

Sequenzierung einzelner DNA-Moleküle in Echtzeit

Jenny Ibach und Susanne Brakmann*

DNA-Polymerase · DNA-Sequenzanalyse · Einzelmoleküldetektion · Fluoreszenz · Nucleotide

Der Markt für schnelle und kostengünstige Technologien zur DNA-Sequenzierung war nie größer als heute: Er umfasst Forschungsinteressen im Bereich der vergleichenden Genomik, der Evolution, der epidemiologischen Studien sowie der Forensik und der medizinischen Diagnostik. Besonders gefragt ist die Identifizierung von Sequenzvariationen (SNP) des menschlichen Genoms, die einen Schlüssel zum Verständnis von Gesundheit oder Krankheit verspricht. Das starke Interesse an diesen biomedizinischen Problemen löste die Entwicklung einer Reihe von Techniken^[1a] für die SNP-Genotypisierung und die DNA-Sequenzierung aus, von denen einige Ergebnisse in Echtzeit liefern und die Möglichkeit einer deutlichen Maßstabsvergrößerung bieten. Ein exzellentes Beispiel stellt die Pyrosequenzierung^[1b] dar, die zwar nur bedingt für die Sequenzanalyse vollständiger Genome geeignet ist, dafür aber in idealer Weise die Sequenzierung kleinerer Fragmente mit bis zu 250 Basen und somit die SNP-Genotypisierung und die Identifizierung multipler SNPs in einem Fragment ermöglicht.

Die Entwicklung von Sequenziertechniken der nächsten Generation erfolgt mit dem Ziel, den enormen Aufwand an Material, Arbeit und Zeit vor allem für die Amplifizierung, Klonierung und Reinigung von DNA zu reduzieren, den die derzeit genutzten Techniken mit sich bringen, und dadurch die Kosten für die Sequenzanalyse eines kompletten Genoms letztlich bis auf 1000 \$ zu senken.^[2] Zu den zurzeit favorisierten Strategien gehören die Sequenzierung durch Hybridisierung (SBH), die Nanoporen-Sequenzierung sowie die Sequenzierung durch Synthese (SBS) und ihr Umkehrprozess, die Sequenzierung durch Abbau.^[3] Unter diesen verschiedenen Ansätzen sind insbesondere diejenigen Strategien interessant, bei denen eine direkte Beobachtung der „Sequenziermaschinen“ mithilfe von Einzelmolekül-Detektionsverfahren möglich ist, weil bei ihnen Amplifizierungs- und Klonierungsschritte unnötig sind und dadurch der gesamte Prozess der DNA-Sequenzierung deutlich beschleunigt und vereinfacht wird. Faszinierend ist ferner die Tatsache, dass sie das Auslesen von DNA-Sequenzen bis zu einigen Tausend Basenpaaren ermöglichen und die Lesegeschwindigkeit – abhängig vom Enzym – auf einige Hundert Basen pro Sekunde erhöhen könnten. Das Konzept der Einzelmolekülsequenzierung durch Synthese beruht auf der Beobachtung des schrittweisen Einbaus von Nucleotiden in eine wachsende DNA-Kette und erfordert eine Technik für die Unterscheidung der vier Monomere A, G, C und T, vorzugsweise durch die Verwendung von spektral separierbaren Fluoreszenzfarbstoffen.

Obwohl die Idee, einzelne DNA-Moleküle zu sequenzieren, bereits vor 20 Jahren aufkam,^[4] war die Realisierung bislang schwierig, weil für eine konsistente, lückenlose Sequenzanalyse der vollständige Austausch aller Nucleotide durch entsprechende fluoreszenzmarkierte Analoga notwendig wäre. DNA-Markierungsverfahren, die auf der Verwendung von Nucleotiden mit Basen-gekoppelten Fluoreszenzfarbstoffen beruhen, sind aufgrund zunehmender sterischer Hinderung während des enzymatischen Einbaus problematisch. Wiederholte Abspalt- und Waschschrte erschweren hingegen Ansätze, bei denen Nucleotide mit abspaltbaren Farbstoffen verwendet werden. Weitere Probleme bereiten die Isolierung und Handhabung einzelner Ziel-DNA-Moleküle sowie die verlässliche spektroskopische Identifizierung der Fluoreszenzfarbstoffe vor dem Hintergrundsignal.

Erst kürzlich gelang Korlach und Turner zusammen mit ihrem Team bei der Firma Pacific Biosciences ein Durchbruch bei der Sequenzierung einzelner Moleküle durch Synthese.^[5] Sie führten den Beweis einer Einzelmolekülsequenzierung in Echtzeit, bei der eine DNA-Polymerase den zu einer Matrize komplementären DNA-Strang komplett aus vier unterscheidbaren, fluoreszenzmarkierten Nucleotiden aufbaut. Diesen Erfolg verdankten sie vor allem zwei ausgeklügelten Techniken: der Verwendung von Nullmode-Lichtwellenleitern (zero-mode waveguides, ZMW) und der kovalenten Bindung von Fluoreszenzfarbstoffen an den 5'-Phosphatrest der Nucleotide.

Konventionelle Methoden für die Untersuchung einzelner Moleküle wie konfokale oder Totalreflexionsmikroskopie nutzen Beobachtungsvolumina im Femtoliter-Bereich (10^{-15} L) und erfordern demnach piko- bis nanomolare Konzentrationen an Fluorophoren, um individuelle Moleküle in Lösung zu isolieren. Polymerasen (und viele andere Enzyme) arbeiten jedoch bei Substratkonzentrationen im mikromolaren Konzentrationsbereich, sodass die Beobachtung individueller Polymerasemoleküle „bei der Arbeit“ mit fluoreszierenden Substraten eine 1000-fache Reduktion des konventionellen Detektionsvolumens voraussetzt. Korlach ging an dieses Problem heran, indem er nanophotonische ZMWs entwickelte. Diese Strukturen bestehen aus auf Quarzdeckgläsern angeordneten Metallfilmen mit Löchern,

[*] J. Ibach, Priv.-Doz. Dr. S. Brakmann
Chemische Biologie/BCMT, Technische Universität Dortmund
Otto-Hahn-Straße 6, 44227 Dortmund (Deutschland)
Fax: (+49) 231-9742-6627
E-Mail: susanne.brakmann@tu-dortmund.de

deren Durchmesser kleiner als die zu verwendende Lichtwellenlänge ist.^[6] Unterhalb einer kritischen Größe, die sowohl von der Wellenlänge des Lichts als auch von der Form des Lichtwellenleiters abhängt, existieren keine sich ausbreitenden Moden innerhalb des Hohlleiters, und man spricht von einem Nullmode-Wellenleiter. Wenn dieser ZMW von unten mit einem Laser angestrahlt wird, nimmt die Lichtintensität im Hohlraum exponentiell ab. Das resultierende Beobachtungsvolumen wird dadurch auf den Atto- biszeptoliterbereich (10^{-18} bis 10^{-21} L) reduziert. Bei der Beobachtung einer Polymerasereaktion, die Nucleotidkonzentrationen von 0.1 bis 10 μM für eine schnelle, genaue und prozessive DNA-Synthese erfordert, läge somit die durchschnittliche molekulare Belegung eines ZMW mit 100 nm Durchmesser bei etwa 0.01 bis 1 Molekül.

Probleme, die mit dem fortlaufenden Einbau von markierten Nucleotiden mit Basen-gekoppelten Farbstoffen verbunden sind, konnten Korlach und Turner durch die Entwicklung von vier Nucleotiden umgehen, die über einen 5'-Oligophosphatrest mit einem Fluorophor verbunden sind (Abbildung 1). Die so modifizierten Nucleotide fungieren als

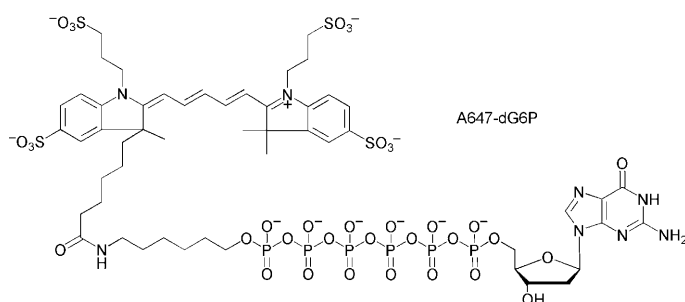


Abbildung 1. Beispiel für ein Nucleotid, das über einen 5'-Oligophosphatrest mit einem Farbstoff verbunden ist.^[7] Die Abspaltung des Farbstoffs erfolgt durch die Aktivität der DNA-Polymerase zwischen P_{α} und P_{β} .

Substrate für die DNA-Polymerase des Bakteriophagen $\Phi 29$, die die Fluorophore von jedem einzelnen Nucleotid während der Knüpfung der Phosphodiesterbindung abspaltet und dadurch natürliche, unmodifizierte DNA erzeugt. Die Wildtypform der $\Phi 29$ -DNA-Polymerase zeigt die höchste Prozessivität und Strangverdrängungsaktivität von allen bekannten DNA-Polymerasen und ist in der Lage, DNA-Moleküle mit mehr als 70 kb zu synthetisieren. Die Anforderung an das Enzym, Nucleotide mit 5'-Phosphatgebundenen Fluorophoren als Substrate zu akzeptieren,

erforderte jedoch ortsspezifische Mutationen sowie eine Anpassung der Länge und der chemischen Eigenschaften des Linkers, um den vollständigen Ersatz der natürlichen Nucleotide zu ermöglichen.

In Kombination haben die interessanten Entwicklungen von Korlach und Turner das exemplarische Auslesen von Sequenzdaten in Echtzeit ermöglicht. Das Prinzip ihres Experiments beruht auf der Emission eines Fluoreszenzsignals, das immer dann entsteht, wenn die Polymerase ein passendes, farbcodiertes Nucleotid im Detektionsvolumen des ZMW zurückhält (Abbildung 2). Da die Aufenthaltsdauer des zurückgehaltenen Nucleotids deutlich größer ist als die eines frei diffundierenden (μs) oder eines nichtkomplementären Nucleotids (< 1 ms), können „Signale“ eindeutig von der niedrigen und konstanten Hintergrundfluoreszenz unterschieden werden.

Eine Echtzeit-Sequenzierung mit vier Farben wurde von Korlach und Turner anhand eines 150 Basen langen, linearen Matrizenstrangs mit einem Zwei-Laser-Aufbau zur Anregung von vier Fluorophoren demonstriert. Die Identifizierung der Fluoreszenzsignale erfolgte nach farbstoffgewichteter Summierung von vier Referenzspektren und automatischem Abgleich mit den gemessenen Spektren nach der Methode der kleinsten Fehlerquadrate. In einem einzigen Durchlauf konnte so aus den aufgezeichneten Signalen eine Sequenz von insgesamt 158 Basen extrahiert werden, von denen 131 korrekt identifiziert worden waren, während die übrigen 27 Signale Fehler darstellten (12 Deletionen, 8 Insertionen und 7 Fehlpaarungen). Die Lesegenauigkeit konnte durch Mittelung von 15 Sequenzierläufen auf 99.3 % und somit auf einen für Sequenzierungen bekannter Zielsequenzen ausreichenden Wert verbessert werden. Die detaillierte Analyse der Daten führte ferner zur Identifizierung der maßgeblichen Fehlerquellen: 1) Zu kurze Intervalle zwischen zwei Einbauereignissen erschienen als Deletionen innerhalb der gelesenen Sequenz. 2) Die Dissoziation des korrekt paarenden

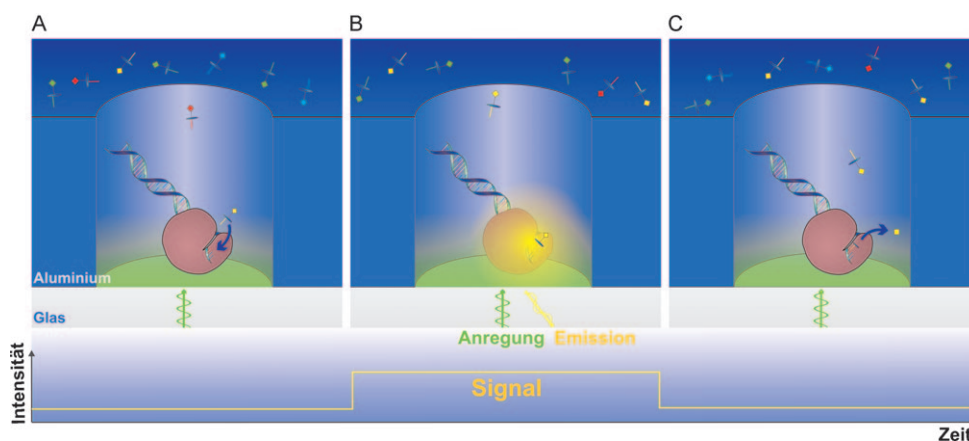


Abbildung 2. Sequenz der Ereignisse beim Einbau eines einzelnen, am Phosphatrest markierten Nucleotids (in Anlehnung an Eid et al.^[5]). Die ZMW-Nanostruktur verringert das Anregungsvolumen auf 10^{-21} L und ermöglicht dadurch die Detektion eines einzelnen fluoreszierenden Substratmoleküls vor dem Hintergrundsignal der Lösung. A) Annäherung des markierten Nucleotids an die komplementäre Base des Matrizenstrangs. B) Erkennung und Assoziation des Nucleotids in der aktiven Tasche des Enzyms rufen ein Fluoreszenzsignal hervor, das auf dem entsprechenden Farbkanal verzeichnet wird. C) Die Knüpfung der Phosphodiester-Bindung führt zur Freisetzung von Farbstoff-Linker-Oligo-/Pyrophosphat, das anschließend aus dem ZMW herausdiffundiert. Dadurch sinkt das Fluoreszenzsignal wieder auf den Wert des Hintergrundsignals.

Nucleotids aus dem Enzym-Substrat-Komplex vor der Bildung der Phosphodiesterbindung erschien als Insertion. 3) Eine falsche Signalzuordnung aufgrund unzureichender spektraler Separation erschien als Fehlpaarung.

Dieselbe Analyse führte aber auch zu einem anderen wichtigen Aspekt der Echtzeitbeobachtung auf Einzelmolekülniveau: Die erhaltenen Daten über Signaldauer und Zeitspanne zwischen zwei Signalen liefern Informationen über die Kinetik des Einbaus jedes einzelnen Nucleotids und offenbaren so die Verteilung kinetischer Daten in Hunderten von verschiedenen Sequenzkontexten. Dadurch werden Informationen über DNA-Polymerasen und generell, über Enzymkinetik, in bislang ungekannter Genauigkeit zugänglich.

Hinsichtlich einer zukünftigen Anwendung ihrer Technik für die Analyse langer Sequenzen untersuchten Korlach und Turner schließlich auch das Potenzial der Φ 29-DNA-Polymerase in „Rolling-circle“-Amplifizierungsreaktionen mit am Phosphatrest markierten Nucleotiden. Das Enzym war dabei mehr als eine Stunde in Gegenwart der modifizierten Nucleotide aktiv und baute 4000 Nucleotide in den wachsenden Strang ein. Folglich würden ungefähr 30000 funktionstüchtige ZMW ausreichen, um die Rohdaten des menschlichen Genoms an einem Tag mit einfacher Abdeckung auszulesen – eine kostengünstige Genomsequenzanalyse ist also in Sicht. Daher sind wir gespannt auf die weitere Entwicklung der Einzelmolekülsequenzierung!

Eingegangen am 17. Januar 2009
Online veröffentlicht am 4. Mai 2009

-
- [1] a) J. Shendure, R. D. Mitra, C. Varma, G. M. Church, *Nat. Rev. Genet.* **2004**, 5, 335; b) M. Ronaghi, M. Uhlen, P. Nyren, *Science* **1998**, 281, 363.
 - [2] Siehe hierzu: <http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html>.
 - [3] M. L. Metzker, *Genome Res.* **2005**, 15, 1767.
 - [4] J. H. Jett, R. A. Keller, J. C. Martin, B. L. Marrone, R. K. Moyzis, R. L. Ratliff, N. K. Seitzinger, E. Brooks-Shera, C. C. Stewart, *J. Biomol. Struct. Dyn.* **1989**, 7, 301.
 - [5] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. de-Winter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, *Science* **2009**, 323, 133.
 - [6] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, W. W. Webb, *Science* **2003**, 299, 682.
 - [7] J. Korlach, A. Bibillo, J. Wegener, P. Peluso, T. T. Pham, I. Park, S. Clark, G. A. Otto, S. W. Turner, *Nucleosides Nucleotides Nucleic Acids* **2008**, 27, 1072.