A Bottom-Up Proteomic Approach to Identify Substrate Specificity of Outer-Membrane Protease OmpT

Sarah E. Wood[+], Gaurav Sinsinbar[+], Sushanth Gudlur, Madhavan Nallani, Che-Fan Huang, Bo Liedberg,* and Milan Mrksich*

Abstract: *Identifying peptide substrates that are efficiently cleaved by proteases gives insights into substrate recognition and specificity, guides development of inhibitors, and improves assay sensitivity. Peptide arrays and SAMDI mass spectrometry were used to identify a tetrapeptide substrate exhibiting high activity for the bacterial outer-membrane protease (OmpT). Analysis of protease activity for the preferred residues at the cleavage site (P1, P1′) and nearest-neighbor positions (P2, P2′) and their positional interdependence revealed FRRV as the optimal peptide with the highest OmpT activity. Substituting FRRV into a fragment of LL37, a natural substrate of OmpT, led to a greater than 400-fold improvement in OmpT catalytic efficiency, with a k_{cat}/K_m value of $6.1 \times 10^6 \, L \, mol^{-1} \, s^{-1}$. Wild-type and mutant OmpT displayed significant differences in their substrate specificities, demonstrating that even modest mutants may not be suitable substitutes for the native enzyme.*

The protease OmpT belongs to the omptin family of proteases, which are present on the outer membrane of most *E. coli* and functions as a defense mechanism for the bacteria by cleaving antimicrobial peptides secreted by the host epithelial cells.[1–3] It also acts as a virulence factor by aiding in the spread of the bacteria within the host cells.[4] Identifying peptide substrates that are efficiently cleaved by OmpT can thus help to improve both the sensitivity of peptide-based assays for *E. coli* detection in complex mixtures and in the design of D-amino acid based peptide inhibitors.

Peptide optimization for target proteases can be achieved using various methods, including phage display,[5,6] internally quenched fluorescent substrate libraries,[7,8] positional scanning synthetic combinatorial libraries (PS-SCL),[9,10] cellular libraries of peptide substrates (CLiPS),[11] and in silico

studies.[12] However, detection of optimal protease targets using most of the above methods relies heavily on substrate labels or tags that can alter the substrate structure and influence reaction rates.[13–15] Self-assembled monolayers for matrix-assisted laser-desorption–ionization mass spectrometry (SAMDI-MS) is an efficient approach that avoids the many limitations of traditional label-based assays and phage display methods.[16] SAMDI-MS uses self-assembled monolayers of long-chain alkanethiolates on gold-coated surfaces to which peptides (or other enzyme substrates) are immobilized (Figure 1).[17] We have shown that the SAMDI assay format is well-suited to quantitatively measure a broad range of enzyme activities and that it is compatible with high throughput formats that use arrays in the common 384 and 1536 spot formats.[18–25]

Herein we demonstrate the use of SAMDI-MS and peptide arrays in identifying FRRV, a tetrapeptide substrate for OmpT that is cleaved with high efficiency. When FRRV was substituted into a fragment of LL37, an antimicrobial peptide and a natural substrate of OmpT, the protease displayed a more than 400-fold improvement in k_{cat}/K_m value
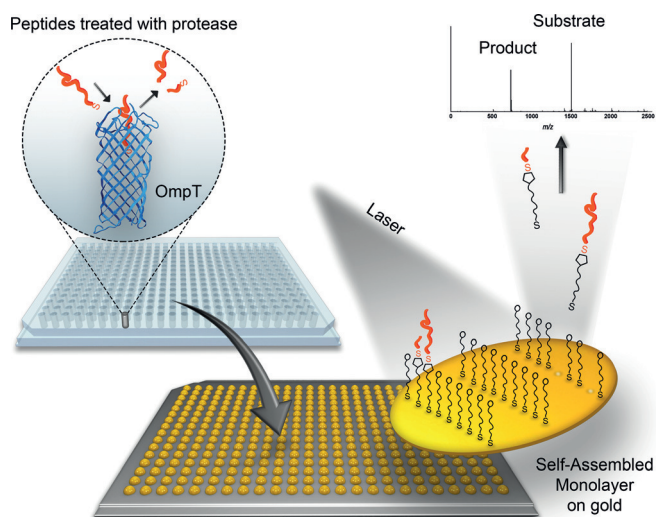
[*] S. E. Wood,[+] C.-F. Huang, Prof. M. Mrksich
Departments of Chemistry and Biomedical Engineering
Northwestern University
2145 Sheridan Road, Evanston, IL 60208 (USA)
E-mail: milan.mrksich@northwestern.edu

G. Sinsinbar,[+] Dr. S. Gudlur, Dr. M. Nallani, Prof. B. Liedberg
Center for Biomimetic Sensor Science
School of Materials Science & Engineering
Nanyang Technological University
50 Nanyang Drive, 637553 (Singapore)
E-mail: bliedberg@ntu.edu.sg

[+] These authors contributed equally to this work.

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.201707535.

Figure 1. Illustration of the method used to identify peptide substrates for the OmpT protease. Peptides from a library are individually pretreated with OmpT in 384-well plates. The solutions are then transfered to monolayer array plates, where the peptide substrate and cleaved product undergo immobilization to the maleimide-terminated monolayer. In SAMDI-MS, the monolayer is irradiated with a laser that dissociates the thiolate–gold bond, resulting in the release of ionized alkanethiolate–peptide species from the surface. The enzyme specificity and activity can be determined from characteristic mass shifts and relative peak intensities in the MS spectrum.

over the original peptide. Additionally, we compare the activity of wild-type (wt) and mutant (mut) OmpT on the same peptide array in order to gain insight into differences in the substrate specificity of the protease.

We began the process of peptide optimization with the rational design of a peptide library composed of 76 individual peptides. The library was based on a known tetrapeptide sequence (ARRA)[26] where each position was sequentially modified with one of nineteen different amino acids with the exception of cysteine, which through its thiol group might interfere with the immobilization reaction. Additionally, this starting sequence was consistent with OmpT's preference for a di-basic residue at the cleavage site (Supporting Information, Figure S1).

We synthesized the 76 peptides using standard protocols and individually treated them with wild-type protease that was recombinantly expressed in *E. coli*, and refolded from inclusion bodies. To avoid autoproteolysis of wtOmpT during its in vitro expression and purification, we and others have used a mutant form of the protease. The mutOmpT, carrying a subtle mutation in its sequence (G216-K217-R218 changed to K216-G217-R218 to remove the dibasic proteolysis site), has a circa 30% lower activity than wtOmpT and has been presumed to share the same substrate specificity.[26–30] However, we recently developed a method that suppresses autoproteolysis in wtOmpT,[31] making it possible to use the protease in high-throughput screening (HTS) assays. This method is based on the addition of excess LPS during the in vitro refolding stage of wtOmpT inclusion bodies. Although we are unable to confirm whether LPS plays a direct role in preventing wtOmpT autoproteolysis, the low levels of autoproteolysis directly translated to higher specific activity.

After treatment of the peptides with wtOmpT, small volumes of each reaction (2 µL) were transferred onto array plates having 384 gold islands that were modified with a monolayer presenting maleimide groups at a density of about 10% against a background of tri(ethylene glycol)-terminated alkanethiolates.[32] The maleimide functionality allows for the selective immobilization of cysteine-terminated peptides, and the glycol functionality is effective in preventing the non-specific adsorption of proteins.[33] The array plates were analyzed by MALDI-TOF MS to obtain a mass spectrum for each spot, and the protease activity was calculated by measuring the relative area under each molecular ion peak (M⁺) from each spectra according to the equation:

$$\text{Activity}(\%) = \text{AUC}_{\text{product}}/(\text{AUC}_{\text{substrate}} + \text{AUC}_{\text{product}}) \cdot 100\%.$$

Hence, the activity represents the percent yield in conversion of the substrate to the product, and though the activities provide a rank-ordering of the substrate turnovers, they do not provide a direct comparison of the substrates (which requires measurement of k_{cat} and K_{m}, as described below). We also note that when the substrate and product have different ionization efficiencies, the activity would be skewed and not give the actual yield of the reaction. However, we confirmed this was not the case in our study by using the peak for the tri(ethylene glycol)-terminated background alkanethiolate as an internal control; this analysis showed that the ionization

efficiencies for substrate and product differed by less than 2% (Supporting Information, Table S1).

Our results are consistent with those reported by Dekker et al.[27] using mutOmpT; the wtOmpT also has an absolute preference for charged basic residues at the P1 and P1′ positions, though with a modest preference for Arg–Arg over a combination of Arg and Lys (Figure 2, rows 2 and 3). As
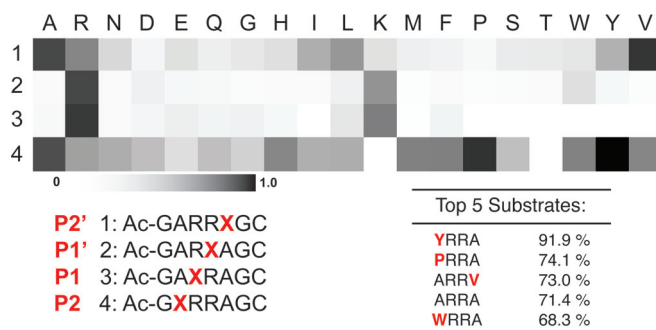

Figure 2. Wild-type OmpT activity screen on the 76 peptide library displayed as a heatmap. White areas correspond to no activity and dark areas correspond to high activity. The five most active peptide substrates are shown in the table together with the observed activity.

expected, charged acidic residues are not favored at the cleavage site nor in its immediate vicinity. This can be explained by the acidic residues in the active site where the catalytic pairs consist of Asp83–Asp85 and His212–Asp210.[28] Beyond the cleavage site, the protease displays less stringent requirements for residues at the P2 position (Figure 2, row 4), and has a preference for small hydrophobic residues like Ala and Val at the P2′ position (Figure 2, row 1). Overall, these results are mostly in agreement with previous reports obtained using mutOmpT,[27,29] thus validating the use of SAMDI-MS in determining the substrate specificity for wtOmpT. Although the protease displayed a sufficiently high activity for ARRA, we observed that three substrates-YRRA, PRRA, and ARRV-displayed higher activity, demonstrating the value of using arrays that include previously unknown substrates.

This first array was useful in determining the preferred residues at each of the four positions. However, it is recognized that there is often an interdependence of amino acids at each position in the peptide sequence, and therefore arrays having two variable positions would be expected to reveal still more active substrates. We designed a larger peptide library that was based on the sequences described above, but where the P2 and P2′ positions were varied simultaneously, with arginine held constant at the P1 and P1′ positions. This array contained 361 peptides of the form ᴬᶜGXRRZGC, where the X and Z residues are one of nineteen amino acids excluding cysteine.

We treated this larger array with wtOmpT and then analyzed each spot with SAMDI-MS to determine the activity (Figure 3, representative spectra are shown in the Supporting Information, Figure S2). Twelve tetrapeptides displayed higher activity for wtOmpT than did the ARRA peptide, which had an activity of 91.0%. Among these, FRRV
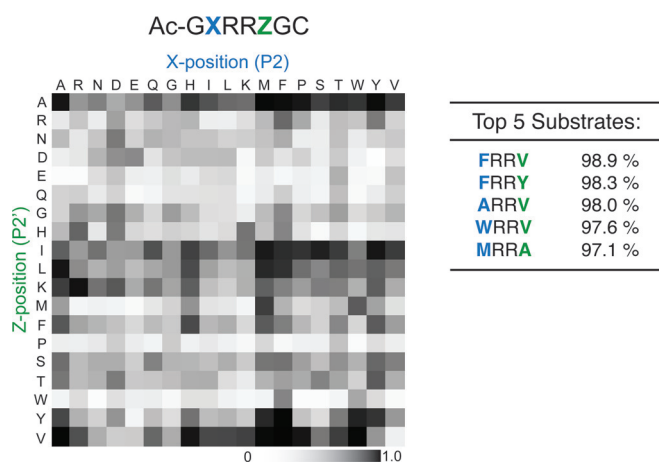
Figure 3. Wild-type OmpT activity screening of a combinatorial peptide library with substrate sequence [Ac]GXRRZGC displayed as heatmap. The top five peptide substrates and corresponding OmpT activities are displayed in the table.

Figure 4. Kinetic comparison of peptide substrates that are inserted into the LL37 fragment. A) The LL37 sequence has 2 dibasic sites that can be cleaved by OmpT. A 14 amino acid peptide from the N-terminal of LL37 was substituted with the FRRV or ARRA sequences and compared to the native sequence having the FKRS motif. B) Kinetic data for cleavage of the FRRV substrate by wtOmpT (data for the two other peptides are included in the Supporting Information, Figure S3). C) Best-fit values for K_m and k_{cat} were used to determine the k_{cat}/K_m turnover numbers for each substrate.

exhibited the highest activity of any peptide in the array, at 98.9%. Moreover, this larger peptide array also confirmed the preference for small hydrophobic residues Ala, Ile, and Val at the P2′ position and for Ala and Phe at the P2 position. Interestingly, SAMDI-MS also revealed a strong positional interdependence for the P2′ and P2 positions. For example, when the P2′ position was occupied by a small hydrophobic residue and the P2 position was occupied by an aromatic residue or methionine, there is high activity, as is the case with the peptide substrate FRRV (Figure 3).

To validate if FRRV was indeed a preferred substrate for the protease, we substituted this tetrapeptide sequence into LL37, a natural substrate for OmpT.[1,34] LL37 is a cationic antimicrobial peptide that is secreted by the epithelial cells of lungs and intestines into the extracellular environment as part of innate immunity. To establish and maintain infection, OmpT present on pathogenic *E. coli* strains is known to cleave LL37 at the two dibasic sites present in its sequence, thus rendering them inactive (Figure 4 A).[1] We synthesized two variants of a fourteen amino acid N-terminal fragment of LL37 in which the original FRKS sequence of LL37 was substituted with either FRRV or ARRA. For the unmodified LL37 fragment, wtOmpT exhibited a catalytic efficiency of 1.4×10^4 L mol^{-1} s^{-1} (Figure 4 C). However, when FRRV was substituted into this sequence, the catalytic efficiency significantly improved to 6.1×10^6 L mol^{-1} s^{-1}, which is approximately 400-fold higher than the original FRKS peptide. When the natural sequence was substituted with ARRA, wtOmpT exhibited a catalytic efficiency of 5.8×10^5 L mol^{-1} s^{-1} that is approximately 40-fold higher than for the original peptide. These experiments establish that the protease is more active for FRRV than for ARRA or the natural sequence FRKS. A sequence homology of LL37 in primates reveals that positions P2-P1-P1′ (F-R-K) for the first di-basic site are highly conserved.[35] The P2′ position is less conserved with the occurrence of either Ala or Val, the same two non-polar residues that are highly favored at this position (Supporting Information, Table S2). However, human LL37 appears to
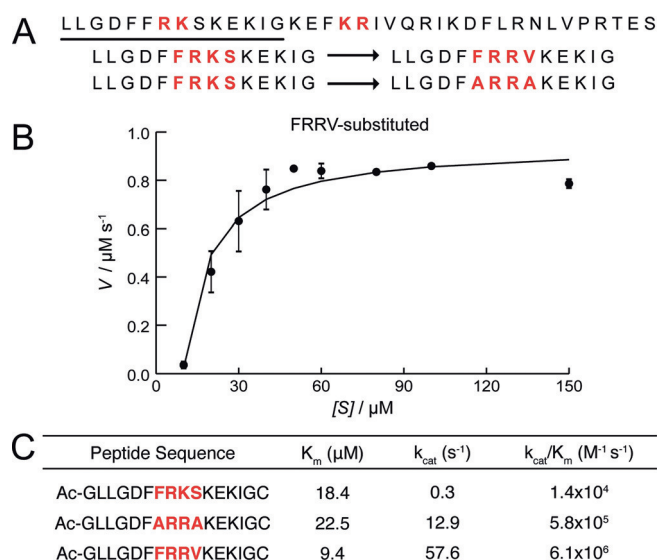
have evolved with Ser, a small polar residue, at the P2′ position and for which the protease has a relatively poor cleavage preference.[35] It may be possible that this change is a consequence of positive selection pressure driven by exposure to various bacterial pathogens.

Finally, we used the large arrays to directly compare the substrate specificities of mutOmpT (Figure 5 A) and wtOmpT (Figure 3). Interestingly, the relative difference heatmap for the two proteases (computed as a difference in activity relative to the activity of wtOmpT for the same substrate; Figure 5 B) reveal trends illustrating differences in their specificities. Despite both OmpT variants exhibiting a high catalytic activity for FRRV, we observed substantial differences in the activities of the two proteases for many of the substrates. Most noticeably, mutOmpT displayed a higher activity for peptides substituted with Ser at the P2′ position (Figure 5 B, row-S). A similar comparison of wtOmpT and mutOmpT activity on the LL37 peptide fragments again revealed a difference in their specificity. While both OmpT variants displayed similar activity for the FRRV and ARRA-substituted LL37 peptide fragments (Figure 5 C), mutOmpT displayed poor activity for the original natural sequence (FRKS). This was in contrast to the high activity of wtOmpT for the above peptide. Such stark difference in properties of the wild-type and mutant forms of the protease illustrates the ways that even modest mutants (that is, interchanging two adjacent residues outside of the active site) can lead to distinct differences in substrate specificity.

In conclusion, the use of SAMDI-MS and peptide arrays in this work was able to reveal trends in OmpT activity and
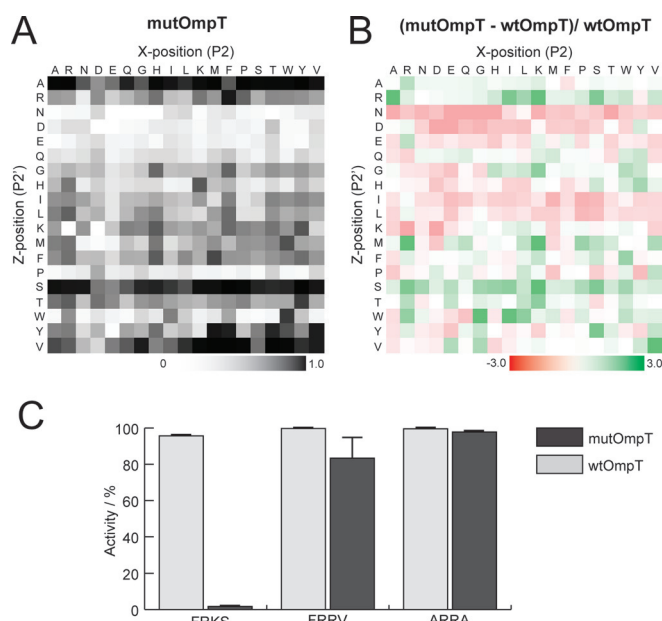
Figure 5. Mutant OmpT displays differential substrate specificity compared to wild-type. A) Heatmap of global mutOmpT on the peptide array. B) Differences in specificity for mutOmpT were determined by dividing the difference in activity of the proteases on each peptide by the activity for the wtOmpT. Peptides that are coded green were more active for mutOmpT and peptides that are coded red showed greater activity for wtOmpT C) Comparison of mutOmpT and wtOmpT on LL37 peptide fragments.

interplay between multiple positions in the peptide substrate sequence. Peptide arrays permit the unbiased evaluation of hundreds to thousands of enzyme substrates. In this example, we demonstrate that peptide arrays and SAMDI-MS are able to discover new short peptide sequences that have catalytic turnover numbers that are approximately 400-fold greater than that for the native sequence and approximately 40-fold greater than that for the previously identified "best" ARRA substrate. Further, we use the peptide arrays to reveal significant differences in substrate specificity for the wild-type and mutant proteases, showing that these forms of the OmpT protease cannot be used interchangeably, as they often have been in OmpT research. We believe this approach will be important for enabling studies of a broader range of proteases.[36,37]

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

[1] J.-L. Thomassin, J. R. Brannon, B. F. Gibbs, S. Gruenheid, H. L. Moual, *Infect. Immun.* **2012**, *80*, 483–492.
[2] S. Stumpe, R. Schmid, D. L. Stephens, G. Georgiou, E. P. Bakker, *J. Bacteriol.* **1998**, *180*, 4002–4006.
[3] J. Haiko, M. Suomalainen, T. Ojala, K. Lähteenmäki, T. K. Korhonen, *Innate Immun.* **2009**, *15*, 67–80.
[4] J. Lin, S. Huang, Q. Zhang, *Microbes Infect.* **2002**, *4*, 325–331.
[5] D. Deperthes, *Biol. Chem.* **2005**, *383*, 1107–1112.
[6] B. Ratnikov, P. Cieplak, J. W. Smith, *Methods Mol. Biol.* **2009**, *539*, 93–114.
[7] L. C. G. Oliveira, V. O. Silva, D. N. Okamoto, M. Y. Kondo, S. M. B. Santos, I. Y. Hirata, M. A. Vallim, R. C. Pascon, I. E. Gouvea, M. A. Juliano, et al., *Anal. Biochem.* **2012**, *421*, 299–307.
[8] I. Y. Hirata, M. H. S. Cezari, C. R. Nakaie, P. Boschcov, A. S. Ito, M. A. Juliano, L. Juliano, *Lett. Pept. Sci.* **1995**, *1*, 299–308.
[9] C. Pinilla, J. R. Appel, E. Borràs, R. A. Houghten, *Nat. Med.* **2003**, *9*, 118–122.
[10] E. L. Schneider, C. S. Craik, *Methods Mol. Biol.* **2009**, *539*, 59–78.
[11] K. T. Boulware, P. S. Daugherty, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7583–7588.
[12] M. Poreba, M. Drag, *Curr. Med. Chem.* **2010**, *17*, 3968–3995.
[13] Y. S. Sun, J. P. Landry, Y. Y. Fei, X. D. Zhu, J. T. Luo, X. B. Wang, K. S. Lam, *Langmuir* **2008**, *24*, 13399–13405.
[14] T. Kodadek, *Chem. Biol.* **2001**, *8*, 105–115.
[15] Y. Fei, Y.-S. Sun, Y. Li, K. Lau, H. Yu, H. A. Chokhawala, S. Huang, J. P. Landry, X. Chen, X. D. Zhu, *Mol. Biosyst.* **2011**, *7*, 3343–3352.
[16] J. Su, M. Mrksich, *Angew. Chem. Int. Ed.* **2002**, *41*, 4715–4718; *Angew. Chem.* **2002**, *114*, 4909–4912.
[17] M. Mrksich, G. B. Sigal, G. M. Whitesides, *Langmuir* **1995**, *11*, 4383–4385.
[18] J. Kim, M. Mrksich, *Nucleic Acids Res.* **2010**, *38*, e2.
[19] J. Su, T. W. Rajapaksha, M. E. Peter, M. Mrksich, *Anal. Chem.* **2006**, *78*, 4945–4951.
[20] Z. A. Gurard-Levin, M. Mrksich, *Biochemistry* **2008**, *47*, 6242–6250.
[21] Z. A. Gurard-Levin, J. Kim, M. Mrksich, *ChemBioChem* **2009**, *10*, 2159–2161.
[22] Z. A. Gurard-Levin, K. A. Kilian, J. Kim, K. Bähr, M. Mrksich, *ACS Chem. Biol.* **2010**, *5*, 863–873.
[23] D.-H. Min, J. Su, M. Mrksich, *Angew. Chem. Int. Ed.* **2004**, *43*, 5973–5977; *Angew. Chem.* **2004**, *116*, 6099–6103.
[24] M. D. Cabezas, C. A. Mirkin, M. Mrksich, *Nano Lett.* **2017**, *17*, 1373–1377.
[25] E. J. Berns, M. D. Cabezas, M. Mrksich, *Small* **2016**, *12*, 3811–3818.
[26] R. A. Kramer, D. Zandwijken, M. R. Egmond, N. Dekker, *Eur. J. Biochem.* **2000**, *267*, 885–893.
[27] N. Dekker, R. C. Cox, R. A. Kramer, M. R. Egmond, *Biochemistry* **2001**, *40*, 1694–1701.
[28] L. Vandeputte-Rutten, R. A. Kramer, J. Kroon, N. Dekker, M. R. Egmond, P. Gros, *EMBO J.* **2001**, *20*, 5033–5039.
[29] J. D. McCarter, D. Stephens, K. Shoemaker, S. Rosenberg, J. F. Kirsch, G. Georgiou, *J. Bacteriol.* **2004**, *186*, 5919–5925.

[30] K. Okuno, M. Yabuta, K. Kawanishi, K. Ohsuye, T. Ooi, S. Kinoshita, *Biosci. Biotechnol. Biochem.* **2002**, *66*, 127–134.

[31] G. Sinsinbar, S. Gudlur, K. Metcalf, M. Mrksich, B. L. Liedberg, M. Nallani, unpublished results.

[32] B. T. Houseman, E. S. Gawalt, M. Mrksich, *Langmuir* **2003**, *19*, 1522–1531.

[33] C. Pale-Grosdemange, E. S. Simon, K. L. Prime, G. M. Whitesides, *J. Am. Chem. Soc.* **1991**, *113*, 12–20.

[34] J. R. Brannon, J.-L. Thomassin, I. Desloges, S. Gruenheid, H. Le Moual, *FEMS Microbiol. Lett.* **2013**, *345*, 64–71.

[35] I. Zelezetsky, A. Pontillo, L. Puzzi, N. Antcheva, L. Segat, S. Pacor, S. Crovella, A. Tossi, *J. Biol. Chem.* **2006**, *281*, 19861–19871.

[36] P. Chen, R. Selegård, D. Aili, B. Liedberg, *Nanoscale* **2013**, *5*, 8973–8976.

[37] X. Liu, Y. Wang, P. Chen, Y. Wang, J. Zhang, D. Aili, B. Liedberg, *Anal. Chem.* **2014**, *86*, 2345–2352.

Angew. Chem. Int. Ed. **2017**, *56*, 16531–16535

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

www.angewandte.org **16535**