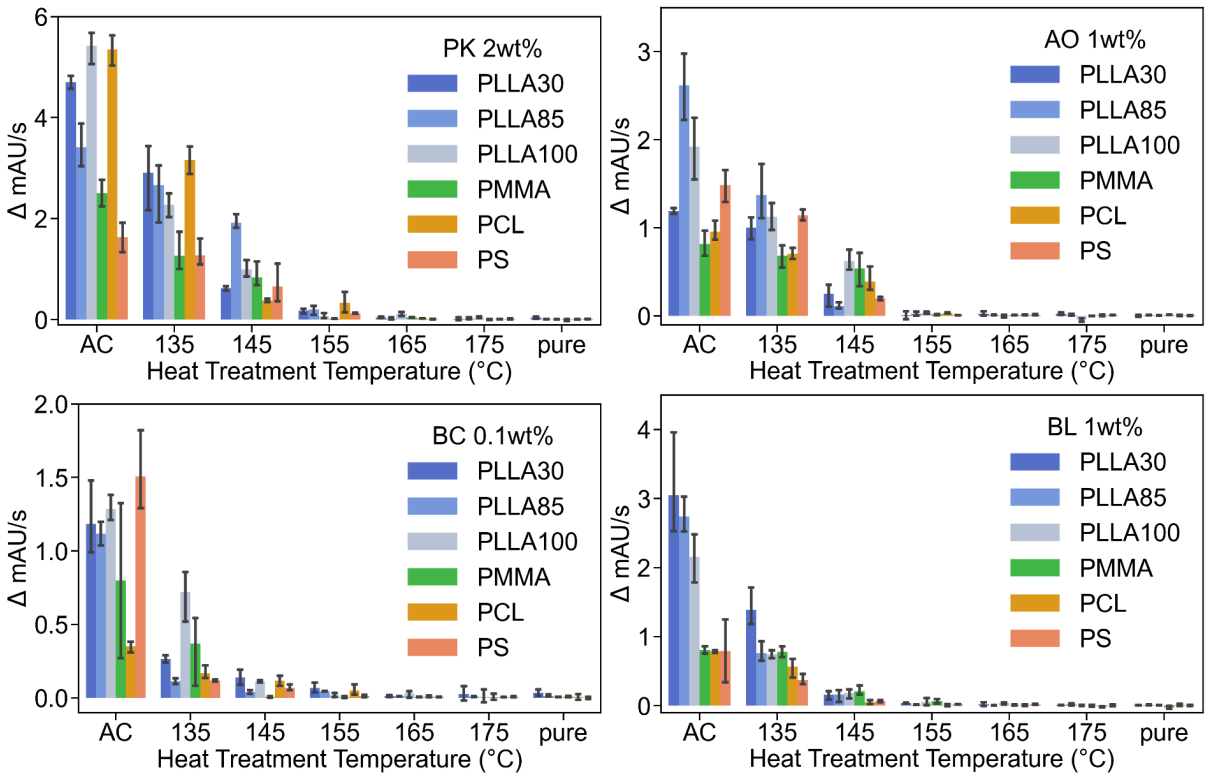


Fig S3.5 | 155°C universal deactivation. Non-As cast (AC) normalized version of **Figure 3.2**

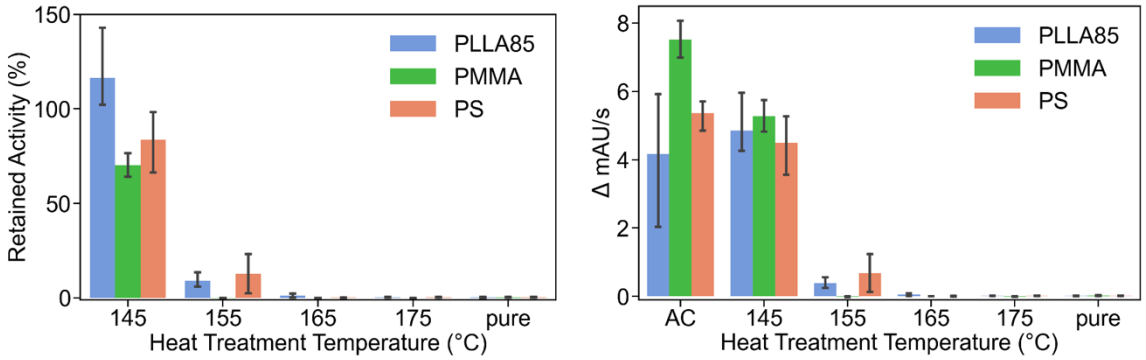


Fig S3.6 | 10wt% PK loading in various plastics. Protein thermostablity does not improve with increased % loading in polymer films.

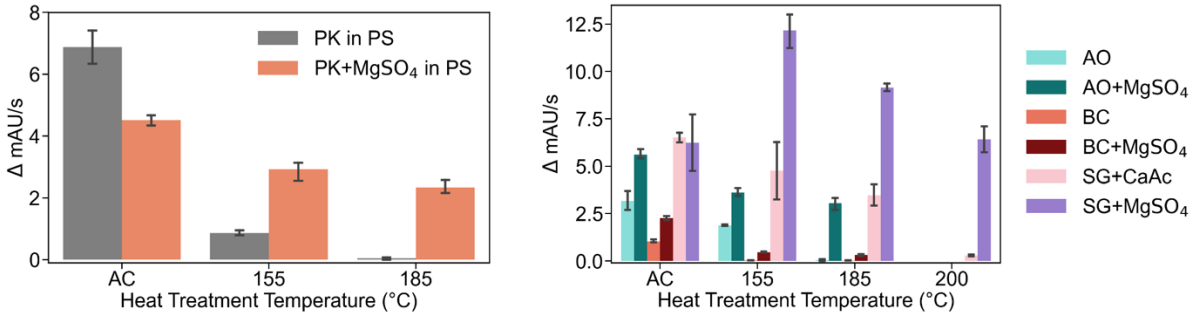


Fig S3.7 | Non-As-cast normalized version of Figures 3.3a and 3.3b.

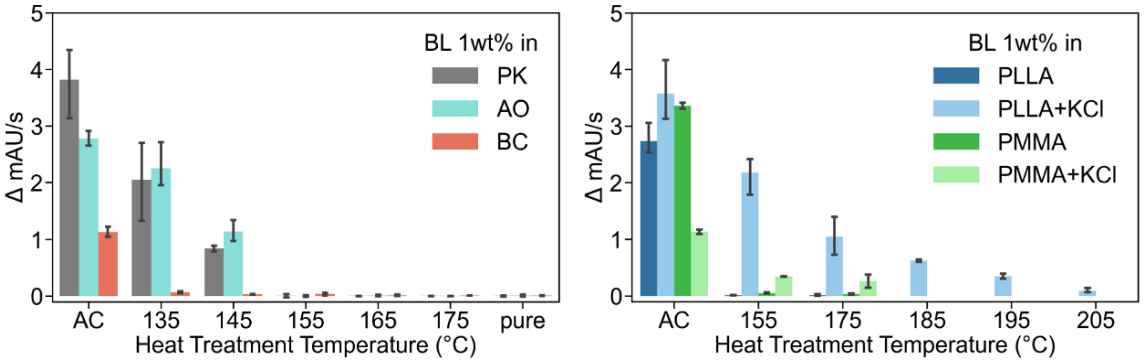


Fig S3.8 | Non-As-cast normalized version of Figures 3.4a and 3.4d.

Ch 4. Biocheminformatic Design of Biomimetic Heteropolymers

Table S4.1 | Table of reactivity ratios (Row-Col reactivity) used in this work. Taken from the RHPapp online database. Tabulated reactivity values should be read as RRRowCol = kRowRow / kRowCol. All self-additions are assumed to be 1, therefore when RRRowCol < 1, for growing polymer chain where the row monomer is terminal, the column monomer is more likely to add on than another row monomer.

	MMA	OEGMA500	2EHMA	3SPMA	STY
MMA	1	0.89	0.89	0.89	0.46
OEGMA500	1.09	1	1	1	0.46
2EHMA	1.09	1	1	1	0.46
3SPMA	1.09	1	1	1	0.46
STY	0.52	0.52	0.52	0.52	1

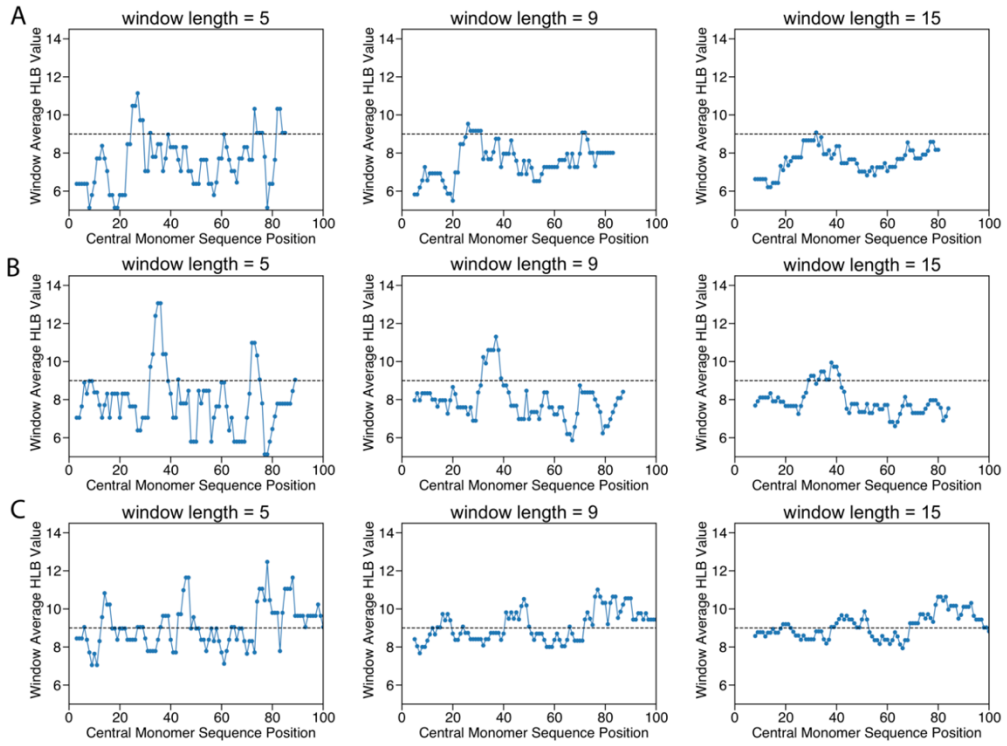


Figure S4.1 | Example segment level sliding window analysis of RHP sequences. Sliding window of 3 varying window lengths applied to random sequences sampled from a simulated 4 monomer RHP of methyl methacrylate (MMA), polyethylene glycol average Mn 500 (OEGMA500), 2-ethylhexyl methacrylate (EHMA) and 3-sulfopropyl methacrylate potassium salt (SPMA) of compositions: A. 10:25:60:5 B. 30:25:40:5 C. 60:25:10:5

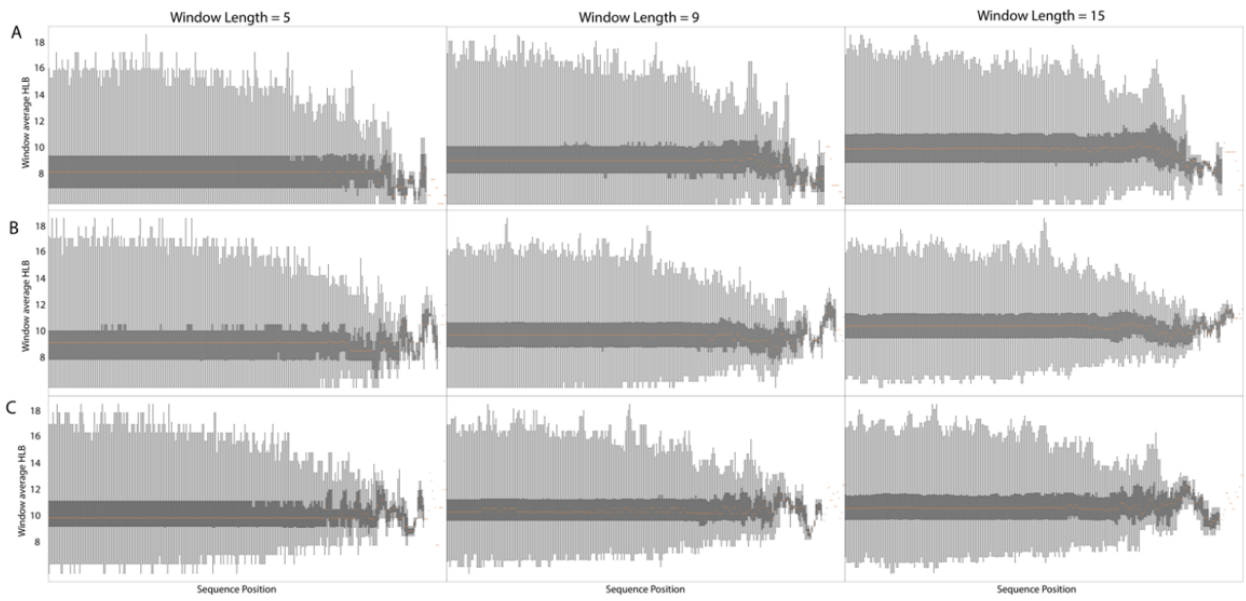


Figure S4.2 | Example sequence-level sliding window analysis of RHP sequences. Sliding window of 3 varying window lengths applied to batches of simulated 4 monomer RHP of methyl methacrylate (MMA), polyethylene glycol average Mn 500 (OEGMA500), 2-ethylhexyl methacrylate (EHMA) and 3-sulfopropyl methacrylate potassium salt (SPMA) of compositions: **A.** 10:25:60:5 **B.** 30:25:40:5 **C.** 60:25:10:5

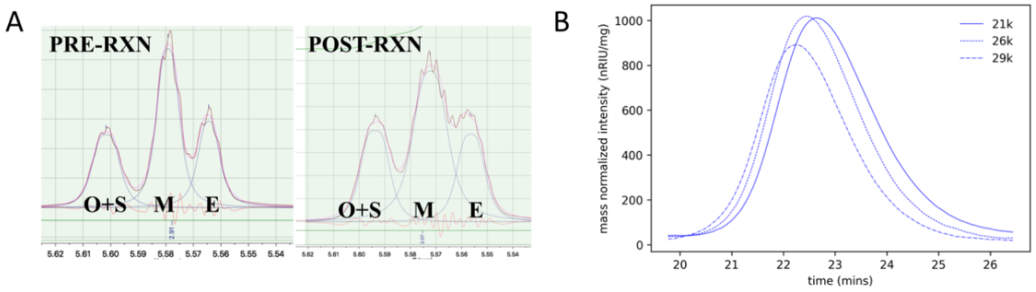


Figure S4.3 | Example of routine experimental characterization of RHPs designed by RHPapp. For a 4 monomer RHP of 50% methyl methacrylate (M), 25% polyethylene glycol average Mn 500 (O), 20% 2-ethylhexyl methacrylate (E) and 5% 3-sulfopropyl methacrylate potassium salt (S): **A.** Proton nuclear magnetic resonance spectroscopy (H-NMR) of the methacrylate monomer vinyl peaks pre- and post- reaction samples that can be integrated with respect to an internal standard to verify global conversion and in some cases monomer specific conversions to confirm an absence of compositional drift. **B.** Gel permeation chromatography (GPC) spectra of 3 RHPs of number average molecular weights (Mw) of 21, 26 and 29 kDa that confirm bounded polydispersity and achieved Mw targeted by RHPapp with respect to a polymethyl methacrylate or polyethylene glycol standard.

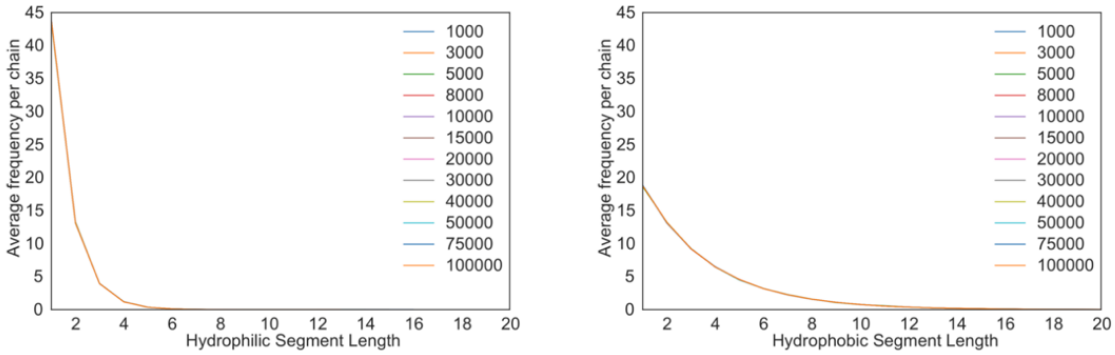


Figure S4.4 | Varying the number of chains (NC) simulated. Batch-level hydrophilic and hydrophobic segment distribution heterogeneities for a simulated 4 monomer RHPs of 50% methyl methacrylate (MMA), 25% polyethylene glycol average Mn 500 (OEGMA), 20% 2-ethylhexyl methacrylate (EHMA) and 5% 3-sulfopropyl methacrylate potassium salt (SPMA), of varying NC.

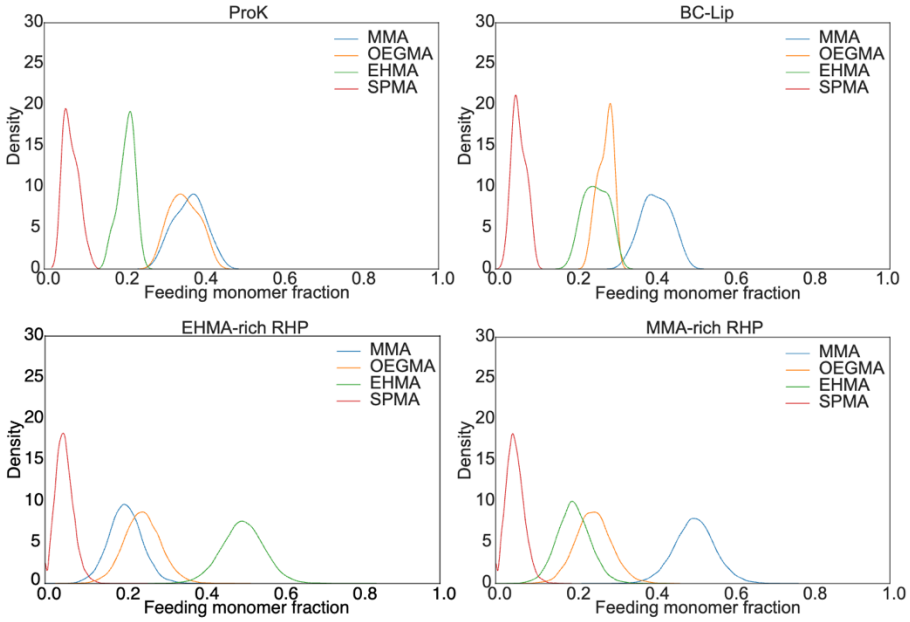


Figure S4.5 | Sequence level monomer distributions of Proteinase K (ProK) and Lipase from Burkholderia cepacia (BC- Lip) convolved into RHP sequence space, segmented to form a batch, and 2 batches of simulated 4 monomer RHPs of methyl methacrylate (MMA), polyethylene glycol average Mn 500 (OEGMA), 2-ethylhexyl methacrylate (EHMA) and 3-sulfopropyl methacrylate potassium salt (SPMA) of monomer feeding ratios 50:25:20:5 (MMA-rich) and 20:25:50:5 (EHMA-rich). All sequences are of average degree of polymerization = 100.

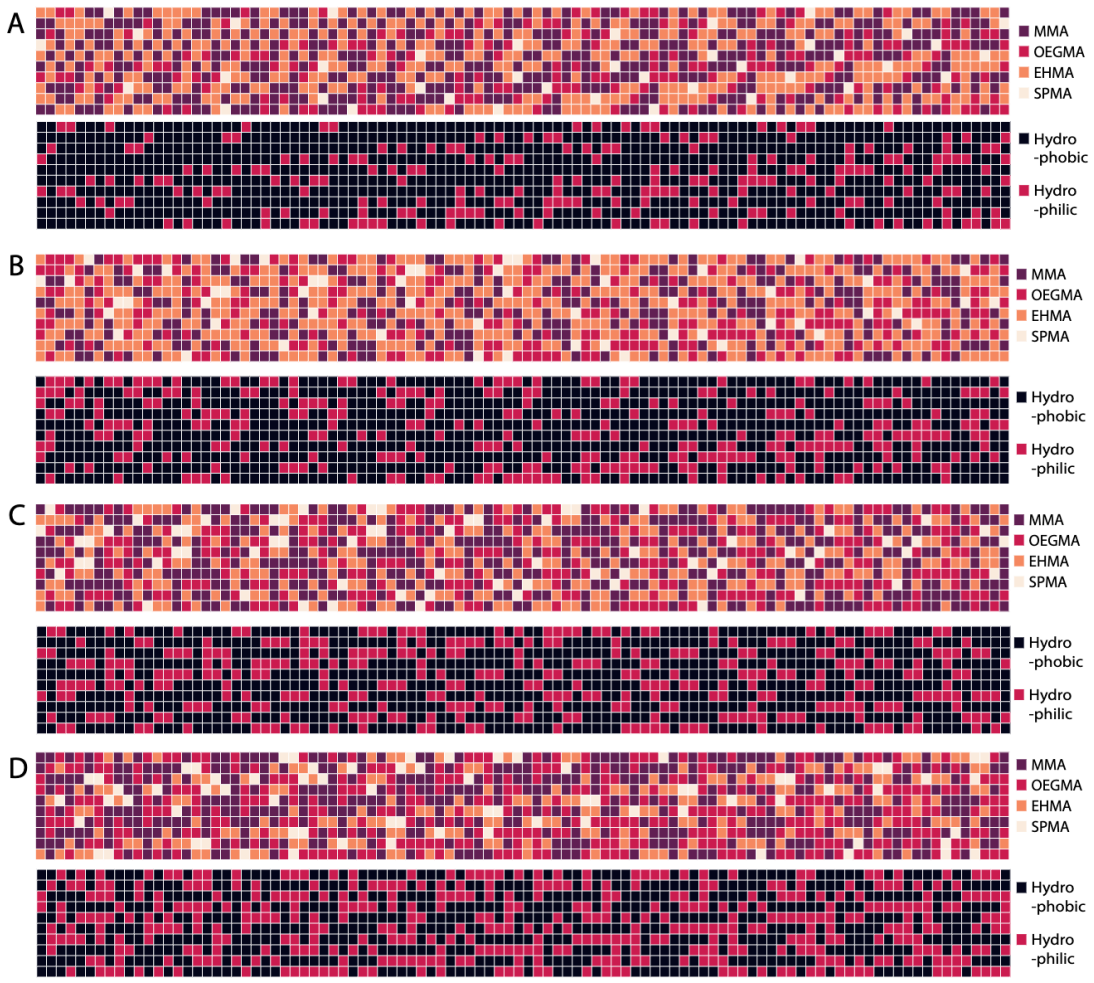


Figure S4.6 | Protein sequences in RHP sequence space. The following proteins are segmented and translated to RHP sequence space, then binarized to hydrophobic and hydrophilic units: (A) AquaporinZ (B) PepTSo (C) BC-Lipase (D) Proteinase K

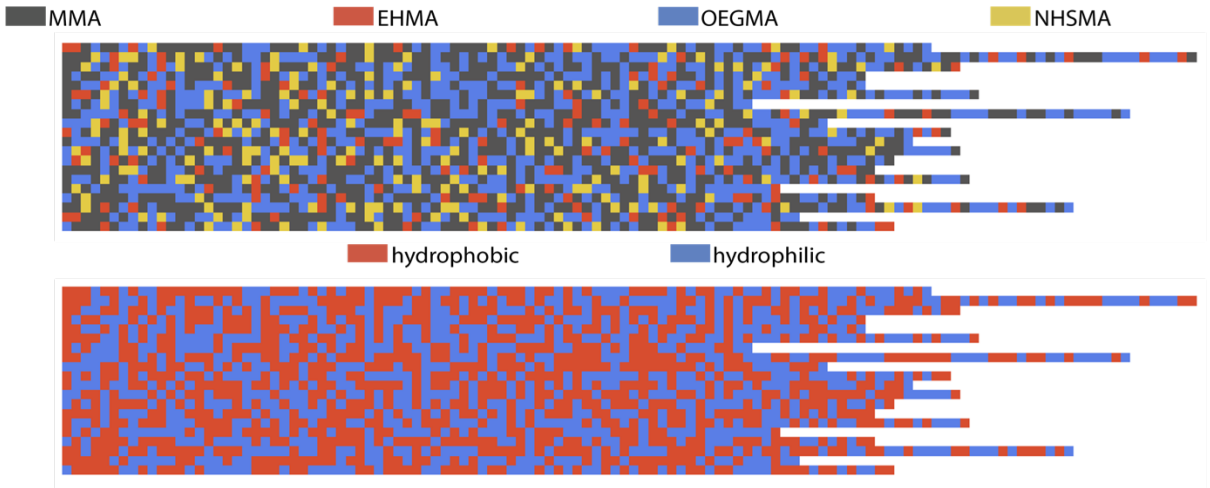


Figure S5.1 | Random sequences sampled from simulated **RHP2** using measured reactivity ratios and targeted monomer conversion. The RHP sequences shown at the bottom were binarized into hydrophobic and hydrophilic units to display the distributions of hydrophobic/hydrophilic segments.

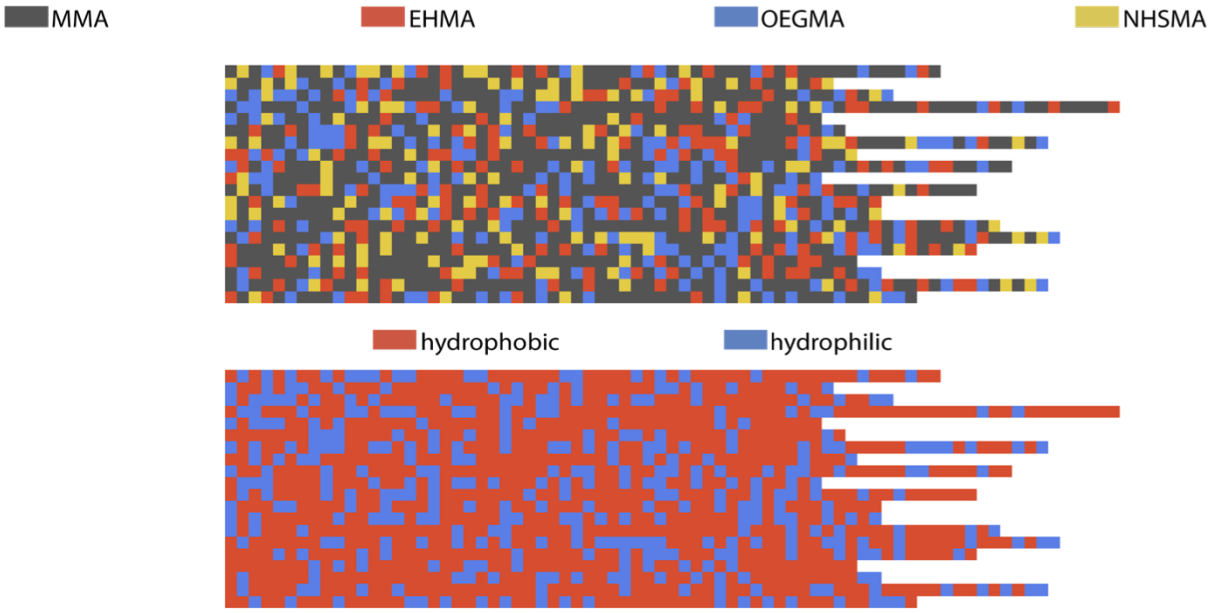


Figure S5.2 | Random sequences sampled from simulated **RHP3** using measured reactivity ratios and targeted monomer conversion. The RHP sequences shown at the bottom were binarized into hydrophobic and hydrophilic units to display the distributions of hydrophobic/hydrophilic segments.

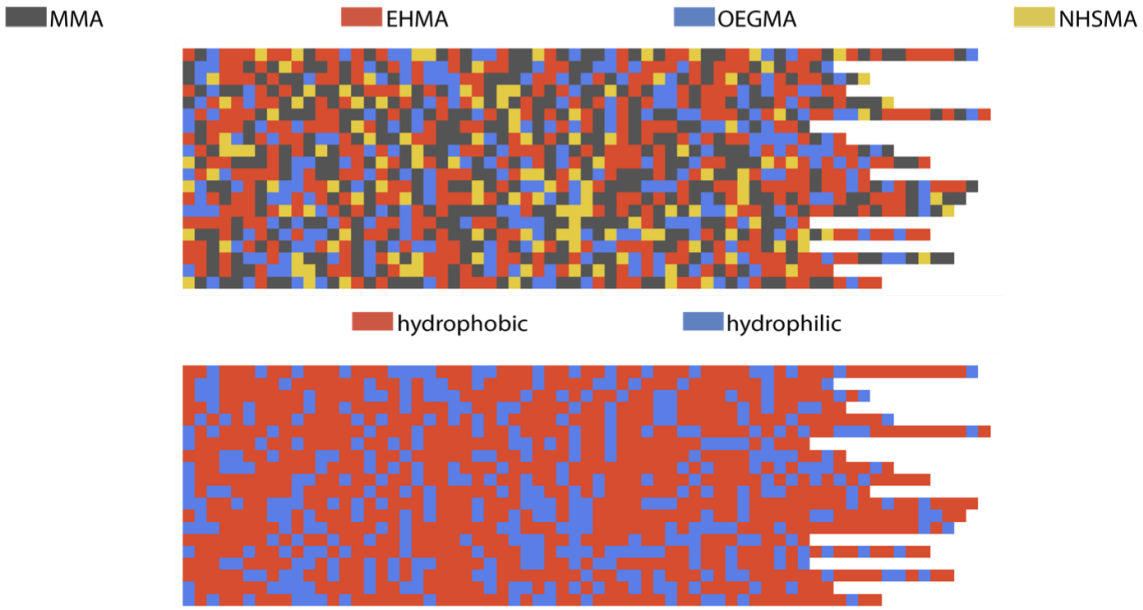


Figure S5.3 | Random sequences sampled from simulated **RHP4** using measured reactivity ratios and targeted monomer conversion. The RHP sequences shown at the bottom were binarized into hydrophobic and hydrophilic units to display the distributions of hydrophobic/hydrophilic segments.

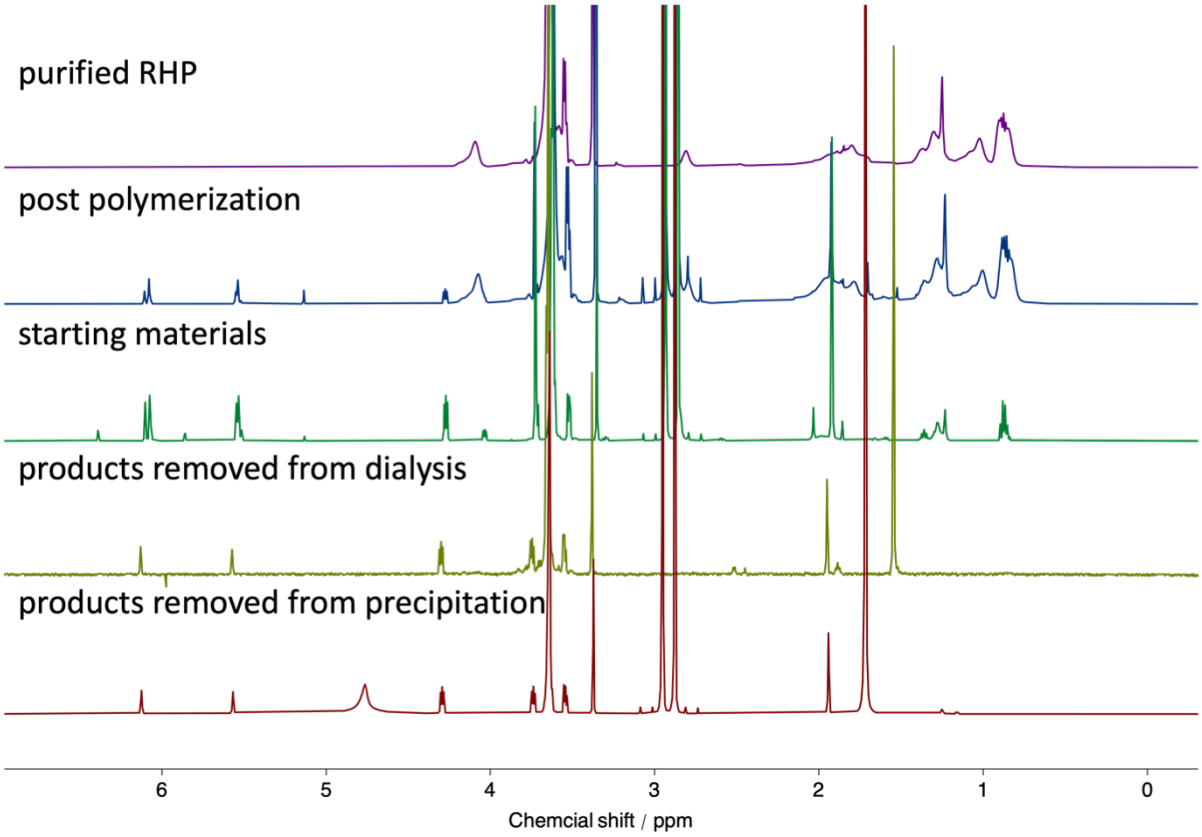


Figure S5.4 | Comparison of ^{1}H NMR spectrum of **RHP-2**, crude products, starting materials, and products removed during purification (500 MHz, $CDCl_3$).

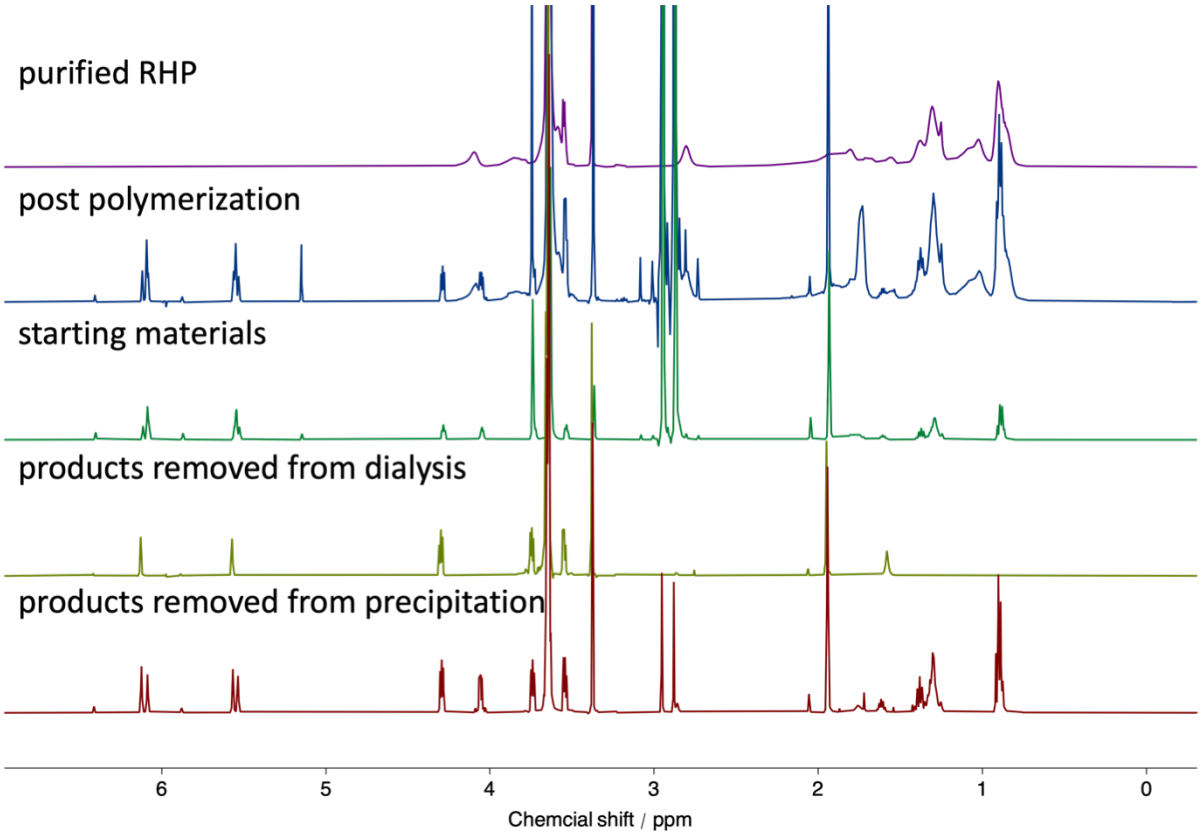


Figure S5.5 | Comparison of ^{1}H NMR spectrum of **RHP-3**, crude products, starting materials, and products removed during purification (500 MHz, $CDCl_3$).

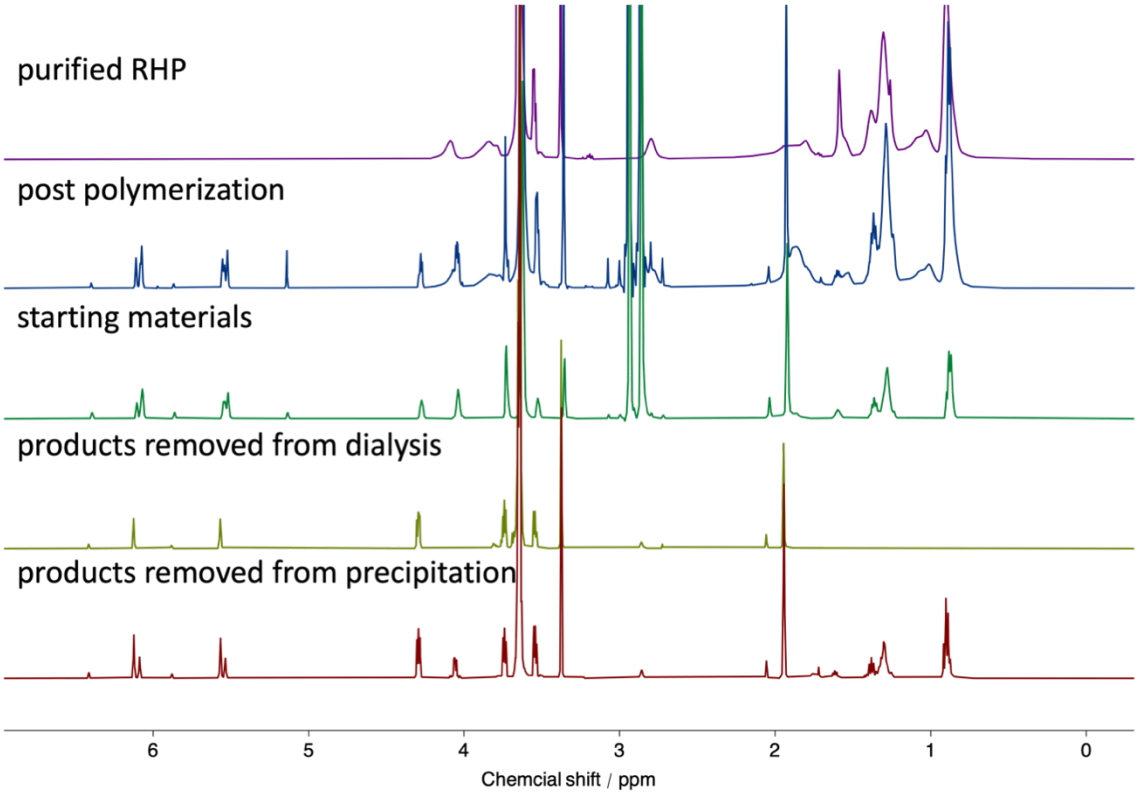


Figure S5.6 | Comparison of 1H NMR spectrum of **RHP-4**, crude products, starting materials, and products removed during purification (500 MHz, CDCl3).

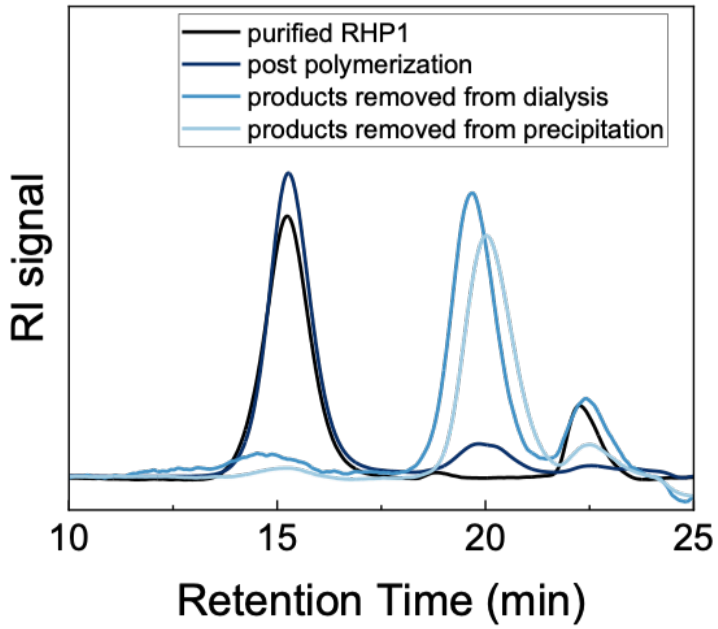


Figure S5.7 | GPC traces of **RHP-1**, crude products, starting materials, and products removed during purification.

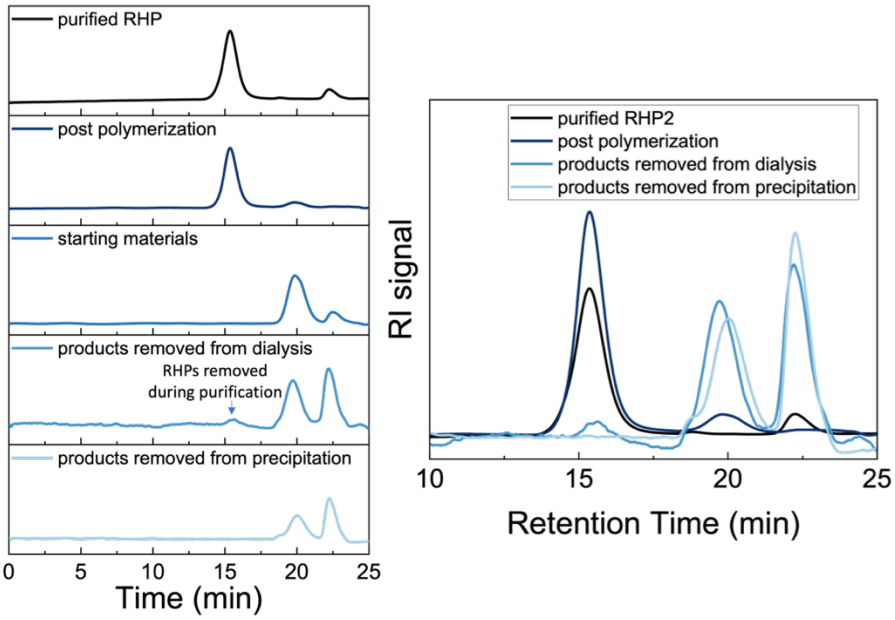


Figure S5.8 | GPC traces of **RHP-2**, crude products, starting materials, and products removed during purification.

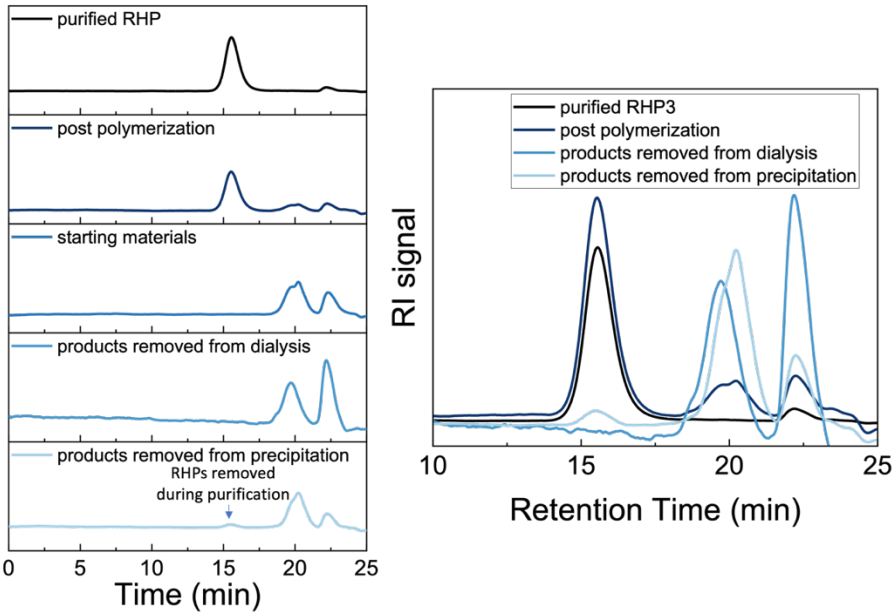


Figure S5.9 | GPC traces of **RHP-3**, crude products, starting materials, and products removed during purification.

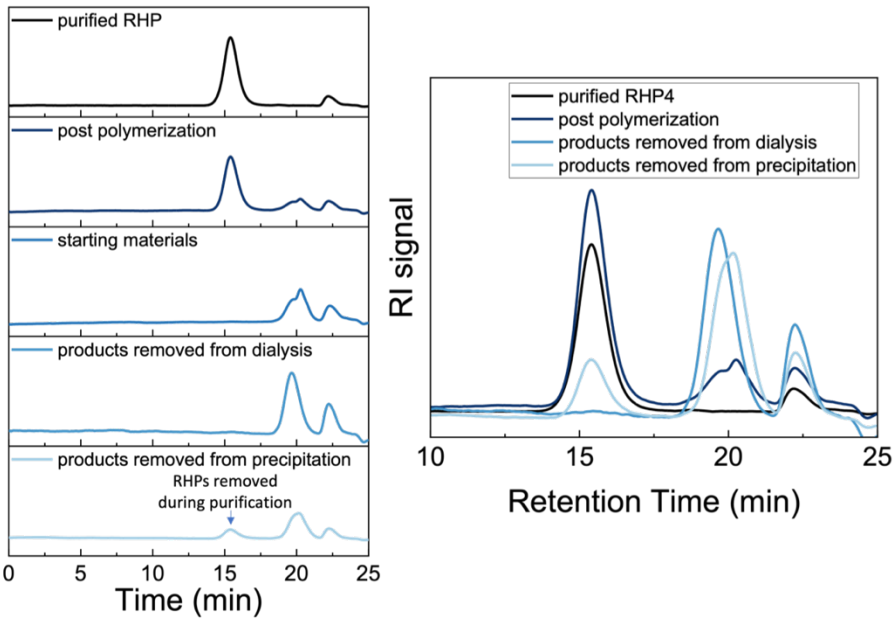


Figure S5.10 | GPC traces of **RHP-4**, crude products, starting materials, and products removed during purification.

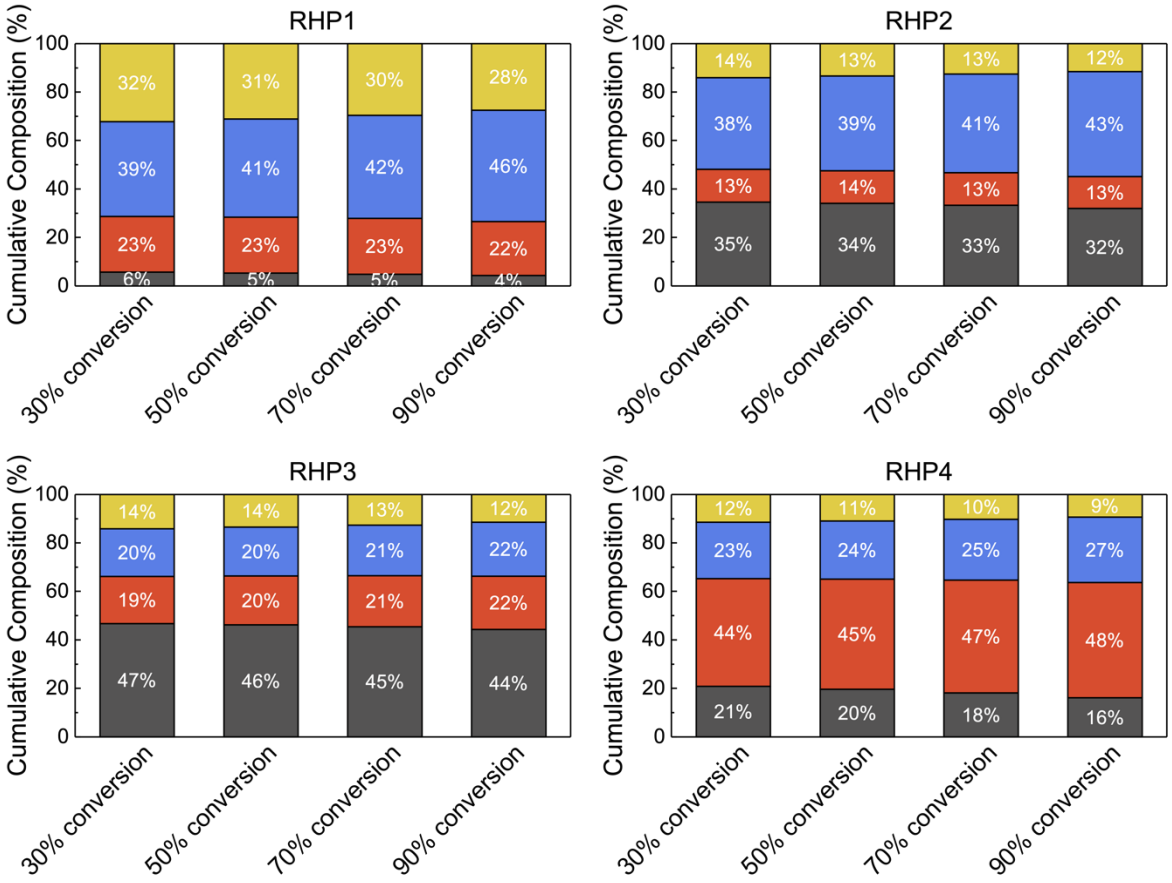


Figure S5.11 | Simulated composition drift during RAFT polymerization at different monomer conversions. Simulation results reflect the average composition of 50,000 simulated individual chains.

Table S5.1 | Summary of composition ratio of raw RHPs calculated from pre/post-polymerization aliquot NMR. (n = 3)[21,22]

# raw RHPs	MMA (%)	EHMA (%)	OEGMA (%)	NHSMA (%)
RHP1	3.3 ± 0.4	20.2 ± 2.4	48.1 ± 2.5	28.3 ± 1.9
RHP2	30.0 ± 2.0	11.8 ± 2.1	46.1 ± 1.6	12.2 ± 0.8
RHP3	41.1 ± 5.3	22.4 ± 2.5	22.0 ± 1.6	13.9 ± 1.2
RHP4	20.1 ± 2.5	41.6 ± 1.3	27.2 ± 1.7	11.1 ± 1.3

Table S5.2 | Summary of composition ratio of raw RHPs simulated using *RHPapp*.

# simulated RHPs	MMA (%)	EHMA (%)	OEGMA (%)	NHSMA (%)
RHP1	4.3	22.3	45.7	27.7
RHP2	32.3	12.6	43.3	11.8
RHP3	45.3	20.8	20.6	13.3
RHP4	19.2	45.8	24.1	10.9

Table S5.3 | Summary of composition ratio of purified RHPs determined by quantitative ^{13}C NMR. (n = 3)

# purified RHPs	MMA (%)	EHMA (%)	OEGMA (%)	NHSMA (%)
RHP1	11.0 ± 1.9	17.2 ± 1.2	47.5 ± 1.2	24.3 ± 1.6
RHP2	31.5 ± 1.0	9.6 ± 6.4	48.1 ± 0.9	11.2 ± 4.7
RHP3	37.6 ± 4.8	23.3 ± 1.3	24.2 ± 1.9	15.0 ± 1.7
RHP4	17.2 ± 0.7	45.7 ± 5.9	23.6 ± 1.0	13.4 ± 7.4

Table S5.4 | Summary of composition ratio of RHPs living chains calculated from RHPs depolymerization experiments. (n = 9)

# living chains in RHPs	MMA (%)	EHMA (%)	OEGMA (%)	NHSMA (%)
RHP1	4.1 ± 2.8	19.8 ± 3.3	45.5 ± 7.4	30.6 ± 11.0
RHP2	24.2 ± 6.1	14.8 ± 5.0	49.8 ± 4.9	11.2 ± 3.0
RHP3	27.4 ± 5.4	24.7 ± 2.9	28.3 ± 1.7	19.6 ± 1.8
RHP4	7.6 ± 1.5	43.0 ± 4.2	32.3 ± 5.1	17.1 ± 1.6

Table S5.5 | Poly(STY-co-GMA) system inputs. Total monomer (M), initiator (I), and chain transfer (RAFT) agent (T) concentrations were 8.1115, 0.008, and 0.0322 M respectively. The target average degree of polymerization (DP) was 252, resulting in total simulated M, I, and T of ~5e9, 5e6, and ~2e7. ~6000 chains were explicitly tracked out of ~2e7 total chains. Initiator fragmentation kinetic constants (k_d, *f*) were 0.533696 and 0.6.

	M1: STY	**M2: GMA**
Feeding concentration (M)	4.8669	3.2446
Feeding Ratio	0.6	0.4
k_p (e6)	4.68	12.1
k_a (e9)	0.36	1.74
k_f (e7)	3.63	1.74
k_t (e11)	11.60	7.02
STY Reactivity ratio ($r_{row/Column}$)	1	0.372
GMA Reactivity ratio ($r_{row/Column}$)	0.683	1

Table S5.6 | Poly(MMA-co-OEGMA-co-EHMA-co-NHSMA) system inputs. Total monomer (M), initiator (I), and chain transfer (RAFT) agent (T) concentrations were 2.2098, 0.005, and 0.0238 M respectively. The target average degree of polymerization (DP) was 92.8, resulting in total simulated M, I, and T of ~2.2e8, 5e5, and ~2.38e6. ~6000 chains were explicitly tracked out of ~2.38e6 total chains. Initiator fragmentation kinetic constants (k_d, *f*) were 0.533696 and 0.6.

	M1: MMA	**M2: OEGMA**	**M3: EHMA**	**M4: NHSMA**
Feeding concentration (M)	0.428	0.595	0.952	0.238
Feeding Ratio	0.1934	0.2689	0.4302	0.1075
k_p (e6)	4.79	6.78	5.83	8.64
k_a (e9)	3.6	3.6	3.6	3.6
k_f (e7)	2.43	2.43	2.43	2.43
k_t (e11)	3.2	0.16	0.64	0.16
MMA Reactivity ratio ($r_{row/Column}$)	1	0.9	1.3	0.5
OEGMA Reactivity ratio ($r_{row/Column}$)	0.7	1	0.4	0.6
EHMA Reactivity ratio ($r_{row/Column}$)	0.4	1.3	1	0.6
NHSMA Reactivity ratio ($r_{row/Column}$)	0.7	1.6	2.0	1

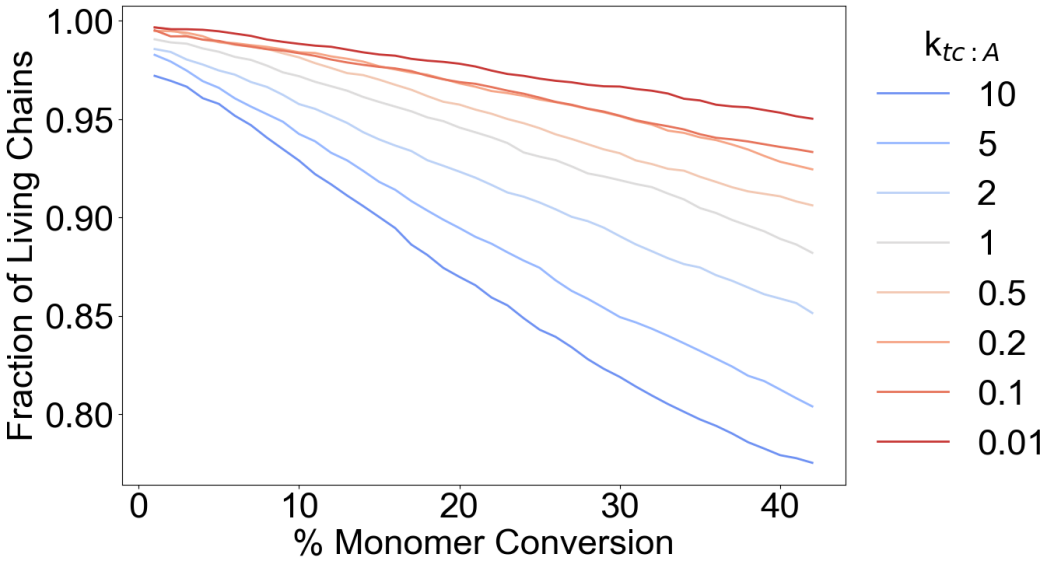


Figure S5.12 | S-G k_{tc} screening. Fraction of living chains with increasing % monomer conversions and varying k_{tc}

Ch 6. Deep learning driven heteropolymer design

Table S6.1 | Compositions of 2- and 4-monomer random heteropolymers (RHPs).

	RHP	MMA	OEGMA	EHMA	SPMA
2 Mon.	A	0	10	90	0
	B	0	30	70	0
	C	0	50	50	0
	D	0	70	30	0
	E	0	90	10	0
4 Mon.	1	70	25	0	5
	2	65	25	5	5
	3	60	25	10	5
	4	50	25	20	5
	5	40	25	30	5
	6	20	25	50	5
	7	0	25	70	5

Table S6.2 | Conversion dictionary mapping amino acids to RHP monomers (*n* = 4)

ID	Hydrophobicity	Amino acid	Monomer equivalent
1	Hydrophobic	Cys, Tyr, Ala, Thr, and Gly	MMA
2	Hydrophilic	Ser, Gln, His, Asn, and Pro	OEGMA
3	Very hydrophobic	Leu, Ile, Phe, Trp, Val, and Met	EHMA
4	Charged	Glu, Asp, Arg, and Lys	SPMA or DMAEMA

Table S6.3 | Conversion dictionary mapping amino acids to RHP monomers (*n* = 2)

ID	Hydrophobicity	Amino acid	Monomer equivalent
1	Hydrophobic	Leu, Ile, Phe, Trp, Val, Met, Cys, Tyr, Ala, Thr, and Gly	EHMA
2	Hydrophilic	Ser, Gln, His, Asn, Pro, Glu, Asp, Arg, and Lys	OEGMA

Table S6.4 | Conversion dictionary mapping amino acids to RHP monomers (*n* = 8)

ID	Hydrophobicity	Amino acid	Monomer equivalent
1	Very hydrophobic 2	Leu, Trp, Ille, Phe	EHMA
2	Very hydrophobic 1	Met and Val	HMA[a]
3	Hydrophobic 2	Tyr, Cys, Ala	BMA[b]
4	Hydrophobic 1	Thr, Gly	MMA
5	Hydrophilic 2	His, Ser and Gln	OEGMA (Mn=300 Da)
6	Hydrophilic 1	Asn and Pro	OEGMA (Mn=500 Da)
7	Positively Charged	Arg and Lys	DMAEMA
8	Negatively Charged	Glu and Asp	SPMA

[a] Hexyl methacrylate.

[b] Butyl methacrylate.

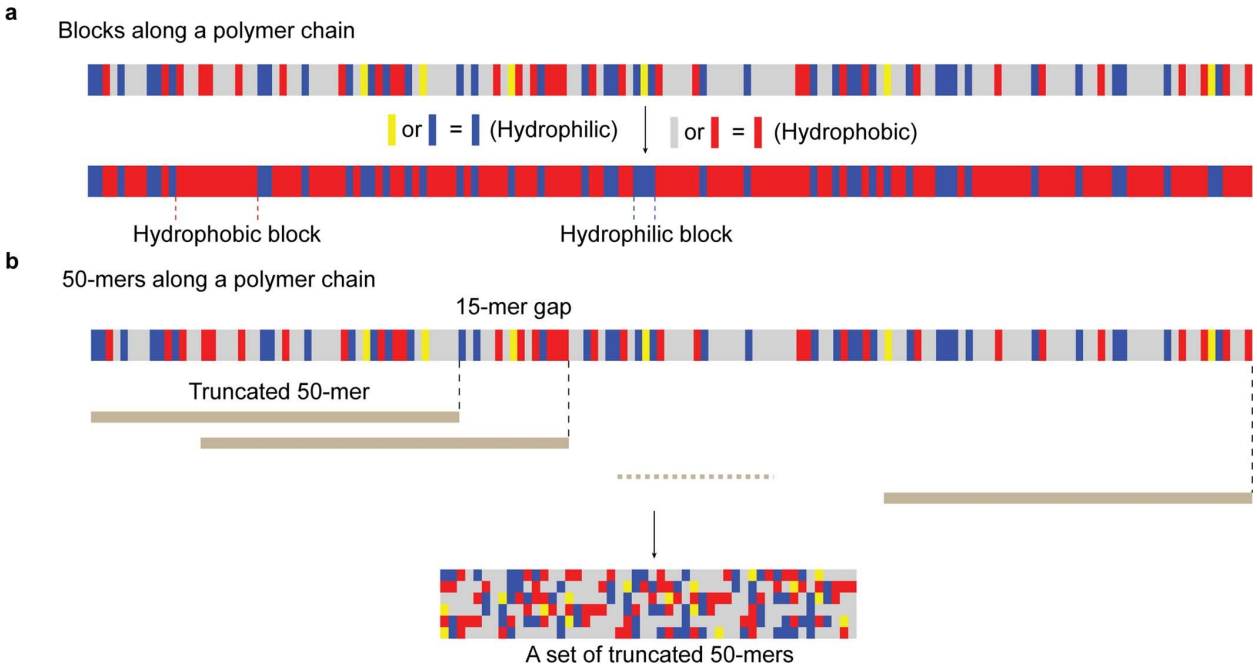


Figure S6.1 | A. Each monomer was reassigned to one of two pseudo-monomers (hydrophobic vs. hydrophilic) based on monomer's hydrophobicity. A block comprises consecutive monomers of a single type. **B.** Polymer chain truncated into a set of 50-mers.


RHP
RHP7
RHP6
RHP5
RHP4
RHP3
RHP2
RHP1
RHP
RHP7
RHP6
RHP5
RHP4
RHP3
RHP2
RHP1
RHP
RHP7
RHP6
RHP5
RHP4
RHP3
RHP2
RHP1
RHP
RHP7
RHP6
RHP5
RHP4
RHP3
RHP2
RHP1

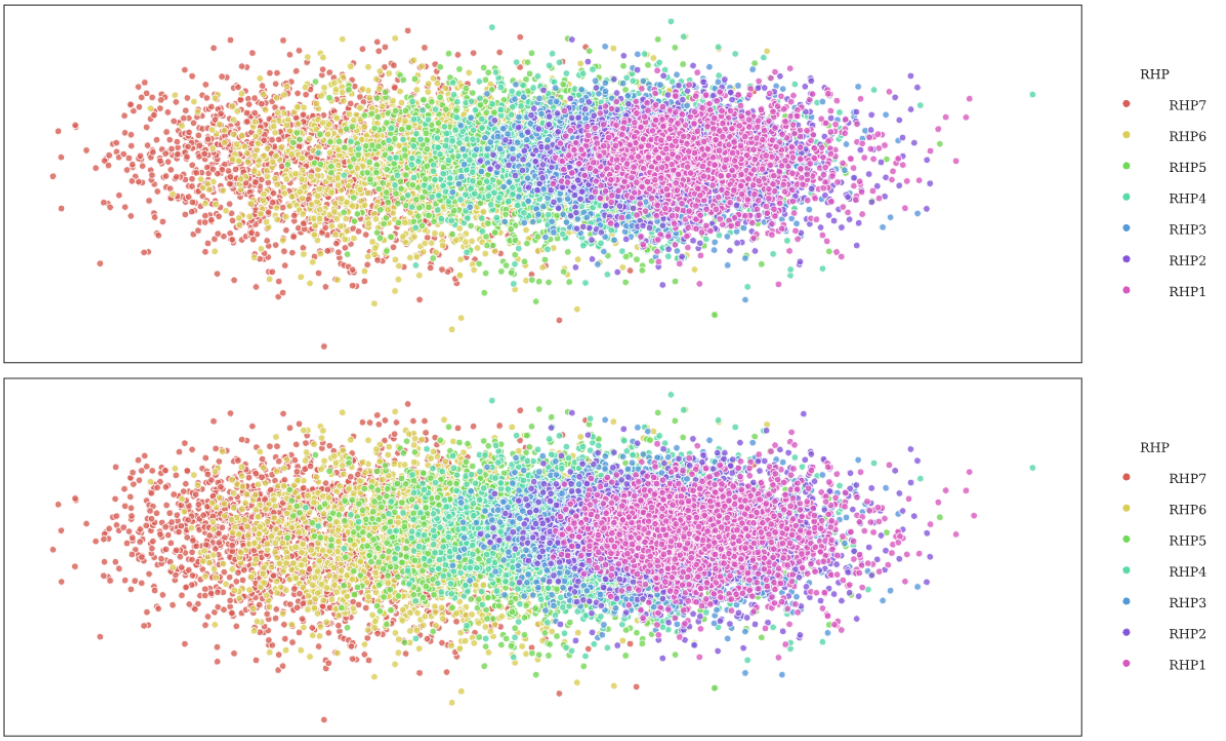


Figure S6.2 | The segment distribution of each RHP ensemble as a function of the number of segments (N) per RHP ensemble. From top to bottom, N is 50, 100, 500, 1000, 1500, and 2000, respectively. Visual inspection of each RHP ensemble shows that the segment distribution converges when N >= 1500 segments per RHP ensemble are analyzed.

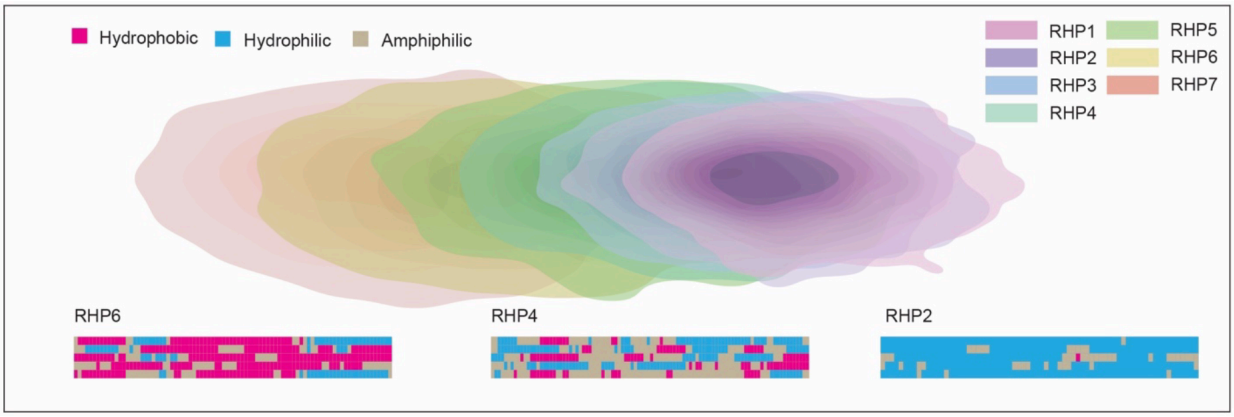


Figure S6.3 | The correlation between the relative abundance of different segments and their distribution in PCA space. The hydrophilic, amphiphilic, and hydrophobic segments are highlighted in blue, grey, and pink, respectively. The segments along an RHP chain are shown for three representative RHP ensembles with five individual chains.

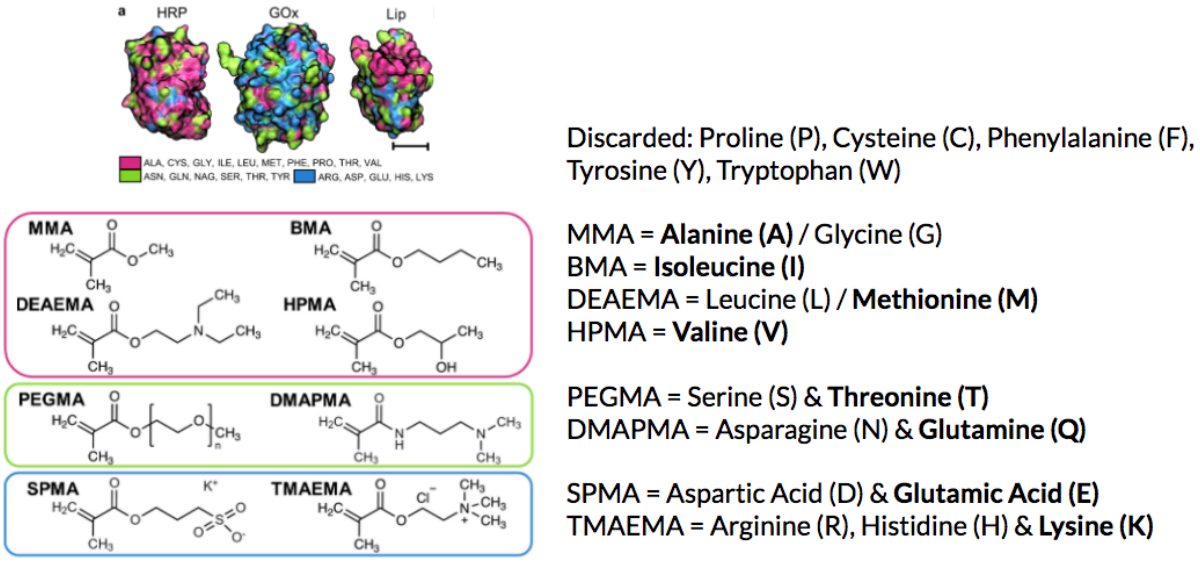


Figure S6.4 | Conversion between random heteropolymer (RHP) monomers and natural protein and random polypeptide (RPP) amino acids.

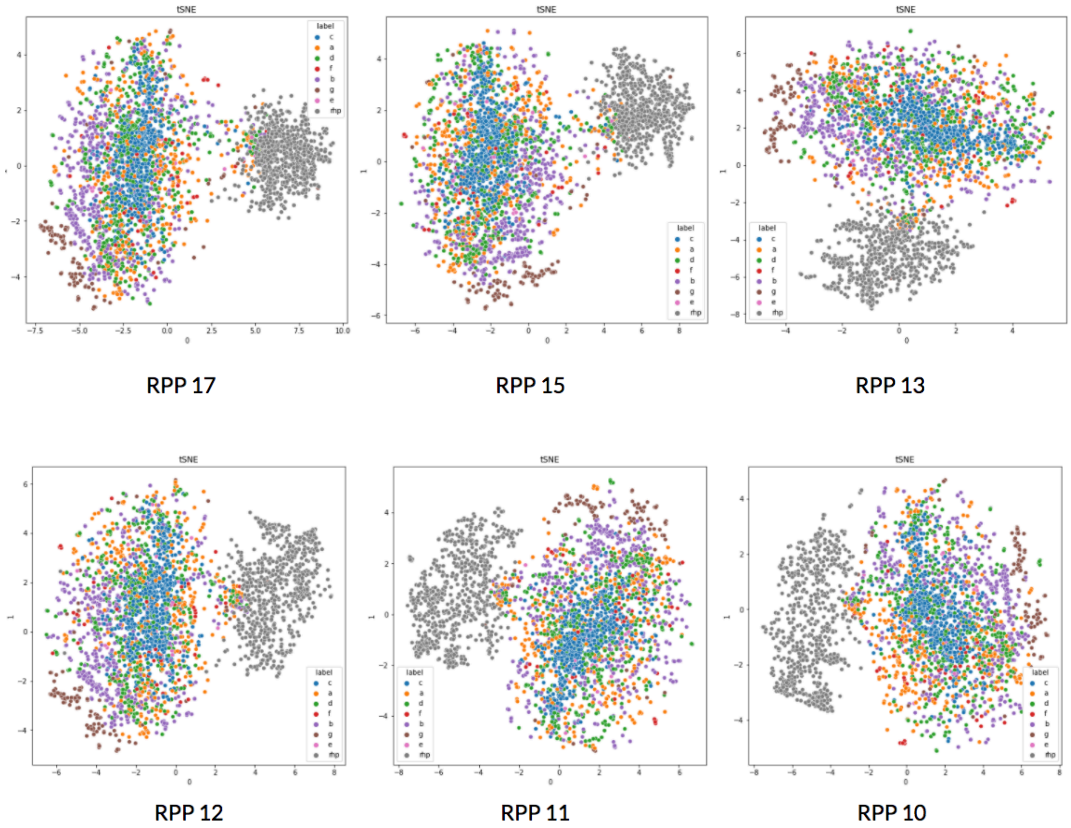


Figure S6.5 | RPPs sampled from natural distribution of decreasing # of amino acids



Figure S6.6 | RPP sampled from top N most abundant amino acids of natural protein distribution, where N = 6,8,10

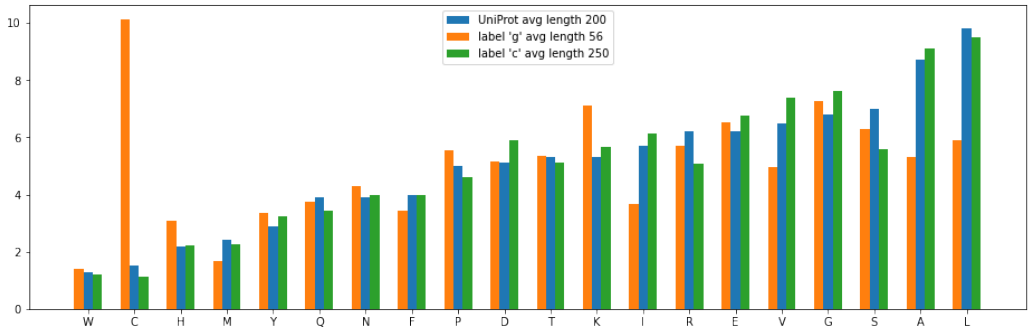


Figure S6.7 | Amino acid occurrence frequencies in natural protein distribution and in 2 structure subdivisions: 'g' = small proteins and 'c' = (a/b) proteins

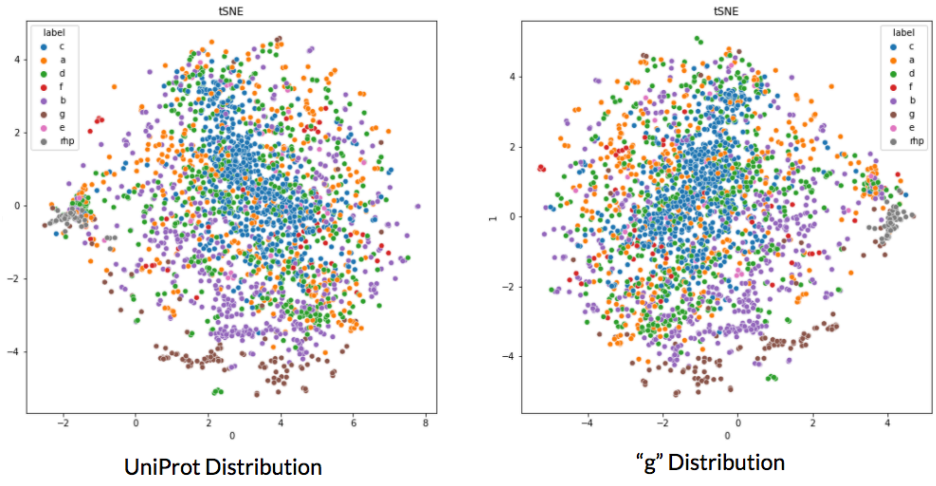


Figure S6.8 | Showing there is negligible difference when sampling from natural protein distribution (length-200) and small protein distribution (length-56)

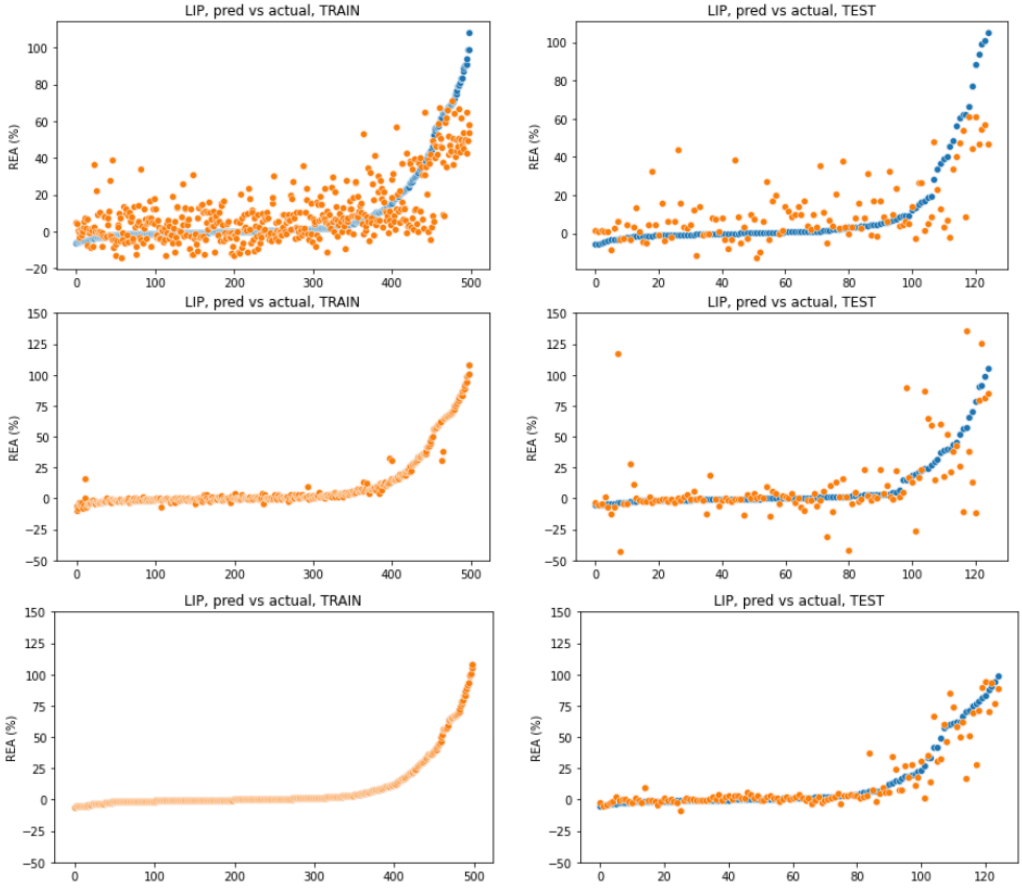


Figure S6.9 | Representative prediction vs actual labels for Lip model: (top) Batch-Ave + RidgeCV (mid) Batch-Ave + GPR, note that y-axis has been truncated to ignore extreme outliers that exist at >600 REA % (bot) RPP-seq encodings + GPR

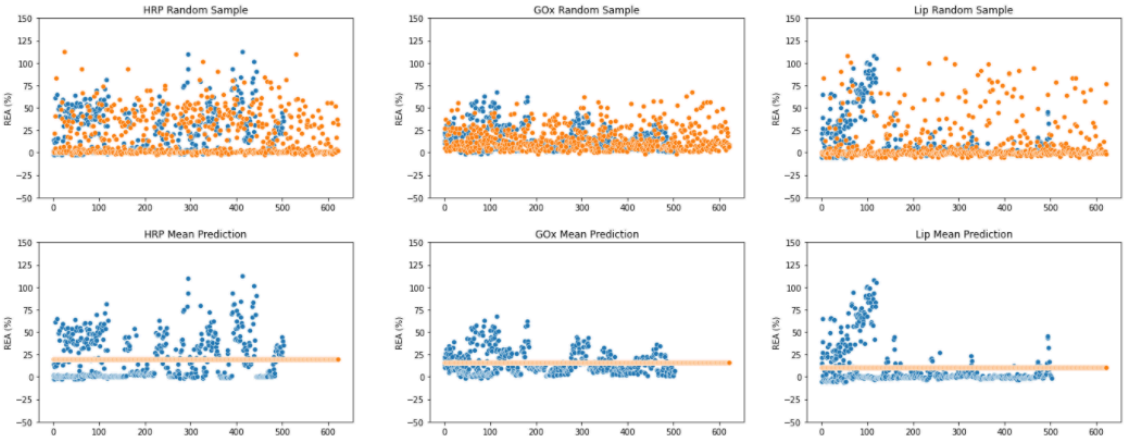
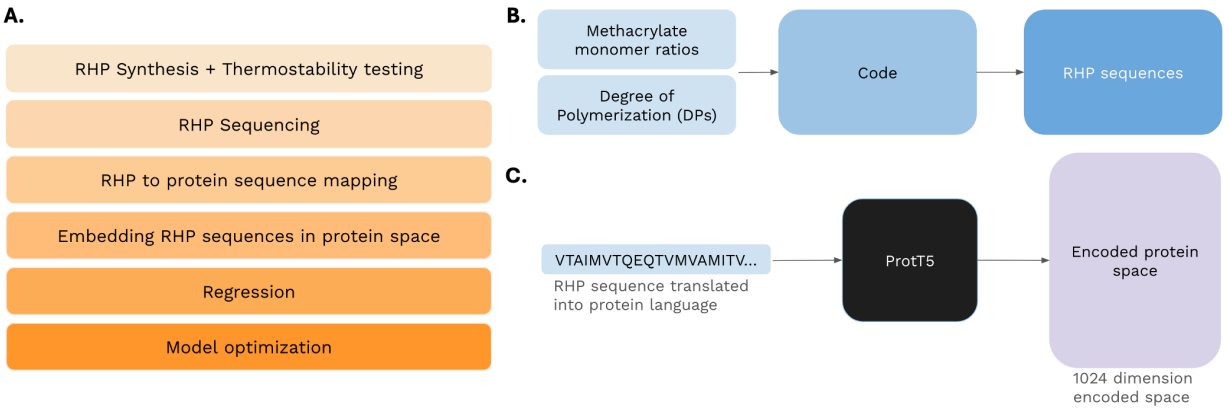


Figure S6.10 | Random sampling (top) and mean prediction (bot) for 3 regression tasks



Scheme S6.1 | Flowcharts showing **A.** Optimized RHP design cycle for protein thermostabilization. **B.** RHP sequence generation. **C.** RHP sequence encoding using ProT5.