

## Eine verblüffende Verzerrung von DNA, hervorgerufen durch eine Methyltransferase (Nobel-Vortrag)\*\*

Richard J. Roberts\*

Ein großer Teil meiner biologischen Arbeiten wurde durch mein frühes chemisches Training beeinflusst. Beim Studium einer neuen chemischen Verbindung ist die erste und wichtigste Aufgabe, ihre molekulare Struktur genau zu bestimmen. Für einen Molekularbiologen bedeutet das gemeinhin, eine DNA zu sequenzieren, denn erst bei genauer Kenntnis der Sequenz können gezielt Experimente zur Aufklärung der Funktion entworfen werden. Ich wurde zum ersten Mal mit der Vorstellung von makromolekularen Sequenzen konfrontiert, als ich als Postdoc bei Jack Strominger in Harvard arbeitete. In dieser Zeit besuchte ich auch Fred Sangers Labor in Cambridge in England, um das RNA-Fingerprinting und -Sequenzieren zu lernen.

Kurz bevor ich ans Cold Spring Harbor Laboratory wechselte, hörte ich einen Vortrag von Dan Nathans über Restriktionsenzyme und war sofort überzeugt, daß hier der Schlüssel zur DNA-Sequenzierung lag. Der Gedanke war, daß durch das Kartieren von Restriktionsschnittstellen und das Sequenzieren kurzer Fragmente längere Sequenzen mit der Länge von Genen zusammengefügt werden könnten. Meine Arbeitsgruppe begann damit, so viele Restriktionsenzyme wie möglich zu isolieren und zu charakterisieren<sup>[1]</sup>. Mit diesen Enzymen fingen wir an, die DNA des Adenovirus-2 zu kartieren<sup>[2]</sup> und kurze Fragmente zu identifizieren, bei denen sich das Sequenzieren lohnen könnte. Eines dieser Fragmente sollte das 5'-Ende einer Adenovirus-2-mRNA zusammen mit dem eukaryontischen Promotor enthalten, der ihre Expression kontrollierte. Diese Jagd nach dem promotortragenden Fragment führte uns zur Entdeckung unterbrochener Gene und des Spleißens von RNA<sup>[3-6]</sup>. Mittlerweile hatten Fred Sanger und Walter Gilbert Methoden zur DNA-Sequenzierung entwickelt, und wir konzentrierten uns auf die Sequenzvoraussetzungen für das Spleißen von RNA. Joe Sambrook, Walter Keller und andere klonierten die dreiteilige Leadersequenz, die auf den späten mRNAs des Adenovirus-2 liegt, und Sayeeda Zain bestimmte die Sequenz<sup>[7]</sup>. Bald darauf gingen wir die gesamte Sequenz des Adenovirus-2-Genoms an, die schließlich in gemeinsamer Arbeit mit der Arbeitsgruppe von Ulf Pettersson aufgeklärt wurde<sup>[8]</sup>.

Schon frühzeitig erkannten wir, daß Computerunterstützung für das Sequenzierungsprojekt unverzichtbar war, und wir entwickelten viele Programme, um DNA-Sequenzen zusammenzusetzen und zu analysieren<sup>[9-14]</sup>. Zur gleichen Zeit begannen wir und andere, die Gene für Restriktionsenzyme und die sie begleitenden Methyltransferasen (MTasen) zu klonieren und zu sequenzieren<sup>[15,16]</sup>. Unsere ursprüngliche, naive Annahme war, daß es in jedem Restriktions-Modifikations-System eine gemeinsame Region in der Proteinsequenz des Restriktionsenzyms und der MTase gibt, die zur Erkennung der DNA-Sequenz führt. Diese Annahme beruhte darauf, daß beide Enzyme exakt die gleiche DNA-Sequenz erkennen müssen. Wenn diese gemeinsame Proteinsequenz einmal identifiziert wäre, hielten wir es für wahrscheinlich, daß wir mit Hilfe von In-vitro-Techniken andere DNA-Sequenzen erkennbar machen könnten, um so durch Protein-engineering neue Restriktionsenzyme herzustellen. Diese Vorstellung erwies sich jedoch als hoffnungslos naiv. Nicht nur, daß keine Ähnlichkeit zwischen den Gensequenzen der Restriktionsenzyme und der MTasen bestand, es gab auch keine Ähnlichkeiten zwischen den Genen für unterschiedliche Restriktionsenzyme. Dagegen gab es bemerkenswerte Sequenzähnlichkeiten unter den Genen für MTasen, besonders bei solchen, die 5-Methylcytosin bilden. Diese Beobachtung bestimmte die letzten Jahre meiner Forschungsanstrengungen und führte zu unserer neuesten Entdeckung, die Gegenstand dieses Vortrages ist.

Die Methylierung von Adenin- und Cytosinresten findet sich im allgemeinen bei Prokaryonten; die Cytosinmethylierung ist bei Eukaryonten weitverbreitet. In Bakterien ist die DNA-Methylierung Teil des Restriktions-Modifikations-Systems<sup>[17]</sup> und des Systems zur Reparatur von Fehlpaarungen<sup>[18]</sup>. In höheren Eukaryonten scheint die Methylierung von Cytosinresten essentiell zu sein<sup>[19]</sup>, um Genexpression, Entwicklungsregulation, genomische Prägung und Inaktivierung des X-Chromosoms zu kontrollieren<sup>[20]</sup>.

Bakterien methylieren ihre DNA auf drei Arten, um sie gegen Restriktionsenzyme zu schützen: an der 5- (m5C) und der N<sup>4</sup>-Position von Cytosin (m4C) sowie an der N<sup>6</sup>-Position von Adenin (m6A). Bis heute wurden 50 unterschiedliche m5C-MTase-Gene sequenziert<sup>[21]</sup>. Als wir 1987 anfangen, die Sequenzen dieser Gene zu vergleichen, war die verfügbare Computer-Software bei vielen Sequenzen noch nicht in der Lage, sie auf Übereinstimmung hin zu vergleichen, weil sich die Ähnlichkeiten zwischen ihnen auf kurze Stücke beschränkten, die wir nun Motive

[\*] Prof. Dr. R. J. Roberts  
New England Biolabs  
32 Tozer Road, Beverly, MA 01915 (USA)  
Telefax: Int. + 508/921-1527

[\*\*] Copyright © The Nobel Foundation 1994. – Wir danken der Nobel-Stiftung, Stockholm, für die Genehmigung zum Druck einer deutschen Fassung des Vortrags.

nennen und die durch untereinander ziemlich verschiedene Sequenzen getrennt sind. Die Ähnlichkeiten konnten jedoch mit bloßem Auge erkannt werden. Janos Pósfai begann in meiner Arbeitsgruppe damit, Algorithmen zu entwickeln, die diese kurzen ähnlichen Stücke bei den Proteinen finden konnten<sup>[22]</sup>. Das Programm suchte nach kurzen Triplettmustern (Abb. 1) und fügte sie zu Motiven zusammen, die hochkonservierte Regionen innerhalb eines jeden Proteinsatzes repräsentierten. Anhand die

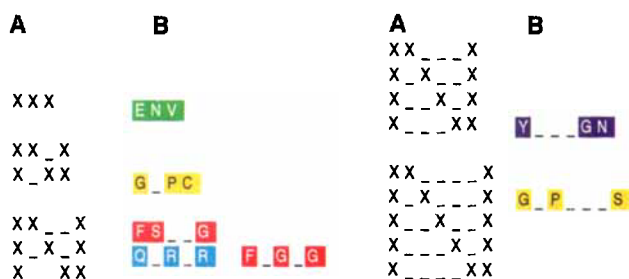


Abb. 1. A) Beispiele für die Triplettmuster, die verwendet wurden, um multiple Sequenzen auf ihre Übereinstimmung hin zu untersuchen. Jedes Muster enthält drei bestimmte Aminosäuren an jeweils festgelegten Positionen (X); an den anderen Positionen ( ) ist jede beliebige Aminosäure erlaubt. Das einfachste Muster (links oben) umfaßt alle möglichen Tripeptide ( $20 \times 20 \times 20 = 8000$  Kombinationen). Alle Sequenzen werden nach jedem dieser Muster abgesucht (bei zehn Aminosäuren 224000 insgesamt), und deren Häufigkeit wird registriert. B) Beispiele gemeinsamer Muster, die in m5C-MTasen gefunden wurden.

ser Motive wurde die Anfangsübereinstimmung fixiert. Durch eine Lockerung der Suchkriterien für die Triplets, die dann nur auf die Sequenzen zwischen den Hauptmotiven angewendet wurden, konnten Motive mit geringerer Übereinstimmung identifiziert werden, die ein weitergehendes Ordnen ermöglichten. Schließlich konnten auch Lücken eingefügt werden, um die Anordnung zu vervollständigen. Wir haben die Programme ständig weiterentwickelt und um eine Benutzeroberfläche mit graphischer Ausgabe ergänzt; diese kann die generelle Architektur einer Gruppe von Proteinen sichtbar machen. In Abbildung 2 ist schematisch die Verwandtschaft zwischen den m5C-MTasen verdeutlicht. Es gibt sechs Regionen mit starker Ähnlichkeit und vier weitere mit geringerer Ähnlichkeit, anhand derer die Übereinstimmung zwischen den Sequenzen definiert werden kann. Dazwischen befinden sich Regionen un-

terschiedlicher Länge und Sequenz. Ähnliche Analysen der m5C-MTasen wurden auch von anderen Arbeitsgruppen durchgeführt<sup>[23]</sup>.

Zwei der sechs hochkonservierten Motive konnte funktionelle Bedeutung zugeordnet werden. Motiv IV, das als Blockschema in Abbildung 3a gezeigt ist, enthält einen Cysteinrest, von dem Dan Santi et al. postulierten, er habe eine katalytische Schlüsselfunktion<sup>[24]</sup>. Sie hatten vorgeschlagen (Abb. 3b), daß in einem der ersten Schritte in einer Michael-Reaktion ein kovalenter Komplex zwischen diesem MTase-Cysteinrest und der Position 6 eines Cytosinrestes gebildet wird. Dies sollte die Posi-

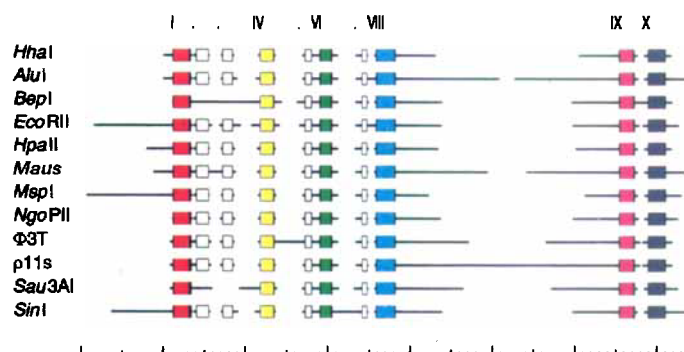


Abb. 2. Schematische Darstellung der Verwandtschaft von zwölf repräsentativen m5C-MTasen. Die sechs hochkonservierten Motive, anhand derer die Anfangsübereinstimmung definiert wurde, sind farblich unterlegt (hellrot: Motiv I, F-G-G; gelb: Motiv IV, PC; grün: Motiv VI, ENV; hellblau: Motiv VIII, Q-R-R; dunkelrot: Motiv IX, RE; dunkelblau: Motiv X, GN). Die offenen Kästchen repräsentieren die vier schwachkonservierten Motive. Links stehen jeweils die Organismen, aus denen die Enzyme gewonnen wurden.

tion 5 des Cytosinrestes aktivieren und die Übertragung einer Methylgruppe vom Cofaktor S-Adenosylmethionin (AdoMet) ermöglichen. Dieser Mechanismus wurde durch eine Reihe biochemischer Befunde gestützt, einschließlich der wichtigen Beobachtung, daß die Reaktion durch 5-Fluorcytosin aus mechanistischen Gründen gehemmt wird<sup>[25]</sup>. In späteren Arbeiten wurde die kovalente Zwischenstufe isoliert<sup>[26-28]</sup> und das Cystein aus Motiv IV als Bindungsstelle nachgewiesen<sup>[26]</sup>. Neuere Versuche haben gezeigt, daß der Ersatz dieses Cysteinrestes durch einen Glycin-, Serin- oder anderen Aminosäurerest die MTase-Aktivität blockiert<sup>[29-31]</sup>.



Richard J. Roberts, geboren 1943 in Derby, England, promovierte 1968 an der Sheffield University in Organischer Chemie. Nach einem Postdoc-Aufenthalt bei J. L. Strominger an der Harvard University arbeitete er von 1972 bis 1992 am Cold Spring Harbor Laboratory, das er als Assistant Director of Research verließ. Derzeit ist er Director of Eukaryotic Research bei den New England Biolabs, Beverly, Massachusetts. Neben intensiven Studien zu den Restriktionsenzymen des Typs II beschäftigte ihn unter anderem das Adenovirus-2. Heute gilt sein Hauptinteresse den DNA-Methyltransferasen. Er war darüber hinaus einer der Pioniere bei der Entwicklung der computergestützten Sequenzanalyse von Proteinen und Nucleinsäuren.

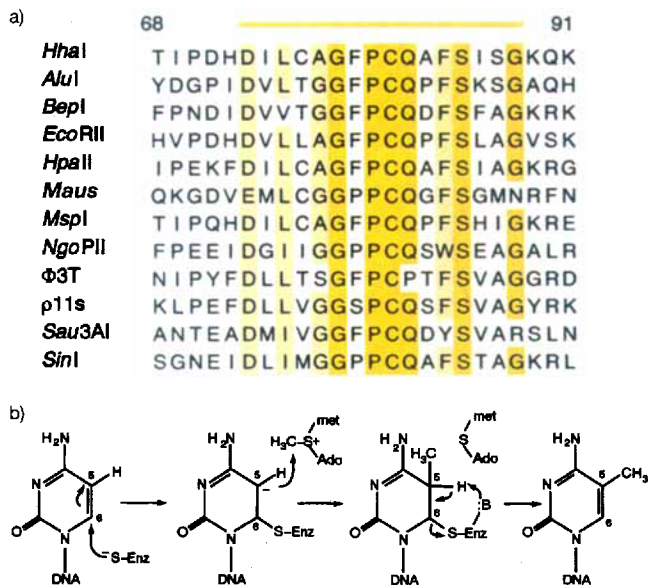


Abb. 3. a) Blockdiagramm des hochkonservierten Motivs IV (PC). Reste, die allen m5C-MTase-Sequenzen gemeinsam sind, sind kräftig gelb unterlegt, solche mit drei oder weniger Abweichungen schwach gelb. b) Schematische Darstellung des Verlaufs der Methylierung von Cytosin.

Auch Motiv I konnte eine Funktion zugeordnet werden. Es enthält drei hochkonservierte Aminosäurereste, FGG (Abb. 4a). Wichtig ist außerdem, daß es deutliche Ähnlichkeit mit Motiven aus anderen MTasen aufweist (Abb. 4b), d. h. die drei Klassen von DNA-MTasen, m5C-, m4C- und m6A-MTase, dazu die Protein- und RNA-MTasen enthalten alle ein verwandtes Motiv. Die funktionelle Gemeinsamkeit besteht darin, daß alle MTasen AdoMet als Methylendon nutzen; daher wurde vermutet, daß Motiv I an der Cofaktorbindung beteiligt ist<sup>[32, 33]</sup>.

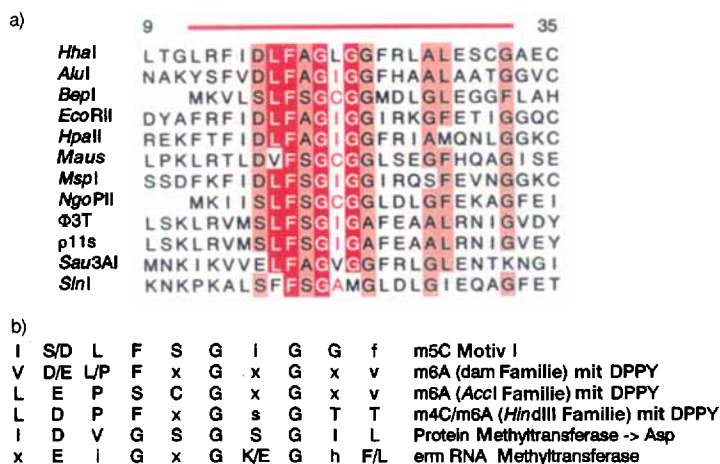


Abb. 4. a) Blockdiagramm des hochkonservierten Motivs I (FGG). Reste, die allen m5C-MTase-Sequenzen gemeinsam sind, sind rot unterlegt, solche mit drei oder weniger Abweichungen orange. b) Konsensussequenzen von Motiv I und seinen Entsprechungen in anderen MTasen (nach [32]).

Unser Interesse war darauf gerichtet, die Region innerhalb der m5C-MTasen zu identifizieren, die für die sequenzspezifische Erkennung des DNA-Substrates verantwortlich ist. Weil die Gemeinsamkeit dieser Gruppe von MTasen die Chemie der katalysierten Reaktion ist, nahmen wir an, daß jede der gemein-

samen Regionen, wie die Motive I und IV, direkt an der Umsetzung beteiligt ist, entweder durch Wechselwirkung mit dem Cofaktor oder dadurch, daß sie Aminosäurereste bereitstellt, die für die Katalyse von Bedcutung sind. Weil dagegen die genaue DNA-Sequenz, die von jedem Enzym erkannt wird, eine andere ist, erwarteten wir, daß diese Erkennung durch eine Region vermittelt wird, die bei allen Enzymen unterschiedlich ist. Nur eine Region, das sogenannte variable Segment zwischen den Motiven VIII und IX, schien ein aussichtsreicher Kandidat dafür zu sein. Eine Untersuchung dieser variablen Regionen ergab, daß sich nicht nur die Sequenz, sondern auch die Länge der Segmente von MTase zu MTase erheblich unterscheidet.

Die meisten bekannten m5C-MTasen sind monospezifisch, erkennen also nur eine einzige Sequenz. Es gibt jedoch einige MTasen, die mehr als eine spezifische Sequenz erkennen, die „multispezifischen“ MTasen, die von mehreren *Bacillus*-Bacteriophagen codiert werden. Diese Enzyme bestehen ebenso wie die monospezifischen aus einer einzigen Polypeptidkette, haben jedoch die ungewöhnliche Eigenschaft, mehrere, ziemlich unterschiedliche Sequenzen zu methylieren. Ein Beispiel ist M. $\Phi 3T$ I, das die Sequenzen GGCC und GCNGC erkennt und methyliert<sup>[34]</sup>. Dieses Enzym ist in Abbildung 2 mit aufgeführt; dort ist auch die ungewöhnlich lange variable Region sichtbar. Eine Reihe eleganter Studien der Arbeitsgruppe von Tom Trautner zeigte überzeugend, daß Punktmutationen innerhalb der variablen Regionen von M. $\Phi 3T$ I und anderen multispezifischen MTasen deren Fähigkeit, eine der spezifischen Sequenzen zu erkennen, unterbinden, während die übrigen nach wie vor erkannt werden<sup>[35, 36]</sup>. Mit den monospezifischen MTasen war ein vergleichbares Experiment nicht möglich, weil Mutationen, die die DNA-Erkennung verhinderten, gleichzeitig auch die MTase-Aktivität blockierten. Meine Mitarbeiter Saulius Klimasauskas und Janise Nelson führten eine Reihe von Experimenten durch, in denen sie einzelne Domänen zwischen zwei der monospezifischen MTasen austauschten<sup>[37]</sup>. Wir entschieden uns für M.*HpaII* (Erkennungssequenz CmCGG) und M.*HhaI* (Erkennungssequenz GmCGC), weil beide Tetranucleotid-Sequenzen erkennen und weil die methylierte Base an der gleichen Position innerhalb der Erkennungssequenzen liegt. Die variablen Regionen, mit oder ohne flankierende Sequenzen, wurden, wie in Abbildung 5 gezeigt, ausgetauscht. In mehreren Fällen entstanden aktive MTasen, und der Vergleich der Hybride zeigte eindeutig, daß die Sequenzspezifität ausschließlich in der variablen Region verankert ist. Überraschend war, daß Hybride, in denen die variable Region zusammen mit dem benachbarten Motiv IX ausgetauscht wurde, eine viel höhere Methylierungsaktivität hatten als Hybride, bei denen nur die variable Region ausgetauscht worden war.

In einer anderen Serie von Experimenten vertauschte Mi Sha die variablen Regionen von M.*HpaII* und M.*MspI*. Beide Enzyme erkennen die Sequenz CCGG, doch M.*MspI* methyliert den ersten Cytosinest der Sequenz<sup>[38]</sup>, M.*HpaII* den zweiten<sup>[39]</sup>. Aus den Ergebnissen dieser Experimente folgte, daß auch die Wahl der zu methylierenden Base von der variablen Region abhängt<sup>[40]</sup>.

Während wir unsere Studien zur Biochemie und Molekularbiologie der MTasen vorantrieben, wurde immer deutlicher, daß eine Kristallstruktur unverzichtbar war, wenn wir den Reaktionsmechanismus vollständig verstehen wollten. Ashok Dubey



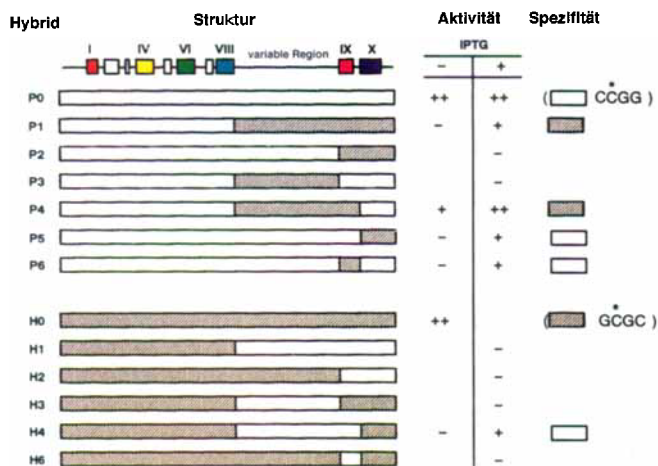


Abb. 5. Schematische Darstellung der Hybride P1-P6 und H1-H6 zwischen *M.HpaII* (P0) und *M.HhaI* (H0). Die MTasen standen unter der Kontrolle eines IPTG-induzierbaren Promotors auf einem pBR322-Derivat. Offene Kästchen symbolisieren Sequenzen von *M.HpaII*, schraffierte solche von *M.HhaI*. In der Spalte „Aktivität“ ist das Ausmaß des Schutzes von Plasmid-DNA in vivo aufgeführt, wie es durch Restriktionsenzymsspaltung in vitro vor (–) oder nach (+) IPTG-Induktion bestimmt wurde (– bedeutet kein Schutz, + bedeutet schwacher Schutz, ++ bedeutet >90% Schutz). Die Spezifitäten der Hybrid-Methyltransferasen sind rechts angegeben.

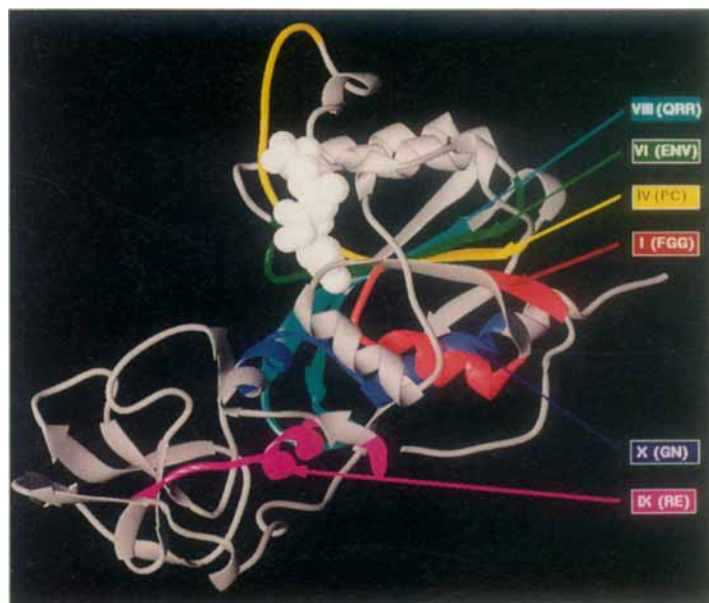


Abb. 6. Struktur von *M.HhaI* im binären Komplex mit seinem Cofaktor AdoMet. Die Motive sind mit den gleichen Farben wie in Abbildung 2 gekennzeichnet. Der Cofaktor AdoMet ist in weiß als Kalottenmodell gezeigt.

aus meiner Gruppe begann eine Zusammenarbeit mit Xiaodong Cheng aus der Arbeitsgruppe von Jim Pflugrath in Cold Spring Harbor. Sie reinigten *M.MspI*, das wir intensiv untersucht hatten, und versuchten, es zu kristallisieren<sup>[41–43]</sup>. Später versuchten sie, *M.HpaII*, das wir ebenfalls untersucht hatten<sup>[44]</sup>, zu kristallisieren. In beiden Fällen hatten sie keinen Erfolg. Erst unser dritter Anlauf mit *M.HhaI* war erfolgreich. Das Enzym ist Teil des *HhaI*-Restriktionssystems von *Haemophilus haemolyticus*, das von meiner Arbeitsgruppe entdeckt worden war<sup>[45]</sup>. Mit 327 Aminosäuren ( $M_r \approx 37$  kDa) ist es eine der kleinsten m5C-MTasen. Das zugehörige Gen war kloniert und sequenziert<sup>[46]</sup>, das Protein in *E. coli* überexprimiert<sup>[37, 47a]</sup> und detaillierten kinetischen Studien unterzogen worden<sup>[25]</sup>. Sanjay Kumar reinigte das Enzym, und im Dezember 1991 kristallisierte es bereitwillig – ein wunderbares Weihnachtsgeschenk<sup>[47b]</sup>! Innerhalb von elf Monaten hatte Xiaodong Cheng die Struktur des binären Komplexes aus *M.HhaI* und AdoMet mit einer Auflösung von 2.5 Å bestimmt<sup>[48]</sup>.

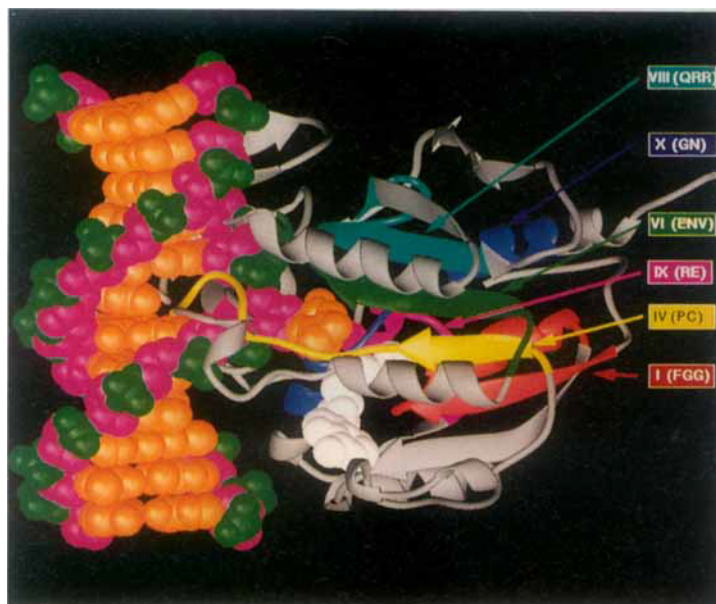
Die Struktur war außergewöhnlich aufschlußreich (Abb. 6); es handelt sich um eine große und eine kleine Domäne, die gemeinsam eine Furche bilden, die ideal geeignet scheint, eine DNA-Helix zu umschließen. Die Beteiligung von Motiv I (FGG) an der AdoMet-Bindung wurde bestätigt, und auch ein anderes der konservierten Motive, Motiv X (GN), scheint bei der AdoMet-Bindung eine Rolle zu spielen. Motiv IV (PC) liegt nahe der AdoMet-Bindestelle auf der gleichen Seite der Furche. Motiv VIII (QRR) befindet sich am Boden der Furche, wo die positiven Ladungen des Motivs möglicherweise für die Bindung des Phosphodiester-Rückgrats der DNA-Helix wichtig sind. Das konservierte Motiv IX (RE) schlängelt sich durch die kleine Domäne, die ansonsten fast aus der gesamten variablen Region besteht. Es bildet offenbar ein Rückgrat, das die Struktur der kleinen Domäne stabilisiert. Das konservierte Motiv VI (ENV) schließlich ist eindeutig an der korrekten Positionierung von Motiv IV beteiligt.

Die Struktur lieferte eine überzeugende Erklärung für unsere frühere Beobachtung, daß Domänen austausche, die Motiv IX und die variable Region umfaßten, aktivere MTasen lieferten als solche, die die variable Region allein betrafen<sup>[37, 40]</sup>. Obwohl Motiv IX bei den MTasen ziemlich hoch konserviert ist, gibt es doch auch Unterschiede zwischen den einzelnen MTasen; von diesen kann man erwarten, daß sie die Fähigkeit eines heterologen Motivs IX beeinflussen, die Struktur der kleinen Domäne zu stabilisieren.

Bald nachdem wir Kristalle des binären Komplexes aus *M.HhaI* und AdoMet erhalten hatten, versuchten wir, Cokristalle zu gewinnen, die auch DNA enthielten. Zwei Arten von Komplexen wurden ins Auge gefaßt. Die eine sollte *M.HhaI* zusammen mit nativer DNA und *S*-Adenosylhomocystein (AdoHcy) enthalten, die andere *M.HhaI* mit AdoMet und einer in der Erkennungssequenz mit 5-Fluorcytosin substituierten DNA-Doppelhelix. Im zweiten Fall würde man zwar eine kovalente Zwischenstufe erwarten, in der die Methylgruppe auf die Position 5 des Cytosinrestes übertragen wurde, die Ablösung der DNA aus dem kovalenten Komplex mit dem Protein sollte jedoch durch die Anwesenheit des Fluoratoms gehemmt sein, da dieses eine sehr schlechte Abgangsgruppe ist. Ein solcher Komplex könnte als Modell für den Übergangszustand der DNA-Methylierung angesehen werden.

Saulius Klimasauskas schloß sich dem Cokristallisationsteam an, und wir erhielten relativ einfach gute Kristalle mit nativer und 5-fluorcytosinhaltiger DNA. Xiaodong Cheng, die mittlerweile ein eigenes Labor in Cold Spring Harbor hat, löste die Struktur des Cokristalls mit 2.8 Å Auflösung (Abb. 7)<sup>[49]</sup>. Wir hatten erwartet, daß die DNA-Helix in irgendeiner Weise verzerrt ist, um die chemische Reaktion zu ermöglichen, denn der Reaktionsmechanismus erfordert, daß sich der Cysteinrest dem Cytosinrest senkrecht zur Ringebene nähert. Wir stellten uns hauptsächlich eine extreme Biegung der DNA vor. Zu unserer Überraschung war die Verzerrung viel mehr als eine Biegung

a)



b)

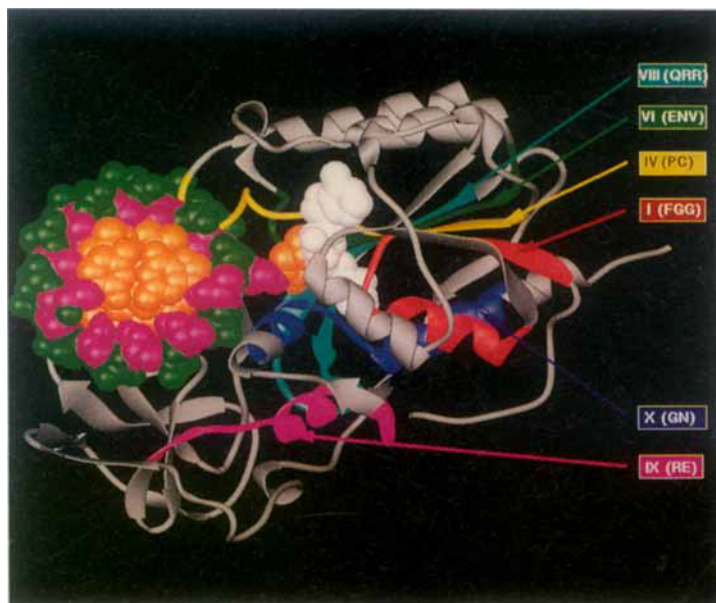


Abb. 7. Struktur von *M.HhaI* im ternären Komplex mit einem DNA-Duplex-Oligonucleotid als Substrat und dem Endprodukt der Reaktion, AdoHcy. Die Motive sind mit den gleichen Farben wie in Abbildung 2 gekennzeichnet. AdoHcy ist in weiß, die DNA-Basen in orange, die Desoxyribosen in pink und die Phosphatreste in grün dargestellt. a) Blick von der Seite auf die DNA-Achse. b) Blick von oben auf die DNA-Achse.

und viel eleganter. Das Zielcytosin gleitet aus der Achse der DNA-Helix heraus und in eine Tasche im Enzym hinein, die den katalytisch aktiven Cysteinrest enthält. Der Rest der Helix bleibt relativ unverzerrt. Eine weitere starke Konformationsänderung hatte in der Schleife des aktiven Zentrums stattgefunden, die Motiv IV enthält. Die gesamte, zwanzig Aminosäuren lange Schleife hatte sich um fast 180° aus ihrer ursprünglichen Position gedreht, um damit den katalytisch aktiven Cysteinrest in Kontakt mit dem Zielcytosin zu bringen (Abb. 8).

Die Position in der DNA-Helix, die ursprünglich vom Zielcytosin eingenommen wurde, wird nun von zwei Resten ausgefüllt.

Der eine ist ein Glutaminrest aus der kleinen Domäne und der andere ein Serinrest aus der Schleife des aktiven Zentrums. Erstaunlicherweise ist keiner dieser Reste bei den anderen m5C-MTasen konserviert, was darauf hindeutet, daß viele andere Reste ebenfalls dazu beitragen können, die DNA-Helix auf die geschilderte Weise zu öffnen. Wie aus Abbildung 7 zu entnehmen ist, kann man sich die Wechselwirkung zwischen *M.HhaI* und DNA so vorstellen, daß das Enzym die DNA „umarmt“, dabei in die Helix eindringt und so das Herausschieben des Cytosinrestes unterstützt.

An den sequenzspezifischen Wechselwirkungen zwischen *M.HhaI* und seinem Substrat sind im wesentlichen zwei Schleifen, die Sequenzerkennungsschleifen der kleinen Domäne, beteiligt. Eine Schleife wechselwirkt mit dem Guaninrest, der zurückbleibt, wenn der Cytosinrest aus der Helix gleitet, und nimmt spezifische Kontakte zu den benachbarten Basen des gleichen Stranges auf. Außerdem stellt sie den Glutaminrest, der die Lücke füllt, die der Cytosinrest hinterläßt. Die zweite Erkennungsschleife wechselwirkt hauptsächlich mit dem Strang, der das Zielcytosin enthält. Die Wechselwirkungen im kovalenten Komplex sollen jedoch als Endpunkt einer komplizierteren Wechselwirkung angesehen werden. Sie müssen daher die einleitenden Schritte, die zur Sequenzerkennung führen, nicht genau widerspiegeln. Genauer über diese einleitenden Schritte kann von Experimenten mit mutierten Proteinen oder mit DNA-Analoga erwartet werden, die Komplexe bilden können, ohne das Zielcytosin aus der Helix zu drängen. In diesem Zusammenhang sollte vermerkt werden, daß keines der gut charakterisierten DNA-bindenden Motive<sup>[50]</sup> in diesem System eine Rolle zu spielen scheint.

Frühere Studien über DNA-Protein-Wechselwirkungen ergaben ziemlich drastische Verzerrungen der DNA durch Proteine, die aber meistens in Verbiegen und Knicken bestanden; eine der stärksten DNA-Konformationsänderungen war dabei die Ab-

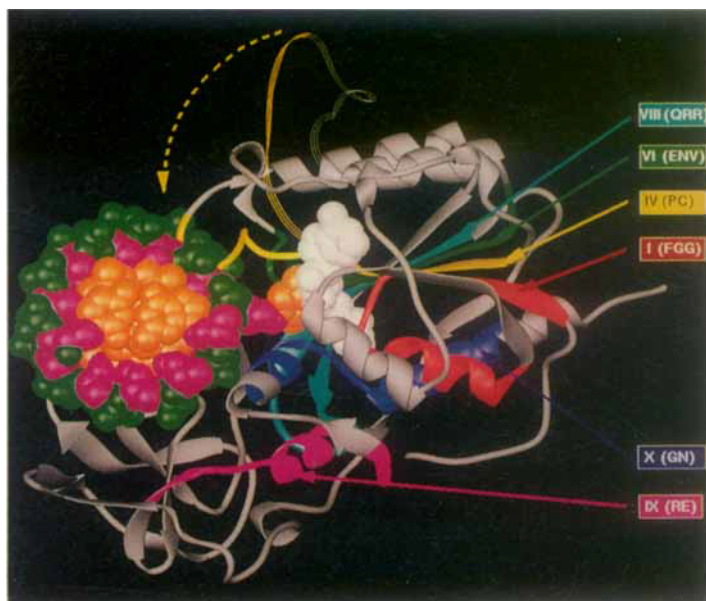


Abb. 8. Überlagerung der Strukturen von *M.HhaI* mit und ohne DNA, um zu verdeutlichen, wie sich die Konformation beim Binden von DNA ändert (Farben wie in Abb. 7). Die Position der Schleife des aktiven Zentrums (gelb) ist im ternären Komplex (mit DNA) als durchgehende Linie, im binären Komplex (ohne DNA) als Band aus drei Linien gezeigt.



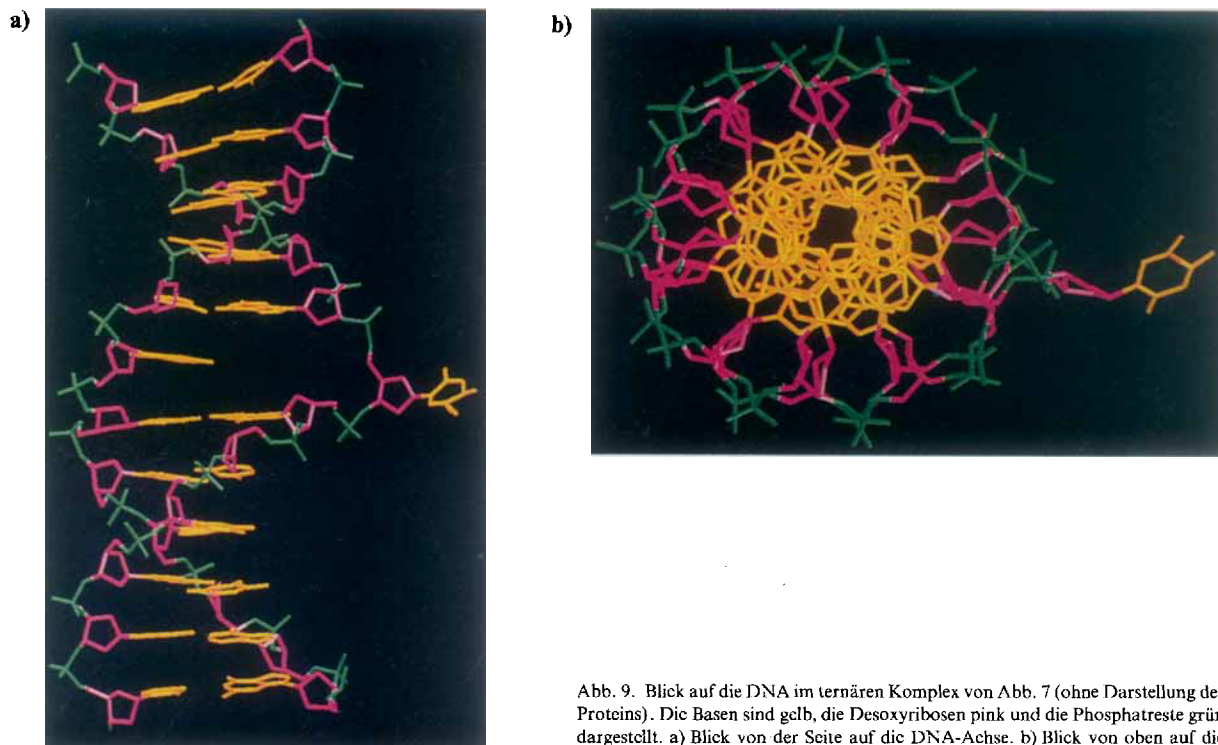


Abb. 9. Blick auf die DNA im ternären Komplex von Abb. 7 (ohne Darstellung des Proteins). Die Basen sind gelb, die Desoxyribosen pink und die Phosphatreste grün dargestellt. a) Blick von der Seite auf die DNA-Achse. b) Blick von oben auf die DNA-Achse.

flachung und Verbiegung, die durch das TATA-Box-bindende Protein hervorgerufen wurde<sup>[51, 52]</sup>. Es hatte jedoch keine Beispiele für eine DNA-Protein-Wechselwirkung gegeben, die das Herausgleiten einer Base aus der Helix verursachte. Da die Basen normalerweise im Inneren der DNA eingebettet sind, bietet der hier vorgestellte Mechanismus eine elegante Möglichkeit, die Basen vollständig zugänglich zu machen. Wir nehmen an, daß auch andere Proteine, die chemische Reaktionen an DNA-Basen durchführen, diesen Mechanismus nutzen. Naheliegende Kandidaten dafür sind die MTasen, die  $N^6$ -Methyladenin oder  $N^4$ -Methylcytosin bilden. Einige der Enzyme, wie die DNA-Glycosylasen, die DNA-Schäden reparieren, könnten die beschädigte Base ebenfalls aus der Helix schieben, bevor sie sie ausschneiden. Viele Proteine, die mit DNA wechselwirken, müssen die Helix öffnen; dazu gehören Topoisomerasen, Helicasen, DNA-Polymerasen und/oder ihre Hilfsproteine, die am Replikationsursprung angreifen, RNA-Polymerasen und Rekombinationsenzyme. In der Struktur in Abbildung 9 sieht man, daß die Phosphodiesterbindungen neben dem Zielcytosin gegenüber ihrer Position in normaler B-DNA verzerrt sind, und man kann sich vorstellen, daß eine weitere Öffnung der Helix von hier aus einfach sein sollte.

Überraschenderweise benötigt die Wechselwirkung zwischen *M.HhaI* und DNA keine externe Energiequelle. Konformationsänderungen innerhalb des Proteins und die spezifischen Wechselwirkungen zwischen Protein und DNA liefern vermutlich genügend Energie, um die Helix zu öffnen. Diese Energie wird dann gespeichert, um für die Rückkehr des Zielcytosins zur Verfügung zu stehen. Es wäre verwunderlich, wenn dieser Mechanismus nicht auch bei anderen Reaktionen verwendet würde. Unterbrochene Gene waren nicht erwartet worden, konnten jedoch leicht gefunden werden, als man einmal von ihrer Existenz wußte. Dieser Mechanismus, eine Base aus der DNA zu

schieben, könnte sich ebenfalls als wichtig und weitverbreitet erweisen als erster Schritt, eine DNA-Helix zu öffnen. Wir sollten keine Zeit verlieren, diese Möglichkeit zu untersuchen.

*Ich bin allen zu großem Dank verpflichtet, die in meiner Gruppe hart gearbeitet haben, um unseren Erfolg sicherzustellen. Nichts von dem Erreichten wäre ohne ihren Beitrag möglich gewesen. Außer denen, die ausdrücklich im Text genannt sind, danke ich Phyllis Myers, Kathy O'Neill, Margaret Wallace, Louise D'Alessandro, Jodie Freyer, Carol Marcincuk, Dana Macelis und Ching Lin, deren Einsatz besonders wichtig war, um mir das Leben zu erleichtern. Ich danke Jim Watson dafür, daß er mich überredet hat, nach Cold Spring Harbor zu gehen, und für viele anregende Gespräche. Von den vielen Kollegen waren mir vor allem Ashok Bhagwat, Mike Botchan, Don Comb, Greg Freyer, Rich Gelinas, Tom Gingeras, David Helfman, Winship Herr, Adrian Krainer, Stu Linn und Ira Schildkraut gute Freunde und wertvolle Quellen der Kritik und des Rats. Ich danke den National Institutes of Health und der National Science Foundation für viel Unterstützung im Laufe der Jahre. Schließlich bin ich meiner Frau Jean und meinen Kindern Alison, Andrew, Christopher und Amanda verpflichtet, die meiner Hingabe an die Wissenschaft hilfreich und verständnisvoll gegenübergestanden haben.*

Eingegangen am 17. Januar 1994 [A 48]  
Übersetzt von Dr. Burkard Neuß, Jülich

- [1] R. J. Roberts, *Crit. Rev. Biochem.* **1976**, *4*, 123-164.
- [2] C. Mulder, J. R. Arrand, H. Delius, W. Keller, U. Pettersson, R. J. Roberts, P. A. Sharp, *Cold Spring Harbor Symp. Quant. Biol.* **1974**, *39*, 397-400.
- [3] R. E. Gelinas, R. J. Roberts, *Cell* **1977**, *11*, 533-544.
- [4] L. T. Chow, R. E. Gelinas, T. R. Broker, R. J. Roberts, *Cell* **1977**, *12*, 1-8.
- [5] R. E. Gelinas, L. T. Chow, R. J. Roberts, T. R. Broker, D. F. Klessig, auf dem *Brookhaven Symposium in Genetic Interaction and Gene Transfer (Brookhaven Symp. Biol.* **1977**, *29*, 345-347).
- [6] T. R. Broker, L. T. Chow, A. R. Dunn, R. E. Gelinas, J. A. Hassell, D. F. Klessig, J. B. Lewis, R. J. Roberts, B. S. Zain, *Cold Spring Harbor Symp. Quant. Biol.* **1978**, *42*, 531-553.

- [7] B. S. Zain, J. Sambrook, R. J. Roberts, W. Keller, M. Fried, A. R. Dunn, *Cell* **1979**, *16*, 851–861.
- [8] R. J. Roberts, G. Akusjarvi, P. Alestrom, R. E. Gelin, T. R. Gingeras, D. Sciaky, U. Pettersson in *Adenovirus DNA. The Viral Genome and Its Expression* (Hrsg.: W. Doerfler), Martinus Nijhoff, Boston, MA, **1986**, S. 1–51.
- [9] T. R. Gingeras, J. P. Milazzo, R. J. Roberts, *Nucleic Acids Res.* **1978**, *5*, 4105–4127.
- [10] T. R. Gingeras, J. P. Milazzo, D. Sciaky, R. J. Roberts, *Nucleic Acids Res.* **1979**, *7*, 529–545.
- [11] T. R. Gingeras, P. I. Rice, R. J. Roberts, *Nucleic Acids Res.* **1982**, *10*, 103–114.
- [12] T. R. Gingeras, R. J. Roberts, *Science* **1980**, *209*, 1322–1328.
- [13] R. M. Blumenthal, P. R. Rice, R. J. Roberts, *Nucleic Acids Res.* **1982**, *10*, 91–101.
- [14] K. Keller, M. Corcoran, R. J. Roberts, *Nucleic Acids Res.* **1984**, *12*, 379–386.
- [15] A. Kiss, G. Pósfai, C. C. Keller, P. Venetianer, R. J. Roberts, *Nucleic Acids Res.* **1985**, *13*, 6403–6421.
- [16] A. S. Bhagwat, A. Sohail, R. J. Roberts, *J. Bacteriol.* **1986**, *166*, 751–755.
- [17] R. J. Roberts, S. S. Halford in *Nucleases*, (Hrsg.: S. M. Linn, R. S. Lloyd, R. J. Roberts), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, **1993**.
- [18] P. Modrich, *Annu. Rev. Genet.* **1991**, *25*, 229–253.
- [19] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915–926.
- [20] J. P. Jost, H. P. Saluz, *DNA Methylation: Molecular Biology and Biological Significance*, Birkhäuser, Basel, **1993**.
- [21] S. Kumar, X. Cheng, S. Klimasauskas, S. Mi, J. Pósfai, R. J. Roberts, G. G. Wilson, *Nucleic Acids Res.* **1994**, *22*, 1–10.
- [22] J. Pósfai, A. S. Bhagwat, G. Pósfai, R. J. Roberts, *Nucleic Acids Res.* **1989**, *17*, 2421–2435.
- [23] R. Lauster, T. A. Trautner, M. Noyer-Weidner, *J. Mol. Biol.* **1989**, *206*, 305–312.
- [24] D. V. Santi, P. V. Danenberg in *Folates and Pterins*, Vol. 1 (Hrsg.: R. L. Blakeley, S. J. Benkovic), Wiley-Interscience, New York, **1984**, S. 343–396.
- [25] J. C. Wu, D. V. Santi, *J. Biol. Chem.* **1987**, *262*, 4778–4786.
- [26] L. Chen, A. M. MacMillan, W. Chang, K. Ezaz-Nikpay, W. S. Lane, G. L. Verdine, *Biochemistry* **1991**, *30*, 11018–11025.
- [27] S. Friedman, N. Ansari, *Nucleic Acids Res.* **1992**, *20*, 3241–3248.
- [28] S. S. Smith, B. E. Kaplan, L. C. Sowers, E. M. Newman, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4744–4748.
- [29] M. W. Wyszynski, S. Gabbara, A. S. Bhagwat, *Nucleic Acids Res.* **1992**, *20*, 319–326.
- [30] S. Mi, R. J. Roberts, *Nucleic Acids Res.* **1993**, *21*, 2459–2464.
- [31] L. Chen, A. M. MacMillan, G. L. Verdine, *J. Am. Chem. Soc.* **1993**, *115*, 5318–5319.
- [32] D. Ingrosso, A. V. Fowler, J. Bleibaum, S. Clarke, *J. Biol. Chem.* **1989**, *264*, 20130–20139.
- [33] S. Klimasauskas, A. Timinskas, S. Menkevicius, D. Butkiene, V. Butkus, A. A. Janulaitis, *Nucleic Acids Res.* **1989**, *17*, 9823–9832.
- [34] M. Noyer-Weidner, S. Jentsch, B. Pawlek, U. Gunthert, T. A. Trautner, *J. Virol.* **1983**, *46*, 446–453.
- [35] T. S. Balgandesh, L. Reinert, R. Lauster, M. Noyer-Weidner, K. Wilke, T. A. Trautner, *EMBO J.* **1987**, *6*, 3543–3549.
- [36] K. Wilke, E. Rauhut, M. Noyer-Weidner, R. Lauster, B. Pawlek, B. Behrens, T. A. Trautner, *EMBO J.* **1988**, *7*, 2601–2609.
- [37] S. Klimasauskas, J. E. Nelson, R. J. Roberts, *Nucleic Acids Res.* **1991**, *19*, 6183–6190.
- [38] R. Y. Walder, C. J. Langtim, R. Catterjee, J. A. Walder, *J. Biol. Chem.* **1983**, *258*, 1235–1241.
- [39] M. B. Mann, H. O. Smith, *Nucleic Acids Res.* **1977**, *4*, 4211–4221.
- [40] S. Mi, R. J. Roberts, *Nucleic Acids Res.* **1992**, *20*, 4811–4816.
- [41] P. M. Lin, C. H. Lee, R. J. Roberts, *Nucleic Acids Res.* **1989**, *17*, 3001–3011.
- [42] A. K. Dubey, B. Mollet, R. J. Roberts, *Nucleic Acids Res.* **1992**, *20*, 1579–1585.
- [43] A. K. Dubey, R. J. Roberts, *Nucleic Acids Res.* **1992**, *20*, 3167–3173.
- [44] C. O. Card, G. G. Wilson, K. Weule, J. Hasapes, A. Kiss, R. J. Roberts, *Nucleic Acids Res.* **1990**, *18*, 1377–1383.
- [45] R. J. Roberts, P. A. Myers, A. Morrison, K. Murray, *J. Mol. Biol.* **1976**, *103*, 199–208.
- [46] M. Caserta, W. Zacharias, D. Nwankwo, G. G. Wilson, R. D. Wells, *J. Biol. Chem.* **1987**, *262*, 4770–4777.
- [47] a) J. C. Wu, D. V. Santi, *Nucleic Acids Res.* **1988**, *16*, 703–717; b) S. Kumar, X. Cheng, J. W. Pflugrath, R. J. Roberts, *Biochemistry* **1992**, *31*, 8648–8653.
- [48] X. Cheng, S. Kumar, J. Pósfai, J. W. Pflugrath, R. J. Roberts, *Cell* **1993**, *74*, 299–307.
- [49] S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **1994**, *76*, 357–369.
- [50] S. C. Harrison, *Nature* **1991**, *353*, 715–719.
- [51] Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature* **1993**, *365*, 512–520.
- [52] J. L. Kim, D. B. Nikolov, S. K. Burley, *Nature* **1993**, *365*, 520–527.