Protein Design

International Edition: DOI: 10.1002/anie.201713377
German Edition: DOI: 10.1002/ange.201713377

Peptide-Guided Assembly of Repeat Protein Fragments

Erich Michel, Andreas Plückthun, and Oliver Zerbe**

Abstract: *Herein, we present the peptide-guided assembly of complementary fragments of designed armadillo repeat proteins (dArmRPs) to create proteins that bind peptides not only with high affinity but also with good selectivity. We recently demonstrated that complementary N- and C-terminal fragments of dArmRPs form high-affinity complexes that resemble the structure of the full-length protein, and that these complexes bind their target peptides. We now demonstrate that dArmRPs can be split such that the fragments assemble only in the presence of a templating peptide, and that fragment mixtures enrich the combination with the highest affinity for this peptide. The enriched fragment combination discriminates single amino acid variations in the target peptide with high specificity. Our results suggest novel opportunities for the generation of new peptide binders by selection from dArmRP fragment mixtures.*

Directed evolution methods have been very successfully used over the past years to create binding proteins that target peptides, proteins, and even small molecules with high affinity and high selectivity.[1–3] These procedures have used both antibody and non-antibody scaffolds, and they aim at modifying loops and/or surfaces to become complementary to the target and thus suitable for forming sufficient intermolecular contacts while retaining the overall fold of the scaffolding protein.[4–7] In this work, we investigated whether the target recognition surface can also be provided by two complementary protein fragments that form a complex capable of binding the target.

Designed armadillo repeat proteins (dArmRPs) form elongated, rod-like molecules that consist of multiple, tightly packed internal modules M, and are terminated at the N- and C-terminal ends by capping modules, Y_{iii} and A_{ii}, respectively.[8–11] Each internal module M contains three tightly packed α-helices H1, H2, and H3 (Figure 1 A). They propagate a right-handed triangular spiral (Figure 1 B), which exposes a supercoiled binding surface consisting of helix H3 of each repeat. Each module specifically recognizes two amino acids of an extended target peptide (Figure 1 B, C).[12,13]

[*] Dr. E. Michel, Prof. O. Zerbe
Department of Chemistry
University of Zurich
Winterthurerstrasse 190, 8057 Zurich (Switzerland)
E-mail: erichxmichel@gmail.com
oliver.zerbe@chem.uzh.ch
Prof. A. Plückthun
Department of Biochemistry
University of Zurich
Winterthurerstrasse 190, 8057 Zurich (Switzerland)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201713377.

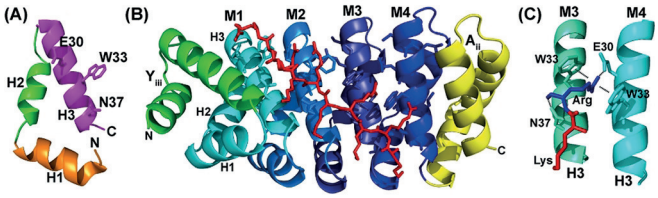

Figure 1. Structural features of designed armadillo repeat proteins. A) dArmRPs contain internal 42-residue M modules, which consist of three α-helices H1, H2, and H3. The capping modules Y_{iii} and A_{ii} consist of 31 and 41 residues, respectively. B) A model of YM_4A based on the crystal structure of YM_5A.[14] The bound $(KR)_4$ peptide is shown in red. C) Detailed view of the modular (KR) dipeptide recognition: Each arginine of the peptide specifically forms salt bridges with Glu30, π–cation interactions with Trp33, and a double hydrogen bond of Asn37 to the peptide backbone.

dArmRPs have been shown to bind $(KR)_n$-type peptides with a K_d value that depends both on the number of M modules and on the number of (KR) dipeptide units in the peptide.[13] We recently discovered that the complementary dArmRP fragments YM_2 and MA assemble with a K_d of 126 nM into the YM_2:MA complex that structurally resembles the full-length YM_3A protein.[14] Importantly, the assembled fragments of a dArmRP binder for the peptide ligand neurotensin retained the ability to bind the peptide.[8]

Based on these observations, we set out to investigate whether mixtures of complementary dArmRP fragments enrich those combinations that constitute the best binder towards a given target peptide. The principle is demonstrated for the protein YM_4A using mixtures of a particular N-terminal fragment with a number of complementary C-terminal fragments that display different affinities towards a target peptide (Figure 2).

Our initial setup comprised YM_2 and M_2A fragments obtained from an inter-modular split of the YM_4A dArmRP. However, the high-affinity interaction of complementary fragments from the inter-modular split, even in the absence of ligand, caused an enrichment bias to fragments with the highest affinity for the respective complementary partner fragment, independent of peptide affinity. This greatly reduced the impact of the ligand to effectively guide the assembly of those fragment combinations that show highest affinity to the ligand (data not shown). We therefore aimed for alternative fragmentation sites that avoid fragment assembly in the absence of the templating peptide altogether. We thus introduced intra-modular split sites between helices H1 and H2 and between helices H2 and H3 (Figure 3 A) and determined the complementation affinities of the corresponding pairs by analytical size exclusion chromatography (Figure 3 B–E).

The stepwise addition of M_2A to YM_2 results in immediate and stoichiometric complementation, evident from the

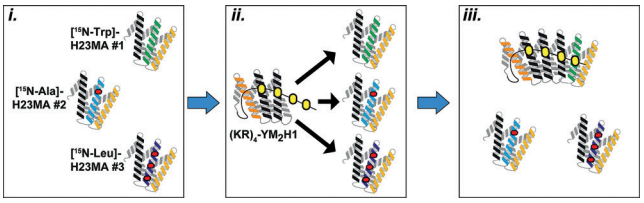

Figure 2. Peptide-guided dArmRP fragment assembly. i) Three C-terminal H23MA dArmRP fragments, in which H23 denotes the second and third helix of a M module, with different affinities to a target peptide (mutations shown in red) are ii) mixed with $(KR)_4$-YM_2H1, which is a fusion of the $(KR)_4$ target peptide with the complementary N-terminal YM_2H1 fragment (H1 here is the first helix of a M module). iii) $(KR)_4$-YM_2H1 enriches the C-terminal H23MA fragment that has the highest binding affinity to the target peptide to form a complex (top) over more weakly binding fragments (bottom). Modules that are kept constant throughout the selection procedure are shown in black while the green, light blue, and dark blue modules represent the variable modules.

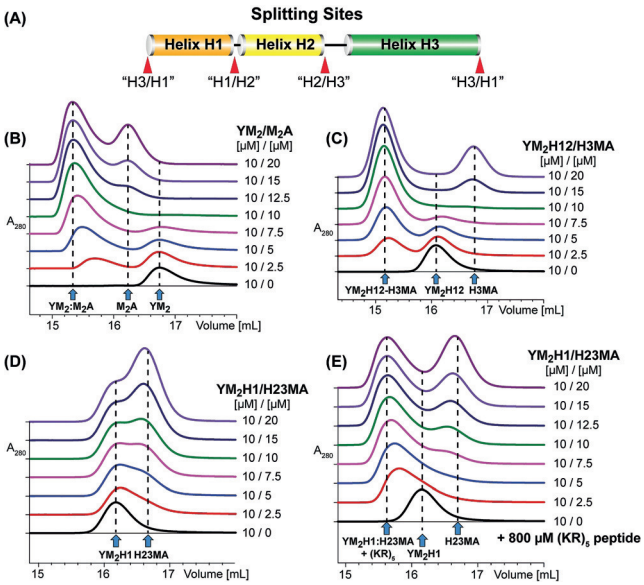

Figure 3. Complementation properties of split dArmRPs. A) Location of the analyzed split sites. B–E) Size-exclusion profiles of the various N-terminal fragments titrated with increasing amounts of the complementary C-terminal fragments: B) YM_2:M_2A (H3:H1), C) YM_2H12:H3MA, D) YM_2H1:H23MA, and E) YM_2H1:M23MA in the presence of 80 equiv of $(KR)_5$. The concentrations of N- and C-terminal fragments are indicated next to each elution profile.

YM_2:M_2A peak eluting at earlier volumes (Figure 3B). The YM_2H12 and H3MA pair displays a similar SEC behavior, indicating a comparable affinity in complementation (Figure 3C). In contrast, the absence of a peak at earlier elution volumes clearly indicates that YM_2H1 and H23MA do not complement each other in the absence of peptide (Figure 3D). Subsequent isothermal titration calorimetry (ITC) measurements indicated a K_d of 593 ± 35 nM for assembly of the YM_2:M_2A pair, a K_d of 234 ± 62 nM for the YM_2H12:H3MA pair, and a K_d of >20 µM for the YM_2H1:H23MA pair. However, when carrying out the SEC experiment with the YM_2H1:H23MA pair in the presence of excess $(KR)_5$ peptide (80 equiv), we observed stoichiometric fragment complementation (Figure 3E). The YM_2H1:H23MA pair therefore represents the desired conditional complementation system that strictly requires the presence of the target peptide.

To facilitate the NMR analysis later on, we converted the trimolecular into a bimolecular complex by fusing $(KR)_4$ via a $(GS)_n$ linker to the N-terminal YM_2H1 fragment. This system maintains the conditional assembly as only C-terminal fragments with a significant interaction with the (KR) peptide assemble with the N-terminal fragment. ITC analysis revealed that the $(KR)_4$-$(GS)_4G$-YM_2H1 fusion binds H23MA with a K_d of 17.5 nM (Table 1), which is roughly three orders of magnitude better than that for YM_2H1:H23MA complementation in the absence of ligand. Systematic variation of the linker length between the peptide and the YM_2H1 domain indicates a minimum of nine alternating Gly and Ser residues, $(GS)_4G$, for an optimal interaction and shows a dramatic reduction in the affinity when the linker is too short ($K_d \approx 1.5$ µM with 8 residues and >8 µM with 7 linker residues). We therefore continued with the $(GS)_4G$ linker construct, which in the following is referred to as $(KR)_4$-YM_2H1.

The design of H23MA variants with reduced affinities towards $(KR)_n$ peptides was inspired by the crystal structure of YM_5A bound to $(KR)_5$, which revealed specific and modular recognition of each arginine through electrostatic interactions with Glu30 and π–cation interactions with Trp33, as well as by the peptide bond forming hydrogen bonds to the conserved Asn37 (Figure 1C).[13] Hence, various residues were replaced by alanine in the most C-terminal M module so that the binding interface to the N-terminal fragment remained unchanged. This resulted in variants [2-W]-H23MA, with mutation in Trp37 to Ala, and [2-EWN]-H23MA, where Glu30, Trp33, and Asn37 have been simultaneously mutated to Ala. ITC analysis of their interaction with $(KR)_4$-YM_2H1 revealed a K_d of 32.9 nM for [2-W]-H23MA and a K_d of 137 nM for [2-EWN]-H23MA, which can be compared to a K_d of 17.5 nM for the interaction with the wild-type H23MA (Table 1).

To also investigate the selectivity of the enriched fragment combination for different target peptide sequences, we

Table 1: Dissociation constants and $\Delta\Delta G$ values for the binding of the $(KR)_4$ target peptide and dArmRP mutants.

Entry	XXXX-KRKR[a]	H23MA	K_d [nM]	$\Delta\Delta G$ to wt [kJ mol^{-1}]	$\Delta\Delta G$ (calc.) [kJ mol^{-1}]
1	wt	wt	17.5 ± 0.1	0	–
2	ARKR	wt	24.9 ± 3.3	0.9 ± 0.3	–
3	KRAA	wt	248 ± 25	6.6 ± 0.2	–
4	ARAA	wt	434 ± 32	8.0 ± 0.2	7.4 (entries 2+3)
5	AAAA	wt	3890 ± 117	13.4 ± 0.1	–
6	AEAA	wt	6070 ± 417	14.5 ± 0.2	–
7	wt	2-W	32.9 ± 28.4	1.6 ± 1.8	–
8	wt	2-EN	84.7 ± 8.6	3.9	–
9	wt	2-EWN	137 ± 38	5.0 ± 0.7	5.5 (entries 7+8)
10	ARAA	2-W	971 ± 13	10.0 ± 0.1	9.5 (entries 4+7)
11	ARAA	2-EN	2340 ± 54	12.1 ± 0.1	11.9 (entries 4+8)
12	ARAA	2-EWN	5090 ± 707	14.0 ± 0.4	13.4 (entries 4+9)

[a] Shown are only sequences of the first four residues, wt = KRKR.

prepared $(KR)_4$-YM_2H1 variants that contained mutations in the N-terminal tetrapeptide KRKR that interacts with the C-terminal H23MA fragment. To isolate the influence of a single target residue, we replaced three target residues by alanine residues and altered the residue of interest that faces the Arg binding pocket of the terminal M module in H23MA. ITC analysis of wt-H23MA with three (AXAAKRKR)-YM_2H1 constructs, with X being either Arg, Ala, or Glu, revealed K_d values of 0.43 μM, 3.89 μM, and 6.07 μM, respectively (Table 1). The clear preferential binding to the Arg variant confirmed the high selectivity in our split dArmRP fragment complementation system. Interestingly, (AEAAKRKR)-YM_2H1 binds even more weakly to H23MA than the Ala variant, which is likely due to electrostatic repulsion between Glu in the peptide and Glu37 of the arginine pocket (Figure 1).

Using the K_d values obtained from the ITC analysis of the $(KR)_4$-YM_2H1 variants with wt-H23MA and the [2-W]-, [2-EN]-, and [2-EWN]-H23MA mutants, we calculated the Gibbs free energy differences, $\Delta\Delta G$, to the wild-type $(KR)_4$-YM_2H1:H23MA interaction. We observed that changes in $\Delta\Delta G$ due to multiple mutations are essentially additive (Table 1). The perturbation of specific interactions between the M module and the Arg residue of the target peptide is additive, in a first approximation, irrespective of whether the mutations occur in the M module ([2-EWN]-H23MA) or in the peptide ((AAAAKRKR)-YM_2H1; Table 1). This feature illustrates the robustness of target specificity, allowing to alter single residues without triggering large and unpredictable structural changes in the dArmRP.

A clear advantage of the conditional assembly is the absence of a complementation bias in the absence of peptide, and that discrimination between C-terminal fragments is solely dictated by the relative energetic contribution of the arginine binding to the terminal M module in the H23MA variants. Nonetheless, the affinity of $(KR)_4$-YM_2H1 to the three C-terminal fragments wt-H23MA, [2-W]-H23MA, and [2-EWN]-H23MA, with K_d values of 17.5 nM, 32.9 nM, and 137 nM, is relatively similar, initially raising the question whether the C-terminal wild-type fragment can still be enriched in the mixture. We thus aimed at using NMR spectroscopy to analyze the populations in an equimolar mixture of $(KR)_4$-YM_2H1 and the three C-terminal fragments wt-H23MA, [2-W]-H23MA, and [2-EWN]-H23MA. To this end, we first prepared uniformly [^{13}C,^{15}N]-labeled H23MA and assigned the protein backbone resonances. Secondary chemical shifts[15] and ^{15}N{^{1}H} NOE data revealed rigid helix formation in helices H2 and H3 of the first truncated module, as well as in the second module and the C-cap (see the Supporting Information, Figures S1 and S2). A subsequent NMR titration of uniformly [^{15}N]-labeled H23MA with an excess of unlabeled N-terminal YM_2H1 revealed only small chemical shift perturbations (CSPs) in the fast-exchange regime (Figure S3), indicative of a K_d in the μM to mM range, which confirms our ITC measurements. In contrast, the addition of the peptide fusion $(KR)_4$-YM_2H1 to uniformly [^{15}N]-labeled H23MA resulted in large CSPs in the slow-exchange regime, which suggests a K_d in the nM range. This slow exchange allows one to relate the disappearance of resonances corresponding to the free state to the formation of the assembled state. Integration of resonances from residues sensitive to complementation for all three C-terminal variants therefore quantifies free and thus bound states of these constructs in the mixture.

To be able to discriminate and quantify the populations of each C-terminal H23MA fragment variant, we [^{15}N]-labeled each of them with a different amino acid type. We chose Ala, Trp, and Leu residues because many of these residues form different contacts in the free and assembled states.

Our home-made *E. coli*-based cell-free expression system[16,17] was used to prepare the [^{15}N-Trp]-labeled wt-H23MA, [^{15}N-Ala]-labeled [2-W]-H23MA, and [^{15}N-Leu]-labeled [2-EWN]-H23MA constructs. Comparison of [^{15}N,^{1}H] HSQC spectra of the individual H23MA variants with a spectrum of an equimolar mixture indicates that the C-terminal fragments do not interact with each other in the mixture (Figure S4).

Stepwise addition of $(KR)_4$-YM_2H1 to an equimolar mixture of the three amino-acid-selectively ^{15}N-labeled H23MA variants revealed preferential binding of the N-terminal fragment to wt-H23MA (Figure 4 and Table 2). For example, the indole resonance of Trp-149, which stems from wild-type H23MA, is affected by addition of $(KR)_4$-YM_2H1 while the amide moiety from Leu-136, which is part of the triple mutant, is not affected (Figure 4). This enrichment is particularly pronounced in the presence of only

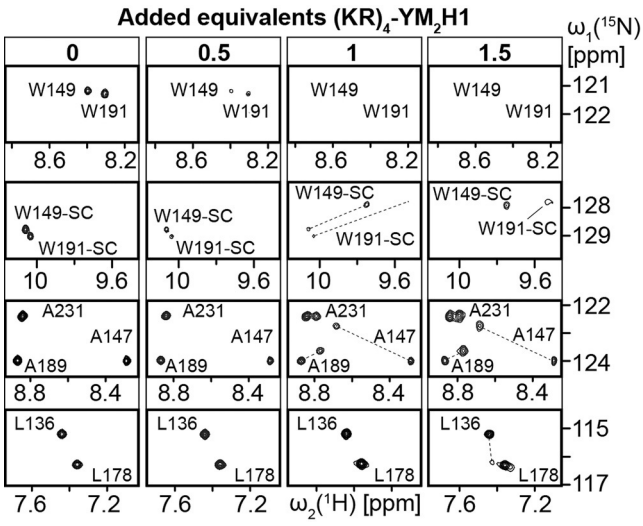

Figure 4. Population analysis in the $(KR)_4$-YM_2H1:H23MA mixture. Expansions of [^{15}N,^{1}H] HSQC spectra of a mixture containing 18 nmol each of [^{15}N-Trp]-wt-H23MA, [^{15}N-Ala]-[2W]-H23MA, and [^{15}N-Leu]-[2EWN]-H23MA in the presence of increasing amounts of unlabeled $(KR)_4$-YM_2H1 are shown for selected regions that display characteristic peaks corresponding to the free and assembled states of H23MA fragment variants. SC denotes the Trp indole side chain resonance.

Table 2: Populations of $(KR)_4$-YM_2H1:H23MA complexes.[a]

$(KR)_4$-YM_2H1 [equiv]	Bound population [%] wt-H23MA	[2W]-H23MA	[2EWN]-H23MA
0.5	29.6 ± 1.4	18.4 ± 5.5	2.0 ± 1.2
1.0	53.6 ± 2.5	42.2 ± 2.2	4.2 ± 2.9
1.5	73.7 ± 0.8	63.5 ± 2.5	12.8 ± 2.2

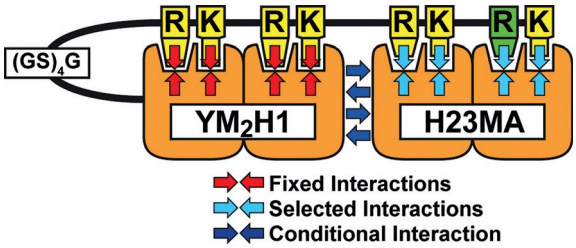

Figure 5. Important interactions for the conditional assembly of the $(KR)_4$-YM_2H1:H23MA complex. Fixed intramolecular peptide–protein interactions that occur in the free N-terminal fragment are shown with red arrows while cyan arrows indicate critical interactions that are required for the conditional assembly of the complementary dArmRP fragments. Additional binding energy is derived from the assembly of the complementary fragments (dark blue arrows) and strongly contributes to the overall stability of the complex. N- and C-terminal caps are omitted for clarity.

0.5 equiv of $(KR)_4$-YM_2H1, which interacts with about 30% wt-H23MA, 18% [2-W]-H23MA, and only 2% [2-EWN]-H23MA. This experiment clearly demonstrates the successful enrichment of the best fragment combination from a mixture, where only the interaction between a single M module and one arginine residue of the target peptide is modulated. Therefore, the particular binding energy of a single interaction with one target residue significantly favors one dArmRP fragment over the others, even for a comparably narrow K_d range of 17.5–137 nM.

Our work has demonstrated that the formation of high-affinity complexes requires the interaction of only four amino acids of the peptide with their respective binding pockets in the C-terminal fragment—usually the binding of a tetrapeptide to dArmRP would result in complexes with very weak affinity.[13] However, here the "locking in" of the side chains of the fitting peptide into the binding pockets also brings both fragments into vicinity, thereby locking the protein binding interfaces into place, releasing additional free energy through formation of the protein–protein complex (Figure 5). Thus the binding process can be considered to show cooperativity when compared to the binding of a tetrapeptide to an unsplit protein. This format leads therefore to a very strong selection power as discrimination is effectively made against a fraction of the target peptide (in our case a tetrapeptide as opposed to the entire octapeptide in an unsplit system), making it possible that the nature of a single residue has decisive discriminatory power.

Optimization of selectivity is often much more difficult to achieve than affinity optimization. We believe that this setup will be particularly useful to optimize the selectivity of binders towards peptides of only slightly different sequences, for example, when optimizing the binding of a single M module to a specific dipeptide unit. The extension of our mixture setup to large libraries will necessitate a coupling strategy of phenotype and genotype to unambiguously identify the amino acid sequence of the enriched binders. This could be achieved by application of commonly used display technologies, such as ribosome, CIS, or phage display.[3,18,19] Besides selection for best binders, the presented fragment-based recognition system could find promising applications in the development of in vivo sensor proteins, and in the formation of well-defined macromolecular assemblies in a protein-origami-type fashion.[20]

Acknowledgements

We acknowledge financial support for E.M. by a SINERGIA grant from the Swiss National Science Foundation (grant 122686) and from the Research Council of the University of Zurich (grant K-73539-01-01). We thank Stefan Vetterli for help with peptide synthesis and Ilian Jelesarov for help with ITC experiments. We acknowledge Randall Watson for helpful discussions at the early stage of the project.

Conflict of interest

The authors declare no conflict of interest.

Keywords: armadillo repeat · peptide binding · protein design · repeat proteins · split proteins

How to cite: *Angew. Chem. Int. Ed.* **2018**, *57*, 4576–4579
Angew. Chem. **2018**, *130*, 4666–4669

[1] M. S. Packer, D. R. Liu, *Nat. Rev. Genet.* **2015**, *16*, 379–394.
[2] K. Josephson, A. Ricardo, J. W. Szostak, *Drug Discovery Today* **2014**, *19*, 388–399.
[3] J. Hanes, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4937–4942.
[4] J. Kuriyan, D. Cowburn, *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 259–288.
[5] J. McCafferty, D. Schofield, *Curr. Opin. Chem. Biol.* **2015**, *26*, 16–24.
[6] M. Gebauer, A. Skerra, *Curr. Opin. Chem. Biol.* **2009**, *13*, 245–255.
[7] A. Plückthun, *Annu. Rev. Pharmacol. Toxicol.* **2015**, *55*, 489–511.
[8] C. Reichen et al., *J. Mol. Biol.* **2016**, *428*, 4467–4489.
[9] E. Conti, M. Uy, L. Leighton, G. Blobel, J. Kuriyan, *Cell* **1998**, *94*, 193–204.
[10] F. Parmeggiani et al., *J. Mol. Biol.* **2008**, *376*, 1282–1304.
[11] P. Alfarano et al., *Protein Sci.* **2012**, *21*, 1298–1314.
[12] C. Reichen, S. Hansen, A. Plückthun, *J. Struct. Biol.* **2014**, *185*, 147–162.
[13] S. Hansen et al., *J. Am. Chem. Soc.* **2016**, *138*, 3526–3532.
[14] R. P. Watson et al., *Structure* **2014**, *22*, 985–995.
[15] D. S. Wishart, B. D. Sykes, *J. Biomol. NMR* **1994**, *4*, 171–180.
[16] E. Michel, K. Wüthrich, *J. Biomol. NMR* **2012**, *53*, 43–51.
[17] E. Michel, K. Wüthrich, *FEBS J.* **2012**, *279*, 3176–3184.
[18] J. McCafferty, A. D. Griffiths, W. Ginter, D. J. Chiswell, *Nature* **1990**, *348*, 552–554.
[19] R. Odegrip et al., *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2806–2810.
[20] Q. Luo, C. Hou, Y. Bai, R. Wang, J. Liu, *Chem. Rev.* **2016**, *116*, 13571–13632.

Manuscript received: December 29, 2017
Accepted manuscript online: February 26, 2018
Version of record online: March 13, 2018

