

Die Sequenzierung der sechsten Base (5-Hydroxymethylcytosin): basenaufgelöste Bestimmung durch selektive DNA-Oxidation**

Peter Schüler und Aubry K. Miller*

5-Hydroxymethylcytosin · DNA · Enzyme ·
Oxidationen · Sequenzierung

Die Methylierung von Cytosin (C) zu 5-Methylcytosin (5mC) ist eine der wichtigsten epigenetischen Modifikationen in eukaryotischer DNA. Wegen ihres Einflusses auf Zellentwicklung und Genexpression wird sie seit Jahrzehnten eingehend untersucht (Abbildung 1). 2009 fanden unabhängig voneinander zwei Forschergruppen 5-Hydroxymethylcytosin

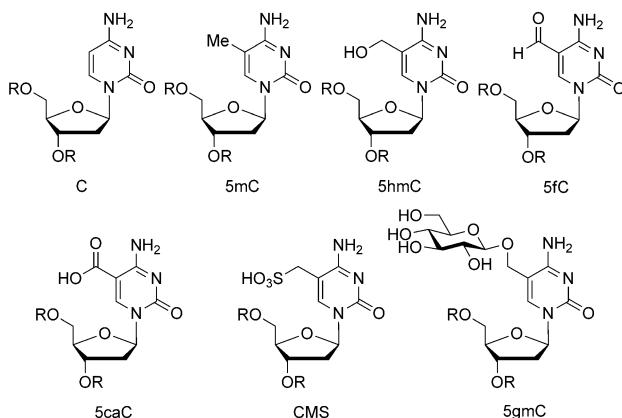


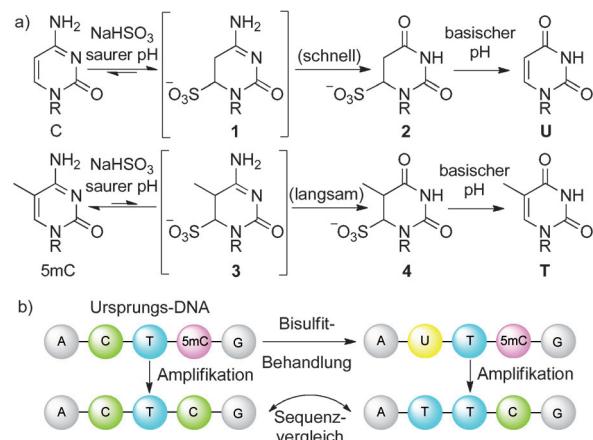
Abbildung 1. Natürliche (C, 5mC, 5hmC, 5fC und 5caC) und nichtnatürliche (CMS und 5gmc) eukaryotische Cytosinanaloge; R = DNA.

(5hmC) in Säuger-DNA und konnten zeigen, dass diese neue DNA-Base durch Enzyme der „ten-eleven-translocation“-(TET)-Oxygenasen aus 5mC hergestellt wird.^[1,2] Im letzten Jahr konnten auch die höher oxidierten Analoga 5-Formylcytosin (5fC) und 5-Carboxycytosin (5caC) als weitere TET-Oxidationsprodukte nachgewiesen werden.^[3] Mit Spannung wird nun erwartet, welche Rolle(n) diese Oxidationsprodukte von 5mC als epigenetische Markierungen oder Intermediate eines Demethylierungsmechanismus spielen.

[*] Dr. P. Schüler, Dr. A. K. Miller
Wirkstoffforschung, Deutsches Krebsforschungszentrum (DKFZ)
Im Neuenheimer Feld 280, 69120 Heidelberg (Deutschland)
E-Mail: aubry.miller@dkfz.de
Homepage: <http://www.dkfz.de/en/drugs/index.php>

[**] Wir danken dem Helmholtz Portfoliothema Wirkstoffforschung (P.S.) und dem Programm „epigenetics@dkfz“ für die finanzielle Unterstützung.

Die Bisulfit-vermittelte Desaminierung von Cytosinen zu Uracilen war ein Meilenstein auf dem Weg zum Verständnis der DNA-Methylierung. Shapiro und Hayatsu berichteten bereits vor mehr als 40 Jahren unabhängig voneinander von dieser Reaktion und untersuchten sie in der Folge detailliert anhand einzelner Nucleotide (Schema 1a).^[4] Sie konnten



Schema 1. a) Bisulfit-vermittelte Desaminierung von C und 5mC. b) Polymerasekettenreaktions(PCR)-Amplifikation eines DNA-Ab schnitts behandelt sowohl C als auch 5mC wie C. Bisulfit-Behandlung des Ursprungsstrangs wandelt C in U um, während 5mC (und alle anderen Basen) unverändert bleiben. Nach Amplifikation und Sequenzierung zeigen die als C gelesenen Basen die Positionen von 5mC an. Ein Vergleich dieser Daten mit den Befunden einer Standardsequenzierung liefert nun die Positionen von C.

zeigen, dass Bisulfit bereitwillig mit C zu Sulfonat **1** reagiert. Dieses hydrolysiert anschließend zu **2**. Erhöht man nun den pH-Wert, eliminiert Bisulfit, und das Desaminierungsprodukt Uridin (U) wird frei. In der Folge wurde deutlich, dass die analoge Reaktion von 5mC zu Thymidin (T) (über **3** und **4**) um beinahe zwei Größenordnungen langsamer verläuft als im Falle von C. Diesen Unterschied in den Geschwindigkeitskonstanten macht man sich bei der so genannten Bisulfit-Sequenzierung (BS-Seq) zunutze. Heute stehen kommerzielle BS-Seq-Kits zur Verfügung, mit deren Hilfe selektiv und mit hoher Genauigkeit C neben 5mC in genomischer DNA desaminiert werden kann (Schema 1b). Ein Abgleich zwi-

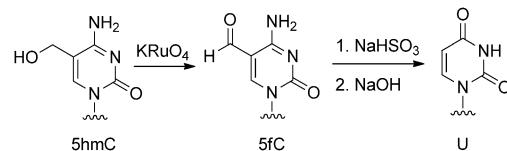
schen den Daten aus Standard- und BS-Sequenzierung ermöglicht somit die exakte Positionsbestimmung von C und 5mC in DNA. Diese Methode hat sich mittlerweile zu einem weit verbreiteten und wertvollen Hilfsmittel in der epigenetischen Forschung entwickelt und wird dazu verwendet, ganze „Methylome“ zu kartieren.^[5]

Nach der Entdeckung von 5hmC zeigte sich, dass BS-Seq nicht in der Lage ist, zwischen 5mC und 5hmC zu unterscheiden.^[6] Wie ursprünglich von Hayatsu beschrieben, reagiert 5hmC mit Bisulfat zu Cytosin-5-methylensulfonat (CMS; Abbildung 1), das noch langsamer als 5mC desaminiert und folglich nach Amplifikation und Sequenzierung gleichfalls als C gelesen wird (Schema 2b).^[7] Dies bedeutet, dass die vorhandenen, durch BS-Seq gewonnenen Genomkarten nicht vollständig korrekt sind – ein Zustand, der nur mithilfe besserer Sequenzierungsvorschriften korrigiert werden kann. Diese wurden nun in Form von modifizierten BS-Seqs durch die Gruppen um Balasubramanian^[8] und He^[9] vorgestellt. Beide Methoden nutzen hierbei selektive chemische Umwandlungen von genomicscher DNA und anschließende BS-Seq. In diesem Highlight möchten wir die Chemie hinter diesen Methoden näher beleuchten.

Im letzten Jahr konnte die Gruppe um He in Zusammenarbeit mit der von Guo-Liang Xu zeigen, dass sich 5caC unter BS-Seq-Bedingungen wie C verhält, d.h. ebenfalls als T erkannt wird.^[3b] Nun zeigten He et al., dass 5mC durch einen Überschuss an rekombinantem Tet1 zu 5caC oxidiert werden kann, während nichtmodifiziertes C nicht reagiert.^[9] Hieraus folgerten sie, dass eine selektive Oxidation von 5mC zu 5caC in Gegenwart von 5hmC in Kombination mit BS-Seq die exakte Bestimmung der 5hmC-Positionen in DNA ermöglichen sollte. Es musste „lediglich“ eine passende „Schutzgruppe“ für 5hmC gefunden werden. Die von ihnen nun veröffentlichte Vorschrift mit dem Namen „Tet-unterstützte Bisulfat-Sequenzierung“ (TAB-Seq) nutzt zwei enzymatische Reaktionen (Schema 2c): Zunächst wird das in genomicscher DNA enthaltene 5hmC selektiv mithilfe von β -Glucosyltransferase (β GT) in β -Glucosyl-5-hydroxymethylcytosin (5gmC; Abbildung 1) überführt. Danach wird die DNA mit einem Überschuss an Tet1 behandelt, der sämtliches 5mC in 5caC überführt. Die anschließende BS-Seq wandelt sämtliches C ebenso wie (dem Modell nach) sämtliches 5caC (siehe unten) in U um, während 5gmC unverändert bleibt. Nach der Amplifikation (5gmC verhält sich hier wie C) und Sequenzierung

zeigen die Positionen, die weiterhin als C gelesen werden, die ursprünglichen 5hmC-Positionen an. Der Vergleich dieser Daten mit denen einer BS-Seq liefert im Anschluss die Positionen von 5mC.

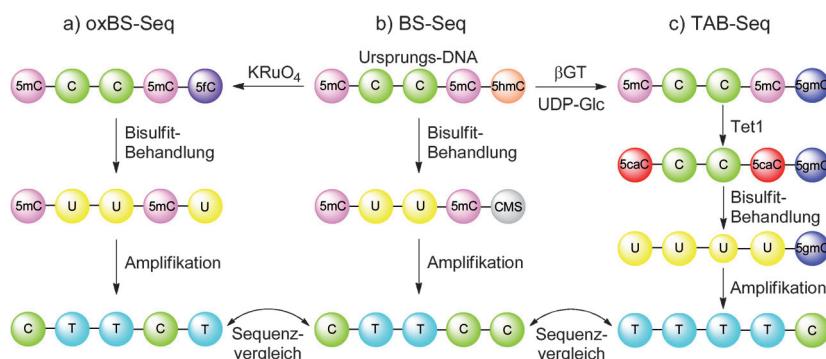
Auch der Ansatz von Balasubramanian basiert auf der selektiven Oxidation eines bestimmten Cytosinanalogs, bevor ebenfalls eine BS-Seq durchgeführt wird. Die Methode beruht auf der Erkenntnis, dass 5hmC durch KRuO₄ zu 5fC oxidiert wird, das im Anschluss einer Bisulfat-vermittelten deformylierenden Desaminierung zu U unterzogen werden kann (Schema 3). Die gesamte Vorschrift, die unter dem



Schema 3. Oxidation von 5hmC und anschließende Bisulfat-vermittelte deformylierende Desaminierung von 5fC.

Namen „oxidative Bisulfat-Sequenzierung“ (ox-BS-Seq) veröffentlicht wurde, beginnt mit einer KRuO₄-Oxidation, die sämtliches 5hmC in einem DNA-Strang zu 5fC oxidiert (Schema 2a). Die herausragende Selektivität der Oxidation ist hierbei bemerkenswert. Eine anschließende Behandlung mit Bisulfat wandelt nun außer dem vorhandenen C auch das neu gebildeten 5fC in U um. Nach Amplifikation und Sequenzierung können alle als C gelesenen Positionen 5mC zugeordnet werden. Ein Abgleich mit den Daten einer BS-Seq ermöglicht nun zusätzlich die Bestimmung der 5hmC-Positionen.

Obwohl sich oxBS-Seq (Oxidation von 5hmC) und TAB-Seq (Schützen von 5hmC und anschließende Oxidation von 5mC) deutlich in Vorgehen und den erhaltenen Daten unterscheiden, lassen sich aus ihnen die gleichen Informationen ableiten (Abbildung 2). Deshalb sind Verlässlichkeit und Handhabbarkeit die entscheidenden Kriterien dafür, welche dieser beiden Methoden sich letztlich durchsetzen wird. Das ox-BS-Seq-Verfahren scheint deutlich einfacher: Es werden keine „Schutzgruppen“, sondern lediglich eine chemische Reaktion und ein Aufreinigungsschritt benötigt, bevor die Bisulfat-Reaktion durchgeführt wird. Das alleinige Reagens,



Schema 2. Vergleich der Vorgehensweisen bei oxBS-Seq (a), BS-Seq (b), TAB-Seq (c).

Base	Standard-Sequenz	BS-Sequenz	TAB-Sequenz	oxBS-Sequenz
C	C	T	T	T
5mC	C	C	T	C
5hmC	C	C	C	T

Abbildung 2. Vergleich der Ergebnisse von TAB-Seq und oxBS-Seq mit Standard- und Bisulfit-Sequenzierungsergebnissen zur Differenzierung von C, 5mC und 5hmC.

KRuO₄, ist eine Basischemikalie und im benötigten Maßstab problemlos handhabbar. Das TAB-Seq-Verfahren hingegen umfasst zwei enzymatische Reaktionen, die beide einen (sehr einfachen) Aufreinigungsschritt nach sich ziehen. Zusätzlich fällt hier die Verfügbarkeit der Enzyme ins Gewicht. Produktion und Verwendung sind im Fall von β GT weitgehend etabliert, nicht hingegen im Fall von Tet1. Dies stellt (noch) eine deutliche Hürde auf dem Weg zu dessen breiterer Anwendung dar. In ihren Untersuchungen griff die Gruppe um He auf die Tet1-Expression in Insektenzellen zurück. Inwiefern sich die Produktion dahingehend optimieren lässt, dass ein hochwertiges Enzym verlässlich und zu einem konkurrenzfähigen Preis hergestellt werden kann, muss hier noch gezeigt werden.

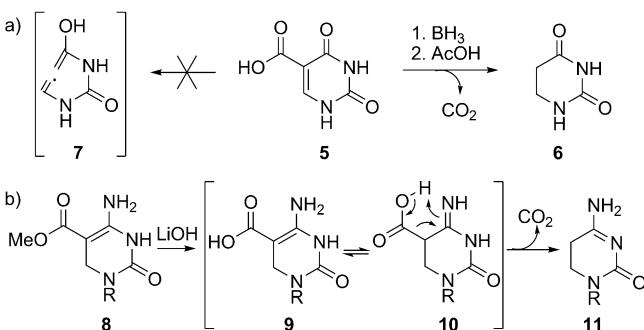
Die Gruppe um Balasubramanian konnte durch qPCR (einer quantitative Sequenzierungsbibliothek aus genomicscher DNA) abschätzen, dass lediglich 0.5% der DNA-Fragmente oxBS-Seq unbeschadet überstehen. Die Autoren ordnen 95 % der Strangbrüche der Perruthenat-Oxidation zu, während zusätzlich 90 % der verbliebenen Stränge im Laufe der Bisulfit-Behandlung brechen. TAB-Seq wurde bisher nicht auf einen derartigen Abbauprozess untersucht. Es ist jedoch naheliegend, dass die Zahl der Strangbrüche bei der Umsetzung mit β GT oder Tet1 geringer ist. Da jedoch beide Sequenzierungstechniken mit 2 μ g genomicscher DNA durchgeführt werden konnten, ist der hohe Grad an DNA-Zerstörung im Falle von oxBS-Seq offenbar kein entscheidender Nachteil.

Mithilfe von quantitativen Sequenzierungsmethoden konnten beide Gruppen die Effizienz ihrer Methoden genauer untersuchen. Die Gruppe um Balasubramanian berichtet von ca. 5 % 5hmCs, die beim Einsatz von oxBS-Seq als 5mC fehlinterpretiert wurden. Die Gruppe um He bestimmte für ihr TAB-Seq-Verfahren einen Wert von ca. 8 %. He et al. beobachteten zusätzlich eine gewisse Abhängigkeit der Glycosylierungsrate von der Nähe zu weiteren 5hmCs, was womöglich auf sterische Einflüsse zurückzuführen ist. Allerdings ist ein minimaler Abstand von drei Basen zwischen zwei 5hmCs ausreichend, um eine Glycosylierungsrate vergleichbar mit jener von isoliertem 5hmC zu erreichen. Die Rate fällt auf 90 % dieses Niveaus, falls lediglich ein G beide 5hmC-Positionen trennt. Auch wenn ein analoger sterischer Einfluss im Fall von oxBS-Seq unwahrscheinlich ist, wäre eine dahingehende Untersuchung durchaus sinnvoll.

Beide Methoden werden in Zukunft sicherlich weiter optimiert werden, gewähren aber bereits jetzt einen weitaus

detaillierteren Blick auf die Methylierungs-/Hydroxymethylierungsmuster von DNA als bisherige Methoden. Obwohl die oxBS-Seq eine erheblich aggressivere Oxidationsmethode aufweist, scheint sie derzeit der TAB-Seq überlegen. Als Hauptgrund ist hier die kommerzielle Verfügbarkeit aller eingesetzten Reagentien zu sehen, wohingegen bei einem Einsatz von TAB-Seq die Enzymproduktion eine zusätzliche Hürde ist.

Es ist naheliegend, dass die Bisulfit-Behandlung von 5caC in DNA letztlich zu einer Decarboxylierung und Desamierung zu U führt, auch wenn dieser Aspekt von He et al. nicht explizit behandelt wird. Diese Reaktion wurde bereits 1972 von Isono am Beispiel von 5-Carboxycytosin beschrieben.^[10] Ein interessantes Detail beider Sequenzierungsvorschriften wird hierbei deutlich: Das Bisulfit-vermittelte Entfernen eines Kohlenstoffatoms von einem Cytosinderivat unter milden Bedingungen. Dieser Prozess verläuft in Form einer Decarboxylierung (TAB-Seq) oder einer Deformylierung (oxBS). Die Mechanismen dieser Reaktionen sind bisher nicht vollständig geklärt, jedoch geben die Arbeiten von Pal und Carell einige Hinweise auf den Verlauf der Decarboxylierung von 5caC: 1984 konnte Pal zeigen, dass die Behandlung von 5-Carboxyuracil (**5**) mit Boran zu 5,6-Dihydouracil (**6**) führt (Schema 4a).^[11] Die Sättigung der C5-C6-Doppelbindung von **5** hat eine Decarboxylierung zur Folge.



Schema 4. a) Sättigung der C5-C6-Doppelbindung von **5** hat eine Decarboxylierung zur Folge. b) Die Verseifung von **9** führt ebenfalls zu Decarboxylierung; R = 3',5'-Bis(OTBS)-2'-desoxyribose, TBS = *tert*-Butyl-dimethylsilyl.

Doppelbindung (durch Hydroborierung) ergibt hier ein Intermediat, das in der Folge als β -Ketosäure CO₂ freisetzen kann. Eine direkte Decarboxylierung von **5** nach einem analogen Mechanismus scheitert hingegen an der vorhandenen C5-C6-Doppelbindung. Ein hypothetisches erstes Intermediat (**7**) müsste hier ein sp-hybridisiertes Kohlenstoffatom in einen sechsgliedrigen Ring zwingen.

Vor kurzem konnte Carell zeigen, dass auch 5-Carboxycytosinderivate in dieser Weise reagieren.^[12] Die Verseifung des Methylesters in Dihydro-5caC **8** bei Raumtemperatur ergab **11** in 10 % Ausbeute, wobei die Decarboxylierung vermutlich über das Tautomer **10** verläuft (Schema 4b). Ferner stellten die Autoren fest, dass ein geringer Anteil 5caC in Gegenwart von Cystein, einer Aminosäure, die unter Umständen intermedial an die C5-C6-Doppelbindung addieren kann, in DNA decarboxyliert.

Bisulfit vermag die C5-C6-Doppelbindung von Cytosin sehr effektiv anzugreifen und fördert die Decarboxylierung von 5caC vermutlich in ähnlicher Weise. Allerdings gibt es – soweit uns bekannt – allenfalls vereinzelt mechanistische Untersuchungen zur Bisulfit-vermittelten Decarboxylierung von 5caC und Deformylierung von 5fC, weshalb hier weitere Experimente notwendig sind. Auch können stereochemische Einflüsse durchaus eine Rolle spielen, da Bisulfit die C5-C6-Doppelbindung von Pyrimidinbasen hoch stereospezifisch angreift, jedoch nur eine geringe Diastereospezifität hinsichtlich des chiralen Ribosegerüsts aufweist. In der Folge bildet sich insbesondere im Falle des 5fC eine Reihe verschiedener diastereomerer Intermediate. Ungeachtet der ungeklärten mechanistischen Abläufe ermöglicht die Kombination aus selektiver DNA-Oxidation und der überraschend vielseitigen Bisulfit-Chemie eine basenaufgelöste Sequenzierung von 5hmC. Dieser Durchbruch macht es nun möglich, die epigenetische Rolle von 5hmC sehr viel eingehender zu untersuchen, und wir blicken gespannt auf die weiteren Entwicklungen in diesem Feld, die möglicherweise auch die noch selteneren C-Analoga 5fc und 5caC für eine Sequenzierung zugänglich machen werden.

Eingegangen am 18. Juni 2012

Online veröffentlicht am 26. September 2012

1972, 126, 781–790. Da die Ergebnisse damals nicht reproduziert werden konnten, wurde auf diesem Feld lange Zeit nicht weiter geforscht.

- [3] a) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, 333, 1300–1303; b) Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, G.-L. Xu, *Science* **2011**, 333, 1303–1307; c) T. Pfaffeneder, B. Hackner, M. Truss, M. Muenzel, M. Mueller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, 123, 7146–7150; *Angew. Chem. Int. Ed.* **2011**, 50, 7008–7012.
- [4] R. Shapiro, R. E. Servis, M. Welcher, *J. Am. Chem. Soc.* **1970**, 92, 422–424; H. Hayatsu, Y. Wataya, K. Kai, *J. Am. Chem. Soc.* **1970**, 92, 724–726.
- [5] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Toni-Tilippani, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, 462, 315–322; S. Balasubramanian, *Angew. Chem.* **2011**, 123, 12612–12616; *Angew. Chem. Int. Ed.* **2011**, 50, 12406–12410.
- [6] Y. Huang, W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLoS One* **2010**, 5, e88888; S.-G. Jin, S. Kadam, G. P. Pfeifer, *Nucleic Acids Res.* **2010**, 38, e125.
- [7] H. Hayatsu, M. Shiragami, *Biochemistry* **1979**, 18, 632–637.
- [8] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, 336, 934–937.
- [9] M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, 149, 1368–1380.
- [10] K. Isono, S. Suzuki, M. Tanaka, T. Nanbata, K. Shibuya, *Agric. Biol. Chem.* **1972**, 36, 1571–1579.
- [11] C. Ghosh, D. G. Schmidt, B. C. Pal, *J. Org. Chem.* **1984**, 49, 5256–5257.
- [12] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Mueller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem.* **2012**, 124, 6622–6626; *Angew. Chem. Int. Ed.* **2012**, 51, 6516–6520.

[1] S. Kriaucionis, N. Heintz, *Science* **2009**, 324, 929–930; M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, 324, 930–935.

[2] 5hmC wurde bereits 1972 in Säuger-DNA gefunden: N. W. Penn, R. Suwalski, C. O’Riley, K. Bojanowski, R. Yura, *Biochem. J.*