

Einzelmolekülsequenzierung von RNA

Sabine Müller*

Einzelmoleküluntersuchungen · Hochdurchsatztechniken · RNA · Sequenzanalyse · Transkriptom

Die Möglichkeit der DNA-Sequenzierung in den späten 1970er Jahren^[1] hat die Forschung in den Lebenswissenschaften stark verändert. Die Kettenabbruchmethode nach Sanger,^[1c] auch als Didesoxymethode bekannt, war jahrzehntelang die dominierende Sequenzierungstechnik und gipfelte in der Entzifferung des menschlichen Genoms.^[2] Heute wird die Sanger-Methode zunehmend von Sequenzierungstechniken der nächsten Generation abgelöst, wie 454/Roche,^[3] Illumina/Solexa,^[4] ABI/SOLiD,^[5] Helicos^[6] oder das kürzlich von Pacific BioSciences entwickelte Verfahren.^[7] Alle diese neuen Techniken folgen dem Prinzip „Sequenzierung durch Synthese“ und bieten eine beträchtliche Erhöhung des kosteneffektiven Durchsatzes, allerdings zu Lasten der Leselänge. Daher wurden diese Sequenzierungstechniken der nächsten Generation, besonders die 454-Sequenzierung, hauptsächlich für die Sequenzierung von cDNA in der Transkriptomanalyse genutzt, und weniger für die Sequenzierung von genomischer DNA.

Die bei der Sanger-Sequenzierung notwendige In-vivo-Amplifizierung wird üblicherweise umgangen durch klonungsfreie In-vitro-Amplifizierung räumlich getrennter DNA-Moleküle, bekannt als Emulsions-Polymerasekettenreaktion (Emulsions-PCR,^[8] 454/Roche und ABI/SOLiD), oder durch Festphasen-Brücken-Amplifizierung von Einzelmolekül-DNA-Templaten (Illumina/Solexa). Dadurch werden Kolonien von DNA-Molekülen generiert, die auf der Festphase analysiert und sequenziert werden. Die von Helicos^[6] und Pacific BioSciences^[7] entwickelten Verfahren beruhen auf Einzelmolekültechniken, sodass überhaupt keine Amplifizierung erforderlich ist. Das Helicos-Verfahren nutzt beispielsweise Primer-Templat-Duplex, die auf Glasplättchen immobilisiert sind, und die Sequenzierung der Einzelmoleküle wird durch schrittweise Zugabe von Farbstoff-markierten Nucleotiden gestartet. Nach jeder Runde der Nucleotidzugabe werden die farbigen Positionen auf dem Glasplättchen visualisiert.

Dieses DNA-Sequenzierungsverfahren wurde nun für die direkte RNA-Sequenzierung (direct RNA sequencing, DRS) adaptiert.^[9] Dies ist eine bedeutende Entwicklung, da mit

DRS prinzipiell Femtomolmengen der Gesamt-RNA einer gegebenen Zellpopulation ohne vorherige Umschrift in cDNA sequenziert werden können. Das ist eine entscheidende Bereicherung der Forschung auf den Gebieten Genexpressionsanalyse, Genomannotierung und Detektion von Umordnungen für die Entdeckung nichtcodierender RNAs und deren Quantifizierung.

So wie die Helicos-DNA-Sequenzierung,^[6] ist auch DRS eine Einzelmolekültechnik. Im ersten Schritt wird Poly(A)-Polymerase (PAPI) aus *E. coli* genutzt, um eine Poly(A)-Sequenz am 3'-Ende der zu analysierenden RNA-Moleküle zu generieren. Diese Prozedur ist für RNAs, die so wie mRNAs von Natur aus eine Poly(A)-Sequenz enthalten, prinzipiell nicht nötig. Durch die Zugabe von ddATP zehn Minuten nach dem Start der Poly(A)-Synthese wird die Länge der Poly(A)-Sequenz kontrolliert und gleichzeitig das 3'-Ende blockiert, um den Einbau von Nucleotiden im Templat während der Sequenzierungsreaktion zu verhindern. Die so behandelten Templatstränge werden an Glasplättchen mit immobilisierten Poly(dT)-Sequenzen hybridisiert und anschließend durch schrittweise Zugabe von Fluoreszenz-markierten 3'-O-geschützten Nucleotiden (virtuelle Terminatornucleotide, VT-Nucleotide, genannt) sequenziert. Um den Sequenzierungsstartpunkt zu definieren, werden die immobilisierten RNA-Template mit dTTP aufgefüllt und durch anschließende Zugabe von VT-A, -C und -G in Startposition gebracht (Abbildung 1A). Nichteingegebene VT-Nucleotide werden durch Waschen entfernt, und anschließend werden die farbigen Bereiche auf dem Chip detektiert. Im Anschluss werden der Fluoreszenzfarbstoff und der Inhibitor abgespalten, sodass eine freie 3'-OH-Gruppe für die nächste Runde der Nucleotidzugabe vorhanden ist (Abbildung 1B). Im weiteren Verlauf werden alternierend die vier Nucleotide (VT-A, -C, -G und -T) zugegeben, gefolgt von Waschschritten, Visualisierung und Abspaltung (Abbildung 1C–F). Nach iterativen Zyklen dieser Schritte werden die gesammelten Abbildungen justiert und zur Bestimmung der Sequenz der individuellen RNA-Moleküle genutzt.

Die Schlüsselemente dieser Technik sind zum einen eine Polymerase, die die modifizierten Fluoreszenz-markierten Nucleotide als Substrate akzeptiert, und zum anderen das Design der 3'-blockierten und markierten Nucleotide. Obwohl die Einzelmolekülsequenzierung bereits 1989 vorgeschlagen wurde,^[10] ist die Machbarkeit der Methode erst kürzlich demonstriert worden.^[11] Die Vorteile der Einzelmolekülsequenzierung bestehen darin, dass nur minimale Mengen der zu sequenziierenden Probe erforderlich sind, und

[*] Prof. Dr. S. Müller

Institut für Biochemie
Ernst-Moritz-Arndt-Universität Greifswald
Felix-Hausdorff-Straße 4, 17487 Greifswald (Deutschland)
Fax: (+49) 3834-864-471
E-Mail: smueller@uni-greifswald.de
Homepage: <http://www.chemie.uni-greifswald.de/~bioorganik>

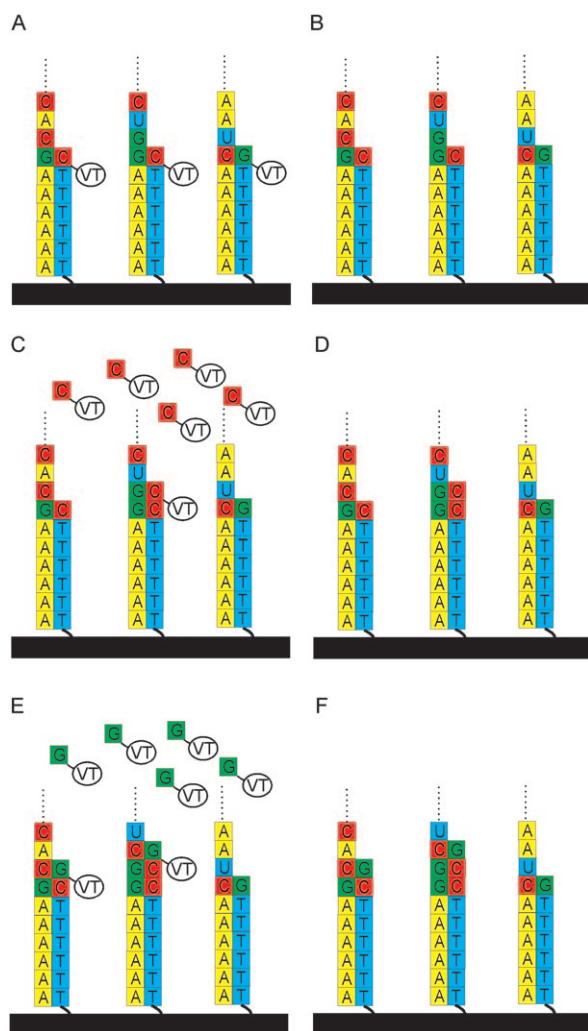


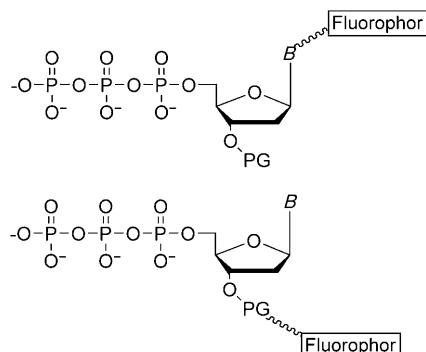
Abbildung 1. RNA-Sequenzierung. A: Polyadenylierte Templatstränge werden an Poly(dT)-beschichteten Oberflächen hybridisiert, mit dTTP aufgefüllt und mithilfe von VT-Nucleotiden in Startposition gebracht. B, D, F: Der Fluoreszenzfarbstoff und die 3'-O-Schutzgruppe werden abgespalten, eine freie 3'-OH-Gruppe wird generiert. C, E: VT-Nucleotide (z. B. VT-C oder VT-G) werden zugegeben, gefolgt von Waschschritten und Visualisierung der Position des Farbstoffes auf dem Chip.

vor allem in der Tatsache, dass jedes einzelne Molekül individuell analysiert wird. Dadurch ist es nicht nötig, dass jeder Nucleotideinbau vollständig erfolgt, was wiederum Fehleinbauten und im Zusammenhang damit die Fehlerraten reduziert. Wegen ihrer deutlich langsameren Reaktionskinetik können nichtkomplementäre Nucleotide nicht konkurrierend während der Zeit eingebaut werden, die zur Polymerisation von 80–90 % der korrekten Nucleotide benötigt wird.^[6,8]

Um die markierten VT-Nucleotide einzubauen, ist eine modifizierte Polymerase nötig. Ozsolak et al.^[9] haben bekannte reverse Transkriptasen und verschiedene DNA-abhängige DNA-Polymerasen mit maßgeschneiderter Reverse-Transkriptase-Aktivität^[12] untersucht. Offensichtlich ist es gelungen, eine Polymerase zu finden, die die VT-Nucleotide effizient einbaut und dabei ausreichend gut zwischen passenden und nicht passenden Nucleotiden unterscheidet. Leider enthält der Artikel keine weiteren Informationen zur

Natur dieser Polymerase. Ebenso verhält es sich mit den VT-Nucleotiden. Es bleibt der Fantasie des Lesers überlassen, was die chemische Natur dieser Nucleotide ist und wie der Farbstoff und die Inhibitorgruppe abgespalten werden.

Eine Möglichkeit ist die Nutzung Fluoreszenz-markierter 3'-O-blockierter Nucleotide, deren Fluoreszenzmarkierung an der heterocyclischen Base positioniert ist (Schema 1,



Schema 1. Mögliche Struktur von VT-Nucleotiden. Die Fluoreszenzmarkierung kann an der heterocyclischen Base (B) lokalisiert sein, was eine zusätzliche 3'-O-Schutzgruppe erforderlich macht (oben). Alternativ kann der Fluorophor direkt mit der 3'-O-Schutzgruppe (PG) verknüpft sein (unten).

oben), sowie Abspaltung der Fluoreszenzmarkierung und der 3'-O-Schutzgruppe unter identischen Bedingungen.^[13] Dies ist eine nahe liegende Strategie, da eine Reihe von Polymerasen Nucleotide mit voluminösen Substituenten akzeptieren kann.^[14] Vermutlich sind auch die von Ozsolak et al.^[9] genutzten VT-Nucleotide mit einem an die Nucleobase gekuppelten Farbstoff und einer separaten 3'-O-Schutzgruppe ausgerüstet, und die Abspaltung beider Gruppen erfolgt unter identischen Bedingungen. Weiterhin ist zu vermuten, dass der Farbstoff in allen vier VT-Nucleotiden identisch ist, da nach jeder Runde des Nucleotideinbaus die Lokalisation der Markierung auf dem Chip analysiert wird, und nicht die Farbe des abgespaltenen Farbstoffs.

Alternativ könnte die Farbstoffmarkierung direkt mit der 3'-O-Schutzgruppe verknüpft sein (Schema 1, unten), sodass die Abspaltung der 3'-O-Schutzgruppe über das Verschwinden der Fluoreszenz detektierbar ist. Es ist fragwürdig, ob natürliche Polymerasen Nucleotide mit derart voluminösen Gruppen an der 3'-Position als Substrate akzeptieren würden. Die modernen Techniken zur Proteinmutagenese und Evolution bieten allerdings die Möglichkeit, solche Polymerasen mit verändertem Substratspektrum zu entwickeln.

Ozsolak et al.^[9] haben mit der neuen DRS-Technik zunächst chemisch synthetisierte 40-mer-Oligonucleotide als Modellsysteme sequenziert, um die zugrunde liegende Chemie zu entwickeln und zu optimieren. Anschließend wurde Poly(A)⁺-RNA aus *Saccharomyces cerevisiae* sequenziert, und die Sequenzsätze wurden entsprechend dem Hefegenom mit bioinformatischen Methoden zugeordnet. Die durchschnittliche Satzlänge betrug 28.7 Nucleotide in 41261 Sequenzsätzen, mit 120 Sequenzierzyklen über drei Tage. Die Methode befindet sich zwar noch in einem frühen Ent-

wicklungsstadium, dieses Ergebnis belegt jedoch, dass DRS eine Hochdurchsatztechnik mit Potenzial für interessante Anwendungen ist. Nach Weiterentwicklung und Verfeinerung könnte die Technik z. B. genutzt werden, um Momentaufnahmen des Transkriptoms einer gegebenen Zellpopulation oder gar einzelner Zellen zu generieren. Leider gibt der Artikel von Ozsolak et al.^[9] keinerlei Informationen zur Chemie, die der Technik zugrunde liegt, sowie zu anderen wissenschaftlichen Details. Dies kann vom kommerziellen Standpunkt aus verständlich sein, ist aber enttäuschend für den Leser, der von einem Artikel in einem angesehenen Wissenschaftsmagazin zu Recht erwartet, alle Informationen zu erhalten, die er für das Verständnis der Arbeit benötigt.

Eingegangen am 10. November 2009
Online veröffentlicht am 13. Januar 2010

- [1] a) F. Sanger, A. R. Coulson, *J. Mol. Biol.* **1975**, *94*, 441–448; b) A. M. Maxam, W. Gilbert, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 560–564; c) F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467.
- [2] a) E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh et al., *Nature* **2001**, *409*, 860–921; b) International Human Genome Sequencing Consortium, *Nature* **2004**, *431*, 931–945; c) J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt et al., *Science* **2001**, *291*, 1304–1351.
- [3] M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen et al., *Nature* **2005**, *437*, 376–380.
- [4] a) S. T. Bennett, *Pharmacogenomics* **2004**, *5*, 433–438; b) S. T. Bennett, C. Barnes, A. Cox, L. Davies, C. Brown, *Pharmacogenomics* **2005**, *6*, 373–382; c) D. R. Bentley, *Curr. Opin. Genet. Dev.* **2006**, *16*, 545–552.
- [5] J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, G. M. Church, *Science* **2005**, *309*, 1728–1732.
- [6] T. D. Harris, P. R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. DiMeo, J. W. Efcavitch et al., *Science* **2008**, *320*, 106–109.
- [7] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman et al., *Science* **2009**, *323*, 133–138.
- [8] D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* **1998**, *16*, 652–656.
- [9] F. Ozsolak, A. R. Platt, D. R. Jones, J. G. Reiffenberger, L. E. Sass, P. McInerney, J. F. Thompson, J. Bowers, M. Jarosz, P. M. Milos, *Nature* **2009**, *461*, 814–818.
- [10] J. H. Jett, R. A. Keller, J. C. Martin, B. L. Marrone, R. K. Moyzis, R. L. Ratliff, N. K. Seitzinger, E. B. Shera, C. C. Stewart, *J. Biomol. Struct. Dyn.* **1989**, *7*, 301–309.
- [11] a) I. Braslavsky, B. Herbert, E. Kartalov, S. R. Quake, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3960–3964; b) W. J. Greenleaf, S. M. Block, *Science* **2006**, *313*, 801.
- [12] a) S. Vichier-Guerre, S. Ferris, N. Auberger, K. Mahiddine, J.-L. Jestin, *Angew. Chem.* **2006**, *118*, 6279–6283; *Angew. Chem.* **2006**, *118*, 6279–6283; *Angew. Chem. Int. Ed.* **2006**, *45*, 6133–6137; b) K. B. M. Sauter, A. Marx, *Angew. Chem.* **2006**, *118*, 7795–7797; *Angew. Chem. Int. Ed.* **2006**, *45*, 7633–7635.
- [13] D. C. Knapp, A. Keller, J. D'Onofrio, A. Lubys, S. Serva, A. Kurg, M. Remm, M. Kwiatkowski, J. W. Engels, *Nucleic Acids Symp. Ser.* **2008**, *52*, 345–346.
- [14] a) S. Jäger, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum, M. Famulok, *J. Am. Chem. Soc.* **2005**, *127*, 15071–15082; b) S. H. Weisbrod, A. Marx, *Chem. Commun.* **2008**, 5675–5685.