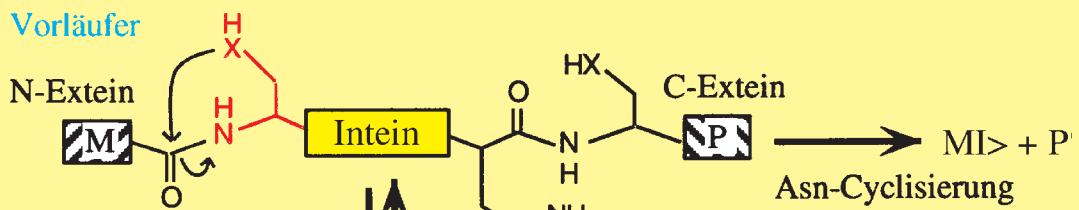
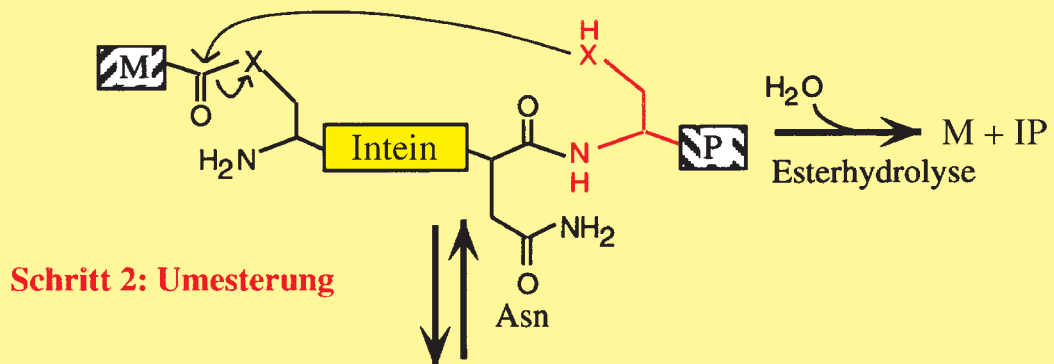


Der gegenwärtig akzeptierte Mechanismus des Protein-Spleißens



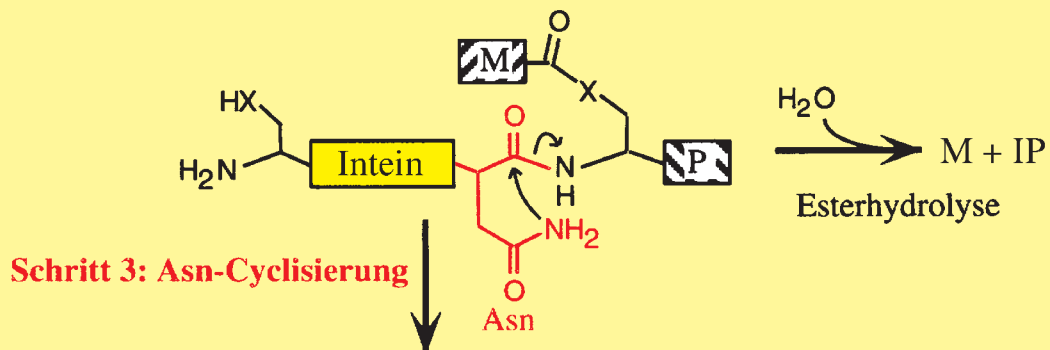
Schritt 1: N-X-Acyl-wanderung

lineares Esterintermediat

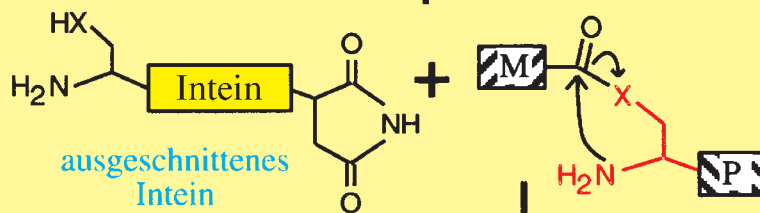


Schritt 2: Umesterung

verzweigtes Intermediat



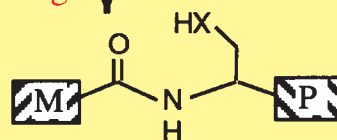
Schritt 3: Asn-Cyclisierung



ausgeschnittenes Intein

Schritt 4: X-N-Acyl-wanderung

verknüpfte Exteine



Protein-Spleißen: Mechanismus und Anwendungen

Christopher J. Noren, Jimin Wang und Francine B. Perler*

In memoriam Holger W. Jannasch

Inteine sind interne Polypeptidsequenzen, die posttranslational aus einer Proteinvorstufe durch eine autokatalytische Spleißreaktion entfernt werden. Die ca. 150 Aminosäuren lange Domäne, die das Spleißen katalysiert, löst die Spaltung der Peptidbindungen an den N- und C-terminalen Spleißstellen und die gleichzeitige Bildung einer neuen Peptidbindung zwischen den flankierenden Polypeptiden (Exteinen) aus. Das Intein, verlängert um eine Aminosäure vom C-Extein, kann zwei Peptidbindungen brechen, eine neue Peptidbindung bilden, DNA spalten und die Mobilisierung des Inteingens auslösen. Fast 100 Inteine sind bereits identifiziert worden; sie tragen Ser oder Cys an der N-terminalen und die Triade His-(Asn/Gln)-(Ser/Thr/Cys) an der C-terminalen Spaltstelle. Die konservierten Reste sind bekannt-

lich an gut untersuchten chemischen Reaktionen anderer Enzyme beteiligt, so dass ein Sammelurteil von Reaktionsmechanismen für das Spleißen vorgeschlagen wurde. In diesem Aufsatz beschränken wir uns auf die drei Jahre, in denen mit einer Kombination unterschiedlicher experimenteller Ansätze die Schritte der Spleißreaktion bei Proteinen aufgeklärt wurden, wodurch viele der vorgeschlagenen Mechanismen widerlegt und andere gestützt wurden. Der Mechanismus des Protein-Spleißens verläuft über vier gekoppelte nucleophile Substitutionen: 1) eine N→O(S)-Acylverschiebung am Ser/Cys des N-Terminus des Inteins; 2) eine Umesterungsreaktion zu einem verzweigten Zwischenprodukt mit zwei N-Termini; 3) die Cyclisierung des C-terminalen Asn/Gln unter Freisetzung des Inteins und 4) eine

O(S)→N-Acylverschiebung von Ser/Thr/Cys unter Bildung einer nativen Peptidbindung zwischen den Exteinen. Wie das Intein diese Reaktionen auf molekularer Ebene erleichtert, beginnt man gerade erst zu verstehen. Dies mündete in die Entwicklung verschiedener Intein-vermittelter Protein-Engineering-Anwendungen, wie Proteinreinigung, Anknüpfen fluoreszierender Biosensoren an Proteine, Expression cytotoxischer Proteine, Protein-Semisynthese und segmentweise Markierung von Proteinen für die NMR-Spektroskopie. Inteine in pathogenen Mikroorganismen stellen neue Zielstrukturen für die Wirkstoffsuche dar.

Stichwörter: Autokatalyse • Bioorganische Chemie • Peptide • Protein-Spleißen • Umesterungen

1. Einleitung

Anfangs dachte man, die Proteinbiosynthese sei ein einfacher Prozess, bei dem die genetische Information aus der DNA direkt in die Messenger-RNA (mRNA) umgeschrieben werde (Transkription), welche ihrerseits die Biosynthese des Proteins steuere (Translation). Allerdings gibt es viele Gene, in denen interne oder terminale Aminosäuren codiert sind, die im reifen Protein nicht mehr vorkommen. Wenn diese

Sequenzen vor der Translation der reifen mRNA aus der RNA entfernt werden, nennt man den Vorgang RNA-Spleißen und die entfernten Sequenzen heißen „Introns“ (Schema 1B). Die posttranslationale Entfernung von Polypeptidsegmenten kann durch Protein-Spleißen, Autoproteolyse oder proteolytische Spaltung geschehen. Beim Protein-Spleißen wird ein spezifischer Typ einer internen Sequenz, Intein genannt,^[1] aus einer im Inneren eines Vorläuferproteins liegenden Stelle ausgeschnitten und die beiden angrenzenden Polypeptide (die Exteine) werden zum reifen Protein verknüpft (Schema 1A). Durch das Protein-Spleißen entstehen zwei stabile Proteine ausgehend von einem einzigen Gen: das ausgeschnittene Intein und die verknüpften Exteine. Dies ist direkt analog zum RNA-Spleißen der Introns.

Inteine gehören zu einer wachsenden Familie autokatalytischer enzymähnlicher Proteine, die chemische Reaktionen nach den gleichen Strategien wie Enzyme katalysieren, aber

[*] Dr. F. B. Perler, Dr. C. J. Noren
New England Biolabs, Inc.
32 Tozer Road, Beverly, MA 01915 (USA)
Fax: (+1) 978-921-1350
E-mail: perler@neb.com
Dr. J. Wang
Department of Molecular Biophysics & Biochemistry
Bass Center for Structural Biology, Yale University
New Haven, CT (USA)

nicht unterschiedliche Substrate umsetzen. Andere Beispiele für Autoprozessierung findet man bei der autokatalytischen Spaltung von Vorstufen von N-terminal nucleophil angreifenden Amidohydrolasen (Glycosyl-Asparaginase, Penicillin-Acylase, Proteasom etc.), bei der Bildung von Pyruvoyl-enzymen, beim Prozessieren von embryonalen Hedgehog-Signalproteinen, bei der RecA-unterstützten autokatalytischen Spaltung des LexA-Proteins oder des Lambda-Repressors sowie bei der Aktivierung von Plasmaproteinen.^[2, 3] All diese Prozesse einschließlich des Protein-Spleißens werden durch einen nucleophilen Angriff der aktivierten Thiol- oder Hydroxygruppe der Seitenkette von Cys bzw. Ser oder Thr auf eine Amidbindung eingeleitet. Protein-Spleißen ist komplexer als die anderen autokatalytischen Reaktionen, da dazu eine schnelle sequentielle und räumliche Kopplung von vier separaten nucleophilen Substitutionen erforderlich ist.

Der Spleißvorgang läuft nach einem intramolekularen Mechanismus ohne exogene Cofaktoren ab.^[4, 5] Alle für die Spaltung notwendigen Informationen sind in der Polypeptidkette des Inteins und der ersten Extein-Aminosäure an der C-terminalen Spleißstelle enthalten.^[4–10] Ein bestimmtes Intein kann sich, wenn auch mit unterschiedlicher Effizienz, aus verschiedenen flankierenden Peptidsequenz-Umgebungen herauschneiden, wenn es sich im Polypeptid vor seinem zugehörigen C-terminalen Extein-Aminosäurerest befin-

det.^[4–10] Um Verwirrungen zu vermeiden,^[11] wurde der Ausdruck „protein-splicing element“ (Protein-spleißendes Element) benutzt, um das zum Spleißen befähigte Intein zusammen mit den C-terminalen Aminosäuren Ser, Thr oder Cys des zugehörigen Exteins zu bezeichnen, da das Intein allein das Protein-Spleißen nicht vermitteln kann.

Eine Reihe von Anwendungen wurde entwickelt, die das umgebungsunabhängige Herausschneiden der Inteine nutzen, darunter die auf dem Spleißen beruhende Proteinsynthese,^[12, 13] selbstspaltende Affinitätsmarker für die Proteinreinigung,^[7, 14–16] ein neues Polypeptid-Verknüpfungssystem für die Protein-Semisynthese,^[15, 17–20] die segmentweise Markierung von Proteinen für die NMR-Spektroskopie^[12, 20, 21] und die Verknüpfung fluoreszierender Biosensoren mit Proteinen.^[17, 22] Die Entwicklung dieser Systeme erforderte unter anderem die genetische und biochemische Manipulation der Protein-spleißenden Elemente, orientiert am bekannten Mechanismus des Spleißvorgangs. In diesem Aufsatz werden wir uns auf die schrittweise Aufklärung des Mechanismus der Spleißreaktion und ihre Modulation zur Entwicklung der neuen Anwendungen konzentrieren. Außerdem werden wir die neuesten Strukturinformationen zu den bekannten mechanistischen Daten in Beziehung setzen und so versuchen, die funktionalen Erfordernisse für jeden Schritt der Spleißreaktion herauszuarbeiten.

Francine B. Perler, geboren in Brooklyn, New York, erhielt 1972 ihren BS-Abschluß in Biologie am Brooklyn College der City University of New York. 1977 promovierte sie an der University of North Carolina, Chapel Hill, bei Clyde A. Hutchinson III zum Ph.D. in Genetik mit einer Arbeit zur Klonierung von ribosomalen RNA-Genen. Anschließend arbeitete sie als Postdoktorandin am Roche Institute of Molecular Biology und dann am Biochemistry Department der Harvard Medical School, wo sie bei Dr. Argiris



F. B. Perler



C. J. Noren



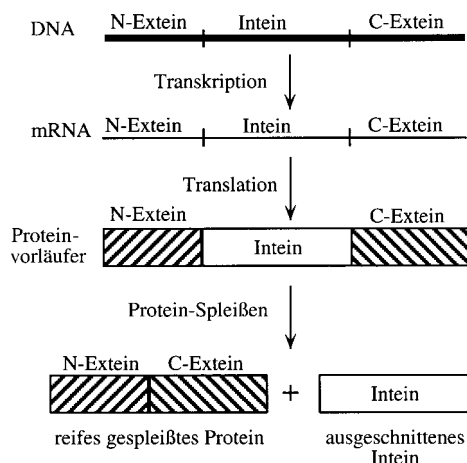
J. Wang

Efstratiadis das RNA-Spleißen untersuchte. Seit 1980 ist sie Senior Research Scientist bei New England BioLabs, wo sie die Molekularbiologie von Filaria und Enzyme aus ultrathermophilen Organismen untersuchte, vor allem die Vent-DNA-Polymerase. 1989 begann sie mit Untersuchungen zum Protein-Spleißen, ihrem derzeitigen Hauptarbeitsgebiet.

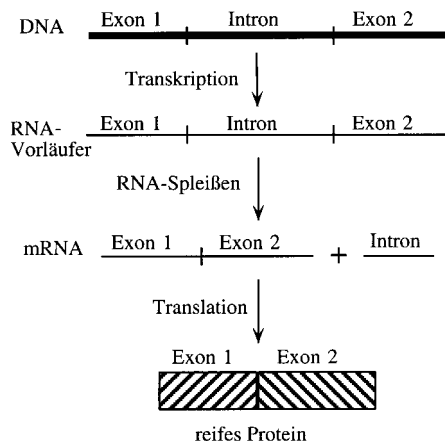
Christopher J. Noren, geboren in Manasquan, New Jersey, erhielt 1984 seinen BS-Abschluß in Chemie am Massachusetts Institute of Technology. 1990 promovierte er zum Ph.D. in Chemie an der University of California, Berkeley, wo er an der Entwicklung der Mutagenesemethode zur Einführung nichtnatürlicher Aminosäuren in Proteine im Labor von Peter G. Schultz beteiligt war. Seit 1990 arbeitet er bei New England BioLabs, seit 1993 als Senior Scientist. Seine Forschungsinteressen umfassen die Untersuchung des Mechanismus des Protein-Spleißens und die Entwicklung einer Reihe innovativer Phagen-Display-Methoden.

Jimin Wang, geboren in Luqiao, Zhejiang, Volksrepublik China, erhielt 1982 seinen BS-Abschluß in Chemie an der Universität Peking mit Arbeiten zur Übergangsmetallkatalyse. 1988 promovierte er an der University of California, San Diego, bei Professor Joseph Kraut mit kristallographischen Untersuchungen der Cytochrom-c-Peroxidase zum Ph.D. in Chemie. Danach wechselte er zur Yale University, Department of Molecular Biophysics and Biochemistry, in die Arbeitsgruppe von Professor Thomas A. Steitz, wo er als Research Scientist arbeitet. Seine Forschungsgebiete umfassen die Biochemie und Strukturbiologie von DNA-Replikation, Transkription, Translation, Proteinfaltung und Proteinabbau.

A) Protein-Spleißen:



B) RNA-Spleißen:



Schema 1. Vergleich zwischen dem RNA-Spleißen zur Entfernung der Introns (A) und dem Protein-Spleißen zur Entfernung der Inteine (B).

1.1. Inteine spleißen als Proteine, nicht als RNA

1990 beobachteten zwei Arbeitsgruppen unabhängig voneinander, dass das offene Leseraster für das Gen der vakuolären ATPase aus *Saccharomyces cerevisiae* (Sce VMA) beinahe doppelt so lang war wie sein Proteinprodukt und dass es eine Sequenz aus 454 Aminosäureresten enthielt, die mit der anderer vakuolärer ATPasen keine Ähnlichkeit hatte.^[23, 24] Es erschien unwahrscheinlich, dass diese eingefügte Sequenz ein Intron war, denn sie hatte keinerlei Ähnlichkeit mit bekannten Introns und befand sich im gleichen Leseraster wie das vorhergesagte reife Protein – eine Vorbedingung, die für das RNA-Spleißen unnötig ist. Bis 1993 wurden fünf weitere Insertionen des gleichen Typs in anderen Genen gefunden.^[4, 25–27] Die schnelle Prozessierung dieser nativen Vorstufen in ihren natürlichen Wirtsorganismen oder in *Escherichia coli* ließ nur den indirekten Nachweis zu, dass die Inteine-Entfernung auf Proteinebene geschah. Zur Modifizierung des zentralen Dogmas der Genexpression waren jedoch direkte Nachweise notwendig. In verschiedenen Ver-

suchsansätzen scheiterte der Nachweis gespleißter RNA.^[10, 23, 24, 27, 28] Stille Mutationen, die vermutete essentielle Intronsequenzen, nicht aber die Proteinsequenz verändern, hatten keinen Einfluss auf das Spleißen,^[26] während eine Störung des Intein-Leserasters das Spleißen blockierte.^[10, 26, 27] Der Durchbruch kam 1993, als Wissenschaftler der New England BioLabs das erste In-vitro-Spleißsystem entwickelten. Damit wurde die Reinigung einer aktiven ungespleißten Proteinvorstufe und der Nachweis des Spleißens in Abwesenheit anderer Proteine oder Cofaktoren möglich.^[4]

1.2. Intein-Mobilität und Homing-Endonuclease-Aktivität

Ähnliche Intein-Allele wurden in der gleichen Position in homologen Genen eng verwandter, aber auch stark divergenter Organismen gefunden,^[8, 29–32, 34] obwohl die Inteine unregelmäßig über die Arten verteilt sind. GC-Gehalt und Codon-Gebrauch der DNA lassen darauf schließen, dass einige Intein-Gene in jüngerer Zeit zwischen Organismen ausgetauscht worden sind.^[33, 34] Die Mobilität eines Intein-Gens wurde experimentell bei *S. cerevisiae* nachgewiesen.^[37]

Einige Intron-Typen sind ebenfalls mobile genetische Elemente. Eine auf einem Intron codierte Endonuclease („homing endonuclease“)^[38] initiiert die Mobilisierung durch die Einführung eines Doppelstrangbruchs; dieser wird dann zum Ausgangspunkt für eine Genkonversions-Reaktionssequenz.^[13, 30, 35] Homing-Endonucleasen sind eine Klasse doppelstrang- und sequenzspezifischer Endonucleasen, die DNA an oder nahe der Intron- oder Intein-Insertionsstelle in homologen Genen spalten, die kein solches Element enthalten.^[35, 38] Viele Inteine sind bifunktionelle Proteine, die auch Homing-Endonuclease-Aktivität aufweisen (siehe Abschnitt 1.3).^[26, 28, 37, 38, 41, 43, 44] Wenn ein Intein-enthaltendes Allel durch geschlechtliche Fortpflanzung, Infektion, Konjugation, Transduktion o. Ä. in eine Zelle gelangt, wird das Intein exprimiert. Die Intein-Endonucleaseaktivität spaltet die Intein-negative Kopie des Gens bei oder nahe der Intein-Insertionsstelle („home“). Da die Erkennungsstelle für die Homing-Endonuclease durch das Intein unterbrochen wird, ist das Intein-positive Gen die einzige intakte Kopie des Allels, durch die das gesplattene Intein-negative Gen repariert werden kann. Die Reparatur des gesplattene Intein-negativen Allels mit Hilfe des intakten Intein-positiven Allels als Matrize führt zu einer gerichteten Genkonversion in Richtung des Intein-positiven Allels. Auf ähnliche Weise können Intein-negative Genabschnitte auf Phagen, Viren, Plasmiden oder anderen episomalen Elementen ein Intein aufnehmen, wenn sie eine Zelle mit einem Intein-positiven Homologen durchlaufen.

Es gibt vier Homing-Endonuclease-Familien, die anhand von konservierten Signaturmotiven unterschieden und benannt werden.^[38] Sequenzvergleiche ergaben, dass die meisten Inteine Homing-Endonucleasen aus der Dodecapeptid-motiv-Familie (DOD) enthalten, während ein Intein eine Homing-Endonuclease der H-N-H-Familie enthält (Abbildung 1 und Abschnitt 1.3).^[1, 26, 29–31, 35–41] Die Intein-Region mit

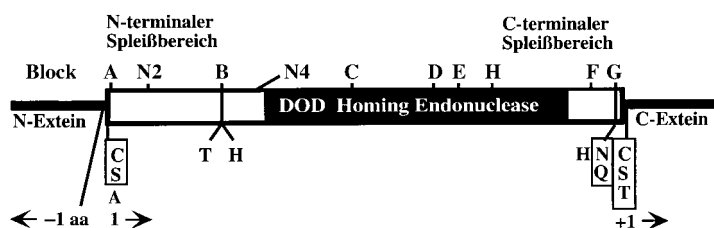


Abbildung 1. Konservierte Eigenschaften der Inteine. Abgebildet ist ein Vorläuferprotein, das ein Intein mit einer DOD-Homing-Endonuclease enthält: Oberhalb der Sequenz sind die konservierten Motive, unterhalb davon die konservierten Aminosäurereste, die an der Katalyse beteiligt sind, gezeigt. Aminosäuren mit nucleophiler Gruppe in der Seitenkette sind eingerahmt. Zu beachten ist, dass einige Inteine mit Ala beginnen. Die Intein-Spleißstellen (weiße Kästchen) sind durch eine Endonuclease-Domäne (wie hier gezeigt) oder durch ein kurzes Verbindungsstück voneinander getrennt. Für die Nummerierung der Aminosäuren in Vorläuferproteinen gilt folgende Konvention: Die Aminosäuren des Inteins werden beginnend mit 1 vom N-Terminus zum C-Terminus nummeriert. Die Aminosäuren des N-Exteins werden mit negativen Zahlen beginnend mit -1 an der N-terminalen Spleißstelle rückwärts zum N-Terminus des Vorläuferproteins hin nummeriert. Die Aminosäuren im C-Extein werden beginnend mit +1 vom N-Terminus (also der C-terminalen Spleißstelle) in Richtung des C-Terminus des Vorläuferproteins hin gezählt, wobei das Pluszeichen das C-Extein kennzeichnet. Die Pfeile deuten die Nummerierungsrichtung an. - aa = Aminosäure.

der DOD-Homing-Endonuclease lässt sich anhand charakteristischer Motive identifizieren, die als Intein-Blöcke C, D, E und H bezeichnet werden.^[29–32, 38, 40] Es gibt Hinweise aus unterschiedlichen Richtungen, die darauf hindeuten, dass die Homing-Endonuclease-Gene während der Evolution in Intron- und Intein-Gene integriert wurden.

1.3. Intein-Verteilung und konservierte Merkmale

Das Spleißen von Proteinen kommt überall in der Natur vor: Bis zum 4. Juli 1999 sind 97 Inteine in einzelligen Organismen aus allen drei Reichen (Eubakterien, Archaeobakterien, Eukaryonten), in einem Virus und einem Prophagen identifiziert worden.^[29–32, 40] Aktuellere Informationen über Inteine sind in InBase, der Intein-Datenbank (<http://www.neb.com/neb/inteins.html>) zu finden.^[29]

Man nimmt von einem Protein an, dass sein Gen von einer Intein-Sequenz unterbrochen wird, wenn 1) Extein-Homologe die eingeschobene Sequenz (auf Gen-Ebene) nicht enthalten, 2) die eingeschobene Sequenz länger als 100 Aminosäurereste ist, 3) die eingeschobene Sequenz Intein-Motive enthält und 4) das beobachtete Proteinprodukt so groß ist wie das vorhergesagte offene Leseraster ohne Intein. Die Nummerierung der Aminosäuren in noch nicht gespleißten Vorstufen eines Proteins ist in Abbildung 1 erläutert.

Inteine haben eine Größe zwischen 134 und 600 Aminosäuren.^[29] Einige der größeren Inteine sind bifunktionelle Proteine mit Endonuclease-Aktivität. Zehn konservierte Sequenzmotive kommen in den meisten Inteinen vor (Abbildung 1).^[29–31, 40] In einem typischen, noch nicht gespleißten Vorläuferprotein wird die zentrale Intein-Region (in der ein

Verbindungsstück oder eine Homing-Endonuclease-Domäne codiert ist) von der N-terminalen Spleiß-Region (Blöcke A, N2, B, N4) und der C-terminalen Spleiß-Region (Blöcke F und G) flankiert. Daran schließen sich nach außen das N-Extein bzw. das C-Extein (Block G) an.^[42] Die ungespleißte Vorstufe ist also ähnlich wie eine russische Matroschkapuppe aufgebaut (Abbildung 1). Im Sce-VMA-Intein wurde ein Teil der Homing-Endonuclease in den N-terminalen Spleißbereich verlagert;^[43–45] eine vergleichende Sequenzanalyse ergab jedoch, dass diese Organisationsform unter den bekannten Inteinen wahrscheinlich selten ist.^[31, 32, 39, 42] Ungefähr 10% der Inteine haben ein kurzes Verbindungsstück anstelle der Homing-Endonuclease-Domäne.^[29]

Die Intein-Spleißdomäne umfasst die Blöcke A, N2, B, N4, F und G (Abbildung 1). Keine einzige Aminosäure ist in allen Inteinen konserviert. Stattdessen enthalten die Inteinblöcke sogar an den höchstkonservierten Positionen nur verwandte Aminosäuren. Alle bekannten Inteine haben ein Ser, Thr oder Cys am C-terminalen Ende beider Spleißstellen, mit Ausnahme von zwei Intein-Familien, die an der N-terminalen Spleißstelle ein Ala enthalten (Abbildung 1). Ein Edman-Proteinabbau vom N-Terminus her ergab, dass das freie *Thermococcus-littoralis* (Tli)-Pol-2-Intein mit Ser beginnt, woraus sich schließen lässt, dass das konservierte Ser oder Cys an der stromaufwärts (upstream) liegenden Spleißstelle der erste Aminosäurerest des Inteins ist.^[26] Alle bekannten Inteine enden mit Asn außer zweien, die ein Gln am Ende tragen. Block B enthält ein Thr-Xxx-Xxx-His-Motiv, das man auch in Serin-Proteasen findet, wo His zur katalytischen Triade gehört.^[46] Ein zweiter konservierter Histidinrest liegt als vorletzter Rest bei ca. 90% der Inteine in Block G.^[29] Es gibt keine Konsensus-Aminosäure am C-Terminus des N-Exteins. Weil verwandte – wenn auch nicht identische – Aminosäuren in konservierten Motiven vorkommen, fungieren unterschiedliche Aminosäuren als Nucleophil oder als unterstützender Rest in jedem Intein. Ein konservativer Austausch von Ser, Thr oder Cys an der Spleißstelle beeinträchtigt oder verhindert jedoch oft den Spleißvorgang. Es sind also subtile Unterschiede im pK_a-Wert, der Nucleophilie und der Konformation bei verschiedenen Inteinen, die möglicherweise die Komplexität eines jeden Protein-Spleißmechanismus trotz eines einheitlichen Grundmusters erhöhen.

Fast alle Kombinationen von Ser-, Thr- und Cys-Resten an den Spleißstellen wurden beobachtet, mit Ausnahme von Thr am N-Terminus des Inteins (Thr1).^[29] Dies mag an der Zahl der untersuchten Inteine liegen, denn auch nach Mutation des N-terminalen Ser1 zu Thr im Tli-Pol-2-Intein ist ein effizientes Spleißen noch möglich.^[28] Wenn man die verschiedenen Intein-Allele als eine Intein-Familie ansieht, gibt es kaum Präferenzen für Kombinationen aus Aminosäuren mit katalytisch aktiven funktionellen Gruppen: Thiol- und Thiolgruppen (36%), Hydroxy- und Hydroxygruppen (18%) oder Thiol- und Hydroxygruppen (46%). Allerdings ist bei den gegenwärtig verfügbaren 39 unterschiedlichen Intein-Allelen mit Ser, Thr oder Cys an beiden Spleißstellen Cys an der N-terminalen Spleißstelle überrepräsentiert (82%) und an der C-terminalen Spleißstelle unterrepräsentiert (36%; hierbei sind Ser und Thr als Hydroxygruppen-enthaltende Aminosäuren zusammengefasst).

1.4. Inteine und die Autoprozessierungsdomänen der Hedgehog-Proteine

Zwar muss das Protein-Spleißen in vielzelligen Organismen erst noch nachgewiesen werden, aber Proteine der Hedgehog-Familie, die zu den embryonalen Signalproteinen gehören, enthalten eine Autoprozessierungsdomäne, die höchstwahrscheinlich mit den Inteinen einen gemeinsamen Vorläufer in der Evolution hat.^[42, 45, 48] Die Hedgehog-Autoprozessierungsdomäne und die Spleißdomäne der Inteine weisen die gleiche Faltung auf (das HINT-Modul),^[45] haben zwei Motive (Block A und B) gemeinsam^[46, 47] und katalysieren ähnliche chemische Reaktionen (siehe Abschnitt 2.6).^[42, 45, 46, 48] Die Vorstufe des Hedgehog-Proteins wird zwischen der N-terminalen Signaldomäne und der C-terminalen Prozessierungsdomäne gespalten. Gleichzeitig wird Cholesterin an den C-Terminus der Signaldomäne gebunden, wodurch diese an der Zelloberfläche verankert wird. Dort vermittelt sie die Musterbildung während der Embryogenese. Ähnliche Autoprozessierungsdomänen wurden in mehreren unterschiedlichen Nematoden-Genen unbekannter Funktion gefunden.^[45, 48, 49]

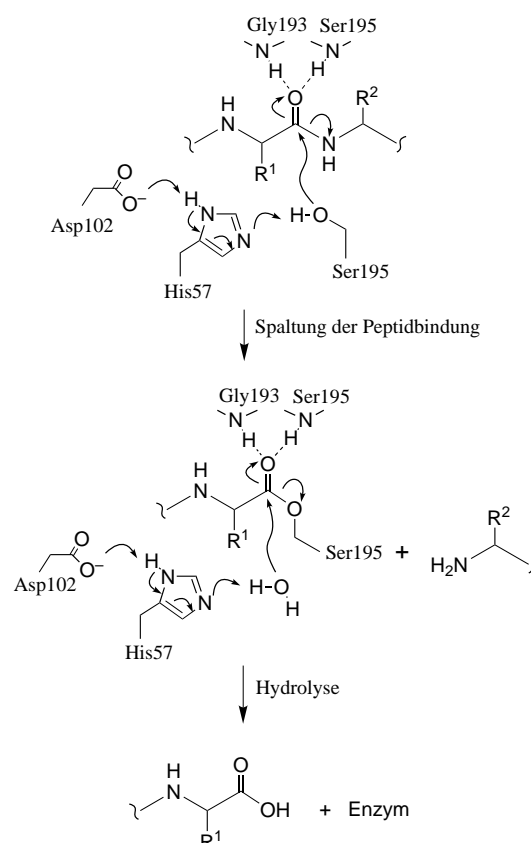
2. Aufklärung des Mechanismus des Protein-Spleißens

Beim Spleißen eines Proteins wird letztlich eine native Peptidbindung in den verknüpften Exteinen gebildet.^[9] Der Spleißvorgang kann in drei voneinander abgrenzbare, aber gekoppelte chemische Reaktionen unterteilt werden: Auf zwei Proteinspaltungsreaktionen folgt eine Verknüpfungsreaktion. Viele Reaktionsmechanismen wurden vorgeschlagen, die auf den bekannten chemischen Eigenschaften der konservierten Aminosäurereste an den Spaltstellen basierten.

2.1. Ähnlichkeit mit Serin- und Cystein-Proteasen

In der Klasse der Serin- und Cystein-Proteasen^[50] greift die nucleophile Hydroxy- bzw. Thiolgruppe in der Seitenkette des jeweiligen Serin- oder Cysteinrestes im aktiven Zentrum die Peptidbindung des Substrats an einem Carbonyl-Kohlenstoffatom der Hauptkette an (Schema 2). Dadurch kommt es zur Spaltung der Peptidbindung, wobei das C-terminale Spaltprodukt freigesetzt wird und ein Acyl-Enzym-Zwischenprodukt gebildet wird, in dem der N-terminale Teil des Substrats kovalent als (Thio-)Ester mit dem Enzym verbunden ist. Durch Hydrolyse der (Thio-)Esterbindung wird das N-terminale Polypeptid freigesetzt und die aktive Protease regeneriert. Treibende Kraft für den vollständigen Ablauf der Proteasereaktion ist die Hydrolyse des Acyl-Enzym-Zwischenprodukts. Dementsprechend können Serin- und Cystein-Proteasen in Peptid-Ligasen umgewandelt werden, wenn Wasser von der Reaktion ausgeschlossen^[51] oder wenn das Nucleophil im aktiven Zentrum zu einer schlechten Abgangsgruppe bei der Hydrolyse des Acyl-Enzym-Zwischenprodukts mutiert wird.^[52, 53]

Die Katalyse des einleitenden nucleophilen Angriffs auf die Peptidbindung wird verstärkt, indem einerseits die Elektro-



Schema 2. Katalysemechanismus der Serin-Proteasen. Die Aminosäurereste der katalytischen Triade (Asp102, His57, Ser195) im aktiven Zentrum von Chymotrypsin sind zusammen mit einem hypothetischen Polypeptid (dem Substrat) gezeigt; die zu spaltende Bindung liegt zwischen den Resten R¹ und R². Der nucleophile Angriff durch Ser195 im aktiven Zentrum wird durch gleichzeitige Deprotonierung durch das benachbarte His57 vermittelt, das seinerseits durch die Wechselwirkung mit dem benachbarten, negativ geladenen Asp102 stärker basisch wird. Die sich aufbauende negative Ladung am Carbonyl-Sauerstoffatom wird durch die Wechselwirkung mit Amidprotonen der Hauptkette von Gly193 und Ser195 (dem „Oxyanionen-Loch“) stabilisiert. Tetraedrische Zwischenprodukte wurden aus Gründen der Übersichtlichkeit weggelassen.

philie der Carbonylgruppe im Substrat (durch Wechselwirkung des Carbonyl-Sauerstoffatoms im Übergangszustand mit benachbarten geladenen Gruppen, dem „Oxyanionen-Loch“) und andererseits die Nucleophilie der angreifenden Hydroxy- oder Thiolgruppe erhöht wird.^[50] Letzteres wird durch ein Ladungsverschiebungssystem erreicht, in dem das Proton des angreifenden Moleküls im Übergangszustand auf einen benachbarten Histidinrest übertragen wird; dessen pK_a-Wert (und dementsprechend seine Basizität) wird durch die Wechselwirkung mit einem benachbarten Aspartat- (in Serin-Proteasen) oder Asparaginrest (in Cystein-Proteasen) erhöht.^[54, 55] Diese katalytische Triade, Asp-His-Ser oder Asn-His-Cys, ist das typische Kennzeichen aller Serin- bzw. Cystein-Proteasen. Die konservierten His-, Asn/Gln- und Ser/Thr/Cys-Reste in Inteinen (Abbildung 1 und Abschnitt 1.3) erinnern an die katalytischen Triaden der Proteasen, weshalb wir postulierten, dass diese Aminosäurereste ähnliche funktionelle Rollen beim Protein-Spleißen übernehmen.^[26, 28]

2.2. Ein In-vitro-Spleißsystem führt zur Entdeckung eines verzweigten Proteinzwischenproduktes

Die weitere Aufklärung des Spleißmechanismus wurde durch die Entwicklung eines In-vitro-Spleißsystems sehr erleichtert, das die zeitaufgelöste Untersuchung des Spleißens einer gereinigten Proteinvorstufe ermöglichte (Abbildung 2).^[4] Das

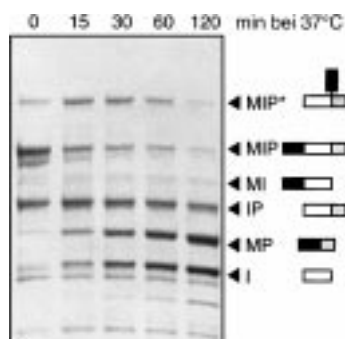


Abbildung 2. Zeitaufgelöste In-vitro-Spaltung des gereinigten MIP-Vorläuferproteins. MIP, ein Fusionsprotein aus MBP, dem Psp-GBD-Pol-Intein und Paramyosin, wurde unter Bedingungen exprimiert, die das Spleißen verhindern, und durch Affinitätschromatographie an Amyloseharz gereinigt.^[4] Die erste Probe enthielt den MIP-Vorläufer (132 kDa), eine kleine Menge einer langsam wandernden MIP-Vorstufe (MIP*) und gespleißtes MP sowie freies Intein (I) und Spaltprodukte aus der Spaltung an nur einer Spleißstelle (M, IP, MI). Die Produkte der zeitaufgelösten Spleißreaktion der MIP-Vorstufe in Abwesenheit von Cofaktoren wurde mit SDS-PAGE untersucht (die Gele wurden mit Coomassie-Blau gefärbt). – Symbole: schwarzes Kästchen: M (N-Extein); weißes Kästchen: I (Intein); graues Kästchen: P (C-Extein).

Konstrukt besteht aus dem *Pyrococcus-sp.* (Psp)-GBD-Pol-Intein der Deep-Vent-DNA-Polymerase, das N-terminal vom Maltose-bindenden Protein (MBP oder M) aus *E. coli* und C-terminal von einem 29 kDa-Fragment des Paramyosins (P) aus *Dirofilaria immitis* flankiert wird. Das Psp-GBD-Pol-Intein hat Serinreste an beiden Spleißstellen. Durch dieses dreiteilige Fusionsprotein, MIP (MBP-Intein-Paramyosin), wurde die affinitätschromatographische Reinigung der ungespleißten Proteinvorstufe an trägergebundener Amylose und die schnelle Identifizierung von Spleiß-Zwischenstufen und -Produkten durch Natriumdodecylsulfat-Polyacrylamidgel-Elektrophorese (SDS-PAGE) oder Westernblots erleichtert. Das MIP-Konstrukt ermöglichte das Spleißen, war aber nicht optimal. Die Reaktion lief bei 12–20 °C sehr langsam ab, schneller aber oberhalb von 30 °C. Diese Temperaturabhängigkeit rührte nicht daher, dass das Psp-GBD-Pol-Intein aus einem extrem thermophilen Organismus stammt, denn das Spleißen des nativen DNA-Polymerase-Vorläufers in *E. coli* lief bei 20 °C zu schnell ab, um diesen isolieren zu können. Ein temperaturabhängiges Spleißen durch Verwendung nichtnativer Exteine erwies sich als verbreitetes Phänomen.^[4, 8, 13] Das Spleißen von MIP war pH-abhängig mit einem pH-Optimum von 4,5–6,5 und einem stetigen Abfall der Reaktionsgeschwindigkeit bis pH 10, von wo an die Reaktion zum Stillstand kam.^[4]

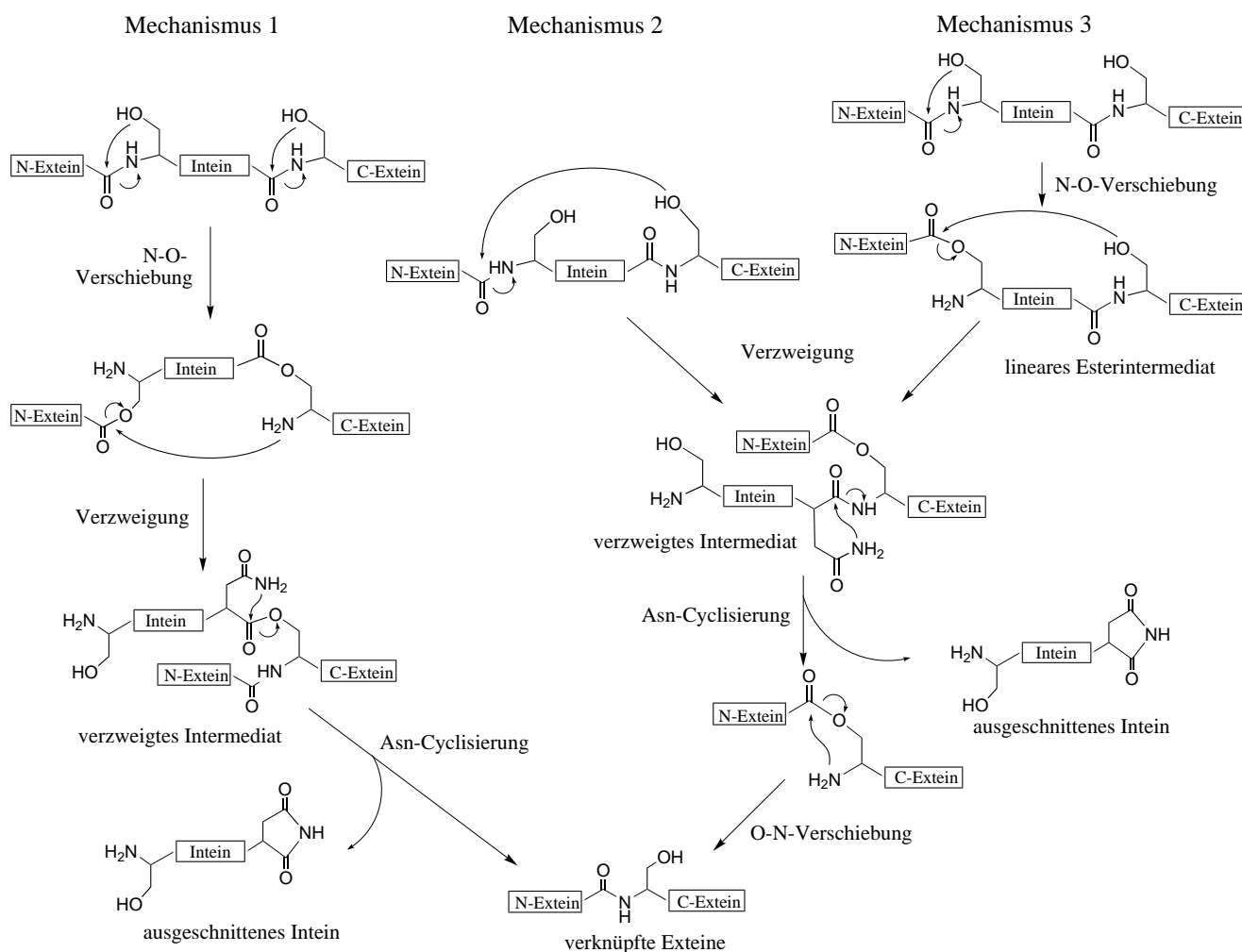
Mit dem MIP-System konnte ein entscheidendes Zwischenprodukt der Reaktion durch SDS-PAGE identifiziert werden. Aus MIP-Proben wurden gespleißte Produkte freigesetzt, MP

(MBP-Paramyosin) und das freie Intein (I), sowie Spaltprodukte, die nur an einer Seite gespleißt waren (siehe Abschnitt 2.8). Eine Spaltung an der N-terminalen Spleißstelle führte zur Bildung von N-Extein und Intein/C-Extein (M + IP), Spaltung an der C-terminalen Spleißstelle zur Bildung von N-Extein/Intein und C-Extein (MI + P). Zusätzlich zu den erwarteten Spleiß- und Spaltprodukten trat in einer frühen Phase des Reaktionsverlaufs eine Spezies mit geringer elektrophoretischer Mobilität auf, die langsam verschwand, wenn der Vorläufer verbraucht und das Endprodukt gebildet wurde (Abbildung 2). Diese Reaktionskinetik ist charakteristisch für das Auftreten eines Zwischenprodukts, und die im Vergleich zum Vorläufer kleine Wanderungsgeschwindigkeit bei der SDS-PAGE ließ an ein kovalent oligomerisiertes oder an ein verzweigtes Zwischenprodukt denken. Bei der Sequenzierung der langsam wandernden Komponente durch N-terminalen Edman-Abbau fanden sich zwei überlappende Sequenzen, die den N-Termini des N-Exteins (MBP) und des Inteins entsprachen.^[4] Bildung und Stabilität des verzweigten Zwischenprodukts waren pH-abhängig: Das Zwischenprodukt wandelte sich bei pH 10 und 37 °C wieder in das Vorläuferprotein um, was implizierte, dass die Bindung am Verzweigungspunkt eine Estergruppe ist, denn das Ester/Amid-Gleichgewicht wird bei hohem pH-Wert in Richtung des Amids verschoben.^[56–58] Die Existenz dieses verzweigten Zwischenprodukts deutet auf eine Umlagerung des Vorläuferproteins hin, bei der die Spaltung der N-terminalen Spleißstelle mit der Acylwanderung des N-Exteins zur Seitenkette eines Restes in der Nachbarschaft der C-terminalen Spleißstelle gekoppelt ist. Das Vorliegen eines verzweigten Zwischenprodukts schloss einen Mechanismus aus, bei dem die Einleitung des Spleißens durch die Cyclisierung von Asn vorhergesagt wurde.^[59]

Die Mutagenese des Ser-, Thr- oder Cys-Nucleophils an einer der Spleißstellen zu Ala verhinderte die Bildung einer Verzweigung, während die Mutagenese des His- oder Asn-Restes am C-Terminus des Inteins wenig Einfluss auf die Bildung der Verzweigung hatte.^[60] Damit wurde sofort ein hypothetischer Mechanismus widerlegt, nach dem das C-terminale Asn das angreifende Nucleophil bei der Bildung der Verzweigung sein sollte.^[61]

2.3. Drei vorgeschlagene Mechanismen auf der Basis eines verzweigten Proteinzwischenprodukts

Drei weitere Mechanismen wurden vorgeschlagen, um Bildung und Weiterreaktion des verzweigten Zwischenprodukts vor dem Hintergrund der vorliegenden Mutagenesedaten zu erklären. Nach Mechanismus 1 (Schema 3), der von Paulus (persönliche Mitteilung in Lit. [4]) und Wallace^[62] unabhängig voneinander vorgeschlagen wurde, gehen beide Spleißstellen eine N → O(S)-Acylwanderung ein, bei der die nucleophilen Hydroxy- oder Thiolgruppen der konservierten Ser/Thr/Cys-Reste an beiden Spleißstellen das Proteinrückgrat an den Carbonylgruppen der vorhergehenden Reste angreifen, wodurch chemisch labile (Thio-)Esterbindungen an beiden Enden des Inteins entstehen. Bei der N → O(S)-Acylwanderung am stromabwärts (downstream) liegenden



Schema 3. Drei Vorschläge für den Mechanismus des Protein-Spleißens. Alle drei Mechanismen sind für ein hypothetisches Intein mit Ser an beiden Spleißstellen demonstriert, obwohl der Mechanismus der gleiche ist, wenn Ser durch Cys oder Thr an einer oder beiden Spleißstellen ersetzt wird. Alle tetraedrischen Zwischenprodukte und Protonenübertragungsschritte sind der besseren Übersicht halber weggelassen. Für Einzelheiten siehe Text.

Ende wird eine freie α -Aminogruppe am ersten Rest des C-Exteins gebildet, die ihrerseits die labile (Thio-)Esterbindung an der N-terminalen Spleißstelle angreift. In dieser Reaktion bricht die Verbindung zwischen Intein und N-Extein auf, und das lineare (Thio-)Ester-Zwischenprodukt wandelt sich in ein verzweigtes (Thio-)Ester-Zwischenprodukt um, in dem die Exteine bereits durch eine Peptidbindung verknüpft sind. Die Weiterreaktion des verzweigten Zwischenprodukts geschieht dann durch die Cyclisierung des C-terminalen Asn aus dem Intein zu einem Succinimidring, bei der die (Thio-)Esterbindung zwischen Intein und C-Extein gespalten wird und so die Polypeptidketten des Inteins und der verknüpften Exteine freigesetzt werden.

Nach Mechanismus 2 (Schema 3)^[63] findet die einleitende Acylwanderung distal und nicht proximal statt. Die nucleophile Hydroxy- oder Thiolgruppe der Seitenkette des C-Extein-Restes + 1 greift die Peptidbindung an der N-terminalen Spleißstelle an, was unmittelbar zu einem verzweigten Zwischenprodukt führt, in dem N-Extein und C-Extein durch eine (Thio-)Ester-Bindung verknüpft sind. Diese Reaktion ist das intramolekulare Analogon des einleitenden Schrittes im Serin/Cystein-Protease-Mechanismus, wobei das verzweigte Zwischenprodukt dem Acyl-Enzym-Zwischenprodukt ent-

spricht. Das verzweigte Zwischenprodukt reagiert durch die Cyclisierung des C-terminalen Asn im Intein weiter, wobei die durch eine (Thio-)Esterbindung verknüpften Exteine freigesetzt werden. Wenn nun die stabilisierende Umgebung des aktiven Zentrums des Inteins fehlt, kommt es zu einer spontanen O(S) \rightarrow N-Acylwanderung, durch die die thermodynamisch stabilere Amidbindung zwischen den Exteinen entsteht.^[56, 57, 64, 65]

Nach Mechanismus 3^[63] (Schema 3) beginnt die Reaktion analog zu Mechanismus 1 mit einer N \rightarrow O(S)-Acylwanderung an der N-terminalen Spleißstelle. Das stromabwärts liegende Ser/Thr/Cys-Nucleophil greift dann die neu entstandene (Thio-)Esterbindung an (anstelle der Amidbindung, wie in Mechanismus 2 vorgeschlagen). Die übrigen Schritte von Mechanismus 2 und 3 sind gleich.

Die Eigenschaften des verzweigten Zwischenprodukts waren maßgeblich für den Ausschluss von Mechanismus 1, da dieser vorsieht, dass die Exteine durch eine Amidbindung und das Intein mit dem C-Extein durch eine (Thio-)Esterbindung verknüpft sind. Nach Mechanismus 2 und 3 sind Intein und C-Extein im verzweigten Zwischenprodukt amidisch verknüpft, und das N-Extein ist durch eine (Thio-)Esterbindung angebunden. Mit dem MIP-System konnte man zwi-

schen den beiden Möglichkeiten unterscheiden, indem man die Labilität der Esterbindungen bei hohen pH-Werten nutzte.^[63] Das vollständig denaturierte verzweigte Zwischenprodukt des MIP-Systems zerfiel bei 65 °C und pH 9 in M und IP, was auf eine im Basischen labile Esterbindung zwischen dem N-Extein (M) und dem Intein/C-Extein (IP) hindeutete.^[63] Dieses Ergebnis schloss Mechanismus 1 aus und stützte die Mechanismen 2 und 3. Ein ähnliches verzweigtes Zwischenprodukt wurde später beim Spleißvorgang des Sce-VMA-Inteins nachgewiesen, allerdings nur nach Mutation von Cys +1 zu Ser.^[66]

2.4. Die Weiterreaktion des verzweigten Zwischenproduktes

Nach allen drei Mechanismen reagiert das verzweigte Zwischenprodukt durch die Cyclisierung von Asn weiter, wobei die benachbarte C-terminale Spleißstelle gespalten wird. Die Cyclisierung von Asn während des Spleißens wurde nachgewiesen, indem fünf Positionen stromaufwärts von der C-terminalen Spleißstelle von MIP ein Methioninrest eingeführt und das gereinigte herausgeschnittene Intein mit dem für Methioninreste spezifischen Spaltungsreagens Bromcyan behandelt wurde, um das C-terminale Tetrapeptid des Inteins freizusetzen.^[63] HPLC-analytisch wurden etwa gleiche Mengen von zwei Peptiden nachgewiesen, die beide die erwartete Zusammensetzung für das C-terminale Intein-Tetrapeptid hatten. Durch Massenspektrometrie wurde gezeigt, dass sich die beiden Peptide in ihrer relativen Molekülmasse um 18 Einheiten (ein Wasser-Äquivalent) unterschieden, wie dies für die hydrolysierte und die intakte Form eines Peptides mit einem C-terminalen Succinimidrest erwartet wurde. Der Succinimidring war in vivo und in vitro relativ stabil.^[63] Die Gegenwart eines Succinimidrestes wurde mit synthetischen Modellpeptiden, Massenspektrometrie und einem colorimetrischen Test im Falle des Psp-GBD-Pol-Inteins^[67] und durch massenspektrometrische Untersuchung von Peptiden beim Sce-VMA-Intein^[66] bestätigt.

Ein analoger Prozess – Spaltung von Peptidbindungen und Bildung eines C-terminalen Succinimidrestes durch Cyclisierung von Asn – findet in Peptiden bei pH 7.4^[68] und bei der Alterung von Proteinen wie dem Rinder- α -Kristallin^[69] statt. Succinimide sind auch Zwischenprodukte bei der häufig vorkommenden spontanen Desaminierung und Isomerisierung von internen Asn-Resten in Proteinen und Peptiden.^[68, 70–73] Anders als beim Spleißen von Proteinen und beim Altern von α -Kristallin wird das Succinimid in den letztgenannten Fällen durch einen Angriff des Amid-Stickstoffatoms aus der Hauptkette auf die Carbonylgruppe in der Seitenkette gebildet. Die Hydrolyse des entstandenen Succinimids in der Seitenkette oder an der Carbonylgruppe der Hauptkette führt dann zur Bildung eines internen Aspartat- oder Isoaspartatrestes. In Peptiden besteht eine Konkurrenzreaktion zwischen dem Angriff des Amid-Stickstoffatoms der Seitenkette auf die Carbonylgruppe der Hauptkette und dem Angriff des Hauptketten-Stickstoffatoms auf die Carbonylgruppe der Seitenkette; in den meisten Peptiden dominiert die erste Reaktion, doch die zweite findet verstärkt statt,

wenn Pro, Ser oder His auf das Asn in der Peptidsequenz folgen.^[68, 72] Man geht davon aus, dass die Succinimidbildung durch die Fähigkeit begrenzt wird, das Amid-Stickstoffatom der Hauptkette zu deprotonieren und seine Nucleophilie zu erhöhen.^[72] Die Asn-Cyclisierung in Proteinen wird auch durch die Konformation eingeschränkt, denn Strukturanalysen lassen vermuten, dass für die Succinimidbildung Diederwinkel $\varphi = -120^\circ$ und $\chi = +120^\circ$ notwendig sind.^[72] Wegen der Entropieeffekte isomerisiert Gln etwa 10-mal langsamer zu einem sechsgliedrigen Ring als Asn zu einem fünfgliedrigen Ring.^[72] Das Spleißen der Inteine läuft um Größenordnungen schneller ab als die Spaltung von Peptiden durch Asn oder Gln (wo die Halbwertslebensdauer in Tagen und nicht in Sekunden gemessen wird), wahrscheinlich deswegen, weil das Spleißelement einem Enzym insofern ähnelt, als es diese Reaktion unter physiologischen Bedingungen erleichtert (siehe Abschnitt 2.1).

2.5. Die Bildung einer Peptidbindung zwischen den verknüpften Exteinen

Nach der Cyclisierung von Asn und dem Ausschneiden des Inteins bleiben die Exteine durch eine (Thio-)Esterbindung zwischen dem Carbonyl-Kohlenstoffatom des Aminosäurerestes –1 des N-Exteins und der Hydroxy- oder Thiolgruppe des Aminosäurerestes +1 des C-Exteins verbunden. Ohne den stabilisierenden Effekt des Inteins findet eine spontane O(S) \rightarrow N-Wanderung unter Bildung der thermodynamisch stabileren Amidbindung zwischen den Exteinen statt. Modellpeptide, die interne Ester- oder Thioesterbindungen an Serin- oder Cysteinresten enthalten, isomerisieren schnell zur Amidform.^[64, 65]

2.6. Die Aktivierung der N-terminalen Spleißstelle für die Spaltung

Das letzte Stück des Puzzles betraf die konservierten nucleophilen Aminosäuren Ser1 oder Cys1 an der N-terminalen Spleißstelle. Anders als bei Mechanismus 3 wird bei Mechanismus 2 keine direkte katalytische Rolle für diesen Rest gefordert (Schema 3). Eine Möglichkeit ist, dass die Hydroxy- oder Thiolgruppe eine essentielle Wasserstoffbrückenbindung bildet, vielleicht zum benachbarten Carbonyl-Kohlenstoffatom. Dies würde den nucleophilen Angriff durch den Aminosäurerest +1 des C-Exteins katalysieren, indem die im Übergangszustand zum ersten tetraedrischen Zwischenprodukt entstehende negative Ladung stabilisiert würde. Die Substitution von Ser1 gegen Cys in den Tli-Pol-2- und Psp-GBD-Pol-Inteinen verringerte die Spleißgeschwindigkeit und die Menge an gespleißtem Produkt nach Inkubation über Nacht,^[28, 60] während die Substitution von Cys1 gegen Ser das Spleißen des Sce-VMA-Inteins verhinderte^[9, 66, 74] und die Substitution von Ser1 oder Cys1 gegen Ala das Spleißen in allen drei Inteinen blockierte. Die Spaltung der N-terminalen Spleißstelle fand aber auch nach Mutation von allen drei konservierten C-terminalen Aminosäuren an der Spleißstelle (His/Asn/Ser) zu Ala noch statt.^[60] Diese Daten lassen sich

eher mit der Vorstellung einer direkten nucleophilen Funktion von Ser1 und Cys1 in Einklang bringen als mit Mechanismus 3.

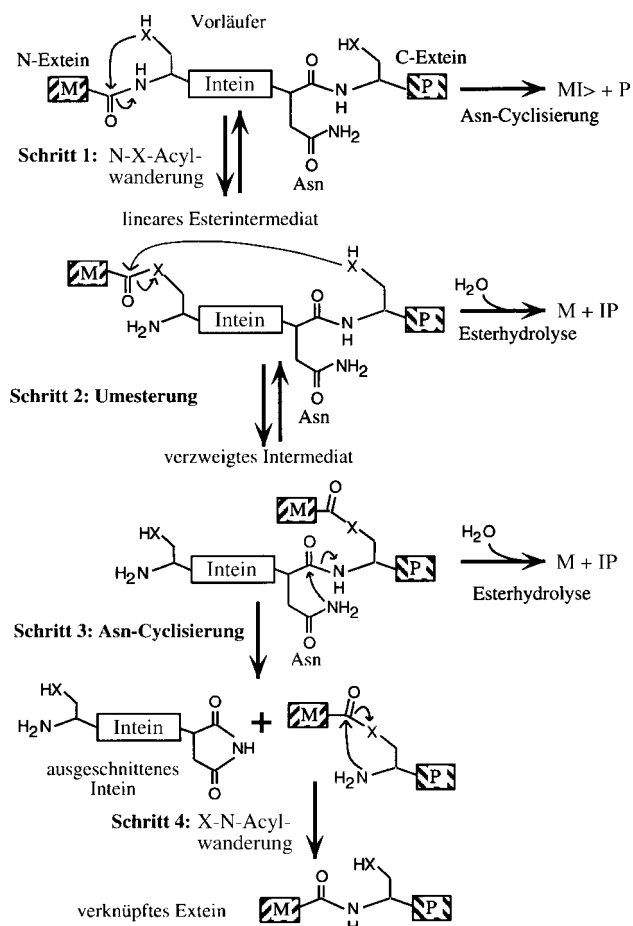
Bei Mechanismus 3 wird postuliert, dass Ser1 oder Cys1 eine $N \rightarrow O(S)$ -Acyl-Umlagerung eingehen können. Das stromabwärts liegende Nucleophil soll dann den entstehenden (Thio-)Ester (statt – wie in Mechanismus 2 gefordert – die Amidbindung) angreifen. Ähnliche $N \rightarrow O(S)$ -Acyl-Umlagerungen, die zur Spaltung der entstehenden (Thio-)Esterbindungen führen, wurden bereits in anderen autokatalytischen Protein-Prozessierungsreaktionen beobachtet, darunter bei der Bildung der Pyruvoylgruppe, der Prozessierung der Hedgehog-Proteine und bei N-terminalen nucleophilen Amidohydrolasen.^[2, 3] Die Pyruvoyl-Bildung in der Prohistidin-Decarboxylase wird durch eine $N \rightarrow O$ -Acylverschiebung an Ser82 eingeleitet. Die Mutation von Ser82 zu Cys verringert die Geschwindigkeit der Pyruvoyl-Bildung stark, worin sich die nicht optimale Übereinstimmung zwischen dem Nucleophil und der elektronischen Umgebung im aktiven Zentrum des Enzyms widerspiegelt.^[75] Ein ähnliches Phänomen trat bei konservativen Substitutionen von Ser, Thr oder Cys an Intein-Spleißstellen auf,^[6, 9, 28, 60, 66, 74, 76, 77] außerdem dann, wenn das Nucleophil im aktiven Zentrum von Serin- oder Cystein-Proteasen zu Cystein bzw. Serin geändert wurde,^[78] oder wenn Thr im für die Autoprozessierung verantwortlichen aktiven Zentrum der *Flavobacterium*-Glycosylasparaginase zu Cys mutiert wurde.^[79]

Die Acylverschiebung in einer Ser82 \rightarrow Cys-Mutante der Prohistidin-Decarboxylase wurde durch Hydroxylamin-Spaltung bestätigt.^[75] Bei neutralem pH-Wert sind interne Thioesterbindungen in Proteinen viel empfindlicher gegenüber Hydroxylamin als Amid- und Esterbindungen.^[58] Im Unterschied zum Wildtyp-Enzym war die Ser82 \rightarrow Cys-Mutante der Prohistidin-Decarboxylase empfindlich gegenüber einer Hydroxylamin-Spaltung. Interne Thioestereinheiten in Proteinen können auch durch niedermolekulare Thiolreagentien umgeestert werden, wodurch das Protein gespalten wird. Eine ähnliche Strategie wurde benutzt, um Thioester-Zwischenprodukte bei der Prozessierung des Hedgehog-Proteins und der *Flavobacterium*-Glycosylasparaginase nachzuweisen.^[48, 60, 79, 80] Die Mutation von Ser1 aus dem Psp-GBD-Pol-Intein zu Cys verringerte das Spleißen und die Bildung des verzweigten Zwischenprodukts; verlängerte Inkubation mit Hydroxylamin, freiem Cystein oder Thiolreagentien führte im Vergleich zum Wildtyp-Vorläuferprotein zu einer Zunahme der Spaltung an der N-terminalen Spleißstelle, was auf die Spaltung einer Thioesterbindung an der N-terminalen Spleißstelle schließen lässt.^[60, 80] Im MIP-System kommt es auch zur Hydroxylamin-induzierten Spaltung an Cys1, wenn durch zusätzliche Mutationen die Bildung eines verzweigten Produkts blockiert ist; demnach scheint also bei der Spaltungsreaktion die Bildung eines linearen (Thio-)Esters der eines verzweigten (Thio-)Esters voranzugehen.^[60] Die Position der Hydroxylamin-Spaltstelle wurde durch massenspektrometrische und HPLC-Analyse eines bei der Bromcyan-Spaltung erhaltenen C-terminalen Fragmentes bestätigt, das aus dem gespaltenen N-Extein stammte und mit einem Modellpeptid verglichen wurde, das den erwarteten C-terminalen Hydroxamatrest enthält.^[80] Ähnliche Ergebnisse wurden später für das Sce-VMA-Intein erhalten.^[66]

Zusammengenommen sprechen diese Daten stark für eine $N \rightarrow O(S)$ -Acylwanderung an der stromaufwärts gelegenen Spleißstelle als einleitenden Schritt des Protein-Spleißens gemäß Mechanismus 3. Zwei Familien vermeintlicher Inteine enthalten allerdings N-terminal einen Alaninrest (siehe Abschnitt 1.3). Es gibt erste Hinweise darauf, dass wenigstens eines dieser Inteine zum Spleißen fähig ist.^[81] Wir schlagen vor, dass diese Ala1-enhaltenden Inteine eine Spleißreaktion entsprechend Mechanismus 2 entwickelt haben, weil alle anderen Inteinmotive intakt sind.

2.7. Der Mechanismus des Protein-Spleißens

Der weithin akzeptierte Mechanismus der Protein-Spleißreaktion bei den meisten Inteinen ist in Schema 4 dargestellt. Er besteht aus vier Schritten: 1) Angriff der Seitenkette des ersten Intein-Aminosäurerestes (Ser, Thr oder Cys) auf die Carbonylgruppe des in der Sequenz unmittelbar vorherge-



Schema 4. Der gegenwärtig akzeptierte Mechanismus des Protein-Spleißens und der Spaltung an einer einzelnen Spleißstelle. X steht entweder für das Sauerstoffatom in der Seitenkette von Ser oder Thr oder für das Schwefelatom in der Seitenkette von Cys. In einigen Inteinen ist Asn durch Gln ersetzt. Inteine mit Ala1 spleißen wahrscheinlich nach Mechanismus 2 (Schema 3). Alle tetraedrischen Zwischenprodukte und Protonenübertragungsschritte sind der besseren Übersicht halber weggelassen. Stellen im Reaktionsweg, an denen die Spaltung an nur einer Spleißstelle stattfinden kann, sind ebenfalls angegeben. Weitere Einzelheiten siehe Text. M steht für das N-Extein, P für das C-Extein. Eine C-terminale Succinimidgruppe ist durch das Zeichen > angezeigt.

henden Aminosäurerestes, wodurch es zur Acyl-Verschiebung vom N-Extein an die Seitenkette des ersten Intein-Aminosäurerests kommt; 2) Umesterungsreaktion, bei der die Hydroxy- oder Thiolgruppe des ersten C-Extein-Aminosäurerests (Ser, Thr oder Cys) die (Thio-)Esterbindung angreift; Ergebnis ist die Übertragung des N-Exteins auf die C-terminale Spleißstelle; 3) Spaltung der Amidbindung am C-Terminus des Inteins durch Cyclisierung von Asn oder Gln unter Freisetzung des Inteins; 4) spontane Umlagerung der (Thio-)Esterbindung zwischen den verbundenen Exteinen zur stabileren Peptidbindung.

Bei der Prozessierung des Hedgehog-Proteins tritt eine interessante Variante der Intein-vermittelten Umesterungsreaktion auf.^[48] Im Hedgehog-System kommt es zur Spaltung der Thioesterbindung zwischen der Signal- und der Autoprocessingdomäne, wenn die Hydroxygruppe von Cholesterin den Thioester nucleophil angreift. Dabei entsteht eine kovalente Bindung zwischen Cholesterin und dem C-Terminus der Signaldomäne, die dazu dient, die Signaldomäne an der Zelloberfläche zu verankern, wo sie wichtige Prozesse der Embryonalentwicklung steuert.^[48]

2.8. Nebenreaktionen

Zusätzlich zu den erwarteten Produkten, dem ausgeschnittenen Intein und den verknüpften Exteinen, entstehen bei Protein-Spleißreaktionen gewöhnlich noch Produkte von Nebenreaktionen, besonders dann, wenn die ungespleißten Vorstufen in heterologen Wirtsorganismen exprimiert werden oder wenn Inteine zwischen fremden Exteinen kloniert werden. Diese Produkte entstehen bei der Spaltung an einer oder beiden Spleißstellen ohne gleichzeitige Ligation und sind eine klare Erweiterung des Mechanismus (Schema 4). Die Spaltung an der N-terminalen Spleißstelle entsteht durch die Konkurrenz zwischen der Hydrolyse des linearen (Thio-)Ester-Zwischenproduktes und der Umesterung oder durch die Hydrolyse des (Thio-)Esters im verzweigten Zwischenprodukt vor der Cyclisierung von Asn oder Gln. Zur Spaltung an der C-terminalen Spleißstelle kommt es bei Asn- oder Gln-Cyclisierung ohne Übertragung des N-Exteins. Mutiert man für die Katalyse wichtige Aminosäuren an einer Spleißstelle, nimmt im allgemeinen die Spaltungs-Nebenreaktion an der anderen Spleißstelle zu.^[6, 9, 14, 60, 66, 74] Ersetzt man zum Beispiel den N-terminalen Aminosäurerest des Inteins durch Ala, kommt es ausschließlich zur Spaltung an der C-terminalen Spleißstelle, während der Austausch des C-terminalen Asn-Restes im Intein gegen Ala zur Spaltung an der N-terminalen Spleißstelle führt.

2.9. Die Rolle der proximalen Extein-Aminosäuren

Die Spleißreaktion verläuft bei chimären Vorläuferproteinen im allgemeinen langsamer als bei nativen Vorstufen und unter Bildung von mehr unkorrekten Spaltungsprodukten, die nur an einer Spleißstelle gespalten sind.^[4–10] Nach unserer Erfahrung ist das Spleißen bei weniger als 25 % aller chimären Vorläuferproteine erfolgreich. Es ist außerdem oft tempera-

turabhängig, was darauf hindeutet, dass die Faltung des Inteins beeinträchtigt ist.^[4, 8, 13] Da alle drei Nucleophile, die beim Spleißvorgang eine Rolle spielen, an den Spleißstellen liegen, ist es wahrscheinlich, dass suboptimale proximale Exteinsequenzen das Spleißen verhindern, indem sie das aktive Zentrum entweder chemisch, sterisch oder durch Störung des Einpassens des Substrats in das aktive Zentrum destabilisieren.

Obwohl der Aminosäurerest –1 an der N-terminalen Spleißstelle nicht aktiv als Nucleophil oder unterstützende Gruppe an der Spleißreaktion beteiligt ist, ist sein Carbonyl-Kohlenstoffatom das Ziel von drei der vier nucleophilen Substitutionen. In verschiedenen Untersuchungen wurde gezeigt, dass proximale N-Exteinreste die Spleißreaktion hemmen oder blockieren oder die Spleißaktivität von Defektmutanten wieder herstellen können.^[6, 8, 14, 76, 77] In zwei systematischen Untersuchungen wurde bestätigt, dass einige –1-Substitutionen das Spleißen der Inteine von Sce-VMA und von Gyrase A aus *Mycobacterium xenopi* (Mxe-GyrA) blockieren können, wobei allerdings Reste, die das Spleißen in einem Intein blockieren, in einem anderen Intein möglicherweise kaum Effekte haben.^[6, 14] Aminosäuren, die den ursprünglichen Resten in der Position –1 am ähnlichsten waren, funktionierten in diesen Untersuchungen am besten. Gewisse Aminosäuren in Position –1 können entweder die Geschwindigkeit der einleitenden Acyl-Verschiebung verringern oder die Umesterung blockieren.^[14] Aufgrund dieser Beobachtungen kann man vermuten, dass Inteine sich zusammen mit ihren Exteinsequenzen dahin gehend entwickelt haben, dass die Spleiß-Effizienz maximiert wurde, und dass es daher am günstigsten ist, Insertionsstellen für die Inteine auszuwählen, die den nativen Insertionsstellen ähnlich sind.

2.10. Die Katalyse erleichternde Faktoren und Inteinstruktur

Wie in Abschnitt 2.1. diskutiert wurde, wird in Serin- und Cystein-Proteasen die Katalyse dadurch erleichtert, dass durch ein „Oxyanionen-Loch“ die Elektrophilie der Substrat-Carbonylgruppe gesteigert wird und über ein Ladungs-Verschiebungssystem die Nucleophilie der angreifenden Hydroxy- oder Thiolgruppe erhöht wird.^[50] In Inteinen ist vermutlich die gleiche Strategie verwirklicht. Aus Mutagenese-Untersuchungen lässt sich folgern, dass der konservierte His-Rest in Block B bei Reaktionen am N-Terminus des Inteins^[77] und beim Hedgehog-Protein^[46, 82] unterstützend wirkt, doch ist es nicht gelungen, Gleiches für die Rolle des His-Restes im vorletzten Block G des Inteins zu zeigen.

Dieser His-Rest ist für den Spleißvorgang und die Spaltung der C-terminalen Spleißstelle des Psp-GBD-Pol-Inteins erforderlich.^[60] Allerdings scheint der Histidinrest aus Block G nicht allgemein für das Spleißen nötig zu sein, obwohl Mutationen dieses Restes normalerweise die Spleißreaktion in unterschiedlichem Ausmaß – abhängig vom Intein, vom Extein und von der substituierten Aminosäure – verhindern.^[6, 9, 60] Ca. 10 % aller Inteine haben sogar an dieser Stelle eine von Histidin verschiedene Aminosäure; wenn also in der Natur eine solche Mutation geschieht, kann das Intein sich

vermutlich so weiterentwickeln, dass es auch ohne Histidinrest in der vorletzten Position gespleißt wird.^[29] Von vier Inteinen, die natürlicherweise Ser, Ala, Phe oder Gly als vorletzten Aminosäurerest enthalten, konnte bei einem Test in *E. coli* nur das ClpP-Intein von *Chlamydomonas* nicht gespleißt werden.^[81, 83–85] Durch die Mutation des vorletzten Glycinrestes (im ClpP-Intein von *Chlamydomonas*)^[83] und des vorletzten Phenylalaninrestes (im Phosphoenolpyruvat-Synthase-Intein von *Methanococcus jannaschii*)^[85] zu Histidin wurde die Spleißreaktion jedesmal deutlich erleichtert. Im Unterschied dazu führte die analoge Mutation (Gly → His) im RNA-Polymerase-A'-Intein von *M. jannaschii* zu einer deutlichen Hemmung des Spleißvorgangs und zu einer Akkumulation des verzweigten Zwischenprodukts.^[85] Wenn man also bei Inteinen, die an vorletzter Position normalerweise andere Aminosäuren als Histidin enthalten, diese Position wieder zu Histidin zurückmutiert, kann dies je nach Intein die Spleißreaktion in *E. coli* verbessern oder verschlechtern. Möglicherweise spiegeln diese widersprüchlichen Ergebnisse Inteine in unterschiedlichen Evolutionsstadien nach Mutation des vorletzten Histidinrestes wider. Zusammengefasst sprechen diese Befunde dafür, dass der vorletzte Histidinrest wahrscheinlich die Spleißreaktion und die Weiterreaktion der Verzweigung durch die Asn-Cyclisierung erleichtert.

Strukturdaten unterstützen die Annahme, dass beide konservierten Histidinreste die Spleißreaktion erleichtern. Von zwei Inteinen und einem Teil der Autoprozessierungs-Domäne eines Hedgehog-Proteins wurde die Kristallstruktur bestimmt.^[45, 86, 87] Das Sce-VMA-Intein und der N-terminale Teil der Autoprozessierungs-Domäne des *Drosophila*-Hedgehog-Proteins wurden nach dem Herausschneiden aus dem Vorläuferprotein kristallisiert, während das Mxe-GyrA-Intein noch einen einzigen N-Extein-Rest (Ala – 1) und die Mutation Cys1 → Ser enthält. Durch Überlagerung des Ca-Rückgrats aller drei Proteine mit einem Großteil der Spleißdomäne^[42, 45, 48, 86, 87] kann man ein konserviertes Faltungsmuster identifizieren, das als HINT(Hedgehog/Intein)-Modul bezeichnet wird.^[45] Der HINT-Modul besteht aus zwei Subdomänen mit Ca-Rückgraten, die sich überlagern lassen; sie haben eine pseudo-zweifache Symmetrieachse und einen einzelnen hydrophoben Kernbereich gemeinsam.^[45] Die N- und C-Termini werden durch die Faltung des HINT-Moduls, das nur aus β -Faltblattsträngen und Schleifen besteht, in Nachbarschaft zueinander gebracht.

Unglücklicherweise fehlen in diesen Strukturen das C-Extein-Nucleophil und die zu spaltenden Bindungen mit Ausnahme der N-terminalen Spleißstelle im Mxe-GyrA-Intein. In dieser sind Ala – 1 und Ser 1 in *cis*-Konformation zueinander, was einen Teil der Triebkraft für das Spleißen darstellen kann.^[87, 88] Mit Alanin in Position – 1 ist die Spleißreaktion um 90% gehemmt und die Anfangsgeschwindigkeit der Thiol-induzierten N-terminalen Spaltung auf ein Fünftel verringert, wenn sich das Mxe-GyrA-Intein im MIP-Fusionsprotein befindet.^[14] Die Position von Ala – 1 und vom C-terminalen Asn ist wegen der Wechselwirkung zwischen der geladenen N-terminalen α -Aminogruppe von Ala – 1 und der Carboxylatgruppe des C-terminalen Asn 198 in dieser Struktur vielleicht anders als in einem größeren Vorläuferprotein (Abbildung 3).

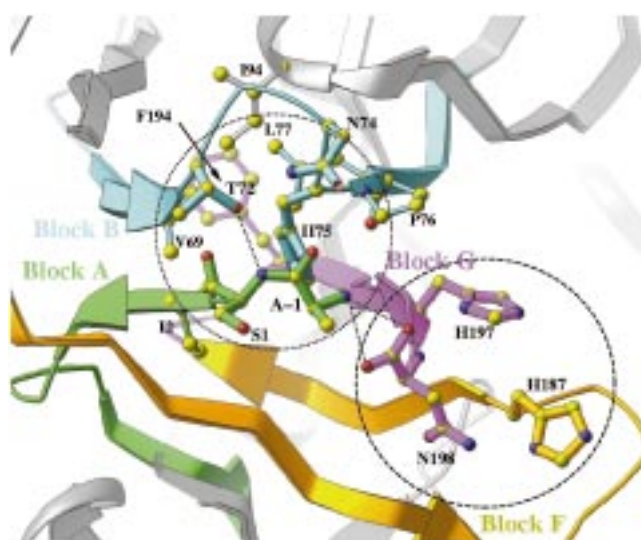


Abbildung 3. Die Struktur des aktiven Zentrums des Mxe-GyrA-Inteins. Aminosäuren, die im aktiven Zentrum des Mxe-GyrA-Inteins liegen, sowie die konservierten Motive sind auf Basis der Mxe-GyrA-Intein-Kristallstruktur^[87] eingezeichnet. Durch gestrichelte Kreise sind die zwei potentiellen Oxyanionen-Löcher angedeutet. Ein Paar von zwei gestrichelten Linien zeigt potentielle Wasserstoffbrückenbindungen zwischen den freien Amino- und Carboxygruppen des Proteins, die die Position dieser Gruppen in diesem Protein gegenüber der Position in einem Vorläuferprotein mit größeren Exteinen verzerren können. Eine weitere gestrichelte Linie deutet eine mögliche Wasserstoffbrückenbindung zwischen der Hydroxygruppe von Thr 72 in Block B und dem Stickstoffatom von Ser 1 bei der zu spaltenden Bindung an.

Die Kristallstrukturen geben dennoch Aufschluss über einige der Gruppen, die die Spleißreaktion unterstützen. Das Intein hat ein vorgeformtes aktives Zentrum mit zwei potentiellen Oxyanionen-Löchern. In allen drei Strukturen liegen verschiedene Reste aus Block B innerhalb der Distanz einer Wasserstoffbrückenbindung vom N-Terminus des HINT-Moduls entfernt und gehen möglicherweise im Vorläuferprotein eine Wechselwirkung mit dem Carbonyl-Sauerstoffatom oder der Amin-Abgangsgruppe der N-terminalen zu spaltenden Bindung ein; dabei bildet sich ein Oxyanionen-Loch, das an die Serin- und Cystein-Proteasen erinnert (siehe Abschnitt 2.1). Zu diesen Resten gehören die konservierten His und Thr (Asn im Sce-VMA-Intein) im Block B, aber auch noch weitere Reste (Abbildung 1 und Abschnitt 1.3). In der Mxe-GyrA-Intein-Struktur liegt His 75 aus Block B innerhalb der Distanz einer Wasserstoffbrückenbindung zum Ala – 1-Carbonyl-Sauerstoffatom, und Thr 72 und Asn 74 aus Block B liegen etwa eine Wasserstoffbrückenbindung von der Stickstoff-Abgangsgruppe der Hauptkette entfernt (Abbildung 3).^[87]

Verschiedene Reste könnten möglicherweise ein zweites Oxyanionen-Loch bilden und damit die Cyclisierung von Asn erleichtern (Abbildung 3). In beiden Inteinen ist der vorletzte Histidinrest so positioniert, dass er eine Wasserstoffbrückenbindung mit der C-terminalen Carboxylatgruppe eingeht und nicht mit dem Stickstoff-Nucleophil der Asn-Seitenkette. Diese Daten lassen vermuten, dass His entweder die sich aufbauende negative Ladung im tetraedrischen Zwischenprodukt der Asn-Cyclisierungsreaktion stabilisiert oder die Qualität der Hauptketten-Aminogruppe als Abgangsgruppe

verbessert und nicht etwa, indem es die Nucleophilie des angreifenden Stickstoffatoms der Seitenkette erhöht.^[45, 86, 87]

Keine Aminosäure steht in Wechselwirkung mit der Hydroxy- oder Thiolgruppe der Seitenkette an der N-terminalen Spleißstelle, um ihre Nucleophilie zu erhöhen.^[45, 86, 87] Alle drei Inteine mit bekannter Struktur enthalten einen N-terminalen Cys-Rest. Es wäre interessant, die Struktur eines Inteins mit einem N-terminalen Ser-Rest zu analysieren, denn solche Inteine brauchen wahrscheinlich wegen des hohen pK_a -Werts von Ser1 Unterstützung, um die Nucleophilie zu erhöhen.

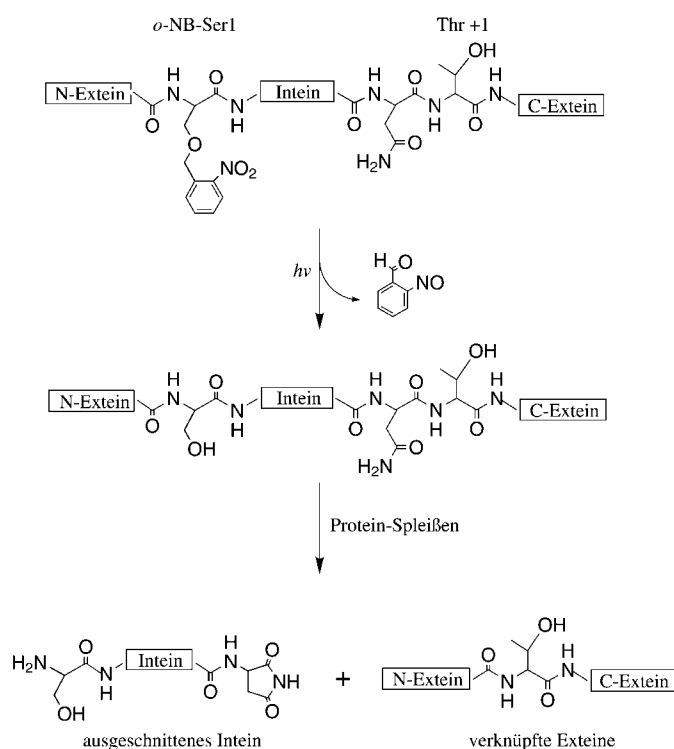
Zur Katalyse des Proteinspleißens gibt es mehr Fragen als Antworten. Welche Schritte des Reaktionsweges werden durch die Bildung des N-terminalen Oxyanionen-Loches an der Spleißstelle unterstützt, und gibt es ein zweites Oxyanionen-Loch an der C-terminalen Spleißstelle? Welche Konformationsänderungen, wenn überhaupt, sind nötig, um die vier gekoppelten nucleophilen Substitutionen zu erreichen, wobei dreimal das gleiche Carbonyl-Kohlenstoffatom angegriffen wird? Müssen sich die Seitenketten bewegen, um jeder neuen nucleophilen Gruppe den Angriff zu ermöglichen? Welche Reste sind an der Ausrichtung des angreifenden Nucleophils zum elektrophilen Zentrum beteiligt? Warum hemmen konservative Substitutionen von Ser, Thr und Cys im allgemeinen das Spleißen? Kommt die ungewöhnliche *cis*-Konformation auch im nativen Vorläuferprotein vor und wenn ja, schränken die entstehenden sterischen Hinderungen die Auswahl der möglichen Aminosäurereste an der Position – 1 ein? Ein fundiertes Verständnis des Protein-Spleißmechanismus umfasst auch die Bestimmung der Struktur der Vorläufer und der Zwischenprodukte von Inteinen, die Cys oder Ser an der Spleißstelle enthalten.

3. Anwendungen

3.1. Kontrolle des Spleißens

Anwendungen unter Kontrolle des Spleißens lassen sich in zwei Kategorien einteilen: pharmazeutische „Schalter“ und Werkzeuge für das Protein-Engineering. Inteine kommen in essentiellen Genen menschlicher Pathogene vor, z.B. in *Mycobacterium tuberculosis*, *Mycobacterium leprae* und *Candida tropicalis*. Weil Inteine-enthaltende Vorstufen inaktiv sind, wären chemische Reagentien, die das Spleißen blockieren, von erheblichem pharmakologischem Wert. Kontrollierbare Inteine, eingeschoben in einen fremden genetischen Kontext, ermöglichen es, ein Enzym auf Wunsch zu aktivieren, um so beispielsweise die Biosynthese cytotoxischer Proteine auszulösen oder das kontrollierte Ausschalten von Genen zu steuern. Das Spleißen wurde erstmals nach Insertion eines Inteins in eine suboptimale Extein-Umgebung oder nach Inteine-Mutation dadurch kontrolliert, dass die Biosynthese des Vorläufers und das Spleißen bei unterschiedlichen Temperaturen abliefen.^[4, 8, 13, 60]

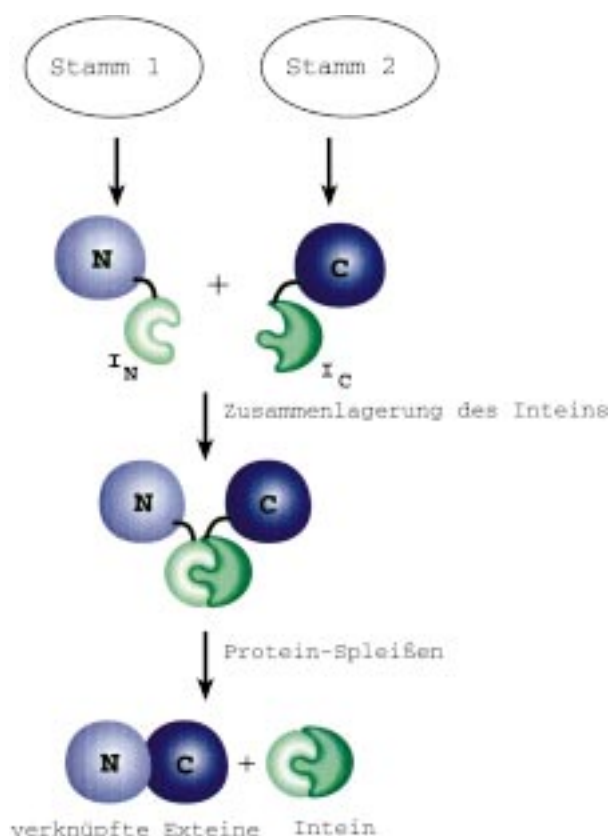
Der Spleißvorgang wurde auch durch neue photochemische Methoden und ortsspezifische Mutagenese kontrolliert (Schema 5).^[89] Mit einer chemisch aminoacylierten Suppressor-tRNA und einem In-vitro-Translationssystem^[90] wurde *O*-(2-



Schema 5. Kontrolle des Spleißens durch Verwendung photolabiler Gruppen. Das Protein-Spleißen des Tli-Pol-2-Inteins, das eine photoreaktive Schutzgruppe trägt, wird durch die Photolyse der 2-Nitrobenzylethergruppe von Ser1 eingeleitet.^[89] – o-NB = o-Nitrobenzyl.

Nitrobenzyl)serin anstelle von Ser1 in das Tli-Pol-2-Intein eingebaut. Die photolabile Schutzgruppe kann durch kurze Bestrahlung mit sichtbarem Licht aus dem Vorläuferprotein abgespalten werden, wobei Ser1 deblockiert wird. Photochemisch aktiviertes Protein-Spleißen sollte eine Vielzahl zeit-aufgelöster Untersuchungen ermöglichen, die für den Mechanismus der Spleißreaktion relevant sind.

Trans-Spleißen *in vitro* ist eine andere Methode, um das Spleißen zu kontrollieren. Dazu wird das Vorläufergen in zwei Fragmenten separat in verschiedenen Kulturen exprimiert (Schema 6).^[12, 21, 91, 92] Nach der Reinigung wurden die Vorläuferfragmente unter nativen oder denaturierenden Bedingungen je nach Löslichkeit und Notwendigkeit für die Reassoziierung gemischt. Bei der Zusammenlagerung des Inteins wird das Spleiß-Element aktiviert, was zur Verknüpfung der Exteine und zur Freisetzung der Inteine-Fragmente führt. In allen heute bekannten Fällen wurde die Vorstufe in der entbehrlichen Endonuclease-Domäne des Inteins geteilt. Bei zwei von drei Spaltstellen war ein funktionelles Trans-Spleißen im Psp-GBD-Pol-Intein möglich, und der freigesetzte geteilte Intein-Komplex behielt seine Homing-Endonuclease-Aktivität.^[12] Trans-Spleißen ermöglicht die Expression extrem cytotoxischer Proteine ebenso wie die segmentweise Modifikation oder Markierung von Proteinen. Trans-Spleißen kann auch als neues Werkzeug für die Untersuchung der Proteinfaltung eingesetzt werden. Bei einem anderen Beispiel, in dem die Natur durch die Biotechnologie nachgeahmt wird (oder umgekehrt), wurde die katalytische Untereinheit der replikativen DNA-Polymerase von *Synechocystis sp.* innerhalb des Inteins durch eine Genominversion geteilt, doch



Schema 6. Trans-Spleißen: Kontrolle des Spleißens durch Fragmentierung des Vorläuferproteins. Das Vorläufergen kann innerhalb der Inteinsequenz geteilt und jedes Fragment in getrennter Kultur exprimiert werden. Die Zusammenlagerung erfolgt ausschließlich *in vitro*. Aus den Intein-Fragmenten ($I_N + I_C$) wird durch Zusammenlagerung ein aktives Spleiß-Element gebildet, das dann die Verknüpfung der Exteinfragmente (N, C) vermittelt.

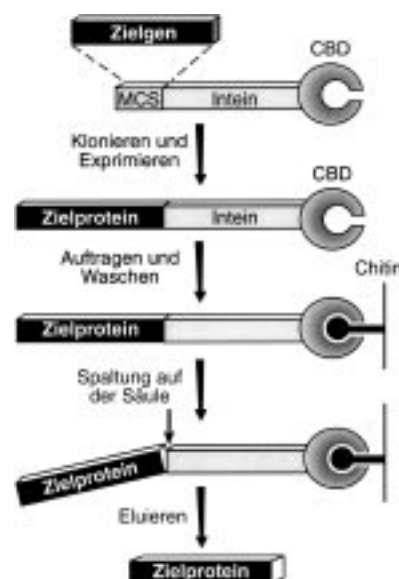
Trans-Spleißen in vivo stellt die DNA-Polymeraseaktivität wieder her und erhält so die Lebensfähigkeit dieses Organismus.^[84]

3.2. Vektoren für die Proteinreinigung

Nachdem man den Mechanismus des Protein-Spleißens verstanden hatte, konnte man ein Intein in einen allgemeineren Typ von selbstspaltendem Element umwandeln, in eine Autoprotease, die eine einzelne benachbarte Peptidbindung spaltet. Die Spaltung nur an der N-terminalen Spleißstelle läßt sich erreichen, indem man an der C-terminalen Spleißstelle einen oder mehrere Aminosäurereste mutiert, um die Asn-Cyclisierung zu verhindern, während die Thioesterbildung weiterhin möglich ist.^[14, 16, 60] Die Thioesterbindung kann dann mit thioispezifischen Reagentien (z.B. freiem Cystein) oder mit stickstoffhaltigen Nucleophilen (z.B. Hydroxylamin) gespalten werden. Die spezifische Aktivierung der C-terminalen Spleißstelle des Inteins wurde mit verschiedenen Methoden erreicht, z.B. durch die Mutation des vorletzten Histidinrestes des Sce-VMA-Inteins^[7] oder die Mutation des Aminosäurerestes an der Position – 1 in einem Temperatorkontrollierbaren Mxe-GvrA-Intein.^[14]

Die Umwandlung eines modifizierten Inteins, das nur noch eine Spleißstelle spalten kann, in einen Vektor für die

Proteinreinigung geschieht in zwei Schritten: 1) Einführung einer Mehrfachklonierungsstelle (multiple cloning site) zur Klonierung des Zielgens und 2) Klonieren des Gens für einen Marker für die Affinitätsreinigung innerhalb des Inteins oder benachbart zur inaktiven Spleißstelle (Schema 7). Vorläufer-



Schema 7. Intein-Reinigungsvektoren. Die Reinigung eines Zielproteins kann durch Klonieren des Zielgens in eine Mehrfachklonierungsstelle (multiple cloning site, MCS) im Leseraster mit der einzelnen aktiven N- oder C-terminalen Splice-Stelle des modifizierten Inteins erreicht werden.^[7, 14-16] Nach der Expression unter Bedingungen, die die Selbstspaltung unterdrücken, wird das Fusionsprotein durch Chromatographie an einer Chitin-Affinitätsäule über die als Affinitätsmarker neben der inaktiven Splice-Stelle eingefügte Chitin-bindende Domäne (CBD) isoliert. Die Spaltung wird dann auf der Säule durch verschiedene Methoden (z.B. Thiolreagentien, Temperaturänderung) induziert, woraufhin das freigesetzte Zielprotein eluiert werden kann.

proteine können in jedem eukaryontischen oder prokaryontischen Wirt exprimiert werden. Die Intein-Spaltung wird direkt auf der Affinitätssäule durchgeführt, wodurch das gewünschte Protein eluiert wird, während der Rest des Fusionsproteins zurückgehalten wird. Zwei Reinigungsvektoren sind im Handel erhältlich (New England BioLabs, Inc.), in denen der C-Terminus des Zielproteins an den aktiven N-Terminus des Inteins und eine 5 kDa große bakterielle Chitin-bindende Domäne als Affinitätsmarker an den inaktiven C-Terminus des Inteins gebunden ist. Das IMPACT-System (*int*ein-mediated *pur*ification with an *aff*inity chitin-binding *tag*) basiert auf einem modifizierten, 454 Aminosäuren langen Sce-VMA-Intein,^[16] während in einem zweiten Vektor ein modifiziertes, 198 Aminosäuren langes Mxe-GyrA-Intein verwendet wird.^[14, 15] In einem dritten käuflichen Reinigungsvektor, der ebenfalls Teil des IMPACT-Systems von New England BioLabs ist, ist eine Mehrfachklonierungsstelle für das Zielprotein neben dem C-Terminus eines modifizierten Sce-VMA-Inteins eingefügt; in diesem Intein wurde die zentrale Endonucleasedomäne durch die 5-kDa-Chitin-bindende Domäne ersetzt, und zehn Codons sind im gleichen Leseraster wie die N-terminale Spleißstelle.^[7] In diesem mutierten Intein wird die Asn-Cyclisierung durch die

Thiol-induzierte Spaltung des Peptids am N-Terminus des Inteins aktiviert.

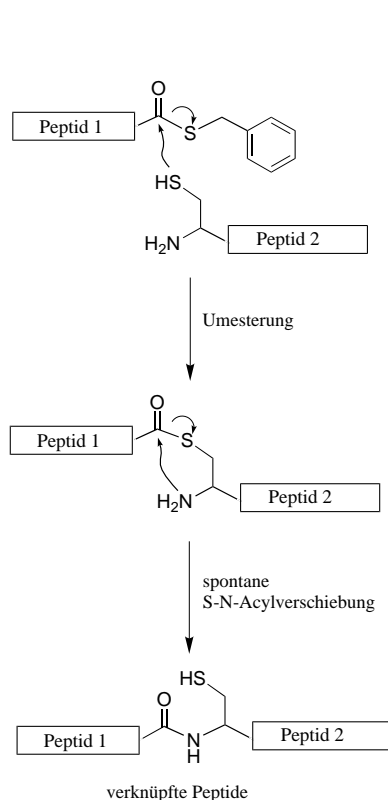
Die Intein-vermittelte Proteinreinigung hat gegenüber anderen Reinigungsmethoden verschiedene Vorteile. Man benötigt keine teure Protease, und die Spaltung findet nur an der Spleißstelle und nicht an anderen proteaseempfindlichen Stellen im Inneren des gewünschten Proteins statt. Außerdem kann die Reinigung des Proteins, das keine Aminosäuren aus dem Vektor enthält, in einem Schritt erreicht werden, da Reinigung und Spaltung auf derselben Säule durchgeführt werden.

3.3. Intein-vermittelte Polypeptidverknüpfung und verwandte Methoden

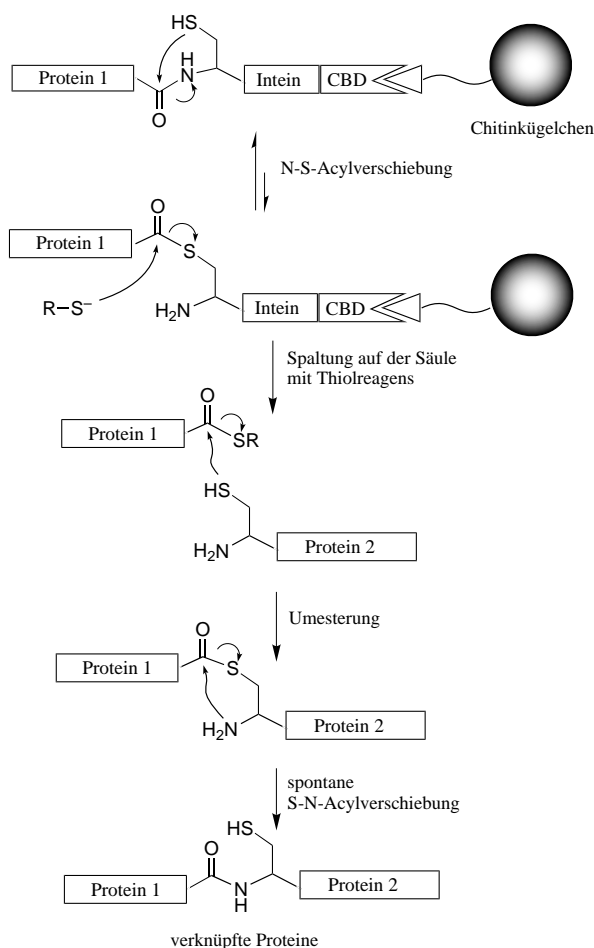
1998 wurden zwei vorher voneinander unabhängige Forschungsfelder zusammengeführt, das Protein-Spleißen und

die Protein-Semisynthese. Das Ergebnis war die Entwicklung einer neuen Protein-Engineering-Methode, die als Verknüpfung exprimierter Proteine (expressed protein ligation, EPL)^[17, 19, 20, 22] oder als Intein-vermittelte Proteinverknüpfung (intein-mediated protein ligation, IPL)^[15, 18] bezeichnet wurde. Die Aufklärung des Mechanismus des Protein-Spleißens erfolgte gleichzeitig mit der Entwicklung eines Verfahrens für die Protein-Semisynthese („native chemische Verknüpfung“).^[65] Bei der nativen chemischen Verknüpfung wird ein synthetisches Peptid mit einer C-terminalen α -Thioestergruppe mit einem zweiten Peptid oder Protein, das einen N-terminalen Cysteinrest enthält, gemischt (Schema 8 A). Es findet dann eine hocheffiziente und chemoselektive intermolekulare Reaktion statt, bei der ein Thioester-gebundenes Zwischenprodukt entsteht, das sich spontan über eine $S \rightarrow N$ -Acylumlagerung unter Bildung einer Peptidbindung zwischen den verknüpften Peptiden umwandelt. So durch-

A) native chemische Verknüpfung



B) Intein-vermittelte Proteinverknüpfung



Schema 8. Vergleich der nativen chemischen Verknüpfung mit der Intein-vermittelten Proteinverknüpfung. A) Die native chemische Verknüpfung^[65] geschieht durch Kupplung von zwei synthetischen Peptiden, von denen das N-terminale Peptid eine C-terminale Thioesterbindung und das C-terminale Peptid (oder Protein) einen N-terminalen Cysteinrest trägt. Nach Mischen der beiden Komponenten entsteht durch Umesterung ein Ligationsprodukt, in dem die beiden Peptide durch eine Thioesterbindung mit einer benachbarten freien α -Aminogruppe miteinander verbunden sind. Die Thioesterbindung lagert sich spontan zur stabileren Amidbindung um. B) Die Intein-vermittelte Proteinverknüpfung^[15, 17–20, 22] verläuft über identische chemische Reaktionen bei der Kupplung wie die native chemische Verknüpfung, aber die α -Thioestergruppe am N-terminalen Segment wird aus einem rekombinanten Protein, das mit einem spaltungskompetenten Intein fusioniert ist, durch Spaltung auf der Säule erzeugt. Dies umgeht die Einschränkungen hinsichtlich der Größe, die bei der chemischen Peptidsynthese bestehen. CBD bezeichnet die Chitin-Bindungsdomäne, die zur Immobilisierung des Inteins dient, und das Dreieck symbolisiert einen harzgebundenen Chitinrest. Alle tetraedrischen Zwischenprodukte und Protonenübertragungsschritte sind der besseren Übersicht halber weggelassen.

läuft die Protein-Semisynthese alle Schritte der Protein-Spleißreaktion mit Ausnahme der Asn-Cyclisierung, die das Intein aus den verknüpften Exteinen freisetzt (Schema 4). Die Protein-Semisynthese ist durch die Möglichkeiten begrenzt, große Peptide durch Festphasensynthese herzustellen, sowie durch die Tatsache, dass es nicht möglich ist, ein Polypeptid mit einer C-terminalen α -Thioestergruppe zu exprimieren oder selektiv eine C-terminale α -Thioestergruppe in ein nicht geschütztes Polypeptid einzuführen.

IMPACT und ähnliche Intein-Vektoren (siehe Abschnitt 3.2) sind einzigartige Werkzeuge für die Biosynthese einer α -Thioestergruppe am C-Terminus eines Proteins; diese kann dann für verschiedene Arten chemoselektiver Kondensationsreaktionen einschließlich der Protein-Semisynthese genutzt werden. Andere Intein-Vektoren liefern ein Polypeptid mit einem N-terminalen Cysteinrest nach Spaltung an der C-terminalen Spleiß-Stelle des Inteins.^[14, 15] Eine Alternative ist die Erzeugung eines N-terminalen Cys-Restes in einem biologisch synthetisierten Polypeptid durch Protease-verdau.^[20, 22] Bislang wurden das Mxe-GyrA-Intein,^[14, 18] das Sce-VMA-Intein^[16–20, 22] und das Intein der Ribonucleosiddiphosphat-Reduktase (RIR1) aus *Methanobacterium thermoautotrophicum*^[15] dazu genutzt, Polypeptide mit einer C-terminalen α -Thioestergruppe zu exprimieren.

Erste Untersuchungen befassten sich unter anderem mit der Kondensation synthetischer Peptide mit Proteinen, die Intein-vermittelt biologisch exprimiert wurden und eine α -Thioestergruppe zwischen dem C-Terminus des Zielproteins und dem N-Terminus des Inteins enthalten (Schema 8B), um cytotoxische Enzyme zu exprimieren,^[18] um den Effekt der Phosphorylierung auf Aktivität und Konformation von Kinasen zu testen^[17] und um Protein-Protein-Wechselwirkungen zu untersuchen.^[19] Die Protein-Semisynthese wurde in einem mehrstufigen Verfahren durchgeführt: 1) Verkürzen des Zielgens unmittelbar vor dem letzten Cystein-Codon oder an einer nichtkonservierten Stelle nahe dem C-Terminus; 2) Exprimieren des verkürzten Enzyms als Fusionsprotein mit dem N-Terminus eines modifizierten Inteins; 3) Reinigen des verkürzten Enzyms durch Affinitätschromatographie; 4) Mischen des verkürzten Enzyms mit einem synthetischen Peptid, das den fehlenden Rest des Enzyms mit einem N-terminalen Cys-Rest enthält, und Kondensieren in Gegenwart eines Thiolreagens wie 2-Mercaptoethansulfonsäure (MESNA) oder Thiophenol. Eine andere Möglichkeit ist, zuerst mit dem Thiolreagens das verkürzte Protein als α -Thioester von der Säule freizusetzen und anschließend das synthetische Peptid hinzuzufügen (Schema 8B). Die verknüpfte RNase A musste zur Aktivierung erneut gefaltet werden, nicht dagegen das Restriktionsenzym HpaI.^[18] Die C-terminalen Src-Kinase-Proteine (Csk; 50 kDa) katalysieren die Phosphorylierung eines konservierten Tyrosinrestes im C-terminalen Teil von Proteinen (SH2-Domäne), die zur Familie der Src-Kinasen gehören.^[17] In einem ähnlichen Synthesansatz mit einem IMPACT-Vektor (Sce-VMA-Intein) und Thiophenol wurde ein phosphoryliertes oder unphosphoryliertes synthetisches Peptid mit einem verkürzten menschlichen Csk-Protein kondensiert. Die resultierende Kinase war messbar aktiver, wenn das phosphorylierte Peptid verwendet wurde, wodurch die Bedeutung von Phosphotyrosin für die Aktivität bestätigt

wurde. An jedes Peptid war ein Fluoreszenzmarker gebunden, der die Analyse der Konformation durch partielle Proteolyse ermöglichte und durch den die Induktion einer Konformationsänderung in Csk durch Phosphotyrosin angezeigt wurde. Beim dritten Beispiel wurde ein C-terminales Peptid mit einer verkürzten σ^{70} -Untereinheit der *E. coli*-RNA-Polymerase ligiert; auf diese Weise wurde gezeigt, dass dieser Bereich des Proteins an einer Protein-Protein-Wechselwirkung im Holoenzym und an der Promotorerkennung beteiligt ist.^[19]

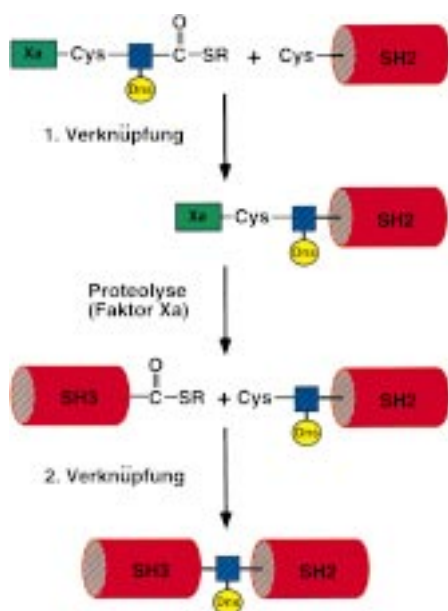
In der zweiten Generation der IPL wird ein biologisch exprimiertes Polypeptid, das einen N-terminalen Cys-Rest enthält, anstelle des synthetischen Peptids benutzt. Der N-terminale Cys-Rest kann mit zwei Intein-Vektoren erzeugt werden, in denen der N-Terminus des Zielproteins mit dem C-Terminus des Inteins fusioniert ist,^[14, 15] nicht aber mit dem kommerziellen IMPACT-Sce-VMA-Intein-System mit einer spaltbaren C-terminalen Spleißstelle^[7] (siehe Abschnitt 3.2). Eine andere Möglichkeit zur Erzeugung eines N-terminalen Cys-Restes besteht darin, ein exprimiertes Polypeptid mit einer Protease zu spalten.^[20, 22] In jedem Fall werden zwei große Polypeptide in Bakterien in löslicher Form exprimiert und kondensiert, ohne dass eine Denaturierung oder Renaturierung nötig wäre.^[15, 20, 22]

Mit den gleichen Techniken kann der N- oder C-Terminus eines Proteins spezifisch mit jedem Typ von Marker versehen werden, der entweder in ein Peptid eingebaut werden kann oder das Intein aus dem α -Thioester-Zwischenprodukt verdrängen kann. So bildet beispielsweise die Autoprozessierungs-Domäne des Hedgehog-Proteins eine ähnliche Thioesterbindung, die das Anfügen von Cholesterin an den C-Terminus der freigesetzten Signaldomäne durch eine Umesterungsreaktion auf die Hydroxygruppe von Cholesterin ermöglicht.^[46, 48]

Ein wichtiger Durchbruch bei der Anwendung der NMR-Spektroskopie für die Strukturaufklärung von Proteinen wurde mit IPL erreicht.^[20] Die NMR-Analyse ist durch die abnehmende Auflösung bei steigender Zahl von Aminosäureresten begrenzt. Durch die abschnittsweise Markierung eines Proteins wird die Komplexität des Spektrums verringert. So konnten große Proteine untersucht werden, deren Strukturen in Lösung bis dahin durch NMR-Spektroskopie nicht zugänglich waren. Mit IPL wird ein Segment oder eine Domäne eines großen Proteins während der Expression in Bakterien mit ^{13}C oder ^{15}N isotoopenmarkiert und dann mit dem restlichen unmarkierten Proteinteil (Segment oder Domäne), der ebenfalls in Bakterien exprimiert wurde, verknüpft. In einer Serie getrennter Experimente werden verschiedene Teile des Proteins vor ihrer Verknüpfung durch IPL markiert, so dass die Struktur der markierten Region im Kontext des gesamten gefalteten Proteins bestimmt werden kann. Diese Technik ermöglicht es auch, den Effekt von Ligandenbindung oder der Wechselwirkung zwischen Domänen durch Beeinflussung der chemischen Verschiebung zu untersuchen und damit Beziehungen zwischen Struktur und Aktivität NMR-spektroskopisch aufzuklären. Ähnliche Strategien abschnittsweiser Markierung sind durch Trans-Spleißen möglich^[12] (Abschnitt 3.1), und es wurde auch schon über die NMR-Analyse kleiner Polypeptide nach Trans-Spleißen

berichtet.^[21] Obwohl man mit Trans-Spleißen und IPL zum gleichen Ergebnis kommt, ist das Trans-Spleißen durch die Schwierigkeiten bei der Suche nach geeigneten Insertionsstellen für das Intein und durch die Notwendigkeit, ein aktives Intein zu rekonstituieren, begrenzt und in manchen Fällen durch den Zwang, die Vorläufer-Fragmente vor dem Zusammenfügen zu denaturieren (Abschnitt 3.1).

In einer anderen eleganten Serie von Experimenten haben Muir et al.^[22] mit der Intein-vermittelten Verknüpfung exprimierter Proteine einen Protein-Biosensor zwischen die Src-Homologie-3(SH3)- und die Src-Homologie-2(SH2)-Domäne der Abelson-Nicht-Rezeptor-Protein-Tyrosinkinase (Abl) eingeführt.^[22] Ein synthetisches Peptid, das den auf Umgebungsänderungen ansprechenden Fluorophor 5-(Dimethylamino)naphthalin-1-sulfonamid (Dns) enthielt, wurde nacheinander unter Verwendung von IPL und nativen chemischen Verknüpfungsstrategien an den C-Terminus der Abl-SH2-Domäne und den N-Terminus der Abl-SH3-Domäne gebunden (Schema 9). Das so entstandene semisynthetische Protein



Schema 9. Sequentielle Kondensation von Proteinfragmenten und Einbau eines Proteinbiosensors. Die sequentielle Kondensation synthetischer Peptide oder biologisch exprimierter Polypeptide durch eine Kombination aus Intein-vermittelter Verknüpfung exprimierter Proteine und nativer Peptidverknüpfung ermöglicht den Einbau eines Dns-Biosensors zwischen die Abl-Kinasedomänen SH2 und SH3.^[22] Nach Kondensation des synthetischen Peptids mit dem C-terminalen Polypeptid wird das Fusionsprotein mit der Protease Faktor Xa behandelt, so dass ein neues C-terminales Cys freigelegt wird. Das Fusionsprotein wird dann mit dem N-terminalen Polypeptid zum reifen Protein umgesetzt. Dieselbe Technik kann benutzt werden, um jede nichtnatürliche Aminosäure oder ein chemisches Reagens an jeder gewünschten Stelle in ein Protein einzubauen. – Xa: Peptid mit einer Spaltstelle für Faktor Xa.

trägt eine umgebungssensitive Dansylgruppe im Verbindungsglied zwischen der SH2- und der SH3-Domäne, die auf Konformationsänderungen durch die Bindung von Liganden reagiert. Diese Strategie des Protein Engineering ist prinzipiell auf die Insertion jeder Art von Marker, sei es ein Fluorophor oder eine nichtnatürliche Aminosäure, in Pro-

teine jeder Größe oder Sequenz anwendbar; die Limitationen von Methoden, die auf chemisch falsch acylierten Suppressor-tRNAs^[90] oder der Festphasensynthese von Peptiden beruhen, spielen in diesem Fall keine Rolle.

Vor kurzem wurden Inteine auch benutzt, um cyclische Proteine^[93] und Proteine mit einem für die Protein-Semisynthese geeigneten C-terminalen Thiocarboxylatrest herzustellen.^[94] Die Thiocarboxylatgruppe wurde mit Hilfe eines 66 Aminosäuren langen Peptides (ThiS) eingeführt, das als Schwefeldonor für die Biosynthese des Thiazolrestes von Thiamin fungiert. Das Peptid wurde mit dem N-Terminus des Sce-VMA-Inteins im IMPACT-CN-Vektor fusioniert, und der so gebildete Thioester wurde mit Ammoniumsulfid gespalten. Fünf Milligramm dieses ThiS-COSH-Peptides wurden aus einem Liter Kulturmedium gewonnen.^[94]

4. Zusammenfassung und Ausblick

Nach weniger als zehn Jahren stellt sich das Protein-Spleißen nicht mehr als biologische Merkwürdigkeit dar, sondern als wohlverstandener, weit verbreiteter Prozess, der die Basis vieler Anwendungen im Protein Engineering geworden ist. Wir sind allerdings noch immer unsicher bezüglich der biologischen Funktion der Inteine. Manche vermuten, dass die Intein-Gene egoistische DNA-Elemente sind, die in andere Gene eindringen, wobei sie beim Ausschneiden kaum Folgen hinterlassen.^[13, 35, 37, 48, 95] Allerdings scheint das Protein-Spleißen als Selbst-Prozessierungsmechanismus wenig Raum für externe Kontrolle zu lassen, vor allem deshalb, weil die Reaktion unter physiologischen Bedingungen schnell abläuft. Im angestammten Wirtsorganismus sind noch keine natürlichen Vorläufermoleküle entdeckt worden, und man konnte auch die Wachstumsbedingungen nicht so beeinflussen, dass dieser Prozess im nativen Organismus moduliert wird. Es wurde auch schon vermutet, dass Inteine vielleicht eine wichtige Rolle in einem frühen Stadium der Evolution gespielt haben.^[95] Trans-Spleißen und Intein-vermittelte Proteinverknüpfung legen nahe, dass Inteine ein Mittel für die statistische Neukombination von Domänen war, um verbesserte Enzyme auf Proteinebene zu erzeugen, besonders vor der Entwicklung ausgefeilter Rekombinationswege für die DNA. Als die DNA-Rekombination erst einmal etabliert war, waren die Intein-Gene vielleicht Inseln der Homologie für die Rekombination von Genen, die kleinere Proteinmodule codieren.

Obwohl ein allgemeiner Mechanismus des Protein-Spleißens aufgeklärt worden ist (Schema 4), gibt es in der Natur Ausnahmen von diesem Mechanismus. Die größte Lücke in unserem Verständnis des Protein-Spleißens ist, wie das Intein die vier nucleophilen Substitutionen erleichtert. Kreative Ansätze bei der Strukturaufklärung der Vorläufermoleküle und der Zwischenprodukte sind nötig, um dieses Problem zu lösen. Wie unterstützen Inteine zum Beispiel den Angriff des Stickstoffatoms der Asn-Seitenkette auf die Carbonylgruppe der Hauptkette anstelle des üblicheren Angriffs des Stickstoffatoms der Hauptkette auf die Carbonylgruppe in der Asn-Seitenkette? Da die Protein-Spleißreaktion aus gut bekannten chemischen Schritten besteht, wurden schnell

Anwendungen entwickelt, bei denen Inteine verwendet werden. Obwohl Proteinreinigungs-Vektoren auf Intein-Basis die erste von Inteinen abgeleitete Anwendung war, werden sich Intein-vermittelte Proteinverknüpfung und verwandte Anwendungen wahrscheinlich als die wichtigsten und am weitesten verbreiteten Techniken herausstellen, die auf Inteinen basieren. Durch die Möglichkeit, mit Intein-Vektoren chemoselektiv große Polypeptide beliebiger Sequenz herstellen zu können, die α -Thioestergruppen oder N-terminale Cysteinreste enthalten, können die Grenzen der Festphasen-Peptidsynthese umgangen werden. Inteine stellen auch ein neues Target für die Wirkstoffsuche bei verschiedenen Krankheiten dar. Im letzten Jahr sind viele neue Protein-Engineering-Anwendungen entwickelt worden, in denen IPL genutzt wird, um cytotoxische Proteine zu synthetisieren, um Fluoreszenzmarker an Zellen anzuheften, um Proteine abschnittsweise für die NMR-spektroskopische Analyse zu markieren, um Proteine zu cyclisieren und um C-terminale Thiocarboxylatgruppen oder Fluorophore in Proteine einzuführen. Die Liste der Anwendungen von Inteinen wird weiterwachsen, wenn immer mehr Chemiker sich der Möglichkeiten von Intein-basierten Methoden bewusst werden.

Wir widmen diesen Artikel dem Gedenken an Prof. Holger W. Jannasch (1927–1998) vom Woods Hole Oceanographic Institute, dessen Erforschung extremer Biotope zur Charakterisierung vieler Mikroorganismen führte, darunter *Thermococcus litoralis* und *Pyrococcus* sp. GB-D, aus denen die inteinhaltigen Gene kloniert wurden, mit denen wir den Protein-Spleißmechanismus aufklären konnten. Wir danken unseren Kollegen von New England BioLabs, Virginia Cornish und Tom Muir für hilfreiche Diskussionen, Karen Sandman, Karen Noren, Tom Evans und Maurice Southworth für die kritische Durchsicht des Manuskripts und Tom Muir für die Erstellung von Schema 9.

Eingegangen am 10. März 1999 [A331]
Übersetzt von Dr. Burkard Neuß, Jülich

- [1] F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner, M. Belfort, *Nucleic Acids Res.* **1994**, 22, 1125–1127.
- [2] F. B. Perler, M. Q. Xu, H. Paulus, *Curr. Opin. Chem. Biol.* **1997**, 1, 292–299, zit. Lit.
- [3] F. B. Perler, *Nat. Struct. Biol.* **1998**, 5, 249–52, zit. Lit.
- [4] M. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra, F. B. Perler, *Cell* **1993**, 75, 1371–1377.
- [5] M. Kawasaki, Y. Satow, Y. Anraku, *FEBS Lett.* **1997**, 412, 518–520.
- [6] S. Chong, K. S. Williams, C. Wotkowicz, M. Q. Xu, *J. Biol. Chem.* **1998**, 273, 10567–10577.
- [7] S. Chong, G. E. Montello, A. Zhang, E. J. Cantor, W. Liao, M. Q. Xu, J. Benner, *Nucleic Acids Res.* **1998**, 26, 5109–5115.
- [8] A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs, Jr., F. B. Perler, *J. Bacteriol.* **1997**, 179, 6378–6382.
- [9] A. A. Cooper, Y. Chen, M. A. Lindorfer, T. H. Stevens, *EMBO J.* **1993**, 12, 2575–2583.
- [10] E. O. Davis, P. J. Jenner, P. C. Brooks, M. J. Colston, S. G. Sedgwick, *Cell* **1992**, 71, 201–210.
- [11] H. Paulus, *Chem. Soc. Rev.* **1998**, 27, 375–386.
- [12] M. W. Southworth, E. Adam, D. Panne, R. Byer, R. Kautz, F. B. Perler, *EMBO J.* **1998**, 17, 918–926.
- [13] V. Derbyshire, D. W. Wood, W. Wu, J. T. Dansereau, J. Z. Dalggaard, M. Belfort, *Proc. Natl. Acad. Sci. USA* **1997**, 94, 11466–11471.
- [14] M. W. Southworth, K. Amaya, T. C. Evans, M. Xu, F. B. Perler, *BioTechniques* **1999**, 27, 110–120.
- [15] T. C. Evans, J. Benner, M. Q. Xu, *J. Biol. Chem.* **1999**, 274, 3923–3926.
- [16] S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus, M. Q. Xu, *Gene* **1997**, 192, 271–281.
- [17] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 6705–6710.
- [18] T. C. Evans, J. Benner, M. Q. Xu, *Protein Sci.* **1998**, 7, 2256–2264.
- [19] K. Severinov, T. W. Muir, *J. Biol. Chem.* **1998**, 273, 16205–16209.
- [20] R. Xu, B. Ayers, D. Dowburn, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **1999**, 96, 388–393.
- [21] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino, H. Nakamura, *J. Am. Chem. Soc.* **1998**, 120, 5591–5592.
- [22] G. J. Cotton, B. Ayers, R. Xu, T. W. Muir, *J. Am. Chem. Soc.* **1999**, 121, 1100–1101.
- [23] P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebel, T. H. Stevens, *Science* **1990**, 250, 651–657.
- [24] R. Hirata, Y. Ohsumi, A. Nakano, H. Kawasaki, K. Suzuki, Y. Anraku, *J. Biol. Chem.* **1990**, 265, 6726–6733.
- [25] E. O. Davis, S. G. Sedgwick, M. J. Colston, *J. Bacteriol.* **1991**, 173, 5653–5662.
- [26] F. B. Perler, D. G. Comb, W. E. Jack, L. S. Moran, B. Qiang, R. B. Kucera, J. Benner, B. E. Slatko, D. O. Nwankwo, S. K. Hempstead, C. K. S. Carlow, H. Jannasch, *Proc. Natl. Acad. Sci. USA* **1992**, 89, 5577–5581.
- [27] H. H. Gu, J. Xu, M. Gallagher, G. E. Dean, *J. Biol. Chem.* **1993**, 268, 7372–7381.
- [28] R. A. Hodges, F. B. Perler, C. J. Noren, W. E. Jack, *Nucleic Acids Res.* **1992**, 20, 6153–6157.
- [29] F. B. Perler, *Nucleic Acids Res.* **1999**, 27, 346–347.
- [30] F. B. Perler, G. J. Olsen, E. Adam, *Nucleic Acids Res.* **1997**, 25, 1087–1093.
- [31] S. Pietrokovski, *Protein Sci.* **1998**, 7, 64–71.
- [32] J. Z. Dalggaard, M. J. Moser, R. Hughey, I. S. Mian, *J. Comput. Biol.* **1997**, 4, 193–214.
- [33] X. Q. Liu, Z. Hu, *Proc. Natl. Acad. Sci. USA* **1997**, 94, 7851–7816.
- [34] H. Fsihi, V. Vincent, S. T. Cole, *Proc. Natl. Acad. Sci. USA* **1996**, 93, 3410–3415.
- [35] M. Belfort, P. S. Perlman, *J. Biol. Chem.* **1995**, 270, 30237–30240.
- [36] A. E. Gorbalenya, *Nucleic Acids Res.* **1998**, 26, 1741–1748.
- [37] F. S. Gimble, J. Thorner, *Nature* **1992**, 357, 301–306.
- [38] M. Belfort, R. J. Roberts, *Nucleic Acids Res.* **1997**, 25, 3379–3388, zit. Lit.
- [39] J. Z. Dalggaard, A. J. Klar, M. J. Moser, W. R. Holley, A. Chatterjee, I. S. Mian, *Nucleic Acids Res.* **1997**, 25, 4626–4638.
- [40] S. Pietrokovski, *Protein Sci.* **1994**, 3, 2340–2350.
- [41] M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, *Nucleic Acids Res.* **1998**, 26, 4409–4412.
- [42] F. B. Perler, *Cell* **1998**, 92, 1–4, zit. Lit.
- [43] W. Grindl, W. Wende, V. Pingoud, A. Pingoud, *Nucleic Acids Res.* **1998**, 26, 1857–1862.
- [44] Z. He, M. Crist, H. Yen, X. Duan, F. A. Quirocho, F. S. Gimble, *J. Biol. Chem.* **1998**, 273, 4607–4615.
- [45] T. M. Hall, J. A. Porter, K. E. Young, E. V. Koonin, P. A. Beachy, D. J. Leahy, *Cell* **1997**, 91, 85–97.
- [46] J. J. Lee, S. C. Ekker, D. P. von Kessler, J. A. Porter, B. I. Sun, P. A. Beachy, *Science* **1994**, 266, 1528–1537.
- [47] E. V. Koonin, *Trends Biochem. Sci.* **1995**, 20, 141–142.
- [48] a) J. A. Porter, S. C. Ekker, W. J. Park, D. P. von Kessler, K. E. Young, C. H. Chen, Y. Ma, A. S. Woods, R. J. Cotter, E. V. Koonin, P. A. Beachy, *Cell* **1996**, 86, 21–34; b) P. A. Beachy, M. K. Cooper, K. E. Young, D. P. von Kessler, W. Park, T. M. T. Hall, D. J. Leahy, J. A. Porter, *Cold Spring Harbor Symp. Quant. Biol.* **1997**, 62, 191–204.
- [49] T. R. Burglin, *Curr. Biol.* **1996**, 6, 1047–1050, zit. Lit.
- [50] A. R. Fersht, *Enzyme Structure and Mechanism*, W. H. Freeman, Reading, **1977**.
- [51] C. F. Barbas III, J. R. Matos, J. B. West, C. H. Wong, *J. Am. Chem. Soc.* **1988**, 110, 5162–5166.
- [52] T. Nakatsuka, T. Sasaki, E. T. Kaiser, *J. Am. Chem. Soc.* **1987**, 109, 3808–3810.

- [53] L. Abrahamsen, J. Tom, J. Burnier, K. A. Butcher, A. Kossiakoff, J. A. Wells, *Biochemistry* **1991**, *30*, 4151–4159.
- [54] A. Warshel, G. Naray-Szabo, F. Sussman, J. K. Hwang, *Biochemistry* **1989**, *28*, 3629–3637.
- [55] D. R. Corey, C. S. Craik, *J. Am. Chem. Soc.* **1992**, *114*, 1784–1790.
- [56] K. Iwai, T. Ando, *Methods Enzymol.* **1967**, *11*, 262–282.
- [57] T. C. Bruice, S. J. Benkovic in *Bioorganic Mechanisms, Vol. 1* (Hrsg.: R. Breslow, M. Karplus), W. A. Benjamin, New York, **1966**, S. 266–297.
- [58] W. P. Jencks, S. Cordes, J. Carriulo, *J. Biol. Chem.* **1960**, *235*, 3608–3614.
- [59] A. A. Cooper, T. H. Stevens, *BioEssays* **1993**, *15*, 667–673.
- [60] M. Xu, F. B. Perler, *EMBO J.* **1996**, *15*, 5146–5153.
- [61] N. D. Clarke, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11084–11088.
- [62] C. J. Wallace, *Protein Sci.* **1993**, *2*, 697–705.
- [63] M. Xu, D. G. Comb, H. Paulus, C. J. Noren, Y. Shao, F. B. Perler, *EMBO J.* **1994**, *13*, 5517–5522.
- [64] Y. Shao, H. Paulus, *J. Pept. Res.* **1997**, *50*, 193–198.
- [65] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776–779.
- [66] S. Chong, Y. Shao, H. Paulus, J. Benner, F. B. Perler, M. Q. Xu, *J. Biol. Chem.* **1996**, *271*, 22159–22168.
- [67] Y. Shao, M. Q. Xu, H. Paulus, *Biochemistry* **1995**, *34*, 10844–10850.
- [68] T. V. Brennan, S. Clarke, *Int. J. Pept. Protein Res.* **1995**, *45*, 547–553, zit. Lit.
- [69] C. E. Voorter, W. A. de Haard-Koekman, P. J. van den Oetelaar, H. Bloemendal, W. W. de Jong, *J. Biol. Chem.* **1988**, *263*, 19020–19023.
- [70] A. V. Klotz, *Bioorg. Chem.* **1993**, *21*, 83–94.
- [71] T. Geiger, S. Clarke, *J. Biol. Chem.* **1987**, *262*, 785–794.
- [72] S. Clarke, R. C. Stephenson, J. D. Lowenson in *Stability of Protein Pharmaceuticals, Part A* (Hrsg.: T. J. Ahern, M. C. Manning), Plenum, New York, **1992**, S. 1–29, zit. Lit.
- [73] R. C. Stephenson, S. Clarke, *J. Biol. Chem.* **1989**, *264*, 6164–6170.
- [74] R. Hirata, Y. Anraku, *Biochem. Biophys. Res. Commun.* **1992**, *188*, 40–47.
- [75] P. Vanderslice, W. C. Copeland, J. D. Robertus, *J. Biol. Chem.* **1988**, *263*, 10583–10586.
- [76] S. Nogami, Y. Satow, Y. Ohya, Y. Anraku, *Genetics* **1997**, *147*, 73–85.
- [77] M. Kawasaki, S. Nogami, Y. Satow, Y. Ohya, Y. Anraku, *J. Biol. Chem.* **1997**, *272*, 15668–15674.
- [78] J. N. Higaki, L. B. Evnin, C. S. Craik, *Biochemistry* **1989**, *28*, 9256–9263.
- [79] C. Guan, T. Cui, V. Rao, W. Liao, J. Benner, C. L. Lin, D. Comb, *J. Biol. Chem.* **1996**, *271*, 1732–1737.
- [80] Y. Shao, M. Q. Xu, H. Paulus, *Biochemistry* **1996**, *35*, 3810–3815.
- [81] M. W. Southworth, F. B. Perler, unveröffentlichte Ergebnisse.
- [82] J. A. Porter, D. P. von Kessler, S. C. Ekker, K. E. Young, J. J. Lee, K. Moses, P. A. Beachy, *Nature* **1995**, *374*, 363–366.
- [83] C. Huang, S. Wang, L. Chen, C. Lemieux, C. Otis, M. Turmel, X. Q. Liu, *Mol. Gen. Genet.* **1994**, *244*, 151–159.
- [84] H. Wu, Z. Hu, X. Q. Liu, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226–9231.
- [85] L. Chen, F. B. Perler, unveröffentlichte Ergebnisse.
- [86] X. Duan, F. S. Gimble, F. A. Quijcho, *Cell* **1997**, *89*, 555–564.
- [87] T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs, Jr., J. C. Sacchettini, *Nat. Struct. Biol.* **1998**, *5*, 31–36.
- [88] B. L. Stoddard, S. Pietrokovski, *Nat. Struct. Biol.* **1998**, *5*, 3–5.
- [89] S. N. Cook, W. E. Jack, X. Xiong, L. E. Danley, J. A. Ellman, P. G. Schultz, C. J. Noren, *Angew. Chem.* **1995**, *107*, 1736–1737; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1629–1630.
- [90] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182–188.
- [91] K. V. Mills, B. M. Lew, S. Jiang, H. Paulus, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3543–3548.
- [92] B. M. Lew, K. V. Mills, H. Paulus, *J. Biol. Chem.* **1998**, *273*, 15887–15890.
- [93] a) T. C. Evans, Jr., J. Benner, M. Q. Xu, *J. Biol. Chem.* **1999**, *274*, 18359–18363; b) J. A. Camarero, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 5597–5598.
- [94] C. Kinsland, S. V. Taylor, N. L. Kelleher, F. W. McLafferty, T. P. Begley, *Protein Sci.* **1998**, *7*, 1839–1842.
- [95] F. B. Perler, *Trends Biochem. Sci.* **1999**, *24*, 209–210, zit. Lit.