RESEARCH ARTICLE

Macromolecular Chemistry and Physics
www.mcp-journal.de

Precision Design of Sequence-Defined Polyurethanes: Exploring Controlled Folding Through Computational Design

*Svetlana Samokhvalova, Jean-François Lutz, and Ivan Coluzza**

This study presents the exploration of sequence-defined polyurethanes (PUs) as a new class of heteropolymers capable of precise conformational control. Utilizing molecular dynamics simulations, the folding behavior of polyurethane chains is investigated of varying lengths (11, 20, and 50 monomers) in both vacuum and aqueous environments. The simulations reveal that the heterogeneous chains systematically refold to approach the designed target structures better than non-designed chains or chains with artificially disrupted hydrogen-bond networks. The subsequent synthesis of an optimized 11-mer sequence (P1) is achieved through solid-phase chemistry, with thorough characterization via NMR, MS, and SEC confirming the accuracy of the predicted sequence and its controlled chain length. Solubility tests showed favorable results across multiple solvents, highlighting the versatility of the designed polymer. This research underscores the potential of sequence-defined polyurethanes to emulate the structural and functional attributes of biological macromolecules, opening new pathways for their application in catalysis, drug delivery, and advanced material design. The findings illustrate a promising direction for the development of synthetic polymers with tailored properties, emphasizing the transformative impact of sequence control in polymer chemistry.

1. Introduction

Natural materials, exemplified by proteins, embody sustainability and recyclability due to their modular nature, which allows biological systems to break them down and repurpose their components. Proteins stand out in nature as nano-machines, possessing a unique modularity encoded by a simple string of amino acids, which underlies their diverse functions and autonomous behavior. The exploration of protein design and folding, alongside the development of protein-mimicking polymers known as foldamers, represents a promising avenue for achieving materials with unparalleled versatility and sustainability.

Foldamers are a class of synthetic molecules that mimic the structural features of biopolymers, such as proteins and nucleic acids, by folding into well-defined 3D shapes.[1–7] Unlike natural polymers, foldamers can be composed of non-natural monomers, offering a broad palette for designing structures with novel properties and functionalities.[8,9] This versatility enables researchers to explore new aspects of molecular recognition, catalysis, and self-assembly, extending the principles of biomolecular architecture into the realm of synthetic chemistry. The ability of foldamers to adopt specific shapes and bind to biological targets with high affinity makes them promising candidates for therapeutic development, material science, and beyond, positioning them at the forefront of molecular design and engineering.

In the past five years, foldamers research has seen significant advancements, pushing the boundaries of chemical and biomedical sciences. Notably, the exploration of self-assembly processes in foldamers has unlocked new possibilities for creating complex structures, as detailed in,[10] highlighting the endless self-assembly modes of foldamers for practical applications. Additionally, the development of helical anion foldamers, discussed by,[7] represents a leap in understanding how single- or multistrand helical complexes can be utilized, particularly in anion recognition and transport. The field of catalytic foldamers, explored by,[11] has also made substantial progress, demonstrating how foldamers can act as catalysts for chemical reactions, opening up new avenues for synthetic chemistry. Moreover, the design of unnatural helical peptidic foldamers as protein segment mimics, as reviewed by,[12] showcases the potential of foldamers to mimic complex biological structures, offering promising strategies for therapeutic and material science applications. These achievements underscore the versatile and transformative nature of foldamers, making them a focal point for future innovations in multiple scientific domains.

S. Samokhvalova
Université de Strasbourg
CNRS
ISIS
8 allée Gaspard Monge, Strasbourg 67000, France
J.-F. Lutz, I. Coluzza
Center for Theoretical Biological Physics
Rice University
Houston, TX 6100, USA
E-mail: ic33@rice.edu

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/macp.202400223

DOI: 10.1002/macp.202400223

Besides classical synthetic foldamers, significant research has recently been dedicated to so-called single-chain nanoparticles, which are individual polymer globules. Single-chain nanoparticles are usually obtained by intramolecular cross-linking of synthetic polymer chains using covalent or non-covalent interactions.[6,13–15] The macromolecules used for the formation of such single-chain objects are not necessarily molecularly uniform as proteins are (i.e. homochiral and sequence-defined) but may exhibit less controlled molecular structures (e.g. polydisperse, atactic). Therefore, in such approaches, the single chains are not precisely folded like a protein or in synthetic foldamer, but are somehow compacted into a structured shape, which can eventually be compartmentalized and may exhibit a function such as catalytic activity.

To date, most of the research on synthetic foldamers and single-chain nanoparticles has been conducted using homopolymers.[4] However, as learned from protein folding, primary structure (i.e. controlled monomer sequence in a heteropolymer) is a key parameter for controlling folding. Over the last fifteen years, significant research has been dedicated to the synthesis of man-made polymers with controlled monomer sequences.[9] Thus, new options have arisen for the design of foldamers or single-chain nanoparticles. For instance, hetero-foldamers promise to push the field beyond repetitive structures controlled exclusively by the backbone chemical bond constraints. Achievements in backbone heterogeneity span various methods, including mixed α-/β-/γ-/δ-peptides and peptide-peptoid systems,[16–18] which allow for the fine-tuning of foldamer conformations. Innovations such as mixed aliphatic-aromatic hybrid foldamers and the incorporation of tricyclic and spirocyclic diproline mimetics into collagen model peptide backbones have shown to maintain or alternative structures effectively. There seems to be a difficulty in extending the design of hetero-foldamers beyond the backbone architectures that closely resemble natural peptides. One possibility is that we still have not identified the set of properties that makes proteins highly designable.

Here we present a study focusing on the design of polyurethanes, polymers that were never used in the past for the design of hetero-foldamers, with the objective of demonstrating a theory about heteropolymer design that we recently introduced and only proved theoretically using computational methods.[19–22] From our previous research on the design of generalized heteropolymers we isolated directional interactions as key for the successful identification of monomer sequences capable of encoding structure and function. In particular, our study identifies polyurea and polyurethane (PU) among others as potential candidates for designable heteropolymers. Polyurea has been already explored in the Foldamer research [23]however it has focused on the properties of the backbone hydrogen bonds to form specific secondary structures and to the best of our knowledge, no hetero-foldamers have been proposed using polyurea. In recent years, interesting strategies have been reported for preparing sequenced-defined polyurethanes both for oligonucleotides[24–26] and long chains.[27–29] These polymers have been used in applications such as data storage and anti-counterfeiting technologies.[30–33] Yet, little is known about the potential folding of this new class of macromolecules. Similarly, to proteins, polyurethanes contain H-bond donor and acceptor sites and can therefore undergo intra- or intermolecular associations (e.g. crystallization).[34–36] Thus, in the present work, the single-chain structuration of sequence-defined polyurethanes was investigated as a first step toward establishing the principles needed to achieve foldamer-like behavior in these systems, although the current findings do not yet fully demonstrate the design of true foldamers with polyurethane.

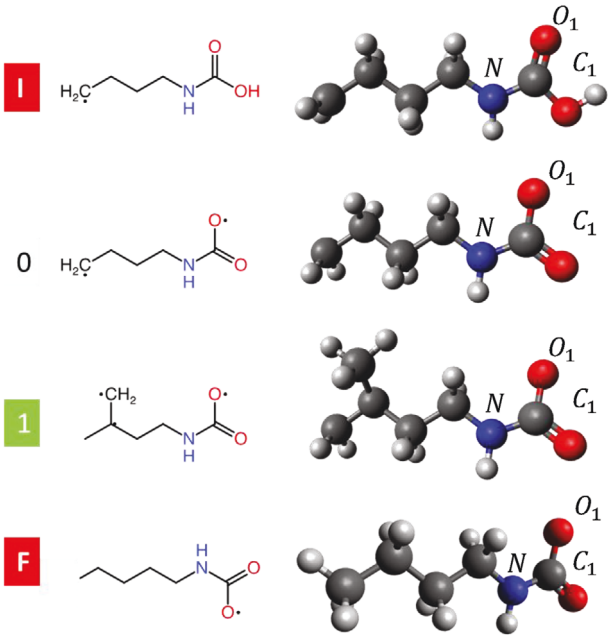

Figure 1. Molecular formulas of the monomers used to construct the hetero polyurethane chains. With I and F we refer to the initial and final monomers of every chain. While the 0 and 1 monomers are the ones that define the sequence. We highlight the carbon C_1, which we used as a reference to assess the impact of the hydrogen bond network on the chain's conformational space.

2. Experimental Section

2.1. Polyurethane Modeling

The polyurethane studied herein was selected from previous experimental studies (Figure S1, Supporting Information). These heteropolymers contain two different repeat units, namely butyl carbamate and methyl-butyl carbamate. The latter one contains an asymmetric carbon and a racemic reagent is usually used for polymer synthesis. Consequently, the formed polymers are not homochiral. The polymers also contain a caproic acid end-group (**I**) that is a residual linker from solid-phase synthesis and a terminal end-group at the other extremity (**F**). For the generation of the initial topology in this study, polyurethane (PU) was meticulously constructed into chains by sequentially assembling monomeric units (UR) according to a predefined pattern. Each UR monomer (**I**, **0**, **1**, and **F** see **Figure 1**) was carefully annotated with atomic details and bonding configurations sourced from the Avogadro minimization, ensuring the accuracy of the molecular structure (See Supporting Information). This assembly process commenced with an **I** monomer. Following the initial I unit, it added either 0 or 1 monomers in accordance with the designated sequence, adding complexity and diversity to the polymer chain. The final component in the chain construction was the **F** monomer, denoting the terminal residue of urethane, which concluded the polymer se-

quence. This systematic assembly of monomers facilitated the creation of PU chains of varying lengths, with detailed attention to the chemical and physical properties inherent to each monomer type. Through this methodological approach, a comprehensive topology was generated that accurately reflects the intricate structure of polyurethane chains. This topology served as the foundation for subsequent molecular dynamics simulations.

2.2. Design Process

Conformation-dependent sequence design was a method for designing the order (sequence) of building blocks (monomers) in a polymer chain to control the overall 3D shape (conformation) that the chain adopts. This approach was inspired by how proteins fold into specific shapes based on the sequence of their amino acids. There were two main applications of conformation-dependent sequence design: i) Protein-like copolymers: by designing the sequence of a copolymer (a polymer chain made from two different types of monomers) to mimic the sequence of a protein, researchers can create copolymers that fold into complex shapes with specific functions. ii) Adsorption-tuned copolymers: By designing the sequence of a copolymer to interact favorably with a surface, researchers can create copolymers that can self-assemble into ordered structures on surfaces.

2.3. Rationale and Process Overview

The design process involves systematically alternating between two key stages: 1) sequence design, where the order of monomers is adjusted based on their role in the target conformation, and 2) folding simulations, where the designed sequences are tested and equilibrated in a simulated environment to evaluate their conformational stability. This iterative process ensures that the final sequence is optimized for the desired folded structure.

2.4. Sequence Design Stage

The initial sequence is typically generated as a homopolymer of one type of monomer (e.g., hydrophobic or hydrophilic). During the first folding simulation, the chain collapses into a globular structure driven primarily by hydrophobic interactions. The positions of monomers in this folded structure were then analyzed, and surface-exposed monomers were replaced with hydrophilic units while buried monomers remained hydrophobic. This targeted substitution aims to promote stable interactions that reinforce the globular conformation in subsequent simulations.

The iterative nature of this process was critical: after each round of sequence adjustments, the chain was refolded, and the positions of monomers were reassessed. This iterative folding-design process continues until the sequence converges, meaning further adjustments did not yield significant changes in the folded structure. This convergence indicated that the sequence had reached a state of equilibrium, and the designed polymer was now stably folded into the target conformation.

2.5. Folding Stage

Folding simulations were conducted using molecular dynamics (MD) in both vacuum and explicit solvent conditions. The initial simulations start with a stretched conformation of the polymer, which was then allowed to collapse naturally based on intramolecular forces. This was followed by multiple rounds of equilibration to ensure that the final structure reflects the true thermodynamic minimum. Each sequence is tested across varying environmental conditions to assess the robustness of the designed fold, including different temperatures and the introduction of chirality variations.

During this stage, the Helmholtz free energy F landscapes are mapped to evaluate the stability of the folded structure relative to other possible conformers. Specifically, the root-mean-square deviation (RMSD) from the target structure was used as a collective variable to track the folding progress. The free energy profiles *F[RMSD]* are carefully analyzed to ensure that the designed sequence consistently favors the desired conformation, even in the presence of potential folding intermediates or misfolded states.

2.6. Folding Simulations

In the molecular dynamic simulations, the GROMACS[37] package was utilized focusing on polyurethane chains with detailed settings inspired by the OPLS-AA[38] force field. This methodology involved simulating heterochain polyurethane chains of different lengths (11, 20, and 50 monomers) in both vacuum and aqueous environments.

Key simulation parameters were as follows: The initial preparation involved minimizing in vacuum starting from a stretched conformation, followed by running a 12 ns simulation in the NVT ensemble at a high temperature of 500 K, then another 12 ns at 300 K, and finally it solvates the system, include ions for neutrality and minimize again to avoid any clashes. At this point, 40 ns was run in explicit water using a 2 fs timestep to check the stability of the folded structure. The water simulations were executed using Metadynamics[39] implemented in PLUMED,[40] utilizing the RMSD from the structure obtained during the design simulations as a reference. Metadynamics was an enhanced sampling technique that drives the system out of local energy minima by adding a history-dependent bias potential to the free energy surface, allowing efficient exploration of the conformational space and overcoming energy barriers that could trap the system in metastable states. The NVT ensemble was chosen for equilibration, ensuring constant volume and temperature, with a specific emphasis on maintaining protein stability through position restraints. The use of the canonical (NVT) ensemble was particularly advantageous in this context as it allows precise control over temperature while preserving the structural integrity of the polymer during equilibration. The effectiveness of sampling in the NVT ensemble is reinforced by the application of the Nose–Hoover robust thermostats, which prevent temperature fluctuations that could destabilize the system. Additionally, position restraints were applied during the initial equilibration stages to avoid perturbations that could lead to sampling inefficiencies or artifacts. By ensuring that the system was well-thermalized before releasing restraints, reliable sampling of the conformational space was achieved, leading to meaningful insights into the equilibrium properties of the folded structures. Output for coordinates, velocities, and energies was configured at 1.0 ps intervals for comprehensive analysis. For bonded interactions, the LINCS algorithm[41] was applied to constrain the lengths of hydrogen-bonded atoms, ensuring stability during the simulation. For non-bonded interactions, the Verlet cutoff scheme was employed for efficient neighbor searching and calculating short-range forces. The Particle Mesh Ewald method[42] addressed long-range electrostatics, while the V-rescale thermostat maintained the system temperature at 300 K except for the short 11-mer where the folding was performed at 200 K. Notably, pressure coupling was omitted to concentrate on temperature effects, employing periodic boundary conditions to simulate an infinite system. Initial velocities were assigned from a Maxwell distribution at the target temperature, ensuring diverse molecular motion dynamics. This condensed simulation strategy enabled a focused investigation into the structural and dynamic properties of polyurethane chains, offering insights into their potential applications and behaviors in various environments. It was not looking for high accuracy in the force field selection but aimed to confirm with higher detail the predictions from the previously published coarse-grained model of patchy-polymers.[20] The OPLS-AA force field provides a reliable proof of concept for this purpose, consistent with the past work.

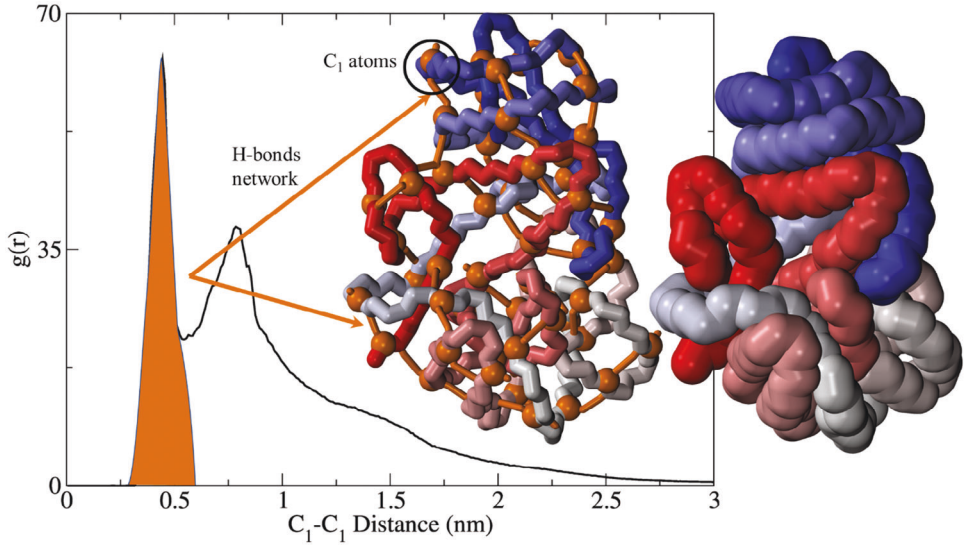

Figure 2. Typical collapsed conformation of a homopolymer chain of polyurethane made of 50 "1" monomers in a vacuum. In the plot, we show the radial distribution g(r) between the C_1 atoms of the chain. The large peak highlighted in orange corresponds to the contact stabilized by the hydrogen bonds. In the inset, we show the network of hydrogen bonds that is reminiscent of beta strands in proteins.

2.7. Polymer Synthesis

The model polymer **P1** was synthesized by stepwise iterative synthesis as previously reported using 4-amino-1-butanol (0) and 4-amino-2-methyl-1-butanol (1) as building blocks. The polymer was prepared on a hydroxy-functionalized Wang resin by repeating two reaction steps. In the first step, the resin (200 mg with the theoretical loading 0.96 mmol g^{-1}, 1 Eq.) was reacted for 1h with DSC (6 Eq.) in the presence of pyridine (12 Eq.) in 4 mL of anhydrous acetonitrile under microwave irradiation (Monowave 300, Anton Paar, 60 °C, 8 W). Afterward, the resin was transferred into a solid-phase extraction (SPE) tube and washed several times with dimethylformamide. In the second step, the activated resin was reacted for 20 min at RT with an excess amino-alcohol (10 Eq. of the corresponding monomer) in the presence of pyridine (20 Eq.) in 4 mL of anhydrous dimethylformamide. After the resin was washed with dimethylformamide and diethyl ether, it was transferred back to a microwave tube. The two steps were repeated ten times. Butan-1-amine was used as a chain-terminating agent in the last step. During this step, before washing with diethyl ether, the solid support was additionally washed with acetone to remove the traces of dimethylformamide. After washing with diethyl ether, the solid support was cleaved as such with 6 mL of trifluoroacetic acid/dichloromethane (1:1 v/v) mixture. The solution was collected via filtering-off the resin and then concentrated in vacuo.

2.8. Polymer Characterization

Polymer **P1** was characterized by nuclear magnetic resonance (NMR), electrospray ionization mass spectrometry (ESI-MS), and size exclusion chromatography (SEC). NMR spectra were recorded using Bruker Avance Neo 500 MHz spectrometer equipped with an Ultrashield Magnet and Cryo-Probe Prodigy X-1H 5m. High-resolution MS and MS/MS experiments were performed using a QqTOF mass spectrometer (QStar Elite, Applied Biosystems SCIEX) with the ESI source for accurate mass measurements in the orthogonal acceleration time-of-flight (oa-TOF) mass analyzer. In this instrument, air was used as nebulizing gas (10 psi) while nitrogen was used as curtain gas (20 psi) and collision gas. SEC analysis was performed using three PLGel individual pore size columns (5 µm, 30 cm, diameter = 7.5 mm): 50, 100, and 500 Å, an Agilent 1260 Infinity II RID refractometer, Agilent 1260 Infinity II VWD detector. Tetrahydrofuran was used as the mobile phase with a flow rate of 1 mL min^{-1} for a 45 min analysis time. The column compartment and the RID refractometer cell were heated at 35 C. Toluene was used as an internal reference. The calibration was based on linear polystyrene standards from Agilent EasiVial.

3. Results

3.1. Hydrogen-Bond Signature of the Homopolymers

Our initial simulations focused on the conformational properties of a homopolymer chain composed of 50 "1" monomers in a vacuum. These simulations revealed the formation of a significant hydrogen-bonding network characterized by a prominent peak in the radial distribution function, $g(r)$, between the C_1 carbon atoms (see **Figure 2**). This network bears a resemblance to the backbone hydrogen bonding patterns found in the structures of proteins, suggesting a tendency for the chain to adopt organized secondary structures. In Figure S4 (Supporting Information) of the supporting information we show the additional g(r)s for the N–N and N-O_1 atom pairs, the figures confirm the indication of the formation of the hydrogen bond network.

Beyond the collapsed conformation depicted in Figure 2, additional folding patterns were observed in the designed sequences. Specifically, in certain cases, the chain exhibits regions of the extended structure, with local ordering that mimics beta-sheet arrangements, as illustrated in Figure S5 (Supporting Information). These configurations arise from the cooperative formation of hydrogen bonds between aligned segments of the polymer chain, analogous to beta-strands.

Furthermore, our analysis of the hydrogen bonding network indicates that these beta-strand-like structures are stabilized by directional interactions, contributing to their persistence throughout the simulation. Although these folded structures do not display the same level of definition or complexity as natural protein folds, they provide important insights into the folding

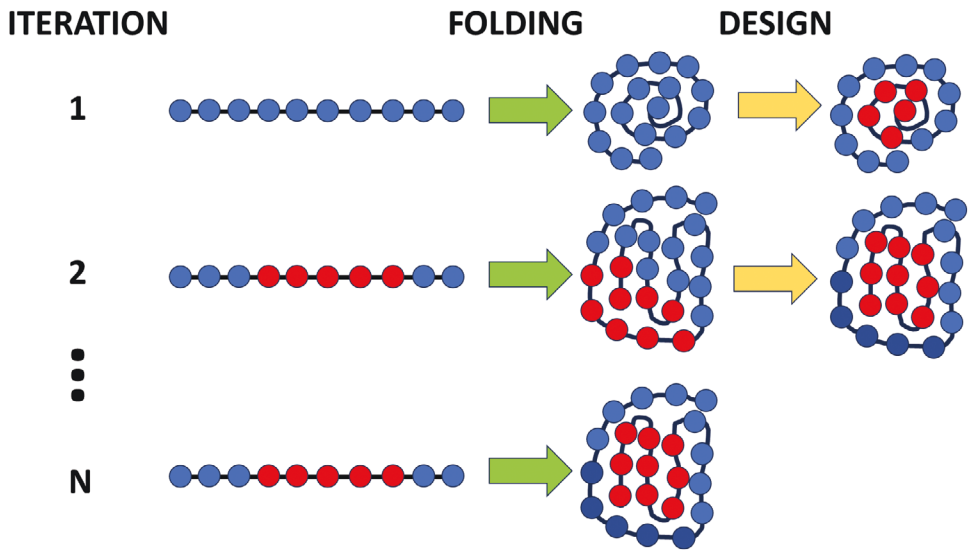

Figure 3. Schematic representation of the conformation-dependent design algorithm over N iterations. In Iteration 1, the polymer begins as a homopolymer of type 0 in a stretched configuration. A folding simulation in a vacuum is performed to collapse the polymer. During the design step, residues that are exposed to the surface are assigned as type 1 while the buried ones are labeled type 0. In Iteration 2, the designed chain is stretched and folded again. If the new equilibrium conformation differs from the one used in Design Step 1, the chain is redesigned and used for the next iteration. This process continues until the sequence remains unchanged.

behavior of sequence-defined polyurethanes. The presence of these distinct, stable conformations highlights the capacity for controlled design in achieving specific structural outcomes, paving the way for more refined sequence-based folding strategies. This supports our claim that, beyond simply collapsed globules, these polyurethanes can indeed form more organized conformations reminiscent of natural macromolecular folds.

3.2. Test Design with Two Letters in Vacuum

In the subsequent phase of our study, we embarked on the conformation-dependent sequence design for polyurethane chains of lengths 11, 20, and 50 monomers. This iterative process, spanning five cycles explained in **Figure 3** consists of alternating folding and design steps until the sequence generated does not change anymore. The design process led to the convergence of the chains to their definitive sequences with their equilibrated native structure, which are catalogued in **Figure 4**c. The crux of our analysis involved conducting free energy calculations, for which we mapped the conformational space onto a collective variable—namely, the Root Mean Square Deviation (RMSD).

The reference structure used in each RMSD plot is the final structure to which the conformation-dependent design is equilibrated to. These experiments are aimed to confirm the consistent folding behaviors observed in vacuum conditions in conditions that are closer to the reality of experiments. Additionally, we explored the impact of chirality by randomizing the side-chain enantiomer of monomer 1 to mimic experimental conditions. This approach allowed us to investigate the influence of stereochemistry on the folding behavior and stability of the polyurethane chains. By incorporating both chiral and racemic configurations, we ensured that our simulations captured a realistic representation of the polymer's behavior under experimental conditions.

3.3. Folding the Designed Sequences in Water

After we reached an equilibrated sequence during the conformation-dependent design, we proceeded to study the folding behavior of the sequence under various conditions. The initial folding simulations were conducted in vacuum, starting from an extended structure, following the protocol outlined in the Experimental Section. Subsequently, the collapsed chain was solvated and further simulated in explicit water for 40 ns using a metadynamics approach. The use of metadynamics was crucial for enhancing the sampling of the free energy landscape and preventing the chain from becoming trapped in local minima, which could lead to misleading conclusions about the equilibrium state of the polymer.

Sampling the free energy *F[RMSD]* is essential for accurately characterizing the equilibrium properties of the folding process. The free energy landscape provides a comprehensive view of both the energetic and entropic contributions to the stability of various conformations. By analyzing the free energy as a function of *RMSD*, we can determine the most thermodynamically favorable structures—the ones that are exponentially more likely to be populated at equilibrium, according to $P = \exp(-F/k_BT)$.

The Metadynamics approach allows for effective exploration of the conformational space, enabling us to overcome energy barriers that would otherwise trap the chain in metastable states. This method improves the reliability of our analysis by ensuring that the simulations explore the full energy landscape, capturing not only the lowest-energy conformations but also intermediate and higher-energy states that might be encountered during the folding process.

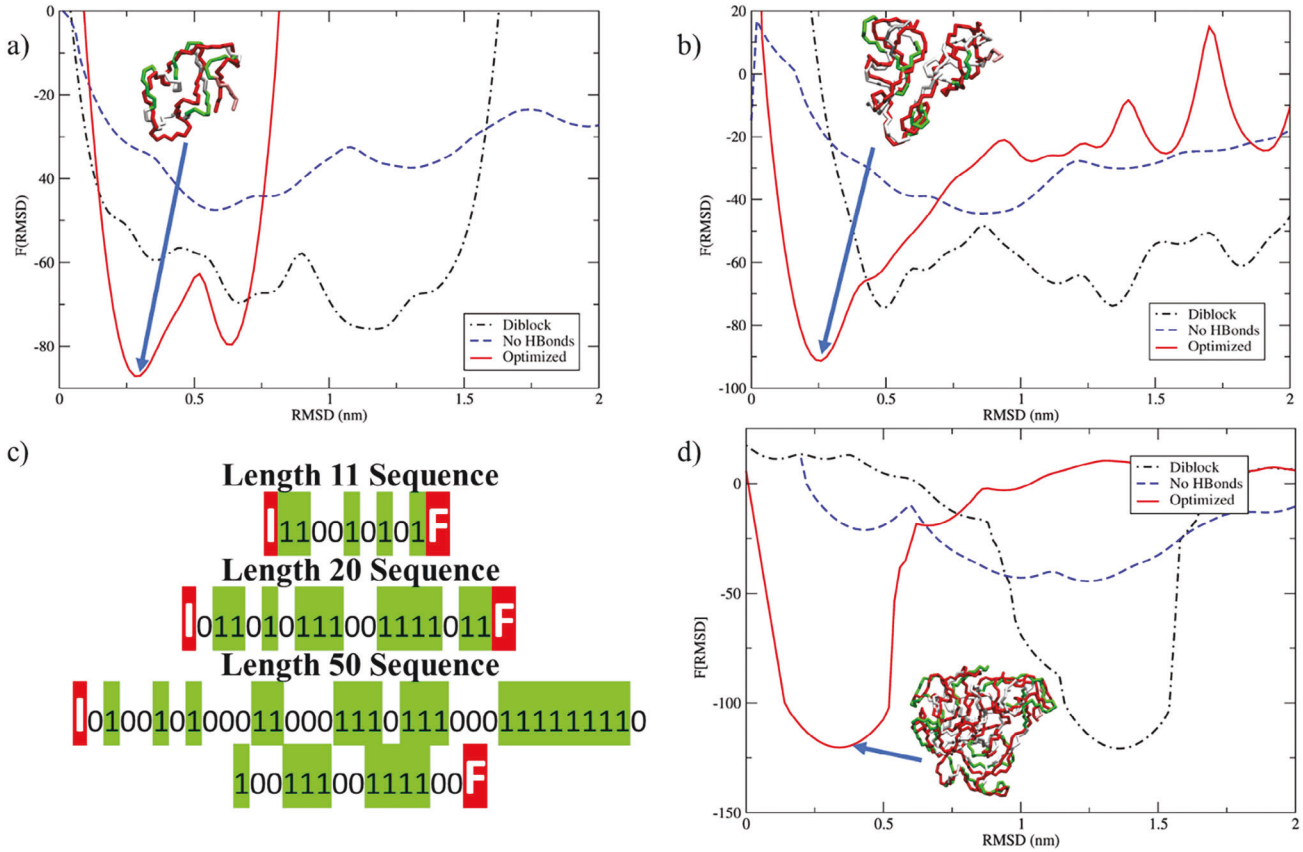

Figure 4. Free energy vs RMSD from the designed structure *F[RMSD]* for chains of lengths: a) 10, b) 20, d) 50. The free energies are in internal units of the simulations, this is because we used them only to show that optimized sequences have equilibrium configuration close to the target structure at RMSD=0. The values should be interpreted at this stage of the model as actual predictions of the thermodynamic stability. c) Sequences obtained from the Conformation-dependent sequence design of hetero polyurethane chains of 10 20, and 50 monomers. Diblock simulations correspond to sequences comprised equally of type 0 and type 1 monomers. Optimized simulations involve the folding of these designed sequences. The chirality of monomer 1 is randomized to mimic experimental conditions. Additionally, the 'No H-Bonds' simulations involve refolding of designed sequences with the hydrogen bonding term disabled in the force field. All refolding simulations were performed in water. These figures demonstrate that the design process systematically drives the chain toward the target structure. Due to the two-letter alphabet used, while higher folding precision is limited, the results confirm the polymer's designability.

Figure 3 encapsulates the free energy landscape for the sequences that were finely tuned through our optimization process and then refolded in explicit water. Notably, the free energy landscape for the 50-mer chain (Figure 4d) shows a deep energy well near RMSD = 0.3 nm, indicating a strong preference for the intended folded conformation. The presence of well-defined energy barriers between this minimum and other higher-energy conformers underscores the stability of the fold, even in the face of perturbations or changes in environmental conditions.

This stability is further validated by the consistent folding behavior observed across different conditions, including both vacuum and aqueous simulations. The depth and sharpness of the energy well in the *F[RMSD]* plot demonstrate that the optimized sequence reliably adopts the desired structure at equilibrium, reinforcing the effectiveness of our design approach. Overall, the sampling of the free energy landscape provides critical insights into the folding process, ensuring that our conclusions are based on robust equilibrium properties rather than transient kinetic effects. The energy landscape also reveals the presence of intermediate states with moderate RMSD values, which correspond to partially folded structures observed during the simulation. These intermediates are separated from the global minimum by significant energy barriers, indicating that spontaneous transitions between the target structure and these misfolded states are unlikely. Such features are characteristic of well-designed sequences, where the energy landscape is "funnel-shaped" toward the desired conformation, minimizing the likelihood of non-native folding.

Crucially, these simulations in aqueous environments revealed that the folding dynamics and the formation of the hydrogen bonding network among the C_1–C_1 carbon atoms occur in a manner analogous to those observed in vacuum. The introduction of chirality variations did not lead to any qualitative differences in the overall behavior of the system, confirming the robustness of the polymer's conformational characteristics across different simulation conditions. For comparative analysis, the figure also includes data on a diblock chain composed of a first block of segments of '0' followed by a second block of '1' monomers and equal length to the first, alongside a variant of the optimized sequence where hydrogen bond interactions were deliberately

excluded. Notably, the optimized sequences demonstrate a pronounced preference for lower RMSD values, indicating free energy minima that signify a strong control over the polymer's conformational space. This is in stark contrast to the other cases, which depict a random globule conformation typical of polymers lacking specific design considerations.

While the precision in folding observed here does not yet match that of proteins—hardly surprising given the simplistic two-letter monomer alphabet employed—the ability to exert substantial influence over the polymer's conformational space is indeed remarkable. This indicates not only the success of our sequence design methodology but also highlights the potential for further refining the control mechanisms to achieve even more precise conformational outcomes.

These results bolster our confidence in the polymer design's stability and its potential application across diverse environments.

3.4. Prediction-Guided Polymer Synthesis

Following in silico predictions, a model heteropolymer **P1** containing the optimized 11-mer sequence **I110010101F** was synthesized by solid-phase chemistry and characterized by mass spectrometry, size-exclusion chromatography, and NMR. Figures S2 and S3 (Supporting Information) show the MS and SEC characterization of this sample. Both techniques evidenced the synthesis of a uniform polymer with a controlled chain-length. Furthermore, MS/MS measurements (Figure S2, Supporting Information) confirmed the formation of the in silico optimized sequence. Preliminary solubility tests and NMR measurements indicated a good solubility at room temperature in chloroform, dichloromethane, dimethyl formamide and dimethyl sulfoxide. In those media, polymer-solvent interactions dominate. This is expected since DMF and DMSO are H-bond-breaking solvents.

Solubility at room temperature is also attainable after heating in methanol, acetone, and tetrahydrofuran (the latter case is highlighted by the SEC results displayed in Figure S3, Supporting Information). In aqueous medium, polymer-polymer interactions strongly dominate at room temperature. Visual solubility can only be attained by heating. Yet, NMR measurements recorded in D_2O at room temperature evidence a small fraction of soluble chains (data not shown).

4. Conclusion

In this study, we have explored the potential of sequence-defined polyurethanes as heteropolymers that could be designed to adopt specific conformations. Through a combination of molecular dynamics simulations and conformation-dependent sequence design, we have provided insights into the ability of polyurethanes to form stable and predictable structures under controlled conditions. While this research is an early-stage investigation, the findings lay a promising foundation for the controlled folding behaviors of these polymers.

Our simulations suggest the formation of hydrogen bonding networks that resemble those observed in natural proteins, hinting at the possibility for polyurethanes to mimic certain aspects of biological folding mechanisms. However, it is important to note that, while our results indicate potential for controlled folding, further experimental validation is needed to conclusively demonstrate that these sequence-defined polyurethanes achieve the precise target structures envisioned.

The successful synthesis and characterization of the model heteropolymer P1, which followed the computationally optimized sequence, offer preliminary validation of our design approach. Solubility tests and NMR measurements revealed that these polymers exhibit favorable properties in various solvents, reinforcing their versatility and applicability. Nevertheless, the experimental evidence presented here serves primarily as an initial proof of concept, highlighting the need for additional studies to fully realize and verify the design principles needed to achieve true foldamer-like behavior in synthetic polyurethanes.

In the spirit of moving closer to structural design with folding precision similar to that of proteins, our approach naturally opens the possibility of translating many protein functionalities to polyurethanes, including catalysis, drug delivery, and the design of self-assembling materials.[43–46] Moreover, by leveraging the principles of sequence control and conformation-dependent design, we can create polymers with specific functions and behaviors similar to natural biomolecules.[47] This approach not only ensures the creation of polymers with desired 3D shapes but also leverages the intrinsic properties of the polymers for specific applications, highlighting its significance in the realm of materials science.

Drawing a parallel to proteins, the reuse of the same building blocks for different chain designs could lead to a more sustainable industry. This study paves the way for future research into the development of advanced materials for applications in catalysis, drug delivery, and molecular recognition among others The versatility and designability of polyurethanes, as demonstrated in this work, position them as promising candidates for a wide range of applications, potentially transforming the landscape of material science and polymer chemistry.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

S.S. and J-F.L. thank Laurence Charles (Aix Marseille Université) for the MS and MS/MS measurements.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

informational Polymers, macromolecular folding, sequence-controlled polymers, single-chain technology

Received: June 29, 2024
Revised: September 3, 2024
Published online: October 15, 2024

[1] J. C. Nelson, J. G. Saven, J. S. Moore, P. G. Wolynes, *Science* **1997**, *277*, 5333.
[2] R. B. Prince, J. G. Saven, P. G. Wolynes, J. S. Moore, *J. Am. Chem. Soc.* **1999**, *121*, 3114.
[3] D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, J. S. Moore, *Chem. Rev.* **2001**, *101*, 3893.
[4] G. Guichard, I. Huc, *Chem. Commun.* **2011**, *47*, 5933.
[5] C. M. Goodman, S. Choi, S. Shandler, W. F. DeGrado, *Nat. Chem. Biol.* **2007**, *3*, 252.
[6] M. Gonzalez-Burgos, A. Latorre-Sanchez, J. A. Pomposo, *Chem. Soc. Rev.* **2015**, *44*, 6122.
[7] E. A. John, C. J. Massena, O. B. Berryman, *Chem. Rev.* **2020**, *120*, 2759.
[8] J.-F. Lutz, *ACS Macro Lett.* **2014**, *3*, 1020.
[9] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1238149.
[10] S. Rinaldi, *Molecules* **2020**, *25*, 3276.
[11] Z. C. Girvin, M. K. Andrews, X. Liu, S. H. Gellman, *Science* **2019**, *366*, 1528.
[12] P. Sang, J. Cai, *Chem. Soc. Rev.* **2023**, *52*, 4843.
[13] Y. Liu, S. Pujals, P. J. M. Stals, T. Paulöhrl, S. I. Presolski, E. W. Meijer, L. Albertazzi, A. R. A. Palmans, *J. Am. Chem. Soc.* **2018**, *140*, 3423.
[14] B. V. K. J. Schmidt, N. Fechler, J. Falkenhagen, J.-F. Lutz, *Nat. Chem.* **2011**, *3*, 234.
[15] N. Giuseppone, J.-F. Lutz, *Nature* **2011**, *473*, 40.
[16] R. P. Cheng, S. H. Gellman, W. F. DeGrado, *Chem. Rev.* **2001**, *101*, 3219.
[17] D. H. Appella, L. A. Christianson, I. L. Karle, D. R. Powell, S. H. Gellman, *J. Am. Chem. Soc.* **1996**, *118*, 13071.
[18] S. H. Gellman, *Acc. Chem. Res.* **1998**, *31*, 173.
[19] G. Magi Meconi, I. R. Sasselli, V. Bianco, J. N. Onuchic, I. Coluzza, *Rep. Prog. Phys.* **2022**, *85*, 086601.
[20] C. Cardelli, F. Nerattini, L. Tubiana, V. Bianco, C. Dellago, F. Sciortino, I. Coluzza, *Adv. Theory Simul.* **2019**, *2*, 1900031.
[21] C. Cardelli, V. Bianco, L. Rovigatti, F. Nerattini, L. Tubiana, C. Dellago, I. Coluzza, *Scientific Rep.* **2017**, *7*, 4986.
[22] I. Coluzza, P. D. J. van Oostrum, B. Capone, E. Reimhult, C. Dellago, *Phys. Rev. Lett.* **2013**, *110*, 075501.
[23] L. Matolyak, J. Keum, L. T. J. Korley, *Biomacromolecules* **2016**, *17*, 3931.
[24] Y. R. Nelli, L. Fischer, G. W. Collie, B. Kauffmann, G. Guichard, *Peptide Sci.* **2013**, *100*, 687.
[25] M. Szatko, W. Forysiak, S. Kozub, T. Andruniów, R. Szweda, *ACS Biomater. Sci. Eng.* **2024**, *10*, 3727.
[26] X. Wang, Q. Gan, B. Wicher, Y. Ferrand, I. Huc, *Angew. Chem.* **2019**, *131*, 4249.
[27] I. Youssef, S. Samokhvalova, I. Sergent, L. Charles, J.-F. Lutz, *Precis. Chem.* **2023**, *1*, 480.
[28] T. Mondal, V. Greff, B. É. Petit, L. Charles, J.-F. Lutz, *ACS Macro Lett.* **2019**, *8*, 1002.
[29] U. S. Gunay, B. E. Petit, D. Karamessini, A. Al Ouahabi, J.-A. Amalian, C. Chendo, M. Bouquey, D. Gigmes, L. Charles, J.-F. Lutz, *Chem.* **2016**, *1*, P114.
[30] I. Youssef, I. Carvin-Sergent, E. Konishcheva, S. Kebe, V. Greff, D. Karamessini, M. Matloubi, A. A.l Ouahabi, J. Moesslein, J.-A. Amalian, S. Poyer, L. Charles, J.-F. Lutz, *Macromolec. Rapid Commun.* **2022**, *43*, 412.
[31] D. Karamessini, S. Poyer, L. Charles, J.-F. Lutz, *Macromolec. Rapid Commun.* **2017**, *38*, 426.
[32] D. Karamessini, T. Simon-Yarza, S. Poyer, E. Konishcheva, L. Charles, D. Letourneur, J.-F. Lutz, *Angew. Chem., Int. Ed.* **2018**, *57*, 10574.
[33] D. Karamessini, B. E. Petit, M. Bouquey, L. Charles, J.-F. Lutz, *Adv. Funct. Mater.* **2017**, *27*, 1604595.
[34] B. É. Petit, B. Lotz, J.-F. Lutz, *ACS Macro Lett.* **2019**, *8*, 779.
[35] C. E. Fernández, M. Bermúdez, S. Muñoz-Guerra, S. León, R. M. Versteegen, E. W. Meijer, *Macromolecules* **2010**, *43*, 4161.
[36] T. Mondal, K. Dan, J. Deb, S. S. Jana, S. Ghosh, *Langmuir* **2013**, *29*, 6746.
[37] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1*, 19.
[38] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 12225.
[39] A. Laio, F. L. Gervasio, *Rep. Prog. Phys.* **2008**, *71*, 126601.
[40] T. P. consortium, *Nat. Methods* **2019**, *16*, 670.
[41] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463.
[42] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.
[43] Q. Shi, Z. Zhang, S. Liu, *Angew. Chem., Int. Ed.* **2024**, *63*, e202313370.
[44] J.-F. Lutz, *Eur. Polym. J.* **2023**, *199*, 112465.
[45] R. Szweda, *Prog. Polym. Sci* **2023**, *145*, 101737.
[46] C. Yang, K. B. Wu, Y. Deng, J. Yuan, J. Niu, *ACS Macro Lett.* **2021**, *10*, 243.
[47] J.-F. Lutz, *ACS Macro Lett.* **2020**, *9*, 185.