JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

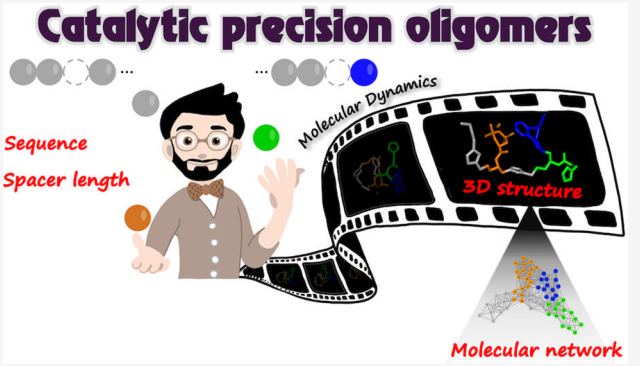
Revealing the Organization of Catalytic Sequence-Defined Oligomers via Combined Molecular Dynamics Simulations and Network Analysis

Sinan Kardas, Mathieu Fosséprés,* Vincent Lemaur, Antony E. Fernandes, Karine Glinel, Alain M. Jonas, and Mathieu Surin*

Cite This: J. Chem. Inf. Model. 2022, 62, 2761−2770 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

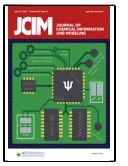
ABSTRACT: Similar to biological macromolecules such as DNA and proteins, the precise control over the monomer position in sequence-defined polymers is of paramount importance for tuning their structures and properties toward achieving specific functions. Here, we apply molecular network analysis on three-dimensional structures issued from molecular dynamics simulations to decipher how the chain organization of trifunctional catalytic oligomers is influenced by the oligomer sequence and the length of oligo(ethylene oxide) spacers. Our findings demonstrate that the tuning of their primary structures is crucial for favoring cooperative interactions between the catalytic units and thus higher catalytic activities. This combined approach can assist in establishing structure−property relationships, leading to a more rational design of sequence-defined catalytic oligomers via computational chemistry.



INTRODUCTION

Nature has the amazing ability to achieve perfect control of monomer sequences in biological macromolecules. Such chemical precision leads to sophisticated structures with remarkable functions such as catalysis, molecular recognition, and information storage.[1−3] Inspired by the precision of biological macromolecules, the design of synthetic sequence-defined polymers constitutes an emerging topic of interest to generate materials with next-generation performances.[4−16] However, despite considerable progress in controlling the primary structure, the by-design folding and assembly of synthetic polymers into predetermined three-dimensional (3D) shapes have not been achieved so far. Experimental methods such as X-ray crystallography have been used to solve the 3D structures of biological macromolecules at the atomic scale, providing valuable insights into their functions.[17−19] Over 175 000 protein structures are now stored in the Protein Data Bank, which has boosted the capability to predict 3D structures by the development of structural bioinformatics methods (homology modeling and fold recognition techniques) or, more recently, the training of neural networks (AlphaFold 2 from DeepMind). However, even minor modifications of the primary structure could strongly affect the function of a protein in some cases, and hundreds of millions of 3D structures of proteins remain to be discovered. Additionally, it remains challenging to obtain information on the dynamics of macromolecules and their assemblies. In this context, molecular dynamics (MD) simulations offer an attractive method to explore at atomistic resolution the dynamics of macromolecules from the femtosecond to the millisecond timescale.[20−23] As the cost of computing has decreased over the past several years, MD simulations have become more and more common to tackle larger molecular systems, leading to the development of new tools to efficiently analyze large data sets. Among these tools, network theory, already applied to a broad variety of systems, from computer science to physics and from social science to biology,[24−32] has helped in understanding complex molecular structures, such as proteins.[33−35] In this reductionist approach, the complex 3D structures of proteins and macromolecules are abstracted into a set of nodes corresponding to atoms (or groups of atoms) interconnected by edges, with each edge representing the (covalent or noncovalent) interactions between the nodes.[36] Combining network theory with 3D structural information obtained from MD simulations is expected to provide insights into the structure and dynamics of macromolecules.[19] For

Received: January 27, 2022
Published: May 24, 2022



https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761−2770

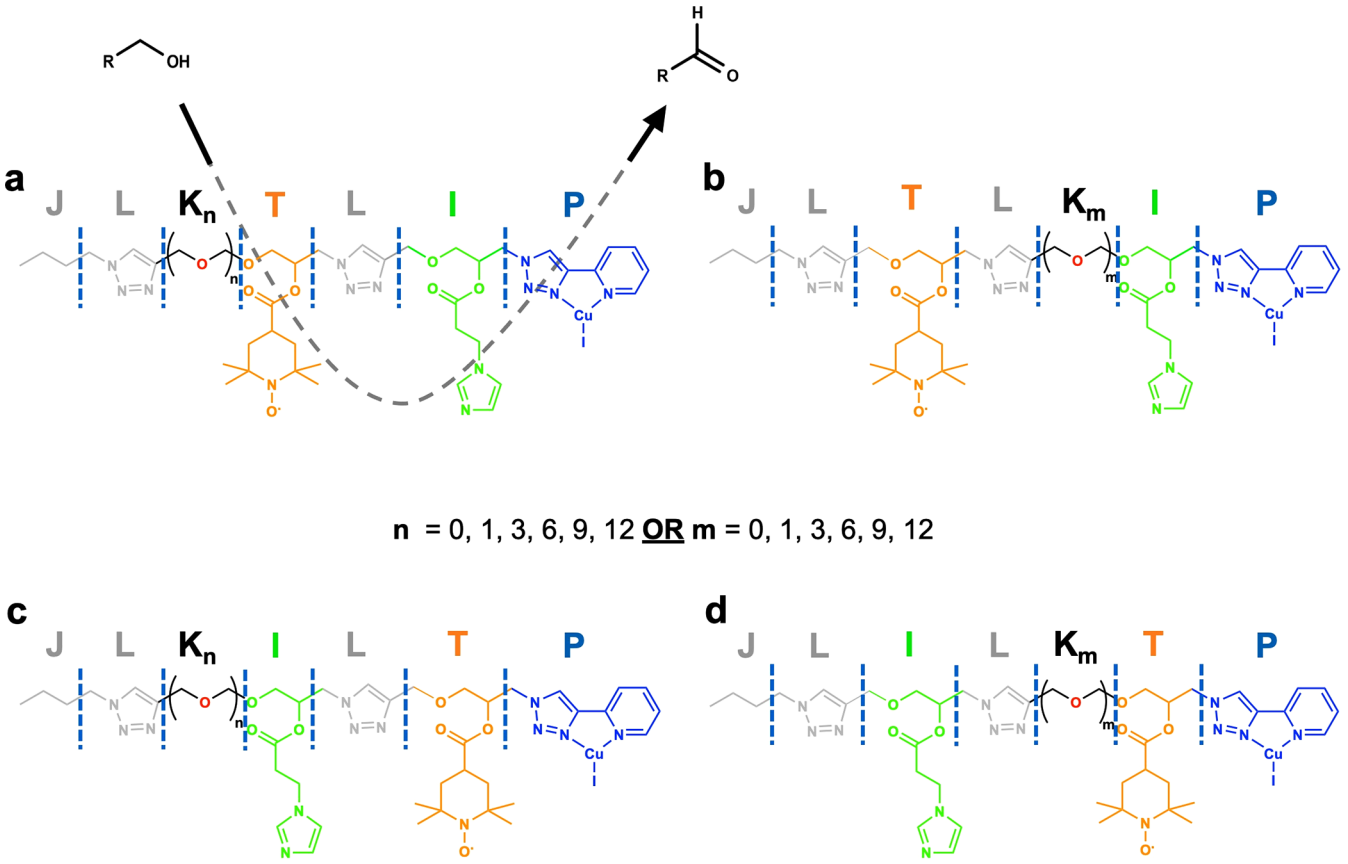


Figure 1. General structures of the catalytic oligomers (**a**) [JLK$_n$TLIP], (**b**) [JLTLK$_m$IP], (**c**) [JLK$_n$ILTP], and (**d**) [JLILK$_m$TP]. The substrate and product of the catalytic reaction are sketched in Figure 1a. The one-letter code is associated with the fragments used to build the molecular models. These oligomers incorporate a catalytic triad comprising TEMPO (**T**), NMI (**I**), and a pyta-Cu^{I} (**P**) and noncatalytic groups comprising a propyl group (**J**) and triazole derivatives (**L**). The arrangement of the three catalytic units **T**, **I**, and **P** is tuned by their positioning in the sequence and their relative spacing through the insertion of oligo(EG) spacers before (**K**$_n$) or in-between (**K**$_m$) the catalytic monomers (n, m = 0, 1, 3, 6, 9, and 12).

instance, Giuliani et al. utilized a network representation of protein structures to identify structural modules of highly connected nodes within and between the regions of the networks.[37] Aier et al. combined molecular docking simulations and residue interaction network analysis to provide insights into the ligand binding mechanism of the EZH2 protein and the conformational effects following specific mutations.[38] In a recent work, we combined MD simulations with network-based approaches at the single-chain level to rationalize the differences in experimental catalytic activities of short-sequence-defined oligomers.[39]

Here, we exploit this approach to the study of other synthetic precision oligomers displaying cooperative catalytic activity.[40−42] These oligomers incorporate a catalytic triad comprising 2,2,6,6-tetramethyl-1-piperidinyloxyl (TEMPO, **T**), *N*-methylimidazole (NMI, **I**), and a pyridyltriazole-Cu^{I} complex (pyta-Cu^{I}, **P**) that synergistically catalyze the aerobic oxidation of alcohols to aldehydes (as sketched in Figure 1a), and noncatalytic groups comprising a propyl group (**J**) and triazole derivatives (**L**), see Figure 1. Although being a very important and timely reaction, the mechanism of the Cu/TEMPO-catalyzed aerobic oxidation of alcohols is still the subject of study and debate. The main elementary steps of the catalytic cycle are well elucidated, however, the reaction pathways and the exact nature of some of the intermediate species are not well ascertained and, moreover, they appear to be substrate dependent, which makes the reaction even more complex. It is, however, obvious that this multicatalytic reaction displays a high degree of cooperativity to achieve significant turnovers and therefore requires an optimal and synergistic orchestration of the catalytic triad **T**, **I**, and **P**.[43−45] From our current understanding, the rate-determining step of the reaction appears mainly to be the catalyst oxidation step (i.e., Cu(I) to Cu(II) oxidation and subsequent TEMPO oxidation) itself, rather than the substrate oxidation. The former reaction requires the formation of an active (pyta)Cu-(NMI) complex that enables the oxygen-mediated oxidation of the Cu species while having the TEMPO site in close proximity. Hence, it is important to design dynamic systems capable of selectively promoting these interactions, especially between the (pyta)Cu and NMI sites − a problem that we propose to address through the rational design of flexible sequence-defined oligomers. In this view, the interactions between the three catalytic units can be tuned by their precise positioning in the sequence and by adjusting their spacing through the insertion of a specific number of ethylene glycol (EG) spacer units before (**K**$_n$) or in-between (**K**$_m$) the catalytic monomers (n, m = 0, 1, 3, 6, 9, and 12 EG units) (Figure 1). Here, we focused on trifunctional oligomers [JLK$_n$TLK$_m$IP] and [JLK$_n$ILK$_m$TP] that differ in the position of the imidazole group (**I**) relative to the pyta-Cu^{I} (**P**) center, which has been experimentally demonstrated to be critical to enabling

https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761−2770

cooperative interactions between the three catalytic groups, leading to higher catalytic activity.[43] Hence, studying the sequence effects is of paramount importance to guide the design of this type of functional oligomer.[43−47] Based on these previous works, we used a computational approach to decipher the structure and dynamics of these sequence-defined catalytic oligomers as a function of the oligomer sequence and the spacer length at the single-chain level.

METHODS

MD Simulations. All-atom MD simulations were performed using the GPU version of the AMBER16 package.[48] The oligomers were built within the LEaP program by joining fragments constructed within Discovery Studio 4.0.[49] The partial atomic charges of each fragment were calculated using the semiempirical AM1-BCC model[50,51] as implemented within the antechamber module[52] of AMBER16, whereas other force field parameters are from the General AMBER Force Field (GAFF2).[53] The structural parameters and charges of the TEMPO fragment were calculated through an ab initio re-parametrization reported by Stendardo et al.[54] Quantum chemical computations were used to obtain geometric parameters and partial atomic charges of the pyridyltriazole (pyta)-Cu^{I} complex (pyta-Cu^{I}) using density functional theory at the B3LYP/6-31G** level. Extra basis sets LanL2DZ and MIDI! were applied to the copper(I) and the iodine atoms, respectively. Molecular mechanics calculations were performed to optimize the geometry of the entire oligomer. For this, 10 000 steps of energy minimization of the system were performed, consisting of 1000 steps of the steepest descent algorithm, followed by 9000 cycles of the conjugate gradient. After this initial energy minimization phase, the oligomers underwent an MD production phase for 500 ns. In order to conduct relevant conformational sampling on the submicrosecond timescale with reasonable computational resources, the MD simulations were achieved in an implicit solvent using the Generalized Born solvation model.[55] A dielectric constant of 36.0 was considered, that is, the dielectric constant of acetonitrile at room temperature. The temperature in all the simulations was set to 300 K and controlled by the Langevin thermostat with a coupling constant of 1.0 ps and combined with a pseudo-random seed generator.[56] Because the systems contain a subatomic particle, that is, a nitroxide radical, no constraints were employed on bond lengths and an integration time step of 1.0 fs was used. Harmonic positional restraints with a force constant of 25 kcal·mol^{-1}·$Å^{-2}$ were applied to the carbon atom of the first methyl fragment of the sequences to mimic the surface grafting of the chains. Frames were collected at 1 ns intervals for a total simulation length of 500 ns, resulting in a set of 500 conformations. To ensure that the MD simulations reached equilibrium and that there are no large conformational changes observed longer, the root-mean-square deviations (rmsds) were calculated (Table S1). Similar average rmsd values for the different sequences and small standard deviation (SD) values (<1 Å) indicate that the simulated systems are stable on a 500 ns timescale. Three replica simulations were performed for all the sequences, each lasting 500 ns (see the Supporting Information for details). Replicas differ only in terms of the random number of seeds selected to assign initial atom velocities. Then, the relationship between the replicas was analyzed by computing the average value of the distances between the catalytic units over all the conformations of each replica (Table S4). Similar average distance values and small SD values (<1 Å) for the different replicas indicate the convergence and the reproducibility of the simulations. The resulting trajectories were visualized using the VMD software package[57] and rendering of MD snapshots was performed using PyMOL.[58] The analyses of the trajectories were performed using the *cpptraj* module available in AmberTools (see the Supporting Information for details).

Molecular Network Analysis. In order to build the oligomer networks from the last conformation of the MD simulations, in-house scripts were used to prepare the network files. For this, each heavy atom (omission of hydrogen atoms and lone pairs) was assigned to only one node, and a cutoff of 5 Å was used to define the distance below which two nodes are connected by an edge. All the edges in the networks are undirected and unweighted; that is, the relationship between the nodes is a simple connection, without a given flow implied and without considering the strength of interactions. The network files were imported within Cytoscape software 3.8.0[59] for subsequent analysis and visualization. The global network measures were obtained with the Network Analyzer plugin of Cytoscape (see the Supporting Information for details).

Modularization of the Molecular Networks. There are several network modularization techniques to study the properties of complex networks.[60] Among them, the flow-based methods rely on the communication patterns of the networks rather than their structural properties to detect the modules of highly connected nodes. In this study, modules were detected using the so-called INFOMap algorithm,[61] a well-known method for detecting flow-based modules based on the compression of information flow in the network. INFOMap minimizes an objective function (referred to as the map equation), which measures how much a modular partition of the network can compress a description of flow in the network. A particular assignment of nodes into modules is called a network partition. The description of flow is modeled by a random walker on the network that moves across nodes via edges. Hence, the trajectory of the walker depends on the hierarchical structure of the network and tends to stay longer in dense areas representing the modules of highly connected nodes (Figure S3). For a given partition of the network, there is an associated information cost (measured in bits) for describing the movements of the random walker. The map equation converts the flow rates within and between the modules to an information-theoretic modular description measure of the random walker's movements on the network. Optimizing the map equation over possible network partitions corresponds to detecting the partition that minimizes the description length of the dynamics of the random walker along the network.[62] The trajectory of the walker should, in principle, be limited to the link structure. However, to ensure that the trajectory is an ergodic solution (i.e., that the average visit rates reach a steady-state solution independent of where the random walker starts and of the length of the trajectory), from each node the random walker has a teleportation rate of 15%.[63] Modules consist of nodes among which flow persists, and the nodes that are highly connected are the ones through which information flows quickly and easily. Here, the modularization of the oligomer networks was obtained with the stand-alone version of INFOMap 1.6.0. The modular networks were considered undirected and the standard value of 15% for the teleportation rate was used. A two-level module description was considered, meaning that INFOMap only returns one layer of modules which are partitioned to minimize the map

https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761−2770

equation. The number of outer-most loops to run before picking the best solution was specified to 10. The web server version of INFOMap was used for the visualization and analysis of the modular networks (see the Supporting Information for details).

RESULTS AND DISCUSSION

Catalytic Units Are Brought into Close Proximity by the Folding Effect of the Spacers. All-atom MD simulations were carried out in an implicit acetonitrile solvent to evaluate the effect of the sequence and of the spacer length on the 3D structure and dynamics of the single-chain oligomers. The flexibility of the oligomers was explored via the calculation of per-monomer root-mean square fluctuations (RMSFs) (Figure 2a–c for RMSF profiles and Table S2 for statistics). For short spacers (<6 EG units), the flexibility of the catalytic segments is related to their position in the sequence. As the spacer length increases before ($\mathbf{K}_n$) or in-between ($\mathbf{K}_m$) the catalytic units, the differences in the flexibility of the catalytic fragments progressively decrease, except for **T** in the [**JLTLK$_m$IP**] sequence (red line in Figure 2a). For longer spacers (>6 EG units), the flexibility profiles are comparable, indicating that the length of the spacer does not influence the intrinsic flexibility of the catalytic monomers. Thus, the results suggest that for short spacers (<6 EG units), interchanging two catalytic units along the sequence strongly impacts the local flexibility of the catalytic units, leading to significantly higher differences in flexibility compared to longer spacers (>6 EG units).

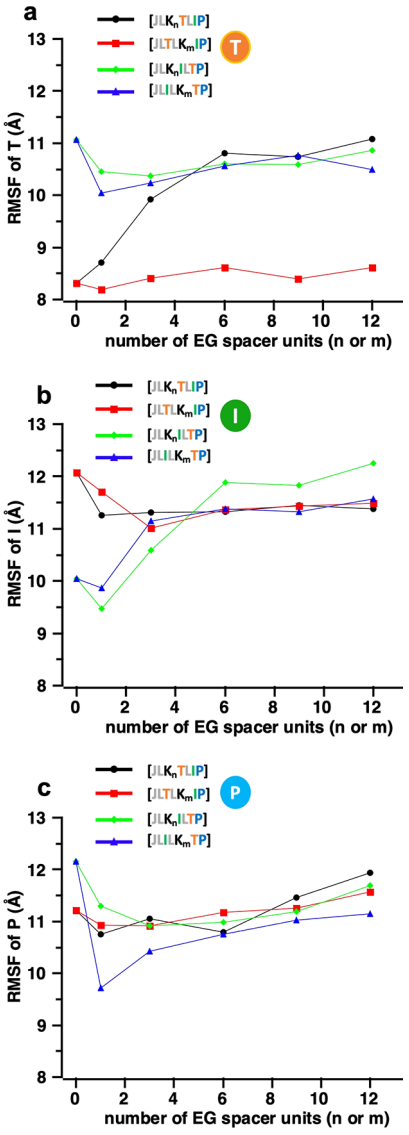


Figure 2. RMSF profiles of the three catalytic monomers (**a**) **T**, (**b**) **I**, and (**c**) **P** as a function of the oligomer sequence and the spacer length $\mathbf{K}_n$ or $\mathbf{K}_m$.

To determine if such different flexibility profiles could influence the compactness of the sequence-defined oligomers, we computed their radii of gyration (R_g) (see the Supporting Information). Both the stability of R_g profiles over simulation time and the unimodal sharp distribution indicate that the average and global size of the oligomers is preserved on the submicrosecond timescale. Moreover, the global size of the macromolecular chains is barely influenced by the oligomer sequence and the spacer length. We observed a peak value around 5–6 Å regardless of the oligomer sequence (Figure S4), and similar average R_g values, from *ca.* 6 Å for shorter ones, to *ca.* 7 Å for longer ones (Table S3). Hence, the oligomers are folded into a compact structure, as driven by the interactions between **T**/**I**/**P**/**L** units, and thanks to the flexibility of the EG spacers ($\mathbf{K}_n$ and $\mathbf{K}_m$). However, it is important to note that R_g values only indicate the average and global size of the oligomers, while no information about their shape such as the deviation from sphericity is included (see the Supporting Information for details).

As similar compactness does not mean that the internal organization of the chains is conserved, we followed the time evolution of the distances between the three catalytic units **T**, **I**, and **P** as an indicator of their proximity during simulations—a crucial piece of information for the cooperative catalytic activity of the oligomeric chains (Figure 3a and Supporting Information). Although important oscillations were observed throughout the simulations, a statistical analysis of distance profiles indicates similar average values associated with low SD values compared to the range of values (i.e., the maximum values minus the minimum values), regardless of the oligomer sequence and the spacer length. These average distances in all the sequences range from 9 to 12.5 Å, with small SD values (<1 Å) for the three replica MD simulations (Table S4). Figure 3b shows the profiles of the average distances between the **T** and **P** units as a function of the oligomer sequence and the spacer length. For the [**JLK$_n$TLIP**], [**JLK$_n$ILTP**], and [**JLILK$_m$TP**] sequences (black, green, and blue solid lines, respectively), the average distance between the **T** and **P** catalytic units does not evolve as the **T** and **P** units are close in the sequence. Surprisingly, increasing the spacer length between the **T** and **P** units from 0 to 12 EG units in the [**JLTLK$_m$IP**] sequences (red dashed line in Figure 3b) does not strongly impact their spatial proximity in the 3D structure, even for the longer oligomers (Figure 3c). The average distance is increased by only ~1.5 Å while it would increase to ~12 × 4 Å (4 Å = approximate length of one elongated EG unit) if the oligomer chain is fully elongated. The two catalytic units that are far apart in the sequence remain close in the 3D structure due to the flexibility of the EG spacers and the interactions between **T**/**I**/**P**/**L** units in these solvation conditions. We observed the same tendencies for the average distances between **I** and **P** and between **T** and **P**, where the spatial proximity of the catalytic units was not influenced by

https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761–2770

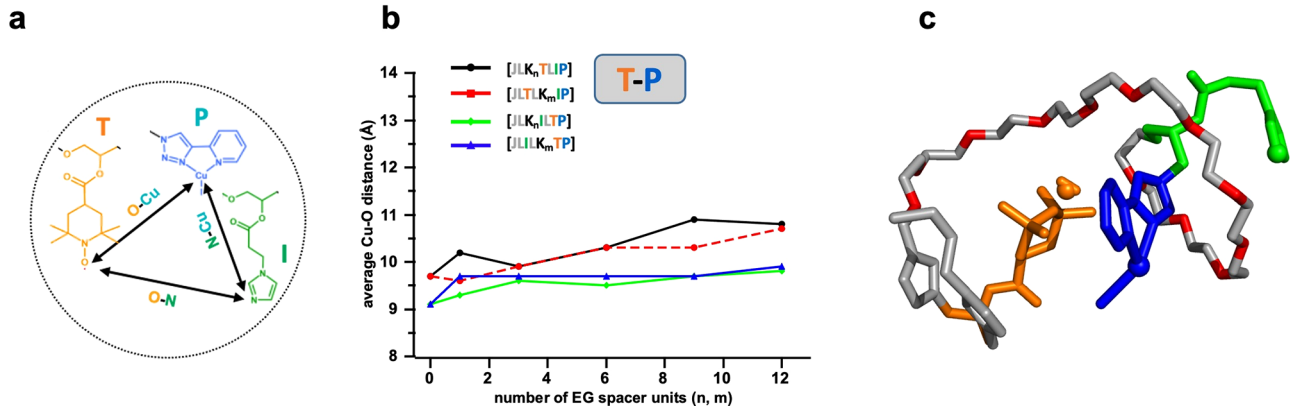


Figure 3. (**a**) Scheme showing the selected atoms to compute the average distances between the three catalytic units, that is, O•, N, and Cu for **T**, **I**, and **P** units, respectively. (**b**) Profiles of the average distances between **T** and **P** units for the different sequences. (**c**) Final MD snapshot (at 500 ns) of the [**JLTLK**$_{12}$**IP**] sequence with 12 EG spacer units between **T** and **I**/**P**. **T**, **I**, and **P** catalytic groups are depicted in orange, green, and blue, respectively. The selected atoms to compute the average distances between the three catalytic units are represented by small spheres. The noncatalytic groups (**J** and **L**) are represented in gray, and the oxygen atoms of the EG spacer units (**K**$_n$ and **K**$_m$) are in red (in stick representation).

their positioning in the sequence (Figure S5). The folding of the chains brings the three catalytic units into close proximity, even though they may be far apart in the sequence (primary structure). To gain insights into the relative positions of the three catalytic units, their collective dynamic behavior was analyzed as the vertices of a triangle formed by the representative atoms of the three catalytic units (see the Supporting Information and Figure S1 for details). For the different sequences, the distribution profiles of the area of the triangle along the simulations indicate a unimodal sharp distribution with a peak value around 40 Å^2 (Figure S6 and Table S5). Thus, the different sequences have similar "catalytic geometries" regardless of the spacer length and the oligomer sequence. The introduction of the flexible oligo(ethylene oxide) spacers before (**K**$_n$) or in-between (**K**$_m$) the catalytic groups in the oligomer chains does not alter the proximity and the geometry of the three catalytic moieties.

Altogether, these results indicate that the compactness of the oligomer chains and the proximity and flexibility of the catalytic moieties do not strongly depend on the oligomer sequence nor on the spacer length in the different sequences tested here. Nevertheless, the three catalytic units can potentially be hindered by the long and flexible EG spacers, whereas they should be accessible to each other to act cooperatively. In the next section, we investigate the overall intramolecular organization of the oligomers to get insights into the spatial arrangement of the catalytic units in the 3D native oligomer structures.

Flexible EG Spacers Influence the Folding of Catalytic Oligomers. The solvent accessible surface area (SASA) profiles of the three catalytic units as a function of the oligomer sequence and the spacer length were calculated to investigate their degree of burial into the 3D structures (Figure S7). The average solvent accessibility of the catalytic units is not much influenced by the oligomer sequence, but increasing the spacer length resulted in higher SASA values. On average, the oligomers gained around 5000 Å^2 of solvent accessibility going from 0 to 12 EG units before (**K**$_n$) or in-between (**K**$_m$) the catalytic units, leading the catalytic units to be more exposed to their chemical environment. This observation is expected from the theory of random coils, for which the average coil density decreases with chain length. To understand how the three catalytic units are arranged in the complex 3D structures of oligomers, a network-based analysis (see the Supporting Information for details) was applied. Based on the rmsd convergence, the last conformations of the MD simulations were used to construct the molecular networks. We analyzed four network parameters as topological descriptors of the oligomer structures; that is, the average degree, the network density, the characteristic path length, and the modularity (see the Supporting Information for details). Because the average distances between the catalytic units only differ by small SD values (<1 Å, Table S4) for the three replicas of the different sequences, the cutoff used to construct the molecular networks (5 Å) is relevant to represent the connectivity of atoms/nodes within the oligomers.

In these networks, the degree of a node is the number of connections it has to other nodes. The average degree $\langle k \rangle$ is the degree averaged over all the nodes in the network and reflects the connectivity of the network. The oligomer networks have an average degree $\langle k \rangle$ ranging from 12 to 18, depending on the sequence and the spacer length (Figure S8). More interestingly, the shape and the peak values of the degree distributions depend on the oligomer sequence and the spacer length. Hence, as the spacer length increases before or in-between the catalytic fragments, the node degrees exhibit broad and heavy-tailed distributions (Figure S9 and Table S6), meaning that the nodes present on average a lower degree of connectivity. Moreover, the fraction of highly connected nodes (nodes with high degrees) depends not only on the spacer length but also on the position of the EG spacers in the sequence. For example, for [**JLILK**$_m$**TP**] sequences, the distributions of the node degrees become gradually narrower when increasing the spacer length (**K**$_m$) between the catalytic units, unlike for the other sequences, where the degree of connectivity in the oligomer networks becomes more heterogeneous when increasing the spacer length (**K**$_n$ or **K**$_m$). These results show that the sequence and the spacer length strongly influence the internal organization of the oligomers and thus the local properties of the molecular networks. More specifically, the degrees of the catalytic nodes are especially influenced by the spacer length and the sequence (Figure S10). Interestingly, the network density ρ(G) of oligomers decreases by around 10% as the number of EG

https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761−2770

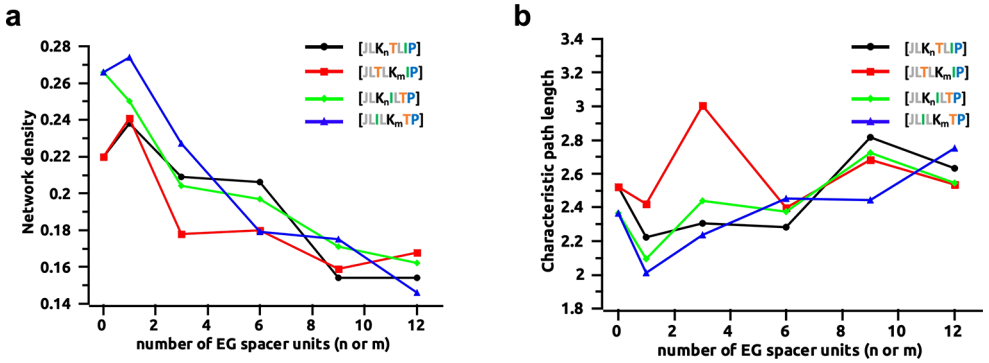


Figure 4. (**a**) Evolution of the network density ρ(G) as a function of the number of EG spacer units (n or m) for the different sequences. The network density decreases as the number of EG spacer units (n or m) increases. (**b**) Characteristic path length (l) as a function of the number of EG spacer units (n or m) for the different sequences.

spacer units increases (Figure 4a), with values ranging from 0.24 to 0.15 for [**JLK$_n$TLK$_m$IP**] sequences, and from 0.27 to 0.15 for [**JLK$_n$ILK$_m$TP**] sequences. Hence, while the sequence barely influences the global size of the oligomeric chains and the proximity of catalytic units (vide infra), the folding clearly depends on the sequence of the oligomers. Both the decrease of ρ(G) and the heterogeneity of degree distribution with the spacer length indicate that some parts of the oligomers are less connected to each other. The smaller density values for longer oligomers are also related to higher SASA values, indicating that the oligomers are more exposed to the surrounding environment (see Figures S12 and S13). Furthermore, we calculated the characteristic path length l for the networks of the sequence-defined oligomers (Figure 4b) to measure how information spreads from a given node to other ones in the network. Increasing the number of EG spacer units resulted in a higher characteristic path length, in correlation with the decrease in the network density and the fact that some regions of the oligomers are more isolated when increasing the spacer length.

Altogether, these results suggest that the oligomer networks are organized into separate modules. Therefore, we deciphered the modularity of the oligomer networks to identify highly connected regions within the oligomers and thus the ability of the three catalytic units to act cooperatively. We modularized the networks by using the so-called INFOMap method,[61] an algorithm based on the compression of information flow in the networks (see the Supporting Information for details). For all the oligomer networks, the number of modules is similar, ranging from 3 to 5 modules, regardless of the sequence and the spacer length (Table S7). However, the size (i.e., the number of nodes in the modules) and the composition of the modules depend on the sequence and the spacer length. The number of nodes in the catalytic modules **T**, **I**, and **P** (i.e., modules composed of nodes corresponding to the catalytic fragments) ranges from 10 to 40 as a function of the sequence and the spacer length (Figures S11 and 5). The number of nodes in the **T** module (Figure S11a) decreases for sequences with three EG spacer units, indicating that the **T** fragment is less connected to other fragments, but further increasing the spacer length results in larger **T** modules. We observed the opposite situation for the **I** module: the number of nodes in the **I** module (Figure S11b) increases for three EG spacer units, meaning that the nodes of the **I** module are more connected to the nodes of other modules, whereas the size of the **I** module decreases when increasing spacer length, indicating that the **I** unit is not hindered by other fragments of the oligomers. The number of nodes in the **P** module (Figure S11c) varies with the sequence and the spacer length, but when the flexibility of the oligomeric chains increases with the spacer length, the catalytic modules are allowed to interact with each other and belong to the same modules (Figures 5 and S12 and S13).

Concerning the composition of the modules, introducing an increasing number of EG units before the catalytic units in the [**JLK$_n$TLIP**] sequences (Figure S12a–f) results in highly connected regions in the networks. For sequences with short EG units, the catalytic modules are close and highly interconnected, as shown by the thick lines between the catalytic modules. For longer EG spacers, the catalytic modules remain highly connected, and the nodes of the spacers can be organized into distinct modules or be part of the catalytic modules without altering the interactions between the catalytic modules. This suggests that increasing the spacer length (**K$_n$**) before the **T** unit does not strongly hinder the accessibility of catalytic units to each other. For the [**JLTLK$_m$IP**] sequences (Figures S12g–k), the catalytic modules remain highly interconnected when increasing the spacer length between **T** and **I**/**P** (Figure S11). Such interconnectivity is due to the flexibility of the chains that allows the **T** unit to interact with the **I**/**P** units. For the [**JLILTP**] sequence ($n = m = 0$) (Figure 5a), while the **T** and **P** nodes are obviously in the same module, the **I** nodes are connected to the **T**/**P** module by a few intermediate nodes corresponding to the noncatalytic groups (**J** and **L**). Although the size and the composition of the **T**/**P** module are not influenced by the introduction of the long and flexible EG spacer before the catalytic units in the [**JLK$_{12}$ILTP**] sequence, the number of spacer nodes in the **I** module increases (Figures 5b and S13a–f for other spacer lengths). This indicates that the long and flexible spacer prevents the interaction between the **I** unit and the other catalytic fragments. Finally, with a long EG spacer between **I** and **T**/**P** units in the [**JLILK$_{12}$TP**] sequence (Figures 5c and S13g–k for other spacer lengths), it allows the **I** module to interact directly with the **T** and **P** modules, as shown by the higher flow rate between the catalytic modules. These findings show that the optimal relative arrangement of the catalytic units and the nearby chemical environment results from a subtle balance between the flexibility of the oligomeric chains and the oligomer sequence that regulates the interactions between the different fragments.

https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761−2770

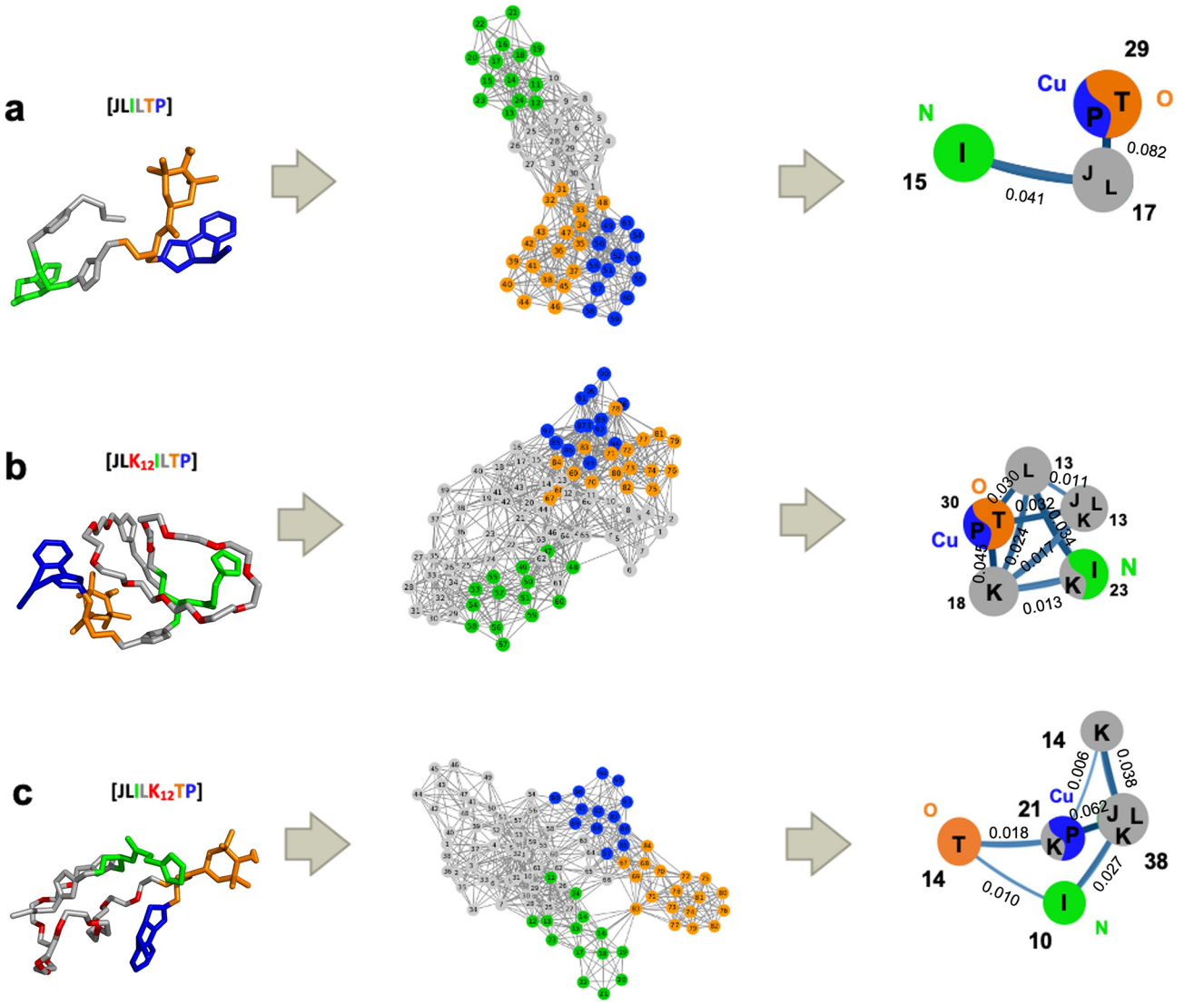


Figure 5. Left: final MD snapshots of the (**a**) [**JLILTP**], (**b**) [**JLK$_{12}$ILTP**], and (**c**) [**JLILK$_{12}$TP**] sequences (in stick representation). **I**, **T**, and **P** units are depicted in green, orange, and blue, respectively. The noncatalytic groups **J** and **L** are represented in gray and the oxygen atoms of the spacer units **K** are highlighted in red. **Middle:** molecular networks of the corresponding sequences (see the Supporting Information for details). Node colors correspond to the colors of the catalytic and noncatalytic units in the final MD snapshots shown on the left. **Right:** modularization of the oligomer networks for the corresponding sequences (see the Supporting Information for details). The letters of each fragment associated with the modules, and the representative atoms of the catalytic units, are shown. Module colors are related to the colors of the catalytic and noncatalytic units in the final MD snapshots shown on the left. The size of the modules (circles) is proportional to the number of nodes comprised in the modules, annotated next to the modules. The thickness of the edges (lines) is proportional to the degree of connectivity (decimal numbers) between the modules.

CONCLUSIONS

In this study, in-depth molecular insights into the organization and dynamics of a model system consisting of (pyta)Cu/TEMPO/NMI catalytic oligomers were obtained from a computational approach. The relative arrangement of the three catalytic units was tuned by their positioning in the sequence and the presence of flexible oligo(ethylene oxide) spacers of various lengths before or in-between the catalytic units. Our MD results showed that, with this specific flexible scaffold, the oligomers adopted a folded and compact conformation (i.e., small radii of gyration) and the average distances between the catalytic units were not influenced by the oligomer sequence nor by the spacer length for the different oligomers studied here; the catalytic units that are far apart in the primary structure remain in close proximity in the 3D structure. However, this simple proximity criterion between catalytic units and its convergence over all the three replica simulations were not sufficient to describe the chain organization and evaluate the sampling quality of the global conformational space (i.e., side-chain conformations and unfolding/folding events of the backbone). Hence, the differences in the topologies of the oligomers, as revealed by the molecular networks of the oligomer 3D structures, emphasized the effects of the oligomer sequence and the spacer length on the conformations of the oligomers. Our analyses suggest that the oligomer networks possess a modular structure and a heterogeneous topology. The catalytic modules may thus not directly interact with each other but via a few spacer modules, depending on the sequence and the spacer length of the oligomeric chains. These results demonstrate that

https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761−2770

folding is regulated by a balance between the flexibility of the chains and the position of the different units in the sequence, which can promote or hinder the interactions between the catalytic units, a key parameter for developing efficient catalysts with a precise control of their sequence. Therefore, by changing the flexibility of the backbone (number/position of spacer units) and the position of the catalytic units, one could promote the interactions between the catalytic units even if they are far from each other in the sequence. This study constitutes the premises of an inverse design approach to identify the molecular parameters of importance in order to optimize the catalytic activity of oligomers. A direct comparison with experiments and assessment of structure−activity relationships is unfortunately not yet possible because this requires a substantial synthetic effort. Besides, further work should also focus on generalizing our approach to interchain interactions. Meanwhile, we believe that this study is helpful for the construction of a more complete view of the spatial organization and interconnections within one single complex molecule, which should contribute to the development of discrete macromolecular systems in which a precise spatial arrangement of functional units is required for optimum properties.

DATA AND SOFTWARE AVAILABILITY

The data generated and analyzed for this study including the starting structures, the input files, the trajectory files, the output files from *cpptraj* and network analyses, and the in-house scripts used to prepare the networks and module files are publicly available at https://github.com/SinanKardas/TIPnetworks. Software RedMD used to prepare the network files is an open source and available at http://bionano.cent.uw.edu.pl/software/redmd/. AmberTools is publicly available free of charge at https://ambermd.org. Cytoscape software is available free of charge at https://cytoscape.org. The standalone and web server versions of INFOMap are available at https://mapequation.org/infomap/.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00101.

Computational details for MD simulations and molecular network analysis (PDF)

AUTHOR INFORMATION

Corresponding Authors

Mathieu Fosséprè − *Laboratory for Chemistry of Novel Materials, Center of Innovation and Research in Materials and Polymers, University of Mons-UMONS, Mons B-7000, Belgium*; Email: mathieu.fossepre@umons.ac.be

Mathieu Surin − *Laboratory for Chemistry of Novel Materials, Center of Innovation and Research in Materials and Polymers, University of Mons-UMONS, Mons B-7000, Belgium*; orcid.org/0000-0001-8950-3437; Email: mathieu.surin@umons.ac.be, bemathieu.surin@umons.ac.be

Authors

Sinan Kardas − *Laboratory for Chemistry of Novel Materials, Center of Innovation and Research in Materials and Polymers, University of Mons-UMONS, Mons B-7000, Belgium; Institute for Complex Molecular Systems, Eindhoven University of Technology-TU/e, Eindhoven 5600 MB, The Netherlands*; orcid.org/0000-0002-4532-8160

Vincent Lemaur − *Laboratory for Chemistry of Novel Materials, Center of Innovation and Research in Materials and Polymers, University of Mons-UMONS, Mons B-7000, Belgium*

Antony E. Fernandes − *Institute of Condensed Matter and Nanosciences, Bio- and Soft Matter, Université catholique de Louvain-UCLouvain, Louvain-la-Neuve B-1348, Belgium; Certech, Seneffe B-7180, Belgium*; orcid.org/0000-0002-7993-2980

Karine Glinel − *Institute of Condensed Matter and Nanosciences, Bio- and Soft Matter, Université catholique de Louvain-UCLouvain, Louvain-la-Neuve B-1348, Belgium*; orcid.org/0000-0002-2000-0169

Alain M. Jonas − *Institute of Condensed Matter and Nanosciences, Bio- and Soft Matter, Université catholique de Louvain-UCLouvain, Louvain-la-Neuve B-1348, Belgium*; orcid.org/0000-0002-4083-0688

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00101

Author Contributions

The manuscript was written through the contributions of all the authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The collaboration between the research groups of Louvain and Mons was supported by the Fonds de la Recherche Scientifique—FNRS and the Fonds Wetenschappelijk Onderzoek—FWO under EOS project no 30650939. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under grant no. 2.5020.11 and by the Wallonia Region. K.G. and M.S. are senior research associates of the F.R.S.-FNRS. The authors thank Elif Kardas for her creative assistance with the TOC figure.

REFERENCES

(1) Badi, N.; Lutz, J.-F. Sequence Control in Polymer Synthesis. *Chem. Soc. Rev.* **2009**, *38*, 3383−3390.

(2) Lutz, J.-F.; Lehn, J.-M.; Meijer, E. W.; Matyjaszewski, K. From Precision Polymers to Complex Materials and Systems. *Nat. Rev. Mater.* **2016**, *1*, 16024.

(3) Rutten, M. G. T. A.; Vaandrager, F. W.; Elemans, J. A. A. W.; Nolte, R. J. M. Encoding Information into Polymers. *Nat. Rev. Chem.* **2018**, *2*, 365−381.

(4) Danilov, D.; Barner-Kowollik, C.; Wenzel, W. Modelling of Reversible Single Chain Polymer Self-Assembly: From the Polymer towards the Protein Limit. *Chem. Commun.* **2015**, *51*, 6002−6005.

(5) Zhang, Z.; Zeng, T.-Y.; Xia, L.; Hong, C.-Y.; Wu, D.-C.; You, Y.-Z. Synthesis of Polymers with On-Demand Sequence Structures via Dually Switchable and Interconvertible Polymerizations. *Nat. Commun.* **2018**, *9*, 2577.

(6) Yang, C.; Wu, K. B.; Deng, Y.; Yuan, J.; Niu, J. Geared Toward Applications: A Perspective on Functional Sequence-Controlled Polymers. *ACS Macro Lett.* **2021**, *10*, 243−257.

(7) Genabeek, B.; Lamers, B. A. G.; Hawker, C. J.; Meijer, E. W.; Gutekunst, W. R.; Schmidt, B. V. K. J. Properties and Applications of

Precision Oligomer Materials; Where Organic and Polymer Chemistry Join Forces. *J. Polym. Sci.* **2021**, *59*, 373−403 February..
(8) Aksakal, R.; Mertens, C.; Soete, M.; Badi, N.; Du Prez, F. Applications of Discrete Synthetic Macromolecules in Life and Materials Science: Recent and Future Trends. *Adv. Sci.* **2021**, *8*, 2004038.
(9) Steinkoenig, J.; Aksakal, R.; Du Prez, F. Molecular Access to Multi-Dimensionally Encoded Information. *Eur. Polym. J.* **2019**, *120*, 109260.
(10) Uvyn, A.; De Geest, B. G. Multivalent Antibody-Recruiting Macromolecules: Linking Increased Binding Affinity with Enhanced Innate Immune Killing. *ChemBioChem* **2020**, *21*, 3036−3043.
(11) Li, J.; Leclercq, M.; Fossepré, M.; Surin, M.; Glinel, K.; Jonas, A. M.; Fernandes, A. E. Discrete Multifunctional Sequence-Defined Oligomers with Controlled Chirality. *Polym. Chem.* **2020**, *11*, 4040−4046.
(12) Martens, S.; Holloway, J. O.; Du Prez, F. E. Click and Click-Inspired Chemistry for the Design of Sequence-Controlled Polymers. *Macromol. Rapid Commun.* **2017**, *38*, 1700469.
(13) Mertens, C.; Soete, M.; Ślęczkowski, M. L.; Palmans, A. R. A.; Meijer, E. W.; Badi, N.; Du Prez, F. E. Stereocontrolled, Multi-Functional Sequence-Defined Oligomers through Automated Synthesis. *Polym. Chem.* **2020**, *11*, 4271−4280.
(14) Holloway, J. O.; Van Lijsebetten, F.; Badi, N.; Houck, H. A.; Du Prez, F. E. From Sequence-Defined Macromolecules to Macromolecular Pin Codes. *Adv. Sci.* **2020**, *7*, 1903698.
(15) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *J. Am. Chem. Soc.* **2013**, *341*, 1238149.
(16) Perry, S. L.; Sing, C. E. 100th Anniversary of Macromolecular Science Viewpoint: Opportunities in the Physics of Sequence-Defined Polymers. *ACS Macro Lett.* **2020**, *9*, 216−225.
(17) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **1958**, *181*, 662−666.
(18) Shi, Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* **2014**, *159*, 995−1014.
(19) O'Rourke, K. F.; Gorman, S. D.; Boehr, D. D. Biophysical and Computational Methods to Analyze Amino Acid Interaction Networks in Proteins. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 245−251.
(20) Srivastava, A.; Nagai, T.; Srivastava, A.; Miyashita, O.; Tama, F. Role of Computational Methods in Going beyond X-Ray Crystallography to Explore Protein Structure and Dynamics. *Int. J. Mol. Sci.* **2018**, *19*, 3401.
(21) Christoforides, E.; Fourtaka, K.; Andreou, A.; Bethanis, K. X-ray crystallography and molecular dynamics studies of the inclusion complexes of geraniol in *β*-cyclodextrin, heptakis (2,6-di-O-methyl)-*β*-cyclodextrin and heptakis (2,3,6-tri-O-methyl)-*β*-cyclodextrin. *J. Mol. Struct.* **2020**, *1202*, 127350.
(22) Kumar, R.; Moche, M.; Winblad, B.; Pavlov, P. F. Combined X-Ray Crystallography and Computational Modeling Approach to Investigate the Hsp90 C-Terminal Peptide Binding to FKBP51. *Sci. Rep.* **2017**, *7*, 1−17.
(23) Nury, H.; Poitevin, F.; Van Renterghem, C.; Changeux, J.-P.; Corringer, P.-J.; Delarue, M.; Baaden, M. One-Microsecond Molecular Dynamics Simulation of Channel Gating in a Nicotinic Receptor Homologue. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 6275−6280.
(24) Molontay, R.; Nagy, M. Two Decades of Network Science as Seen through the Co-Authorship Network of Network Scientists. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM*, 2019; pp 578−583.
(25) Riaz, F.; Ali, K. M. Applications of Graph Theory in Computer Science. *Proceedings of 3rd International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2011*, 2011; pp 142−145.
(26) Zegarac, A.; Caravelli, F. Memristive Networks: From Graph Theory to Statistical Physics. *Epl* **2019**, *125*, 10001.
(27) Borgatti, S. P.; Mehra, A.; Brass, D. J.; Labianca, G. Network Analysis in the Social Sciences. *Science* **2009**, *323*, 892−895.
(28) Liu, C.; Ma, Y.; Zhao, J.; Nussinov, R.; Zhang, Y.-C.; Cheng, F.; Zhang, Z.-K. Computational Network Biology: Data, Models, and Applications. *Phys. Rep.* **2020**, *846*, 1−66.
(29) Mason, O.; Verwoerd, M. Graph Theory and Networks in Biology. *IET Syst. Biol.* **2007**, *1*, 89−119.
(30) Prathik, A.; Uma, K.; Anuradha, J. An Overview of Application of Graph Theory. *Int. J. ChemTech Res.* **2016**, *9*, 242−248.
(31) Newman, M. E. J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45*, 167−256.
(32) Barabási, A.-L. The Network Takeover. *Nat. Phys.* **2012**, *8*, 14−16.
(33) Amitai, G.; Shemesh, A.; Sitbon, E.; Shklar, M.; Netanely, D.; Venger, I.; Pietrokovski, S. Network Analysis of Protein Structures Identifies Functional Residues. *J. Mol. Biol.* **2004**, *344*, 1135−1146.
(34) Heal, J. W.; Bartlett, G. J.; Wood, C. W.; Thomson, A. R.; Woolfson, D. N. Applying Graph Theory to Protein Structures: An Atlas of Coiled Coils. *Bioinformatics* **2018**, *34*, 3316−3323.
(35) Yan, W.; Hu, G.; Shen, B. Network Analysis of Protein Structures: The Comparison of Three Topologies. *Curr. Bioinform.* **2016**, *11*, 480−489.
(36) Chakrabarty, B.; Naganathan, V.; Garg, K.; Agarwal, Y.; Parekh, N. NAPS Update: Network Analysis of Molecular Dynamics Data and Protein-Nucleic Acid Complexes. *Nucleic Acids Res.* **2019**, *47*, W462−W470.
(37) Giuliani, A.; Krishnan, A.; Zbilut, J.; Tomita, M. Proteins As Networks: Usefulness of Graph Theory in Protein Science. *Curr. Protein Pept. Sci.* **2008**, *9*, 28−38.
(38) Aier, I.; Varadwaj, P. K.; Raj, U. Structural Insights into Conformational Stability of Both Wild-Type and Mutant EZH2 Receptor. *Sci. Rep.* **2016**, *6*, 1−10.
(39) Li, J.; Qin, Q.; Kardas, S.; Fosséprè, M.; Surin, M.; Fernandes, A. E.; Glinel, K.; Jonas, A. M. Sequence Rules the Functional Connections and Efficiency of Catalytic Precision Oligomers. *ACS Catal.* **2022**, *12*, 2126−2131.
(40) Hoover, J. M.; Ryland, B. L.; Stahl, S. S. Mechanism of Copper(I)/TEMPO-Catalyzed Aerobic Alcohol Oxidation. *J. Am. Chem. Soc.* **2013**, *135*, 2357−2367.
(41) Hoover, J. M.; Stahl, S. S. Highly Practical Copper(I)/TEMPO Catalyst System for Chemoselective Aerobic Oxidation of Primary Alcohols. *J. Am. Chem. Soc.* **2011**, *133*, 16901−16910.
(42) Ryland, B. L.; McCann, S. D.; Brunold, T. C.; Stahl, S. S. Mechanism of Alcohol Oxidation Mediated by Copper(II) and Nitroxyl Radicals. *J. Am. Chem. Soc.* **2014**, *136*, 12166−12173.
(43) Chandra, P.; Jonas, A. M.; Fernandes, A. E. Sequence and Surface Confinement Direct Cooperativity in Catalytic Precision Oligomers. *J. Am. Chem. Soc.* **2018**, *140*, 5179−5184.
(44) Chandra, P.; Jonas, A. M.; Fernandes, A. E. Spatial Coordination of Cooperativity in Silica-Supported Cu/TEMPO/Imidazole Catalytic Triad. *ACS Catal.* **2018**, *8*, 6006−6011.
(45) Chandra, P.; Jonas, A. M.; Fernandes, A. E. Synthesis of Discrete Catalytic Oligomers and Their Potential in Silica-Supported Cooperative Catalysis. *RSC Adv.* **2019**, *9*, 14194−14197.
(46) Fernandes, A. E.; Riant, O.; Jensen, K. F.; Jonas, A. M. Molecular Engineering of Trifunctional Supported Catalysts for the Aerobic Oxidation of Alcohols. *Angew. Chem. Int. Ed.* **2016**, *55*, 11044−11048.
(47) Fernandes, A. E.; Jonas, A. M. Design and Engineering of Multifunctional Silica-Supported Cooperative Catalysts. *Catal. Today* **2019**, *334*, 173−186.
(48) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.;

https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761−2770

Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. *AMBER 16*; University of California: San Francisco, 2016.
(49) Discovery Studio Modeling Environment v. 4.0.BIOvIA, D. S.. *Discovery Studio Modeling Environment*; Dassault Systèmes, Release: San Diego; Vol. *4*, 2015., 2015.
(50) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(51) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(52) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
(53) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(54) Stendardo, E.; Pedone, A.; Cimino, P.; Cristina Menziani, M.; Crescenzi, O.; Barone, V. Extension of the AMBER Force-Field for the Study of Large Nitroxides in Condensed Phases: An Ab Initio Parameterization. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11697−11709.
(55) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(56) Cerutti, D. S.; Duke, R.; Freddolino, P. L.; Fan, H.; Lybrand, T. P. A Vulnerability in Popular Molecular Dynamics Packages Concerning Langevin and Andersen Dynamics. *J. Chem. Theory Comput.* **2008**, *4*, 1669−1680.
(57) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.
(58) Delano, W. L. *The PyMOL Molecular Graphics System*; Version 1.8.4.0, Schrödinger, LLC; 2002.
(59) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498−2504.
(60) Yang, Z.; Algesheimer, R.; Tessone, C. J. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Sci. Rep.* **2016**, *6*, 30750.
(61) Rosvall, M.; Bergstrom, C. T. Maps of Random Walks on Complex Networks Reveal Community Structure. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1118−1123.
(62) Farage, C.; Edler, D.; Eklöf, A.; Rosvall, M.; Pilosof, S. Identifying Flow Modules in Ecological Networks Using Infomap. *Methods Ecol. Evol.* **2021**, *12*, 778−786.
(63) Rosvall, M.; Bergstrom, C. T. Multilevel Compression of Random Walks on Networks Reveals Hierarchical Organization in Large Integrated Systems. *PLoS One* **2011**, *6*, No. e18209.

https://doi.org/10.1021/acs.jcim.2c00101
J. Chem. Inf. Model. 2022, 62, 2761−2770