

Gensynthese auf Mikrochips**

Joachim W. Engels*

Stichwörter:

Chemische Biologie · Hybridisierungen · Mikrochips ·
Oligonukleotide · Polymerasenkettenreaktion

Mithilfe der Gensynthese wird man in Zukunft ganze Genome synthetisieren und damit auch grundlegende biologische Fragen experimentell beantworten können, z.B. wie chemische Reaktionen in der Zelle ablaufen. Der erste Schritt, die Sequenzierung von Genomen und insbesondere des menschlichen Genoms, hat eine gewaltige Fülle an Daten erbracht. Wir wissen heute, dass der Mensch etwa 25 000 Gene besitzt. Gene sind chemisch gesehen nichts anderes als eine lange Folge definierter DNA-Sequenzen. In den letzten 40 Jahren wurden viele Methoden entwickelt, um DNA chemisch zu synthetisieren.^[1-3] Diese Verfahren beruhen auf der schrittweisen Kupplung einzelner Nucleotidbausteine durch chemische oder enzymatische Verknüpfung der Phosphorsäureester zu Oligonucleotiden, die aufgrund ihrer Basenkomplementarität zu doppelsträngiger DNA zusammengeführt werden können. Die Hybridisierung zu DNA-Duplexen wird dabei durch Wasserstoffbrücken vermittelt. Auf diese Weise gelang es bereits 1970 Khorana und Mitarbeitern, ein Gen der Hefe tRNA^{Ala} zu synthetisieren.^[3] Seit dieser Zeit haben sich die Synthesemethoden verfeinert, besonders durch Anwendung der Polymerasenkettenreaktion (PCR),^[4-6] und die Ziele wurden anspruchsvoller. Folgende Gene wurden synthetisiert: 1985 das für α -Interfe-

ron,^[7] 1987 das für den Interleukin-2-Rezeptor α^[8] und 1990 das für ein Plasmid.^[9] 2002 folgte die Synthese einer Poliovirus-DNA^[10] mit etwa 7500 Basenpaaren, die biologisch aktiv, d. h.

infektiös war. 2003 gelang die Synthese der DNA des Bakteriophagen Φ X174 mit 5386 Basenpaaren,^[11a] des Moleküls also, das Sanger als erstes mit der Di-desoxymethode sequenziert hatte.^[11b]

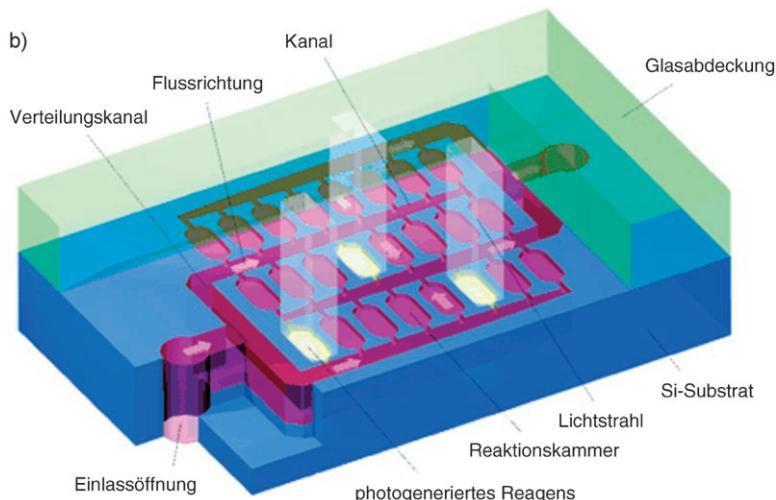
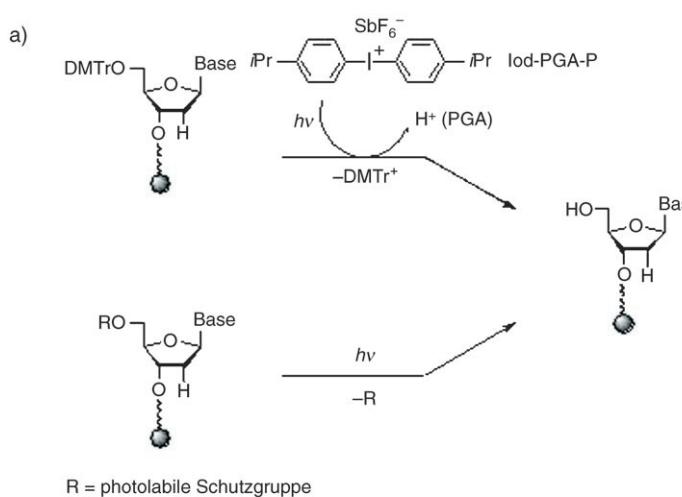


Abbildung 1. Phosphoramidit-Oligonucleotidsynthese auf einem Chip. a) Erster Schritt der Oligonucleotidsynthese, die 5'-Entschützung: Synthese mit photogenerierter Säure (PGA) oder photolabiler Schutzgruppe; b) Aufbau des Chips: PicoArray-Reaktor für parallele Oligonucleotidsynthesen (University of Michigan). 128 × 31 (3968) individuelle Reaktionskammern mit einem internen Volumen von 270 pL (Gesamtvolumen 10 µL). Die digitale Lichtprojektion zeigt die PGA-kontrollierten Reaktionen. Aus Lit. [17] mit freundlicher Genehmigung von Oxford University Press.

[*] Prof. Dr. J. W. Engels

Prof. Dr. J. W. Engels
Institut für Organische Chemie
und Chemische Biologie
Johann Wolfgang Goethe-Universität
Marie-Curie-Straße 11
60439 Frankfurt am Main (Deutschland)
Fax: (+49) 69-798-29148
E-mail: joachim.engels@chemie.
uni-frankfurt.de

[**] Ich danke A. Klöpffer, K. Strube und C. Engels für technische Hilfe.

Als Längenrekord folgte dann 2004 die Synthese des 32 000 Basenpaare langen Genclusters der Polyketid-Synthase.^[12] Auch hier war das Gen nach Transformation in *E. coli* biologisch aktiv und ergab 6-Desoxyerythronolid.

Durch kombinatorische oder parallele Synthese ist eine Vielfalt chemischer Strukturen zugänglich, deren Funktionen sich in Testsystemen analysieren lassen. Gene im Bereich von Megabasen, wie die menschliche DNA, können heute chemisch oder chemoenzymatisch synthetisiert werden. 20- bis 100mere Oligonucleotide sind für etwa 0.10 Euro pro Nucleotid kommerziell erhältlich. Sie haben Fehlerraten von etwa 1:100 bis 1:400, die im Wesentlichen aus Problemen bei der chemischen Synthese resultieren. Ursächlich sind vor allem unvollständige Kupplungen und Entschüttungen sowie Nebenreaktionen der verwendeten Reagentien, z.B. die Michael-Addition von Acrylnitril an Thymin. Hochdurchsatz-Synthesen sind aus Gründen der Kosten und der Fehlerraten Grenzen gesetzt, da Klonierung und Sequenzierung die Kosten mindestens verzehnfachen.

Hier haben neuerdings Parallelsynthesen auf Mikrochips Wege aufgezeigt, um Kosten zu reduzieren. Als problematisch erweist sich allerdings die Menge der synthetisierten Oligonucleotide

im Atto- und Femtomolbereich, denn diese 10^5 - 10^9 Moleküle müssen durch PCR auf 10^9 - 10^{12} Moleküle vermehrt werden.^[13] Wie Gao, Church und Mitarbeiter nun demonstrierten,^[14] können ca. 4000 Oligonucleotide auf z.B. mit Licht programmierten Chips zeitgleich synthetisiert werden (Abbildung 1).^[15-17] Mit photolabilen Schutzgruppen oder photogenerierten Protonen und der Dimethoxytritylschutzgruppe (DMTr) wird auf dem Chip direkt das Oligonucleotid nach der Phosphoramiditmethode von Caruthers synthetisiert.^[1] Da wie oben erwähnt Fehler bei der Synthese auftreten, wurde von den Autoren eine stringente Hybridisierung mithilfe eines so genannten Qualitäts-Chips mit den komplementären Oligonucleotiden als Qualitätskontrollen ausgeführt. Die fehlerhaften Oligonucleotide, z.B. Punktmutationen, wurden durch Affinitätschromatographie an immobilisierten, kurzen komplementären Oligonucleotiden gereinigt. Für eine simultane Trennung ist es hierbei wichtig, die Schmelzpunkte (T_m -Werte) der Doppelstränge durch Variation der Oligonucleotidlängen isotherm zu wählen.

Ein wichtiger Punkt beim Gen-Design ist die Wahl der Restriktionsenzymeschnittstellen. Es wurden Restriktionsenzyme des Typs IIS gewählt, die außerhalb ihrer Erkennungssequenz

schnieden und so eine Freisetzung beliebiger interner Sequenzen – hier der Gensequenzen – ermöglichen (Abbildung 2). Die Autoren verwendeten die Enzyme BsaI mit der Erkennungssequenz 5'-GGTCTC(1/5), die einen 4-Basen-5'-Überhang bildet, und BseRI mit der Erkennungssequenz 5'-GAG-GAG(10/8) mit einem 2-Basen-3'-Überhang beliebiger Sequenz. Zusätzlich zu den Gensequenzen werden noch Selektionsoligonucleotide benötigt, die durch Affinitätsmarker wie Biotin derivatisiert und durch Streptavidinbindung immobilisiert sind (Abbildung 3). Nur die unmodifizierten, fehlerfreien Oligonucleotide, in diesem Fall 50mere, werden selektiert. In zwei Schritten werden die exakt passenden Oligonucleotide mithilfe von je 25meren Oligonucleotiden R und L („Einfangsonden“) herausgefiltert. Anschließend können die Gene parallel in Mikrotiterplatten durch PAM-Reaktionen („polymerase assembly multiplexing“) aufgebaut werden. Hierbei wird die durch Wasserstoffbrücken zusammengehaltene dsDNA ohne Primer mithilfe von Polymerase zum Gen zusammengeführt. Die Qualität der zur Gensynthese eingesetzten Oligonucleotide wurde in drei verschiedenen Formaten getestet: 1) ungereinigt (wie synthetisiert), 2) nach Reinigung mit Gelektrophorese und 3) nach Rei-

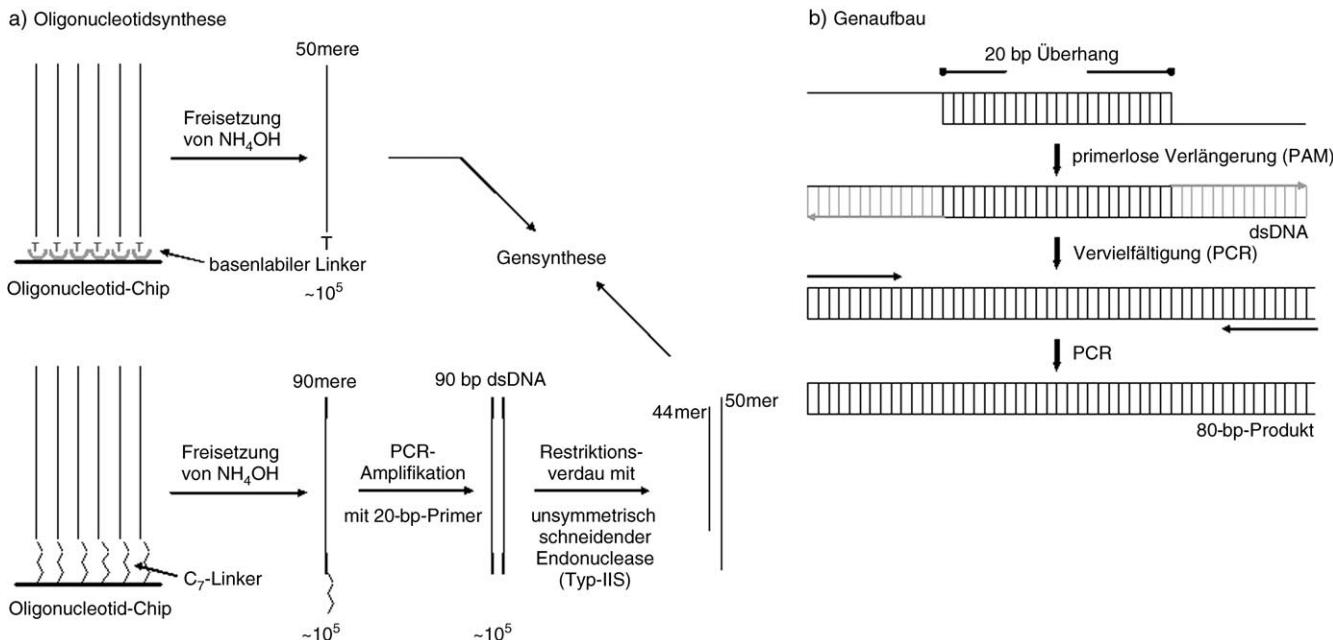


Abbildung 2. Oligonucleotid- und Gensynthese. a) Oligonucleotidsynthese (50mere) auf einem Chip mit und ohne PCR-Amplifikation und Verdau der Primersequenz; b) Polymeraseunterstützte Gensynthese: Auffüllreaktion (primerlose Verlängerung) und PCR der beiden 50meren Oligonucleotide.

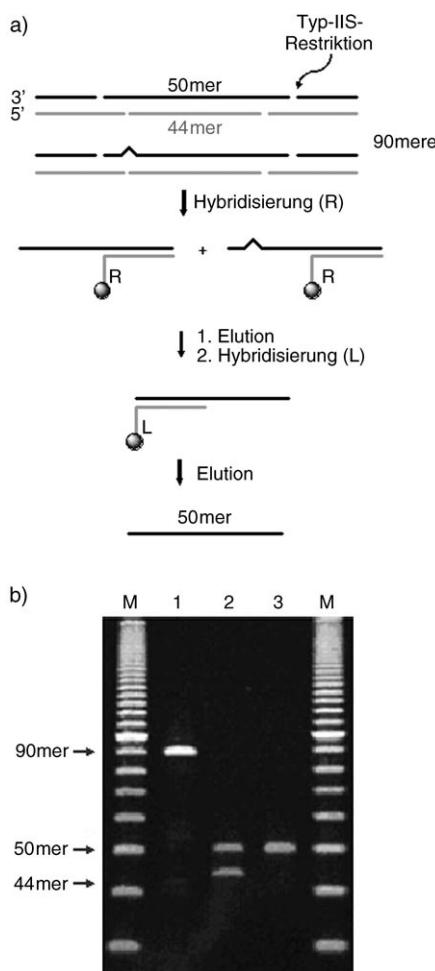


Abbildung 3. Selektion synthetischer Oligonukleotide durch Hybridisierung: a) Schema der Hybridisierung der durch PCR amplifizierten 90mere durch immobilisierte Oligonukleotide L und R; b) Polyacrylamid(PAA)-Gel der mit BsAl und BseRI geschnittenen Oligonukleotide (50- und 44mer in Spur 2) und selektierten 50mere (Spur 3). Aus Lit. [14] mit freundlicher Genehmigung von Nature Publishing.

nigung durch affinitätschromatographische Hybridisierung.^[18] Letztere Methode ergab in diesem Experiment die geringste Fehlerrate mit 1:1394 bp; das ist etwa um einen Faktor 10 besser als die konventionelle Methode der chemischen Oligonukleotidsynthese mit nachfolgender enzymatischer Ligation.

Als Beispiel für die Anwendung dieser Methode in der synthetischen Biologie synthetisierten Gao, Church und Mitarbeiter^[14] codonveränderte Gene für 21 Proteine der kleinen Ribosomenuntereinheit 30S aus *E. coli*.^[19] Es ist bekannt, dass die Expressionsraten die-

ser 21 Proteine in vitro gering sind und dass die Translationseffizienz durch geeignete Codonwahl gesteigert werden kann. Als Ursache für die geringe Syntheseeffizienz werden Sekundärstrukturen der mRNA diskutiert. Der Ansatz bestand nun darin, die Bildung dieser Sekundärstrukturen durch Verringern des G/C-Gehaltes einzuschränken. Auch hier wurden 50mere Oligonukleotide gewählt, die in 70mere inkorporiert wurden, um eine Amplifikation durch PCR zu ermöglichen. Mit Computerunterstützung wurde eine PAM-Strategie gefunden, die sicherstellte, dass jedes Oligonukleotid seinen richtigen Partner findet. Als Ergebnis konnten alle 21 Gene, die abgeleitete RNA und die Proteine erfolgreich detektiert werden, wobei ein Gesamtoperon von 14.6 kb aller 21 Gene resultierte (Abbildung 4). Es wurde eine Fehlerrate bei den Oligonukleotiden von nur 1:7300 Basenpaaren errechnet. Somit ist klar, dass die Fehlerrate der synthetisierten Oligonukleotide bei den kurzen Sequenzen entscheidend ist – und nicht etwa die Genauigkeit der Polymerase, die bei etwa 1:100000 liegt.

Dieses Experiment verdeutlicht die Möglichkeiten der On-Chip-Synthese,

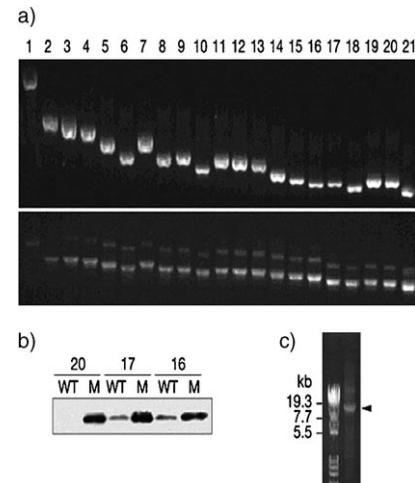


Abbildung 4. Synthetische Gene und ihre Transkriptionsprodukte (mRNAs) und Translationsprodukte (Proteine) auf Polyacrylamidgelen. a) Synthetische Gene zu 21 ribosomalen Proteinen (oberes Gel) und ihre in vitro transkribierte RNA (unteres Gel); b) Western-Blot ausgewählter translatierter Gene der Proteine 16, 17 und 20 (WT Wildtyp, M synthetisierte Variante); c) zusammengesetztes 14.6-kb-Operon aller 21 ribosomalen Proteine. Aus Lit. [14] mit freundlicher Genehmigung von Nature Publishing.

und wenn man die heutige Kapazität der Mikrochips berücksichtigt, ist das anfängliche Ziel von Gensynthesen im großen Maßstab näher gerückt. Eine Hochrechnung ergibt, dass bei einer Oligonukleotiddichte von 100000 pro Chip die Kosten der Gensynthese von derzeit 10 bp pro Euro auf 20 kbp pro Euro reduziert werden könnten. Mikrochips der Firma NimbleGen (www.nimblegen.com) mit Dichten von 95000–382000 Oligonukleotiden pro Chip sollten demnach 2–18 Mbp DNA erreichen.

Zur weiteren Verbesserung der Fehlerrate kommt eine Korrektur der synthetisierten Oligonukleotide oder der Genfragmente durch Proteine des DNA-Reparatursystems wie MutS in Betracht. In einem Experiment mit fixiertem MutS aus *Thermus aquaticus* entfernten Jacobson et al. aus einem synthetisierten Gen die Mismatch-Produkte und erreichten so eine Fehlerrate von 1:10000.^[20] Eine weitere vielversprechende Methode zur Herstellung möglichst fehlerfreier DNA ist die Hochdurchsatz-Sequenzierung, mit der sich tausende von Sequenzen simultan lesen lassen.^[21]

Fazit ist: Die Komponenten der hier vorgestellten Gensynthesemethode waren schon weitestgehend bekannt, sie wurden aber neu kombiniert, um fehlerfreie Gene schneller und billiger zu synthetisieren. Die Methode umfasst folgende Einzelschritte (ungefährer Zeitaufwand in Klammern):^[14]

- Modellierung und Design der Oligonukleotide für die Gensynthese (2 h).
- Oligonukleotidsynthese in Mikrofluidreaktoren (8 h).
- Amplifikation der von den Chips freigesetzten Oligonukleotide (14 h).
- Verdau der amplifizierten Oligonukleotide durch Typ-IIS-Restriktionsenzyme auf die gewünschte Größe (50mere) (v.s.).
- Selektieren der gewünschten Sequenzen durch Hybridisierung (30 h).
- Zusammenfügen der Gene durch Polymerase-Auffüllreaktion und Amplifikation durch PCR (4 h).
- In-vitro-Transkription und -Translation und Nachweis der mRNAs und Proteine.

Mögliche Anwendungen sind die Synthese von Genclustern für Polyketide oder die Resynthese einer Bakterienzelle, z.B. Mycoplasma. Die Gensynthese bietet so einen vielversprechenden Zugang zu neuen Sekundärmetaboliten wie Polyketiden oder Peptidantibiotika, um nichtnatürliche Naturstoffe zu erzeugen. Sie ist eine zentrale Methode für die funktionelle Genomik, und sie kann zur vollständigen Zuordnung aller chemischen Reaktionen in der Zelle führen. Das letztendliche Ziel der biologischen Synthese ist es, aus der Vielfalt von Komponenten (Genpool) neue Arrangements zu finden und so zu versuchen, neue Lösungen für neue Produkte zu erhalten.

- [1] a) S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, 22, 1859–1862; b) M. H. Caruthers, *Science* **1985**, 230, 281–285.
- [2] J. W. Engels, E. Uhlmann, *Angew. Chem.* **1989**, 101, 733–752; *Angew. Chem. Int. Ed. Engl.* **1989**, 28, 716–734.
- [3] K. L. Agarwal, H. Büchi, M. H. Caruthers, N. Gupta, H. G. Khorana, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. Rajbhandary, J. H. Van de Sande, V. Sgaramella, H. Weber, T. Yamada, *Nature* **1970**, 227, 27–34.

- [4] P. J. Dillon, C. A. Rosen, *BioTechniques* **1990**, 9, 298–300.
- [5] W. P. C. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, H. L. Heyneker, *Gene* **1995**, 164, 49–53.
- [6] S. R. Casimiro, P. E. Wright, H. J. Dyson, *Structure* **1997**, 5, 1407–1412.
- [7] M. Edge, A. R. Greene, G. R. Heathcliffe, P. A. Meacock, W. Schuch, D. B. Scanlon, T. C. Atkinson, C. R. Newton, A. Markham, *Nature* **1981**, 292, 756–762.
- [8] J. Engels, W. Becker, T. Beckers, D. Hüskens, W. Wetekam, *Pure Appl. Chem.* **1987**, 59, 437–444.
- [9] W. Mandecki, M. A. Hayden, M. A. Shallcross, E. Stotland, *Gene* **1990**, 94, 103–107.
- [10] J. Cello, A. V. Paul, E. Wimmer, *Science* **2002**, 297, 1016–1018.
- [11] a) H. O. Smith, C. A. Hutchison, C. Pfannkoch, J. C. Venter, *Proc. Natl. Acad. Sci. USA* **2003**, 100, 15440–15445; b) F. Sanger, A. R. Coulson, T. Friedmann, G. M. Air, B. G. Barrell, N. L. Brown, J. C. Fiddes, C. A. Hutchinson III, P. M. Slocombe, M. Smith, *J. Mol. Biol.* **1978**, 125, 225–246.
- [12] S. J. Kodumal, K. G. Patel, R. Reid, H. G. Menzella, M. Welch, D. V. Santi, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 15573–15578.
- [13] K. E. Richmond, M.-H. Li, M. J. Rodesch, M. Patel, A. M. Lowe, C. Kim, L. L. Chu, N. Venkataraman, S. F. Flanagan, J. Kayser, P. J. Belshaw, M. R. Sussman, F. Cerrina, *Nucleic Acids Res.* **2004**, 32, 5011–5018.
- [14] J. Tian, H. Gong, N. Sheng, X. Zhou, E. Gulari, X. Gao, G. Church, *Nature* **2004**, 432, 1050–1053.
- [15] X. Gao, P. Yu, E. LeProust, L. Sonigo, J. P. Pellois, H. Zhang, *J. Am. Chem. Soc.* **1998**, 120, 12698–12699.
- [16] X. Gao, E. LeProust, H. Zhang, O. Srivannavat, E. Gulari, P. Yu, C. Nishiguchi, Q. Xiang, X. Zhou, *Nucleic Acids Res.* **2001**, 29, 4744–4750.
- [17] X. Zhou, S. Cai, A. Hong, Q. You, P. Yu, N. Sheng, O. Srivannavat, S. Muranjan, J. M. Rouillard, Y. Xia, X. Zhang, Q. Xiang, R. Ganesh, Q. Zhu, A. Matejko, E. Gulari, X. Gao, *Nucleic Acids Res.* **2004**, 32, 5409–5417.
- [18] R. G. Eason, N. Poumand, W. Tongprasit, Z. S. Herman, K. Anthony, O. Jeelowo, R. W. Davis, V. Stolc, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 11046–11051.
- [19] G. M. Culver, H. F. Noller, *RNA* **1999**, 5, 832–843.
- [20] P. A. Carr, J. S. Park, Y.-J. Lee, T. Yu, S. Zhang, J. M. Jacobson, *Nucleic Acids Res.* **2004**, 32, e162.
- [21] J. Shendure, R. D. Mitra, C. Varma, G. M. Church, *Nat. Rev. Genet.* **2004**, 5, 335–343.