Molecular Recognition Hot Paper

How to cite: *Angew. Chem. Int. Ed.* **2022,** *61,* e202201061
International Edition: doi.org/10.1002/anie.202201061
German Edition: doi.org/10.1002/ange.202201061

The Protein-Templated Synthesis of Enzyme-Generated Aptamers

*Julio Vacacela, Anna-Lisa Schaap-Johansen, Patricia Manikova, Paolo Marcatili, Marta Prado, Yi Sun, and Jon Ashley**

Abstract: Inspired by the chemical synthesis of molecularly imprinted polymers, we demonstrated for the first time, the protein-target mediated synthesis of enzyme-generated aptamers (EGAs). We prepared pre-polymerisation mixtures containing different ratios of nucleotides, an initiator sequence and protein template and incubated each mixture with terminal deoxynucleotidyl transferase (TdT). Upon purification and rebinding of the EGAs against the target, we observed an enhancement in binding of templated-EGAs towards the target compared to a non-templated control. These results demonstrate the presence of two primary mechanisms for the formation of EGAs, namely, the binding of random sequences to the target as observed in systematic evolution of ligands by exponential enrichment (SELEX) and the dynamic competition between TdT enzyme and the target protein for binding of EGAs during synthesis. The latter mechanism serves to increase the stringency of EGA-based screening and represents a new way to develop aptamers that relies on rational design.

In the world of molecular recognition, molecular imprinted polymers (MIPs) and DNA aptamers have emerged as the most attractive synthetic alternatives to natural antibodies. Despite their common goal, little synergy has been achieved with these two technologies.[1] Researchers have attempted to develop hybrid aptamer MIPs by mixing the nucleic acid bases with a polymer backbone through extensive engineering of synthetic monomers or incorporate existing aptamers into hydrogels.[2,3] We can clearly demonstrate the gulf of differences between these two technologies through the methodologies used to develop them. Aptamers are selected by incubating a target with a random library of DNA and undertaking repeated rounds of partitioning, PCR amplification and regeneration of the library through the SELEX methodology.[4,5] Upon library enrichment, one can perform sequencing of the aptamers to determine their sequence. MIPs on the other hand, are formed by incubating the target (template) with a mixture of synthetic monomers and a crosslinker to form non-covalent interactions.[6] Upon crosslinking of the polymers through the addition of an initiator molecule, a cavity (recognition site) is formed which is complimentary to the shape of the template.

Using terminal deoxynucleotidyl transferase (TdT) as a catalyst, a short oligonucleotide sequence as an initiator and a mixture of nucleotides as the functional monomers, we demonstrated for the first time, the protein-templated synthesis of nucleic acid based enzyme-generated aptamers (EGAs). TdT is a unique polymerase enzyme capable of adding individual nucleotides onto an elongating oligonucleotide initiator without the need for a DNA template.[7] Interest in the use of TdT enzyme has increased in the last few years as a possible alternative method for the de novo solid-phase synthesis of DNA.[8] The TdT catalyzed formation of polynucleotides can occur through a living polymerization mechanism provided that a poly (T) initiator sequence and only the corresponding nucleotide (dTTP) are used.[9] We recently demonstrated that enriched libraries of EGAs of broad variable size (vsDNA), which are visible on a native gel as a DNA smear, can be rationally designed towards each target to allow for the non-evolutionary screening of protein binding sequences.[10]

Analogous to a molecular imprinting, we found that we can tune the apparent binding properties of EGAs by forming the EGAs in the presence of the template protein referred to as templated-EGAs. We achieved this through the adjustment of a number of reaction parameters in each pre-polymerisation mixture such as incubation time, type of divalent metal ion used, the ratios of nucleotides to template protein and the ratios of template protein to the initiator sequence.

We first demonstrated an enhancement in the apparent binding by incubating the model protein template (human lactoferrin) with pre-polymerisation mixtures containing different ratios of dNTPs, the initiator oligonucleotide and selection buffer (Table S1–S3). Upon initiation of the

[*] J. Vacacela, A.-L. Schaap-Johansen, P. Manikova, P. Marcatili, Y. Sun, J. Ashley
Department of Health Technology
Technical University of Denmark
Ørsteds Pl. 345C, Kgs. Lyngby, 2800 (Denmark)
E-mail: j.ashley@ljmu.ac.uk

M. Prado
International Iberian Nanotechnology Laboratory (INL)
Av. Mestre José Veiga Braga 4715-330 (Portugal)

J. Ashley
School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Byrom Street, Liverpool, L3 3AF (UK)

reaction through the addition of TdT enzyme (1 U μL^{-1}), the mixtures were incubated for 0.5–2 hrs at room temperature. We stopped the reaction by incubation of the mixtures at 75 °C for 10 minutes.

The removal of the protein template and purification of the unreacted nucleotides from template-mediated EGAs can be easily achieved through a PCR purification kit. The observed size distributions of the libraries **3a** and **3b**, are shown as characteristic poly-disperse smears, visualized on a 5% denaturing gel (Figure 1A). The resultant protein-templated EGA mixtures (**T-EGA**) and corresponding non-templated control mixtures (**NT-EGA**), where the aptamers are formed in the absence of the target demonstrated similar size profiles, when the ratio of the initiator and template are kept at a 1:1 ratio. These broad sized products result from the fact that each pre-polymerisation mixture contains all four dNTPs and initiator sequence. From our previous next generation sequencing (NGS) and bioinformatics studies, we showed that the incorporation of rationally designed nucleotide mixtures gives rise to enzyme generated aptamers which display diverse range of secondary structures (G-quadruplexes or 1–2 way junctions).[10] This serves to reduce the rate of incorporation and slow down the kinetics of incorporation to a point that each strand of polynucleotide elongates at a different kinetic rate and deviates from the living polymerisation mechanism observed when incorporating a single type of nucleotide.[11,12] The template protein may also act to inhibit further elongation of the polynucleotides through competition for binding with the TdT enzyme. We would also expect a kinetic bias for the incorporation of each nucleotide by TdT enzyme.[12] The kinetics of incorporation and the bias of TdT can effectively be controlled through the incorporation of different divalent metal ions, the time of incubation and the ratios of initiator sequence to each type of nucleotide.[12]

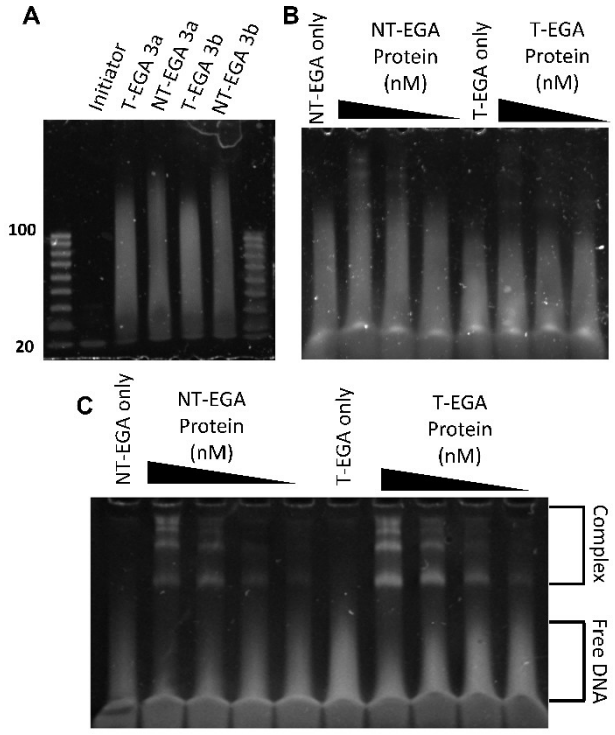

Figure 1. A) 5% denaturing gel of synthesized templated-EGAs and corresponding non-templated EGAs; B) 5% EMSA of synthesized templated-EGAs and corresponding non-templated EGAs with composition **3a** and C) EMSA of synthesized templated-EGAs and corresponding non-templated EGA of composition **3b**.

In the case of the divalent metal ions used in the reaction buffer, the kinetics of TdT incorporation of nucleotides can change remarkably. For instance, in most commercial buffers, the use of Co^{2+} results in the formation of longer polynucleotides with a broader range of sizes compared to Mg^{2+} ions. The inclusion of Co^{2+} in the reaction buffer also results in the increased preference to incorporate the pyrimidines (dCTP and dTTP) whereas the presence of Mg^{2+} has a preference to incorporate the purines (dGTP and dATP). Mg^{2+} was chosen based on the slower observed kinetics and the preference for dGTP.

The time of incubation also affects the resultant size distribution of the formed EGAs. As such, the size distribution of the polynucleotide aptamers correlates well with the time of incubation. The ratios of the initiator sequence to nucleotide concentrations can also profoundly affect the size distribution of EGAs. As the ratio increases from 1:10 to 1:50 ratio [initiator: nucleotide], the size distribution of the resultant polynucleotide aptamers and non-templated control increases with 1:100 ratio resulting in larger size distributions. The 1:50 ratio was chosen for all subsequent enzymatic synthesis due to the observed sizes obtained from the denaturing gel, which still allowed for adequate separation resolution on 5% EMSA gels.

Next, we performed rebinding studies on the resultant mixtures to compare the apparent binding of the protein against both **T-EGAs** and **NT-EGA** mixtures. Both **T-EGAs** and **NT-EGA** mixtures with compositions **3a** and **3b** were incubated with the protein for 1 hour and analyzed for apparent binding on a 5% EMSA. **3a** showed very little binding affinity towards the protein for both the templated aptamers and non-templated control due to the higher proportion of dATP and dCTP (Figure 1B**)**. In contrast, **3b** shows a slightly higher degree of binding towards the **T-EGAs** compared to the **NT-EGAs (**Figure 1C**)**, which suggests that an additional mechanism for the formation of EGAs was at play. Although, we observed complexes for both the **T-EGAs** and **NT-EGAs** as is often seen when comparing MIPs and non-imprinted polymers (NIPs), the observed complexes seen in the **NT-EGA** mixtures come about from random sequences interacting with the target protein. This type of binding is usually observed in a normal SELEX based method and from the rationally designed EGA libraries from our previous study.[10] However, unlike MIPs, this type of binding towards the target shouldn't be considered a source of "non-specific binding" as we can easily separate the target bound EGAs from our unbound EGAs and elucidate individual sequences through sequencing, whereas with MIPs, the presence of "non-specific binding" is seen as detrimental to the molecular imprinting process.

We then performed a comprehensive feedback loop study by synthesizing different batches of **T-EGA** mixtures, purifying them and testing their apparent binding. By rationally designing and testing each of the parameters such as the relative nucleotide ratios, initiator concentration, template protein concentration and time of reaction, we optimized the enhanced binding of the **T-EGA** mixture towards the protein target compared to the **NT-EGA** mixture. Once we optimized the conditions for the synthesis of **T-EGAs**, we performed comprehensive binding studies on the batch with composition **5a** which showed the highest apparent binding affinity according to EMSA analysis (Figure S1). Using surface plasmon resonance (SPR), we estimated the binding affinity (K_D) of the **T-EGA** (**5a**) towards human lactoferrin to be about 12 ± 1.3 nM (Figure 2A). We subtracted response of the **NT-EGA** mixture from that of the **T-EGA** mixture and fitted it using a bivalent kinetic model. Absolute responses on the SPR sensor demonstrate the degree of non-specific binding between the templated and non-templated EGA mixtures (Figure 2B).

We also used the FIDA 1 platform, which is capable of measuring binding affinities of complex mixtures in solution using capillary-based flow dispersion analysis to confirm the binding kinetics.[13] Using FAM labelled **T-EGA** and **NT-EGA** mixtures of composition **5a**, we estimated the binding affinity (K_D) of the **T-EGA** mixture to be about 9.96 nM while the **NT-EGA** mixture showed a binding affinity of about 64.2 nM (Figure S3). The qualitative binding of **T-EGA** and **NT-EGA** mixtures of **5a** were also confirmed on a 5 % EMSA in Figure 3A, B. To test the specificity of the synthesized **T-EGAs** and **NT-EGA** mixtures, we incubated both mixtures with micromolar amounts of human serum albumin (HSA) trypsin (Ty). Figure 3C shows that both the **T-EGA** and **NT-EGA** mixtures demonstrated specific binding towards human lactoferrin and no binding towards the control proteins. The specificity of templated-enzyme generated aptamer mixtures was reconfirmed by SPR (Figure S2) and revealed that the **T-EGA** mixtures demonstrated at least a 1000× higher binding towards human lactoferrin compared to HSA and trypsin. This confirms that the synthesized **T-EGAs** mixtures show highly specific binding towards human lactoferrin compared to the **NT-EGA** mixtures.

We separated, extracted and pooled protein:DNA complexes from a preparative native gel and used rapid amplification of variable ends (RAVE) assay to amplify them. The extracted dsDNA product of the **T-EGA** and **NT-EGA** complexes were confirmed on a DNA analyzer (Figure S4A and Figure S5A) and then sequenced using an Illumina HiSeq platform. Sequences were ranked by copy number and size (Figure S4B, C and Figure S5B, C). NGS data and bioinformatics studies revealed that both **T-EGAs** and **NT-EGA** sequences were G-rich and both libraries had similar size distributions. Individual sequences were chosen based on their copy number, their Gibbs free energy, retention of secondary structures upon removal of the initiator sequence using mfold and their ability to form G-quadruplex structures from using QGRS Mapper.[14] At least 10 sequences, which fulfilled the criteria were resynthesized (with both the initiator sequence and polyA region omitted) and underwent further aptamer binding affinity studies.

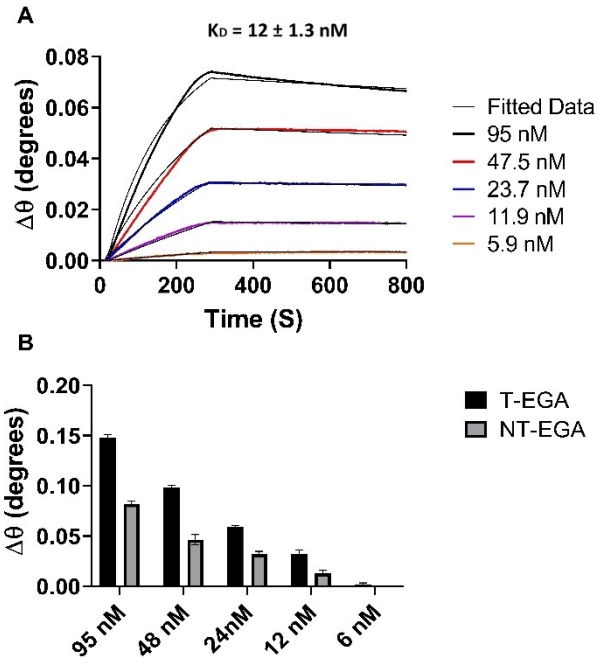

Figure 2. A) SPR sensorgrams showing the relative response of **5a** human lactoferrin towards immobilized EGAs (NT-EGAs reference signal subtracted); B) the maximum absolute SPR responses of human lactoferrin towards templated-EGAs and corresponding non-templated EGAs.

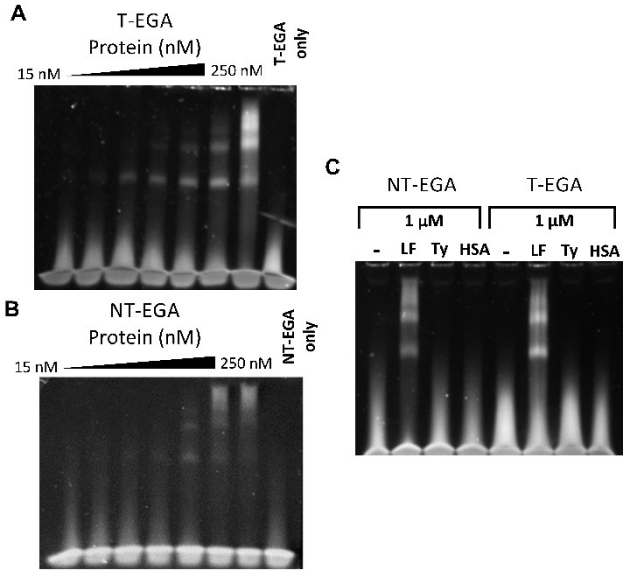

Figure 3. EMSA (5 %) of batch **5a** templated-EGAs and corresponding (A) and non-templated EGAs (B) towards human lactoferrin using the optimized imprinting conditions and (C) specificity of templated-EGAs and corresponding non-templated-EGAs towards human lactoferrin (LF); (pI: 8.7), trypsin (Ty); (pI: 10.5) and human serum albumin (HSA); (pI: 4.7).

Our lead individual sequence **EGA 4T** (Figure 4 and Table S4) demonstrated low nanomolar binding (5.4 ± 1.9 nM) and at least a 10× selectivity towards human lactoferrin compared to bovine lactoferrin (59 ± 9.8 nM), while the corresponding NT-control (**EGA 4NT**) demonstrated a binding affinity of 57 ± 7.1 nM towards human lactoferrin. These differences in binding affinities of **EGA 4T** towards human and bovine lactoferrin may be due to the difference in the degree of glycosylation between bovine and human lactoferrin.[15] We confirmed using the FIDA 1 platform that **EGA 4T** and **EGA 4NT** showed binding affinities of about 9.55 nM and 70.4 nM respectively towards human lactoferrin (Figure S6A, B). **EGA 4T** demonstrated micromolar binding affinities towards HSA and trypsin (Figure S7A, B). A qualitative EMSA gel also confirmed the specificity of **EGA 4T** compared to bovine lactoferrin, trypsin, HSA, lysozyme and hemoglobin (Figure S7C).

Overall, we confirmed that the formation of EGAs in the presence of a template protein is possible using TdT enzyme and this appears to enhance the overall binding of EGA mixtures and individual sequences towards the protein template. In addition, through carefully altering the compositions of the pre-polymerisation mixtures, we can effectively use bottom-up rational design in the development of enzyme-generated aptamers giving us unprecedented control over their development.

This raises the question as to what the mechanism is for the synthesis of templated-EGAs and their apparent enhancement in binding towards the target. The classical mechanism for the formation of molecularly imprinted polymers involves the formation of non-covalent static interactions between monomers and the template molecule. Upon addition of an initiator, the monomers are cross-linked together to form the molecular recognition site. Recently, researchers have suggested that this classical mechanism for molecular imprinting is incorrect due to the weak binding affinities formed between small molecules templates and synthetic monomers in pre-polymerisation mixtures.[16,17] Instead, they suggested that polymer oligomers are formed, which start to interact with the template through dynamic mechanisms, although the k_{off} rate is high due to the flexibility of synthetic polymers. As the polymer network grows through crosslinking, the conformation around the template is fixed into place forming the recognition site.

In the case of templated-EGAs, the absence of any crosslinker, is made up for by the ability of polynucleotides to form distinct conserved 3D conformations due to the phosphate backbone. This apparent "imprinting effect" or template driven synthesis observed for EGAs may occur in a similar manner as the latter alternative model described for MIPs. As the oligonucleotides are formed, they start to form dynamic interactions or switch structures upon binding to the target. Eventually, some of the polynucleotides can form strong enough interactions with the template to prevent the incorporation of further nucleotides by TdT. Those polynucleotide sequences with weak binding affinities continue to grow in size, through further additions of single nucleotides leading to further changes in conformation, which in turn may alter the binding properties towards the target.

The use of TdT and bottom-up rational design removes the current size limitation set on fixed length aptamer libraries (≤100 nt) synthesized using classical phosphoramidite chemistry and the difficulties in synthesizing GC rich sequences. It also allows us to avoid the so-called PCR bias effect where sequences are selected based on their ability to be amplified rather than their binding affinity as observed for SELEX.[18] We can scale up the synthesis, meaning that I can continue generating EGAs against each target rather than using a pre synthesized library containing a fixed number of random sequences. TdT enzyme can also incorporate chemically modified nucleotides making this a potentially simple way to develop base modified EGAs to introduce non-natural interactions between EGAs and the target as well as increase resistance to enzyme degradation.[19,20] These results demonstrate a unique and promising new methodology to synthesize longer polynucleotides (with possible multivalent binding) as a potential new tool for the rapid development of robust chemical antibodies.

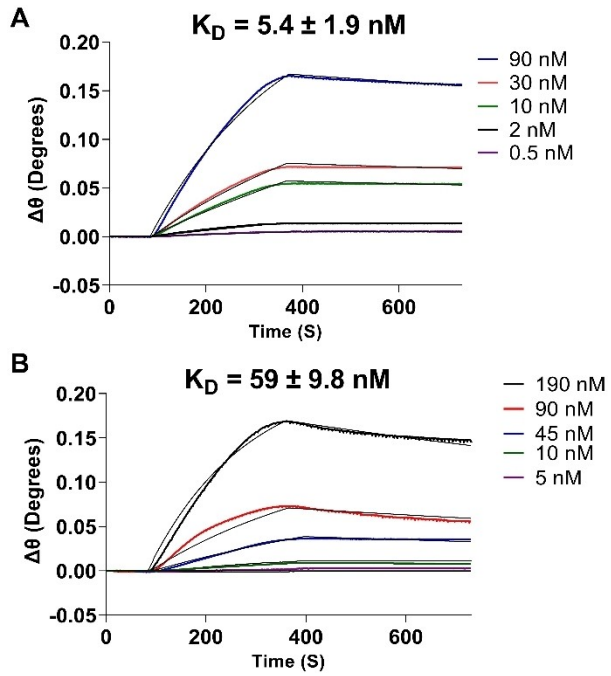

Figure 4. SPR Binding affinity studies of the individual LF candidate sequence LF_EGA 4T towards human lactoferrin (A) (scrambled sequence signal subtracted) and bovine lactoferrin (B) using surface plasmon resonance (scrambled sequence signal subtracted).

Acknowledgements

The authors acknowledge the funding from the VILLUM FONDEN, grant number (00022912). The authors would like to thank Brian Sørensen, Henrik Jensen, Hasse Hedeby and Emil Dandanell Agerschou from FidaBio for providing

access to the FIDA 1 and resources to perform additional affinity studies.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

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

Keywords: Aptamers · Enzymes · Enzyme-Generated Aptamers · Proteins · Terminal Deoxynucleotidyl Transferase

[1] M. Naseri, M. Mohammadniaei, Y. Sun, J. Ashley, *Chemosensors* **2020**, *8*, 32.
[2] A. Poma, H. Brahmbhatt, H. M. Pendergraff, J. K. Watts, N. W. Turner, *Adv. Mater.* **2015**, *27*, 750–758.
[3] W. Li, Q. Zhang, Y. Wang, Y. Ma, Z. Guo, Z. Liu, *Anal. Chem.* **2019**, *91*, 4831–4837.
[4] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822.
[5] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505–10.
[6] R. Arshady, K. Mosbach, *Makromol. Chem.* **1981**, *182*, 687–692.
[7] E. A. Motea, A. J. Berdis, *Biochim. Biophys. Acta Proteins Proteomics* **2010**, *1804*, 1151–1166.
[8] S. Palluk, D. H. Arlow, T. De Rond, S. Barthel, J. S. Kang, R. Bector, H. M. Baghdassarian, A. N. Truong, P. W. Kim, A. K. Singh, N. J. Hillson, J. D. Keasling, *Nat. Biotechnol.* **2018**, *36*, 645–650.
[9] L. Tang, L. A. Navarro, A. Chilkoti, S. Zauscher, *Angew. Chem. Int. Ed.* **2017**, *56*, 6778–6782; *Angew. Chem.* **2017**, *129*, 6882–6886.
[10] J. Ashley, A.-L. Schaap-Johansen, M. Mohammadniaei, M. Naseri, P. Marcatili, M. Prado, Y. Sun, *Nucleic Acids Res.* **2021**, *49*, 1065–1074.
[11] K. H. Leung, B. He, C. Yang, C. H. Leung, H. M. D. Wang, D. L. Ma, *ACS Appl. Mater. Interfaces* **2015**, *7*, 24046–24052.
[12] S. Deshpande, Y. Yang, A. Chilkoti, S. Zauscher, *Methods Enzymol.* **2019**, *627*, 163–188.
[13] N. N. Poulsen, N. Z. Andersen, J. Østergaard, G. Zhuang, N. J. Petersen, H. Jensen, *Analyst* **2015**, *140*, 4365–4369.
[14] O. Kikin, L. D'Antonio, P. S. Bagga, *Nucleic Acids Res.* **2006**, *34*, W676–W682.
[15] C. Conesa, M. Calvo, L. Sánchez, *Biotechnol. Adv.* **2010**, *28*, 831–838.
[16] T. Cowen, E. Stefanucci, E. Piletska, G. Marrazza, F. Canfarotta, S. A. Piletsky, *Macromolecules* **2020**, *53*, 1435–1442.
[17] D. A. Spivak, *MRS Proc.* **2003**, *787*, G2.6.
[18] N. Komarova, A. Kuznetsov, *Molecules* **2019**, *24*, 3598.
[19] M. Hollenstein, *Org. Biomol. Chem.* **2013**, *11*, 5162–5172.
[20] R. Gu, T. Oweida, Y. G. Yingling, A. Chilkoti, S. Zauscher, *Biomacromolecules* **2018**, *19*, 3525–3535.

Manuscript received: January 20, 2022
Accepted manuscript online: February 15, 2022
Version of record online: March 4, 2022