

# DNA-Computing – ein funktionales Modell im laborpraktischen Experiment

E. Stoschek, M. Sturm, T. Hinze

Technische Universität Dresden, Fakultät Informatik, Institut für Theoretische Informatik, D-01062 Dresden, (e-mail: {sturm,hinze}@tcs.inf.tu-dresden.de)

Eingegangen am 14. Februar 2000 / Angenommen am 26. Oktober 2000

**Zusammenfassung.** Im Zentrum der Betrachtungen zum DNA-Computing steht die Frage nach den Chancen und Grenzen dieses neuen Berechnungsmodells, nachdem in den letzten Jahren eine rasante Entwicklung auf das Thema aufmerksam machte. Neben beachtlichen theoretischen Untersuchungen zum „Rechnen im Reagenzglas“ werden auch laborpraktische Implementierungen favorisiert. An der TU Dresden wurde in interdisziplinärer Arbeit ein Integer-Rucksackproblem mittels eines DNA-Algorithmus im Labor gelöst und dabei eine Vielzahl molekularbiologischer Operationen analysiert. Mit Hilfe dieses Satzes von Operationen gelang eine universelle und labornahe Modellierung des DNA-Computing. Hierbei angewandte Techniken und Methoden werden vorgestellt und bewertet. Die Beschreibung des DNA-Algorithmus zeigt, wie sich Einzeloperationen vorteilhaft zu Operationsfolgen zusammensetzen lassen und gemeinsam mit einer geeigneten DNA-Kodierung der Eingangsdaten zur Lösung des Problems im Labor führen. Erstmals wurden hierbei natürliche Zahlen verarbeitet. Die Arbeitsgemeinschaft DNA-Computing Dresden konzentriert sich auf Aufgabenstellungen, die formale Modelle des DNA-Computing mit überzeugenden Laborimplementierungen verbinden.

**Schlüsselwörter:** DNA-Computing, Rekombinationstechnik, Molekularbiologie, Berechnungsmodell, funktionale Sprache, NP-Problem, Rucksackproblem, SIMD, massive Datenparallelität

**Abstract.** DNA computing as a new model for computation is worldwide considered with respect to its chances and limits. During the last years a huge progress in this field of research could be observed. Both remarkable theoretical studies about “Calculating inside the reaction tube” and lab-practical implementations are focussed. An instance of the integer knapsack problem was solved at Dresden University of Technology in an interdisciplinary manner. In this context, a multiplicity of molecular biological operations was analyzed. Using the applicable set of operations it was possible to model a universal description of DNA computing close to the laboratory. The methods and techniques the description is based on are introduced and evaluated. The specification of the DNA algorithm shows a way how single operations can

be combined advantageously to a sequence of operations. This aspect and an appropriate DNA encoding of input data led to a successful solution in the laboratory. For the first time natural numbers were processed. The Dresden DNA Computation Group concentrates on objectives that connect formal models of DNA computing with convincing implementations in the laboratory.

**Key words:** DNA computing, recombinant technologies, molecular biology, model for computation, functional language, NP problem, knapsack problem, SIMD, massive data parallelism

**CR Subject Classification:** C.1.2, D.1.1, D.3.3, F.1.2, F.2, F.4.3, G.2.1

## 1 Einführung

Elektronische Rechner haben in den vergangenen Jahrzehnten die Welt verändert. Sie lösten eine Welle der Automatisierung und Beschleunigung vieler Arbeitsabläufe aus. Computertechnik flankiert nahezu alle Bereiche des Lebens, in einer Reihe von anspruchsvollen Anwendungsgebieten dominiert der Einsatz von Supercomputern. Es zeichnet sich jedoch ab, daß konventionelle Rechentechnik bei der Bearbeitung extrem rechenintensiver Aufgaben und beim Einsatz unter außergewöhnlichen Bedingungen zunehmend an ihre Grenzen stößt. Kombinatorische Suchprobleme (NP-Probleme) erweisen sich als bedeutende mathematische Problemstellungen hoher Praxisrelevanz, deren Lösung konventionelle Rechentechnik wegen ihres exponentiellen Ressourcenverbrauchs vor neue Herausforderungen stellt. In diesem Zusammenhang untersucht die Forschung weltweit alternative Hardwaremöglichkeiten, um auf deren Basis neuartige Computing-Konzepte zu etablieren. Molecular Computing stellt einen vielversprechenden Ansatz dar, sich diesen Herausforderungen zu stellen und zusätzliche Anwendungsgebiete zu erschließen, die gegenwärtig noch als realitätsfern gelten. Als Beispiele seien biologische Minicomputer genannt, die Spezialaufgaben im menschlichen Körper

übernehmen können, sowie Hochleistungsrechner mit enormer Rechengeschwindigkeit (bis zu  $1,2 \cdot 10^{18}$  ops/s), Speicherkapazität (bis zu  $10^{21}$  bit) und Speicherdichte ( $1 \text{ bit/nm}^3$ ) genannt, die sich neben einem sehr niedrigen Energieverbrauch ( $2 \cdot 10^{19}$  ops/J) auch durch vollständige Recyclbarkeit auszeichnen [11].

Molecular Computing verfolgt die Idee, organische Moleküle als Datenträger einzusetzen und mittels chemischer Reaktionen und physikalischer Prozesse die Operationen nachzubilden, um damit Rechenvorgänge zu realisieren. Dies begründet eine neuartige Rechnerarchitektur, bei der die Eingabedaten in Form geeigneter Moleküle bereitgestellt werden und die Ergebnisausgabe durch Analyse der aus der Bearbeitung hervorgegangenen Moleküle erfolgt. Prominentester Vertreter dieses Computing-Konzeptes ist derzeit das DNA-Computing, bei dem Desoxyribonucleinsäure (DNA) als Datenträger sowie moderne Rekombinationstechniken aus der Molekularbiologie im Einklang mit durchdachten, für den speziellen Einsatzzweck konstruierten Versuchsaufbauten zur Anwendung gelangen. Das Potential des DNA-Computing liegt in der massiven Datenparallelität, denn eine Operation wirkt gleichzeitig auf eine Vielzahl von DNA-Strängen. Forschung auf dem Gebiet des DNA-Computing vereint innovativ die Molekularbiologie mit der Informatik.

Die ersten praktischen Arbeiten zum DNA-Computing gehen in das Jahr 1994 zurück, als Leonard Adleman erfolgreich im Labor eine Instanz des Hamiltonkreis-Problems löste, bekannt unter dem Namen Adleman-Experiment [1]. Dieser erste Erfolg gab den entscheidenden Anstoß für zahlreiche Forschergruppen weltweit, die Idee des DNA-Computers aufzugreifen und theoretisch wie auch experimentell weiterzuentwickeln. Begleitet werden diese Aktivitäten von speziellen internationalen Workshops (DNA Based Computers, [2], [3], [28], [12], [4], [13]) sowie der Integration des Forschungsgebietes in eine Disziplin der Theoretischen Informatik, die sich speziell mit unkonventionellen Modellen der Berechenbarkeit befaßt. DNA-Computing-Modelle leisten einen großen Beitrag zu dieser Disziplin.

Die Entwicklung auf dem Gebiet des DNA-Computing umfaßt im wesentlichen folgende Etappen:

- In Konsequenz des Adleman-Experimentes entstanden Filtering-Modelle in verschiedenen Stufen (Modell nach Lipton [15], Modell Parallel Associative Memory nach Reif [19], Sticker-Modell [22], DNA-Pascal nach Roß/Wagner [21]) als Abstraktion des DNA-Computing und als massiv datenparallele Berechnungsmodelle. Der Grundgedanke dieser Modelle benutzt eine Strategie zur Lösung von NP-Problemen (Brute-Force-Ansatz): Aus einem initial bereitgestellten umfassenden DNA-Pool, der den gesamten Suchraum DNA-kodiert, werden schrittweise all jene DNA-Stränge entfernt (herausgefiltert), die keine Problemlösung darstellen. Ein finaler Test auf Vorhandensein von DNA liefert schließlich das Ja-Nein-Ergebnis.
- Von Anfang an zeigte sich die Seiteneffektanfälligkeit von molekularbiologischen Operationen als Hemmnis für große Anwendungen und damit als Herausforderung für die Schaffung entsprechend optimierter Operationen und verwendeter DNA-Sequenzen. Seiteneffekte sind dabei unerwünschte Wirkungen der DNA-Operationen.

Diese äußern sich vorwiegend in der Bildung von DNA-Strängen, die in ihrer Sequenz und/oder ihrer Struktur verfälscht sind, sowie in unvollständigen Reaktionsabläufen, bei denen ein bestimmter Teil der Eingangs-DNA unverarbeitet bleibt. Als Ausweg zur Eindämmung solcher Seiteneffekte wurden spezielle DNA-Kodierungen erdacht, wie beispielsweise der Einsatz von synthetisierten DNA-Einzelsträngen, die generell nur drei der vier möglichen Basen verwenden [5], oder die Verwendung zellulärer doppelsträngiger DNA aus Organismen [24]. Des Weiteren wurden steuerbare Reaktionsparameter (z.B. Temperaturverläufe, DNA-Konzentration, ausgewählte Reagenzien) hinsichtlich des DNA-Computing optimiert [9].

- Zunächst konnte DNA-Computing jeweils ein spezielles Problem mittels eines ebenso speziell zugeschnittenen Laboraufbaus lösen. Seit etwa 1997 trat die Idee eines universellen DNA-Computers in den Vordergrund, der über eine minimale Anzahl von exakt definierten Basisoperationen verfügt, die lediglich in ihrer Abfolge variiert werden können. Ein universeller DNA-Computer erreicht dabei die Berechnungsstärke der Turingmaschine (Splicing-Systeme in verschiedenen Ausprägungen [16], [18], [25], DNA-Haskell [10], P-Systeme [17]).
- Gegenwärtig werden universelle DNA-Computing-Modelle auf Bio-Hardware implementiert. In diesem Zusammenhang erfolgte die Konstruktion des ersten Roboters, der Laborarbeiten automatisiert [8], das Design erster DNA-Chips [20] sowie die erfolgversprechende Verknüpfung von DNA mit RNA (Ribonucleinsäure) in Einheit mit Strangfixierungen an Trägersubstanzen, damit durch Verringerung des Redundanzgrades (Anzahl Strangduplikate) sowohl eine Minimierung der eingesetzten DNA-Mengen als auch eine Erhöhung der verarbeitbaren Problemgrößen erzielt werden kann [7]. Aktuelle Experimente arbeiten mit einer Problemgröße von  $n \approx 20$  [6], vgl. Adleman-Experiment:  $n = 7$ .

Alle aufgeführten Arbeiten dienen dem gleichen Ziel, dem Aufbau eines universellen DNA-Computers, wobei zum jetzigen Zeitpunkt noch keine klare Aussage getroffen werden kann, wann dieser Computer Anwendern zur Verfügung stehen wird. Eine ähnliche Situation konnte man auch im Vorfeld der elektronischen Rechentechnik beobachten: Während das Modell der Turingmaschine bereits in den 30er Jahren in der Theorie anerkannt war, stammt der erste darauf basierende elektronische von-Neumann-Computer aus dem Jahr 1952. Einer kommerziellen Anwendung des DNA-Computers stehen derzeit vor allem die hohen Kosten für die Laborausrüstungen sowie die noch bestehenden Seiteneffekte entgegen. Dennoch besteht ein großes wirtschaftliches und wissenschaftliches Interesse an DNA-Computern. Letzteres bezieht sich auf die Formalisierung paralleler Rechnerarchitekturen und deren beschreibende Kalküle, die semantische Konzeption von datenparallelen Programmiersprachen, und verleiht der Komplexitäts- und Berechenbarkeitstheorie neue Impulse. Die Molekularbiologie profitiert von den Erkenntnissen aus laborpraktischen Experimenten. Das wirtschaftliche Interesse forciert vor allem den einsatzfähigen universellen Bio-Computer mit seinen für möglich gehaltenen

nen technischen Leistungsparametern und Anwendungsfeldern.

Die Arbeiten der Verfasser konzentrieren sich darauf, die Lücke zwischen den formalen Modellen des DNA-Computing und ihrer laborpraktischen Implementierung zu schließen, indem eine labornahere Modellierung favorisiert wird. Die Dresdner Arbeitsgemeinschaft DNA-Computing (<http://www.tcs.inf.tu-dresden.de/dnacom/>) analysierte die Prozesse des DNA-Computing auf molekularer Ebene und konstruierte einen DNA-Algorithmus zur Lösung des Integer-Rucksackproblems, wobei die Verarbeitung natürlicher Zahlen im Mittelpunkt stand. Die dabei gewonnenen Erkenntnisse und Erfahrungen führten zur Spezifikation der DNA-Operationen im formalen Modell DNA-Haskell. Alle Modelloperationen können unmittelbar im Labor ausgeführt werden. Gegenwärtig wird eine umfassende PC-gestützte Simulation von DNA-Haskell umgesetzt, welche die Vorbereitung von weiteren Experimenten unterstützen soll. Im vorliegenden Artikel werden die dem Modell zugrundeliegenden Operationen mit ihrem molekularbiologischen Hintergrund vorgestellt. Ein Beispielalgorithmus zur Lösung eines Integer-Rucksackproblems zeigt, wie sich aus diesen Operationen ein DNA-Algorithmus zusammensetzen läßt und wie dieser Algorithmus im Labor abgearbeitet wird. Diese Laborarbeiten motivierten die Verfasser zu einem eigenen Ansatz für einen universellen DNA-Computer, der genau auf denselben untersuchten Operationen aufbaut [25].

## 2 Molekularbiologische Grundlagen zum Modell und zur Sprache DNA-HASKELL

DNA ist in zellularen Organismen das Speichermedium für genetische Informationen, das in Form strangartiger organischer Moleküle auftritt [27]. DNA-Stränge bestehen aus Sequenzen von Nucleotiden, den DNA-Grundbausteinen, von denen sich vier Typen (A, C, T, G) aufgrund der in ihnen verwendeten Basen (Adenin, Cytosin, Thymin, Guanin) unterscheiden lassen. Neben einer dieser Basen enthält ein Nucleotid einen Desoxyribose-Ring sowie eine Phosphatgruppe. Letztere verkettet benachbarte Nucleotide über eine kovalente, thermisch stabile Phosphodiesterbindung zu sogenannten DNA-Einzelsträngen. Innerhalb jeder Phosphatgruppe tritt eine elektrisch negative Ladung auf, die in Summe die negative Gesamtladung von DNA ergibt, welche für spezielle Analysezwecke ausgenutzt wird. Jeder Desoxyribose-Ring enthält fünf Kohlenstoffatome, die gemäß ihrer Position mit 1' bis 5' bezeichnet werden. Die Kohlenstoffatome an der 3'- und der 5'-Position dienen zur Anlagerung der benachbarten Phosphatgruppen. An den Strangenden sind diese beiden Positionen nicht durch eine solche Verbindung abgesättigt, so daß diese Enden in Konsequenz als 3'- bzw. 5'-Ende bezeichnet werden. Da sich diese Enden chemisch unterscheiden, sind DNA-Stränge richtungsbehaftet. Entsprechend der Orientierung definiert man die Leserichtungen sense (von 5' nach 3') und antisense (von 3' nach 5'). An die DNA-Strangenden, die naturgemäß mit einer Hydroxylgruppe (-OH) markiert sind, können auch andere spezifische chemische Gruppen oder Moleküle angelagert werden, wie beispielsweise 5'-Phosphat, 5'-Biotin oder 3'-Phosphat. Diese sogenannten

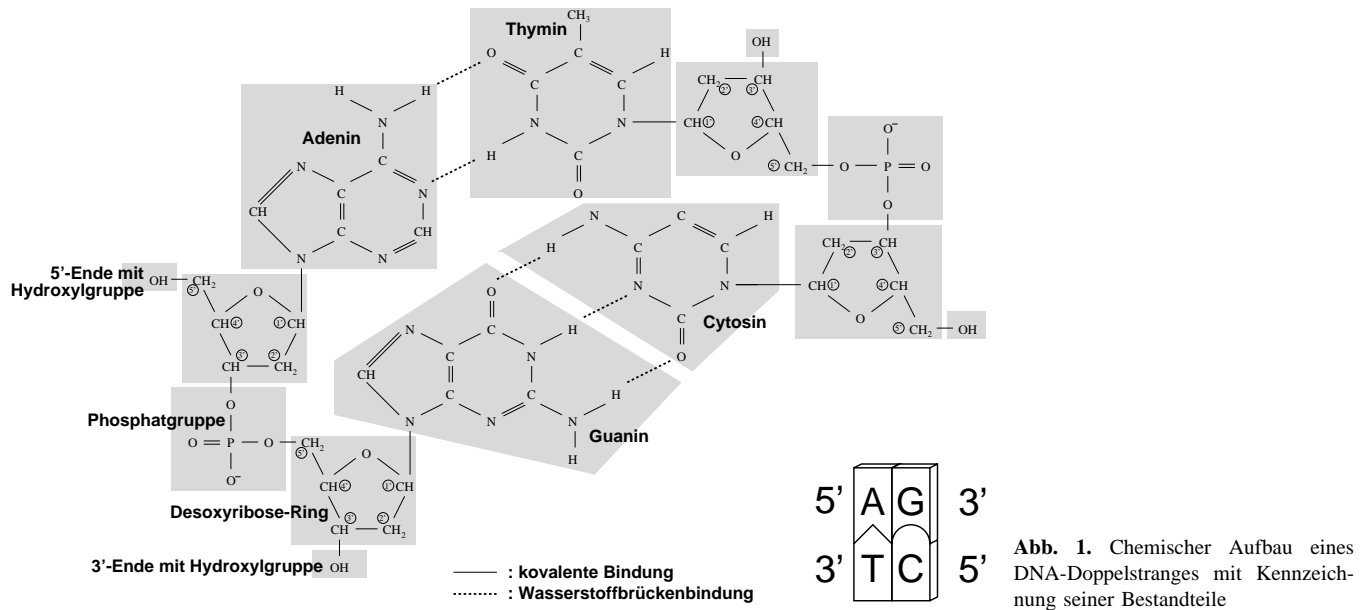
Strangenden-Label ermöglichen spezielle Separationstechniken.

Neben DNA-Einzelsträngen können auch DNA-Doppelstränge verarbeitet werden. Ein DNA-Doppelstrang besteht aus mindestens zwei DNA-Einzelsträngen. Sie werden durch Wasserstoffbrücken zusammengehalten. Die thermisch instabilen Wasserstoffbrücken verknüpfen die Basen in den Nucleotiden zweier Stränge. Diese Bindungen können aber nur zwischen ganz bestimmten Nucleotiden eingegangen werden. T kann sich nur mit A verbinden und C nur mit G. Aus der Sequenz der Nucleotide des einen beteiligten DNA-Einzelstranges kann die Sequenz des gegenüberliegenden Stranges im Doppelstrang aufgebaut werden. Die so gepaarten DNA-Einzelstränge werden als komplementär bezeichnet. Sie verlaufen in entgegengesetzter Richtung, der eine Strang von 5' nach 3', der andere von 3' nach 5' (antiparallel). Werden die beiden komplementären Stränge um eine Achse gewunden, gestaltet sich die Struktur einer Doppelhelix. Die Doppelhelix ist das klassische Modell eines DNA-Doppelstranges. Es wurde im Jahre 1953 durch Watson/Crick veröffentlicht [26] und gilt seitdem als Grundlage in der Molekularbiologie und Medizin. Abbildung 1 zeigt den Aufbau eines DNA-Doppelstranges mit vier Nucleotiden. DNA-Stränge verfügen über eine Reihe von Eigenschaften, die ihre Nutzung als Datenträger im Molecular Computing vorteilhaft erscheinen lassen:

- DNA-Stränge sind im Vergleich zu herkömmlichen Datenspeichern sehr klein und kompakt. DNA erlaubt eine Speicherdichte von bis zu  $10^{21}$  bp/l (Basenpaare pro Liter), das entspricht etwa  $1 \text{ bit/nm}^3$ .
- DNA-Stränge sind unter geeigneten Bedingungen beliebig lange konservierbar und damit als persistentes Speichermedium geeignet.
- DNA-Stränge lassen sich unter Verwendung einer Vielzahl spezifisch wirkender Enzyme rekombinieren, d.h. verarbeiten, wodurch ein großes Spektrum an biochemischen Operationen existiert.
- DNA-Stränge sind richtungsbehaftet, was eine Kodierung und Dekodierung von Daten in DNA-Sequenzen erleichtert und die Modellierung von DNA-Strängen durch Zeichenketten erlaubt.
- DNA-Stränge können in beliebiger Nucleotidsequenz synthetisiert werden, und die Sequenz von DNA-Strängen kann bestimmt werden. Es existiert somit ein Instrumentarium zur Kodierung und Visualisierung von DNA-Daten.
- DNA-Stränge können auch außerhalb von Zellen verarbeitet und aufbewahrt werden. Dies ermöglicht ein effizientes in-vitro-Handling im Labor.

Nachfolgend werden ausgewählte molekularbiologische Prozesse vorgestellt, die Operationen im DNA-Computing bilden. Ziel ist es, einen Überblick über die Wirkung und Ausführung dieser Prozesse zu geben und auf mögliche Seiteneffekte sowie Fehlerquellen hinzuweisen. DNA-Operationen werden i.a. in einem Reaktionsgefäß (Tube) ausgeführt, wofür sich der Begriff in vitro etabliert hat. Im Gegensatz dazu steht der Begriff in vivo, der Abläufe in der lebenden Zelle charakterisiert.

Ein Tube kann eine beliebig große, aber endliche Menge an DNA-Strängen enthalten, die dort in wässriger Lösung

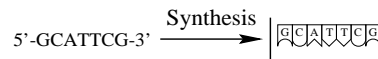


vorliegen. Übliche Reaktionsansätze beinhalten bis zu etwa 20 nmol DNA-Moleküle, die in bis zu etwa 2 ml Reinstwasser gelöst sind. Jeder DNA-Strang wird durch eine Vielzahl identischer DNA-Moleküle repräsentiert, d. h. jeder DNA-Strang existiert in einer Anzahl von bis zu mehreren Millionen Duplikaten im Tube. Dadurch ist eine hohe Redundanz der Informationsspeicherung gewährleistet. Die Eingangsoperanden einer DNA-Operation liegen als Menge von DNA-Strängen in einem Tube vor, wobei jede DNA-Operation gleichzeitig idealerweise auf die Gesamtheit aller DNA-Stränge in diesem Tube wirkt. Im Ergebnis entsteht ein modifizierter Tubeinhalt. DNA-Operationen sind somit massiv datenparallel. Die als DNA-Operationen bezeichneten molekularbiologischen Prozesse sind parameterbehaftet. Über die Belegung von Parametern erfolgt die Prozeßsteuerung. Typische Parameter sind Ausgangsstoffe und ihre Konzentrationen, Reihenfolge und Zeitpunkte ihrer Zusammenführung, der Temperatur-Zeit-Verlauf des Reaktionsansatzes, pH-Wert sowie von außen einwirkende Kräfte (z.B. Schütteln, Zentrifugieren, Durchmischen und Ruhelage). Neben den Parametern gibt es auch eine Reihe unkontrollierbarer Einflußgrößen, welche die Reproduzierbarkeit von DNA-Operationen beeinträchtigen und die Seiteneffektanfälligkeit bedingen. Durch die Konstruktion des Versuchsaufbaus als möglichst geschlossenes System und die exakte Protokollierung des gesamten Ablaufes einschließlich einer gewissenhaften Auswertung ist man bestrebt, die Seiteneffektanfälligkeit zu minimieren.

DNA-Operationen lassen sich in mehrere Klassen einteilen: Operationen zum Gewinnen von DNA-Strängen, zum Knüpfen und Aufbrechen von Wasserstoffbrückenbindungen, zum Mischen von DNA-Lösungen, zum Separieren und Analysieren nach verschiedenen Merkmalen sowie enzymatische Reaktionen.

### 2.1 Gewinnen von DNA-Strängen

Die DNA-Einzelstrangsynthese ist eine von mehreren Möglichkeiten, DNA-Stränge zu gewinnen. Alternativ gibt es



**Abb. 2.** Beispiel für Synthesis

auch Verfahren, DNA aus Gewebe, Blut, Bakterien etc. zu isolieren, wodurch man zumeist DNA-Doppelstränge mit einer vorgegebenen Nucleotidpaarsequenz und Länge erhält.

**2.1.1 Synthesis.** Unter Synthesis versteht man das Erzeugen von DNA-Einzelsträngen frei wählbarer Nucleotidsequenz, siehe Abb. 2. Die ersten erfolgreichen Bestrebungen, DNA-Einzelstränge künstlich herzustellen, liegen mehr als 30 Jahre zurück. Heute gibt es eine Reihe unterschiedlicher Verfahren, von denen sich die im Jahre 1975 durch eine Forschergruppe um Letsinger entwickelte Methode durchsetzen konnte [14]. Mit ihr lassen sich DNA-Einzelstränge bis zu etwa 100 Nucleotiden in großer Reinheit (weniger als 5% fehlerhafte Stränge) generieren. Die Synthese erfolgt in antisense-Richtung. Das Verfahren arbeitet nach dem Prinzip der „wachsenden Kette“: Fixiert an einem Träger wird der DNA-Einzelstrang beginnend am 3'-Ende zyklisch Nucleotid für Nucleotid aufgebaut. Zum Anfügen eines neuen Nucleotids dient ein Zyklus, der aus vier Reaktionsschritten besteht. Nachdem die identischen DNA-Einzelstränge vollständig aufgebaut sind, werden sie vom Träger abgelöst und von verschiedenen Reaktionsresten bereinigt. Es sind Geräte verfügbar, die das Synthetisieren vollautomatisch übernehmen können. Die DNA-Einzelstrangsynthese gilt allgemein als zuverlässig.

### 2.2 Knüpfen und Aufbrechen von Wasserstoffbrückenbindungen

DNA kann in Form von DNA-Einzelsträngen und DNA-Doppelsträngen vorliegen. Beide Formen sind ineinander überführbar. DNA besitzt eine spezifische Schmelztemperatur, unterhalb derer sich die Wasserstoffbrücken in der

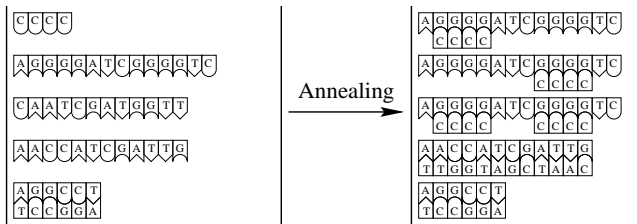


Abb. 3. Beispiel für Annealing

beschriebenen Weise ausbilden und oberhalb derer sie aufbrechen. Die Schmelztemperatur wird maßgeblich durch die Stranglänge und den Guanin-Cytosin-Anteil bestimmt. Der Temperatur-Zeitverlauf ist der wichtigste Parameter der DNA-Operationen in dieser Klasse.

**2.2.1 Annealing.** Unter Annealing (Hybridisieren, Erstarren) versteht man das Zusammenlagern von mindestens zwei antiparallelen und komplementären DNA-Einzelsträngen oder Einzelstrangüberhängen zu DNA-Doppelsträngen unter Bildung aller Anlagerungsmöglichkeiten, siehe Abb. 3. Annealing wird als ein langsam verlaufender Abkühlungsprozeß laborpraktisch realisiert. Hierzu erhitzt man die DNA-Probe auf eine Temperatur, die mit Sicherheit oberhalb der Schmelztemperatur liegt (Richtwert: 94° C), um bereits vorhandene Wasserstoffbrücken abzubauen und hinreichend viel kinetische Energie zuzuführen, damit sich die DNA-Moleküle neu formieren können. Anschließend wird die DNA-Probe über mehrere Stunden auf Raumtemperatur abgekühlt. Je langsamer dieser Vorgang erfolgt, desto mehr Zeit bleibt zur Ausbildung einer optimalen Basenpaarung, bei der alle möglichen Wasserstoffbrücken auch tatsächlich entstehen. Anderenfalls kann es geschehen, daß sich DNA-Einzelstränge nicht an allen komplementären Stellen verbinden und sich sogenannte Hairpins und Loops, d. h. nichtlineare DNA-Fragmente, bilden.

**2.2.2 Melting.** Unter Melting (Denaturieren) versteht man das Aufspalten von DNA-Doppelsträngen in die zugrundeliegenden DNA-Einzelstränge, siehe Abb. 4. Melting ist die Umkehrung des molekularbiologischen Ablaufs beim Annealing. Zum Denaturieren wird die DNA üblicherweise bis über die Schmelztemperatur erhitzt. Dabei brechen die thermisch instabilen Wasserstoffbrückenbindungen auf. Die so gewonnenen DNA-Einzelstränge liegen jedoch nur oberhalb der Schmelztemperatur vor. Diese kann bei DNA-Strängen, die für das DNA-Computing typisch sind, durchaus mehr als 70° C betragen. Soll eine solch hohe Temperatur nach dem Melting nicht aufrechterhalten werden, besteht alternativ die Möglichkeit, DNA-Doppelstränge zu denaturieren, indem man sie alkalischen Bedingungen (hoher pH-Wert) aussetzt.

### 2.3 Mischen von DNA-Lösungen

Das Mischen von DNA-Lösungen beinhaltet sowohl die Bereitstellung von Tubes mit einer exakten DNA-Konzentration in wäßriger Lösung als auch das Vereinigen oder Aufteilen von Tubeinhalten.

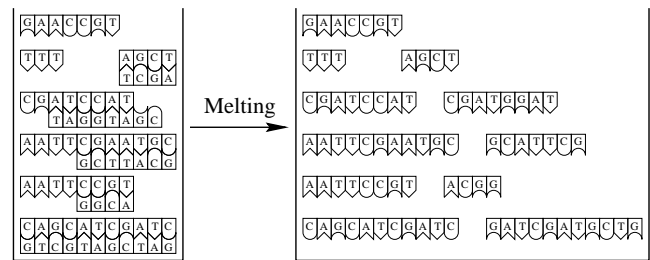


Abb. 4. Beispiel für Melting

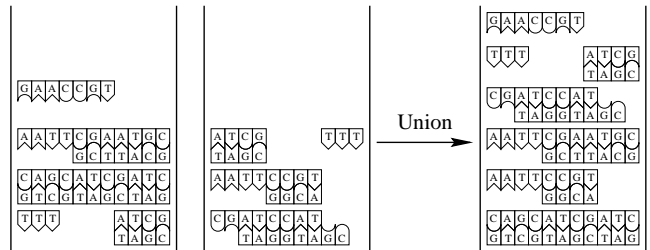


Abb. 5. Beispiel für Union

**2.3.1 Merging.** Unter Merging (Union) versteht man das Zusammenführen mehrerer Tubeinhalte in ein gemeinsames Tube, siehe Abb. 5. Üblicherweise werden hierzu die zu vereinigenden Tubeinhalte nacheinander mittels einer Pipette aufgesaugt, in das Zieltube überführt und anschließend der Zieltubeinhalt gründlich durchmischt, um einer DNA-Clustering vorzubeugen.

### 2.4 Enzymatische Reaktionen

Enzymatische Reaktionen verkörpern die Rekombinationstechnik im engeren Sinne. Sie besitzen eine große, wenn nicht überhaupt die zentrale Bedeutung in der Molekularbiologie. Chemische Umwandlungen in lebenden Zellen werden durch Enzyme gefördert und millionenfach beschleunigt. Enzyme, die von der Natur in großer Mannigfaltigkeit hervorgebracht wurden, gelten als „Katalysatoren des Lebens“. Ohne Enzyme würden molekularbiologische Prozesse, die sich als Lebensvorgänge auffassen lassen, derartig langsam ablaufen, daß die Organismen nicht lebensfähig wären. Im chemischen Sinne sind die weitaus meisten Enzyme Proteine, d. h. Aminosäureketten, deren Glieder über Amino- und Carboxylgruppen miteinander verbunden sind. Man unterscheidet hierbei 20 Aminosäuren, von denen jede durch ihre Seitenketten spezifische chemische Eigenschaften besitzt. Jedes Protein ist durch seine Aminosäuresequenz und die dreidimensionale Faltung der Kette bestimmt. Die dreidimensionale Faltung, hervorgerufen durch verschiedene chemische Bindungen, formt die Kette zu einem Knäuel und verleiht dem Protein eine charakteristische Oberflächenstruktur. Aus dieser ergibt sich seine Wirkung. Ein Protein kann mehr als 1000 Aminosäuren enthalten.

Enzyme arbeiten nach dem Prinzip von „Schlüssel und Schloß“: Ein Teil der Enzymoberfläche ist komplementär zum Substratmolekül – der DNA – geformt. Dadurch kann das Enzym an das Substratmolekül und ggf. an weitere Reagenzien andocken und sie eng umschließen. Auf diese Weise führen Enzyme die Reaktionspartner zusammen und

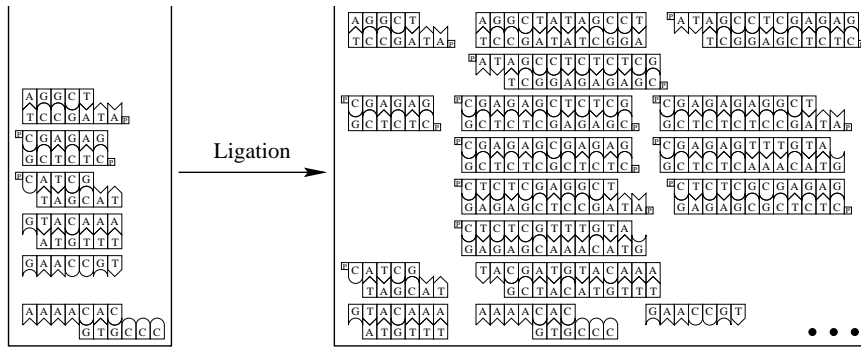


Abb. 6. Beispiel für Ligation

senken darüber hinaus aufgrund bestimmter Eigenschaften ausgewählter Aminosäureseitenketten die Aktivierungsenergie für die auszuführende Reaktion ab. Daraus ergibt sich ihre katalytische Wirkung, welche die exakte Einhaltung eines bestimmten pH-Wertes erfordert. Dieser pH-Wert wird in der DNA-Lösung durch Hinzugabe spezieller sogenannter Puffersubstanzen im Vorfeld der Reaktion eingestellt. Die eigentliche enzymatische Reaktion (Inkubation), bei der innerhalb des Substratmoleküls kovalente chemische Bindungen schrittweise ungeordnet werden, geschieht bei einer durch das Enzym festgelegten Temperatur und Zeitdauer. Das Enzym selbst geht unverändert aus dieser Reaktion hervor und löst sich wieder vom Substratmolekül ab. Über eine anschließende Erhitzung (thermische Deaktivierung) wird die dreidimensionale Faltung des Enzyms zerstört, so daß es nachfolgende molekularbiologische Prozesse nicht mehr unerwünscht beeinflussen kann. Für das DNA-Computing sind besonders die Enzyme interessant, die auf DNA als Substratmoleküle einwirken und diese zielgerichtet verändern können. Beispiele hierfür sind Ligasen, Restriktionsendonucleasen, Polymerasen, Kinasen und Phosphatasen.

**2.4.1 Ligation.** Die Ligation ist eine biochemische Reaktion, die durch das Enzym DNA-Ligase katalysiert wird und das fortgesetzte Verketteten von DNA-Doppelsträngen an ihren Enden bewirkt, siehe Abb. 6. Die Enden müssen hierzu kompatibel, also paßgenau sein und eine Phosphatgruppe besitzen. Es können sowohl DNA-Doppelstrangenden mit zueinander komplementären Einzelstrangüberhängen (sogenannte sticky-Enden) miteinander verkettet werden wie auch DNA-Doppelstrangenden ohne Einzelstrangüberhänge (sogenannte blunt-Enden).

Häufig werden aus Bakterien gewonnene und kommerziell angebotene DNA-Ligasen verwendet, wie z.B. die T4-DNA-Ligase. Sie entfaltet ihre maximale Aktivität in einem pH-Bereich von etwa 7,5 bis 8,0. Eine in-vitro-Ligation führt man gewöhnlich bei einer Temperatur zwischen 16° C und 20° C durch, wobei die Inkubationszeit i.a. bis zu 16 h beträgt. Die thermische Deaktivierung von T4-DNA-Ligase erfolgt bei 65° C für etwa 10 min.

Die Ligation ist mit mehreren Seiteneffekten behaftet. Infolge intramolekularer Ligation können sich DNA-Ringe bilden. Desweiteren ist es möglich, daß die DNA-Konzentrationen der durch die Ligation entstandenen DNA-Fragmentketten-Kombinationen stark voneinander abweichen. Während von einer bestimmten Verkettung sehr viele identische DNA-Moleküle gebildet werden, kann eine an-

dere Verkettung nur durch sehr wenige DNA-Moleküle repräsentiert sein. Generell gilt, daß das deaktivierte Enzym als Verunreinigung im Tube verbleibt und daß alle Reagenzien im Tube gut durchmischt (gleichmäßig verteilt) vorliegen sollten, um die räumliche Nähe der Reaktionspartner sicherzustellen.

**2.4.2 Digestion.** Eine Reihe von Enzymen ermöglicht das gezielte Zerschneiden von DNA-Doppelsträngen. Die wichtigsten Vertreter sind die Restriktionsendonucleasen, die in Organismen die Aufgabe haben, eindringende Fremd-DNA zu zerstören. Restriktionsendonucleasen lagern sich an einer durch das jeweilige Enzym bestimmten kurzen DNA-Sequenz, der sogenannten Erkennungssequenz, an den DNA-Doppelstrang an und spalten ihn definiert innerhalb dieser Erkennungssequenz oder dicht daneben. Man kennt heute etwa 500 Restriktionsendonucleasen, von denen etwa 100 kommerziell angeboten werden. Sie sind nach ihren Erkennungssequenzen katalogisiert. Benannt werden sie durch eine Abkürzung ihres Herkunftsbakteriums, wobei man mehrere Enzyme aus demselben Bakterium durch eine fortlaufende römische Numerierung unterscheidet. Ein Beispiel ist HinPII aus dem Bakterium *Haemophilus influenzae* P<sub>1</sub> mit der Erkennungssequenz  $5' - \text{GCGC} - 3'$  und den Spaltstellen  $5' - \text{G}^{\downarrow} \text{CGC} - 3'$  und  $3' - \text{CGC} - 5'$ .

Die durch Restriktionsendonucleasen katalysierte Reaktion ist die Hydrolyse der Phosphodiesterbindungen, welche die Nucleotide zusammenhalten, an den Spaltstellen. Unter einer Hydrolyse versteht man eine Reaktion, bei der unter Zugabe von Wasser eine kovalente Bindung des Substrats aufgebrochen wird. Die Phosphatgruppen verbleiben nach dem Zerschneiden des DNA-Doppelstrangs an den entstandenen 5'-Enden. DNA-Stränge, welche die Erkennungssequenz der verwendeten Restriktionsendonuclease nicht enthalten, gehen unverändert aus der Operation hervor, siehe Abb. 7.

Bei der Ausführung im Labor werden die zu schneidenden DNA-Doppelstränge, der enzymespezifische Puffer sowie die Restriktionsendonuclease in den vom Anbieter angegebenen Endkonzentrationen in ein gemeinsames Tube gebracht, durchmischt und mit Reinstwasser verdünnt. Die weiteren Reaktionsparameter richten sich nach Herstellerangaben, üblich sind etwa 3 h Inkubation bei 37° C und eine 10minütige Enzymdeaktivierung bei 65° C.

Häufige Seiteneffekte sind eine unvollständige Reaktion, bei der ungeschnittene DNA-Doppelstränge trotz enthaltener Erkennungssequenz zurückbleiben, aber auch zusätzliche,

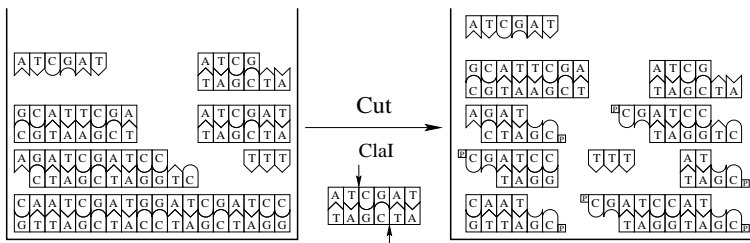


Abb. 7. Beispiel für Digestion, wobei die Restriktionsendonuclease ClaI eingesetzt wird

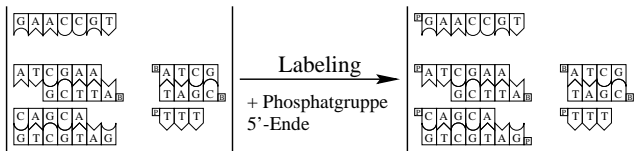


Abb. 8. Beispiel für Labeling

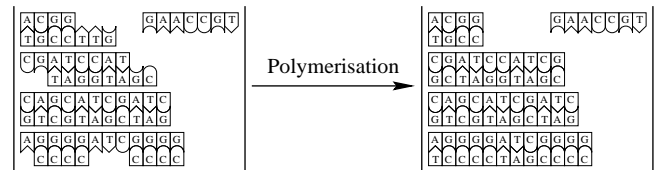


Abb. 9. Beispiel für Polymerisation

willkürliche Schnitte an nicht spezifizierten Stellen (star activity). Hinzu kommen die bereits bei der Ligation genannten Seiteneffekte für alle enzymatischen Reaktionen.

**2.4.3 Labeling.** Unter Labeling versteht man eine enzymatische Reaktion, bei der eine geeignete chemische Gruppe oder ein Molekül (Label) an bestimmten Enden von DNA-Strängen entweder angelagert oder entfernt wird, siehe Abb. 8. Das gewählte Enzym legt fest, ob an den 3'-Enden oder an den 5'-Enden operiert bzw. ob ein Anfügen oder ein Abbauen des Labels vorgenommen werden soll. Die Operation wirkt gleichermaßen auf DNA-Einzelstränge wie auf DNA-Doppelstränge. Das Anfügen geschieht dann und nur dann, wenn das betreffende Ende frei (mit einer Hydroxylgruppe versehen) ist. Das Abbauen markiert das betreffende Ende immer mit einer Hydroxylgruppe. Als Enzyme finden je nach gewünschter Wirkung verschiedene Transferasen oder Hydrolasen Verwendung.

Für den Einsatz im DNA-Computing nutzt man vorwiegend 5'-Phosphorylierung, 5'-Dephosphorylierung, 5'-Biotinylierung sowie verschiedene Fluoreszenzmarkierungen an 3'-Enden.

Mit Hilfe von 5'-Phosphorylierungen, z.B. durch das Enzym T4-Polynucleotide-Kinase, lassen sich Phosphatgruppen an freie 5'-Enden anlagern, um dadurch gezielt eine Voraussetzung für eine mögliche Ligation dieser DNA-Stränge zu schaffen. Im Gegensatz dazu gestattet die 5'-Dephosphorylierung, z.B. mittels einer alkalischen Phosphatase als Enzym, durch den Abbau der Phosphatgruppen an den 5'-Enden den Ausschluß der betreffenden DNA-Stränge von einer Ligation. Folglich läßt sich die Ligation über das Setzen bzw. Entfernen von Phosphatlabeln steuern.

Die 5'-Biotinylierung fügt an jedes freie 5'-Ende ein Biotin-Molekül an. Biotin ist ein Vitamin des Vitamin-B-Komplexes. Es hat die Eigenschaft, sich sehr leicht an Streptavidin binden zu können. Streptavidin kann als Träger-substanz für die Fixierung von biotinylierten DNA-Strängen an einer Oberfläche eingesetzt werden. Die fixierten DNA-Stränge sind dadurch im Tube nicht mehr frei beweglich, sondern mit einem Ende an die Oberfläche der Streptavidin-Kügelchen geheftet. Über diese Art des Labelns lassen sich weiterführend DNA-Stränge in getrennte Tubes separieren.

Fluoreszenzmarkierungen sind aus technischen Gründen notwendig, um die Nucleotidabfolge von DNA-Einzelsträngen effizient analytisch bestimmen zu können.

Seiteneffekte äußern sich unabhängig vom konkret gewählten Enzym vor allem in einer unvollständigen Reaktion, bei der unbehandelte Enden zurückbleiben und somit eine Menge von DNA-Strängen schaffen, welche die gewünschten Labeleigenschaften nicht besitzen. Diese DNA-Stränge können nachgeschaltete Reaktionen negativ beeinflussen. Vielfach erfordert das Labeling neben dem Enzym noch zusätzliche Reagenzien wie z.B. Adenosindiphosphat, deren Reste als weitere Verunreinigung im Tube verbleiben.

**2.4.4 Polymerisation.** Diese Operation wirkt auf DNA-Doppelstränge mit Einzelstrangüberhang (sticky-Ende) und konvertiert sticky-Enden zu blunt-Enden unter Benutzung einer DNA-Polymerase als Enzym. Konkret werden 5'-Überhänge von DNA-Doppelsträngen zu blunt-Enden aufgefüllt sowie 3'-Überhänge zu blunt-Enden abgebaut, siehe Abb. 9. Die aufzufüllenden 3'-Enden an 5'-Überhängen müssen mit Hydroxylgruppen markiert sein.

Im Zusammenhang mit der Zellteilung kommt der DNA-Polymerase eine besondere Bedeutung für die Duplikation von DNA-Doppelsträngen zu. Die Duplikation verläuft semikonservativ, d. h. der DNA-Doppelstrang wird in zwei komplementäre DNA-Einzelstränge aufgespalten und anschließend jeder dieser DNA-Einzelstränge wieder zu einem DNA-Doppelstrang vervollständigt. An diesem Vorgang sind DNA-Polymerasen maßgeblich beteiligt. Die spezielle synthetisierende Wirkung von DNA-Polymerasen läßt sich auch vorteilhaft im DNA-Computing einsetzen. So ist es beispielsweise möglich, zueinander kompatible DNA-Doppelstrangenden zu erzeugen, die sich miteinander ligieren lassen. Darüber hinaus gestattet die Polymerisation auch die Generierung von DNA-Doppelstrangabschnitten, die als Erkennungssequenzen für eine Digestion dienen können.

Der Ablauf einer Polymerisation erfordert neben der zu bearbeitenden DNA, der Puffersubstanz und der Polymerase auch die Nucleotide, die weiterführend in den DNA-Strang eingebaut werden sollen. Diese Bausteine stellt man in hinreichender Menge durch Zugabe von dNTP, einem Gemisch

aus Verbindungen zwischen Desoxyribose mit jeweils einer der vier Basen sowie aus Triphosphat, zur Verfügung.

Es gibt mehrere DNA-Polymerasen, die sich hinsichtlich ihrer Reaktionsparameter und ihrer Wahrscheinlichkeit, an jeder Position das richtige komplementäre Nucleotid einzusetzen, unterscheiden. Am zuverlässigsten arbeiten hierbei jene Polymerasen, die ihre katalytische Wirkung bei einer vergleichsweise hohen Temperatur entfalten, wie z.B. die Vent-DNA-Polymerase. Sie entstammt dem Bakterium *Thermococcus litoralis*, das in heißen Quellen unter der Meeresoberfläche lebt und gegenüber einer Umgebungstemperatur von bis zu 98°C resistent ist. Dieses Enzym hat die gewünschte Nucleotidtransferase-Aktivität zum schrittweisen Einbau von DNA-Bausteinen und besitzt gleichzeitig eine sogenannte 3'-5'-Exonuclease-Aktivität, die zum kontrollierten nucleotidweisen Abbau der 3'-Überhänge führt. Die maximale katalytische Wirkung der Vent-DNA-Polymerase liegt bei einem pH-Wert von 7,4, der vor Beginn der Reaktion über einen geeigneten Puffer eingestellt wird. Die anschließende Inkubation, während der das Enzym wirkt, erfolgt bei 72°C. Üblich ist ein Zeitraum von etwa 30 min. Vent-DNA-Polymerase kann infolge ihrer Temperatur-Stabilität nicht thermisch deaktiviert werden, das Enzym liegt somit nach Ausführung der Operation noch reaktionsbereit im Tube vor. Soll eine Polymerisation zu einem späteren Zeitpunkt ausgeschlossen werden, muß eine Reinigung des Tubeinhaltes vom Enzym erfolgen. Dazu eignet sich der Einsatz einer Gel-Elektrophorese.

Neben den Seiteneffekten, mit denen alle enzymatischen Reaktionen behaftet sind, zeigt sich bei der Polymerisation der Einbau falscher Nucleotide, die zu nicht komplementären Stellen im DNA-Doppelstrang führen. Die Fehlerate ist in Enzymbeschreibungen mit etwa 1:800 angegeben. Außer der DNA-Polymerase verbleiben auch dNTP-Reste als Verunreinigungen im Tube. Bei der Verwendung sehr kurzer DNA-Stränge besteht die Gefahr, daß sie bei der hohen Inkubationstemperatur denaturieren können.

**2.4.5 Polymerase-Kettenreaktion.** Die Polymerase-Kettenreaktion (Polymerase Chain Reaction, PCR) ist eine Technik, die es gestattet, DNA-Stränge millionenfach zu kopieren. Sie wurde 1985 von einer Forschergruppe um K. B. Mullis vorgestellt und setzt sich aus der wiederholten Anwendung einer Operationsfolge bestehend aus Melting, Annealing und Polymerisation zusammen. Die PCR gilt als effektive Methode zur exponentiellen Vervielfältigung (Amplifikation) eines DNA-Strangabschnittes, ohne ihn dafür aus dem Gesamtstrang heraustrennen zu müssen, siehe Abb. 10. Für Anwendungen in der medizinischen Diagnostik hat sich die PCR erfolgreich etabliert.

Das Anfangs- und das Endstück des zu kopierenden DNA-Strangabschnittes muß zur Durchführung einer PCR bekannt sein und in Form entsprechend kurzer DNA-Einzelstränge (Primer) vorliegen. Typische Primer haben eine Länge von etwa 20 Nucleotiden. Zwischen den beiden begrenzenden Primern kann ein DNA-Strangabschnitt beliebiger Sequenz und mit einer Länge bis zu mehreren tausend Nucleotiden liegen.

Die Grundidee der PCR besteht in der zyklischen An-

wendung der bereits genannten drei Operationen. Jeder Zyklus verdoppelt die Anzahl der Duplikate des durch die beiden Primer begrenzten DNA-Strangabschnittes. Üblich sind bis zu 40 Zyklen, mit deren Hilfe etwa  $10^{12}$  Duplikate erzeugt werden können. Danach tritt eine Sättigung ein, und die exponentielle Vervielfältigung stoppt.

Jeder Zyklus beginnt mit dem Denaturieren der DNA-Doppelstränge, die in ihre komplementären DNA-Einzelstränge aufgespalten werden. Als Richtwert für diesen Vorgang gilt eine Temperatur von 94°C, die 60 s auf das Tube einwirkt. Anschließend erfolgt die Anlagerung der Primer an die zuvor erzeugten DNA-Einzelstränge mittels Annealing. Da die Primer gegenüber dem zu duplizierenden DNA-Strang sehr kurz sind, besitzen sie eine im Vergleich niedrigere Schmelztemperatur, die gewöhnlich 60°C nicht übersteigt. Zur Primeranlagerung senkt man die Temperatur im Tube auf einen Wert etwa 2°C unterhalb der Primerschmelztemperatur und hält diese für etwa 50 s aufrecht. Dadurch ist gewährleistet, daß sich die Primer an die „langen“ DNA-Einzelstränge binden können, jedoch nicht die DNA-Einzelstränge miteinander. Im dritten Schritt erfolgt die Primerverlängerung zu kompletten DNA-Doppelsträngen durch Polymerisation. Entsprechend der Synthesegeschwindigkeit der Polymerase gelten für diesen Schritt 72°C über 40 s als Richtwert. Aus jedem zu Beginn des Zyklus vorliegenden, beidseitig primersequenzbegrenzten DNA-Doppelstrang sind nunmehr zwei identische DNA-Doppelstränge entstanden, die in einen erneuten Zyklus eingehen können.

Charakteristisches Merkmal der PCR gegenüber anderen enzymatischen Reaktionen ist der Temperatur-Zeit-Verlauf: Jeder Zyklus verwendet drei unterschiedliche Temperaturen, die in kurzer Folge hintereinander eingestellt und in jedem Zyklus wiederholt werden müssen. Zu diesem Zweck wurden spezielle Laborgeräte, sogenannte Cycler, entwickelt, die mehrere Tubes aufnehmen können und diese einem weitgehend frei programmierbaren Temperatur-Zeit-Verlauf aussetzen. Leistungsfähige integrierte Heiz- und Kühlaggregate sorgen für einen schnellen Temperaturübergang. Thermisch stabile DNA-Polymerasen, wie z.B. die bereits beschriebene Vent-DNA-Polymerase, wurden gezielt für eine Anwendung in der PCR ausgewählt, denn diese Enzyme müssen das Denaturieren der DNA-Doppelstränge überstehen, ohne an Wirkung zu verlieren.

Die PCR vereint die Seiteneffekte der Operationen, aus denen sich jeder Zyklus zusammensetzt. Zusätzlich besteht die Gefahr, daß komplementäre Primer miteinander „annealen“. Die verwendeten Primer sollten deshalb nicht – auch nicht abschnittsweise – komplementär sein. Demgegenüber strebt man jedoch eine ähnliche (gleiche) Schmelztemperatur der Primer an, um eine möglichst gleich gute Qualität ihrer Anlagerung sicherzustellen. Aufgrund der Beschreibung des Temperatur-Zeit-Verlaufes und seiner Abhängigkeit von den konkret verwendeten Primern besitzt die PCR eine Vielzahl von Reaktionsparametern, deren Optimierung eine schwierige Aufgabe darstellt. I. a. benutzt man im voraus berechnete Reaktionsparameter in erster Näherung, die dann experimentell für eine möglichst große Ausbeute von Strangkopien feinabgestimmt werden.



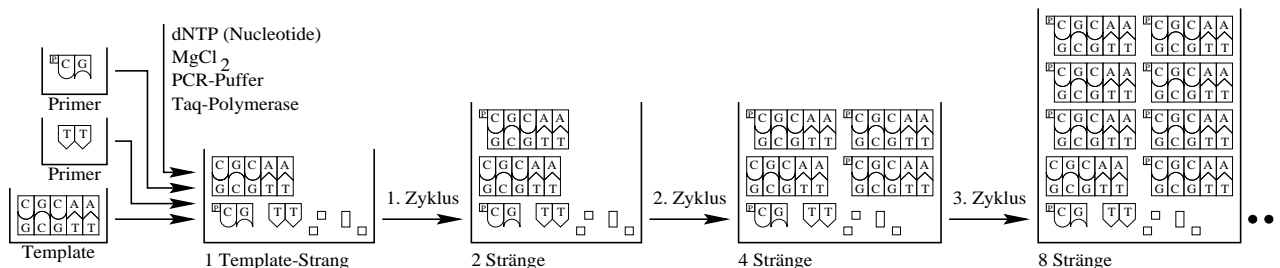


Abb. 10. Beispiel für Polymerase-Kettenreaktion

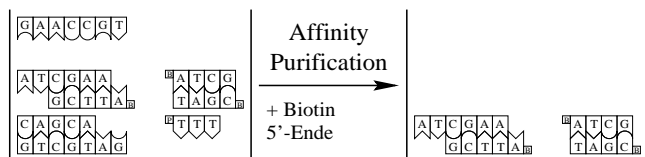


Abb. 11. Beispiel für Affinity Purification

## 2.5 Separieren und Analysieren von DNA-Strängen

Mit bloßem Auge läßt sich nicht feststellen, ob ein Tube DNA-Stränge enthält und wenn ja, aus welchen Nucleotidfolgen diese aufgebaut sind. Folglich benötigt man im DNA-Computing eine Möglichkeit zur Ergebnisausgabe. Eine solche Ausgabe kann darin bestehen, die Längen der in einem Tube enthaltenen DNA-Stränge zu bestimmen oder ihre Nucleotidsequenz zu ermitteln. Die hierzu eingesetzten Techniken werden unter dem Begriff der Analyse von DNA-Strängen zusammengefaßt. Einige dieser Techniken in Kombination mit weiteren Methoden gestatten zudem eine Separation eines DNA-Pools nach verschiedenen Merkmalen. Beispielsweise können all jene DNA-Stränge in ein gesondertes Tube extrahiert werden, die eine bestimmte Länge oder Subsequenz besitzen. Die Separations- und Analysetechniken sind im Zuge der laborpraktischen Arbeit mit organischen Molekülen entstanden und haben kein Vorbild in der lebenden Zelle.

**2.5.1 Affinity Purification.** Unter Affinity Purification versteht man eine Separationstechnik, die es ermöglicht, 5'-biotinylierte DNA-Stränge von nicht 5'-biotinylierten DNA-Strängen räumlich zu trennen und in verschiedene Tubes zu extrahieren, siehe Abb. 11. In Erweiterung läßt sich die Affinity Purification ebenfalls zur Separation nach Subsequenz nutzen, wobei von den in einem Tube befindlichen DNA-Strängen diejenigen selektiert werden, die eine bestimmte, frei wählbare Nucleotidfolge an beliebiger Position enthalten.

Die Separation nach 5'-Biotinylierung beruht auf der Eigenschaft des Biotins, sich leicht an die Oberfläche von Streptavidin-Kügelchen binden zu können, die zu Beginn in ein leeres Tube gegeben werden. Anschließend fügt man diesem Tube das zu separierende Gemisch aus 5'-biotinylierten und nicht 5'-biotinylierten DNA-Strängen hinzu. Durch die dabei erfolgende Bindung des Biotins an das Streptavidin werden die entsprechenden DNA-Stränge an ihrem Ende fixiert und sind deshalb im Tube nicht mehr frei beweglich. Die nicht 5'-biotinylierten DNA-Stränge „schwimmen“

hingegen weiterhin in der wäßrigen Lösung. Die eigentliche Separation wird durch Zentrifugieren bewerkstelligt: Die Streptavidin-Kügelchen mit den daran gehefteten DNA-Strängen setzen sich am Boden des Tubes ab, während sich die frei beweglichen DNA-Stränge im Überstand, d. h. in der oberhalb des Streptavidins konzentrierten wäßrigen Lösung, ansammeln und mittels einer Pipette vorsichtig in ein anderes Tube abgesaugt werden können. Die 5'-biotinylierten DNA-Stränge verbleiben gemeinsam mit dem Streptavidin am Tubeboden. Die Affinity Purification wird folglich durch äußere Kräfte, die auf das Tube einwirken (Zentrifugieren), gesteuert. Der gesamte Prozeß einschließlich der Biotin-Streptavidin-Anbindung verläuft bei Raumtemperatur.

Die Affinity Purification ist das Kernstück einer Operationsfolge, die eine Separation nach Subsequenz bewirkt. All jene DNA-Stränge eines Tubes, die eine frei wählbare vorgegebene DNA-Sequenz enthalten, werden herausgefiltert und in einem gesonderten Tube aufgefangen. Die Subsequenz ist durch einen (kurzen) DNA-Doppelstrang ohne Einzelstrangüberhänge und -abschnitte beschrieben. Die beiden zueinander komplementären DNA-Einzelstränge der Subsequenz werden synthetisiert, 5'-biotinyliert, in ein gemeinsames Tube zusammengeführt und dort an Streptavidin gebunden. Auf diese Weise ist eine Matrix aus fixierten DNA-Einzelsträngen, vergleichbar mit „Fangarmen“, entstanden, an die sich andere DNA-Einzelstränge aus dem zu separierenden DNA-Pool „annealen“ können. Zu diesem Zweck denaturiert man den zu separierenden DNA-Pool und gibt die entstandenen DNA-Einzelstränge in das Tube, das die Matrix enthält. Dort wird durch langsames Abkühlen ein Annealing-Vorgang ausgelöst, durch den genau die DNA-Einzelstränge aus dem Pool an die Matrix andocken, welche die Subsequenz enthalten. Jene DNA-Stränge, die die Subsequenz nicht enthalten, bleiben frei beweglich. Sie lassen sich gesondert absaugen und können entweder in einem anderen Tube aufgefangen oder verworfen werden. Um die DNA-Stränge mit der Subsequenz von der Matrix abzulösen, nutzt man die Operation Melting. Die dadurch wieder frei beweglichen DNA-Einzelstränge mit der Subsequenz werden nunmehr ebenfalls abgesaugt, in einem gesonderten Tube aufgefangen und abschließend wieder „annealt“.

Der bei der Affinity Purification dominierende Seiteneffekt besteht in einer unscharfen Separierung, d. h., in jedem Ergebnistube befindet sich ein Anteil unerwünschter DNA-Stränge, der das jeweilige Selektionsmerkmal (z.B. 5'-Biotinylierung, Subsequenz) nicht besitzt. Die „Perfektion“ des Absaugens, den Überstand exakt, vollständig und ohne Verwischung der Trennschicht zum Streptavidin auf-

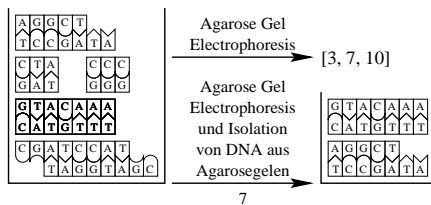


Abb. 12. Beispiel für Gel-Elektrophorese

zunehmen, bestimmt maßgeblich die Separierschärfe. Dieser Seiteneffekt zieht in Konsequenz einen Verlust erwünschter DNA-Stränge nach sich und verursacht folglich eine Verringerung ihrer Konzentration. Reaktionsparameter nachfolgender DNA-Operationen müssen auf diese Konzentration abgestimmt werden. Bei der Separation nach Subsequenz summieren sich die Seiteneffekte der Affinity Purification mit den Seiteneffekten der Operationen Synthesis, Labeling, Union, Annealing und Melting.

**2.5.2 Gel-Elektrophorese.** Die Gel-Elektrophorese ist ein physikalischer Prozeß, der die Längenbestimmung von DNA-Strängen ermöglicht sowie eine Längenseparation bewirkt, welche die DNA-Stränge eines Tubes entsprechend ihrer Nucleotidanzahl räumlich voneinander trennt. In Kombination mit einer weiteren Labortechnik lassen sich aus einem DNA-Pool sogar genau jene Stränge extrahieren, die eine bestimmte vorgegebene Länge besitzen, siehe Abb. 12. Darüber hinaus benutzt man Gel-Elektrophoresen auch dazu, DNA von Fremdstoffen im Tube (wie z.B. von Enzymresten) zu reinigen.

Das Wirkprinzip aller Gel-Elektrophoresen beruht auf der Elektrolyse: Elektrisch negativ geladene Moleküle (wie DNA) wandern im elektrischen Feld innerhalb eines geeigneten Trägermediums zur positiven Elektrode. Die Wanderungsgeschwindigkeit ist maßgeblich von der Molekülgröße und -gestalt abhängig, im Falle von DNA folglich von der Stranglänge. Als Medium, durch das sich die DNA bewegen kann, sind Gele am besten geeignet. Ihre feste Struktur gleicht einem dreidimensionalen Netzwerk aus miteinander verbundenen Fasern, zwischen denen sich Poren befinden. Die Porengröße ist ein Maß für die Trennschärfe eines Gels und derart bemessen, daß sich die DNA-Stränge hindurch bewegen können. Das Gel sollte derart beschaffen sein, daß es nicht mit der DNA reagiert, leicht herstellbar ist und idealerweise eine Rückgewinnung der DNA zuläßt. In der Laborpraxis finden Agarose- und Polyacrylamidgele Verwendung. Die Gel-Elektrophoreseapparatur besteht aus einer Gelkammer, die das vollständig in eine Pufferlösung eingebettete Gel aufnimmt, den beiden Elektroden und einem Spannungsversorgungsgerät. Zur Durchführung der Gel-Elektrophorese gibt man die DNA-Probe in eine Tasche (Ausparung, Slot) des Gels nahe der negativen Elektrode und erzeugt durch Anlegen einer Gleichspannung ein elektrisches Feld, das die Elektrolyse auslöst. Je kürzer ein DNA-Strang ist, desto schneller wandert er durch das Gel. Die kürzesten DNA-Stränge in der Probe legen somit die größte Entfernung im Gel zurück und bestimmen den Zeitpunkt, an dem die Elektrolyse beendet wird, um ein Austreten dieser DNA-Stränge an der gegenüberliegenden Gelseite zu verhindern. Nach Abschalten der Spannung liegen die DNA-Stränge entsprechend

ihrer Länge räumlich verteilt im Gel vor. Um sie sichtbar zu machen und mit Hilfe einer Skala ihre Längen abschätzen zu können, wird die DNA vor dem Einfüllen in die Geltasche mit einem speziellen Farbstoff versetzt, der sich an die DNA bindet und unter ultraviolettem Licht fluoresziert. Zur Auswertung betrachtet man das Gel bei Bestrahlung in diesem Wellenlängenbereich. Die Stellen, an denen sich die mit dem Farbstoff angereicherte DNA befindet, zeichnen sich dabei deutlich in Form sogenannter Banden ab. In jeder Bande sind DNA-Stränge gleicher Länge angereichert. Banden lassen sich mittels eines Skalpell aus dem Gel ausschneiden, und die darin enthaltene DNA kann durch nachfolgende Anwendung entsprechender Methoden aus dem Gel isoliert (rückgewonnen) werden. Auf diese Weise ist neben der Längenbestimmung auch eine Längenseparation möglich.

Die Agarosegel-Elektrophorese hat sich für DNA-Doppelstränge in der Praxis durchgesetzt. Das i.a. nicht denaturierende Gel besteht dabei aus Agarose, dessen Grundsubstanz Agar aus Meeresalgen stammt. Über die Agarosekonzentration kann die Porengröße und in Konsequenz das Auflösungsvermögen variiert werden. Die maximal erreichbare Auflösung liegt bei etwa 10 Nucleotidpaaren, d. h., Stranglängendifferenzen von 10 Nucleotiden sind gerade noch unterscheidbar, kleinere Längendifferenzen verschmelzen zu einer gemeinsamen Bande. Agarosegele können eine DNA-Laufweite bis etwa 30 cm aufweisen und werden meist Spannungen zwischen 80 V und 100 V ausgesetzt. Der Elektrophorese-Vorgang dauert zwischen 30 min und mehreren Stunden. Als Farbstoff für die Fluoreszenzmarkierung dient i.a. Ethidiumbromid, das sich an die Wasserstoffbrücken doppelsträngiger DNA bindet. Zur Isolation von DNA aus Agarosegelen sind verschiedene Methoden bekannt, welche die unterschiedliche Dichte von DNA und Agarosegele ausnutzen und durch Gelverflüssigung mittels Erwärmen (55° C) sowie anschließendes Zentrifugieren eine räumliche Trennung von Gel und DNA bewirken. Der Rückgewinnungsgrad der DNA kann bis ca. 80% betragen.

Hochauflösende Polyacrylamidgele sind feinporiger als Agarosegele und ermöglichen eine Auftrennung von DNA-Strängen, deren Längenunterschied nur ein einziges Nucleotid beträgt. Dadurch kann die Längenbestimmung unter geeigneten Bedingungen exakt vorgenommen werden. Die Polyacrylamidgel-Elektrophorese ist wesentlicher Bestandteil des Arbeitsablaufes zur Sequenzierung von DNA, bei der die konkrete Nucleotidabfolge bestimmt wird. Die mit der Polyacrylamidgel-Elektrophorese einhergehende starke Erwärmung bewirkt eine Denaturierung, so daß immer die Länge der dem DNA-Pool zugrundeliegenden DNA-Einzelstränge gemessen wird. Eine Rückgewinnung von DNA aus Polyacrylamidgelen läßt sich i.a. nicht realisieren.

Die Gel-Elektrophorese ist mit einer Reihe von Seiteneffekten behaftet: Cluster und Luftblasen im Gel können zu unscharfen und verwischten Banden führen. Derselbe Effekt ergibt sich auch infolge einer unscharfen Separation, hervorgerufen durch die begrenzte Trennleistung jedes Gels. Der verwendete Farbstoff impliziert eine Sichtbarkeitsgrenze, d. h., Banden mit zu geringer DNA-Konzentration sind nicht mehr wahrnehmbar. Die Reibung der DNA bei der Bewegung durch das Gel führt zur Erwärmung und kann auch bei Agarosegelen eine Denaturierung der DNA nach sich ziehen. Ringförmige und überspiralisierte DNA zeigt

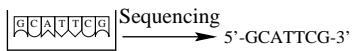


Abb. 13. Beispiel für Sequenzieren

gegenüber linearer DNA ein anderes Laufverhalten, das die Auswertung des Bandenmusters erschwert.

**2.5.3 Sequenzieren.** Die Bestimmung der konkreten Nucleotidabfolge eines DNA-Einzelstranges wird als Sequenzieren bezeichnet, siehe Abb. 13. Seit 1975 sind hierzu die Methode nach Maxam/Gilbert sowie die heute dominierende Methode nach Sanger bekannt. Letztere wird auch als Didesoxy- oder Kettenabbruch-Methode bezeichnet und ist nachfolgend beschrieben.

DNA-Sequenzierung nach Sanger läuft nahezu vollautomatisiert ab. Entsprechende Sequenziergeräte erreichen einen Durchsatz von 80 Nucleotiden pro Stunde und können bis etwa 800 Nucleotide umfassende DNA-Einzelstränge verarbeiten. Der Eingangstubeinhalt sollte möglichst ausschließlich aus identischen DNA-Strängen bestehen. DNA-Doppelstränge müssen vor dem Sequenzieren denaturiert werden. Die Methode nach Sanger stellt eine Kombination aus einer Polymerisation und einer Polyacrylamidgel-Elektrophorese dar.

Das Anfangsstück (etwa 20 Nucleotide) des zu sequenzierenden DNA-Einzelstranges muß bekannt sein. Daran wird zu Beginn ein Primer angelagert. Ein Primer ist ein kurzer, zum Anfangsstück komplementärer DNA-Einzelstrang. Die bei der Primeranlagerung entstandenen DNA-Stränge verteilt man gleichmäßig in vier Tubes. Dort erfolgt anschließend parallel eine modifizierte Polymerisation, zu der außer den Nucleotidbausteinen (dNTP) auch in geringerem Verhältnis sogenannte Kettenabbruch-Nucleotide ( $A^*$ ,  $C^*$ ,  $G^*$ ,  $T^*$ , sogenannte Didesoxynucleotide) gegeben werden, tubeweise getrennt nach den in ihnen enthaltenen Basen. Beispielsweise erhält Tube 1 nur Kettenabbruch-Nucleotide mit der Base Adenin ( $A^*$ ), Tube 2 nur solche mit Cytosin ( $C^*$ ), Tube 3 nur solche mit Guanin ( $G^*$ ) und Tube 4 nur solche mit Thymin ( $T^*$ ). dNTP ist ein Gemisch aus den vier DNA-Grundbausteinen, die zur Komplettierung der DNA-Doppelstränge während der Polymerisation nucleotidweise eingebaut werden. Die Kettenabbruch-Nucleotide besitzen zusätzlich die Eigenschaft, daß dahinter kein weiteres Nucleotid mehr an den Strang angefügt werden kann. Bei der Polymerisation wird in jedem der vier Tubes anstelle eines Nucleotids randomisiert auch das entsprechende Kettenabbruch-Nucleotid eingesetzt, das die Beendigung der Strangverlängerung bewirkt. Im Ergebnis liegen dann DNA-Doppelstränge aller nucleotidweisen Komplettierungsstufen, tubeweise getrennt nach der letzten Base, vor. Beispielsweise enthält Tube 1 all jene DNA-Doppelstränge, deren Primerverlängerung mit  $A^*$  endet. Diese Primerverlängerungen existieren in sehr vielen unterschiedlichen Längen. Die Anzahl unterschiedlicher Längen ist mit der Anzahl der Vorkommen des Nucleotids A in der Primerverlängerung identisch. Analog dazu befinden sich in den Tubes 2, 3 und 4 all jene Primerverlängerungen, die jeweils mit  $C^*$ ,  $G^*$  und  $T^*$  enden. Die vier Tubeinhalte werden anschließend in einem hochauflösenden Polyacrylamidgel in vier parallel verlaufenden Bahnen denaturiert und elektrophoretisch

aufgetrennt. Stranglängendifferenzen von einem Nucleotid können in diesem Gel unterschieden werden. Das resultierende Bandenmuster gibt Aufschluß über die Nucleotidfolge der Primerverlängerung: Der kürzeste DNA-Einzelstrang bestimmt das erste Nucleotid, und je nachdem, in welcher der vier Bahnen die Bande des um ein Nucleotid längeren DNA-Einzelstranges existiert, kann das Nachfolgenucleotid zugeordnet werden, bis die Primerverlängerung vollständig ausgelesen ist. Mit der Primerverlängerung ist auch die Nucleotidfolge des komplementären zu sequenzierenden DNA-Einzelstranges bekannt.

Das Sequenzieren vereint die Seiteneffekte der Gel-Elektrophorese mit den Seiteneffekten der Operationen Union, Melting, Annealing und Polymerisation. Darüber hinaus kann auch der verwendete, zumeist synthetisierte Primer Mutationen aufweisen oder in nicht ausreichender Menge bereitstehen. Bei der Zuordnung der Nucleotide aus dem Bandenmuster können Mehrdeutigkeiten auftreten, die sich allein durch die Gelanalyse nicht bereinigen lassen und eine Wiederholung der gesamten Sequenzierung notwendig machen. Das eingesetzte DNA-Material steht nach der Sequenzierung nicht mehr zur Verfügung, denn es kann aus dem Polyacrylamidgel nicht mehr zurückgewonnen werden. Trotz der Seiteneffekte gilt die Sequenzierung als weitgehend zuverlässig, wobei i.a. mehrere Sequenzierläufe von derselben DNA-Probe durchgeführt werden.

### 3 Modellierung der DNA-Operationen

Der Modellierungsaspekt von DNA-Operationen reflektiert die wesentlichen Entwicklungsetappen auf dem Forschungsfeld des DNA-Computing. Zunächst nutzte man die Molekularbiologie, um ausgewählte Vertreter der Klasse der NP-Probleme experimentell zeiteffizient zu lösen. Ein Repräsentant dieser Klasse ist das Integer-Rucksackproblem, definiert im Abschnitt 4. Motiviert durch die Pionierfolge im Labor entstand der Wunsch, formale Berechnungsmodelle zum DNA-Computing zu schaffen. Speziell in Dresden gab das DNA-Experiment zum Integer-Rucksackproblem den Impuls zur Entwicklung eines Modells, in dessen Zentrum die molekularbiologischen Operationen aus Abschnitt 2 stehen. Eine spezifische Auswahl aus diesem Spektrum löst genau das Integer-Rucksackproblem. Mit den genannten DNA-Grundoperationen kann man ein Berechnungsmodell aufbauen, welches sogar universell ist, d. h. es vermag auch jedes berechenbare Problem zu lösen. Infolge der direkten Entsprechung zwischen Modelloperationen und molekularbiologischen Prozessen ist ein solches DNA-Computing-Modell besonders flexibel. Es zeichnet sich vor allem durch folgende Merkmale aus:

- Andere bereits bekannte DNA-Computing-Modelle sowie konventionelle Berechnungsmodelle können vollständig mit Hilfe dieses Modells labornah abgebildet werden [25].
- Die bisher durchgeführten Experimente im DNA-Computing lassen sich in diesem Modell simulieren. Der Fakt, daß die zugrundeliegenden DNA-Operationen Bestandteile des Modells sind, stützt diese Aussage.

- Mit Hilfe des Modells ist es möglich, die Konstruktion von DNA-Algorithmen kompakt und effizient, sowie ihre Notation transparent vorzunehmen.
- Das Modell ist auf eine unmittelbare Implementierung im Labor ausgerichtet, es existiert kein Freiraum zwischen Modelloperation und Labortechnik. Damit ist der grundlegende Anspruch an einen universellen Bio-Computer erfüllt.

Neben den molekularbiologischen Prozessen als Operationen und der DNA als Basisdatenstruktur benötigt ein DNA-Computing-Modell eine geeignete Kontrollstruktur. Prinzipiell unterscheidet man imperative und deklarative Programmiersprachen. In einer imperativen Sprache besteht das Programm aus einer sequentiellen Operationsfolge an den Rechner, die sich aus Steueranweisungen ergibt. Im Gegensatz dazu beschreibt ein Programm, notiert in einer deklarativen Sprache, die zu lösende Problemstellung ohne Vorgabe konkreter Steueranweisungen. Programmieren heißt in diesem Zusammenhang anzugeben, *was* und nicht *wie* gelöst werden soll. Innerhalb der deklarativen Programmiersprachen haben sich funktionale Programmiersprachen erfolgreich etabliert. Ein bekannter Vertreter ist HASKELL<sup>1</sup>. Die Motivation, sich für eine funktionale Sprache zu entscheiden, ist damit begründet, daß das Abstraktionsniveau eines funktionalen Programms nahe der Ebene von reinen Problemspezifikationen liegt und damit einer mathematischen Behandlung einfacher zugänglich ist. Nur mit dieser Eigenschaft sind Korrektheitsbeweise, Verifikation und Analyse von Programmeigenschaften uneingeschränkt möglich. DNA-Computing als modernes Berechnungsmodell kann besonders dann überzeugen, wenn eine formale Beschreibung der tatsächlichen laborpraktischen Umsetzung vorgenommen werden kann.

Es ist naheliegend, ein DNA-Computing-Modell, dessen Basis in der Programmiersprache HASKELL liegt, als DNA-HASKELL zu bezeichnen. Nachfolgend sollen wichtige Aspekte aus der Konzeption von DNA-HASKELL kurz umrissen werden.

Die DNA muß mit allen für die Ausführung der Modelloperationen relevanten Bestandteilen und Eigenschaften in DNA-HASKELL abgebildet werden. Die Modellierung verwendet ein Abstraktionsniveau, auf dem jedes einzelne Nucleotid und Labelmolekül als Grundbaustein aufgefaßt wird. Dies gestattet eine sehr labornah, sequenz- und labelgetreue Abbildung der zu verarbeitenden DNA-Stränge, und die Wirkung der Modelloperationen kann bis auf die Ebene einzelner Nucleotide und Labelmoleküle spezifiziert werden, wodurch zahlreiche Eigenschaften der laborpraktischen DNA-Operationen in DNA-HASKELL einfließen.

In DNA-HASKELL wird jeder DNA-Strang als linearisierte endliche Sequenz von Nucleotiden und/oder Nucleotidpaaren dargestellt, deren Enden mit jeweils einem Labelmolekül versehen sind. Formen nichtlinearer DNA, wie beispielsweise Plasmidringe, müssen geeignet linearisiert werden.

Unter einem Tube versteht man in DNA-HASKELL eine endliche Menge unterschiedlicher DNA-Stränge. Im Trivialfall kann ein Tube leer sein, d. h., keine DNA-Stränge enthal-

ten. Bei der laborpraktischen Ausführung des DNA-Computing liegen im Tube zahlreiche Duplikate identischer DNA-Stränge vor, wodurch eine hohe Redundanz der DNA-Daten existiert. DNA-HASKELL abstrahiert von dieser Redundanz und stellt mehrere identische DNA-Stränge im gleichen Tube durch genau einen Repräsentanten dar (nichtrestriktives Modell). Zwei DNA-Stränge sind genau dann identisch, wenn sie durch Verschiebung oder Drehung im Raum um beliebige Achsen zur Deckung gebracht werden können, d. h., in richtungsbehafteter Nucleotidfolge und allen Labeln positionsgenau übereinstimmen.

Zusätzlich zu den genannten benutzerdefinierten DNA-Datentypen benötigt man auch Datentypen für ganze Zahlen, für Zeichenkonstanten sowie daraus gebildete Listen. Elemente dieser Datentypen werden als Parameter bestimmter DNA-Operationen benötigt und ermöglichen Ein- und Ausgaben außerhalb der DNA-Datentypen. Dadurch können die Kodierung von Eingangsdaten einer Problemstellung in DNA-Stränge sowie die Dekodierung des Ergebnisses in DNA-HASKELL selbst vorgenommen werden.

Die verwendeten Datenstrukturen, die spezifizierten Operationen sowie die zugrundeliegende Kontrollstruktur beschreiben jedes DNA-Computing-Modell vollständig. Seine Implementierung in einer Programmiersprache stellt eine formale Notation dar. Insbesondere DNA-HASKELL besitzt den Vorteil, daß eine Abarbeitung entsprechender DNA-Algorithmen sowohl im Labor auf Bio-Hardware als auch in silicio möglich ist. Unter Benutzung eines labornahen DNA-Computing-Modells können Laborexperimente kostengünstig im Vorfeld geplant und simuliert werden, um eine optimale Operationsfolge für die Lösung einer algorithmischen Aufgabenstellung zu finden. Die formale Notation eines Modells ist Voraussetzung, um das DNA-Computing als unkonventionelles Computing-Konzept auf ein sauberes theoretisches Fundament zu stellen. Die in-silicio Simulation eines DNA-Computing-Modells gestattet eine Visualisierung von Rechenvorgängen im Tube, so daß die fortlaufende Modifikation von Tubeinhalten im Detail nachvollziehbar ist. Neben diesen pragmatischen Aspekten sind insbesondere komplexitätstheoretische Betrachtungen der DNA-Computing-Modelle sinnvoll. Auf diese Weise können verschiedene Computing-Konzepte hinsichtlich ihres Ressourcenverbrauches gegenübergestellt werden. Eine wichtige Eigenschaft des DNA-Computing besteht beispielsweise darin, daß infolge der massiven Datenparallelität der exponentielle Aufwand zur Lösung von NP-Problemen von der Zeit zum Speicherplatz verschoben wird. Die Operationen Annealing und Ligation bewirken maßgeblich eine Steigerung der Zeiteffizienz von DNA-Algorithmen bei laborpraktischer Abarbeitung. Diese beiden Operationen zeichnen sich dadurch aus, daß bei ihrer Ausführung fortlaufende Interaktionen zwischen verschiedenen DNA-Strängen des Eingangstubes auftreten können und berücksichtigt werden müssen. Annealing und/oder Ligation sind auch genau diejenigen Operationen, die zur Generierung einer kombinatorischen Vielfalt von DNA-Strängen herangezogen werden. Operationen, die keine Interaktionen zwischen verschiedenen DNA-Strängen des Eingangstubes hervorrufen, besitzen eine polynomielle Zeitkomplexität.

<sup>1</sup> Compiler für HASKELL wurden in Yale, Glasgow und Göteborg entwickelt und sind frei verfügbar (<http://www.haskell.org>).

## 4 Eine experimentelle Lösung des Integer-Rucksackproblems

Eine experimentelle Lösung des Integer-Rucksackproblems erfordert über die DNA-Algorithmen von Adleman und Lipton hinaus die Abbildung natürlicher Zahlen in DNA-Strängen sowie deren Addition und Vergleich auf Gleichheit. Deshalb unterscheidet sich der DNA-Algorithmus des Integer-Rucksackproblems grundlegend im Vorgehen von Adleman und Lipton. DNA-Stränge kodieren Informationen über ihre Sequenz *und* ihre Länge, der Aufbau der kombinatorischen Vielfalt wird schrittweise vollzogen, und bei der Selektion der Lösung wird die Fähigkeit der Gel-Elektrophorese zur Längenseparation ausgenutzt.

Das Integer-Rucksackproblem gehört zur Klasse der NP-Probleme. Gegeben sind  $n$  Gegenstände mit den Gewichten  $a_i \in \mathcal{N}$ ,  $1 \leq i \leq n$  sowie eine Zahl  $b \in \mathcal{N}$ . Das Integer-Rucksackproblem besteht in der Entscheidung, ob es eine Teilmenge  $I \subseteq \{1, 2, \dots, n\}$  gibt, so daß  $\sum_{i \in I} a_i = b$

[23]. Der Parameter  $n$  fungiert als Problemgröße. Mit  $n$  Gegenständen existieren  $2^n$  Packmöglichkeiten des Rucksacks.

Zur Implementierung im Labor wurde das Integer-Rucksackproblem mit drei Gegenständen, den Gewichten  $a_1 = 719$ ,  $a_2 = 393$ ,  $a_3 = 270$  und dem angestrebten Gesamtgewicht des Rucksackinhaltes  $b = 1112$  ausgewählt.

### 4.1 Algorithmische Idee für beliebig viele Gegenstände

Jedes Gegenstandsgewicht wird durch einen DNA-Doppelstrang spezifischer Länge kodiert. Zwischen jedem Gegenstandsgewicht  $a_i$  und der Länge des korrespondierenden DNA-Doppelstranges in Basenpaaren (bp) besteht dabei ein linearer Zusammenhang. Mithilfe der Ligation lassen sich DNA-Doppelstränge miteinander verketten, wobei die Phosphorylierung eine gezielte Steuerung der Ligation ermöglicht. Auf diese Weise erzeugt man eine kombinatorische Vielfalt an DNA-Strängen, deren Längen die Rucksackgewichte aller Packmöglichkeiten verkörpern. Nachdem für alle Rucksackbelegungen die kodierenden DNA-Doppelstränge erzeugt wurden, muß nur noch entschieden werden, ob es eine Rucksackbelegung mit dem Gesamtgewicht  $b$  gibt. Basierend auf dieser Idee ist ein DNA-Algorithmus wie folgt konstruierbar:

1. Stelle für jedes Gegenstandsgewicht  $a_i$  mit  $i = 1, \dots, n$  spezifische DNA-Stränge in gesonderten Tubes bereit.
2. Erzeuge alle Packmöglichkeiten aus den Gegenständen 1 und 2.
3. Nimm Gegenstand 3 hinzu und verdopple dadurch die Anzahl der Packmöglichkeiten.
4. Nimm Gegenstand 4 hinzu und verdopple dadurch die Anzahl der Packmöglichkeiten.
- ⋮ (weitere Schleifendurchläufe)
- n. Nimm Gegenstand  $n$  hinzu und verdopple dadurch die Anzahl der Packmöglichkeiten
- (n+1). Prüfe, ob es eine Packmöglichkeit mit dem Gesamtgewicht  $b$  gibt. Falls ja, lautet die Lösung „ja“, anderenfalls „nein“.

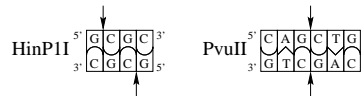


Abb. 14. Schnittstellen der Restriktionsenzyme PvuII und HinP1I

### 4.2 Kodierung der Eingangsgrößen in DNA-Stränge

Es ist üblich, spezifische DNA-Doppelstränge durch Synthese komplementärer DNA-Einzelstränge, die anschließend „annealt“ werden, zu generieren. Bei der Verwendung kurzer DNA-Stränge ( $\leq 100$  Basenpaare) ist dieses Vorgehen auch laborpraktisch realisierbar. Die DNA-Einzelstränge der im Beispiel benötigten Längen lassen sich jedoch nicht mehr zuverlässig synthetisieren, so daß sie zur Implementierung des Algorithmus im Labor auf andere Weise gewonnen werden müssen.

Mikroorganismen – wie z.B. Bakterien – besitzen ringförmige DNA im Zellplasma, die zusätzlich zum Erbmateriale vorkommt und leicht isoliert werden kann. Solche ringförmige DNA bezeichnet man als Plasmid. Bakterien können ein bis viele hundert Plasmide enthalten. Zur Verwendung im Experiment wurde das Plasmid pQE30 ausgewählt, ein doppelsträngiger DNA-Ring bekannter Sequenz, der aus 3462 Basenpaaren besteht, siehe Abb. 15. Die Auswahl erfolgte nach dem Kriterium, mit möglichst wenigen Restriktionsenzymen signifikante Fragmentlängen und -enden zu erzeugen.

Mittels der Restriktionsenzyme PvuII und HinP1I ist das Plasmid derart spaltbar, daß genau die DNA-Fragmente der gewünschten Längen entstehen. PvuII gehört zu den blunt-end-Cuttern, d. h., beim Schnitt des DNA-Doppelstranges entstehen keine Einzelstrangüberhänge an den Schnittstellen, siehe Abb. 14. Die Schnittsequenz von PvuII kommt in pQE30 genau einmal vor. HinP1I ist ein sticky-end-Cutter. Dies bedeutet, daß an den Schnittstellen Einzelstrangüberhänge entstehen. Die Schnittsequenz von HinP1I ist in pQE30 mehrfach vorhanden. Das Verketten von kompatiblen sticky-Enden durch Ligation verläuft laborpraktisch effektiver als das Verketten von blunt-Enden. Die Verbindungsstellen der DNA-Doppelstränge für die Gegenstandsgewichte zum Aufbau der kombinatorischen Vielfalt sollten deshalb sticky-Enden besitzen.

Durch den Schnitt mit HinP1I und PvuII zerfällt das pQE30 in DNA-Fragmente der Längen 719, 561, 393, 337, 332, 270, 174, 141, 109, 103, 100, 93, 67, 33 und 30 Basenpaare. Die 719 bp- und die 561 bp-Fragmente besitzen jeweils ein blunt-Ende und ein sticky-Ende, alle anderen Fragmente beidseitig sticky-Enden.

Zur laborpraktischen Ausführung wurden das Plasmid pQE30 (QIAGEN) und die Restriktionsenzyme HinP1I sowie PvuII (New England BioLabs) verwendet. Die Reagenzien wurden unter geeigneten Pufferbedingungen für 4 h bei  $+37^\circ\text{C}$  inkubiert und die Enzyme anschließend thermisch deaktiviert.

Die durch den Schnitt entstandenen DNA-Enden sind 5'-phosphoryliert. Diese Phosphatgruppen müssen mittels Dephosphorylieren von den 5'-Enden abgebaut werden. Die Schnittfragmente und alkalische Phosphatase CIP (Boehringer Mannheim) wurden unter geeigneten Pufferbedingungen für 3 h bei  $+37^\circ\text{C}$  inkubiert und die DNA-Fragmente unter

CTC GAG AAA TCA TAA AAA ATT TAT TTG CTT TGT GAG CGG ATA ACA ATT ATA ATA GAT TCA	60
ATT GTG AGC GGA TAA CAA TTT CAC ACA GAA TTC ATT AAA GAG GAG AAA TTA ACT ATG AGA	120
GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC GCA TGC GAG CTC GGT ACC CCG GGT CGA CCT	180
GCA GCC AAG CTT AAT TAG CTG AGC TTG GAC TCC TGT TGA TAG ATC CAG TAA TGA CCT CAG	240
AAC TCC ATC TGG ATT TGT TCA GAA CGC TCG GTT GCC GCC GGG CGT TTT TTA TTG GTG AGA	300
ATC CAA GCT AGC TTG GCG AGA TTT TCA GGA GCT AAG GAA GCT AAA ATG GAG AAA AAA ATC	360
ACT GGA TAT ACC ACC GTT GAT ATA TCC CAA TGG CAT CGT AAA GAA CAT TTT GAG GCA TTT	420
CAG TCA GTT GCT CAA TGT ACC TAT AAC CAG ACC GTT CAG	
CTG GAT ATT ACG GCC TTT TTA	480
AAG ACC GTA AAG AAA AAT AAG CAC AAG TTT TAT CCG GCC TTT ATT CAC ATT CTT GCC CGC	540
CTG ATG AAT GCT CAT CCG GAA TTT CGT ATG GCA ATG AAA GAC GGT GAG CTG GTG ATA TGG	600
GAT AGT GTT CAC CCT TGT TAC ACC GTT TTC CAT GAG CAA ACT GAA ACG TTT TCA TCG CTC	660
TGG AGT GAA TAC CAC GAC GAT TTC CCG CAG TTT CTA CAC ATA TAT TCG CAA GAT GTG GCG	720
TGT TAC GGT GAA AAC CTG GCC TAT TTC CCT AAA GGG TTT ATT GAG AAT ATG TTT TTC GTC	780
TCA GCC AAT CCC TGG GTG AGT TTC ACC AGT TTT GAT TTA AAC GTG GCC AAT ATG GAC AAC	840
TTC TTC GCC CCC GTT TTC ACC ATG GGC AAA TAT TAT ACG CAA GGC GAC AAG GTG CTG ATG	900
CCG CTG GCG ATT CAG GTT CAT CAT GCC GTC TGT GAT GGC TTC CAT GTC GGC AGA ATG CTT	960
AAT GAA TTA CAA CAG TAC TGC GAT GAG TGG CAG GGC GGG GCG TAA TTT TTT TAA GGC AGT	1020
TAT TGG TGC CCT TAA ACG CCT GGG GTA ATG ACT CTC TAG CTT GAG GCA TCA AAT AAA ACG	1080
AAA GGC TCA GTC GAA AGA CTG GGC CTT TCG TTT TAT CTG TTG TTT GTC GGT GAA CGC TCT	1140
CCT GAG TAG GAC AAA TCC GCC GCT CTA GAG CTG CCT CG	
C GCG TTT CCG TGA TGA CCG TGA	1200
AAA CCT CTG ACA CAT GCA GCT CCC GGA GAC GGT CAC AGC TTG TCT GTA AGC GGA TGC CCG	1260
GAG CAG ACA AGC CCG TCA GGG	
CGC GTC AGC GGG TGT TGG CCG GTG TCG GGG	
CGC AGC CAT	1320
GAC CCA GTC ACG TAG CGA TAG CGG AGT GTA TAC TGG CTT AAC TAT GCG GCA TCA GAG CAG	1380
ATT GTA CTG AGA GTG CAC CAT ATG CCG TGT GAA ATA CCG CAC AGA TGC GTA AGG AGA AAA	1440
TAC CGC ATC AGG	
CGC TCT TCC GCT TCC TCG CTC ACT GAC TCG CTG	
CGC TCG GTC TGT CGG	1500
CTG CGG CGA GCG GTA TCA GCT CAC TCA AAG GCG GTA ATA CCG TTA TCC ACA GAA TCA GGG	1560
GAT AAC GCA GGA AAG AAC ATG TGA GCA AAA GGC CAG CAA AAG GCC AGG AAC CGT AAA AAG	1620
GCC GCG TTG CTG GCG TTT TTC CAT AGG CTC CGC CCT GAC GAG CAT CAC AAA AAT CGA	1680
CGC TCA AGT CAG AGG TGG CGA AAC CCG ACA GGA CTA TAA AGA TAC CAG GCG TTT CCC CCT	1740
GGA AGC TCC CTC GTG	
CGC TCT CCT GTT CCG ACC CTG CCG CTT ACC GGA TAC CTG TCC GCC	1800
TTT CTC CCT TCG GGA AGC GTG G	
CG CTT TCT CAA TGC TCA CGC TGT AGG TAT CTC AGT TCG	1860
GTG TAG GTC GTT CGC TCC AAG CTG GGC TGT GTG CAC GAA CCC CCC GTT CAG CCC GAC CGC	1920
TG	
C GCC TTA TCC GGT AAC TAT CGT CTT GAG TCC AAC CCG GTA AGA CAC GAC TTA TCG CCA	1980
CTG GCA GCA GCC ACT GGT AAC AGG ATT AGC AGA GCG AGG TAT GTA GGC GGT GCT ACA GAG	2040
TTC TTG AAG TGG TGG CCT AAC TAC GGC TAC ACT AGA AGG ACA GTA TTT GGT ATC TG	
C GCT	2100
CTG CTG AAG CCA GTT ACC TTC GGA AAA AGA GTT GGT AGC TCT TGA TCC GGC AAA CAA ACC	2160
ACC GCT GGT AGC GGT GGT TTT TTT GTT TGC AAG CAG CAG ATT ACG	
CGC AGA AAA AAA GGA	2220
TCT CAA GAA GAT CCT TTG ATC TTT TCT ACG GGG TCT GAC GCT CAG TGG AAC GAA AAC TCA	2280
CGT TAA GGG ATT TTG GTC ATG AGA TTA TCA AAA AGG ATC TTC ACC TAG ATC CTT TTA AAT	2340
TAA AAA TGA AGT TTT AAA TCA ATC TAA AGT ATA TAT GAG TAA ACT TGG TCT GAC AGT TAC	2400
CAA TGC TTA ATC AGT GAG GCA CCT ATC TCA GCG ATC TGT CTA TTT CGT TCA TCC ATA GCT	2460
GCC TGA CTC CCC GTC GTG TAG ATA ACT ACG ATA CCG GAG GGC TTA CCA TCT GGC CCC AGT	2520
GCT GCA ATG ATA CCG CGA GAC CCA CGC TCA CCG GCT CCA GAT TTA TCA GCA ATA AAC CAG	2580
CCA GCC GGA AGG GCC GAG	
CGC AGA AGT GGT CCT GCA ACT TTA TCC GCC TCC ATC CAG TCT	2640
ATT AAT TGT TGC CGG GAA GCT AGA GTA AGT AGT TCG CCA GTT AAT AGT TTG CGC AAC GTT	2700
GTT GCC ATT GCT ACA GGC ATC GTG GTG TCA CGC TCG TCG TTT GGT ATG GCT TCA TTC AGC	2760
TCC GGT TCC CAA CGA TCA AGG CGA GTT ACA TGA TCC CCC ATG TTG TGC AAA AAA GCG GTT	2820
AGC TCC TTC GGT CCT CCG ATC GTT GTC AGA AGT AAG TTG GCC GCA GTG TTA TCA CTC ATG	2880
GTT ATG GCA GCA CTG CAT AAT TCT CTT ACT GTC ATG CCA TCC GTA AGA TGC TTT TCT GTG	2940
ACT GGT GAG TAC TCA ACC AAG TCA TTC TGA GAA TAG TGT ATG CCG CGA CCG AGT TGC TCT	3000
TGC CCG GCG TCA ATA CCG GAT AAT ACC G	
CG CCA CAT AGC AGA ACT TTA AAA GTG CTC ATC	3060
ATT GGA AAA CGT TCT TCG GGG CGA AAA CTC TCA AGG ATC TTA CCG CTG TTG AGA TCC AGT	3120
TCC ATG TAA CCC ACT CGT GCA CCC AAC TGA TCT TCA GCA TCT TTT ACT TTC ACC AGC GTT	3180
TCT GGG TGA GCA AAA ACA GGA AGG CAA AAT GCC GCA AAA AAG GGA ATA AGG GCG ACA CGG	3240
AAA TGT TGA ATA CTC ATA CTC TTC CTT TTT CAA TAT TAT TGA AGC ATT TAT CAG GGT TAT	3300
TGT CTC ATG AGC GGA TAC ATA TTT GAA TGT ATT TAG AAA AAT AAA CAA ATA GGG GTT CCG	3360
CGC ACA TTT CCC CGA AAA GTG CCA CCT GAC GTC TAA GAA ACC ATT ATT ATC ATG ACA TTA	3420
ACC TAT AAA AAT AGG CGT ATC ACG AGG CCC TTT CGT CTT CAC	3462

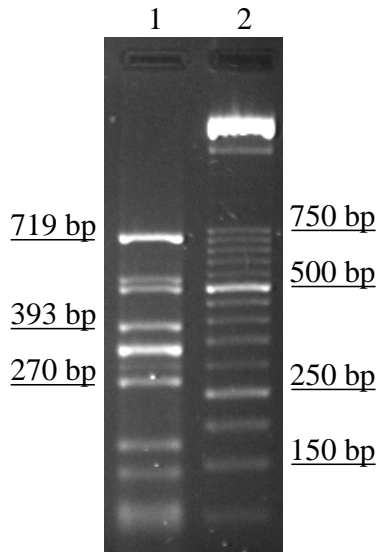
**Abb. 15.** DNA-Sequenz des Plasmids pQE30 in 5'-3'-Leserichtung

Verwendung von EDTA, Glycogen, Natriumazetat, Isopropanol und Ethanol gereinigt.

Mittels Agarosegel-Elektrophorese (3% Agarose in TAE mit Ethidiumbromid) trennt man die durch den Schnitt ent-

standenen DNA-Fragmente entsprechend ihrer Länge auf, siehe Abb. 16.

Die Extraktion der 761, 393 und 270 Basenpaare umfassenden DNA-Fragmente aus dem Agarosegel in gesonderte



**Abb. 16.** Kodierung der Gegenstandsgewichte durch längenspezifische DNA-Doppelstränge (Lane1: dephosphorylierte Schnittfragmente, Lane2: 50 bp-Längenstandard)

Tubes liefert jene DNA-Doppelstränge, welche die Gegenstandsgewichte kodieren. Die Isolierung der DNA erfolgte unter Benutzung einer Methode, deren aufeinander abgestimmte Reagenzien kommerziell angeboten werden.

### 4.3 Laborimplementierung

Das eingangs vorgestellte Integer-Rucksackproblem wurde laborpraktisch gelöst. Das angestrebte Rucksackgesamtwicht  $b = 1112$  kann durch unzulässige Packmöglichkeiten, bei denen gleiche Gegenstände mehrfach enthalten sind, nicht auftreten. Das heißt, die Gleichung  $393i + 270j + 719k = 1112$  hat keine Lösung für  $i > 1, j > 1, k > 1; i, j, k \in \mathcal{N}$ . Deshalb kann im Beispiel auf die Verwendung eines 5'-biotinylierten Starter-DNA-Doppelstranges, der allen Packmöglichkeiten vorangestellt wird und es ermöglicht, zulässige von unzulässigen zu unterscheiden, verzichtet werden.

**4.3.1 Erzeugung der DNA-Stränge für alle Packmöglichkeiten.** Der Aufbau der DNA-Stränge für alle Packmöglichkeiten erfolgt durch Anwendung molekularbiologischer Rekombinationstechniken. Das Ziel besteht darin, die DNA-Doppelstränge miteinander zu kombinieren, welche die Gegenstandsgewichte kodieren. Das Kombinieren erfolgt durch gezieltes Aneinanderketten von DNA-Doppelsträngen durch Ligation. Die Ligation ist eine biochemische Reaktion, in deren Verlauf endenkompatible und zugleich 5'-phosphorylierte DNA-Doppelstränge verbunden werden. „Endenkompatibel“ bedeutet, daß die Einzelstrangüberhänge bei sticky-Enden komplementär sein müssen. Blunt-Enden sind generell miteinander verkettbar, die Reaktion läuft jedoch nicht so effektiv ab wie die sticky-Enden-Ligation. An jeder „Klebestelle“ muß mindestens einer der beiden beteiligten DNA-Doppelstränge am 5'-Ende phosphoryliert sein, um sie verketteten zu können. Mittels 5'-Phosphorylierung läßt sich somit die Ligation steuern. Die Phosphatgruppe ist – anschau-

lich betrachtet – das Bindeglied zwischen jeweils zwei Bausteinen im DNA-Strang. Im Verlauf der Ligation verbinden sich immer mehr DNA-Doppelstränge miteinander, so daß zunehmend längere DNA-Fragmente entstehen. Ligiert man DNA-Stränge der Länge  $m$  Basenpaare, die alle geforderten Eigenschaften besitzen, dann entstehen DNA-Stränge der Längen  $i \cdot m$ ,  $i \geq 2, i \in \mathcal{N}$  Basenpaare. Das zu Beginn der Reaktion vorhandene DNA-Material wird dabei aufgeteilt.

Zunächst wird eine Teilmenge (ein Aliquot) der 719 bp-Fragmente 5'-phosphoryliert. Hierzu wurde die DNA-Probe mit Adenosintriphosphat und T4-Polynukleotid-Kinase (New England BioLabs) für 2 h bei  $+37^\circ \text{C}$  unter geeigneten Pufferbedingungen inkubiert und das Enzym danach thermisch deaktiviert.

Anschließend ligiert man folgende DNA-Doppelstränge:

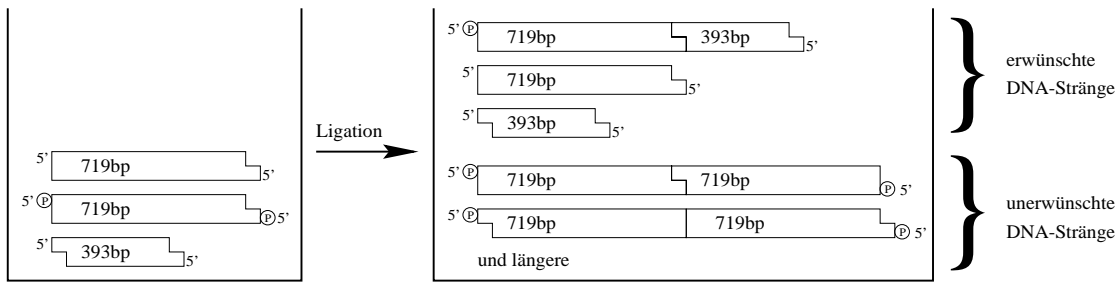
- die 5'-dephosphorylierten 719 bp-Fragmente,
- die 5'-phosphorylierten 719 bp-Fragmente und
- die 5'-dephosphorylierten 393 bp-Fragmente

Alle 719 bp-Fragmente besitzen ein blunt-Ende und ein sticky-Ende. Die 5'-dephosphorylierten 719 bp-Fragmente sind aufgrund der fehlenden 5'-Phosphatgruppe nicht in der Lage, sich mit anderen Fragmenten zu verbinden. Sie verkörpern genau jene Rucksackbelegung, die ausschließlich Gegenstand 1 enthält. An das sticky-Ende der 5'-phosphorylierten 719 bp-Fragmente kann sich genau ein 393 bp-Fragment anhängen. Als unerwünschter Nebeneffekt entstehen durch die Ligation auch Fragmente der Längen  $i \cdot 719$  Basenpaare,  $i > 1; i \in \mathcal{N}$ , indem sich mehrere 719 bp-Fragmente miteinander verbinden. Diese Fragmentlängen befinden sich jedoch außerhalb des Bereichs, der alle Packmöglichkeiten kodiert. Abb. 17 zeigt die zur Ligation eingesetzten DNA-Stränge (links) und die im Verlauf der Reaktion gebildeten DNA-Stränge (rechts). Die genannten DNA-Proben wurden unter Verwendung von Polyethylenglykol und T4-DNA-Ligase (New England BioLabs) unter geeigneten Pufferbedingungen in einer Inkubationszeit von 12 h bei  $+16^\circ \text{C}$  ligiert.

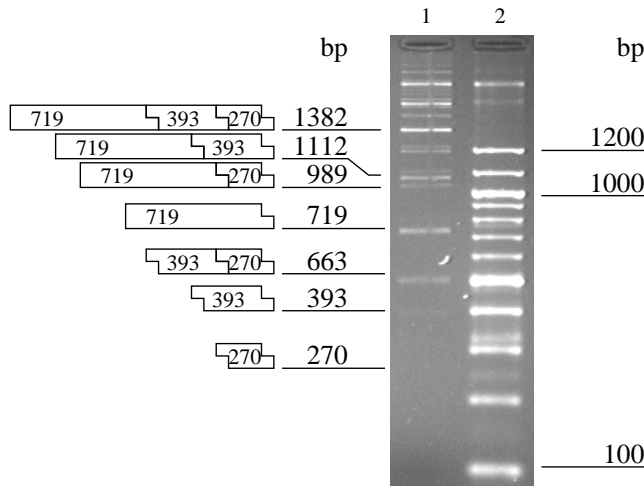
Mittels Phosphorylierung und anschließender Ligation entsteht somit aus den beiden DNA-Fragmenten der Gegenstände 1 und 2 die kombinatorische Vielfalt von 4 Rucksackbelegungsmöglichkeiten, wobei der leere Rucksack jedoch nicht DNA-kodiert ist. Damit ist der erste Schleifendurchlauf vollzogen, und im nächsten Schleifendurchlauf wird die Anzahl zulässiger Packmöglichkeiten von 4 auf 8 verdoppelt.

Im zweiten Schleifendurchlauf wird eine Teilmenge des Ligationsproduktes nach dem genannten Protokoll 5'-phosphoryliert. Das phosphorylierte Ligationsprodukt, das nicht phosphorylierte Ligationsprodukt und das 270 bp umfassende DNA-Fragment für Gegenstand 3 dienen als DNA-Material für eine erneute Ligation, die unter gleichen Bedingungen wie die erste Ligation abläuft. Dabei können auch DNA-Stränge der Längen  $i \cdot 393$ ,  $i \cdot 393 + 270$  sowie  $i \cdot 393 + 2 \cdot 270$  Basenpaare, jeweils  $i > 1, i \in \mathcal{N}$ , entstehen. Diese Nebenprodukte beeinträchtigen das Ergebnis jedoch nicht, da die Länge von 1112 bp auf diese Weise nicht gebildet werden kann.

**4.3.2 Selektion und Dekodierung der Problemlösung.** Die Gel-Elektrophorese separiert DNA-Stränge entsprechend ih-



**Abb. 17.** Wirkung der ersten Ligation



**Abb. 18.** DNA-Repräsentation der Rucksackbelegungen (Lane1: DNA-Stranglängen, welche die Rucksackgewichte repräsentieren, Lane2: 100bp-Längenstandard). Im vorliegenden Beispiel enthält der finale DNA-Pool DNA-Stränge der Länge  $b$ , und zwar genau jene, welche die Packmöglichkeit „Gegenstand1 und Gegenstand2“ kodieren. Die Lösung des Integer-Rucksackproblems lautet somit „ja“. Im Falle der Lösung „nein“ entstehen keine DNA-Stränge der Länge  $b$

rer Länge, wobei durch Vergleich mit einer Skala – gebildet aus DNA-Strängen bekannter Längen – eine Längenbestimmung möglich ist. Unter Benutzung dieser Methode gilt es nunmehr zu prüfen, ob DNA-Stränge existieren, deren Länge dem angestrebten Gesamtgewicht  $b$  des Rucksackinhaltes entspricht. Gibt es keine derartigen DNA-Stränge, so lautet die Lösung des Rucksackproblems „nein“. Sind hingegen DNA-Stränge der entsprechenden Länge vorhanden, muß weiterführend geprüft werden, ob es sich dabei um eine zulässige oder eine unzulässige Rucksackbelegung handelt. Diese Prüfung erfolgt dadurch, daß die DNA-Stränge der entsprechenden Länge aus dem Gel isoliert und anschließend sequenziert werden. Als Ergebnis erhält man die Nucleotidsequenz. Mit deren Kenntnis läßt sich leicht feststellen, ob eine zulässige oder eine unzulässige Rucksackbelegung vorliegt. Im Falle einer unzulässigen Rucksackbelegung lautet die Problemlösung „nein“, anderenfalls „ja“, und es ist darüber hinaus auch bekannt, welche Gegenstände sich im Rucksack befinden.

Im Beispiel wurde eine Agarosegel-Elektrophorese (1,5% Agarose in TAE mit Ethidiumbromid) durchgeführt, siehe Abb. 18. Die in der Beispielimplementierung entstandenen DNA-Stränge lassen sich anhand ihrer Länge den korrespondierenden Packmöglichkeiten zuordnen, wobei die eingesetzte Agarosegel-Elektrophorese hinsichtlich Auflösung

und Trennleistung begrenzt ist. Die Lösung der Beispielinstantz lautet „ja“, die entsprechende DNA-Bande ist erkennbar. Abb. 19 zeigt die angewandte Operationsfolge.

#### 4.4 Auswertung

Mit dem DNA-Algorithmus zur Lösung des Rucksackproblems wurde der Versuch unternommen, natürliche Zahlen durch DNA-Stränge spezifischer Länge darzustellen und Additionen natürlicher Zahlen durch gezieltes Verkettens dieser DNA-Stränge zu realisieren. Die laborpraktische Implementierung des DNA-Algorithmus anhand des gewählten Beispiels hat die prinzipielle Realisierbarkeit des Verfahrens bei kleinen Problemgrößen nachgewiesen. Die Abarbeitung des DNA-Algorithmus im Labor beanspruchte etwa 40h.

**Komplexitätstheoretischer Zeitaufwand:** Die Anzahl erforderlicher DNA-Operationen verhält sich linear zur Anzahl  $n$  der Gegenstände.

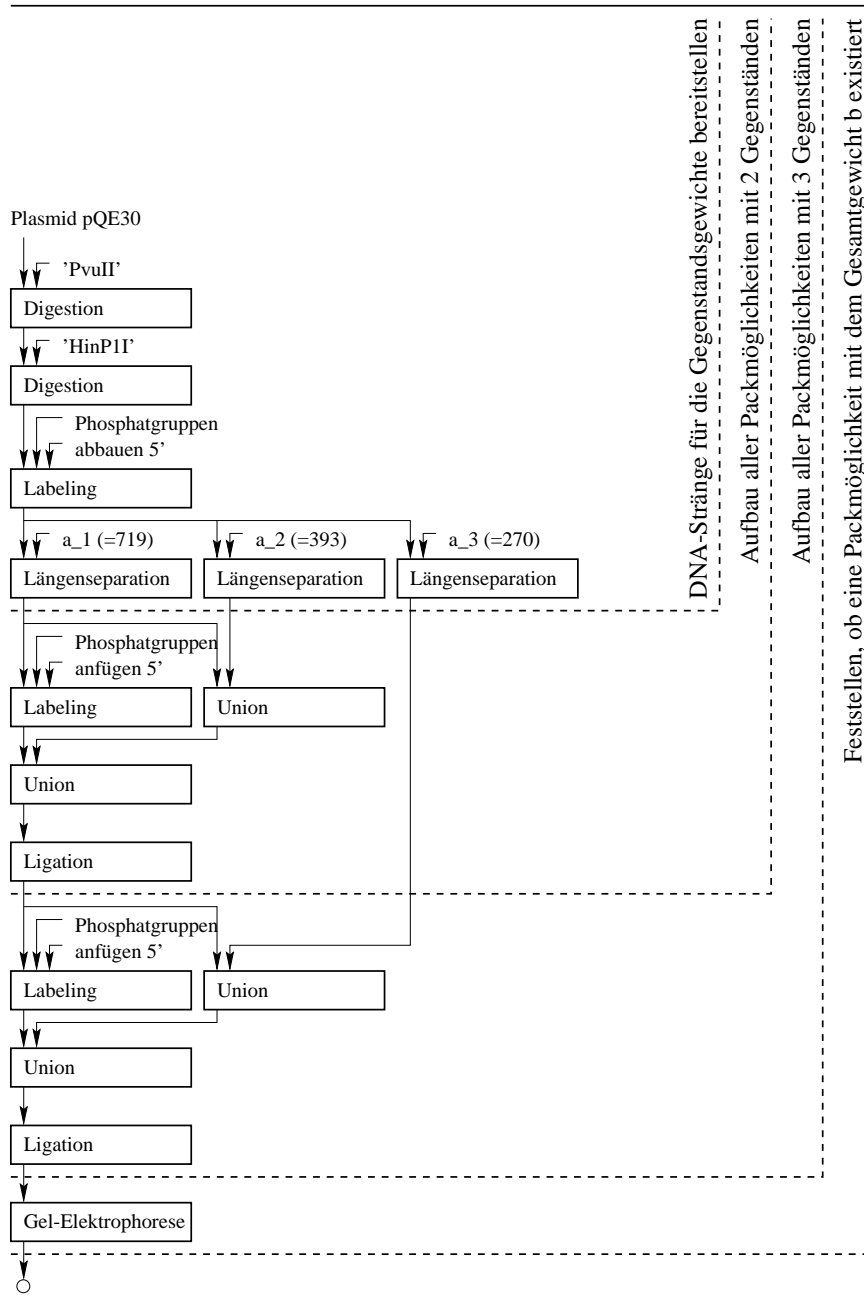
**Speicherplatzbedarf:** Im Verlaufe der Abarbeitung werden  $2^n$  informationstragende unterschiedliche DNA-Stränge gebildet, wenn von der Verwendung des Starters ausgegangen wird.

**Skalierbarkeit:** Der DNA-Algorithmus ist bei seiteneffektfreier Ausführung der DNA-Operation nicht in der verarbeitbaren Problemgröße beschränkt. Die laborpraktische Implementierung erfordert mit linear steigendem  $n$  jedoch exponentiell steigende Reaktionsvolumina, damit genügend nachweisbares DNA-Material zur Verfügung steht. Neben der kombinatorischen Vielfalt von DNA-Strängen, die alle zulässigen Rucksackbelegungen kodieren, können jedoch auch unerwünschte Verkettungen entstehen, bei denen ein und derselbe Gegenstand mehrfach im Rucksack vorhanden ist. Die Entstehung unzulässiger Rucksackbelegungen bei der laborpraktischen Implementierung kann zwar eingedämmt, aber nicht vollständig unterbunden werden.

## 5 Ausblick

Die Forschungsarbeiten zum DNA-Computing in Dresden sind darauf fokussiert, den Ablauf geeigneter molekularbiologischer Prozesse detailliert zu beobachten und zu analysieren. Die im Abschnitt 2 aufbereiteten und zusammenfassend beschriebenen Grundoperationen wurden bereits untersucht, weitere biochemische Reaktionen werden diesen Wissensfundus im Einklang mit der Entwicklung in der Molekularbiologie künftig schrittweise ergänzen. Mit diesen Erkenntnissen gelang die Beschreibung der Abläufe





**Abb. 19.** Ablaufplan zur Lösung eines Integer-Rucksackproblems. Durch Iteration der Schrittfolge Labeling-Union-Union-Ligation unter Einbeziehung weiterer DNA-kodierter Gegenstandsgewichte kann die verarbeitbare Problemgröße von drei auf  $n$  skaliert werden

mit formalen Mitteln. Dadurch war es möglich, komplexitätstheoretische Untersuchungen zu Abläufen des DNA-Computing vorzunehmen und vergleichende Aussagen zwischen dem Ressourcenbedarf im Labor und im Modell abzuleiten. Zusätzlich zu der dem DNA-Computing innewohnenden massiven Datenparallelität kann die Zeiteffizienz von DNA-Algorithmen auch dadurch beträchtlich gesteigert werden, daß verschiedene DNA-Operationen zeitparallel in mehreren Tubes zur Ausführung gelangen. Diese Idee entstammt verteilten Rechnerarchitekturen und läßt sich auch auf das DNA-Computing übertragen. Sie führt zu Mehrtubesystemen, bei denen DNA-Stränge zwischen den Tubes über „spezifische Filter“ zielgerichtet und ergebnisorientiert verteilt werden. Die Konstruktion derartiger Mehrtubesysteme erfolgt mit dem Anspruch, daß ihre laborpraktische Im-

plementierung immer gewährleistet bleibt. Das in Dresden entwickelte labornahe DNA-Computing-Modell unterstützt aufgrund seiner Konzeption diese Zielsetzung. Das Modell unterliegt einer stetigen Weiterentwicklung im Hinblick auf die Integration möglichst vieler Seiteneffekte sowie die statistische Auswertung ihres Auftretens, um seine Annäherung an die molekularbiologische Realität immer weiter voranzutreiben. Ergänzt wird dieser Prozeß durch die Optimierung und Feinabstimmung der Reaktionsparameter.

Gegenwärtig wird eine Simulation des Modells auf dem PC vorbereitet, bei der unter Berücksichtigung aller Operationsspezifikationen und bisher erfaßten Seiteneffekte Operationsfolgen auf Nucleotidebene visualisiert werden können. Diese Arbeiten sollen einen Beitrag zur Überwindung des Freiraums zwischen Modelloperation und zugeordneter La-

borteknik leisten. Besonders motivierend wirkt dabei, daß der Aufbau komplexer Bio-Operationen möglich wird, die jeweils eine Einheit bilden und sich aus speziellen reproduzierbaren Abfolgen von DNA-Grundoperationen zusammensetzen. Diese komplexen Bio-Operationen besitzen die Funktion von Bausteinen und entsprechen einer mathematischen Definition auf hohem Abstraktionsniveau. Durch fortlaufende Zusammenschaltung derartiger Bausteine wird die Gewinnung immer umfassenderer Bio-Operationen angestrebt, die in finaler Konsequenz den universellen Bio-Computer repräsentieren.

## Literatur

1. Adleman L.M.: Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994)
2. American Mathematical Society: Proceedings of the 1st DIMACS Workshop on DNA Based Computers, held at Princeton University, 1995, April 4, American Mathematical Society, Series in Discrete Mathematics and Theoretical Computer Science, 1995
3. American Mathematical Society: Proceedings of the Second Annual Meeting on DNA Based Computers, held at Princeton University, 1996, June 10-12, American Mathematical Society, Series in Discrete Mathematics and Theoretical Computer Science, 1996
4. American Mathematical Society: Proceedings of the Fifth International Meeting on DNA Based Computers, held at Massachusetts Institute of Technology, 1999, June 14-15, American Mathematical Society, Series in Discrete Mathematics and Theoretical Computer Science, 1999
5. Amos M., Gibbons A., Hodgson D.: Error-resistant implementation of DNA computations. <http://www.csc.liv.ac.uk/~martyn/princeton.ps> (1996)
6. Braich R.S., Johnson C., Rothmund P.W.K., Hwang D., Chelyapov N., Adleman L.M.: Solution of a Satisfiability Problem on a Gel-Based DNA Computer. In: [13]
7. Faulhammer D., Cukras A.R., Lipton R.J., Landweber L.F.: Molecular Computation: RNA Solutions to Chess Problems. *Proc. Natl. Acad. Sci. USA* **97**, 1385–1389 (2000)
8. Hagiya M.: From Molecular Computing to Molecular Programming. In: [13]
9. Hartemink A.J., Gifford D.K.: Thermodynamic simulation of deoxyoligonucleotide hybridization for DNA computation. In: [28]
10. Hinze T., Sturm M.: A universal functional approach to DNA computing and its experimental practicability. In: [13]
11. Kari L.: From Micro-soft to Bio-Soft: Computing with DNA. <http://www.csd.uwo.ca/~lila>, 1997
12. Kari L., Rubin H., Wood D.H. (eds): Proceedings 4th DIMACS Workshop on DNA Based Computers, held at the University of Pennsylvania, Philadelphia, USA, 1998, June 15-19, American Mathematical Society, Series in Discrete Mathematics and Theoretical Computer Science, 1998
13. Leiden center for natural computing: Proceedings 6th DIMACS Workshop on DNA Based Computers, held at the University of Leiden, Leiden, The Netherlands, 2000, June 13-17, Series in Discrete Mathematics and Theoretical Computer Science, 2000
14. Letsinger R.L., Finnan J.L., Heavner G.A., Lunsford W.B.: Phosphate coupling procedures for generating internucleotide links. *J. Am. Chem. Soc.* **97**, 3278 (1975)
15. Lipton R.J.: Using DNA to solve NP-complete problems. Technical report, Princeton University, 1995
16. Paun G., Rozenberg G., Salomaa A.: DNA Computing. *New Computing Paradigms*. Berlin Heidelberg New York: Springer 1998
17. Paun G.: Computing with Membranes. TUCS Technical Report No. 208, Turku Centre for Computer Science, 1998
18. Paun G.: SPLICING – a challenge for formal language theorists. *Journal of Automata, Languages and Combinatorics* **4**(1), 3–16 (1999)
19. Reif J.H.: Parallel molecular computation: Models and simulations. *Algorithmica*, Special issue on Computational Biology (1998)
20. Reif J.H., LaBean T.H.: Computationally Inspired Biotechnologies: Improved DNA Synthesis and Associative Search Using Error-Correcting Codes and Vector-Quantization. In: [13]
21. Rooß D., Wagner K.W.: On the power of DNA-computers. Technical Report, University of Würzburg, 1995
22. Roweis S., Winfree E., Burgoyne R., Chelyapov N.V., Goodman M.F., Rothmund P.W.K., Adleman L.M.: A sticker based model for DNA computation. In: [3]
23. Schöning U.: *Theoretische Informatik kurz gefaßt*. Mannheim, Leipzig, Wien, Zürich: BI Wissenschaftsverlag 1992
24. Stoschek E., Sturm M., Hinze T.: Technical Report TUD-FI99-02, Technische Universität Dresden, 1999
25. Sturm M., Hinze T.: Distributed Splicing of  $\mathcal{RE}$  with 6 Test Tubes. Workshop on Multiset Processing, 2000, August 21-25, Curtea de Arges, Romania, 2000
26. Watson J.D., Crick F.H.: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737 (1953)
27. Watson J.D., Gilman M., Witkowski J., Zoller M.: *Recombinant DNA*. New York, NY: Scientific American Books (2nd edition), 1992
28. Wood D. (ed): Proceedings of the 3rd DIMACS Workshop on DNA Based Computers, held at the University of Pennsylvania, 1997, June 23-25, American Mathematical Society, Series in Discrete Mathematics and Theoretical Computer Science, 1997



*Erwin P. Stoschek*, geboren 1936. Studium der Elektrotechnik, Physik und Kerntechnik an der TU Dresden. 1965 Promotion an der TU Dresden, Fachgebiet Theoretische Elektrotechnik. Von 1965 bis 1972 Industrietätigkeit im Kombinat Robotron. 1974 Habilitation auf dem Gebiet Informatik. Von 1978 bis 1990 Dozent an der TU Dresden. Seit 1990 Inhaber des Lehrstuhls Algorithmen- und Automatentheorie an der Fakultät Informatik der TU Dresden, Institut für Theoretische Informatik. Forschungsgebiete: Algorithmenkonstruktion, Computergraphik, Future Computing.



*Monika Sturm*, geboren 1954 in Naumburg/Saale. Informatikstudium an der TU Dresden, Sektion Informationsverarbeitung, Diplom 1977. Bis 1989 Forschungs- und Entwicklungsarbeiten bei Robotron Elektronik Dresden. 1989 Promotion an der TU Dresden, Fachgebiet Angewandte Informatik. Seit 1989 wissenschaftliche Mitarbeiterin an der TU Dresden, Fakultät Informatik, Institut für Theoretische Informatik. Forschungsgebiete: DNA-Computing, Funktionale Programmierung, Modelle der Berechenbarkeit, Formale Sprachen.



*Thomas Hinze*, geboren 1971 in Halle/Saale. Von 1988 bis 1991 Berufsausbildung zum Facharbeiter für Betriebs-, Meß-, Steuer- und Regelungstechnik mit Abitur bei der Chemie AG Bitterfeld-Wolfen. Ab 1992 Informatikstudium an der TU Dresden, Diplom 1997. Bis 1998 Einsatz als Software-Entwickler bei CAD & LAN GmbH Bautzen. Seit 1998 Arbeit an der Promotion zum Thema DNA-Computing an der TU Dresden, Fakultät Informatik. Forschungsgebiete: DNA-Computing, Modelle der Berechenbarkeit, Algorithmentheorie, Komplexitätstheorie.