

Cite this: *Polym. Chem.*, 2025, **16**,
4622

Molecular dynamics as a tool for unveiling protein-like folding behavior in urethane-based macromolecules: the effect of chain length on the secondary structure of oligourethanes

Kasper Witruk,^{a,b} Sara Njoku,^c Eduardo Castellanos,^b Róża Szweda^{*b} and Tadeusz Andruniów^{ID *a}

Sequence-defined and stereocontrolled polymers represent an emerging class of macromolecules, offering the potential to design protein-like molecular architectures based on abiotic polymer backbones. These systems hold promise for applications requiring specific intermolecular interactions, which can be exploited, for instance, in catalysis or sensing. Realizing this potential necessitates a deeper understanding of their folding behavior beyond aqueous solutions, which resemble physiological conditions. In this study, the influence of chain length on the folding of model oligourethanes – comprising four, eight, and twelve monomer building blocks, is investigated *in silico* by molecular dynamics (MD) and experimentally using circular dichroism spectroscopy (CD). It is demonstrated that stabilizing interactions constrain the expansion of their conformational space while also revealing the inherent limitations of such processes. The increase of chain length from four to eight units results in the stabilization of the helical secondary structure. Further elongation leads to the formation of more complex structures, which appear as a combination of looping helix motifs, resembling tertiary structures. Importantly, to learn about the structural details of studied systems, a set of methodological adaptations is proposed to enhance the applicability of conventional data analysis techniques, which are typically applied to amide-based macromolecules, to study secondary structure of urethane-based sequence-defined systems.

Received 26th June 2025,
Accepted 23rd September 2025

DOI: 10.1039/d5py00635j

rsc.li/polymers

1 Introduction

Sequence-defined and stereocontrolled macromolecules, such as proteins, represent the pinnacle of functional sophistication among polymers. The unique properties of proteins are dictated by the precise sequence of amino acids, which direct the folding of their backbone into well-defined secondary and tertiary structures. While the concept of replicating these properties using abiotic polymers was proposed prior to the turn of the millennium, achieving true biomimicry using non-polyamide backbones remains an unresolved challenge.^{1,2}

Significant progress has been made in the development of foldamers, particularly those based on protein-like polyamide

skeleton. Non-natural peptides incorporating β -³ and γ -⁴ or cyclic amino acids,⁵ along with various other derivatives, have been extensively investigated,^{6–8} leading to many important advances in control over folding. Among the most important examples we can mention the tertiary structure of aromatic oligoamides,⁹ forming two helices connected by turn, that were joined together through hydrogen bonding. The helices design was based on rigid aromatic rings, restricting conformational dynamics of the structure, therefore enforcing formation of desired secondary structure. A simple tertiary structure (helix–turn–helix) were achieved for oligourethanes chains.^{10,11} Even more interesting observation was made for sequence-defined oligomers based on triazine, where more complex folding pattern was observed for sufficiently long oligomers.¹² However, the potential of alternative backbones offered by abiotic polymers has yet to be fully explored.^{13–15}

Some degree of structural control has been observed in copolymers,^{16–18} where functional groups positioned along the chain interact with one another.¹⁹ Polymers with controlled but undefined monomer sequences have been shown to fold into predetermined cyclic topologies through the formation of covalent bridges at specific locations.²⁰ This approach has

^aInstitute of Advanced Materials, Faculty of Chemistry, Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland.

E-mail: tadeusz.andruniow@pwr.edu.pl

^bCenter for Advanced Technologies, Adam Mickiewicz University in Poznań, Uniwersytetu Poznańskiego 10, 61-614 Poznań, Poland.

E-mail: roza.szweda@amu.edu.pl

^cLukasiewicz Research Network-PORT Polish Center for Technology Development, Stalówicka 147, 54-066 Wrocław, Poland



been extended to star-shaped polymer topologies, resulting in multicyclic conformations.²¹ Despite these advances, achieving the level of precision required to mimic the specificity of biomacromolecules remains elusive, as absolute control over dispersion, sequence, and stereocenter configuration is demanding to attain.²²

Recent developments have led to methods for the multistep fabrication of sequence-defined and stereocontrolled polymers.²³ The primary prospects arising from these advances include catalysis, achieved by emulating the functions of natural enzymes, selective sensing, facilitated by the creation of specific binding sites, and potential medical applications linked to their bioactivity.^{24–27} The dependence of physicochemical properties of polymers depends on stereocontrol. For instance, by introducing stereocontrol we enhance rigidity of the backbone,²⁸ as well as potential for tuning backbone-ligand interactions.²⁹ However, due to the enormous number of potential configurations—an issue compounded by the exponential increase in possibilities with sequence length—the realization of these prospects is not feasible without a thorough understanding of the sequence-structure relationship. This highlights the importance of studying the folding process and the relationship between sequence and structure, starting with secondary structures. Therefore, the development of computational methodologies that enables the *in silico* pre-screening of enormous sequence space is of great importance. Such knowledge is critical for understanding the arrangement of secondary and tertiary structures, formed with elongation of polymer chain, and their subsequent applications.

Molecular mechanics-based simulations represent a promising approach to address this challenge, as they provide detailed insights into the structural properties of the studied species at a relatively low computational cost. However, their application to the study of new materials presents several significant challenges that must be addressed. First, for novel materials, it is often the case that a well-established force field is not available. Consequently, the force field must be parameterized specifically for the material in question, which introduces an additional layer of complexity in performing accurate simulations. In fact, the choice of the force field remains an actively studied issue not only for proteins,^{30,31} but also for other molecular systems ranging from non-standard peptides^{32,33} towards synthetic polymers.^{34,35} Similarly to protein folding studies, the choice of a force field and the quality of its parameters strongly influence the observed behavior, however not only the force-field components describing the studied species are important for obtaining reliable results, but also those of the medium.³⁴ Furthermore, studies of novel materials suffer from two additional shortcomings: limited range of suitable parametrization tools and availability of experimental data based on which the force field parameters could be optimized. Nonetheless, this does not negate the possibility of utilizing molecular simulations in these studies, but rather stresses the necessity of experimental validation and need for improvement of the methodology over time.

Second, the presence of high energy barriers can significantly complicate simulations by increasing the computational cost required to achieve adequate sampling for a reliable description of system dynamics. While several strategies have been proposed to address this issue,^{36–38} these methods are primarily designed for the study of biological systems, usually proteins, and are not directly applicable to abiotic polymers. Third, many analysis techniques that are routinely used for biological systems may be insufficient or inappropriate for studying new materials without modification.^{39,40} For example, Ramachandran plots,⁴¹ commonly employed in protein structure analysis, would not be suitable for oligourethanes—such polymers contain three active torsional angles rather than the two typically found in proteins.²⁹ Thus, it is essential to assess which methodologies developed for biological systems can be adapted for the study of abiotic macromolecules, which may need to be adjusted or even replaced with entirely new approaches.

The main goal of this study was to decipher a relationship between the number of monomeric units in the oligourethane backbones and their ability to fold into stable secondary structures. Since this diverges from the standard approach, a new methodology was developed based on clustering and Principle Component Analysis (PCA)⁴² in torsional angle dataspace of monomeric units, considering folding as a function of oligomer's torsional angles, and validated with experimental circular dichroism data. Additionally, in order to describe dynamic behavior of studied oligomers and to gather data that could potentially enable prediction of possible folding pathways, an analysis inspired by the Markov State Models (MSMs),⁴³ yet created for individual monomers, was designed. Such analysis grants more detailed representation of the active processes occurring in studied oligomers. This study provides comprehensive overview of isotactic oligourethanes folding in acetonitrile. At the same time, the introduction of novel methodologies to investigate the conformational space of abiotic urethane-based sequence-defined polymers, which may prove critical to shed light on sequence-structure relationship and as a consequence accelerate the process for *de novo* design of functional polymer capable to fold into well-defined structures.

2 Methods

2.1 Computational details

Molecular dynamics. Structures have been generated in two steps: (i) oligomer SMILES⁴⁴ were defined, and then (ii) their structure was generated using Open Babel.⁴⁵ Force field parameters (showcased in the SI, Tables S1–S5) were obtained using AcPyPe⁴⁶ (based on Antechamber^{47,48}), setting GAFF for force field and BCC for charges. In MD simulations, acetonitrile solvent was represented explicitly using structure and force field parameters derived by Coleman *et al.*⁴⁹ All further steps, aside from data analysis, have been performed using GROMACS.⁵⁰



Simulations carried out in this work followed the Multiple Simulated Annealing – Molecular Dynamics (MSA-MD) procedure proposed by Hao *et al.*³⁸ and recently utilized in the study of polyurethane systems by our group.²⁹ The MSA-MD procedure here was applied to 300 initially identical systems for each macromolecule studied. Systems were annealed starting from 300 K, assigned with random velocity for each system, heated in 50 ps to 500 K and maintained at this temperature for another 50 ps, to be at last cooled to 0 K in next 100 ps. Annealing was followed by the two-step equilibration procedure, 500 ps at constant volume followed by another 500 ps with pressure coupling enabled. During the first equilibration step, the v-rescale⁵¹ thermostat was used and during the second one, the c-rescale⁵² barostat was applied, while pressure was set to 1 bar. In the production phase, the temperature was set to 300 K and the pressure to 1 bar, however, the temperature control was achieved using the Nosé–Hoover^{53,54} thermostat and pressure control using the Berendsen⁵⁵ barostat. At all steps $r_{\text{cut-off}}$ was set to 1.2 nm and the Particle–Mesh Ewald method⁵⁶ was applied to both electrostatic and van der Waals interactions. Additionally LINCS⁵⁷ constraints were used to freeze lengths of all bonds with hydrogen atoms, in order to enable timestep length of 2 fs. In all stages of MD simulations, periodic boundary conditions were applied in all spatial directions. Using this approach, 300 series of 10 ns simulations was prepared, yielding 1.5×10^6 data points (data was written to file every 10^3 simulation steps) and 3 μs of total trajectory time length.

Noise Molecular dynamics trajectory analysis. GROMACS⁵⁰ was used to extract all relevant data from the trajectory files. The main tool used for structure visualisations was the Visual Molecular Dynamics (VMD) software.⁵⁸ Most analyses were done using Python 3.10,⁵⁹ with additional use of NumPy⁶⁰ for most computations, Matplotlib⁶¹ for data visualisations, as well as PANDAS⁶² for data management support and SciPy⁶³ for additional numerical methods.

The core of the simulation data analysis was conformational space clustering. The angle definitions have been chosen expanding upon our previous work.⁶⁴ A conformational space analysis was based on Scikit-learn's⁶⁵ implementation of Density-Based Spatial Clustering of Applications with Noise (DBSCAN).⁶⁶ The advantage of using the DBSCAN over other clustering algorithms is the fact that it assigns points into categories based on their density in given spaces, which eliminates the need to predefine what the conformational space should look like. This was considered especially valuable here, since the preliminary data about urethane-based macromolecules is limited. To analyze the conformational space a two stage approach has been implemented:

- clustering of torsional angle in dataspace, incorporating decorrelation of datapoints in time (in strict and relaxed substages):

$$\varepsilon = \sqrt{4R^2 \sum_{\text{angles}}^i \sin^2\left(\frac{\Delta\theta_i}{2}\right) + (\alpha\Delta t)^2} \quad (1)$$

where ε represents distance between points in the constructed phase space, $\Delta\theta_i$ represents difference between the angles, R is a scaling parameter, Δt difference in time coordinate, α is a parameter defining how significantly points should be decorrelated in time during trajectory clustering.

- clustering of previously registered clusters based on just their mean angle values:

$$\varepsilon = \sqrt{4R^2 \sum_{\text{angles}}^i \sin^2\left(\frac{\Delta\theta_i}{2}\right)} \quad (2)$$

where the maximum tolerated distance between points ε_{tol} is defined for the first step by:

$$\varepsilon_{\text{tol}} = \sqrt{4R^2 \sin^2\left(\frac{\Delta\theta_{\text{tol}}}{2}\right) + (\alpha\Delta_{\text{mint}})^2} \quad (3)$$

and for the second step by:

$$\varepsilon_{\text{tol}} = \sqrt{4R^2 \sin^2\left(\frac{\Delta\theta_{\text{tol}}}{2}\right)} = 2R \sin\left(\frac{\Delta\theta_{\text{tol}}}{2}\right) \quad (4)$$

All parameters were empirically chosen: α , which corresponds to decorrelation (by increasing distance in constructed space) due to time passed between the registered points, was set to 0.2, $\Delta\theta_{\text{tol}}$, which stands for the maximum allowed difference between the angle values in between the points, took the following values: 20° and 30° for subsequently strict and relaxed parts of the first stage and 10° for the second stage. R scaling factor $\left(\frac{180^\circ}{\pi}\right)$ arises due to periodicity of each angle.

In some cases, intermediary conformations were observed with such high frequency that it was necessary to manually split some of the core clusters defined by DBSCAN.

Furthermore, to gain understanding of system dynamics, and effects caused by the increase of chain length, Markov State Models (MSMs) were constructed,⁴³ to analyze conformation transition probabilities. In order to verify the quality of the approximation used for such analysis, three different tests, standard for MSMs, were performed:

- comparison of transition matrix's stationary vector to conformation distribution registered through simulations,
- lag time (τ) validations (based on region where function $\frac{-\tau}{\log(\mu)}$ stabilizes; μ is a corresponding eigenvalue),
- Chapman–Kolmogorov test.

The first test yielded the most promising results, while in case of lag time analysis, the function $\frac{-\tau}{\log(\mu)}$ stabilised even before 25 ps for some eigenvalues, while continued growing through whole considered range for others. Lastly the mean error from Chapman–Kolmogorov test, calculated as a square root of Frobenius norm of matrix of difference between real and predicted matrix elements, divided by square root of number of matrix elements, in only few cases exceeded 0.05. The results presented in this study were prepared using lag time of 100 ps. It is important to note that the transition



models used here are not proper MSMs, as the transition probabilities between conformations are not determined solely by the currently assumed conformation, but also depend on the state of the oligomer as a whole. For example, hydrogen bonding or spatial constraints may influence the transition probabilities.

Additionally, analyses based on Free Energy Landscape (FEL) were prepared using Principal Component Analysis (PCA) in torsional angle dataspace. This analysis was inspired by work of Maisuradze *et al.*,⁶⁷ but the active torsional angle space was used in order to improve separation of unique conformations for small oligomers. The algorithmic part of the analysis used Scikit-learn's implementation of PCA,⁶⁵ while the dataspace was constructed based on the same torsional angle representations as used for the clustering. Histograms presented in this paper were prepared using 250 bins. This analysis was supplemented with RMSD- R_g distributions, though in their case values of free energy were not calculated and distributions were considered only in the context of occurrences, setting 10^{-12} m as a bin size.

To analyze the meaning of different conformational states and conformational transitions, model structures based on idealized backbone dodecamer conformations have been created and visualised using PyMOL.⁶⁸ Computation of helicity along with parameters based on approximation of the shape as an elliptical helix allowed for analysis of two radii r_a and r_b , giving a better idea about the helix profile, and pitch, for which the change of sign allows to identify handedness changing transformations.

Electronic circular dichroism (ECD) spectra prediction. To obtain ECD spectra we performed TD-DFT^{69–81} calculations (electronic excitation energies and rotary strengths) for 20 lowest singlet excited states with M06-2X⁸² functional, 6-311+G* basis set,^{83,84} GD3 empirical dispersion correction,⁸⁵ and the Self-Consistent Reaction Field (SCRF),⁸⁶ with parameters corresponding to acetonitrile as the solvent. All calculations were done in Gaussian 16.⁸⁷

To estimate the contributions of local conformational dynamics to the spectra, two computational approaches were employed for the tetramer: (i) simulations of selected conformations pre-optimized at the DFT level of theory with the same set-up as above, and (ii) simulations of ten structures with geometry as extracted from the molecular dynamics trajectory in the vicinity of selected FEL minima (for spectra and discussion see section B.VII of the SI). The ECD spectra were simulated using the ChimeraX software⁸⁸ with the SEQCROW extension,⁸⁹ employing Gaussian peak shapes and a full width at half maximum (FWHM) of 0.50 eV. Further data processing—such as averaging over individual signals and shifting the spectra—was carried out in Python, using modules described in the paragraph detailing the data analysis.

2.2 Experimental details

Studied oligourethanes (tetramer, octamer and dodecamer) were synthesized using following materials: (S)-2-amino-1-pro-

panol (S, 98%, Apollo Scientific, ee: 97%, optical rotation: $[\alpha]_{20/D} + 18^\circ$), *N,N'*-disuccinimidyl carbonate (dSu, >98%, TCI), di-*tert*-butyl dicarbonate ((Boc)₂O, 95%, Angene), pyridine (99.8%, anhydrous, Sigma-Aldrich), acetonitrile (99.8%, anhydrous, Sigma-Aldrich), and ethyl acetate (99.5%, Chemsolute), trifluoroacetic acid (99%, Sigma-Aldrich), dichloromethane (DCM, >99.8%, Honeywell), di-*tert*-butyl dicarbonate (99%, Sigma-Aldrich) as received from the suppliers. The synthesis and characterization followed previously described procedures.^{28,90,91}

In the first step, the amino group of the amino alcohol monomer (S) was protected in the reaction with di-*tert*-butyl dicarbonate leading to Boc-S. Afterwards the free hydroxyl group of the Boc-S has been activated by dSu and coupled with amine group of consecutive monomer S forming dimer Boc-SS. Extension of oligomer chain was performed by repetitive steps of activation of the oligomer hydroxyl end-group followed by chemoselective coupling of amino alcohol monomer. After reaching tetramer, the product was isolated by extraction with water and ethyl acetate.

The Boc protecting group from Boc-SSSS was removed using 50% trifluoroacetic acid in DCM, the resulting deprotected tetramer (SSSS) was then coupled with Boc-SSSS tetramer to yield the octamer Boc-(SSSS)₂. For the dodecamer synthesis, Boc-protected octamer was coupled with deprotected tetramer (SSSS) under similar conditions yielding Boc-(SSSS)₃. Oligourethanes after synthesis were isolated by extraction with water and ethyl acetate and purified by recrystallization from a water/acetonitrile mixture. The structures of the products Boc-SSSS, Boc-(SSSS)₂ and Boc-(SSSS)₃ were confirmed by Liquid Chromatography-Mass Spectrometry (LC-MS), Size Exclusion Chromatography (SEC), and Nuclear Magnetic Resonance (NMR) analyses as shown in the SI (Fig. S1–S9). The characterization of secondary structures was performed using ECD spectroscopy (the complementary UV/Vis spectra are shown in Fig. S10). The spectra were measured using a Jasco J-810 spectropolarimeter with optical path 0.2 cm (quartz cell cuvette, Hellma) at spectroscopic grade acetonitrile (gradient grade for HPLC, VWR chemicals) and temperature of 300 K.

3 Results and discussion

Through this study, in order to gain insight into conformational details and influence of the chain length on folding and stability of the folded structures, three different lengths of isotactic oligourethanes were studied using MSA-MD approach. The studied structures, consisting of monomeric units having an absolute stereoconfiguration S, tetramer, octamer and dodecamer, with *tert*-butyloxycarbonyl protecting group at their chain end, are shown in Fig. 1. The results of MSA-MD calculations were verified by comparing them to experimental data from ECD analysis of synthesized oligourethanes.

3.1 Conformational analysis

Gaining an understanding of emergent behavior of urethane-based building blocks that arises from connecting them into





Fig. 1 Chemical structure of studied isotactic oligourethanes.

oligomeric or polymeric chains and immersing them in solvent, can be considered the first step towards intelligent design of functional macromolecules. In order to achieve that, a detailed look into conformational affinities and dynamics should prove to be the most beneficial approach. Because the torsional angles are the fundamental factor defining the final three dimensional structure assumed by the macromolecule, and folding process can be expressed as exploration of conformational space. In this work we analyzed the impact of: (i) the length of the oligomeric chain and (ii) the monomer's position in the oligomeric chain on the conformation characteristic, as

these are expected to change with individual monomeric unit surroundings.

As can be seen from Fig. 2, DBSCAN analysis (results from the second stage of clustering are showcased in Fig. S11–S13) reveals that a single conformation is generally preferred by all monomers no matter the chain length or position: $\phi \approx -120^\circ \pm 30$; $\omega' \approx -60^\circ \pm 10$; $\omega'' \approx 180^\circ \pm 10$; the only observed exception being the fifth monomer of the longest chain. The preferred values of the torsional angles ϕ and ω' being negative was expected, as these torsional angles are directly neighbouring the stereogenic centres. In-depth analysis of ϕ angle reveals that it oscillates between two different energy minima at around -60° and -150° , giving rise to its higher standard deviation ± 30 vs. ± 10 for other torsional angles.

Interestingly, although it would have been the dynamics of the ω'' dihedral, that would be expected to be the most active degree of freedom, as it should have relatively small rotation barriers, data presented in Fig. 2 suggests that in some regions it is the ω' : $-60^\circ \rightarrow 60^\circ$ transition that might be more important. The transition dynamics, that result in such picture, will be further discussed in a paragraph about kinetic analysis based on Markov chain analysis.

A Prevalent torsional angle values

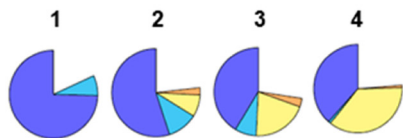


$$\phi \approx -120^\circ \quad \omega' \approx -60^\circ \quad \omega'' \approx 180^\circ \quad \phi \approx -120^\circ \quad \omega' \approx 60^\circ \quad \omega'' \approx 180^\circ$$

$$\phi \approx -120^\circ \quad \omega' \approx -60^\circ \quad \omega'' \approx 90^\circ \quad \phi \approx -120^\circ \quad \omega' \approx -60^\circ \quad \omega'' \approx -90^\circ$$

B Dominant torsional configurations at each unit

Boc-SSSS



Boc-(SSSS)₂



Boc-(SSSS)₃



Fig. 2 (A) Time-average torsional angle values for the most populated conformational states found in tetra-, octa-, and dodecamer. (B) Overview of the most important conformational state occupancy of each individual monomer. Noise and rarely observed conformations (as there we identified 18 unique local conformations) have been excluded. The structures presented on the right are model tetramer backbones, assuming idealized conformations considered in conformational makeup analysis based on ϕ , ω' , and ω'' torsional angle values.



The highest stability of the main conformation for the first monomer in each chain arises from Boc group symmetry. Its rotation by 120° results in arriving at the same orientation, thus it mostly disables one degree of freedom for the first monomer. Additionally, the relative size and mass of the Boc group might also have influence on restricting the dynamics of the first monomer. Noticeably lower stability of the main conformation was observed for terminal monomeric units in the tetramer, as well as around the middle of the chain for octa- and dodecamer. In the tetramer, the conformation differing from the dominant one by $\omega'' = 180^\circ \rightarrow 90^\circ$ increased its presence significantly, becoming effectively bistable with the dominant conformation at the fourth monomer. It is also important to notice that for third and fourth monomeric unit of the tetramer, the time averaged value of ϕ angle at the alternative conformation changed from around -120° to around -90° . Such values of the ϕ angles seem to correspond structurally with optimal rotation for formation of intrachain hydrogen bond between second and fourth monomeric unit. In octamer third and fourth monomeric unit had the most reduced occupancy of the dominant conformation, with ω' : $-60^\circ \rightarrow 60^\circ$ being the most important on the third monomeric unit, while $\omega'' = 180^\circ \rightarrow -90^\circ$ on the fourth. Similar situation can be observed for the dodecamer, where fifth and sixth monomeric unit underwent similar change of trend, although even to greater extent, with the alternative conformation slightly overtaking the dominant one on the fifth monomeric unit and dominant conformation being assumed by the sixth monomeric unit barely over 20% of the time. Interestingly for both octamer and dodecamer, the shift in conformational affinity occurred for monomers with indices $(\frac{n}{2} - 1$ for $n = 8$) and $(\frac{n}{2}$ for $n = 12$), respectively. Lastly, the increase of occupancy for the alternative conformation differing from the dominant one by $\omega'' = 180^\circ \rightarrow 90^\circ$ has been observed in all cases at the terminal monomer.

To gain an insight into how hydrogen bonds can alter the conformational landscape, we analyzed the dependence of affinity towards certain conformations on donated/accepted hydrogen bonds (hydrogen bond cutoff criteria are illustrated in Fig. S14 and S15).

As shown in Fig. 3, the hydrogen bonding scheme for the tetramer species is fairly simple and resembles model helical structure, as presented in Schemes E1 and E2. However, it is also important to notice, that even though interactions between the first and the third monomeric units as well as second and fourth are dominating (next to binding of hydroxyl group to the third mer's acceptor groups), the relative number of interactions between the first and fourth monomeric units is also significant. Considering additionally that monomers can spend non-negligible portion of the time without neither donating nor accepting hydrogen bonds, it can be concluded, that although the tetramer in acetonitrile has affinity for forming helical structure, it should not be expected to assume single well defined structure for extended amounts of time.

The observed hydrogen bonding scheme results from more than one type of structure being able to exist in the system.

Out of them, two main reference helix types can be distinguished, which are 'wide' helix - 4_{22} and 'tight' helix - 2_{12} . In tetramer, only 2_{12} can be properly stabilized through hydrogen bonding, as the tetramer chain is too short to stabilize the 4_{22} helix. Nonetheless, as data from the local conformational affinities suggests (Fig. 2), even without standard hydrogen bond stabilizing it, conformation corresponding to 4_{22} helix is often assumed by the tetramer, as the 'wide' helix corresponds to the preferable conformational arrangement of torsional angles. Another significant interaction involves the donation of a hydrogen bond by the terminal OH group, which may also correlate with the formation of a hydrogen bond between the fourth and second monomeric units, ultimately resulting in a right-handed, tight helix 2.6_{14} .

The hydrogen bonding scheme gets significantly more complicated for longer sequences. For octamer, different helical arrangements, mainly corresponding with 'wide' helices are possible.

At the same time, for dodecamer, it seems like the monomeric units from 9^{th} to 12^{th} have fairly clear affinity towards helical arrangements, while the oligomer segment of monomers between the first and 8th seems to assume a hydrogen bonding scheme more similar to sheet-like, as shown in E4 segment of the Fig. 3, though highly imperfect. That said, looking at hydrogen bonding maps alone not only does not allow left- and right-handed helices to be reliably distinguished, but can lead to other misinterpretations when secondary structures have sufficiently similar corresponding hydrogen bonding schemes. In fact, conformational analysis data as well as visual interpretation of trajectories paint a different picture, as the structure does not seem to be sheet-like structures, but could be more accurately described as a helix 'looping back on itself'.

Although information gathered from the analyses based on local conformations and hydrogen bonding reveals some important structural features of the studied oligomers, it does not represent fully which conformations dominate in the studied systems. To supplement for that, Free Energy Landscapes (FEL) created with Principal Component Analysis (PCA) in the torsional angle dataspace was performed (more detailed figures, comparison against PCA performed in the Cartesian coordinate space as well as variance ratios can be seen in Fig. S19–S24). The PCA-FEL analysis in case of tetramer shows a relatively simple picture, as presented in Fig. 4A. The most important portion of the landscape is constituted by the dynamics between the 'wide' helix and the 2_{12} helix, which appears in the top-left region. Between the minima corresponding with the two helices, there exists a shallower one corresponding to an intermediate structure. This part of the landscape shows a mostly bistable type of behavior, yet the tetramer can also reach other conformations, that are outside this region and one of the conformations it can assume corresponds to a minimum of a comparable depth to the two helical conformations. In that minimum one can find misfolded structures stabilized by hydrogen bond involving the terminal OH group.





Fig. 3 (A) Hydrogen bonding maps portraying the relative frequency of hydrogen bond formation between donor (indicated by the row index) and acceptor monomers (column indices), for tetramer, octamer and dodecamer, respectively. The values presented on the maps are relative to the highest signal in each oligomer, for absolute number see SI – Fig. S16–S18. Hydrogen bonding schemes presented under the hydrogen bonding maps have secondary structure domains highlighted with colors: cyan for left-handed helix and magenta for 'helix looping back on itself'. (B) Model maps of ideal hydrogen bonding schemes with related secondary structures. E(1, 2) correspond with helices and E(3, 4) with sheet-like structures. It is important to notice that these maps show only donor acceptor relationship, therefore are insufficient for differentiating between right- and left-handed helices.

More complex behaviors are observed for both the octamer and the dodecamer, as shown in Fig. 4B and C. Expectedly, in both cases, the FEL minima are not as deep as those observed for tetramer, although the difference is not necessarily substantial (up to 1.05 kJ mol^{-1}). In the case of the octamer, even though the clustering and the hydrogen bonding analyses already pointed to the existence of imperfect helical conformations rather than a single properly folded conformation, it is more clearly visible on the PCA-FEL map, as in the bottom left region a series of FEL minima can be observed. In the case of the dodecamer, as the clustering data suggested, full helices

are not playing a significant role in the folding of dodecamers. Instead, structures corresponding to the major minima contain an array of torsion angles leading to chain reversion at a selected point. This tends to cause the post-twist part of the chain to interact hydrogen-bonded with the pre-twist part, usually locally retaining features corresponding to helical structures. FEL based on mapping of radius of gyration (R_g) against root-mean-square-deviation (RMSD) (see SI, Fig. S25) generated for oligourethanes studied here support the results derived from PCA analysis discussed above. Interestingly, despite dodecamer being approximately 1.5 times longer than





Fig. 4 PCA-based FEL for tetramer (A), octamer (B) and dodecamer (C). In (A) three significant local minima can be observed for tetramer, two of which are types of helices ('wide' helix 4_{22} – structure on the left, 2_{12} helix – structure in the middle), while the third is a misfolded species stabilized by hydrogen bond donated by terminal OH group (shown on the right).

octamer, the range of typically assumed R_g values does not increase proportionally. This is an effect of the change in propensity from helix in the octamer to helix looping on itself in the dodecamer, strengthening our conclusions from previous analyses.

Even though the studied oligomers do not exhibit long-range structural order in acetonitrile, they clearly exhibit significant preferences towards an ensemble of structural arrangements, accompanied by short-range structural order. For the design of functional macromolecules based on building blocks of oligomers studied here, long-range order needs to be achieved. Understanding the relationship between local conformations and conformation of whole chain should be the first step towards designing an approach for stabilization of desired globular structures. A geometric analysis of structures built from monomers assuming defined, idealised local conformations might prove beneficial towards that goal. It was

realized through creation of backbone structures with specified torsional angles in two configurations: structures constructed from monomers assuming solely the specified conformation and structures when only specified mer assumed this conformation, while the remaining monomeric units assumed the generally preferred conformations. Proceeding from these specifications, two tables were prepared, presenting theoretical, idealised backbones of dodecamers.

In Table 1, for a set of torsional angles, that were chosen to grant the best idea about the local dynamics of the dominant conformation, parameters describing a helix are presented (the negative value of the pitch indicates that the helix is left-handed). Firstly, as can be seen from the significance of changes corresponding with different torsional angles: the dynamics of ϕ and ω' angles seem to correspond mainly with stretching and contracting of the helix (changes of the pitch), while the dynamics of ω'' seems to correspond with helix



Table 1 Characteristics of local dynamics of the dominant conformation^a

| | Dihedrals [°] | Pitch [Å] | Radius <i>a</i> [Å] | Radius <i>b</i> [Å] | Structure ^a |
|---|---|-------------------|---------------------|---------------------|--|
| 1 | $\phi = -120$ $\omega' = -60$ $\omega'' = 180$ | -11.57 ± 0.01 | 4.23 ± 0.11 | 4.16 ± 0.26 |  |
| 2 | $\phi = -90$ $\omega' = -60$ $\omega'' = 180$ | -3.41 ± 0.00 | 3.80 ± 0.13 | 4.08 ± 0.24 |  |
| 3 | $\phi = -150$ $\omega' = -60$ $\omega'' = 180$ | -16.74 ± 0.02 | 3.27 ± 0.10 | 3.69 ± 0.46 |  |
| 4 | $\phi = -120$ $\omega' = -50$ $\omega'' = 180$ | -7.56 ± 0.01 | 4.35 ± 0.11 | 4.32 ± 0.29 |  |
| 5 | $\phi = -120$ $\omega' = -60$ $\omega'' = -170$ (190) | -11.02 ± 0.02 | 4.85 ± 0.07 | 4.86 ± 0.23 |  |
| 6 | $\phi = -120$ $\omega' = -60$ $\omega'' = 170$ | -11.29 ± 0.01 | 3.65 ± 0.12 | 3.65 ± 0.25 |  |

^a Structure – model structures of backbones generated based on torsional angles. These model structures were generated to be twelve monomers long (dodecamers). The dihedral angles listed in the first column correspond to idealized values employed in the calculations. The pitch refers to the axial distance covered by one complete turn of the helix. The radii *a* and *b* are obtained by fitting the helical profile to an elliptical shape, thereby conveying both the overall width of the helix and its deviation from a perfectly circular profile. The table presents calculated pitch and radii for unweighted centers of backbone atoms in the model structures, and uncertainties are measures of fit wellness.

radius dynamics. It is also important to notice that with the high values of the pitch, formation of intrachain hydrogen bonds will be restricted. This provides insight into both why hydrogen bonds at many parts of the chains are not necessarily stable and why mean values of ϕ and ω' are often shifted towards lower values (reducing the helix pitch). This suggests that additional forces contributing to reduction of the pitch value, such as propensity towards reducing contact surface with solution, would act in synergy helping to stabilize the hydrogen bonds, in effect further stabilizing the helical fold.

Expectedly, far more significant structural changes can be observed for changes of conformational states. As presented in Table 2, conformational transitions have usually very significant effects on overall structure of the oligomer, especially as in few cases they can even lead to handedness reversal. The main transitions from the dominant conformation, corresponding to the helix reversal are $\omega' \rightarrow 60$ and $\omega'' \rightarrow -90$, and as anticipated they contribute the most significant changes in the oligomer. If they occur at monomer sufficiently far enough from the terminal monomeric unit, they can allow for interactions of different parts of the chain, by altering the direction of chain progression in space (Table S6). Additionally, when

occurring at the same time, their effect can be considered even more drastic, as it could lead to a locally non-helical structure. Moreover, the structural change introduced by transition $\omega'' \rightarrow 90$, not only retains handedness, but also differs significantly from what was observed for dynamics of the main conformation presented in Table 1 only in values of the minor and major radii. Presented information can be used to discern which torsional angle transitions have significant structural impact and which do not. Consequently, it can be seen that the most important alternative conformational states observed for tetramer and at the terminal monomeric unit of every studied oligomer only slightly alter the helix, while the alternative conformational states observed at the fifth and sixth monomeric unit of the dodecamer are the ones that can lead to interactions between different parts of the oligomer, allowing for formation of structure described as ‘helix looping back on itself’. This could offer an explanation for hydrogen bonding scheme observed for dodecamer, that was shown in Fig. 3. Interestingly, observable change of folding tendency being observed for oligomers after exceeding certain length is not unique for the oligomers studied here, as a fairly similar behavior in triazine-based oligomers was reported by Ahn and Grate.¹²



Table 2 Characteristics of structural changes corresponding with conformational transitions^a

| | Dihedrals [°] | Pitch [Å] | Radius a [Å] | Radius b [Å] | Structure ^a |
|---|--|-----------------------|-----------------|-----------------|---|
| 1 | $\phi = -120$ $\omega' = 60$ $\omega'' = 180$ | 11.31 ± 0.01 | 1.57 ± 0.09 | 1.60 ± 0.23 |  |
| 2 | $\phi = -120$ $\omega' = -60$ $\omega'' = 90$ | -8.10 ± 0.00 | 1.79 ± 0.09 | 1.78 ± 0.16 |  |
| 3 | $\phi = -120$ $\omega' = 60$ $\omega'' = 90$ | -6.50 ± 0.00 | 2.47 ± 0.06 | 2.56 ± 0.17 |  |
| 4 | $\phi = -120$ $\omega' = 60$ $\omega'' = -90$ | Non-helical structure | | |  |
| 5 | $\phi = -120$ $\omega' = -60$ $\omega'' = -90$ | 22.60 ± 0.02 | 4.64 ± 0.11 | 4.72 ± 0.16 |  |

^a Structure – model structures of backbones generated based on torsional angles. These model structures were generated to be twelve monomers long (dodecamers). The dihedral angles listed in the first column correspond to idealized values employed in the calculations. The pitch refers to the axial distance covered by one complete turn of the helix. The radii a and b are obtained by fitting the helical profile to an elliptical shape, thereby conveying both the overall width of the helix and its deviation from a perfectly circular profile. The table presents calculated pitch and radii for unweighted centers of backbone atoms in the model structures, and uncertainties are measures of fit wellness. Notice: the investigated model structures are purely theoretical, as it is extremely unlikely for dodecamer to assume them. However, for longer chains characterizing such structures might still prove a useful reference.

All this information should already provide insights into general behavior of studied oligomers, but the acquired understanding of the chain dynamics is still limited. To directly gather information about the chain dynamics of studied oligomers, it is necessary to examine the torsional transitions observed in molecular dynamics simulations. Based on this data, Markov chains, inspired by MSMs, were constructed. The Fig. 5 presents distributions of transitions probability of conformation changes, using the violin plot for representation. The distributions are extremely wide, with a number of points outside of cap range, which most likely results from inadequate statistical sampling of less frequently assumed local conformations. Despite the limitation, some important information can still be extracted from it, namely:

- rotation of the ϕ angle (outside of standard dynamic range around -120°) is extremely rare, and if it would be necessary to be considered, computational methods for overcoming energy barriers might be necessary to study the folding process,
- the ω' torsional angle of value around 60° is the most unstable, but it is seen more often than 180° because probability of oligomer assuming corresponding conformation is much higher,
- the ω'' torsional angle is the most active one as far as conformational transitions are considered,

- the ω'' torsional angle is more predisposed for transition from 180° to -90° in general, but far more prone to come back from it to 180° , while even though transition towards 90° is less likely, its retention is more favorable.

These trends in random exploration illustrate why local conformations that differ by more than a single torsional angle from the dominant conformation are relatively rare, causing the oligomer to spend a significant portion of the simulation time in conformations classified as wide helices and those closely related to them.

Results from Markov chains constructed for conformational transitions can be also grouped in a different manner. In Fig. 6, we display transition probability vs. chain length for six most important conformational transitions. In all instances, a clear downward trend of transition probabilities can be seen. This suggests that even though longer chains have higher probability of locally assuming different conformation, conformational transitions themselves are less likely to occur, which should lead towards two consequences. Firstly, what is the most important for material design, with the increase of the length of oligourethane chains the fold stability should also improve. Secondly, what is more important for modelling of such chains, without applying efficient conformational landscape scanning methods, modelling of such materials might





Fig. 5 Violin plots showcasing transition probabilities for different torsional angles.



Fig. 6 Violin plots comparing transition probabilities for different chain lengths.

get prohibitively expensive, as both the conformational landscape will increase its complexity and exploration itself will proceed more slowly (not even mentioning the increased computational cost arising from studying larger systems). With this in mind, the comparison of conformational transition probability shown in this work, as well as comparison of general conformational affinities for the monomers, it should be possible to choose sets of starting conformations for efficient scanning conformational landscape of larger oligomers and polymers based on oligourethane backbone. If this interpretation is correct, then design approach focusing on local torsional affinities, with correction for overall trends observed for larger oligomers, should yield satisfying results for formulating sequences of functional macromolecules based on oligourethanes. That being said, to verify how well this highly bottom-up approach work, just looking at molecular dynamics trajectories with single method is insufficient.

This specific behavior might be a result of, what was theorized by Ahn and Grate,¹² that with increase of chain length, the number of possible conformations and interactions that

can stabilise them increases. As a results, the structure preferred in short oligomeric chains, might cease to be the most beneficial in the longer ones. This probably plays an important role in the observed behaviour of oligomers studied here, but it does not directly explain why certain conformations are preferred over others. Although this could be partly explained by properly accounting for intra-chain interactions, there is another perspective that should be considered. Recognizing that the acetonitrile is a polar solvent and that oligomers investigated by us posses nonpolar sections, polarity dominated solvophobic effect might play a role here. The solvophobic effect would push oligomers to minimize the contact surface with the solvent, promoting combinations of local conformations allowing for that. In fact, the rough estimate of the energy required to increase the surface of interaction between oligomer and solvent at 300 K ($272\text{--}518\text{ J nm}^{-2}\text{ mol}^{-1}$ – see section B.VIII of the SI for details) appears to be considerably lower than the average molar energy of hydrogen bonds. However, it can still alter the free energy landscape, especially considering that differences between different folded states may reach several nm^2 even for as short oligomers as studied here. This suggests that different folding patterns may be observed in solvents that more closely match the mean oligomer's polarity, or with higher dielectric constant to diminish this effect. In such cases, other interactions, such as dispersive forces, should play significantly greater role, however, as this falls outside of the scope of this study, it will not be further explored in this work.

3.2 Experimental characterization of oligourethanes: verification of MD data

In order to assess the predicted trends in secondary structure formation and their transition with increasing oligomer length, the chiroptical properties of the Boc-SSSS, Boc-(SSSS)₂, and Boc-(SSSS)₃ were studied using circular dichroism (ECD) spectroscopy at UV range (Fig. 7A). The urethane monomeric unit is a weak UV-active chromophore, where the signal comes from $\pi\text{--}\pi^*$ transition of the carbonyl group, which appears around 190 nm, within the transparent spectral window for the selected solvent ($>185\text{ nm}$). The spectra were recorded at 0.1 mg mL^{-1} in ACN. Due to limited solubility of dodecamer, its spectrum was recorded in ACN with 30% addition of water.

All studied compounds exhibited optical activity, showing a negative Cotton effect centered at 185–190 nm and an absorption maximum at 190 nm (Fig. S10). While the UV pattern does not show critical differences, the pattern of the ECD spectrum shows changes depending on the oligomer length. Comparing spectra of tetramer and octamer at the same weight concentration, we observed a significant increase in ECD signal intensity and shift of its maximum towards longer wavelengths. This change suggests different conformational preferences of the backbone as a consequence of chain extension that is in line with MSA-MD data. This change may reflect variation in contribution of distinct conformers forming a dynamic equilibrium. Thus, the variations observed could be attributed to changes in the organized folded structure,



subject will be further explored and investigated in our future work.

Author contributions

TA and RS designed and supervised the study. KW did computational analyses. EC performed ECD and UV analyses. SN synthesised oligourethanes and contributed to their characterisation (LC-MS, SEC, NMR). All authors contributed to manuscript preparation.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). Experimental details on LC-MS, SEC, and spectroscopy techniques (NMR, UV-Vis, ECD), as well as detailed description of MD-based analyses. Experimental characterization (LC-MS, SEC, NMR, UV), further details about MD simulation analyses including clusterings, hydrogen bonding, free energy landscapes, solvophobicity model and ECD predictions. See DOI: <https://doi.org/10.1039/d5py00635j>.

Acknowledgements

The authors acknowledge the National Science Centre, Poland, for financial support under Grants No. 2021/43/I/ST4/01294, 2018/31/D/ST5/01365, and ERC Starting Grant No. 101116700. The authors thank Krzysztof Zwoliński for conducting the mass spectrometry (MS) analyses and Rafał Kowalczyk for recording the NMR spectra of compound Boc-(SSSS)₃. Computational resources were provided by the Wrocław Centre for Networking and Supercomputing (<https://wcss.pl>) and the Poznań Supercomputing and Networking Center (<https://www.pcss.pl/>).

References

- S. Gellman, *Acc. Chem. Res.*, 1998, **31**, 173–180.
- J.-F. Lutz, J.-M. Lehn, E. W. Meijer and K. Matyjaszewski, *Nat. Rev. Mater.*, 2016, **1**, 16024.
- L. K. A. Pilsl and O. Reiser, *Amino Acids*, 2011, **41**, 709–718.
- R. P. Cheng, *Curr. Opin. Struct. Biol.*, 2004, **14**, 512–520.
- F. Fülöp, T. A. Martinek and G. K. Tóth, *Chem. Soc. Rev.*, 2006, **35**, 323–334.
- C. M. Goodman, S. Choi, S. Shandler and W. F. DeGrado, *Nat. Chem. Biol.*, 2007, **3**, 252–262.
- S. Rinaldi, *Molecules*, 2020, **25**, 3276.
- G. Guichard and I. Huc, *Chem. Commun.*, 2011, **47**, 5933–5941.
- D. Mazzier, S. De, B. Wicher, V. Maurizot and I. Huc, *Chem. Sci.*, 2019, **10**, 6984–6991.
- N. Pendem, Y.-R. Nelli, L. Cussol, C. Didierjean, B. Kauffmann, C. Dolain and G. Guichard, *Chem. – Eur. J.*, 2023, **29**, e202300087.
- S. Wang, L. Allmendinger and I. Huc, *Angew. Chem., Int. Ed.*, 2024, **63**, e202413252.
- S.-H. Ahn and J. W. Grate, *J. Phys. Chem. B*, 2019, **123**, 9364–9377.
- L. Fischer and G. Guichard, *Org. Biomol. Chem.*, 2010, **8**, 3101–3117.
- S. Kohmoto, H. Takeichi, K. Kishikawa, H. Masu and I. Azumaya, *Tetrahedron Lett.*, 2008, **49**, 1223–1227.
- Z. Luo, N. Zhu and D. Zhao, *Chem. – Eur. J.*, 2016, **22**, 11028–11034.
- Y. Vo, R. Raveendran, C. Cao, R. Y. Lai, M. Lossa, H. Foster and M. H. Stenzel, *ACS Appl. Mater. Interfaces*, 2024, **16**, 59833–59848.
- Z. Han, Z. Li, M. H. Stenzel and R. Chapman, *J. Am. Chem. Soc.*, 2024, **146**, 22093–22102.
- A. J. DeStefano, S. D. Mengel, M. W. Bates, S. Jiao, M. S. Shell, S. Han and R. A. Segalman, *Macromolecules*, 2024, **57**, 1469–1477.
- B. A. Laurent and S. M. Grayson, *Chem. Soc. Rev.*, 2009, **38**, 2202–2213.
- B. V. K. J. Schmidt, N. Fehler, J. Falkenhagen and J.-F. Lutz, *Nat. Chem.*, 2011, **3**, 234–238.
- Y. Mato, K. Honda, B. J. Ree, K. Tajima, T. Yamamoto, T. Deguchi, T. Isono and T. Satoh, *Commun. Chem.*, 2020, **3**, 97.
- J.-F. Lutz, M. Ouchi, D. R. Liu and M. Sawamoto, *Science*, 2013, **341**, 1238149.
- R. Szweda, *Prog. Polym. Sci.*, 2023, **145**, 101737.
- P. Nanjan and M. Porel, *Polym. Chem.*, 2019, **10**, 5406–5424.
- E. Laurent, R. Szweda and J.-F. Lutz, in *Synthetic Polymers with Finely Regulated Monomer Sequences: Properties and Emerging Applications*, JWS, Ltd, 2022, pp. 1–34.
- Q. Shi, Z. Deng, M. Hou, X. Hu and S. Liu, *Prog. Polym. Sci.*, 2023, **141**, 101677.
- Z. Deng, Q. Shi, J. Tan, J. Hu and S. Liu, *ACS Mater. Lett.*, 2021, **3**, 1339–1356.
- W. Forysiak, S. Kozub, Ł. John and R. Szweda, *Polym. Chem.*, 2022, **13**, 2980–2987.
- M. Szatko, W. Forysiak, S. Kozub, T. Andruniów and R. Szweda, *ACS Biomater. Sci. Eng.*, 2024, **10**, 3727–3738.
- K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, *PLoS One*, 2012, **7**, 1–6.
- M. U. Rahman, A. U. Rehman, H. Liu and H.-F. Chen, *J. Chem. Inf. Model.*, 2020, **60**, 4912–4923.
- P. Kührová, A. De Simone, M. Otyepka and R. Best, *Biophys. J.*, 2012, **102**, 1897–1906.
- A. Wacha, Z. Varga and T. Beke-Somfai, *J. Chem. Inf. Model.*, 2023, **63**, 3799–3813.



- 34 S. J. Rukmani, G. Kupgan, D. M. Anstine and C. M. Colina, *Mol. Simul.*, 2019, **45**, 310–321.
- 35 R. Nkpesu Mbitou and F. Bedoui, *Comput. Mater. Sci.*, 2025, **253**, 113861.
- 36 Y. I. Yang, Q. Shao, J. Zhang, L. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2019, **151**, 070902.
- 37 R. Qi, G. Wei, B. Ma and R. Nussinov, in *Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example*, ed. B. L. Nilsson and T. M. Doran, Springer New York, New York, NY, 2018, pp. 101–119.
- 38 G.-F. Hao, W.-F. Xu, S.-G. Yang and G.-F. Yang, *Sci. Rep.*, 2015, **5**, 15568.
- 39 S. Samokhvalova, J.-F. Lutz and I. Coluzza, *Macromol. Chem. Phys.*, 2024, **225**, 2400223.
- 40 D. Dellemme, S. Kardas, C. Tonneaux, J. Lernoold, M. Fossépré and M. Surin, *Angew. Chem., Int. Ed.*, 2025, e202420179.
- 41 O. Carugo and K. Djinović-Carugo, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2013, **69**, 1333–1341.
- 42 A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé and A. Laio, *Chem. Rev.*, 2021, **121**, 9722–9758.
- 43 B. E. Husic and V. S. Pande, *J. Am. Chem. Soc.*, 2018, **140**, 2386–2396.
- 44 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
- 45 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 33.
- 46 A. W. Sousa da Silva and W. F. Vranken, *BMC Res. Notes*, 2012, **5**, 367.
- 47 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 48 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.
- 49 C. Caleman, P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa and D. van der Spoel, *J. Chem. Theory Comput.*, 2012, **8**, 61–74.
- 50 E. Lindahl, M. Abraham, B. Hess and D. van der Spoel, GROMACS 2021 Manual, 2021. DOI: [10.5281/zenodo.4457591](https://doi.org/10.5281/zenodo.4457591).
- 51 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
- 52 M. Bernetti and G. Bussi, *J. Chem. Phys.*, 2020, **153**, 114107.
- 53 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.
- 54 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
- 55 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
- 56 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 57 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 58 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 59 P. S. Foundation, *C API Reference Manual*, 2021, <https://docs.python.org/3.10/c-api/index.html#c-api-index>.
- 60 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.
- 61 J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 90–95.
- 62 The pandas development team, *pandas-dev/pandas: Pandas*, 2020, Zenodo. DOI: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134).
- 63 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, *Nat. Methods*, 2020, **17**, 261–272.
- 64 A. F. Perez Mellor, J. Brazard, S. Kozub, T. Bürgi, R. Szweda and T. B. M. Adachi, *J. Phys. Chem. A*, 2023, **127**, 7309–7322.
- 65 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
- 66 M. Ester, H.-P. Kriegel, J. Sander and X. Xu, Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, 1996, pp. 226–231.
- 67 G. G. Maisuradze, A. Liwo and H. A. Scheraga, *J. Mol. Biol.*, 2009, **385**, 312–329.
- 68 Schrödinger, LLC, Open-source build, available at <https://pymol.org>.
- 69 R. Bauernschmitt and R. Ahlrichs, *Chem. Phys. Lett.*, 1996, **256**, 454–464.
- 70 M. E. Casida, C. Jamorski, K. C. Casida and D. R. Salahub, *J. Chem. Phys.*, 1998, **108**, 4439–4449.
- 71 R. E. Stratmann, G. E. Scuseria and M. J. Frisch, *J. Chem. Phys.*, 1998, **109**, 8218–8224.
- 72 C. Van Caillie and R. D. Amos, *Chem. Phys. Lett.*, 1999, **308**, 249–255.
- 73 C. Van Caillie and R. D. Amos, *Chem. Phys. Lett.*, 2000, **317**, 159–164.
- 74 F. Furche and R. Ahlrichs, *J. Chem. Phys.*, 2002, **117**, 7433–7447.
- 75 G. Scalmani, M. J. Frisch, B. Menucci, J. Tomasi, R. Cammi and V. Barone, *J. Chem. Phys.*, 2006, **124**, 094107.
- 76 T. Helgaker and P. Jørgensen, *J. Chem. Phys.*, 1991, **95**, 2595–2601.
- 77 K. L. Bak, P. Jørgensen, T. Helgaker, K. Ruud and H. J. A. Jensen, *J. Chem. Phys.*, 1993, **98**, 8873–8887.
- 78 K. L. Bak, A. E. Hansen, K. Ruud, T. Helgaker, J. Olsen and P. Jørgensen, *Theor. Chim. Acta*, 1995, **90**, 441–458.
- 79 J. Olsen, K. L. Bak, K. Ruud, T. Helgaker and P. Jørgensen, *Theor. Chim. Acta*, 1995, **90**, 421–439.



- 80 A. Hansen and K. Bak, *Enantiomer*, 1999, **4**, 455–476.
- 81 J. Autschbach, T. Ziegler, S. J. A. van Gisbergen and E. J. Baerends, *J. Chem. Phys.*, 2002, **116**, 6930–6940.
- 82 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
- 83 A. D. McLean and G. S. Chandler, *J. Chem. Phys.*, 1980, **72**, 5639–5648.
- 84 R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650–654.
- 85 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 86 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3094.
- 87 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian, ~16 Revision C.01*, Gaussian Inc., Wallingford, CT, 2016.
- 88 E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris and T. E. Ferrin, *Protein Sci.*, 2023, **32**, e4792.
- 89 A. J. Schaefer, V. M. Ingman and S. E. Wheeler, *J. Comput. Chem.*, 2021, **42**, 1750–1754.
- 90 P. Cwynar, P. Pasikowski and R. Szweda, *Eur. Polym. J.*, 2023, **182**, 111706.
- 91 W. Forysiak, A. Lizak and R. Szweda, *ChemPhysChem*, 2024, **25**, e202400366.

