Nucleobase Variations

DOI: 10.1002/anie.201103229

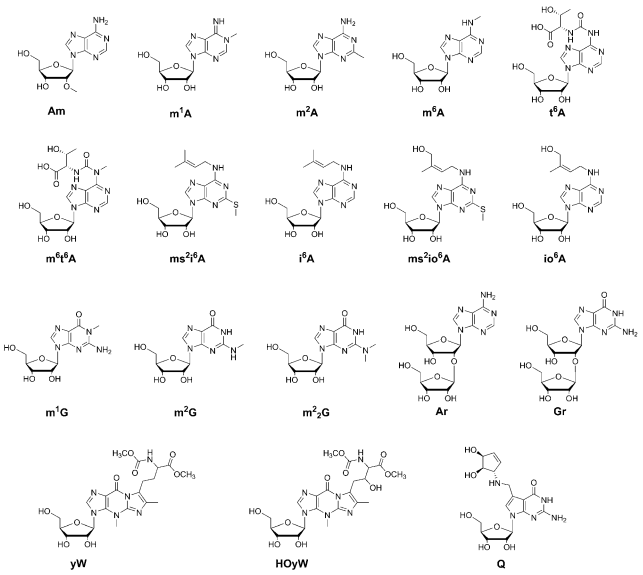
Systems-Based Analysis of Modified tRNA Bases**

*Daniel Globisch, David Pearson, Antje Hienzsch, Tobias Brückl, Mirko Wagner, Ines Thoma, Peter Thumbs, Veronika Reiter, Andrea Christa Kneuttinger, Markus Müller, Stephan A. Sieber, and Thomas Carell**

The genetic system contains several levels of information. Firstly, the sequence of the canonical bases A, C, G, and T/U in DNA and RNA encodes amino acids through specific base triplets. Secondly, the methylation status of the cytosine base in DNA imprints epigenetic information into the genetic system, thereby contributing to the division of genes into active and inactive elements. Both information layers are chemically well investigated. Less is known about a putative third level of information associated with the chemical modification of RNA nucleobases. Although RNA, and in particular tRNA, is known to contain more than 100 different modified nucleosides,[1] the exact type of information added by base modification is largely unknown. A number of common modifications have been shown to improve the maintenance of the reading frame,[2] influence RNA stability,[3] and to be involved in proof-reading by tRNA synthetases.[4]

Recently, it was discovered that the collective set of modified tRNA nucleosides is a regulated component of stress response and gives us a first hint that cells may actively adjust the modification pattern in response to external factors.[5] To learn more about the function of modified nucleobases we have investigated relationships between species by quantification of the modification content. By using a parallel systems-type approach we discovered that the collective set of modified bases is highly species-specific and linked to phylogeny. These data then enabled us to calculate a detailed phylogenetic tree that is consistent with those obtained from traditional data such as the homology of rRNA sequences,[6] conserved orthologous genes,[7] sequences of tRNA synthetases,[8] and tRNA-dependent amidotransferases.[9] The result shows that the set of base modifications is not universal, but rather a highly species-specific code under evolutionary selection to appropriately match base triplets with the corresponding cognate amino acids.

For the study, we applied our recently developed LC-MS-based method for the quantification of modified nucleosides by using isotopically labeled standards.[10] For the parallel quantification we synthesized 18 tRNA modifications in both their natural and isotopically labeled forms (Scheme 1). A number of these nucleosides are present in the 3′-position to the anticodon in position 37, while others are distributed through the tRNA structure. The modifications studied are involved in a range of biological processes such as structural stabilization, codon binding, and translation initiation.[2a,3,10b,11]



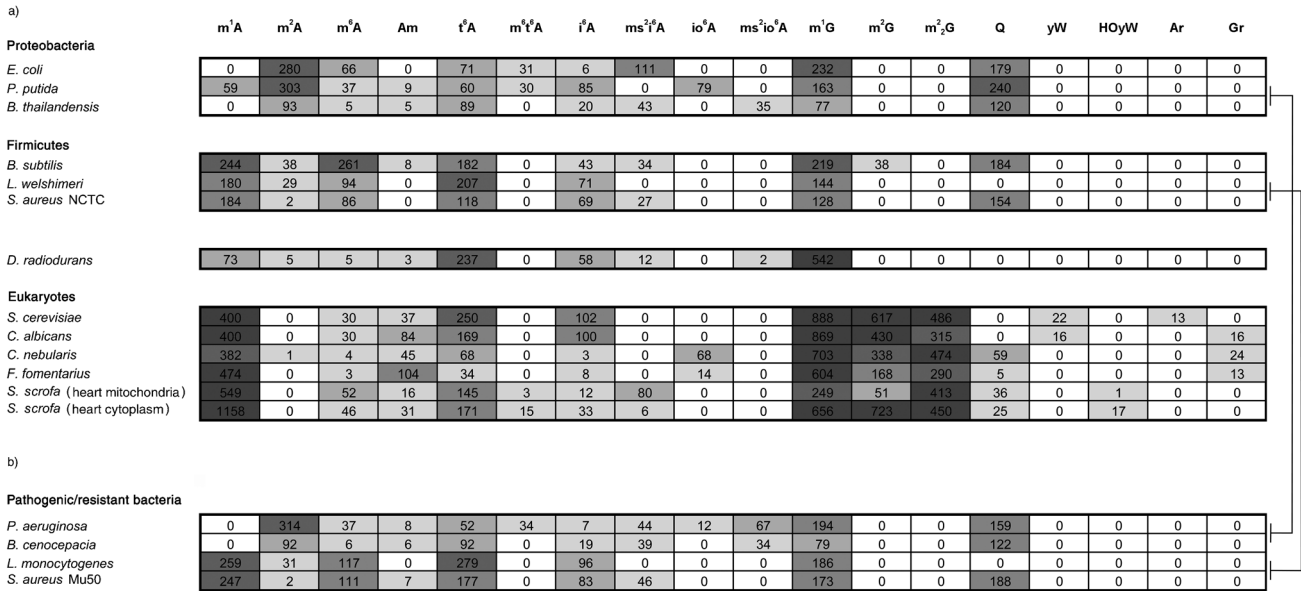
Scheme 1. Modified nucleosides synthesized in both natural and isotopically labeled forms for our parallel quantification study.

With these tRNA nucleosides in hand, we analyzed the tRNA modification pattern of 16 species, so as to cover several branches of the phylogenetic tree. These species include five eukaryotes as well as five Gram-negative proteobacteria and five Gram-positive bacteria of the firmicutes. In addition we studied the bacterium *Deinococcus radiodurans*, which has a somewhat ambiguous classification. *D. radiodurans* is typically identified as Gram-positive, but possesses additional cell walls similar to Gram-negative species.[12] Bulk tRNA was extracted from freshly collected

[*] Dr. D. Globisch, Dr. D. Pearson, Dipl.-Chem. A. Hienzsch, Dr. T. Brückl, Dipl.-Chem. M. Wagner, I. Thoma, Dipl.-Chem. P. Thumbs, Dipl.-Chem. V. Reiter, A. C. Kneuttinger, Dr. M. Müller, Prof. S. A. Sieber, Prof. T. Carell
Center for Integrated Protein Science, Department of Chemistry, Ludwig Maximilian University Munich
Butenandtstrasse 5-13, 81377 Munich (Germany)
E-mail: thomas.carell@cup.uni-muenchen.de
Homepage: http://www.carellgroup.de

[**] We thank the Excellence Cluster CiPS[M], the Deutsche Forschungsgemeinschaft (CA-275 8/4), and the SFB749 for financial support. We thank Prof. Wolfgang Steglich for providing the fungal species and their taxonomy. We thank Kerstin Kurz and Dilek Özden for technical assistance.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201103229.



a)

	m^1A	m^2A	m^6A	Am	t^6A	m^6t^6A	i^6A	ms^2i^6A	io^6A	ms^2io^6A	m^1G	m^2G	m^2_2G	Q	yW	HOyW	Ar	Gr
Proteobacteria																		
E. coli	0	280	66	0	71	31	6	111	0	0	232	0	0	179	0	0	0	0
P. putida	59	303	37	9	60	30	85	0	79	0	163	0	0	240	0	0	0	0
B. thailandensis	0	93	5	5	89	0	20	43	0	35	77	0	0	120	0	0	0	0
Firmicutes																		
B. subtilis	244	38	261	8	182	0	43	34	0	0	219	38	0	184	0	0	0	0
L. welshimeri	180	29	94	0	207	0	71	0	0	0	144	0	0	0	0	0	0	0
S. aureus NCTC	184	2	86	0	118	0	69	27	0	0	128	0	0	154	0	0	0	0
D. radiodurans	73	5	5	3	237	0	58	12	0	2	542	0	0	0	0	0	0	0
Eukaryotes																		
S. cerevisiae	400	0	30	37	250	0	102	0	0	0	888	617	486	0	22	0	13	0
C. albicans	400	0	30	84	169	0	100	0	0	0	869	430	315	0	16	0	0	16
C. nebularis	382	1	4	45	68	0	3	0	68	0	703	338	474	59	0	0	0	24
F. fomentarius	474	0	3	104	34	0	8	0	14	0	604	168	290	5	0	0	0	13
S. scrofa (heart mitochondria)	549	0	52	16	145	3	12	80	0	0	249	51	413	36	0	1	0	0
S. scrofa (heart cytoplasm)	1158	0	46	31	171	15	33	6	0	0	656	723	450	25	0	17	0	0

b)

	m^1A	m^2A	m^6A	Am	t^6A	m^6t^6A	i^6A	ms^2i^6A	io^6A	ms^2io^6A	m^1G	m^2G	m^2_2G	Q	yW	HOyW	Ar	Gr
Pathogenic/resistant bacteria																		
P. aeruginosa	0	314	37	8	52	34	7	44	12	67	194	0	0	159	0	0	0	0
B. cenocepacia	0	92	6	6	92	0	19	39	0	34	79	0	0	122	0	0	0	0
L. monocytogenes	259	31	117	0	279	0	96	0	0	0	186	0	0	0	0	0	0	0
S. aureus Mu50	247	2	111	7	177	0	83	46	0	0	173	0	0	188	0	0	0	0

Figure 1. tRNA modification pattern of all the samples investigated. The level of tRNA modification is presented as the number of modified nucleosides per 1000 tRNAs, with an average standard deviation of 7%. Darker shading represents a higher modification level. a) Eukaryotic and nonpathogenic/resistant bacterial species grouped on the basis of taxonomy. b) Pathogenic/resistant bacteria. The lines on the right side link these bacterial strains with the corresponding nonpathogenic/resistant strains.

tissues as well as from bacteria and yeast samples grown in complex media under optimum conditions. This was followed by digestion and subsequent analysis of the resulting nucleoside mixture by using our established quantitative LC-MS method.[10a]

The results for all the species investigated are depicted in Figure 1. The first key result from this data is that the modification levels vary dramatically between species. For example, $\mathbf{m^2A}$ varies from 1 (modification per 1000 tRNA molecules) in the fungus *Clitocybe nebularis* to 314 in the Gram-negative bacterium *Pseudomonas aeruginosa*, and $\mathbf{m^2G}$ varies from 38 in the Gram-positive bacterium *Bacillus subtilis* to 723 in the mammal *Sus scrofa*. High levels of certain modifications characterize particular groups of organisms. For example, $\mathbf{m^2A}$ is generally high in Gram-negative bacteria, while $\mathbf{m^1A}$, $\mathbf{m^1G}$, $\mathbf{m^2G}$, and $\mathbf{m^2_2G}$ are all abundant in eukaryotes. These initial observations show evolutionary relationships related not only to the presence, but also to the abundance of each tRNA modification.

As four N^6-isopentenyladenosine derivatives ($\mathbf{i^6A}$, $\mathbf{ms^2i^6A}$, $\mathbf{io^6A}$, and $\mathbf{ms^2io^6A}$) were quantified in parallel, the relationship between these modifications can be analyzed in detail. While $\mathbf{i^6A}$ itself is present in all the species investigated, the more complex $\mathbf{i^6A}$ derivatives are not always observed. The $\mathbf{ms^2i^6A}$ modification is present in all bacteria except *Listeria* and *P. putida*, and is found in eukaryotes in *S. scrofa* but not in any of the fungal species investigated. According to the literature, the hydroxy derivatives $\mathbf{io^6A}$ and $\mathbf{ms^2io^6A}$ are predominantly present in γ-proteobacteria such as *Pseudomonas*.[13] Surprisingly, we detected large quantities of these compounds also in the β-proteobacteria *Burkholderia*. Additionally, traces of $\mathbf{ms^2io^6A}$ were found in *D. radiodurans*, which is surprising as the sequence of the modifying enzyme MiaE was not found in genomic analysis.[14] This result indicates that an enzyme with a significantly different sequence might be responsible for this modification. Both complex fungi *C. nebularis* and *Fomes fomentarius* predominantly utilize $\mathbf{io^6A}$, which differentiates them from the other eukaryotes analyzed.

Our data show that the $\mathbf{m^1G}$ and $\mathbf{t^6A}$ modifications are present at high levels in all the species studied. In particular, these modifications are abundant in the unusual bacterium *D. radiodurans*, where they represent the major proportion of all the modifications quantified in this species. These two modified nucleosides are the only tRNA nucleosides present in all three kingdoms of life as well as in mitochondria, and chloroplasts.[15] We, therefore, conclude from our study that they belong to the oldest RNA modifications, which likely contributed to the very early development of life.[16] The observation that these nucleosides are highly abundant in *D. radiodurans* suggests an ancient evolutionary divergence of this bacterial species, in agreement with evolutionary conclusions based on other data.[17]

We applied a hierarchical clustering algorithm to the data using the programs *Cluster* and *Treeview* to statistically analyze the differences in modification levels between species.[18] These calculations group together species with similar modification levels. We used an Euclidean distance algorithm without normalization, and since this measures absolute differences in abundance, it should, therefore, represent overall tRNA functionality and assess fine variations in modification levels between species. Fascinatingly, the clustering result for our quantitative modification data (Figure 2a) produces a highly accurate phylogenetic tree. Bootstrap support is very strong for the bacterial clustering, and somewhat weaker for the clustering of the eukaryote species

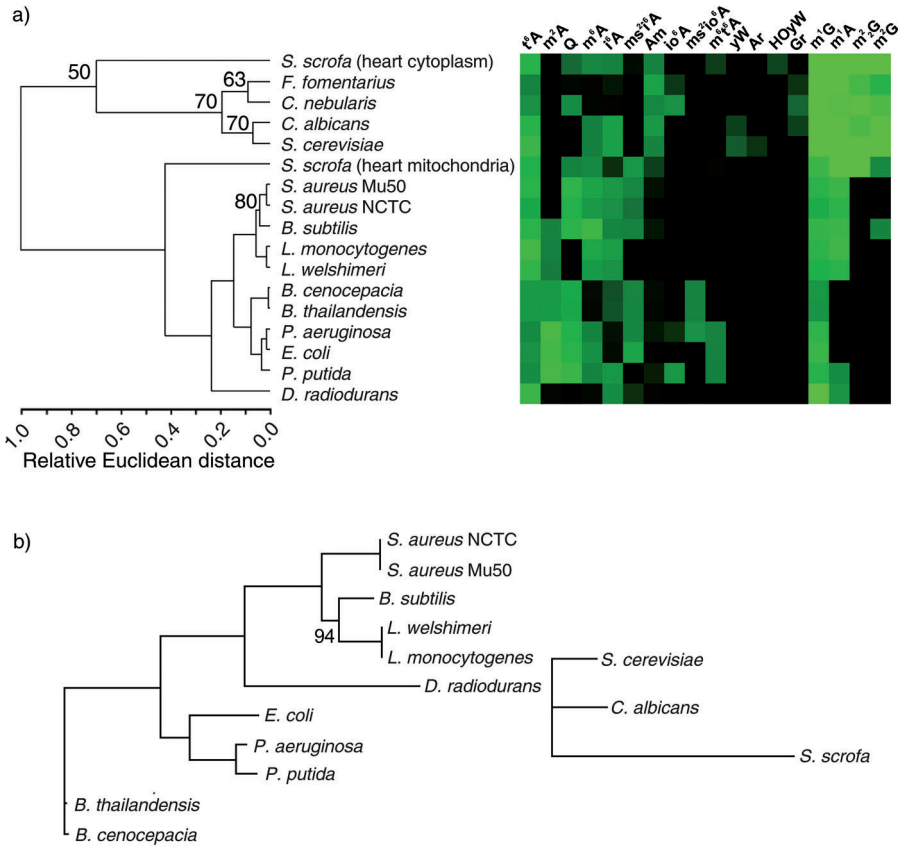


Figure 2. Cluster analysis of species variation. a) Species clustering based on our quantitative tRNA modification data. Branch point labels represent smooth bootstrap values for the least strongly supported clusters. All unlabeled branch points have very strong bootstrap support (>95). The horizontal scale represents the relative Euclidean distance between clusters at each branching point. b) Phylogenetic trees based on analysis of a set of orthologous genes. Branch point labels represent resampling bootstrap values for less strongly supported clusters. All unlabeled branch points have very strong bootstrap support (100). The length of the branch represents relative frequency of amino acid substitution between species/branch points.

analyzed here. To compare with established methods, we also calculated a conventional phylogenetic tree based on genetic variation (Figure 2b). This tree was constructed on the basis of a reported method[7a] using genome data for all possible species present in our study.

Surprisingly, the two clustering experiments produce very similar grouping patterns, thus confirming that modified nucleoside levels are closely linked to genetic variation of species. Eukaryotes are clearly separated from bacteria, and there is also a clear distinction between Gram-negative and Gram-positive bacteria. At a more detailed level, distinct differentiation of the firmicute Gram-positive bacteria (*Bacillus*, *Listeria*, and *Staphylococcus*) demonstrates the resolution and accuracy of our analysis. Similarly, the eukaryotes cluster in excellent correlation with the phylogenetic tree, with a distinction between mammals and fungi, and even a separation between complex fungi and yeasts. The Gram-negative bacteria clustering also shows similarity to the phylogenetic groupings, with the exception of *Pseudomonas aeruginosa*, which clusters with *E. coli* rather than the more closely related *P. putida*. However, the closely related bacterial classes of γ-proteobacteria (*Escherichia* and *Pseudomonas*) and β-proteobacteria (*Burkholderia*) are distinguishable. Interestingly, *D. radiodurans* does not fit into either of the two bacterial groups, thus highlighting its ambiguous character.

Finally, modification levels in mitochondrial tRNA from *S. scrofa* heart tissue were found to cluster with the bacterial data, while the cytoplasmic modification levels clustered, as expected, with the other eukaryotes. These results show that mitochondrial tRNA retains prokaryotic character in terms of tRNA modification levels, in line with qualitative analyses of tRNA sequence and modification.[19] Our result additionally suggests that mitochondria may retain prokaryotic systems for the regulation of modification levels.

The high resolution of the analysis indicated that quantification of tRNA modifications could be applicable for use in reliably distinguishing pathogenic bacteria from related nonpathogenic species. To assess the potential of this approach, we compared the three pairs of pathogenic and nonpathogenic bacteria from the genera *Pseudomonas*, *Burkholderia*, *Listeria*, and one pair of methicillin-resistant and non-resistant *S. aureus* strains (Figure 1). These bacteria represent a selection of the most dangerous clinical pathogens, which are responsible for many deaths.

Indeed, pathogenic and nonpathogenic bacteria are clearly distinguishable. The two *Pseudomonas* species even contain different sets of modified tRNA nucleosides. While the pathogenic species *P. aeruginosa* contains all four **i^6A** derivatives (**i^6A**, **ms^2i^6A**, **io^6A**, and **ms^2io^6A**), the nonpathogenic species *P. putida* contains only **i^6A** and **io^6A**, which lack the methylthio group. Additionally, nucleoside **m^1A** is only present in the nonpathogenic species. In the case of *Listeria* and *Staphylococcus*, the modification content was considerably higher for the pathogenic and the resistant species, respectively. This hints at an altered translational process with an increased need for modifications. Only the *Burkholderia* species were not significantly distinguishable by their modification pattern. However, in general our method provides a novel possibility for differentiation between pathogenic and nonpathogenic bacteria.

In summary, we have measured 18 tRNA modifications in 16 species quantitatively. The presented results offer a deeper insight into the evolution of tRNA modifications, and shows that they characterize species at a very fine level and are linked to phylogenetic variation. Additionally, the data can be

used to differentiate between species, and even pathogenic from nonpathogenic bacteria.

Received: May 11, 2011
Published online: August 31, 2011

Keywords: bioanalytical methods · isotopic labeling · molecular evolution · nucleosides · tRNA

[1] a) H. Grosjean, R. Benne, *Modification and Editing of RNA*, *Vol. 1*, ASM, **1998**; b) F. Jühling, M. Mörl, K. Hartmann, M. Sprinzl, F. Stadler, J. Pütz, *Nucleic Acids Res.* **2009**, *37*, D159–D162.

[2] a) P. F. Agris, F. A. Vendeix, W. D. Graham, *J. Mol. Biol.* **2007**, *366*, 1–13; b) A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* **2007**, *3*, 29–35; c) J. F. Atkins, G. R. Björk, *Microbiol. Mol. Biol. Rev.* **2009**, *73*, 178–210.

[3] Y. Motorin, M. Helm, *Biochemistry* **2010**, *49*, 4934–4944.

[4] L. A. Sylvers, K. C. Rogers, M. Shimizu, E. Ohtsuka, D. Söll, *Biochemistry* **1993**, *32*, 3836–3841.

[5] C. T. Chan, M. Dyavaiah, M. S. DeMott, K. Taghizadeh, P. C. Dedon, T. J. Begley, *PLoS Genet.* **2010**, *6*, e1001247.

[6] a) G. E. Fox, E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, C. R. Woese, *Science* **1980**, *209*, 457–463; b) H. J. Hori, *Mol. Evol.* **1975**, *7*, 75–86.

[7] a) F. D. Ciccarelli, T. Doerks, C. von Mering, C. J. Creevey, B. Snel, P. Bork, *Science* **2006**, *311*, 1283–1287; b) M. Wu, J. A. Eisen, *Genome Biol.* **2008**, *9*, R151–R151.

[8] K. Shiba, H. Motegi, P. Schimmel, *Trends Biochem. Sci.* **1997**, *22*, 453–457.

[9] K. Sheppard, D. Söll, *J. Mol. Biol.* **2008**, *377*, 831–844.

[10] a) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem.* **2009**, *121*, 8074–8077; *Angew. Chem. Int. Ed.* **2009**, *48*, 7932–7934; b) D. Pearson, A. Hienzsch, M. Wagner, D. Globisch, V. Reiter, D. Özden, T. Carell, *Chem. Commun.* **2011**, *47*, 5196–5198; c) M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem.* **2010**, *122*, 5503–5505; *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377.

[11] C. Förster, K. Chakraburtty, M. Sprinzl, *Nucleic Acids Res.* **1993**, *21*, 5679–5683.

[12] J. R. Battista, *Annu. Rev. Microbiol.* **1997**, *51*, 203–224.

[13] J. J. Janzer, J. P. Raney, B. D. McLennan, *Nucleic Acids Res.* **1982**, *10*, 5663–5672.

[14] K. H. Kaminska, U. Baraniak, M. Boniecki, K. Nowaczyk, A. Czerwoniec, J. M. Bujnicki, *Proteins* **2008**, *70*, 1–18.

[15] A. Morin, S. Auxilien, B. Senger, R. Tewari, H. Grosjean, *RNA* **1998**, *4*, 24–37.

[16] G. R. Björk, K. Jacobsson, K. Nilsson, M. J. O. Johansson, A. S. Byström, O. P. Persson, *EMBO J.* **2001**, *20*, 231–239.

[17] a) R. S. Gupta, *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1435–1491; b) K. S. Makarova, L. Aravind, Y. I. Wolf, R. L. Tatusov, K. W. Minton, E. V. Koonin, M. J. Daly, *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 44–79.

[18] a) M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14863–14868; b) N. Jessani, Y. Liu, M. Humphrey, B. F. Cravatt, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10335–10340.

[19] a) H. Grosjean, V. de Crécy-Lagard, C. Marck, *FEBS Lett.* **2010**, *584*, 252–264; b) K. Watanabe, *Proc. Jpn. Acad. Ser. B* **2010**, *86*, 11–39.

