Discrete multifunctional sequence-defined oligomers with controlled chirality†

Jie Li, Maxime Leclercq, Mathieu Fossepré, Mathieu Surin, Karine Glinel, Alain M. Jonas and Antony E. Fernandes

Discrete sequence-defined oligomers are synthetic mimics of short peptides, with possible applications in catalysis, information storage, drug delivery or self-assembly. Here, we report on the efficient and stereo-controlled synthesis of sequence-defined poly(triazole-urethane) oligomers incorporating a very large range of functional side groups, including alkyl, phenyl, pyridyl, hydroxyl, amine, imidazole and carboxylic acid functions. The route involves the alternation of chloroformate-mediated carbamate bond formation and Cu-catalyzed azide–alkyne cycloaddition reactions, expeditiously leading to oligomers up to octamers by an iterative exponential growth strategy. Additionally, the chirality of the monomers is straightforwardly controlled and maintained during the elongation process, leading to optically-active multifunctional oligomers of controlled molecular configuration, as confirmed by $^1$H NMR, ESI-MS/MS and circular dichroism. Our strategy provides a versatile platform to efficiently synthesize discrete oligomers with programmable stereochemistry, sequentiality and multifunctionality.

Introduction

Sequence and stereoregularity are central to the functioning of biomacromolecules such as proteins or nucleic acids. Reaching a similar degree of control over primary structure and chirality in synthetic polymers has been a long endeavor, including the catalytic control of tacticity, the development of reversible deactivation routes for the control of molar mass, and the more recent emergence of synthetic strategies towards sequence-defined polymers. For that latter purpose, iterative stepwise methods, including iterative sequential growth (ISG) and iterative exponential growth (IEG), have generally been developed. In particular, IEG strategies provide rapid access to longer precision oligomers as the chain length is doubled after each iterative cycle, even though they are restricted to the production of symmetric products in which not every repeat unit is freely selectable. Hence, monomers bearing orthogonal protective groups, or latent functional groups, are activated divergently, releasing complementary-activated units, to be convergently reunified. It is however possible to go beyond the repetitive or palindromic primary structures typically obtained by IEG. For instance, the IEG+ strategy developed by Johnson and coll. combines both exponential chain growth and side chain functionalization.

Although such routes are effective in terms of sequence control, they rarely achieve a simultaneous control over stereochemistry, and most often have a limited tolerance towards the more reactive functional side groups typically found in natural polymers, responsible for catalytic activity for instance. Since one of the possible fields of application of discrete sequence-defined polymers is precisely multifunctional or cascade catalysis, as we demonstrated recently by showing that sequence order controls the catalytic activity of trifunctional catalytic oligomers, it becomes critical to develop more tolerant routes leading to chains of controlled sequence. Additionally, controlled stereochemistry, as typically present in enzymes, is also a very desirable feature for catalysis with the aim to further control the tridimensional arrangement of cooperative catalytic sites.

Here, we concentrate on the development of a synthetic route to reach this goal, and propose an innovative strategy leading to precision oligomers of controlled sequence and stereochemistry capable to incorporate a wide range of functional groups – such as those found in proteins – based on the copper-catalyzed azide–alkyne cycloaddition [CuAAC] “click” reaction. The CuAAC reaction, due to its high efficiency and chemoselectivity, was previously demonstrated to be a promising tool for the synthesis of sequence-defined...
This archetypal “click” reaction was for instance successfully used to obtain sequence-defined and stereocontrolled polymers, yet mainly limited to side groups of lesser reactivity. Building onto these previous works, we demonstrate a new strategy leading to discrete stereocontrolled and sequence-defined oligomers capable of accommodating a wide range of functional side groups alike natural peptides, albeit at a lower chain density.

**Experimental**

Experimental and simulation details are given in the ESI.

**Results and discussion**

The main building blocks were prepared starting from TBDMS-protected glycidyl propargyl ether 1 either in its racemic (rac-1) or enantiopure ((R)-1 or (S)-1) forms (TBDMS = tert-butyldimethylsilyl) (Scheme 1a). To prove the validity of our concept, the strategy was first developed starting from racemic rac-1. Side groups were inserted following opening of the oxirane ring of rac-1 with sodium azide and subsequent CuAAC reaction of 2 with alkynes 3a-d to provide functional monomers 4a-d with overall yields varying between 87% and quantitative. Noteworthy, alkynes 3a-d bearing respectively phenyl, pyridyl, methoxy and t-butyl groups were carefully selected to provide unique, unambiguous 1H NMR footprints in order to trace the composition of the oligomers.

![Scheme 1](image-url)

Scheme 1. Iterative exponential growth (IEG) strategy to generate discrete multifunctional stereocontrolled sequence-defined oligomers. (a) Synthetic scheme and one-letter code. (b) Structure of octamer FEPBFEPB obtained after 3 IEG cycles. (c) Simplified illustration of the obtained oligomers FEPB, FEPBFEPB and AOIK.
ing the functionalities used for installing the side chain groups and extending the chain backbone, which is a minor modification; and, much more crucially, by replacing ester groups by the considerably more stable carbamate groups, which is required for the installation of a much wider range of functional groups on the chains than could be done before. Indeed, we have recently shown that ester groups are not suitable in such a scheme, being not stable enough under hydrolytic and TBAF (tetrabutylammonium fluoride) conditions.39,40

Meanwhile, removal of the TBDMS group in 4c and 4d using TBAF provided free alkynes 6c and 6d (71 and 90% yield, respectively). Coupling of the obtained azides and alkynes yielded the dimeric species that allowed pursuing an IEG strategy towards higher oligomers. For clarity, a simple one-letter code was adopted for the naming of the different functional side groups/monomers that is inspired from the nomenclature used for natural amino acids (Scheme 1a).37 For instance, reaction of 5a with 6c under CuAAC conditions gives dimer FE (93% yield) bearing a phenyl (F) and a methoxy (E) side group, while reaction of 5b with 6d bearing a pyridyl (P) and tert-butyl (B) side group, respectively, afforded dimer PB (80% yield). A subsequent IEG cycle provided tetramer FEPB (72% yield); the later allowing access to octamer FEPBFEPB (Scheme 1b) in 57% yield through another IEG cycle. The overall yield of the octamer, starting from the protected monomers 4a–d, was 27%, providing access to milligrams of chains in our conditions. Of note, we did not face solubility issues for our oligomers despite their high content in triazole groups, possibly due to the functional similarity of such groups with amides.38

The stacked 1H NMR spectra of monomers 4a–d show that the signals corresponding to the pendant phenyl, pyridine, methoxy and tert-butyl protons do not overlap, as anticipated (Fig. 1a–d). Additionally, the triazole proton experiences different chemical environments in monomers 4a–d, which translates into different chemical shifts; this provides an additional way to identify the monomer units after their insertion in the oligomers. Precisely, the 1H NMR spectra of tetramer FEPB and octamer FEPBFEPB evidence the presence of the expected characteristic peaks of each monomeric unit together with the additional resonances corresponding to the parent azidopropylamine moiety in the main chain (Fig. 1e and f).

In addition to 1H NMR analysis, electrospray ionization mass spectrometry (ESI-MS) confirmed the structure of the monomers and oligomers (ESI†). For instance, ESI-MS of the octamer FEPBFEPB displayed peaks at m/z = 2952.5932, 1476.8147, 984.8832, 738.9161 and 591.3353 corresponding to the theoretical values for [M + H]+, [M + 2H]2+, [M + 3H]3+, [M + 4H]4+ and [M + 5H]5+, respectively (Fig. S128†). The molecular sequence of the oligomers was deciphered using tandem mass spectrometry (ESI-MS/MS).

Cleavage of the ether bond in the main chain, yielding to two series of complementary fragments, was detected as the most prominent fragmentation mechanism in tetramer FEPB (Fig. S134†) and octamer FEPBFEPB (Fig. 2). According to this mechanism, the sequence of FEPBFEPB reading from left to right were recorded with m/z = 536.32, 887.50, 1271.70, 1634.91, 2018.06, 2369.11 and 2753.04 corresponding respectively to the sequence F, FE, FEP, FEPB, FEPBF, FEPBF and FEPBFEPF. Additionally, reading the sequence from the right side displayed a peak at m/z = 182.14 corresponding to monomer B. Thus, the sequence in the octamer FEPBFEPB was unraveled by ESI-MS/MS, which was confirmed from the right to left side as well. Overall, the fragmentation patterns unambiguously confirm that all the synthesized oligomers have the desired sequence.

In order to further illustrate the versatility of our strategy, oligomers comprising a wider variety of functional side groups were synthesized. Thus, inspired by functional side groups found in natural amino acids, monomers comprising pendant amine (A), hydroxyl (O), imidazole (I) and carboxylic acid (K) side groups were synthesized (Scheme 1). Owing to the very chemistry involved in our approach, it was necessary to protect the nucleophilic side groups in order not to interfere with the elongation steps. Hence, A, O and K were protected with tert-butoxycarbonyl (Boc), benzoyl (Bz) and ethyl (Et) groups, respectively. Following the IEG procedure described above, tetramer A(Boc)O(Bz)IK(Et) was obtained in 35% overall yield after two IEG cycles. Final treatment of A(Boc)O(Bz)IK(Et) with trifluoroacetic acid and KOH provided the targeted AOIK tetramer in quantitative yield without hydrolysis of the main chain.

The structure of AOIK and its protected precursor A(Boc)O(Bz)IK(Et) was confirmed by 1H NMR (Fig. S99–S101†) and ESI-MS (Fig. S129–S130†). As previously, the sequence
information was confirmed by tandem mass spectrometry (Fig. S139† for AOIK, and S138 for its protected version). Interestingly, the negative mode ESI-MS/MS was successfully developed to trace the sequence from the right to the left side of AOIK owing to the contribution of carboxylic acid group in the side chain (Fig. S140†). Especially, preservation of the integrity of the oligomers (composition and sequence) after complete cleavage of the protecting groups was confirmed, demonstrating the tolerance of our synthetic strategy to a wide variety of chemistries and functional groups.

Although synthetic methods are flourishing to access sequence-defined oligomers, a very limited number of synthetic strategies have been reported to control both sequence and chirality,15,27,28,39 a situation that is ever more challenging when multiple functional groups are involved within the chain. Therefore, a set of enantiopure monomers bearing F, E, P and B side groups was synthesized starting from chiral TBDMS-protected glycidyl propargyl ether (R)-1 or (S)-127 (Scheme 1a). The enantiomeric excess (ee) of the monomers was traced by chiral HPLC, in which 4a was carried out as the example (Fig. S141†). The ee of (R)-4a and (S)-4a were 90 and 88%, respectively, showing that the chirality was well controlled under the synthetic processes. Stereocontrolled sequence-defined oligomers \((R,R,R)\text{-FEPB, (S,S,S)\text{-FEPB and (R,R,R,R,S,S,S)\text{-FEPBFEPB}}\) were straightforwardly prepared employing the IEG strategy developed for rac-I (Scheme 1a). As expected, their \(^1\)H NMR, ESI-MS and ESI-MS/MS spectra were identical to the ones of the racemic oligomers (ESI†). The maximum content of the \((R,R,R,R,S,S,S)\text{-FEPBFEPB}\) octamer in oligomers of lower length was estimated from the ratio of \(^1\)H NMR peaks between 0.86 and 0.96 ppm (\(t\)-butyl silyl-) and 5.00–5.34 ppm (protons of the asymmetric carbons of all monomer units except the terminal one); it was found to be at most 5% by number, and very likely much lower. HPLC confirmed the very high purity of the stereocontrolled tetramers and octamer (Fig. S149–S151†).

Circular dichroism (CD) spectroscopy provided qualitative information on the stereochemistry of the chains. The CD spectra of monomers 4a–d and dimers FE and PB (Fig. S142†) with (R)- and (S)-configurations show opposite signals as expected for enantiomers. Likewise, the CD spectra of dilute solutions of the tetramers \((R,R,R)\text{-FEPB and (S,S,S,S)\text{-FEPBFEPB}}\) show opposite signals peaking at 212 nm (Fig. 3a and Fig. S143†), confirming that the two tetramers have mirror configurations. In contrast, octamer \((R,R,R,R,S,S,S,S)\text{-FEPBFEPB}\) exhibits a significantly weaker CD signal (Fig. 3a) because it contains identical amounts of tetrameric (S) and (R) sequences; the non-suppression of the CD signal results from the fact that the octamer nevertheless lacks centers and planes of symmetry. Overall, the CD results thus clearly indicate that the stereochemistry is utterly controlled, and racemization does not occur during the IEG process. Of note, since the enantiomeric excess of our starting monomer units is currently ca. 90% only, our oligomers are still mixtures of a series of...
stereochromical configurations, which does not prevent them from displaying a strong circular dichroism. An even narrower distribution of stereochromical configuration could be obtained by starting from a more enantiopure form of 2, as discussed elsewhere.\(^1\)\(^4\)

Interestingly, at a five times higher concentration (Fig. 3b and Fig. S144†), the CD spectra of \((R,R,R,R)^{-}\) and \((S,S,S,S)^{-}\)FEPB display additional mirror peaks at 238 and 283 nm, which is likely due to aggregation of the chains and suggests that the aggregated states of the oligomers remain chiral. As before, the \((R,R,R,S,S,S)^{-}\)FEPBFEPB octamer only exhibits a weak signal. To further understand the aggregation behavior of the oligomers, all-atom Molecular Dynamics (MD) simulations were carried out from the single chain level to assemblies of a few chains (see ESI† for computational details). MD simulations of a single chain provided structural and dynamical aspects of the octamer \((R,R,R,S,S,S)^{-}\)FEPBFEPB.

The octamer rapidly folds into a compact globular structure (Fig. 4a), characterized by a low radius of gyration \((R_g)\) of 7.4 Å in average, very low compared to the end-to-end distance of the fully elongated conformation (around 100 Å, see ESI S145†). The stability of \(R_g\) profile over time (ESI S146†) indicates that the globular folding is preserved on the microsecond timescale. Although a total of 93 hydrogen bonds were detected for all conformations of the MD, most of them are not persistent: 88 H-bonds are present for less than 10% of the MD time, and none was detected for more than 50% of the MD time (see ESI S147†). This large number of possible H-bonds arises from each repeat unit containing one carbamate group, plus two triazole groups which are good mimetics of secondary amide groups with respect to H-bonding donor/acceptor capabilities.\(^3\)\(^8\)\(^,\)\(^4\)\(^0\) The network of H-bonds is constantly evolving and important oscillations in the number of H-bonds are observed throughout MD time (Fig. 4b), varying from 0 to 7 H-bonds with an average of only 2–3 H-bonds per conformation. Therefore, H-bonds are only weakly participating to the compact globular folding of the octamer. In contrast, the numerous aromatic moieties located on the side chains and along the backbone, labeled on top of Fig. 4, provide many possibilities of \(\pi\)-stacking interactions. The distances between the centers-of-mass (COM) of all depicted aromatic pairs were estimated. As for H-bonds, the network of aromatic interactions is dynamic and a criterion (at least 50% of MD time at a distance below 5.0 Å) was considered to point out the most persistent aromatic interactions. Some of these most persistent stacking interactions are depicted on the final MD snapshot in Fig. 4c and d. With this criterion, six \(\pi\)-stacking interactions are persistent: one is involving two triazole moieties of the backbone (labelled as 13 and 15 in Fig. 4), two interactions concern backbone/side-chain aromatic moieties (5a–12 and 5b–14 pairs) and three interactions are occurring between side-chains (5a–8, 3a–6, and 3b–6). These persistent \(\pi\)-stacking interactions between aromatic groups far in the octamer sequence but close in space evidently contribute to stabilize the compact globular conformations of the oligomer.

MD simulations involving either two chains or five chains were performed to investigate the driving force and dynamics for the aggregation of the octamer. For both MD simulations, octamer chains rapidly form a supramolecular complex that remains stable on the microsecond time scale, i.e. no dissociation between the chains was observed. Interestingly, self-assemblies of 2 chains or 5 chains do not show any intertwining between the chains (see MD snapshots in Fig. 4e and f). In comparison to the MD simulations of a single chain, only a minor increase of the average \(R_g\) is observed in the assembly of two chains (ESI S146†). The compactness of both chains in the assembly is similar as the average \(R_g\) of both chains converge to identical values. In the assembly of 5 chains, compact conformations are again observed, but with a slightly higher average \(R_g\) (around 10 Å). Concerning the dynamics of the octamer, the flexibility profile along the chain of the octamer (as estimated with Root-Mean-Square Fluctuations estimates, see ESI S148†) differs for the single chain and for the assemblies. Importantly, the flexibility of a single octamer chain can be discriminated inside the aggregate. For example, the octamer #1 corresponding to the chain colored in black in Fig. 4f is more rigid than the other octamers, because of its more compact folding (lower average \(R_g\)). Altogether, our results show that \((R,R,R,R,S,S,S)^{-}\)FEPBFEPB octamers fold into very compact structures, mainly stabilized by persistent \(\pi\)-stacking interactions, whereas hydrogen bonds are not persistent. Given the numerous possibilities of stacking interactions, stable assemblies of several chains are observed in which structural aspects and the dynamics of octamers can be influenced by the number of chains within the assembly. Altogether, this indicates that the oligomers studied in this work are prone to compact folding and assembly into...
Conclusions

In conclusion, we report on a straightforward and efficient approach for the synthesis of discrete multifunctional stereo-controlled sequence-defined oligomers. The method provides rapid access to optically-active poly(triazole-urethane) oligomers bearing triazole-linked side chains, with a broad range of possible functional side groups reminiscent of those found in proteins, albeit at a smaller linear density. Another difference is that the folding and aggregation of oligomers is most likely driven by π-stacking interactions, as revealed by MD simulations, whereas the numerous possible H-bonding interactions play a much more minor role. These oligomers offer attractive perspectives for multifunctional catalysis and self-assembly, applications for which octameric sequences are more than long enough as we demonstrated previously.29

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge the Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek under EOS project no. 30650939 for financial support. A.E.F. acknowledges the European Regional Development Fund (ERDF) and Wallonia Operational Program “Wallonia-2020.EU”. M.L. and M.S. acknowledge the Wallonia Region and FNRS under grant no. F.4532.16. The authors acknowledge Gabriella Barrozino-Consiglio for 500 MHz NMR analysis, Raoul Rozenberg for APCI-MS, Hervé Degand for ESI-MS/MS analysis and Laurent Collard for HPLC analysis. Computational resources have been provided by the Consortium des équipements de Calcul Intensif (CéCI), funded by the FNRS under grant 2.5020.11.
KG and MS are Senior Research Associates of the F. R. S.–FNRS.

References