

# Fortschritte in der Entwicklung von Gensynthese-Strategien

Eine neuartige Methode



► Prof. Dr. Markus Fischer  
Institute of Biochemistry  
and Food Chemistry  
Universität Hamburg

Die *de novo* Synthese von Genen findet in der modernen Biotechnologie vielfältige Anwendungen. Synthetische Gene werden beispielsweise auf dem Gebiet der heterologen Expression von Proteinen oder im Bereich des Protein-Engineerings eingesetzt.

## Synthese von Genen mit einzelsträngigen Oligonukleotiden

In den letzten 30 Jahren wurden eine Reihe von Methoden publiziert, die sich mit Genynthese im weitesten Sinne beschäftigen. Üblicherweise werden bei gängigen Syntheseverfahren die jeweiligen Gene oder Genabschnitte in kurze einzelsträngige, sich überlappende Oligonukleotide aufgeteilt, welche mittels automatischer Festphasensynthese hergestellt werden. Diese Oligonukleotide müssen anschließend gereinigt und zu den jeweiligen größeren Fragmenten mit Hilfe von DNA-Ligasen und DNA-Polymerasen assembliert werden. Dabei entstehen zunächst Subfragmente, die ihrerseits über spezifische Überhänge in der richtigen Reihenfolge über Ligation miteinander verknüpft werden.

Diese Technologie wurde seit 1972, als sie zum ersten Mal von Khorana und Mitarbeitern zur Synthese eines vollständigen tRNA-Gens eingesetzt wurde, kontinuierlich verbessert [1].

Spricht man über Gensynthese, sollte der Name Heyneker nicht ungenannt bleiben. Eine seiner Arbeiten dokumentierte die biologische Aktivität eines chemisch synthetisierten genetischen Elements, des 21 bp langen lac operators [2]. Der Erfolg dieser Arbeiten, motivierte diese Gruppe, das Gen für das Hormon Somatostatin (56 bp, acc. no. M10835) zu synthetisieren. Hierzu wurden acht Oligonukleotide mit einer Länge zwischen 11 und 16 Basen zu einem vollständigen Gen assembliert [3].

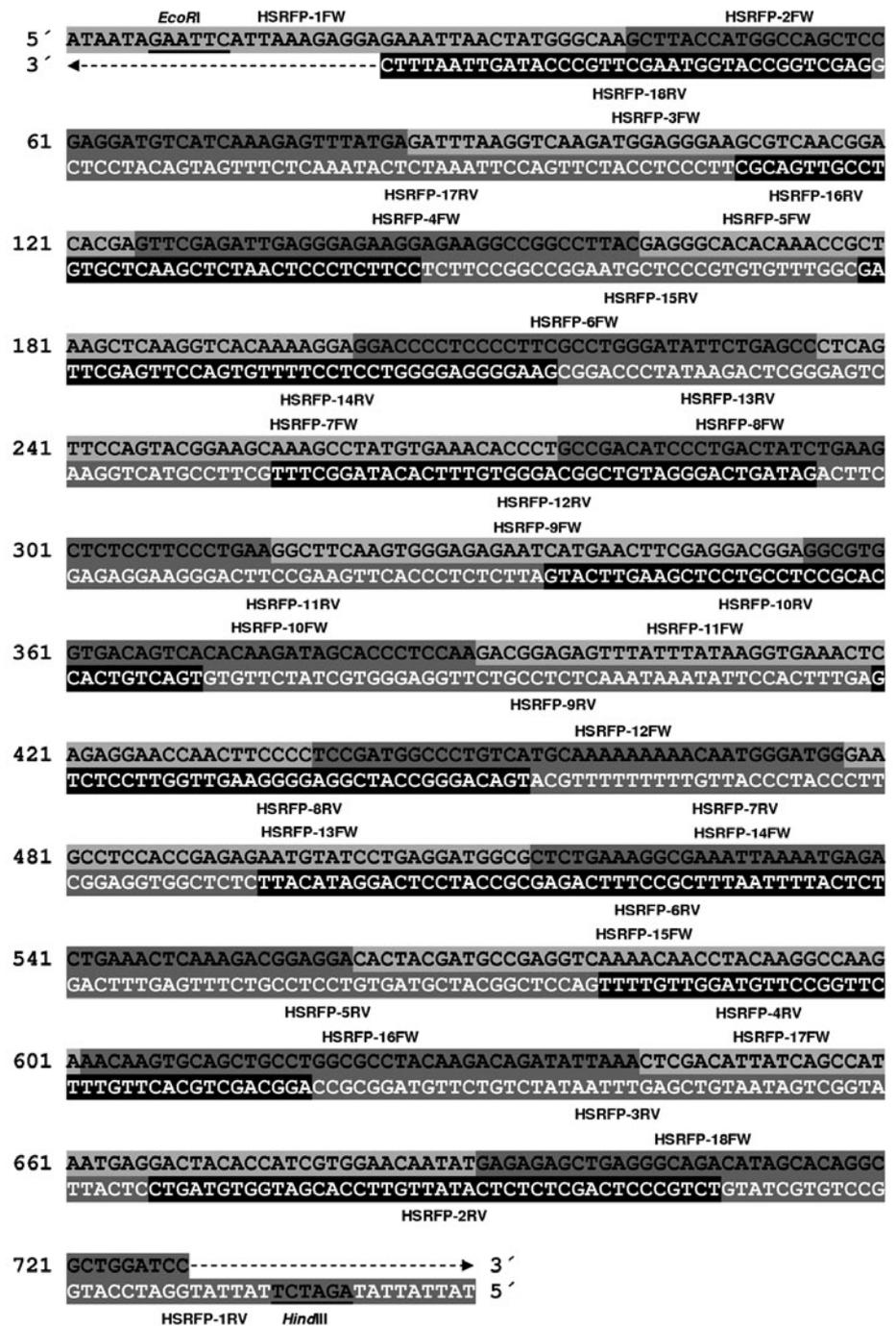


Abb. 1. Assemblierung von 36 Oligonukleotiden mit einer Länge von 32 bis 50 Basen zu einem Gen mit einer Länge von 750 bp [8].

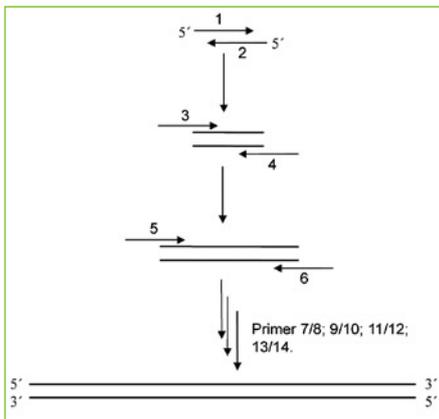


Abb. 2. Verfahren zum Aufbau eines synthetischen Gens. Zunächst wird mit zwei partiell überlappenden Oligonukleotiden (1/2) eine erste PCR durchgeführt. Anschließend wird das PCR-Amplifikat (im gewählten Beispiel 110 bp) gereinigt und dann als Matrize für eine weitere PCR eingesetzt (mit Primern 3 und 4). Im vorliegenden Fall wurden insgesamt 14 Oligonukleotide zur Synthese eines Gens mit einer Länge von 661 bp eingesetzt [10].

In den späten 1970er Jahren konnte Khorana und Kollegen die vollständige *de novo* Synthese eines 207 bp langen suppressor tRNA Gens mittels Ligation von hybridisierten Oligonukleotidpaaren zeigen [4]. Dieselbe Gruppe hat in nahezu 50 weiteren Veröffentlichungen systematisch verschiedene Aspekte der Gensynthese untersucht [5]. In der gleichen Zeit wurden an anderer Stelle auch humanes Insulin und das humane Wachstumshormon auf der Basis eines synthetischen Gens rekombinant in *E. coli* exprimiert [6, 7].

Alternativ dazu können DNA Konstrukte auch dadurch erhalten werden, indem ein Gemisch an einzelsträngigen Oligonukleotiden partiell hybridisiert und verbleibende Lücken mittels DNA-Polymerase aufgefüllt werden. Nach Abschluss der enzymatischen Auffüllreaktion, werden die verbliebenen Einzelstrangbrüche mit DNA-Ligase verschlossen. Dieses Verfahren wird als „Gap filling“ Methode bezeichnet.

Ebenso kann die gesamte Gensequenz in eine definierte Anzahl an relativ kurzen Oligonukleotiden (30–60 Basen) aufgeteilt werden. Das gesamte Gemisch wird über iterative Zyklen hybridisiert und anschließend über flankierende Primer amplifiziert. Diese Methode beruht nicht auf einer Ligation, sondern ausschließlich auf einer DNA-Polymerase-Reaktion (Abb. 1) [8, 9].

Eine weitere Möglichkeit besteht im Aufbau eines längeren Genfragmentes über hintereinander geschaltete Polymerasekettenreaktionen. Es werden, ausgehend von der Mitte des zu synthetisierenden Gens, zunächst zwei teilweise überlappende Oligonukleotide (60–100 Basen) hybridisiert und die einzelsträngigen Bereiche mit DNA-Polymerase aufgefüllt. Das entstandene PCR-Produkt wird gereinigt und dient seinerseits als Matrize für die folgende PCR mit weiteren teilweise überlappenden Oligonukleotiden. Dieser Vorgang wird solange durchgeführt, bis das Gen

vollständig aufgebaut ist (Abb. 2) [10–15]. Bei größeren Genen baut man die vollständige Sequenz nach diesem Verfahren gewöhnlich aus mehreren

dieser, meistens subklonierter, Teilfragmente zusammen [13].

Neben den genannten Standardverfahren wurde 2004 eine Methode publiziert, die Oligonukleotide verwendet, welche auf einem programmierbaren Mikrochip synthetisiert wurden. Entsprechende Oligonukleotide wurden in einer Multiplexreaktion zur Synthese von 21 Genen, die für die entsprechenden Untereinheiten der 30S Untereinheit des *E. coli* Ribosoms kodieren, verwendet [16].

Ein grundsätzliches Problem bei allen genannten Verfahren besteht in der Fehlerhaftigkeit der Synthese von Oligonukleotiden. Das Ergebnis eines Gensynthese-Experiments hängt wesentlich von der Qualität der eingesetzten Oligonukleotide ab. Die Kopplungseffizienz der Phosphoramidit-Bausteine bei der chemischen Oligonukleotidsynthese spielt hierbei eine herausragende Rolle. Beträgt die Kopplungseffizienz bei der Herstellung eines 40-mer Oligonukleotids 99 %, so haben nur 67,5 % der Produkte ( $0,99^{39} \times 100$ ) eine Länge von 40 Nucleotiden. Fällt die Kopplungseffizienz unter 98 %, ist die Länge der synthetisierten Oligonukleotide, die man in ausreichender Menge erhalten kann, entsprechend begrenzt. Es ist leicht ersichtlich, dass diese Rohprodukte nicht ohne vorgeschaltete Reinigungsprozeduren, z.B. high performance liquid chromatography (HPLC) oder Polyacrylamidgelelektrophorese (PAGE), verwendet werden können. Falls 95 % der verwendeten Oligonukleotide die richtige Sequenz aufweisen und 5 % fehlerhaft sind, so kann bei einem typischen Assemblierungsprodukt aus 30 einzelsträngigen Oligonukleotiden lediglich mit einer Wahrscheinlichkeit von  $0,95^{30} \times 100$  (21,5 %) eine korrekte Sequenz erreicht werden. Längere einzelsträngige Bausteine, die bevorzugt in der klassischen Gensynthese Verwendung finden, sind entsprechend umso anfälliger für Fehler.

Werden solche fehlerhaften Oligonukleotide für die Synthese verwendet oder wird auf aufwendige Reinigungsschritte verzichtet, erfordert dies eine sorgfältigere Kontrolle bzw. Fehlerkorrektur auf der Ebene des synthetisierten Fragmentes. Typischerweise wird dies durch umfangreiche DNA-Sequenzierungen sowie zeitaufwendige molekularbiologische Arbeitsschritte erreicht.

Zur Verringerung der Fehlerrate wurden einige Selektionsverfahren entwickelt, die allerdings teilweise sehr arbeitsintensiv sind und einen zusätzlichen Zeit- und Kostenaufwand bedeuten. In der Literatur beschrieben sind z.B. der Einsatz von Fishing-Oligonukleotiden [16], mismatch-

binding Enzymen der mutS-Familie [17, 18] oder spezifische Endonukleasen aus Bakterien und Phagen [19].

Neben der Qualität der eingesetzten Oligonukleotide liegt ein weiteres Problem in der Assemblierung der einzelsträngigen Bausteine zu längeren doppelsträngigen DNA-Fragmenten. Vielfach ist es unmöglich, die Bildung von Sekundärstrukturen zu vermeiden, welche beispielsweise durch inverted repeats, einen aussergewöhnlich hohen oder niedrigen GC-Gehalt oder durch repetitive Strukturen verursacht werden. Üblicherweise müssen diese Probleme durch ein modifiziertes und dadurch wiederum aufwendigeres Design, z.B. bei der Assemblierung von Subfragmenten, gelöst werden. Vielfach wird aber auch eine Anpassung der Codon-Usage verwendet, um solche Sequenz-abhängigen Schwierigkeiten zu eliminieren und so die Synthese eines Fragmentes zu ermöglichen. Ein Kompromiss zur optimalen Expressionsoptimierung muss dabei in Kauf genommen.

### Gensynthese mit Hilfe doppelsträngiger Bausteine

Im Gegensatz zu herkömmlichen Gensynthesemethoden, arbeitet die Slonomics Methode mit einem Baukasten von vorgefertigten und universell einsetzbaren Triplet-Bausteinen aus doppelsträngiger DNA (Abb. 3). Die Verwendung einer

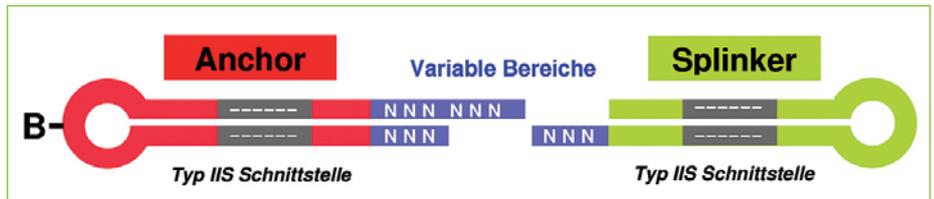


Abb. 3: Universelle Triplet-Bausteine, die bei der Slonomics-Methode eingesetzt werden.

definierten Anzahl von standardisierten Bausteinen macht die Synthese projektspezifischer Oligonukleotide überflüssig. Die universellen Triplet-Bausteine werden zunächst in großem Maßstab auf dem Wege einer herkömmlichen Oligonukleotidsynthese hergestellt. Aufgrund ihrer speziellen Sequenz bilden diese durch intramolekulare Hybridisierung eine Haarnadelstruktur aus. Jedes dieser Standardelemente besteht aus einer kurzen Schleife, einem konstanten doppelsträngigen Bereich und einem Triplet-Überhang (Abb. 3). Der konstante Bereich enthält eine Erkennungssequenz für eine Typ-II-S-Restriktionsendonuklease.

Typ-II-S-Restriktionsenzyme sind sog. „outside cutters“, d.h. sie erkennen asymmetrische Sequenzbereiche und schneiden die DNA außerhalb ihrer Erkennungssequenz. Ein wesentlicher Vorteil liegt darin, dass lediglich die Erkennungssequenz definiert vorhanden sein muss, jedoch die zu schneidende Sequenz, die bei der Typ-II-S-Enzymklasse außerhalb der Erkennungssequenz liegt, variabel sein kann.

Es gibt zwei unterschiedliche Klassen an Bausteinen, die als Splinker und Anchor bezeichnet werden; beide Moleküle werden im Syntheseprozess miteinander kombiniert. Die beiden Klassen unterscheiden sich unter anderem in der Art der verwendeten Schnittstelle innerhalb des konstanten Sequenzabschnitts (Abb. 3).

Im Gegensatz zu den Splinker-Molekülen, enthält jedes Anchor-Molekül zusätzlich ein Biotin-Molekül in der Schleife der Haarnadel, das eine Immobilisierung des Anchors an eine avidin- oder streptavidinbeschichtete Oberfläche erlaubt. Die Anchor-Moleküle unterscheiden sich untereinander lediglich in ihrem variablen Bereich aus sechs Basen, der aus einem doppelsträngigen Basentriplett sowie einem Dreibasenüberhang gebildet wird. Splinker-Moleküle sind dagegen nur

durch ihren variablen Dreibasenüberhang unterscheidbar. Um nun eine komplette Bibliothek mit allen möglichen Varianten aufzubauen, müssen 64 (4<sup>3</sup>) unterschiedliche Splinker und 4096 (4<sup>6</sup>) Anchor-Moleküle bereitgestellt werden (Abb. 3).

Durch die Kombination von Anchor- und Splinker-Molekülen können sehr einfach und effizient längere, doppelsträngige Genfragmente aufgebaut werden. Hierzu wird prinzipiell ein nicht immobilisierter Anchor mit einem Splinker durch Hybridisierung und anschließender Ligation der zueinander komplementären Dreibasenüberhänge verbunden. Das entstandene Ligationprodukt wird über Biotin an eine entsprechend beschichtete Oberfläche in einer 96er Mikrotiterplatte gebunden. Im nächsten Schritt werden die immobilisierten Ligationprodukte mit dem entsprechenden Restriktionsenzym auf der Anchor-Seite so geschnitten, dass dabei der variable Sequenzanteil des Anchors (= Donor) auf den Splinker (= Akzeptor) übertragen wird. Im Überstand befindet sich nun ein um drei Basenpaare verlängertes Splinker-Molekül, das in einem weiteren Zyklus als Akzeptor für ein neues Anchor-Molekül dienen kann. Das verkürzte Anchor-Molekül bleibt an der Festphase zurück. Durch eine zyklische Wiederholung dieser Reaktionsschritte, die jeweils mit einem um drei Basen verlängerten Splinker-Molekül und einem entsprechenden, gezielt ausgewählten Anchor-Molekül durchgeführt werden, lassen sich so definierte Gensequenzen synthetisieren.

Wie aus Abbildung 4 hervorgeht, besteht der gesamte Slonomics-Gensyntheseprozess aus mehreren Phasen. Mit Hilfe eines Computerprogramms wird zunächst die Zielsequenz in kleinere Fragmente aufgeteilt und die dafür benötigten Anchor- und Splinker-Moleküle festgelegt.

Während der anschließenden Elongation werden zunächst kurze Subfragmente (Elongationsblöcke oder E-Blöcke) der zu synthetisierenden Sequenz durch mehrfache Wiederholung der oben beschriebenen Reaktionszyklen aus Ligation und Restriktion hergestellt. Da viele Elongationsschritte parallel durchgeführt werden können, wird das entsprechende Gen, wenn auch in kurzen Fragmenten, bereits in dieser Phase vollständig synthetisiert. Der Zusammenbau der E-Blöcke erfolgt in der nachfolgenden

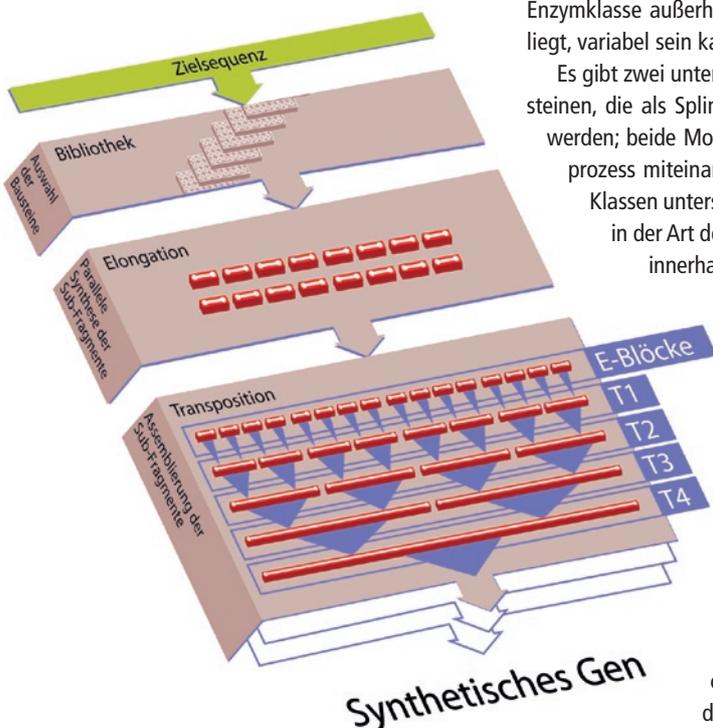


Abb. 4: Schematische Darstellung der gesamten Slonomics-Methode. 4096 Anchor-Elemente und 64 Splinker-Elemente dienen als Ausgangsmaterial (Bibliothek) bei der Synthese. Kleine Subfragmente des entsprechenden Ziel-Gens (E-Blöcke) werden durch wiederholte Ligation, Immobilisierung und Restriktion gebildet. Die entsprechenden E-Blöcke werden anschließend in der Transpositionsphase zu größeren Fragmenten (T1, T2, etc.) assembliert.

Transposition. Hierbei werden die E-Blöcke nach wechselseitigem Verdau mit einer entsprechenden Typ-III-Restriktionsendonuklease paarweise miteinander verbunden, was im jeweiligen Transpositionsschritt zu einer Verdoppelung der Sequenzlänge führt.

## Fazit

Herkömmliche Gensynthesen arbeiten immer mit einzelsträngigen Oligonukleotiden, die entsprechend der zugrunde liegenden Gensequenz für jedes Konstrukt individuell synthetisiert werden müssen. Die Slonomics-Gensynthesetechnologie basiert auf etablierten molekularbiologischen Techniken und der Verwendung einer universell einsetzbaren Bibliothek aus Bausteinen, vergleichbar mit einem Lego-System. Für alle denkbaren Genkonstrukte ist eine Bibliothek von maximal 4160 unterschiedlichen Oligonukleotiden notwendig. Die entsprechenden Elemente können einfach in hoher Qualität und Menge mittels Standardverfahren produziert werden. Da für jede Synthese dieselbe Bausteinbibliothek verwendet werden kann, ist eine Qualitätskontrolle jeweils nur einmal erforderlich. Durch die Möglichkeit, sowohl die Qualität der Ausgangsmaterialien sowie die Durchführung des biochemischen Herstellungsverfahrens auf hohem Niveau zu standardisieren, lässt sich die Technologie vollständig auf eine speziell dafür entwickelte Roboter-Plattform übertragen.

Durch das Konzept, eine Gensynthese mit Hilfe doppelsträngiger Bausteine durchzuführen, wird es zukünftig leicht möglich sein, auch schwierige DNA-Sequenzen, wie beispielsweise sehr GC-reiche oder hoch repetitive Sequenzen, herstellen zu können. Darüber hinaus eignet sich dieses Verfahren im Gegensatz zu den Standardverfahren wesentlich besser zur Sequenzoptimierung, weil hierbei keine Anpassung der Sequenz aus Gründen der Synthesevereinfachung notwendig ist.

Neben der Erzeugung synthetischer Einzelkonstrukte, bildet die vorgestellte Methode die technologische Grundlage für die kontrollierte Erzeugung von Mutanten eines Gens, die in Mutantenbibliotheken zusammengefasst werden können Slonomax. Durch die Verwendung von definierten Gemischen der verwendeten Triplett-Bausteine bei der Synthese wird ein paralleler Einbau unterschiedlicher Codons an jeder beliebigen Position einer Gensequenz ermöglicht. An jeder Sequenzposition lassen sich so ganz gezielt beliebige Codons für alle gewünschten Aminosäuren einsetzen, was eine maximale Vielfalt von Varianten auf Proteinebene mit einer minimalen Anzahl an dafür benötigten Genvarianten ermöglicht.

## Literatur

- [1] Agarwal, K. L.; Yamazaki, A.; Cashion, P. J.; Khorana, H. G.: Chemical synthesis of polynucleotides, *Angew Chem Int Ed Engl* 11, 451–9 (1972)
- [2] Heyneker, H. L.; Shine, J.; Goodman, H. M.; Boyer, H. W.; Rosenberg, J.; Dickerson, R. E.; Narang, S. A.; Itakura, K.; Lin, S.; Riggs, A. D.: Synthetic lac operator DNA is functional in vivo, *Nature* 263, 748–52 (1976)
- [3] Itakura, K.; Hirose, T.; Crea, R.; Riggs, A. D.; Heyneker, H. L.; Bolivar, F.; Boyer, H. W.: Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin, *Science* 198, 1056–63 (1977)
- [4] Sekiya, T.; Takeya, T.; Brown, E. L.; Belagaje, R.; Contreras, R.; Fritz, H. J.; Gait, M. J.; Lees, R. G.; Ryan, M. J.; Khorana, H. G.; Norris, K. E.: Total synthesis of a tyrosine suppressor transfer RNA gene. XVI. Enzymatic joinings to form the total 207-base pair-long DNA, *J Biol Chem* 254, 5787–801 (1979)
- [5] Khorana, H. G.: Total synthesis of a gene, *Science* 203, 614–25 (1979)
- [6] Goeddel, D. V.; Heyneker, H. L.; Hozumi, T.; Arentzen, R.; Itakura, K.; D. Yansura, G.; Ross, M. J.; Miozzari, G.; Crea, R.; Seeburg, P. H.: Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone, *Nature* 281, 544–8 (1979)
- [7] Goeddel, D. V.; Kleid, D. G.; Bolivar, F.; Heyneker, H. L.; Yansura, D. G.; Crea, R.; Hirose, T.; Kraszewski, A.; Itakura, K.; Riggs, A. D.: Expression in *Escherichia coli* of chemically synthesized genes for human insulin, *Proc Natl Acad Sci U S A* 76 106–10 (1979)
- [8] Fischer, M.; Haase, I.; Wiesner, S.; Müller-Taubenberger, A.: Visualizing cytoskeleton dynamics in mammalian cells using a humanized variant of monomeric red fluorescent protein, *FEBS Lett* 580 2495–502 (2006)
- [9] Stemmer, W. P.; Cramer, A.; Ha, K. D.; Brennan, T. M.; Heyneker, H. L.: Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides, *Gene* 164, 49–53 (1995)
- [10] Echt, S.; Bauer, S.; Steinbacher, S.; Huber, R.; Bacher, A.; Fischer, M.: Potential anti-infective targets in pathogenic yeasts: structure and properties of 3,4-dihydroxy-2-butanone 4-phosphate synthase of *Candida albicans*, *J. Mol. Biol.* 341, 1085–96 (2004)

Weitere Literatur ist beim Autoren erhältlich.

## ► KONTAKT

**Prof. Dr. Markus Fischer**  
Universität Hamburg  
Institute of Biochemistry and Food Chemistry  
Tel.: 040/42838-4357  
Fax: 040/42838-4342  
markus.fischer@chemie.uni-hamburg.de  
www.chemie.uni-hamburg.de/lc/