

# Die Variabilität der Plastizität

**Proteinvialt durch subtiles alternatives Spleißen an NAGAG-Motiven! Wissenschaftler des Nationalen Genomforschungsnetzes (NGFN) und des Jenaer Centrums für Bioinformatik (JCB) sind auf eine bisher weitgehend unbeachtete Form der Nutzung von alternativen Spleißstellen gestoßen**

**Matthias Platzer, Michael Hiller, Klaus Huse, Karol Szafranski, Jochen Hampe, Rolf Backofen, Stefan Schreiber (JemalFreiburg/Kiel)**



Die vielleicht größte Überraschung des Humanen Genom Projekts war die Entdeckung, dass der Mensch nur über circa 20.000-25.000 Protein-codierende Gene verfügt (IHGSC 2004), d. h. über nur wenig mehr als der Fadenwurm *C. elegans* und die Fruchtfliege *D. melanogaster* und sogar über weniger als die Ackerschmalwand *A. thaliana*. 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.

## Alternatives Spleißen als Quelle der Proteomkomplexität

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 Zytoplasma 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-Erkennungscode am Anfang des Introns meist mit dem Dinukleotid GT und die Spleißstelle am Ende des Introns weist fast immer die Sequenz AG auf. Neue Schätzungen gehen davon aus, dass mehr als 75 Prozent der menschlichen Gene alternativ gesplissen werden (Johnson et al. 2003). Dies geschieht durch Überspringen von Exons,

Benutzen von alternativen Donoren oder Akzeptoren und Einhalten von Introns. Alternatives Spleißen erfolgt oftmals in Abhängigkeit vom Gewebe, Zelltyp und/oder Entwicklungsstadium. Das Spleißverhalten verändert sich auch in Abhängigkeit von externen Stimuli (z.B. Stress, Hitzeschock, Ionenkonzentrationen). Alternatives Spleißen spielt eine wichtige Rolle in vielen biologischen Prozessen, z.B. der neuronalen Entwicklung, der Geschlechtsbestimmung bei der Fruchtfliege oder der Gewebedifferenzierung. Findet alternatives Spleißen im codierenden mRNA-Abschnitt statt, führt das zur Synthese von unterschiedlichen Proteinen (Isoformen).

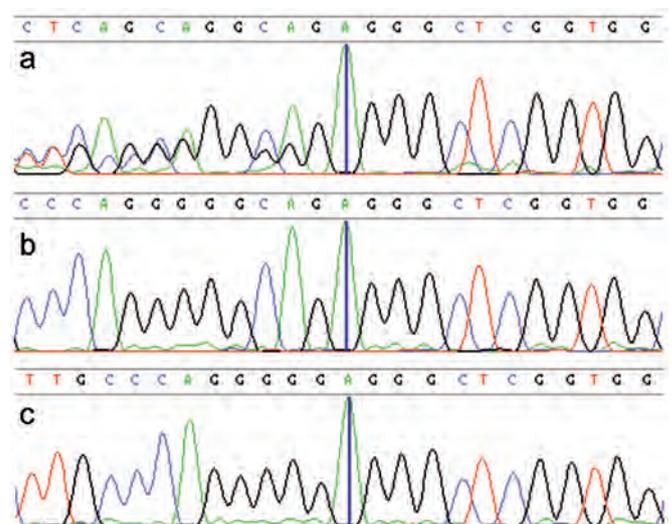
## Kleine Ursache – große Protein-Vielfalt

Wissenschaftler des Nationalen Genomforschungsnetzes (NGFN) und des Jenaer Centrums für Bioinformatik (JCB) sind auf eine bisher weitgehend unbeachtete Form der Nutzung von alternativen Spleißstellen gestoßen (Hiller et al. 2004). Es begann damit, dass bei der Suche nach Gendefekten bei chronisch entzündlichen Darmerkrankungen die Sequenzierung der in DNA umgeschriebenen mRNA

(=cDNA) von Patienten und Kontrollpersonen 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 NAG Triplet (N steht für A, C, G oder T). Nach dem Vergleich von mRNA- und genomischer Sequenz klärte sich das Phänomen weiter auf. Das alternative NAG lag unmittelbar neben einem weiteren NAG und beide bildeten eine Tandem-Spleißstelle NAGNAG (Abbildung 2)!

Genomweiten bioinformatischen Analyse zeigten, dass man nun diese NAGNAG-Sequenz bei ca. 8.000 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. Durch dieses minimal alternative Spleißereignis verändert sich die Sequenz der mRNA in jedem Fall nur um drei Nukleotide. Das weit verbreit-

Abb. 1: Sequenzierung der in DNA umgeschriebenen alternativen NAGAG-Spleißprodukte. In (a) überlagern sich die Sequenz-Signale des alternativen Transkripts mit CAG Triplet (b) und ohne CAG (c), so dass sich das Sequenzierungsergebnis bis zur Exon-Grenze nicht zweifelsfrei auswerten lässt.



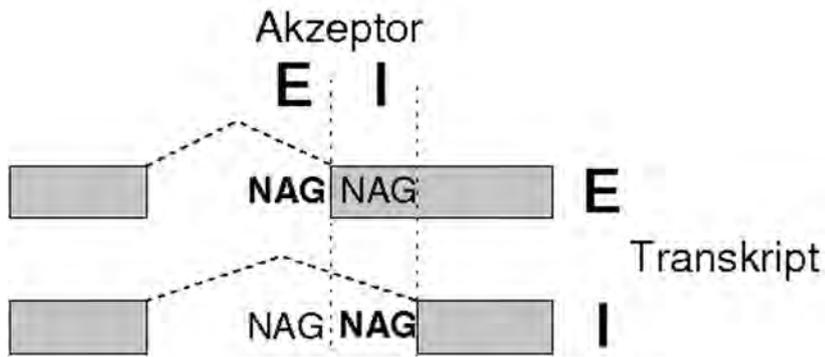


Abb. 2: Nomenklatur der NAGNAG-Spleißstellen und Transkripte. E: extern, zweites NAG wird Teil des Exons; I: intern, zweites NAG bleibt Teil des Introns.

tete 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 Austausch von einer Aminosäure durch zwei völlig verschiedene Aminosäuren (Abbildung 3b). Als noch gravierendes 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 Vielfalt von Ausprägungen und vergrößern damit die Plastizität des Proteoms.

### **NAGNAG-Akzeptoren sind nicht zufällig im Genom verteilt**

Diese Tandem-Spleißstellen weisen eine Reihe von Besonderheiten auf. Zum einen sind sie vorzugsweise in Introns, die ein Codon nach dem ersten Nucleotid unterbrechen (Abbildung 3, Mitte). Zum anderen konnten wir 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 gemischt 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 von Bedeutung, da viele dieser Proteine mit ande-

ren Proteinen oder mit Nucleinsäuren Wechselwirkungen eingehen. Geringe Veränderungen dieser Wechselwirkungen könnten Auswirkungen auf komplexe Prozesse haben.

Interessanterweise konnte außerdem gezeigt werden, 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-Akzeptoren sind auch phylogenetisch weit verbreitet**

NAGNAG-Motive finden sich nicht nur im menschlichen Genom, sondern auch in anderen Spezies wie Maus oder Fruchtfliege. Zusätzlich sind viele Tandem-Spleißstellen des Menschen in der Maus vorhanden und machen etwa die Hälfte aller konservierten alternativen Akzeptoren aus. Diese Fakten sprechen für eine biologische Relevanz der beschriebenen subtilen Form alternativen Spleißens.

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, chronische Entzündungen, Übergewicht und Neurodegeneration sein. Beispielsweise weisen Gene, deren Mutationen zu Fettsucht führen können, Tandem-Spleißstellen auf (Leptin, Leptinrezeptor, Ghrelin).

### **Einzelbasen-Variationen als Modulatoren des alternativen Spleißens**

Fehlerhafte oder veränderte Spleißmuster sind die Ursache für zahlreiche menschliche Krankheiten. Aus diesem Grund haben wir kürzlich das menschliche Erbgut systematisch nach Variationen (SNP, „single nucleotide polymorphism“) an diesen NAGNAG-Spleißstellen in der menschlichen Population durchsucht (Hiller *et al.* 2006). Ergebnis: bei manchen Menschen kann an bestimmten Positionen im Erbgut alternativ gespleißt werden, bei anderen aber nicht. Die identifizierten SNPs zerstören entweder einen bekannten NAGNAG-Akzeptor, erschaffen ein neues NAGNAG oder verändern ein 'N' in einem bekannten NAGNAG-Akzeptor. SNPs, die in diese drei Gruppen fallen, können für das Spleißen an Tandemakzeptoren relevant sein. Deshalb haben wir ihre statistischen, evolutionsbiologischen und funktionellen Charakteristika bestimmt (Intronphase, Aminosäureereignisse, Konservierung in anderen Spezies, Aminosäureverteilung an den Exongrenzen, Pfam-Domänenverteilung, Auftreten der Proteine in Proteininteraktionsdatenbanken).

An ausgewählten Beispielen haben wir anschließend die Veränderung im NAGNAG-Spleißen experimentell verifiziert. Da NAGNAG-zerstörende Allele „natürliche“ „Knock-out“-Experimente darstellen, konnten wir daraus ableiten, dass das NAGNAG-Motiv für das von uns untersuchte Phänomen notwendig ist. Durch Vergleich des genomischen Kontextes zum Schimpansen wurde das wahrscheinlich ancestrale Allel bestimmt und gefunden, dass die neu entstandenen NAGNAG-Akzeptoren dem gleichen Selektionsdruck unterliegen müssen, den wir auch schon für nicht-polymorphe, in der menschlichen Spezies fixierte Tandemakzeptoren beobachtet hatten (Hiller *et al.* 2004). Diese NAGNAG-generierenden Allele sind zudem der Beweis, dass im Umfeld eines Standardakzeptors die Entstehung eines NAGNAG-Motivs auch hinreichend für das Auftreten von entsprechenden alternativen Spleißformen ist. Da 28% der von uns identifizierten NAGNAG-SNPs in bekannten „Krankheitsgenen“ liegen, sind diese besonders interessant für weiterführende funktionelle Untersuchungen. Außerdem erscheint es bemerkenswert, dass eine Reihe der SNPs im I-Akzeptor, die gleichzeitig auch die Sequenz des E-Proteins verändern, im heterozygoten Status zur Expression von drei Proteinisoformen führen (siehe Abbildung 4). Damit gewinnt die Komplexität des Proteoms

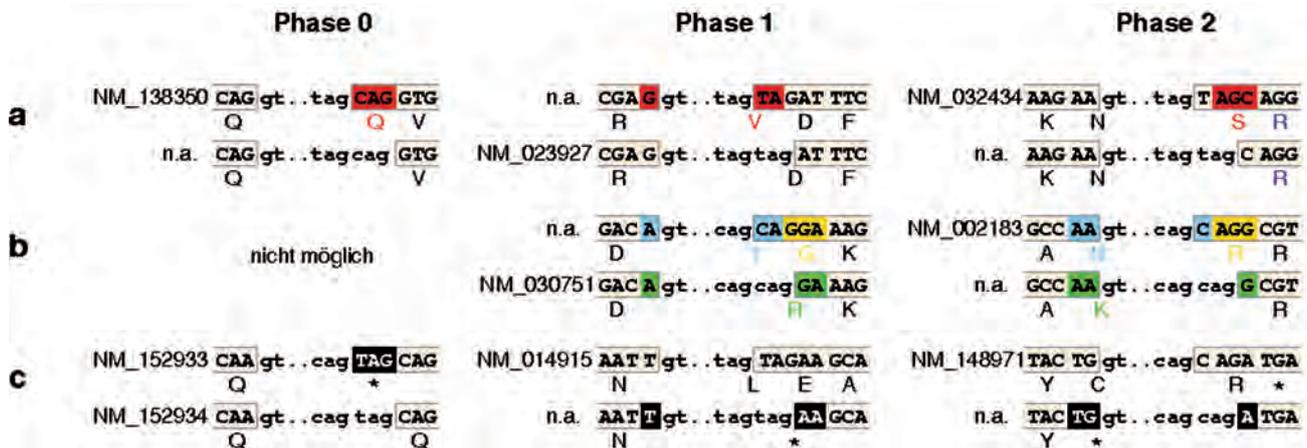


Abb. 3: Wie durch alternatives Spleißen an NAGNAG-Motiven Proteinviefalt entsteht: Die jeweils obere Zeile stellt einen Ausschnitt der Gensequenz am Exon/Intron-Übergang bzw. am Intron/Exon-Übergang dar. Großbuchstaben repräsentieren Nukleotide, die nach dem Spleißen in der mRNA verbleiben. Klein geschriebene Buchstaben stellen Nukleotide dar, die aus der Protein-Bauanleitung herausgeschnitten werden. Je drei Gen-Buchstaben (Triplet) bilden den Code für einen Eiweiß-Baustein (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 Triplettposition das Intron eingefügt ist. Das Intron kann sich direkt zwischen zwei Triplet-Codons befinden (Phase 0) oder es ist nach dem ersten bzw. nach dem zweiten Nukleotid eines Codons eingefügt (Phase 1 bzw. Phase 2). 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.

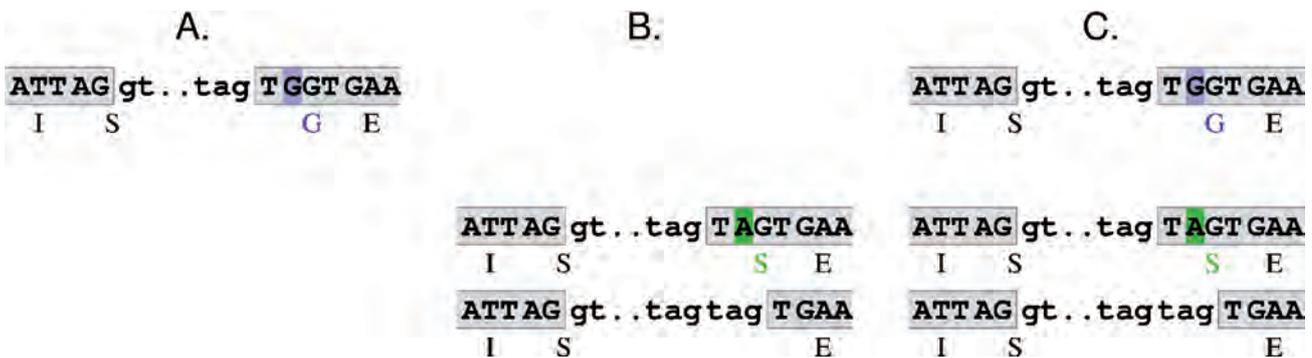


Abb. 4: Beispiel eines SNP, der den I-Akzeptor und die Sequenz des E-Proteins verändert (rs2275992 in ZFP91). Das homozygotes G-Allel ohne NAGNAG führt zur Expression nur eines Proteins (A), Homozygotie des A-Allel mit NAGNAG gestattet die Expression von zwei (B) und Heterozygotie von drei Isoformen (C). Großbuchstaben repräsentieren Nukleotide der Exons, klein geschriebene Buchstaben die von Introns. Die von den Triplets codierten Aminosäuren sind unter der zweiten Codonposition durch Buchstaben symbolisiert.

einen weiteren Freiheitsgrad und bietet die Möglichkeit für die Herausbildung eines Heterozygotenvorteils. Vor dem Hintergrund des generellen Mangels an verlässlichen bioinformatischen Methoden zur Vorhersage von spleißrelevanten SNPs, ist unser, auf NAGNAG-Motive bezogener, heuristischer Ansatz hoch effizient bei der Identifikation von Variationen im Spleißprozess.

Damit besteht die Aufgabe, gleichzeitig sowohl den Mechanismus und die Regulation des NAGNAG-Spleißens aufzuklären, als auch zu untersuchen, ob Störungen z.B. 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.

**Literatur**

- IHGSC 2004, Finishing the euchromatic sequence of the human genome. Nature 431:931-945
- Johnson et al. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302: 2141-2144

**Originalveröffentlichungen**

- Hiller et al. 2004, Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. Nature Genet 36:1255-1257

- Hiller et al. 2006, Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. Am J Hum Genet 78: 291-302

**Kontakt**

Matthias Platzer  
Leibniz-Institut für Altersforschung – Fritz-Lipmann-Institute (FLI/ehemals IMB)  
mplatzer@fli-leibniz.de

Michael Hiller  
Albert-Ludwigs-Universität, Freiburg  
hiller@informatik.uni-freiburg.de