

GENOMFORSCHUNG

Alternatives Spleißen an NAGNAG-Motiven: kleine Ursache – große Protein-Vielfalt

Das alternative Spleißen ist ein molekularer Vorgang, der es ermöglicht, aus einem Gen mehrere Bauanleitungen für unterschiedliche Proteine herzustellen. Wissenschaftler sind nun auf eine bisher weitgehend unbeachtete Form des alternativen Spleißens gestoßen. Dabei können an Spleißstellen mit dem Sequenzmotiv NAGNAG zwei verschiedene Boten-Moleküle (mRNA) entstehen, die entweder das zweite NAG in der Eiweiß-Bauanleitung enthalten oder nicht. Überraschenderweise verfügen circa 30 Prozent unserer Gene über mindestens einen dieser „Vielfalt-Generatoren“. Die NAGNAG-Motive sind nicht zufällig im Genom verteilt. Da auch viele „Krankheitsgene“ betroffen sind, könnten Veränderungen des alternativen NAGNAG-Spleißens zu neuen diagnostischen und therapeutischen Ansätzen führen.

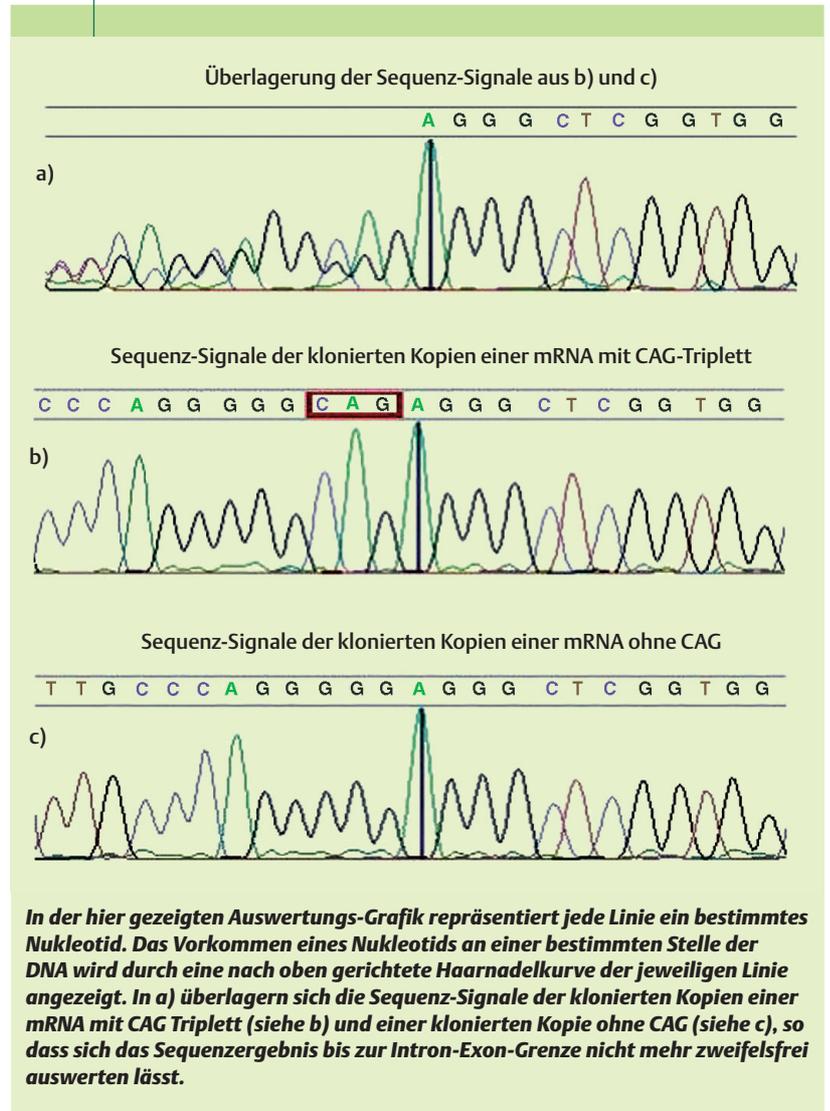
Die vielleicht größte Überraschung des Humanen Genom Projekts war die Entdeckung, dass der Mensch nur über circa 20.000 - 25.000 Gene verfügt [1]. Dies widerspricht scheinbar der geschätzten Anzahl von menschlichen Proteinen, die bei etwa 100.000 liegt: Vor noch nicht allzu langer Zeit ging man davon aus, dass ein Gen in der Regel nur ein Eiweiß codiert. Mittlerweile sind aber eine Reihe von Mechanismen bekannt, die die Herkunft verschiedener Proteine von einem Gen erklären und damit die alte Regel mehr und mehr zur Ausnahme machen.

Einen großen Beitrag zur Vergrößerung der Gesamtheit aller Proteine eines Organismus – des Proteoms – liefert das alternative Spleißen. Bei Eukaryoten muss die Synthesevorschrift für Eiweiße aus dem Zellkern zum Ort der Proteinbiosynthese im Cytoplasma transportiert werden. Die relevanten Sequenzen liegen im Zellkern aber gestückelt auf so genannten Exons vor. Durch den Prozess des „Spleißens“ werden die Exons miteinander verbunden, indem die sie unterbrechenden Sequenzen (Introns) herausgeschnitten werden. An den Nahtstellen von Exons und Introns gibt es Spleiß-Erkennungsstellen, die markieren, welche Teile der mRNA herausgeschnitten werden sollen. Genomisch beginnt der Spleiß-Erken-

nungscod am Anfang des Introns meist mit dem Dinukleotid GT und die Spleißstelle am Ende des Introns weist stets die Sequenz NAG auf. Das N repräsentiert einen der Buchstaben des genetischen Alphabetes (A, C, G oder T). Es gibt also vier mögliche Spleißmotive am Intron/Exon-Übergang (AAG, CAG, GAG und TAG).

Neue Schätzungen gehen davon aus, dass mehr als 75 Prozent der menschlichen Gene alternativ gespleißt werden. Findet alternatives Spleißen im codierenden mRNA-Abschnitt statt, führt das zur Synthese von unterschiedlichen Proteinen (Isoformen).

ABB. 1 SEQUENZIERUNG DER IN DNA UMGESCHRIEBENEN ALTERNATIVEN NAGNAG-SPLEISSPRODUKTE

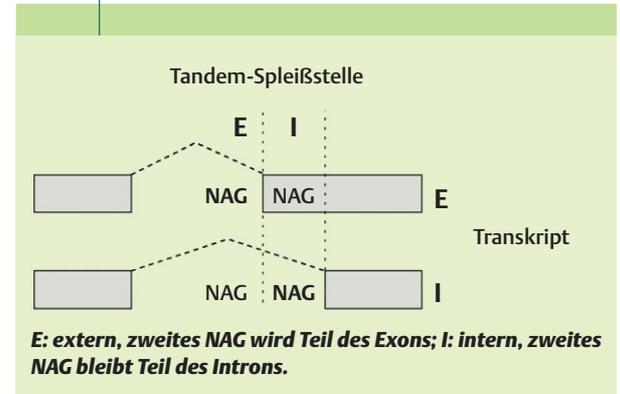


Kürzlich sind Wissenschaftler des Nationalen Genomforschungsnetzes (NGFN) und des Jenaer Centrums für Bioinformatik (JCB) auf eine bisher weitgehend unbeachtete Form der Nutzung von alternativen Spleißstellengesteßen [2]. Es begann damit, dass bei der Suche nach Gendefekten bei chronisch entzündlichen Darmerkrankungen (siehe dazu auch [3]) Wissenschaftler der Arbeitsgruppen von Matthias Platzer am Jenaer Institut für Molekulare Biotechnologie und von Stefan Schreiber von der Universität Kiel immer wieder mit ihren Ergebnissen unzufrieden waren. Die Sequenzierung der in DNA umgeschriebenen mRNA (= cDNA) von Patienten und Kontrollpersonen konnte von bestimmten Stellen an in vielen Proben nicht mehr eindeutig ausgewertet werden (Abbildung 1a). Nach Ausschluss aller möglichen labor- und gerätetechnischen Fehlerquellen stellte sich heraus, dass sich diese „schlechten“ Ergebnisse zweifelsfrei auf die Überlagerung der Signale von zwei mRNA-Molekülen zurückführen ließen (Abbildung 1b, c). Beide mRNAs unterschieden sich aber nur durch ein CAG Triplet. Nach dem Vergleich von mRNA- und genomischer Sequenz klärte sich das Phänomen weiter auf. Das alternative CAG lag unmittelbar neben einem weiteren CAG und beide bildeten eine Tandem-Spleißstelle CAGCAG! Soweit war das zwar interessant, aber noch nicht besonders aufregend, denn Kollegen hat ähnliches schon an anderen Genen beschrieben. Als diese Beobachtung jedoch auch an weiteren Kandidatengen für chronische Entzündungen gemacht wurde, stellte sich die Frage, ob es sich bei diesen Tandem um mehr als um ein sehr spezifisches und seltenes Motiv handelt.

Zur genomweiten bioinformatischen Analyse kam die Arbeitsgruppe von Rolf Backofen der Friedrich-Schiller-Universität Jena ins Boot. Gemeinsam untersuchten die Jenaer Molekulargenetiker und Bioinformatiker alle derzeit bekannten circa 20.000 für Proteine codierenden Gene des Men-

schen. Dabei stellte sich heraus, dass sich das beobachtete Phänomen nicht nur auf CAGCAG-Motive beschränkt, sondern auch bei allen anderen Spleiß-Motiven am Intron/Exon-Übergang vorkommen kann. Das C kann also durch jeden anderen Buchstaben des genetischen Alphabets (N=A, C, G oder T) ersetzt werden. Noch viel überraschender war, dass man nun diese NAGNAG-Sequenz bei circa 8000 Genen, d.h. in 30 Prozent aller menschlichen Gene an mindestens einer Intron/Exon-Grenze des codierenden Bereichs fand. Bei Vorliegen dieser Sequenz erkennt der Spleißapparat manchmal das erste NAG als Intronende, in anderen Fällen das zweite NAG (Abbildung 2). Durch dieses minimal alternative Spleißereignis verändert sich die Sequenz der mRNA in jedem Fall nur um drei Nucleotide.

ABB. 2 NOMENKLATUR DER NAGNAG-SPLEISSSTELLEN UND TRANSKRIPTE



Das weit verbreitete Vorkommen von Tandem-Spleißstellen in Genen erlaubt die Expression von zwei Proteinen, die sich in vielen Fällen nur durch eine Aminosäure unterscheiden (Abbildung 3a). Allerdings kommt es in anderen Fällen zum Aus-

ABB. 3 PROTEINVIELFALT DURCH ALTERNATIVES SPLEISSEN AN NAGNAG-MOTIVEN

	Intron zwischen zwei Codons	Intron nach dem ersten Nucleotid eines Codons	Intron nach dem zweiten Nucleotid eines Codons
a)	<p>CAG gt . . tag CAG GTG Q Q V</p> <p>CAG gt . . tagcag GTG Q V</p>	<p>CGA G gt . . tag TAGAT TTC R V D F</p> <p>CGA G gt . . tagtag AT TTC R D F</p>	<p>AAG AA gt . . tag TAGC AGG K N S R</p> <p>AAG AA gt . . tagtag C AGG K N R</p>
b)	nicht möglich	<p>GAC A gt . . cag CAGGA AAG D T G K</p> <p>GAC A gt . . cagcag GA AAG D R K</p>	<p>GCC AA gt . . cag C AGG CGT A N R R</p> <p>GCC AA gt . . cagcag G CGT A K R</p>
c)	<p>CAA gt . . cag TAG CAG Q *</p> <p>CAA gt . . cag tag CAG Q Q</p>	<p>AAT T gt . . tag TAGAA GCA N L E A</p> <p>AAT T gt . . tagtag AA GCA N *</p>	<p>TAC TG gt . . cag C AGA TGA Y C R *</p> <p>TAC TG gt . . cagcag A TGA Y *</p>

Die jeweils obere Zeile stellt einen Ausschnitt der Gensequenz am Exon/Intron-Übergang beziehungsweise am Intron/Exon-Übergang dar. Großbuchstaben repräsentieren Nucleotide, die nach dem Spleißen in der mRNA verbleiben. Klein geschriebene Buchstaben stellen Nucleotide dar, die aus der Protein-Bauanleitung herausgeschnitten werden. Je drei Gen-Buchstaben (Triplet) bilden den Code für eine Aminosäure. Die von den Triplets codierten Aminosäuren sind in der unteren Zeile durch Buchstaben symbolisiert (Q steht für die Aminosäure Glutamin, V steht für Valin usw.). Welche Protein-Varianten durch das alternative Spleißereignis entstehen, hängt zum einen von der Sequenz des Tandem-Motivs ab, und zum anderen davon, an welcher Triplettsposition das Intron eingefügt ist. Das Intron kann sich direkt zwischen zwei Triplet-Codons befinden (Spalte 1) oder es ist nach dem ersten beziehungsweise nach dem zweiten Nucleotid eines Codons eingefügt (Spalte 2 und Spalte 3). Drei Proteinvariationen sind auf diese Weise möglich: a) eine Aminosäure wird eingefügt oder fehlt, b) Austausch von einer Aminosäure durch zwei völlig verschiedene Aminosäuren, c) Translationsstopp wird eingefügt oder fehlt.

tausch von einer Aminosäure durch zwei völlig verschiedene Aminosäuren (Abbildung 3b). Als noch gravierenderes Ereignis kann durch alternatives Spleißen an NAGNAG-Motiven ein Stopp der Translation eingefügt werden, wodurch ein verkürztes Protein entsteht (Abbildung 3c). Obwohl das Ereignis auf Nukleotidebene recht simpel erscheint, ermöglichen die Tandem-Spleißstellen auf Proteinebene eine Vielzahl von Ausprägungen und vergrößern damit die Plastizität des Proteoms.

Tandem-Spleißstellen weisen eine Reihe von Besonderheiten auf. Zum einen sind sie nicht zufällig im humanen Genom verteilt, sondern vorzugsweise in Introns, die ein Codon nach dem ersten Nukleotid unterbrechen (vgl. Abbildung 3, Mitte). Zum anderen konnten die Wissenschaftler zeigen, dass vor allem Proteinbereiche, die Aminosäuren mit basischen oder sauren Seitenketten enthalten, betroffen sind. Dabei ist bemerkenswert, dass durch das alternative Spleiß-Ereignis an diesen Stellen wieder basische oder saure Aminosäuren eingefügt werden, so dass diese Proteinbereiche nur geringfügig verändert werden. Das wird als weiterer Hinweis interpretiert, dass diese Tandem-Spleißstellen genomisch nicht zufällig verteilt sind und dass dieses Spleißereignis in der Regel nur zu subtilen

Proteinveränderungen führt. Diese subtilen Veränderungen können trotzdem von Bedeutung sein. So sind schon Fälle bekannt, wo NAGNAG-Isoformen funktionelle Unterschiede aufweisen. Das ist insbesondere deshalb wichtig, weil viele dieser Proteine mit anderen Proteinen oder mit Nukleinsäuren Wechselwirkungen eingehen. Geringe Veränderungen dieser Wechselwirkungen könnten Auswirkungen auf komplexe Prozesse haben.

Interessanterweise konnten die Jenaer Wissenschaftler außerdem zeigen, dass das Spleißen an Tandem-Sequenzmotiven gewebespezifisch ist. Für einen Teil der untersuchten Gene wurden Gewebe gefunden, in denen ausschließlich das erste NAG, während in anderen ausschließlich das zweite NAG benutzt wird. Die Zellen haben also die Möglichkeit, nur eine der beiden Proteinformen herzustellen. Welche der beiden NAG-Spleißstellen vorrangig verwendet wird, hängt von der Art des Gewebes, vom Entwicklungsstadium und von Umweltfaktoren ab.

NAGNAG-Motive finden sich nicht nur im menschlichen Genom, sondern auch in anderen Spezies wie Maus oder Fruchtfliege. Zusätzlich sind viele Tandem-Spleißstellen zwischen Mensch und Maus konserviert, was wiederum auf eine funktionelle

Bedeutung dieser Spleißstellen hindeutet.

Das weit verbreitete Vorkommen des alternativen NAGNAG-Spleißens ist wahrscheinlich von weitreichender und grundsätzlicher Bedeutung für die Biologie von komplexen Organismen. Es könnte auch relevant für die Entwicklung und/oder den Verlauf von komplexen Erkrankungen wie Bluthochdruck, chronischen Entzündungen, Übergewicht und Neurodegeneration sein. Beispielsweise weisen Gene, deren Mutationen zu Fettsucht führen können, Tandem-Spleißstellenakzeptoren auf (Leptin, Leptinrezeptor, Ghrelin). Damit stehen die Wissenschaftler nun vor der Aufgabe, gleichzeitig sowohl den Mechanismus und die Regulation des NAGNAG-Spleißens aufzuklären, als auch zu untersuchen, ob Störungen beispielsweise im Fettstoffwechsel mit Störungen in Struktur und Funktion der Tandem-Spleißstellen in Verbindung gebracht werden können. Neue diagnostische und therapeutische Ansätze könnten am Ende dieses noch langen Weges stehen.

[1] IHGSC, *Nature* **2004**, 431, 931-945.

[2] M. Hiller, *Nature Genet.* **2004**, 36, 1255-1257.

[3] M. Stoll, *Biol. Unserer Zeit* **2004**, 4, 208-209.

Matthias Platzer,

Klaus Huse, Michael Hiller, Jena

PROJEKTE DES NGFN: SYSTEMATISCHE GENOMFORSCHUNG



Im Nationalen Genomforschungsnetz (NGFN) wurden insgesamt elf Systematisch-Methodische Plattformen (SMPs) eingerichtet, um umfassend das menschliche Genom, die Aktivität der Gene sowie Struktur und Funktion der Proteine zu untersuchen.

So wird beispielsweise von Wissenschaftlern der SMP „DNA“ erforscht, welche regulatorischen Prozesse auf DNA-Ebene krankheitsrelevant sind. Ein Schwerpunkt ist hier die Analyse von Promotoren. Diese in unserem Erbgut enthaltenen DNA-Sequenzen steuern, welche Gene in welchen Geweben, zu welcher Zeit und in welchem Krankheitsstadium abgelesen werden. Ein Teilprojekt der Plattform beschäftigt sich mit der Frage, welche Erbgut-Veränderungen das Alternative Spleißen der mRNA beeinflussen. Störungen bei diesen Prozessen könnten mit bestimmten Erkrankungen in Verbindung stehen.

Um die funktionellen Folgen von krankheitsrelevanten Gen-Veränderungen zu verstehen, wurden die SMPs „Tiermodelle für menschliche Krankheiten“ und „RNAi“ etabliert. Die Wissenschaftler dieser beiden Plattformen verfolgen zwei unterschiedliche Ansätze – die chemisch-induzierte Mutagenese beziehungsweise die RNA-Interferenz –, um die Funktion einzelner Gene auszuschalten oder zu verändern. Mit diesen beiden Methoden werden die biologischen Systeme an ganz spezifischen Stellen gestört und man kann Rückschlüsse auf die Funktion des Gens und dessen Rolle bei Krankheitsprozessen ziehen.

Informationen zu weiteren SMPs und allgemeine Informationen zum Nationalen Genomforschungsnetz stehen im Internet unter www.ngfn.de sowie unter www.ngfn.de/science.