Biocatalysis

DOI: 10.1002/anie.201410047

A Synthetic Adenylation-Domain-Based tRNA-Aminoacylation Catalyst**

Tobias W. Giessen, Florian Altegoer, Annika J. Nebel, Roman M. Steinbach, Gert Bange,* and Mohamed A. Marahiel**

Abstract: *The incorporation of non-proteinogenic amino acids represents a major challenge for the creation of functionalized proteins. The ribosomal pathway is limited to the 20–22 proteinogenic amino acids while nonribosomal peptide synthetases (NRPSs) are able to select from hundreds of different monomers. Introduced herein is a fusion-protein-based design for synthetic tRNA-aminoacylation catalysts based on combining NRPS adenylation domains and a small eukaryotic tRNA-binding domain (Arc1p-C). Using rational design, guided by structural insights and molecular modeling, the adenylation domain PheA was fused with Arc1p-C using flexible linkers and achieved tRNA-aminoacylation with both proteinogenic and non-proteinogenic amino acids. The resulting aminoacyl-tRNAs were functionally validated and the catalysts showed broad substrate specificity towards the acceptor tRNA. Our strategy shows how functional tRNA-aminoacylation catalysts can be created for bridging the ribosomal and nonribosomal worlds. This opens up new avenues for the aminoacylation of tRNAs with functional non-proteinogenic amino acids.*

In nature, aminoacyl-tRNAs are generated by the ancient and ubiquitous enzyme family of aminoacyl-tRNA synthetases (aaRSs) in a two-step process:[1] Firstly, the respective amino acid is selected and activated in an ATP-dependent reaction resulting in the generation of a reactive adenylate intermediate. Secondly, this intermediate is then transferred to the terminal 3′ adenosine residue of the appropriate tRNA. In eukaryotes and archaea aaRSs associate into large multisynthetase complexes (MSCs) which in addition to the synthetases themselves contain other accessory protein factors.[2] Eukaryotic MSCs can contain between two (*S. cerevisiae*) and nine different synthetases (mammals) and up to three noncatalytic protein components.[3] The protein factors present in MSCs have been shown to be important for complex formation and stability by acting as scaffolds for complex assembly.[2,3] In addition, the accessory protein Arc1p has been shown to facilitate tRNA recruitement to the yeast MSC, thus increasing its overall aminoacylation efficiency.[3b,4] Domains homologous to Arc1p have been identified as C-terminal appendages in different eukaryotic aaRSs, thus highlighting the fusion of protein domains as an important mechanism for molecular evolution which can lead to improved functionalities.[3e,4,5] In contrast to ribosomal protein synthesis, the nonribosomal assembly of peptides, found in the secondary metabolism of many microbes, is carried out by large multimodular enzyme complexes called nonribosomal peptide synthetases (NRPSs).[6] NRPSs represent another important example of the successive fusion of protein domains leading to highly sophisticated enzyme catalysts over evolutionary time.[7] Numerous important bioactive compounds are synthesized by NRPS-dependent pathways and are used as last-resort antibiotics (e.g. vancomycin, NRPS),[8] immunosuppressants (e.g. rapamycin, PKS/NRPS),[9] or antineoplastic compounds (e.g. epothilone D, PKS/NRPS).[10] Each NRPS core module contains adenylation (A) domains responsible for monomer recognition and activation. The two reactions catalyzed by NRPS A-domains closely resemble those carried out by aaRSs (Figure 1a). In both cases, the initial activation of a building block results in the formation of a high-energy adenylate intermediate which is subsequently attacked by a reactive nucleophile. In the case of A-domains, a nucleophilic attack by the thiol group of the 4′-phosphopantethein cofactor (4′-PPant) of the downstream peptidyl carrier protein (PCP) domain links the monomer to the assembly line through a thioester linkage. In aaRS catalysis, the reactive adenylate intermediate is attacked by the 2′- or 3′-hydroxy group of the terminal adenosine residue of a tRNA molecule, thus forming an oxoester bond between amino acid and tRNA. Thus, from a chemical perspective, the only difference between the reactions catalyzed by NRPS A-domains and aaRSs is the final nucleophile which is used as the acceptor of the initially selected monomer. However, aaRSs are restricted to the 20–22 proteinogenic amino acids, while A-domains can select from hundreds of proteinogenic and non-proteinogenic building blocks.[11] Inspired by this notion we envisioned to exploit the large monomer repertoire of NRPSs and to utilize it within the context of the ribosomal assembly pathway, thus combining the ribosomal and nonribosomal worlds of peptide biosynthesis.

[*] Dr. T. W. Giessen, A. J. Nebel, R. M. Steinbach, Dr. G. Bange, Prof. Dr. M. A. Marahiel
Department of Chemistry, Philipps-University Marburg
Hans-Meerwein-Strasse 4, 35032 Marburg (Germany)
E-mail: tobias.giessen@chemie.uni-marburg.de
marahiel@staff.uni-marburg.de

Dr. T. W. Giessen, F. Altegoer, Dr. G. Bange, Prof. Dr. M. A. Marahiel
LOEWE Center for Synthetic Microbiology (Synmikro)
Philipps-University Marburg
Hans-Meerwein-Strasse, 35032 Marburg (Germany)
E-mail: gert.bange@synmikro.uni-marburg.de

[**] We thank the European Synchrotron Radiation Facility (ESRF), Grenoble, France for support during data collection. T.W.G., G.B., and M.A.M. thank the LOEWE program of the state of Hesse for supporting this work. M.A.M. and G.B. thank the Sonderforschungsbereich SFB987 for support.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201410047.

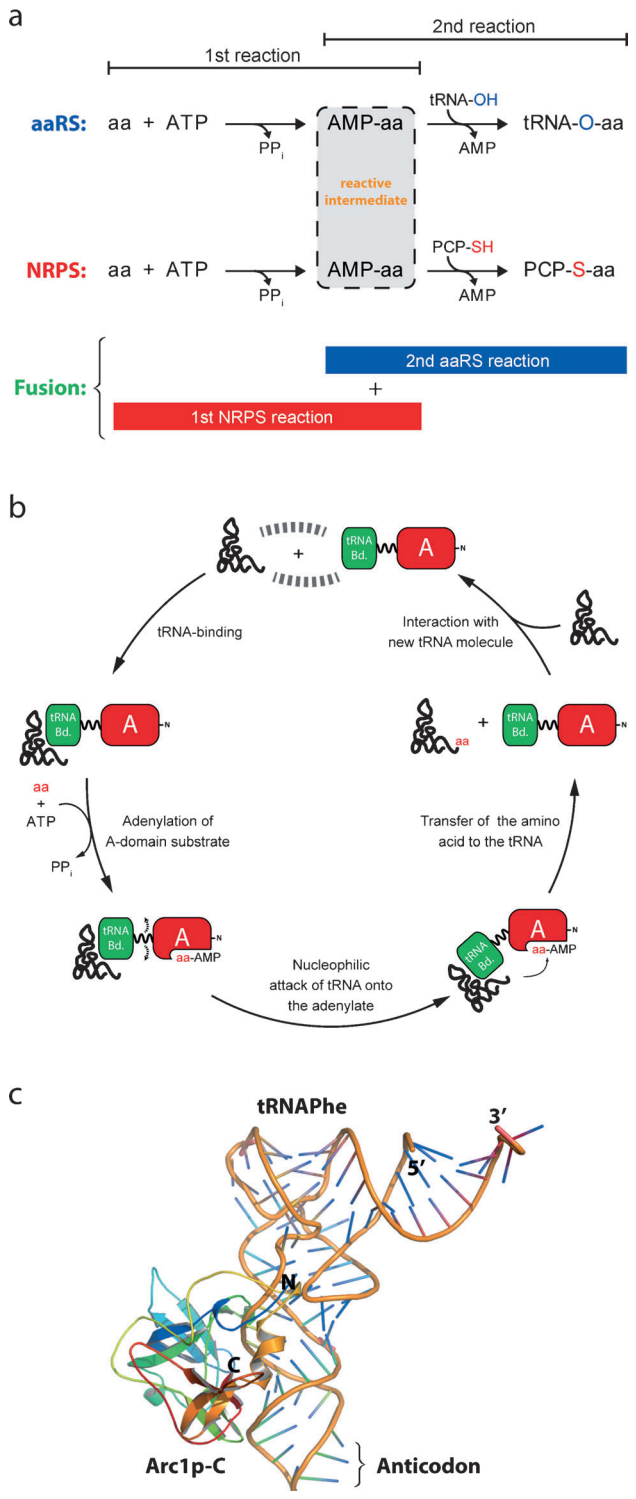

Figure 1. Design strategy for the construction of synthetic adenylation-domain-based fusion proteins for tRNA-aminoacylation based on structural insights and molecular models. a) Comparative overview of the two reactions catalyzed by aaRSs and NRPS A-domains, respectively. b) Exemplary catalytic cycle of A-domain-based tRNA-aminoacylation. Note that the shown order of tRNA-binding and adenylate formation only represents one of several possibilities. c) Docking model showing the interaction of tRNA with Arc1p-C. Shown is the functionally most reasonable model, where both the acceptor stem and the anticodon are accessible. Arc1p-C is colored from blue to yellow (N- to C-terminus).

We set out to rationally design a synthetic tRNA-aminoacylation catalyst by exploiting the analogous chemistry of aaRSs and NRPS A-domains (Figure 1b). From a design perspective, every tRNA-aminoacylation catalyst needs to possess two important functionalities: Firstly, the capability to activate monomers because oxoester formation between unactivated carboxylic acids and alcohols does not proceed efficiently under physiological conditions.[25,26] Secondly, it needs the ability to facilitate the binding and positioning of reactants (monomers and tRNAs), thus leading to a local increase in effective molarity. This local increase in reactant concentration is able to elevate reactant molarities from the nano- and micromolar range to the millimolar and molar range , respectively.[12] In our design, the well-characterized A-domain PheA from *Bacillus brevis* is used as the catalytic domain. PheA has been shown to be active outside of the context of its cognate NRPS module as a lone-standing enzyme.[13] Additionally, PheA can form reactive adenylate intermediates of both L- and D-phenylalanine (Phe) equally well,[13a] thus giving us the opportunity to test if it is possible to load a non-proteinogenic D-amino acid onto tRNAs using our synthetic fusion-protein-based catalyst. To confer tRNA-binding capabilities to our synthetic fusion protein, the functionally characterized C-terminal part of Arc1p (*S. cerevisiae*) was employed (referred to as Arc1p-C),[3e,4] which was previously used to construct a MetRS-Arc1p-C fusion protein in yeast that showed improved aminoacylation activity.[14] To now rationally design a functional tRNA-aminoacylation catalyst based on the above-mentioned components, we initially aimed at constructing a molecular model. With the crystal structure of PheA known,[13b] we still needed to solve the crystal structure of Arc1p-C to build an accurate model which would allow us to rationally design expression constructs for activity testing. The crystal structure of Arc1p-C was determined at 1.40 Å resolution (see Figure S1 and Table S1 in the Supporting Information; atomic coordinates and structure factors have been deposited with the Protein Data Bank (PDB) under the accession code 4R1J). Arc1p-C superimposes very well with the C-terminal domain of human EMAPII (see Figures S2 and S3) with RMSD values of 0.71 Å (over 163 Cα atoms).[15,16] To acquire a deeper understanding of the Arc1p-C–tRNA interaction, we used the DUCK-BP server employing the FTdock algorithm[17] for docking experiments using yeast $tRNA^{Phe}$. While five docking sets were generated, only the highest-scoring set was reasonable from a functional perspective, leaving the acceptor stem accessible for aminoacylation (Figure 1c; see Figures S4 and S5). For all sets, the tRNA molecule interacted precisely with the previously proposed groove and loop regions of Arc1p-C (Figure S4).[18] Additionally, a C-terminally appended EMAPII domain in human TyrRS, homologous to Arc1p, was shown to very likely interact with $tRNA^{Tyr}$ in the same way as suggested here.[19] Interestingly, a glycerol molecule is found inside the hydrophobic pocket of Arc1p-C, likely mimicking the interaction with a tRNA ribose (Figure S1). Our results suggest that the observed promiscuity of Arc1p with regard to tRNA binding likely stems from the particular interaction mode proposed here where no specificity-conferring structural determinants (anticodon-loop or acceptor

stem) are directly involved in the interaction. With both structures in hand and a proposed tRNA-binding mode established, we were able to generate a molecular model where both functional domains (PheA and Arc1p-C) were connected by flexible glycine-serine-linkers of various lengths ($[GSSG]_n$, $n = 0$–8, ≈0–120 Å, Figure S6). Our models suggest that all catalysts, except for the linkerless version, should in principle allow the 3′-end of bound tRNA to be in proximity to the active site of PheA. However, it is evident that the longer linkers would likely lead to too much flexibility, thus drastically lowering the likelihood of aminoacylation. Based on our models, we expect that a linker length of $n = 1$ or 2 should be optimal for both reactant positioning and increase in local reactant concentration. Guided by the insights gained through our molecular models, we designed various expression constructs for activity testing (see Figure S7). We focused on constructs where Arc1p-C was C-terminally fused to PheAl (PheA-Arc1p-C). The reasoning was that the opposite domain configuration (Arc1p-C-PheA) would block the observed tRNA binding site of Arc1p-C. As additional controls, single PheA and Arc1p-C constructs, not connected by a linker, were used to make sure that any observed activities were actually a result of the fusion of both domains.

All proteins used in this study were recombinantly produced in *E. coli* as His_6-tagged fusion constructs (see Figure S8). Prior to carrying out aminoacylation assays, the ability of all PheA constructs to adenylate L-Phe was tested to ensure that fusion of Arc1p-C did not disturb PheA catalysis. All investigated constructs showed similar levels of Phe activation compared to the stand-alone PheA construct (see Figure S9). To test the designed constructs for their ability to aminoacylate tRNAs, we initially used commercially available $tRNA^{Phe}$ (*E. coli*) and carried out standard aminoacylation assays using either L- or D-Phe.[20] To determine aminoacylation levels, we employed a formerly established method based on size-exclusion chromatography, hydrolysis of bound amino acids, and subsequent LC-MS analysis of the released previously tRNA-bound monomers (see Figures S10 and S11).[21] To validate this approach we employed an additional analysis strategy where ^{14}C-labled amino acids were used in aminoacylation assays, followed by phenol extraction, size-exclusion chromatography, and scintillation counting of elution fractions (see Figure S12). Phenylalanyl-tRNA synthetase (PheRS) from *E. coli* was employed as a positive control. By using our analysis strategy, good aminoacylation levels could be observed for PheRS which were comparable to values reported in the literature.[1] In general, all fusion constructs were able to load both L- and D-Phe onto $tRNA^{Phe}$ (Figure 2; see Figures S13 and S14). Varying linker length revealed a clear influence of domain distance on the observed aminoacylation levels, with the 2x-construct, possessing an eight-amino-acid linker ($[GSSG]_2$), showing the highest aminoacylation efficiency and reaching 11% loaded tRNAs after 30 minutes (corresponding to 275 pmol of Phe-$tRNA^{Phe}$). The 8x-construct ($[GSSG]_8$) showed only about 3% loaded tRNAs after 30 minutes, while the linkerless fusion constructs showed similar results. All other constructs yielded intermediate levels of aminoacylation. Importantly, assays where the single PheA and Arc1p-C constructs were used, showed only background levels of aminoacylation, thus clearly indicating that the fusion proteins did facilitate tRNA-loading through spatial control and reactant positioning. The observation that the 1x- and 2x-constructs exhibit the highest aminoacylation efficiency is in agreement with our molecular models. Although, other artificial tRNA-aminoacylation systems have been previously devised using self-acylating ribozymes,[22] our approach is the first one to exploit the inherent modularity of protein domains to create a synthetic protein-based tRNA-aminoacylation catalyst. Our results highlight the importance of interdomain distance and domain configuration, which is often observed for synthetic fusion constructs.[23] The time-

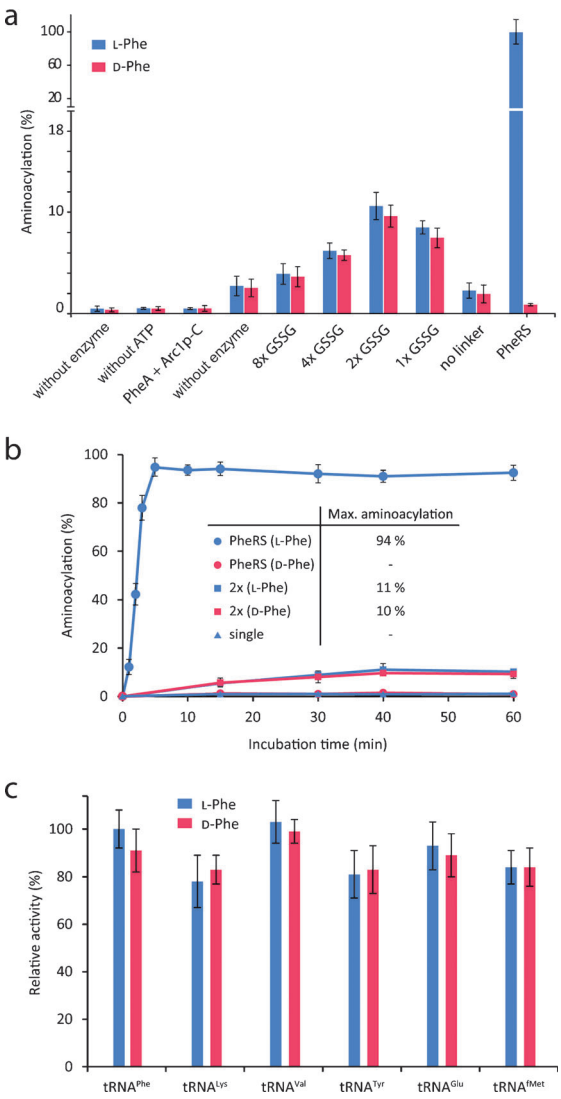

Figure 2. Functional characterization of fusion proteins. Results for assays containing L-Phe are shown in blue, while results for D-Phe-containing assays are shown in red. Error bars represent the standard deviation of three independent experiments. a) Observed aminoacylation activity of fusion constructs and controls. b) Time-dependence of aminoacylation of the 2x-construct compared with single and positive controls. Maximal aminoacylation is shown in the inset. c) Substrate specificity with regards to the tRNA acceptor.

dependence of aminoacylation showed that both L- and D-Phe loading levels reached a maximum after 30 minutes with no further increase in aminoacylation levels over time (Figure 2b).

To address the question of whether our 2x-construct would also be able to utilize other acceptor tRNAs as substrates, we carried out aminoacylation assays with both L- and D-Phe using five additional *E. coli* tRNAs ($tRNA^{Lys}$, $tRNA^{Val}$, $tRNA^{Tyr}$, $tRNA^{Glu}$ and $tRNA^{fMet}$). Our results indicate that, in agreement with previous studies where Arc1p-C was shown to bind to various tRNA species, all the tested tRNAs could be aminoacylated similarly well (Figure 2c).[3e,4] Although, previous competition experiments using total yeast tRNA suggest that Arc1p may exhibit limited tRNA specificity, favoring $tRNA^{Glu}$ and $tRNA^{Met}$ (part of the yeast MSC),[24] no preference could be observed here. This lack of preference may indicate that not tRNA binding, but aminoacylation represents the rate-limiting step in fusion protein catalysis.

To determine at which position the 2x-construct attaches L- and D-Phe to tRNAs, we used a combined approach based on P1 nuclease digestion coupled with Antarctic phosphatase 5′-dephosphorylation and subsequent high-resolution MS, together with a functional analysis of the resulting aminoacyl-tRNAs (Figure 3a). It could be shown that both L- and D-Phe were coupled to adenosine residues (Figure 3b; see Figures S15–S18) while no other Phe-nucleoside adducts could be detected (Figure S14 and S15). The intensities observed for the phenylalanine-adenosine (Phe-A) species in the assays containing the 2x-construct or PheRS correlate well with the previously determined aminoacylation levels, where the 2x-construct generates about ten percent of the aminoacyl-tRNAs synthesized by PheRS. To finally verify that the loaded tRNAs produced by our 2x-construct are functional, meaning that Phe would be bound to the terminal 3′-adenosine residue, we devised an assay where an enzyme which exclusively uses aminoacyl-tRNAs as substrates was employed. The enzyme used was AlbC (*Streptomyces noursei*), belonging to the family of cyclodipeptide synthases (CDPSs).[25] CDPSs are able to use loaded tRNAs to carry out two successive peptide-bond formations, thus resulting in the generation of a cyclic dipeptide scaffold (Figure 3c). AlbC in particular is able to use Phe-$tRNA^{Phe}$ (*E. coli*) to form cyclic diphenylalanine (cFF) as one of its main products in vivo and in vitro. We were able to show that the aminoacyl-tRNAs resulting from assays containing both the 2x-construct and L-Phe could be used by AlbC to form cFF, while loaded tRNAs resulting from D-Phe assays were no valid substrates for AlbC (Figure 3d; see Figure S19). Considering the structure of the active site pocket of AlbC and its catalytic ping-pong

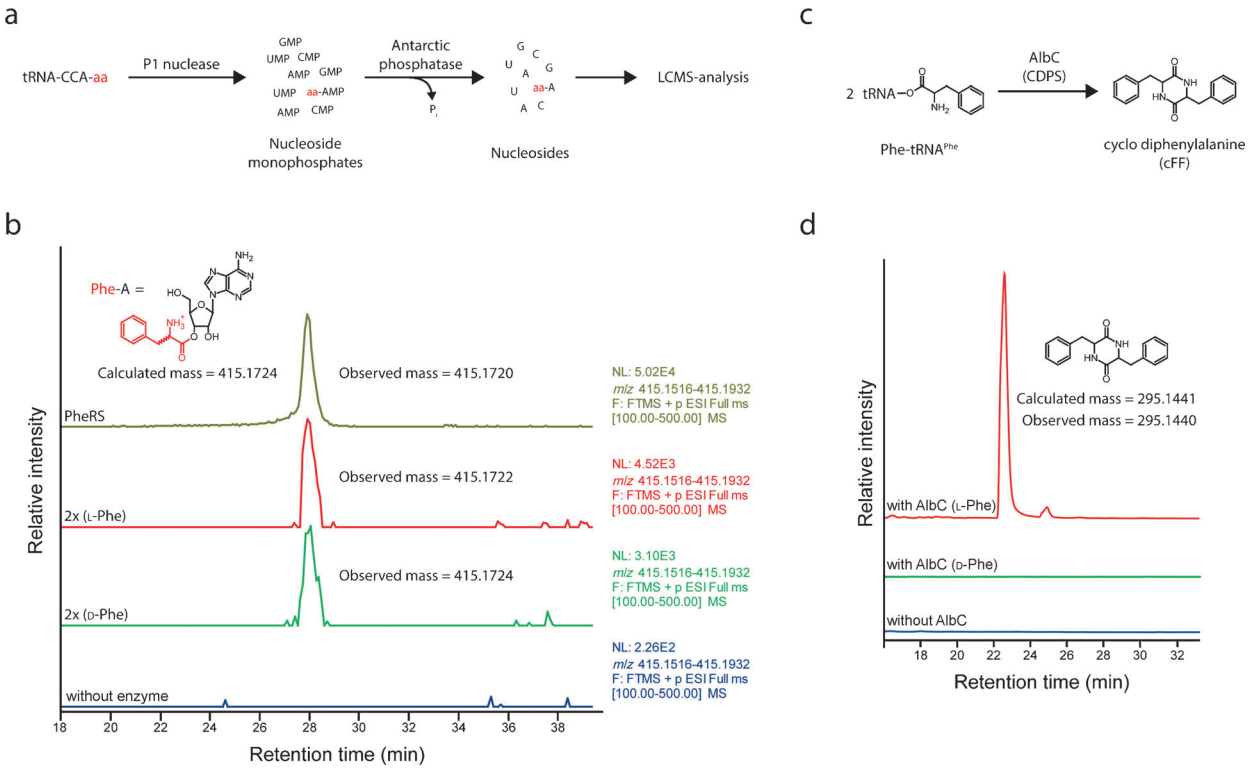

Figure 3. a) The analysis strategy used to determine the aminoacylation position. All steps are carried out under acidic conditions to minimize hydrolysis of tRNA-bound monomers. b) LC-MS analysis of nucleoside solutions resulting from successive P1 nuclease and Antarctic phosphatase treatment. Shown are the results for the positive control PheRS (ochre), assays containing the 2x-construct using L- and D-Phe (colored in red and green, respectively), and a control without enzyme (blue). To the right of the chromatograms peak intensities are shown [NL: normalized intensity level (counts per second)]. c) Reaction catalyzed by the CDPS AlbC. d) LC-MS analysis of aminoacylation assays treated with AlbC. Shown are chromatograms (single-ion monitoring: m/z 295.1233 to 295.1649) of 2x-construct assays in the presence of L- and D-Phe (shown in red and green, respectively) and a control resulting from an L-Phe/2x-construct assay without the addition of AlbC (blue).

mechanism, it can be easily rationalized that only L-configured amino acids represent substrates for cyclic dipeptide formation.[26] Taken together, those results indicate that our synthetic fusion-protein-based aminoacylation catalyst is able to generate functional aminoacyl-tRNAs.

In summary, we introduced a new modular design for protein-based tRNA-aminoacylation catalysts starting from rationally designed models. Our catalysts are able to load proteinogenic and non-proteinogenic monomers onto various tRNAs. It can be easily imagined that many other catalysts could be created based on this design, thus opening up a new approach to produce tRNAs loaded with specific functional non-natural monomers.

Received: October 13, 2014
Revised: December 1, 2014
Published online: January 12, 2015

Keywords: biocatalysis · peptides · protein engineering · protein structure · RNA

[1] M. Ibba, C. Francklyn, S. Cusack, Landes Bioscience, Georgetown, Texas, **2005**.
[2] S. W. Lee, B. H. Cho, S. G. Park, S. Kim, *J. Cell Sci.* **2004**, *117*, 3725 – 3734.
[3] a) J. C. Robinson, P. Kerjan, M. Mirande, *J. Mol. Biol.* **2000**, *304*, 983 – 994; b) S. Quevillon, F. Agou, J. C. Robinson, M. Mirande, *J. Biol. Chem.* **1997**, *272*, 32573 – 32579; c) S. Quevillon, M. Mirande, *FEBS Lett.* **1996**, *395*, 63 – 67; d) M. Kaminska, S. Havrylenko, P. Decottignies, P. Le Marechal, B. Negrutskii, M. Mirande, *J. Biol. Chem.* **2009**, *284*, 13746 – 13754; e) G. Simos, A. Segref, F. Fasiolo, K. Hellmuth, A. Shevchenko, M. Mann, E. C. Hurt, *EMBO J.* **1996**, *15*, 5437 – 5448.
[4] G. Simos, A. Sauer, F. Fasiolo, E. C. Hurt, *Mol. Cell* **1998**, *1*, 235 – 242.
[5] D. A. Liberles et al., *Protein Sci.* **2012**, *21*, 769 – 785.
[6] a) G. H. Hur, C. R. Vickery, M. D. Burkart, *Nat. Prod. Rep.* **2012**, *29*, 1074 – 1098; b) M. Strieker, A. Tanovic, M. A. Marahiel, *Curr. Opin. Struct. Biol.* **2010**, *20*, 234 – 240; c) M. A. Fischbach, C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468 – 3496.
[7] a) A. Gallo, M. Ferrara, G. Perrone, *Toxins* **2013**, *5*, 717 – 742; b) H. Jenke-Kodama, E. Dittmann, *Nat. Prod. Rep.* **2009**, *26*, 874 – 883; c) M. A. Fischbach, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4601 – 4608.
[8] M. L. Avent, V. L. Vaska, B. A. Rogers, A. C. Cheng, S. J. van Hal, N. E. Holmes, B. P. Howden, D. L. Paterson, *Intern. Med. J.* **2013**, *43*, 110 – 119.
[9] V. C. McAlister, K. Mahalati, K. M. Peltekian, A. Fraser, A. S. MacDonald, *Ther. Drug Monit.* **2002**, *24*, 346 – 350.
[10] S. Rosenberg, V. T. DeVita, T. Vincent, S. Hellman, *Cancer: Principles & Practice of Oncology*, 7th ed., Lippincott Williams & Wilkins, Hagerstown, MD, **2005**.
[11] S. Caboche, M. Pupin, V. Leclere, A. Fontaine, P. Jacques, G. Kucherov, *Nucleic Acids Res.* **2008**, *36*, D326 – 331.
[12] X. Li, D. R. Liu, *Angew. Chem. Int. Ed.* **2004**, *43*, 4848 – 4870; *Angew. Chem.* **2004**, *116*, 4956 – 4979.
[13] a) B. W. Stevens, R. H. Lilien, I. Georgiev, B. R. Donald, A. C. Anderson, *Biochemistry* **2006**, *45*, 15495 – 15504; b) E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997**, *16*, 4174 – 4183.
[14] E. Karanasios, H. Boleti, G. Simos, *J. Mol. Biol.* **2008**, *381*, 763 – 771.
[15] S. Kawaguchi, J. Muller, D. Linde, S. Kuramitsu, T. Shibata, Y. Inoue, D. G. Vassylyev, S. Yokoyama, *EMBO J.* **2001**, *20*, 562 – 569.
[16] L. Renault, P. Kerjan, S. Pasqualato, J. Menetrey, J. C. Robinson, S. Kawaguchi, D. G. Vassylyev, S. Yokoyama, M. Mirande, J. Cherfils, *EMBO J.* **2001**, *20*, 570 – 578.
[17] H. A. Gabb, R. M. Jackson, M. J. Sternberg, *J. Mol. Biol.* **1997**, *272*, 106 – 120.
[18] K. Nakanishi, Y. Ogiso, T. Nakama, S. Fukai, O. Nureki, *Nat. Struct. Mol. Biol.* **2005**, *12*, 931 – 932.
[19] X.-L. Yang, J. Liu, R. J. Skene, D. E. McRee, P. Schimmel, *Helv. Chim. Acta* **2003**, *86*, 1246 – 1257.
[20] S. E. Walker, K. Fredrick, *Methods* **2008**, *44*, 81 – 86.
[21] Y. G. Chen, W. E. Kowtoniuk, I. Agarwal, Y. Shen, D. R. Liu, *Nat. Chem. Biol.* **2009**, *5*, 879 – 881.
[22] a) H. Murakami, H. Saito, H. Suga, *Chem. Biol.* **2003**, *10*, 655 – 662; b) J. Xu, B. Appel, D. Balke, C. Wichert, S. Muller, *ChemBioChem* **2014**, *15*, 1200 – 1209.
[23] X. Chen, J. L. Zaro, W. C. Shen, *Adv. Drug Delivery Rev.* **2013**, *65*, 1357 – 1369.
[24] K. Deinert, F. Fasiolo, E. C. Hurt, G. Simos, *J. Biol. Chem.* **2001**, *276*, 6000 – 6008.
[25] a) P. Belin, M. Moutiez, S. Lautru, J. Seguin, J. L. Pernodet, M. Gondry, *Nat. Prod. Rep.* **2012**, *29*, 961 – 979; b) M. Gondry et al., *Nat. Chem. Biol.* **2009**, *5*, 414 – 420.
[26] L. Sauguet, et al., *Nucleic Acids Res.* **2011**, *39*, 4475 – 4489.

