

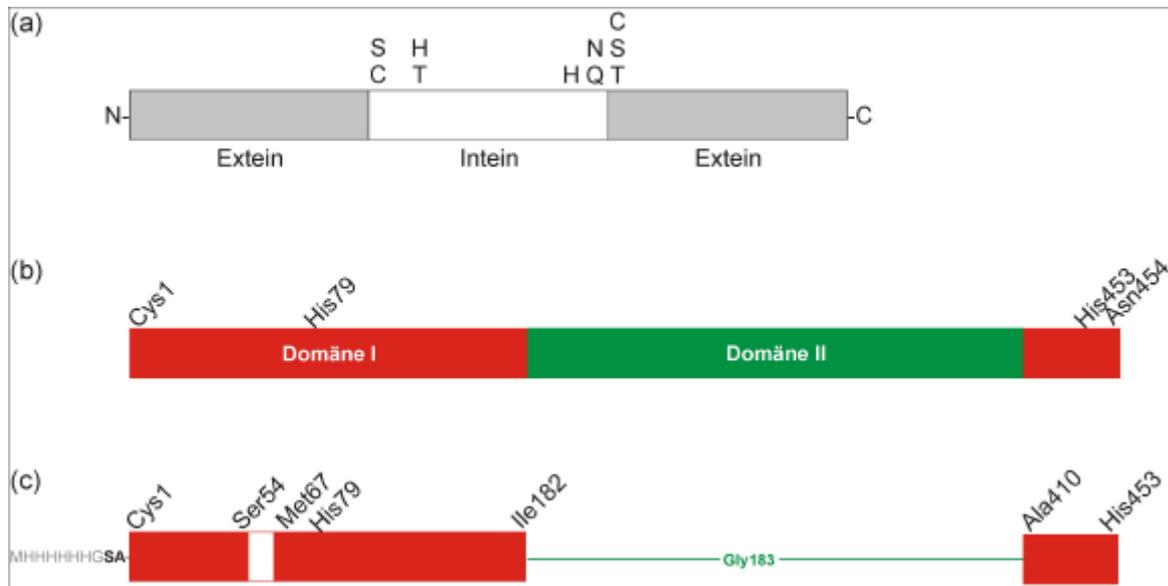
## 5. Die Homing Endonuklease PI-SceI von *Saccharomyces cerevisiae*

### 5.1. Einleitung

#### 5.1.1. Inteine und Homing Endonukleasen

Inteine, oder *interne Proteine*, sind Proteinsequenzen, die im Leseraster einer Vorläufer-Sequenz eingebettet sind und im Laufe der Zellreifung auf der Stufe des Primären Translationsproduktes und nicht auf der mRNA-Stufe herausgeschnitten werden. Daher der an Intron angelehnte Name. Die externen *Proteine* werden Exteine genannt, der Schneideprozess Protein Splicing. Inteine wurden in allen drei phylogenetischen Königreichen, also den Eukarya, den Prokaria und den Archaea, gefunden. Homing Endonukleasen, die im Allgemeinen mit dem Präfix „**I**-“ vor dem Drei-Buchstaben-Kürzel der Spezies bezeichnet werden, sind teilweise freie Inteine. Sie erhalten dann das Präfix „**PI**-“ für **P**rotein **I**nsert oder „protein intervening sequence“ (Perler *et al.*, 1994). Für Reviews siehe (Belfort und Roberts, 1997; Chevalier und Stoddard, 2001; Jurica und Stoddard, 1999; Mueller *et al.*, 1993). Homing Endonukleasen binden an bis zu 35 Basenpaare lange, asymmetrische Erkennungssequenzen und induzieren in Anwesenheit von  $Mg^{2+}$  einen Doppelstrang-Bruch. Dieser Schnitt geschieht an dem Allel, das defizient ist für dasjenige Intron, das die Homing Endonuklease codiert und induziert die gerichtete Integration des eigenen Gens („Homing“) (Belfort und Perlman, 1995; Lambowitz und Belfort, 1993). Nach den konservierten Sequenzmotiven klassifiziert werden vier Familien von Homing Endonukleasen unterschieden, die LAGLIDADG-, die GIY-YIG-, die H-N-H- und die His-Cys-Box-Familie (Belfort und Roberts, 1997; Jurica und Stoddard, 1999). Inteine besitzen einige konservierte Charakteristika, die für den Protein-Splicing-Vorgang von Bedeutung sind. So ist die N-terminale Aminosäure des Inteins (der erste Rest nach der „upstream splice junction“) stets ein Cystein oder Serin, während an der „downstream splice junction“ einem Histidin ein Asparagin oder Glutamin als C-Terminus des Inteins folgt. Der N-terminale Rest des

C-terminalen Exteins ist ein Cystein, ein Serin oder ein Threonin. Außerdem gibt es eine konservierte Stelle innerhalb des Inteins, die entweder von einem Histidin oder einem Threonin eingenommen wird (Abbildung 5.1a). (Noren *et al.*, 2000).



**Abbildung 5.1: Konservierte Aminosäuren von Homing Endonukleasen**

(a) Konservierte Intein-Charakteristika. (b) Native Form von PI-SceI.  
(c) PI-SceI Domäne I, Konstrukt nach Grindl *et al.* (1998).

Das in dieser Arbeit untersuchte Intein besitzt in seiner nativen Form die konservierten Reste Cystein1 und Histidin79, während Histidin453 und Asparagin454 den C-Terminus des Inteins bilden. Das N-terminale Extein beginnt mit einem Cystein (Abbildung 5.1b).

### 5.1.2. Die Homing Endonuklease PI-SceI

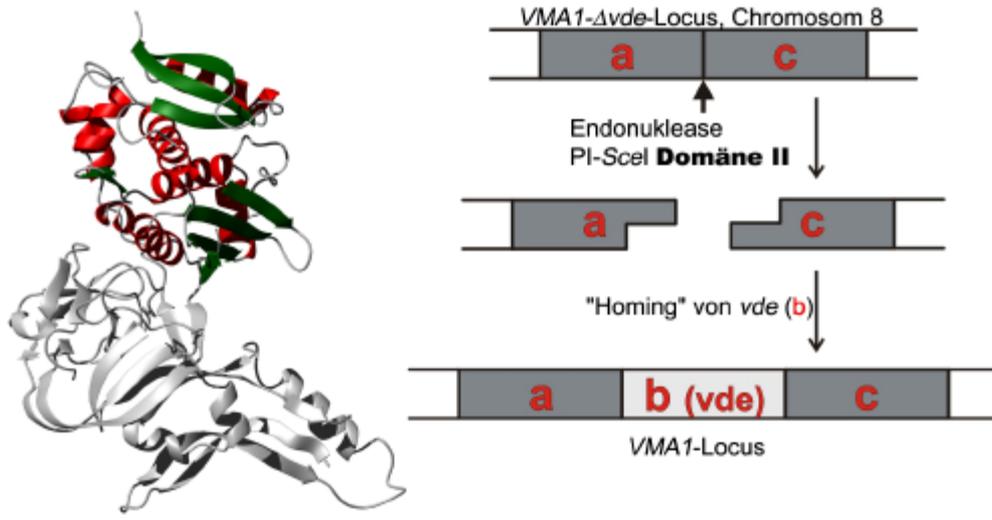
Das Protein Insert aus *Saccharomyces cerevisiae* (der Bäckerhefe) PI-SceI ist ein 50 kDa-Protein und besteht aus zwei Domänen. Die Kristallstruktur von PI-SceI wurde erstmals 1997 von Duan *et al.* (1997), PDB-Code 1vde, gelöst. Domäne I, die Protein-Splicing-Domäne, schneidet PI-SceI aus einem Vorläufer-Protein heraus (Kapitel 5.1.4, Seite 69) und Domäne II, die Endonuklease-Domäne, schneidet die DNA-Erkennungssequenz und initiiert damit den Homing-Prozess des eigenen Gens (Kapitel 5.1.3). D.h. PI-SceI ist am gleichen Prozess

sowohl auf DNA- als auch auf Protein-Ebene beteiligt. PI-SceI wurde erstmals 1990 beschrieben (Hirata *et al.*, 1990; Kane *et al.*, 1990) und wird maßgeblich in den Gruppen von Frederic Gimble und Alfred Pingoud charakterisiert und untersucht (Gimble und Thorner, 1992; Gimble und Thorner, 1993; Pingoud *et al.*, 1998; Wende *et al.*, 1996).

In dieser Arbeit wurde mit der Domäne I von PI-SceI gearbeitet, einem Konstrukt, in dem die Domäne II durch Glycin183 ersetzt ist (Grindl *et al.*, 1998), Abbildung 5.1c. Durch einen N-terminalen His<sub>6</sub>-tag konnte die Domäne I mit Affinitätschromatographie leicht aufgereinigt werden. Dem Konstrukt fehlt daher Asn454, damit die Protein-Splicing-Aktivität den tag nicht abspaltet. Obwohl es die Protein-Splicing-Domäne ist, ist sie hauptverantwortlich für die Bindungsstärke des Komplexes mit der DNA-Erkennungssequenz, während Domäne II, die Endonuklease, die DNA nur locker bindet (Gimble und Wang, 1996; Grindl *et al.*, 1998; He *et al.*, 1998; Wende *et al.*, 1996).

### 5.1.3. PI-SceI Domäne II: Endonuklease-Aktivität

PI-SceI bindet an den *VMA1-Dvde* Locus des Chromosoms 8 von *S. cerevisiae*. Dieser kodiert für die vakuoläre Membran-H<sup>+</sup>-ATPase (bzw. eine 69 kDa Untereinheit davon) und ist defizient (Symbol Δ) für das Gen *vde* („**V**MA1-**d**erived **e**ndonuclease“ = PI-SceI) Das endonukleolytische Zentrum von Domäne II nimmt den Doppelstrangbruch vor und initiiert damit die Insertion des gesamten Gens von PI-SceI, den Homing-Prozess (Abbildung 5.2) (Gimble und Thorner, 1992; Gimble und Thorner, 1993).

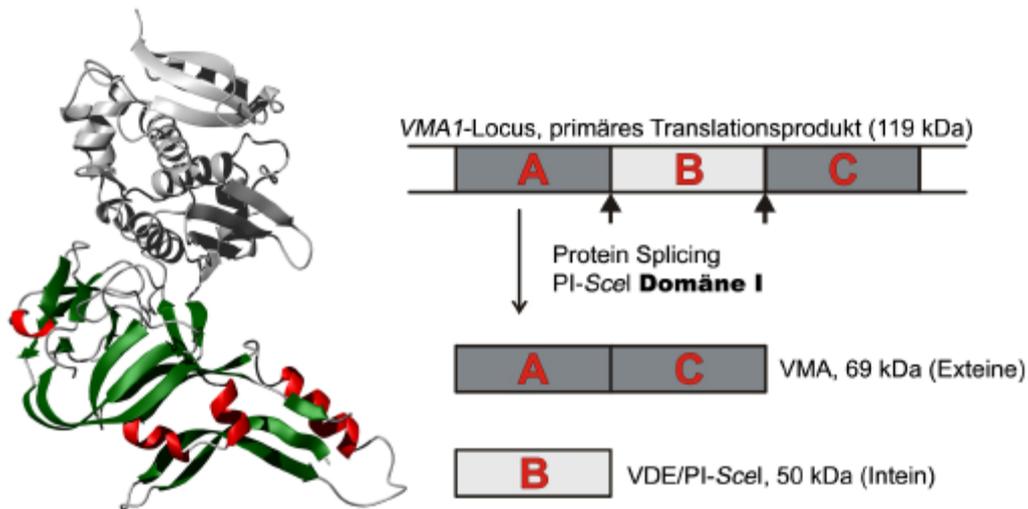


**Abbildung 5.2: Endonuklease-Aktivität von PI-SceI Domäne II**

Die Genabschnitte a und c codieren für die 69-kDa-Untereinheit der vakuolären Membran- $H^+$ -ATPase, das Gen b für PI-SceI = VDE.

Domäne II nimmt die homing-endonuklease-artige Faltung ein (alpha-beta(2)-alpha-beta(2)); zwei Ebenen a/b; antiparalleles Faltblatt der Topologie 1-2-4-3) und ist Mitglied in der gleichnamigen Superfamilie (Klassifizierung nach SCOP (Murzin *et al.*, 1995), CATH ID 3.10.28.10 (Orengo *et al.*, 1997)). Domäne II gehört zur Familie der Intein-Endonukleasen, die eine Tandem-Wiederholung dieses Faltungstypes besitzen. Eine weitere Familie bilden die Gruppe I mobile Intron-Endonukleasen. Ein Mitglied dieser Familie, I-CreI, ist insbesondere hilfreich für die Untersuchung des Endonuklease-Reaktionsmechanismus' von PI-SceI. Die Kristallstruktur des I-CreI-DNA-Komplexes (PDB-Code 1bp7 (Jurica *et al.*, 1998)) kann mit Domäne II von PI-SceI (PDB-Code 1vde (Duan *et al.*, 1997)) überlagert werden (r.m.s.d. 1.53 Å, Überlagerung von 324 C $\alpha$ -Atomen mit dem Swiss-Pdb-Viewer (Guex und Peitsch, 1997)). Beide gehören zur LAGLIDADG-Familie von Homing-Endonukleasen, wobei I-CreI ein über eine zweizählige Achse verbundenes Homodimer mit je einer LAGLIDADG-Sequenz ausbildet (Heath *et al.*, 1997) während das Monomer von PI-SceI Domäne II beide LAGLIDADG-Sequenzen enthält. Diese konservierten Aminosäuresequenzen bilden die beiden zentralen Helices aus, die hier strukturell sehr ähnlich sind und die an deren Ende die jeweiligen Aminosäuren platzieren, die für die endonukleolytische Aktivität verantwortlich gemacht werden (Abbildung 5.3). Im Fall von PI-SceI sind das Asp326 und Asp218, im Fall von I-CreI jeweils Asp20. Es wird aufgrund der strukturellen Analogie und biochemischer Experimente für PI-SceI angenommen, dass mit Hilfe des





**Abbildung 5.4: Protein-Splicing-Aktivität von PI-SceI Domäne I**

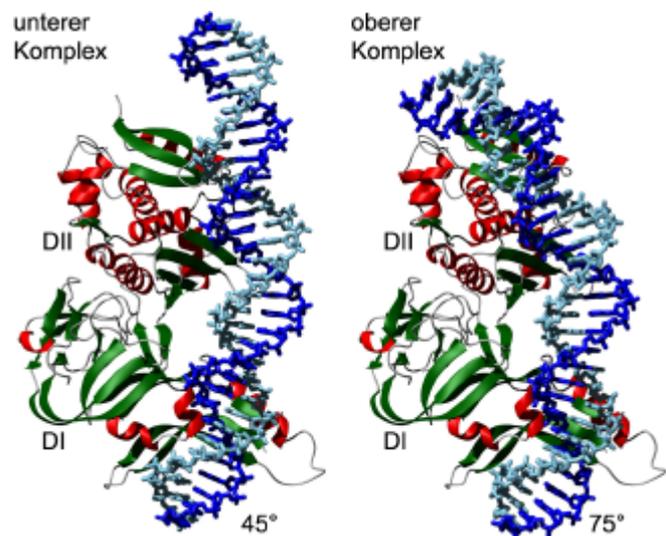
### 5.1.5. DNA-Bindung

PI-SceI bindet eine Erkennungssequenz am *VMA1-Dvde* Locus von Chromosom 8 in *S. cerevisiae* und führt einen Doppelstrangbruch herbei. Die minimale Bindungssequenz ist 17 Basenpaare lang (+5 bis +21) und enthält nicht die Schnittstellen. Nummeriert werden die Basenpaare ausgehend vom Zentrum der Schnittstelle, die einen Überhang von vier ungebundenen Nukleotiden hinterlässt. Die minimale Sequenz, die für den Schnitt nötig ist, ist 31 Basenpaare lang (-10 bis +21), die Region, die vor DNase-I-Verdau geschützt ist, geht etwas darüber hinaus, Abbildung 5.5 (Gimble und Stephens, 1995; Gimble und Wang, 1996). Die endonukleolytische Spaltung, der Bruch der Bindung zwischen Phosphat und 3'-OH des vorherigen Nukleotids, geschieht im oberen Strang zwischen den Nukleotiden +2 und +3 und im unteren Strang zwischen -2 und -3 (Christ *et al.*, 1999; Gimble und Stephens, 1995; He *et al.*, 1998; Wende *et al.*, 1996). Die Bezeichnungen „oberer“ und „unterer“ Strang beziehen sich auf die in der Abbildung 5.5 gezeigten Anordnung, mitunter sind in der Literatur auch die Begriffe „links“ und „rechts“ auf diese Anordnung bezogen.

**Abbildung 5.5: Erkennungssequenz von PI-SceI**

Die minimale Bindungssequenz ist dunkelgrau hinterlegt, die vor DNase I geschützte Region hellgrau. Die minimale Sequenz, die für die endonukleolytische Spaltung nötig ist, ist durch den schwarzen Rahmen gekennzeichnet, die schwarzen Linien zeigen die Schnittstellen.

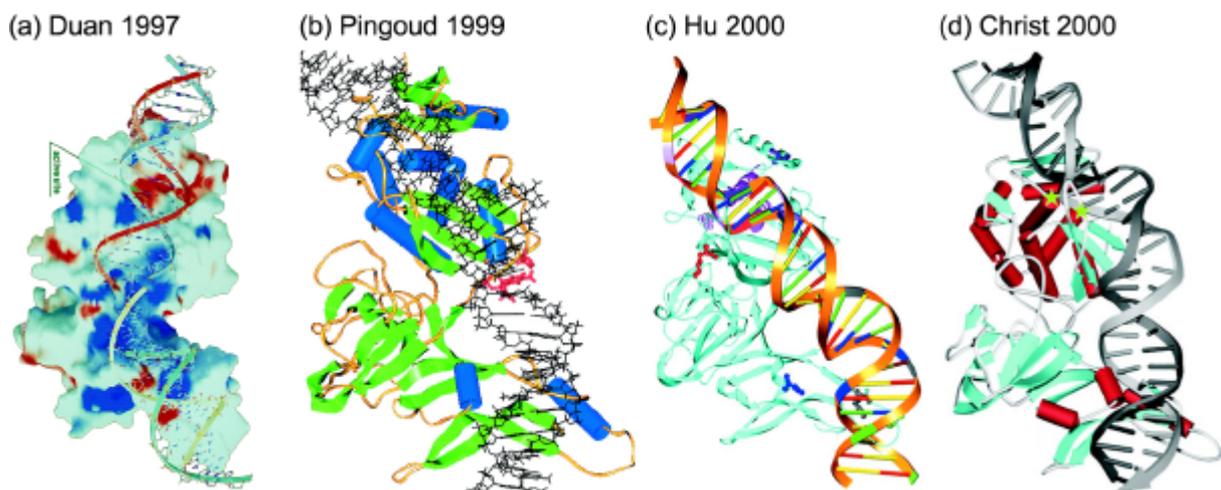
Mit 31 Basenpaaren ist die spezifische Erkennungssequenz eine der längsten bekannten und zudem asymmetrisch. Sie unterscheidet sich somit deutlich von denen vieler Restriktionsenzyme. Diese erkennen meist 4 bis 8 Basenpaare, deren Symmetrie sich auch im bindenden Protein widerspiegelt. Der Großteil der Bindungsenergie kommt von Domäne I, der Protein-Splicing-Domäne, während Domäne II das endonukleolytische Zentrum enthält, aber nur locker an die Erkennungssequenz bindet (Gimble und Wang, 1996; Grindl *et al.*, 1998; He *et al.*, 1998; Wende *et al.*, 1996). Um das gesamte Protein zu bedecken, muss die DNA eine Biegung eingehen und tut dies in zwei Schritten (Abbildung 5.6).

**Abbildung 5.6: Biegung der PI-SceI-Erkennungssequenz**

Im initialen Komplex bindet die Domäne I zunächst die minimale Bindungsregion (Basenpaare +5 bis +21) mit Gleichgewichtskonstanten im nM-Bereich (Grindl *et al.*, 1998). Der Name „unterer Komplex“ ist aus den Gelshift-Experimenten hergeleitet und bezieht sich auf die untere von zwei beobachteten Banden, wenn PI-SceI – DNA Komplexe in einem nativen Gel aufgetrennt werden. Die DNA hat hier einen Biegungswinkel von 40 bis 45°, der möglicherweise durch Verstärkung einer intrinsischen Biegung zustande kommt. Im „oberen Komplex“ wird der Rest

der Erkennungssequenz (bis etwa Basenpaar -10) von Domäne II gebunden, was eine weitere Biegung der DNA auf insgesamt dann 60 bis 75° bewirkt (Gimble und Wang, 1996; Wende *et al.*, 1996). Auf diese Weise werden die Schnittstellen in die Nähe des aktiven Zentrums der Domäne II gebracht und die endonukleolytische Spaltung kann stattfinden. Da Domäne I den einen (den „rechten“) Teil der Erkennungssequenz noch länger festhält, ist die Umsatzrate von PI-SceI gering (Wende *et al.*, 2000).

Strukturelle Daten eines PI-SceI-DNA-Komplexes gibt es bisher nicht, weshalb auf biochemische Daten zurückgegriffen werden muss, um die Feinheiten der DNA-Bindung zu verstehen. Diese Daten sind zahlreich vorhanden, in Abbildung 5.8 zusammengefasst und in deren Legende erläutert. Sie geben einen Überblick, welche Aminosäuren und Basenpaare an der Bindung beteiligt sind und weisen im Falle der Photo-Quervernetzungen direkte Bindungspartner nach. Auf dieser Basis wurden einige (und mittlerweile auch generell konsistente) Modelle entwickelt (Christ *et al.*, 2000; Duan *et al.*, 1997; Hu *et al.*, 2000; Pingoud *et al.*, 1999; Wende *et al.*, 2000), die in Abbildung 5.7 gegenübergestellt sind. Ihnen allen ist gemeinsam, dass sie PI-SceI als rigiden Körper behandeln und somit mögliche Bewegungen der Domänen oder Subdomänen gegeneinander nicht berücksichtigen. Lediglich von der Schleife 55 bis 66 wird angenommen, dass sie sich in Richtung der DNA bewegen könnte (Wende *et al.*, 2000). Im Rahmen dieser Arbeit wird durch ein geometriebasiertes Docking die Möglichkeit der Bewegung einer Subdomäne von Domäne I erläutert und begründet.



**Abbildung 5.7: Bindungsmodelle**

Die Abbildungen sind aus den Original-Publikationen entnommen. (a) (Duan *et al.*, 1997), (b) (Pingoud *et al.*, 1999), (c) (Hu *et al.*, 2000), (d) (Christ *et al.*, 2000).

### 5.1.6. Warum Struktur Nummer Fünf?

Die Struktur von PI-SceI wurde erstmals von Duan *et al.* (1997) mit einer maximalen Auflösung von 2.4 Å gelöst. Drei Jahre später wurden Kristalle beobachtet, die aus einer Protein-DNA-Mischung hervorgingen, jedoch nur PI-SceI alleine (wenn auch in einer anderen Raumgruppe) enthielten (Hu *et al.*, 2000). Weitere zwei Strukturen enthalten außer dem Intein auch noch Extein-Reste und hatten die strukturelle Aufklärung des Protein-Splicing-Mechanismus zum Ziel, eine davon erst kürzlich veröffentlicht (Mizutani *et al.*, 2002; Poland *et al.*, 2000). Allen Strukturen ist gemeinsam, dass in keinem Fall die vollständige Proteinsequenz der Elektronendichte zugeordnet werden konnte. Die maximale Auflösung dieser Strukturen betrug 2.0 Å. Zudem ist es offenbar noch keinem gelungen, PI-SceI-DNA-Cokristalle zu erzeugen.

Zusammen mit den Kooperationspartnern Wolfgang Wende und Alfred Pingoud, Universität Gießen, sollten beide Defizite durch die Konzentration auf den für die DNA-Bindung wichtigen Teil, die Domäne I, überkommen werden. Unglücklicherweise schlug dies fehl. Die hier präsentierte Struktur deckt ebenfalls nicht die vollständige Sequenz ab (die Schleife 55 bis 66 ist nicht sichtbar) und Komplex-Kristalle wurden auch nicht beobachtet. Aber da die Struktur der Domäne I durch die hohe Auflösung (maximal 1.35 Å) viele Details erkennen lässt, insbesondere auch in den Bereichen, die in anderen Strukturen schlecht definiert waren, trägt sie dennoch dazu bei, die biologische Aktivität von PI-SceI besser zu verstehen.

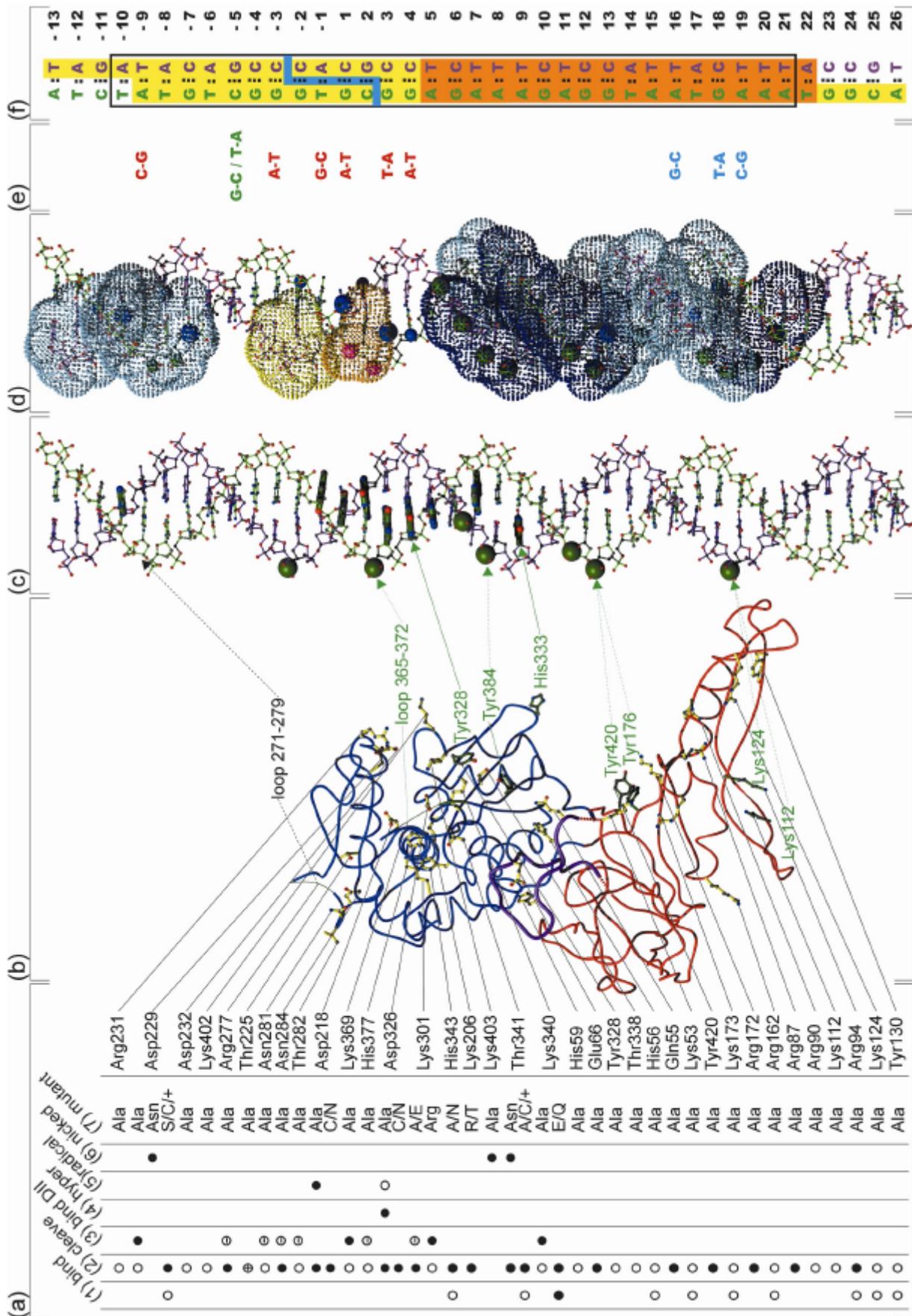


Abbildung 5.8: DNA-Bindung von PI-SceI, Zusammenfassung der biochemischen Daten

zu Abbildung 5.8: DNA-Bindung von PI-SceI – Biochemische Daten

**(a)** Zusammenfassung von Bindungsaktivität, endonukleolytischer Spaltungsaktivität, Hydroxylradikal-Abhängigkeit und der Präferenz von Einzelstrangbrüchen verschiedener Mutanten von PI-SceI (Christ *et al.*, 1999; Gimble und Wang, 1996; He *et al.*, 1998; Hu *et al.*, 1999; Schöttler *et al.*, 2000; Wende *et al.*, 2000).

- (1) ○ Reduzierte Bindungsaktivität.
- (2) ○ Reduzierte, ● verloren gegangene, ⊕ gesteigerte endonukleolytische Aktivität.
- (3) ⊖ Keine, ● festere Bindung der DNA an Domäne II.
- (4) ● Keine Hypersensibilität der DNA gegen Hydroxyl Radikale im Protein-DNA-Komplex.
- (5) ○ Reduzierte, ● keine DNA-Strangbruch-Aktivität von Hydroxyl-Radikalen im Protein-DNA-Komplex.
- (6) ● Präferenz für die endonukleolytische Spaltung von DNA mit Einzelstrangbruch.
- (7) Untersuchte Mutante(n). Für Asp229 wurden außerdem die Mutanten D229E, D229Q und D229T untersucht, für Thr341 außerdem T341D und T341S (jeweils gekennzeichnet durch ein plus).

Weiterhin untersucht, jedoch ohne oder nur mit geringem Effekt wurden die folgenden Mutanten: R57A, K60A, D62A, R65A, R91A, K97A, E121A, R147A, K148A, K152A, Y154A, E160A, Y176A, L220A, S221A, D222A, R223A, S230A, E237A, K253A, D254A, R255A, K256, E257A, Q259A, K262A, T263A, N265A, K297A, D298A, P304A, E316A, T330A, D331A, E332A, H333A, K336A, E366A, D371A, N373A, T375A, K376A, K378A, Y384A, R414A, R417A.

**(b)** Schematische Darstellung des PI-SceI-GV-Konstruktes (siehe Kapitel 5.3.8, Seite 103) mit den Seitenketten in der Kugelstab-Darstellung, deren Mutanten unter (a) aufgelistet sind (gelbe Bindungen) oder mit denen ein Photo-Quervernetzung zur DNA beobachtet wurde (grüne Bindungen und Aminosäure Kennzeichnungen, siehe (c)). Die fehlenden Schleifen in Domäne II (Reste 271 bis 279 und 369 bis 374) sind als dünne Linien dargestellt.

**(c)** Zusammenfassung der Photo-Quervernetzungsdaten (Christ *et al.*, 2000; Hu *et al.*, 2000; Pingoud *et al.*, 1999; Posey und Gimble, 2002). Der obere Strang ist in grün dargestellt, der untere in violett (ebenfalls in (d) und (f)). Grüne Kugeln repräsentieren Phosphate, die eine Photo-Quervernetzung via p-Azidophenacylphosphorothioate eingehen. Ist die betreffende Aminosäure bekannt, ist sie durch einen gepunkteten Pfeil mit dem Phosphat verbunden. Die Nukleobasen mit den dicken Bindungen gehen nach Austausch gegen 5-Iodo-Pyrimidin eine Photo-Quervernetzung ein, bekannte Verbindungen sind mit durchgezogenen Pfeilen gekennzeichnet. Der gestrichelte Pfeil repräsentiert einen Strangbruch ausgelöst durch FeEDTA.

**(d)** Zusammenfassung der Methylierungs- und Ethylierungs-Störungs-Assays sowie der Protektion vor Hydroxyl-Radikalen (Christ *et al.*, 1999; Gimble und Stephens, 1995; Gimble und Wang, 1996). Kugeln (groß: starke Störung; mittlere Größe: schwächere Störung) repräsentieren Atome, wo eine Methylgruppe (blau; Methylierung von Guanin-N7 oder Adenin-N3) oder eine Ethylgruppe (grün; Phosphate) die DNA-Bindung stört. Gepunktete Oberflächen (hellblau: leichte, dunkelblau: starke Protektion; gelb: leichte, orange: starke Hypersensitivität) stellen die Ergebnisse der Assays mit Hydroxyl-Radikalen dar.

**(e)** Mutationen in der PI-SceI-Erkennungssequenz (Gimble und Wang, 1996). In rot Mutanten, die die endonukleolytische Spaltung verhindern (loss-of-cleavage), in blau diejenigen, die die Bindung der DNA an das Protein verhindern (loss-of-binding). Mutanten in grün dargestellt erhöhen die endonukleolytische Aktivität.

**(f)** Erkennungssequenz von PI-SceI (analog Abbildung 5.5, Seite 71). Die minimale Bindungssequenz ist orange, die Sequenz, die für die endonukleolytische Spaltung nötig ist in gelb hinterlegt, die Schnittstelle ist in blau gekennzeichnet. Der „obere“ Strang ist links (grün), der „untere“ rechts (violett).

## 5.2. Material und Methoden

### 5.2.1. Genexpression und Proteinreinigung

Das Plasmid pHisPI-SceI-DI, welches die cDNA der *S. cerevisiae* Homing Endonuklease PI-SceI Domäne I enthält (codierend für die Aminosäuren 1 bis 182 und 410 bis 453), wurde von den Kooperationspartnern Wolfgang Wende und Alfred Pingoud, Universität Gießen zur Verfügung gestellt (Grindl *et al.*, 1998). Im Vergleich zur SWISSPROT-Sequenz (PID: P17255; (Bairoch und Apweiler, 2000)) fehlt Asn454 (zur Vermeidung der Protein-Splicing-Reaktion), Domäne II ist durch Gly183 ersetzt, und die natürlich vorkommenden Variationen Arg44Ser, Val67Met und Ile132Val sind (wie auch im DH1-1A Wildtyp-Hefestamm (Gimble, 2001)) vorhanden. Es enthält zudem eine Sequenz, die für 6 Histidine und einen Linker codieren und nach der Expression im Protein einen N-terminalen His<sub>6</sub>-tag von 10 Aminosäuren Länge ergeben. Das Plasmid wurde mit der CaCl<sub>2</sub>-Methode in *E. coli* BL21(DE3) Zellen (Stratagen) transfiziert und die Genexpression in LB Medium mit 0.5 mM IPTG bei 22°C über Nacht induziert (Sambrook *et al.*, 1989). Die bei 9500 rpm für 15 min zentrifugierte (Beckman JA10; 4°C) Zellmasse entsprach etwa 7 g Feuchtmasse pro Liter Kultur (ca. 5000 OD600 pro Liter) und wurde bei -70°C zwischengelagert. Die Zellwände der im Aufschlusspuffer (50 mM HEPES pH 7.5, 0.5 M KCl, 15 mM Imidazol, 5 % (v/v) Glycerin, 0.4 mM Pefabloc, 5 mM β-Mercaptoethanol) wieder aufgetauten Zellen wurden per Ultraschall zerstört (5 x 25 Pulse, Branson Sonifier 450; auf Eis). Zentrifugation bei 19500 rpm für eine Stunde (Beckmann JA20; 4°C) ergab eine klare, lösliche Fraktion (Überstand). PI-SceI-DI darin wurde an Ni-NTA (Qiagen; 1 ml pro Liter Kultur) bei 4°C während einer Stunde gebunden (batch modus), anschließend in eine Glassäule gefüllt und gewaschen mit 12 Säulenvolumen Waschpuffer (50 mM HEPES pH 7.5, 200 mM KCl, 15 mM Imidazol, 5 % Glycerin). Die Elution des Proteins erfolgte durch Erhöhung der Imidazol-Konzentration auf 250 mM. Durch Dialyse (Serva Spectra/Por, 5000 MWCO) wurde PI-SceI-DI in den Lagerungspuffer (50 mM NaAcetat, 100 mM NH<sub>4</sub>Acetat, pH 4.6) überführt, in dem es mehrere Wochen stabil blieb. Die Ausbeute betrug etwa 8 mg Protein pro Liter Zellkultur. Die Abtrennung von hoch- und niedermolekularen Verunreinigungen erfolgte über Gel-

Chromatographie (Pharmacia Superdex75 16/60), wobei der Laufpuffer mit dem Lagerungspuffer identisch war und zusätzlich 2 mM  $\beta$ -Mercaptoethanol enthielt. Einheitlich eluiertes Protein wurde in einer Millipore Ultrafree-4 (MWCO 5000) Einheit bei 4000 g konzentriert auf 13.3 mg/ml.

### 5.2.2. DNA-Bindung

DNA-Oligonukleotide, hergestellt und gefriergetrocknet geliefert von der Firma BioTeZ (Berlin-Buch), wurden in 10 mM Tris-Cl pH 7.5 gelöst und die beiden komplementären Stränge im molaren Verhältnis von etwa 1 : 1 hybridisiert. Hierfür wurde die Mischung der Oligonukleotide auf die theoretische Schmelztemperatur plus 5 °C erwärmt und durch Abschalten des Heizblockes langsam abgekühlt (ca. 2 bis 3 Stunden). Die Doppelstränge wurden von ungebundenen Einzelsträngen durch Ionenaustauschchromatographie (Pharmacia MonoQ 5/5) getrennt. In einem Puffer von 50 mM Tris-Cl pH 8.5 werden Einzelstränge in einem linearen Gradienten von 300 bis 700 mM NaCl über 20 Säulenvolumen bei etwa 450 mM NaCl eluiert, Doppelstränge bei etwa 600 mM NaCl. Konzentration und Pufferaustausch zu 0.1 M Tris-Cl pH 7.5 fand in Millipore Ultrafree-0.5 Konzentrationseinheiten entsprechend den Herstellerangaben statt. Die Magnesium-Konzentration der gereinigten Doppelstrang-Lösungen wurde mit MgCl<sub>2</sub> auf 200 mM eingestellt. Eine Lösung von konzentriertem PI-SceI-DI in Lagerungspuffer wird im gewünschten molaren Verhältnis hinzu gegeben und bei Raumtemperatur 30 Minuten stehen gelassen. Für eine Gelretardierungs-Experiment (Gelshift) wird eine (entsprechend verdünnte) Lösung mit final 15 % Glycerin und 0.1 % (w/v) Bromphenolblau versetzt. Ein 6 % natives kontinuierliches Polyacrylamidgel (Sambrook *et al.*, 1989) wird nach 30 Minuten Vorlauf bei 80 Volt mit dem Komplex beschickt (ca. 5 bis 10 pmol Komplex pro Probe) und anschließend die Elektrophorese weitere 50 bis 60 min bei 80 Volt fortgeführt. Die Silberfärbung (Sambrook *et al.*, 1989) macht bei rechtzeitigem Abbruch sowohl die DNA als auch das Protein sichtbar. Absorptionsspektren wurden mit dem Photometer Beckman DU7400 aufgenommen.

### 5.2.3. Kristallisation

PI-SceI-DI kristallisiert in einer Reihe von PEG-Bedingungen des Hampton-Screens (Jancarik und Kim, 1991) in der hängende-Tropfen Methode durch Konzentrationsausgleich über die Dampfphase (hanging-drop vapour-diffusion, siehe Kapitel 4.2.3, Seite 23). Eine große Anzahl Kristalle von PI-SceI-DI bildet sich teilweise innerhalb von Minuten, etwas weniger, dafür aber größere Kristalle spätestens nach 2 bis 3 Tagen. Sie erreichen meist nur eine Größe von ca. 400 x 100 x 100  $\mu\text{m}$  und benötigen weder eine Verfeinerung der Kristallisationsbedingungen noch eine Kryoprotektion. Der für die Strukturbestimmung verwandte Kristall entspringt einem Versuch, den Komplex von PI-SceI-DI mit dem 18mer der minimalen Erkennungssequenz (bp +5 bis +22) zu kristallisieren. Zwei  $\mu\text{l}$  der Komplexlösung enthalten PI-SceI-DI und das 18mer im molaren Verhältnis 1 zu 1.05 (0.35 mM Komplex) in einem Puffer aus final 54 mM  $\text{NH}_4\text{Acetat}$ , 27 mM  $\text{NaAcetat}$  pH 4.7, 38 mM  $\text{Tris-Cl}$  pH 7.5 und 188 mM  $\text{MgCl}_2$  und wurde gemischt mit einem  $\mu\text{l}$  Reservoir-Lösung, die 30 % PEG4000, 0.1 M  $\text{NaCitrat}$  pH 5.6 und 0.2 M  $\text{NH}_4\text{Acetat}$  enthält (Hampton Screen Nr. I, Lösung # 9).

### 5.2.4. Röntgendiffraktion, Datenreduktion und Strukturbestimmung

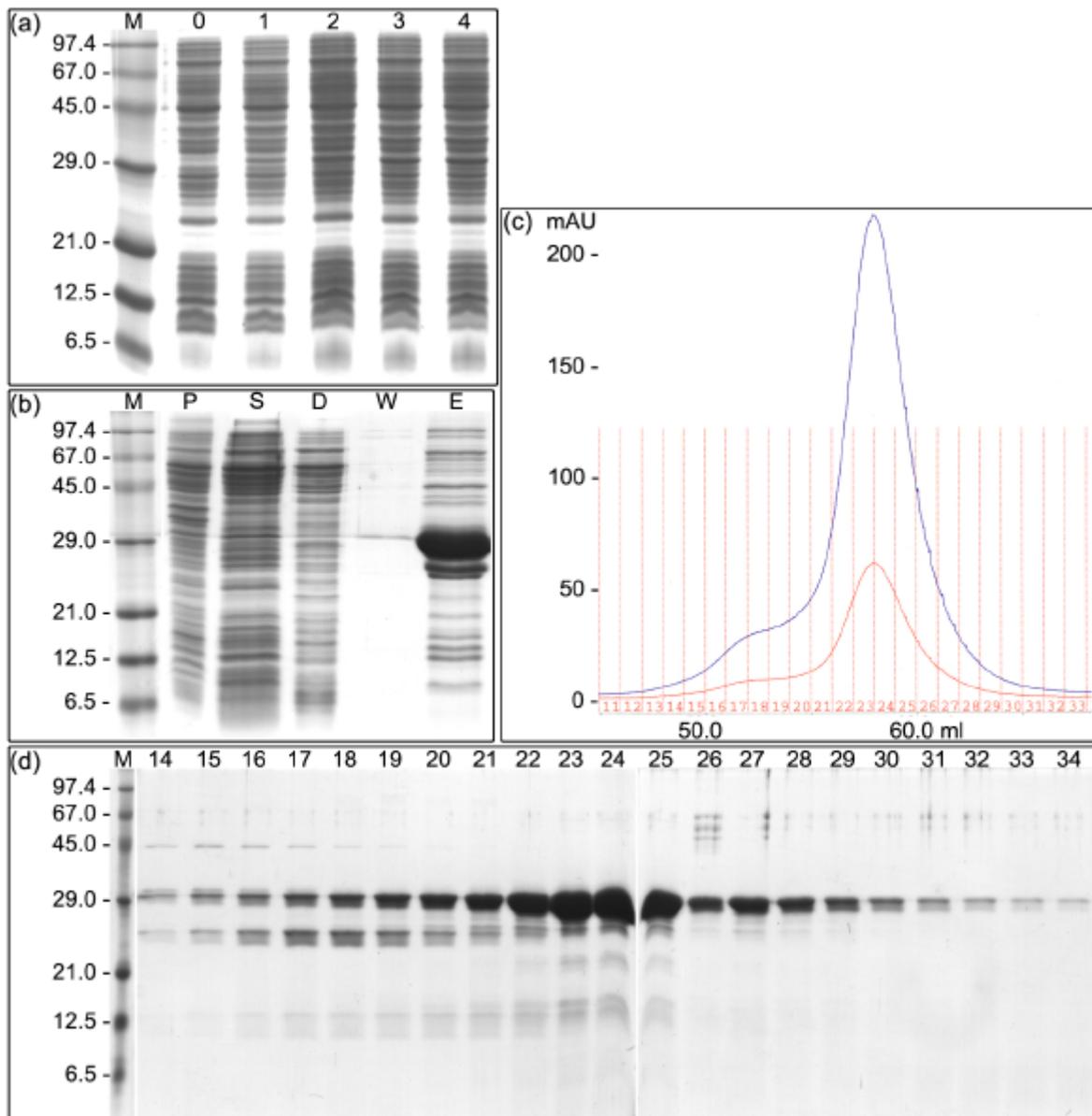
Komplette Datensätze von diesem Kristall wurden sowohl an der hauseigenen Drehanode (Rigaku, Detektor: MAR180 ImagePlate) mit einer maximalen Auflösung von 2.02  $\text{\AA}$  als auch an der Messstation BW7B EMBL/DESY (Detektor: MAR345 ImagePlate) Hamburg mit einer maximalen Auflösung von 1.35  $\text{\AA}$  aufgenommen. Die Daten wurden jeweils mit den Programmen DENZO und SCALEPACK (Otwinowski und Minor, 1997) reduziert. Die Strukturbestimmung gelang mit der Methode des Molekularen Ersatzes und dem Programm AMoRe (Navaza, 1994). Als Modell wurde die Struktur von Duan et al. (1997), PDB-Code 1vde, benutzt. Das Modell wurde gegen den ersten (2.02  $\text{\AA}$ -) Datensatz mit dem Programm CNS (Brünger *et al.*, 1998) vorverfeinert, nachdem einige Reste entfernt waren, die sich hier offenbar mit symmetrieverwandten Molekülen überlappten (Reste 55 bis 69, 129 bis 138 und 410 bis 418). Das resultierende Modell wurde für den Molekularen Ersatz mit dem zweiten (1.35  $\text{\AA}$ -) Datensatz benutzt. Aufgrund der hohen Auflösung konnte die automatische Funktion des

Programms ARP (Lamzin und Wilson, 1993) genutzt werden. ARP unternahm sowohl die Interpretation der Dichte, den Modellbau, suchte automatisch nach den richtigen Seitenketten und später nach Wassermolekülen. Der manuelle Modellbau und die Verfeinerung wurden mit den Programmen ONO (Jones *et al.*, 1991) bzw. REFMAC5 (Collaborative Computational Project, 1994) durchgeführt. Eine individuelle anisotrope *B*-Faktor-Verfeinerung wurde durchgeführt und 303 Wassermoleküle konnten in der Elektronendichte platziert werden.

## 5.3. Ergebnisse und Diskussion

### 5.3.1. Expression, Proteinreinigung und Kristallisation

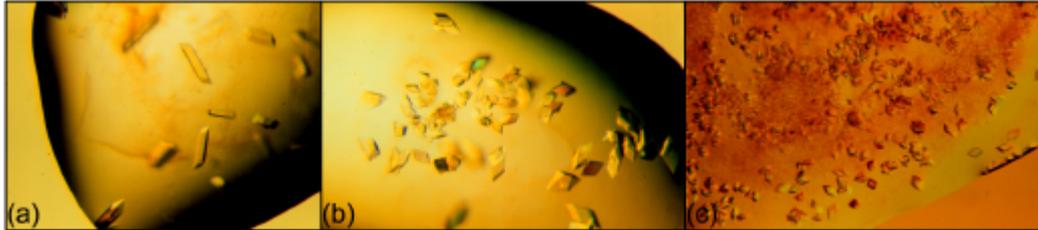
PI-SceI-DI kann in *E. coli*-Zellen rekombinant in ausreichenden Mengen hergestellt und in zwei Schritten zu zufrieden stellender Qualität gereinigt werden, Abbildung 5.9. Obwohl im SDS-Gel Verunreinigungen noch deutlich sichtbar sind, kristallisiert das Protein in dieser Reinheit in einer Vielzahl von PEG-Bedingungen mit tadelloser Morphologie. Die geringe Kristallgröße (ca. 400 x 100 x 100 µm) hat hier jedoch wenig Bedeutung, denn die Diffraktion war in vielen Fällen sehr regelmäßig und bis zu hohen Auflösungen möglich. Schnell wachsende Kristalle (teilweise innerhalb von Minuten) waren meistens sehr zahlreich und noch kleiner, Abbildung 5.10. Offenbar neigt PI-SceI-DI stark zur Selbstanlagerung, was in Lösung und Kristallansätzen bisweilen auch als Präzipitat sichtbar wurde. Beim Aufkonzentrieren wurde teilweise ein gelbliches Öl beobachtet, welches nicht mehr für Kristallisationsversuche geeignet war. Im Kristall wurde dann auch tatsächlich eine enge Kristallpackung entdeckt (siehe Kapitel 5.3.4, Seite 87). Protein-Protein-Kontakte sind hier offensichtlich bevorzugt im Vergleich zu Protein-DNA-Wechselwirkungen. Dafür spricht, dass aus einer Protein-DNA-Lösung das Protein alleine kristallisiert, obwohl immerhin 50 % „Verunreinigung“ (die DNA) vorhanden ist. Dies wird gemeinhin als großes Hindernis für Kristallwachstum betrachtet.



**Abbildung 5.9: Expression und Proteinreinigung von PI-SceI Domäne I**

(a) 16 %ige diskontinuierliche SDS-Gelelektrophorese nach Färbung mit Coomassie Blau von 0.15 OD<sub>600</sub> Zellpellet pro Spur vor (Kennzeichnung 0) und jeweils 1, 2, 3 und 4 Stunden nach Induktion mit 0.5 mM IPTG (Spur 1, 2, 3 bzw. 4). M: Größenstandard mit den links bezifferten Molekulargewichten in kDa. PI-SceI-DI erscheint als Bande knapp über dem 29 kDa-Marker und bleibt schon nach 2 Stunden konstant. (b) SDS-Gel der Proben während des Zellaufschlusses und der Affinitätschromatographie per Ni-NTA. M: Größenstandard, P: Pellet nach Zellaufschluss, S: lösliche Fraktion nach Zellaufschluss, D: nicht an Ni-NTA gebundene Proteine, W: beim Waschen eluierte Proteine, E: mit 250 mM Imidazol eluierte Proteine. Es ist deutlich zu sehen, dass sich PI-SceI-DI sowohl im Pellet als auch in der löslichen Fraktion nach Zellaufschluss befindet, also nur teilweise in löslicher Form in den Zellen synthetisiert wird. Der überwiegende Teil wird über den His-tag an Ni-NTA gebunden und auch nicht herausgewaschen. Die Elution mit hoher Konzentration Imidazol ist vollständig und die Lösung hochkonzentriert. Zahlreiche Fremdproteine (oder auch Abbauprodukte) sind noch enthalten, wenn auch diese Spur sehr überladen ist an Protein. (c) Chromatogramm der Gelfiltration (Größenausschlusschromatographie mit Superdex75). Die blaue Linie entspricht der Absorption bei 278 nm, die rote der bei 251 nm. Die roten Striche mitsamt Nummern entsprechen den aufgefangenen Fraktionen, wovon ein Anteil auf dem (d) SDS-Gel aufgetragen wurde (Spuren entsprechen den Fraktionsnummern). Durch diesen

Reinigungsschritt konnten zwei, etwa 25 und 45 kDa große Verunreinigungen weitgehend abgetrennt werden, kleinere Mengen blieben jedoch enthalten. Für die Kristallisations- und DNA-Bindungsexperimente wurden die Fraktionen 22 bis 38 verwandt.

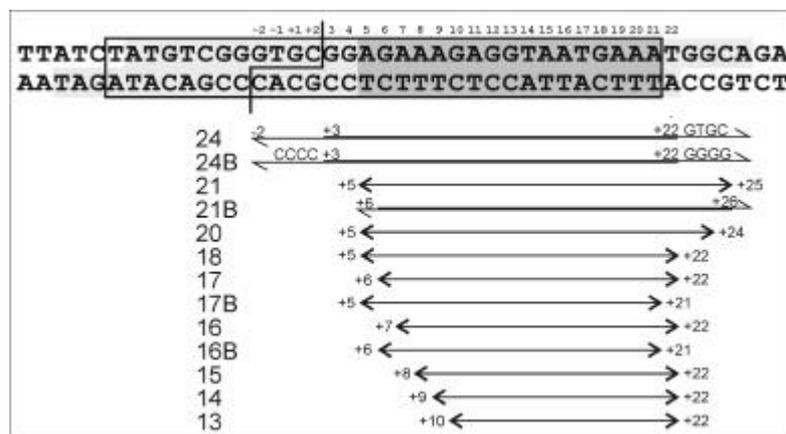


**Abbildung 5.10: Kristalle von PI-SceI-DI**

Fotos verschiedener Kristallansätze. Die größten Kristalle (zentral in a) sind 400 x 100 x 100 µm groß. Auch Gemische von sehr vielen kleinen und größeren Kristallen wurden beobachtet. PI-SceI-DI kristallisiert in einer quaderförmigen Art mit tadelloser Morphologie.

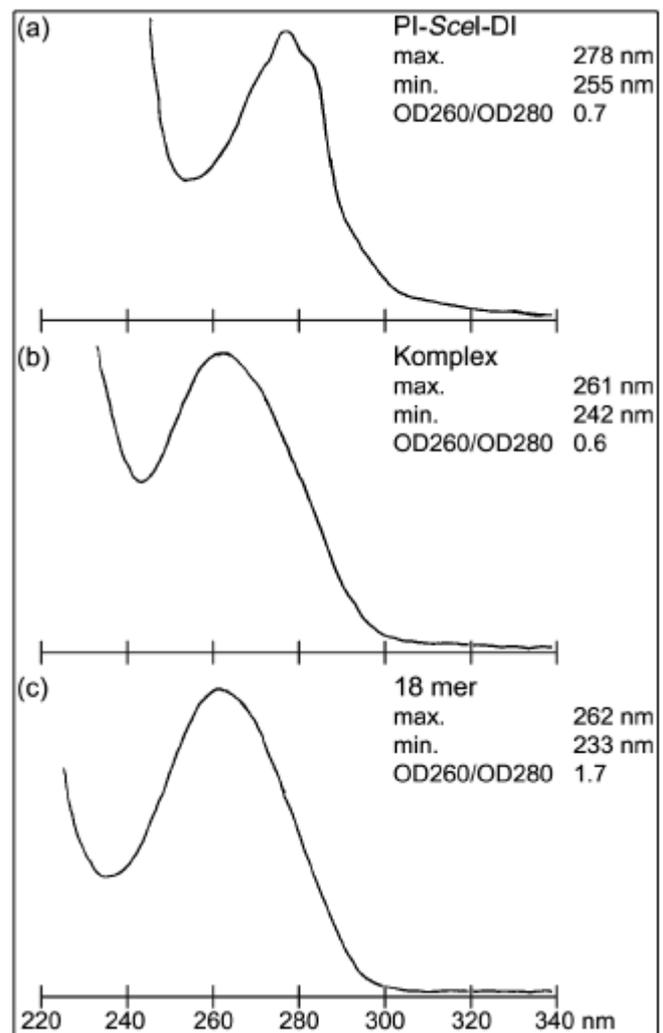
### 5.3.2. DNA-Bindungsversuche

PI-SceI bindet seine 31 bp lange Erkennungssequenz mit Bindungskonstanten im nM Bereich (He *et al.*, 1998; Wende *et al.*, 1996) wofür im Wesentlichen Domäne I verantwortlich ist (Grindl *et al.*, 1998). DNA-Bindungsexperimente wurden mit 13 verschiedenen Ausschnitten der Erkennungssequenz durchgeführt, die in Abbildung 5.11 skizziert sind. Sie decken im Wesentlichen die minimale Bindungsregion ab (Basenpaare +5 bis +21) und wurden teilweise auch im Hinblick auf eine möglicherweise „unendliche“ Doppelhelix im Kristall mit ungepaarten Nukleotiden verschiedener Länge konstruiert.



**Abbildung 5.11: DNA-Fragmente für Bindungsexperimente**

Die verschiedenen Komponenten der Mischung unterscheiden sich durch ihr Absorptionsspektrum (Abbildung 5.12). Während das Protein ein Absorptionsmaximum bei 278 nm und ein Minimum bei 255 nm besitzt ( $OD_{260}/OD_{280} = 0.7$ ), liegen diese für das 18mer bei 262 bzw. 233 nm ( $OD_{260}/OD_{280} = 1.6$ ), typische Werte für Nukleinsäuren. Obwohl DNA deutlich stärker absorbiert als das Protein (Größenordnungen in der Abbildung sind nicht proportional), sieht das Spektrum der Komplexlösung nicht genauso aus wie das der DNA. Das Absorptionsminimum ist mit 242 nm deutlich in Richtung des Minimums des Proteins verschoben und das Verhältnis  $OD_{260}/OD_{280}$ , wichtiger Parameter für die Abschätzung der Reinheit für Nukleinsäuren, ist fast identisch mit dem des reinen Proteins. Ob sich eine Komplex gebildet hat oder ob sich die Einzelspektren einfach addieren, kann hier jedoch nicht unterschieden werden.



**Abbildung 5.12:**  
**Absorptionsspektren**

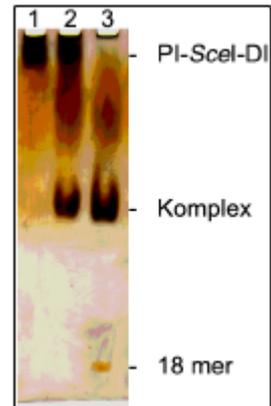
Absorptionsspektren von (a) PI-SceI-DI, (b) einer Mischung von PI-SceI-DI mit dem 18mer und (c) dem 18mer alleine. Das Absorptionsmaximum des Komplexes entspricht weitgehend dem der DNA, das Minimum liegt zwischen denen der Einzelkomponenten und das Verhältnis  $OD_{260}/280$  ähnelt dem des reinen Proteins.

Der Nachweis der Komplexbildung erfolgt durch Gelretardierungsexperimente (Abbildung 5.13) und war im Fall von PI-SceI-DI mit allen beschriebenen DNA-

Fragmenten erfolgreich. Dennoch konnten in keinem Fall Cokristalle beobachtet werden. Vielmehr wuchsen vielfach lediglich Kristalle des Proteins alleine (siehe auch Kapitel 5.3.1, Seite 79).

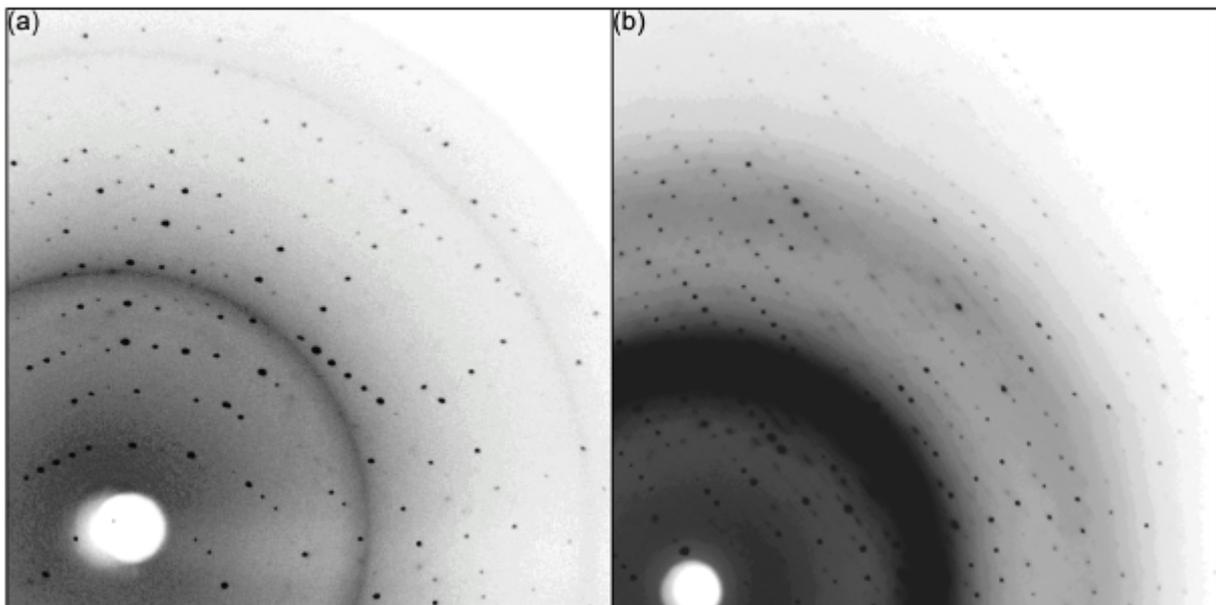
**Abbildung 5.13: Gelshift-Experiment**

6% kontinuierliche, native Polyacrylamid-Gelelektrophorese. Spur 1:nur Protein (5 pmol PI-SceI-DI), Spur 2: DNA und Protein im molaren Verhältnis 1 zu 2 (5 pmol PI-SceI-DI), Spur 3: DNA und Protein im molaren Verhältnis 3 zu 1 (5 pmol PI-SceI-DI). Gel nach Silberfärbung.



**5.3.3. Diffraktion und Strukturbestimmung**

Datensätze von demselben Kristall wurden sowohl an der hausinternen Anlage als auch am DESY, Hamburg, aufgenommen. Die Diffraktionsmuster (Abbildung 5.14) zeichneten sich durch hohe Auflösung und große Schärfe der Reflexe aus, was sich in der Datensammlungsstatistik (Tabelle 5-1) widerspiegelt.



**Abbildung 5.14: Diffraktionsmuster des PI-SceI-DI-Kristalls**

(a) Rigaku, MAR180 Detektor.(b) DESY, Hamburg, Strahlführung BW7B, MAR345 Detektor.

Datensammlung	Rigaku	DESY BW7B
Auflösung [Å]	20 – 2.02 (20.09 – 2.02)	20 – 1.35 (1.37 – 1.35)
Anzahl der beobachteten Reflexe	150910	359618
Anzahl der unabhängigen Reflexe	16032 (1533)	51602 (1141)
Vollständigkeit [%]	95.6 (92.2)	94.1 (64.1)
durchschnittliches $I/sI$	38.4 (13.3)	34.5 (2.5)
$R_{sym}^*$ [%]	3.5 (10.6)	2.9 (16.4)
$R_{r.i.m.}^*$ [%]	4.0 (12.2)	4.1 (20.8)
$R_{p.i.m.}^*$ [%]	1.8 (5.9)	2.7 (13.4)
Raumgruppe	C2 (#5, monoklin)	
Zellparameter [Å]	$a = 102.9$ $b = 47.6$ $c = 60.2$ $b = 121.4$	
Anzahl der Monomere pro asymmetrischer Einheit	1	

**Tabelle 5-1: Datensammlungsstatistik des PI-SceI-DI Kristalls**

Die Werte in Klammern beziehen sich auf die äußerste Auflösungsschale.

$$* R_{sym} = \frac{\sum_{hkl} \sum_i |I_i - \langle I \rangle|}{\sum \langle I \rangle}$$

$$R_{r.i.m.} = \frac{\sum_{hkl} (N/(N-1))^{1/2} \sum_i |I_i - \langle I \rangle|}{\sum \langle I \rangle}$$

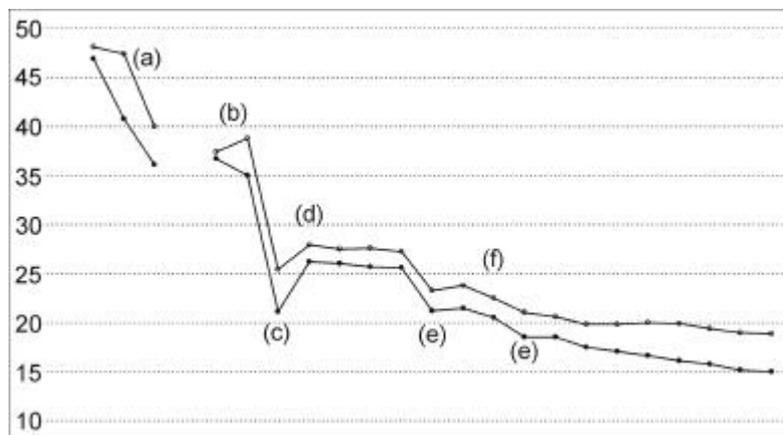
$R_{p.i.m.} = \frac{\sum_{hkl} (1/(N-1))^{1/2} \sum_i |I_i - \langle I \rangle|}{\sum \langle I \rangle}$  mit  $I_i$  der Intensität der Beobachtung des Reflexes  $hkl$ ,  $\langle I \rangle$  der durchschnittlichen Intensität dieses Reflexes und  $N$  der Redundanz (Weiss, 2001).

Die Strukturbestimmung gelang mittels Molekularem Ersatz. Hierfür wurden die Koordinaten der Struktur von Duan *et al.* (1997), PDB-Code 1vde, als Modell benutzt. Auch die anderen Strukturen waren geeignet, ergaben jedoch weniger deutliche Lösungen. Der Modellbau gelang weitgehend automatisch durch Verwendung des Programms ARP (Lamzin und Wilson, 1993). Von den 237 Aminosäuren des Proteins (einschließlich des His-tag) konnte ARP bereits 202 korrekt platzieren (durchschnittlicher Zuverlässigkeits-Index von 99 %). Weitere 15 konnten manuell hinzugefügt werden. Lediglich die Reste Gln55 bis Glu66 sowie Teile des His-tag (Reste -10 bis -3) konnten aufgrund positioneller Dislokation keiner Elektronendichte zugeordnet werden. Die Verfeinerung mit individuellen anisotropen  $B$ -Faktoren erreichte  $R_{cryst}$ - bzw.  $R_{free}$ -Werte von 15.0 % bzw. 18.9 %. Die Verfeinerungsstatistik ist in Tabelle 5-2 zusammengefasst, der Verlauf der  $R$ -Faktoren während der Verfeinerung ist in Abbildung 5.15 zu sehen.

Strukturbestimmungsmethode	MR (1vde)
Anzahl der Proteinatome (ohne Wasserstoff) pro AU	1758
$R_{\text{cryst}}^a$ [%]	15.0
$R_{\text{free}}^b$ [%]	18.9
r.m.s.d. Bindungslängen [Å]	0.018
r.m.s.d. Bindungswinkel [°]	1.83
durchschnittlicher $B$ -Faktor <sup>c</sup> für alle Proteinatome [Å <sup>2</sup> ]	23.3
durchschnittliche r.m.s.d. für alle Hauptketten $B$ -Faktoren <sup>c</sup> [Å]	0.83
durchschnittliche r.m.s.d. für alle Seitenketten $B$ -Faktoren <sup>c</sup> [Å]	2.33
Anzahl der Wassermoleküle	303
durchschnittlicher $B$ -Faktor <sup>c</sup> für alle Solvensatome [Å <sup>2</sup> ]	37.1

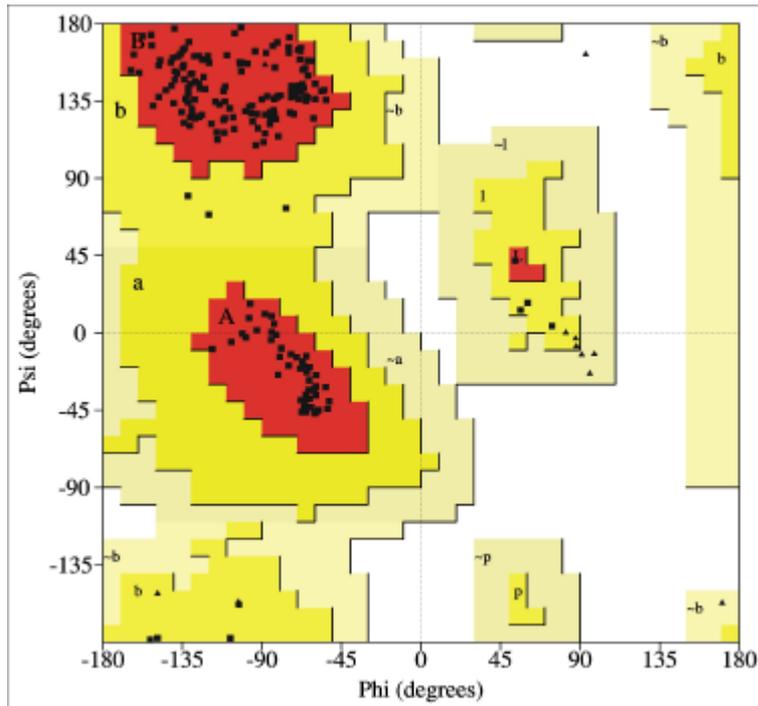
**Tabelle 5-2: Verfeinerungsstatistik von PI-SceI-DI**

<sup>a</sup>  $R_{\text{cryst}} = \frac{\sum |F_{\text{obs}} - F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$ , <sup>b</sup> berechnet mit 7 % zufällig ausgewählten Reflexe;  
<sup>c</sup> berechnet mit BAVEGAGE (Collaborative Computational Project, 1994).

**Abbildung 5.15:  $R_{\text{cryst}}$  und  $R_{\text{free}}$  im Verlauf der Verfeinerung**

$R_{\text{cryst}}$  (●) und  $R_{\text{free}}$  (○) der Verfeinerungsschritte mit Refmac5 [%]. (a) Daten der hauseigenen Anlage, CNS-Verfeinerung. (b) Ab hier Verwendung der DESY-Daten, erster Verfeinerungsschritt mit CNS. (c) Automatische Dichteinterpretation mit ARP. (d) Entfernung der Wassermoleküle, ab hier: ONO-Modellbau/REFMAC5-Verfeinerungszyklen. (e) Wassersuche mit ARP. (f) Ab hier anisotrope  $B$ -Faktor-Verfeinerung.

Im Ramachandran-Diagramm, Abbildung 5.16, der Verteilung der Peptidbindungswinkel, liegen 181 Reste (94.8 % der 191 Nicht-Glycin und Nicht-Prolin-Aminosäuren) im am meisten favorisierten Bereich, die restlichen 10 (5.2 %) im zusätzlich erlaubten Bereich. Der Ausschnitt der Elektronendichte in Abbildung 5.17 reflektiert die Qualität der Struktur, in der bei einer Auflösung von 1.35 Å viele einzelne Details sichtbar sind.

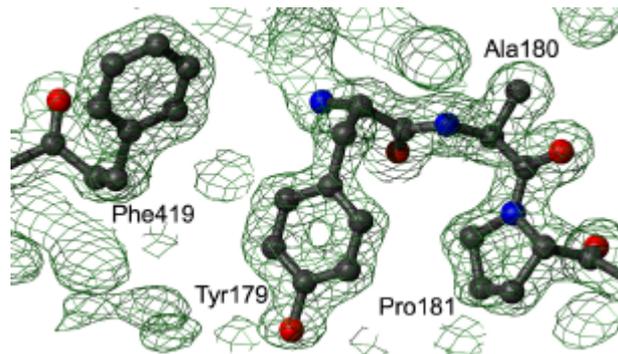


**Abbildung 5.16:**  
**Ramachandran-Diagramm**  
**der Struktur von PI-SceI-DI**

Erzeugt mit dem Programm  
Procheck (Collaborative  
Computational Project, 1994).

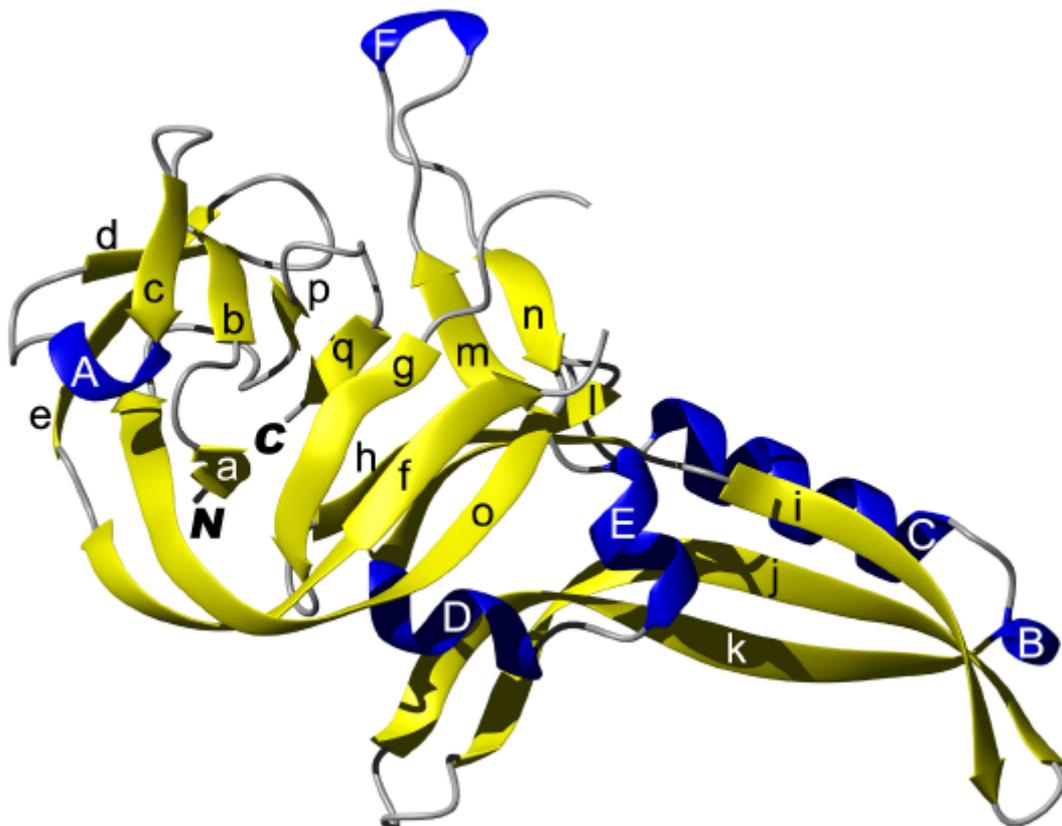
**Abbildung 5.17: Elektronendichte**

$2F_o - F_c$  Elektronendichte des Bereiches Phe419, Tyr179, Pro181, konturiert mit  $1.5 \sigma$ . Das Proteinmodell ist in der Kugelstab-Darstellung zu sehen mit Kohlenstoffatomen und Bindungen in grau, Sauerstoffatomen in rot und Stickstoffatomen in blau.

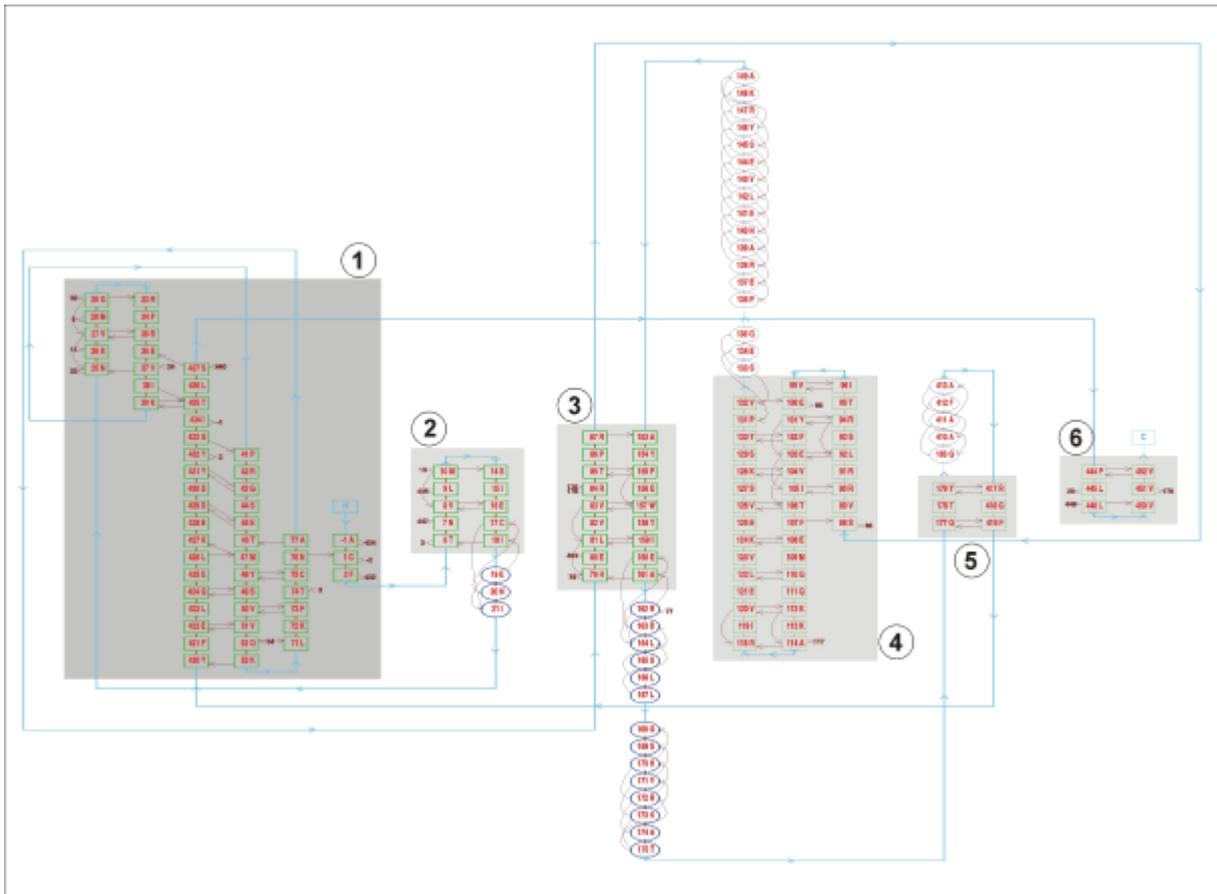


### 5.3.4. Struktur der Domäne I von *S. cerevisiae* PI-SceI

Die Struktur der Domäne I der Homing Endonuklease von *S. cerevisiae* besteht hauptsächlich aus  $\beta$ -Strängen (51 % der in der Elektronendichte definierten Struktur) und zu einem geringeren Anteil (14 %) aus Helices. Insgesamt 17  $\beta$ -Stränge bilden 6 antiparallele Faltblätter. In einem davon bilden die Stränge **a**, **d**, **e**, **f**, **g** und **o** ein verzerrtes Fass (Barrel). Von den sechs Helices sind zwei (**C** und **E**)  $\alpha$ -Helices und vier (**A**, **B**, **D** und **F**) vom  $3_{10}$ -Typ. Die Stränge **i**, **j** und **k** bilden zusammen mit den Helices **B** und **C** eine zangenartige Subdomäne (Reste 86 bis 154), die mit der Kern-Subdomäne über ein scharnierartiges Gelenk verbunden ist. Möglicherweise kann auch die Region, die von den Helices **D** und **E** und dem Strang **l** gebildet wird, dazugezählt werden. Abbildung 5.18 zeigt eine schematische Darstellung der Struktur, Abbildung 5.19 ein Schema der Sekundärstruktur.



**Abbildung 5.18: Bänder-Modell der Struktur von PI-SceI Domäne I**  
Helices sind in blau,  $\beta$ -Stränge in gelb dargestellt.

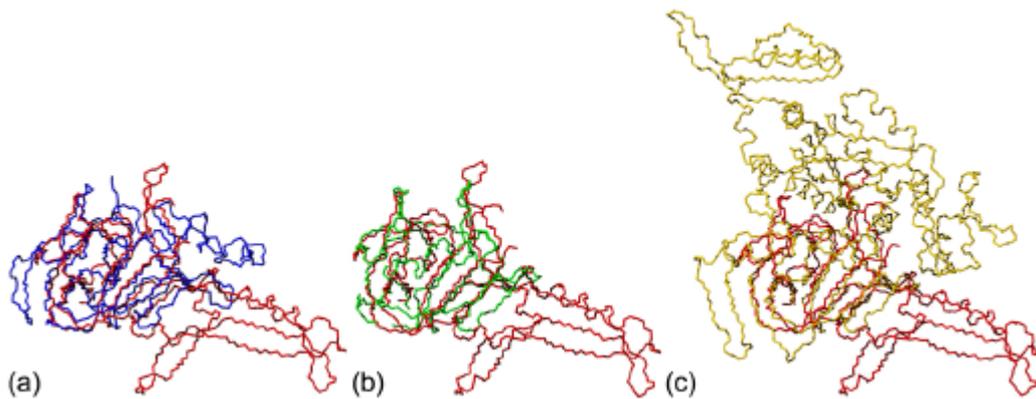


**Abbildung 5.19: Sekundärstruktur-Schema von PI-SceI-DI**

Sekundärstruktur-Diagramm des Programms PROMOTIF (Hutchinson und Thornton, 1996). Die Aminosäuren, deren Nummern in grünen Kästen stehen gehören zu  $\beta$ -Strängen, die in blauen Ellipsen zu Helices. Schwarze Pfeile symbolisieren Wasserstoffbrückenbindungen zwischen den verbundenen Resten, hellblaue Pfeile die Schleifen-Regionen zwischen den Sekundärstrukturelementen (sowie die Laufrichtung der Peptidkette). Grau unterlegt sind die  $\beta$ -Faltblätter, bezeichnet mit Zahlen in Kreisen; in dunkelgrau das „Barrel“.

PI-SceI Domäne I nimmt die Hint-Faltung der Hedgehog/Intein-Domäne ein (komplexe Faltung bestehend aus fünf  $\beta$ -Hairpin-Einheiten und zusätzlichen Elementen) und ist ein Mitglied der Superfamilie mit demselben Namen (in SCOP (Murzin *et al.*, 1995) klassifiziert; CATH ID 1.170.16.10 (Orengo *et al.*, 1997)). Domäne I gehört wie PI-PfuI und GyrA zu der Familie der Inteine (Protein-Splicing-Domäne). Sie kann mit diesen Protein strukturell überlagert werden, genauso wie mit der C-terminalen (Hog) Autoprocessing-Domäne von Hedgehog, einem Mitglied der anderen Familie in dieser Superfamilie. R.m.s.-Abweichungen der strukturellen Überlagerungen betragen 2.2 Å mit PI-PfuI von *Pyrococcus furiosus* (14 % Sequenzidentität, PDB-Code 1dq3 (Ichiyanagi *et al.*, 2000), 2.0 Å mit GyrA von *Mycobacterium xenopi* (19 % Sequenzidentität, PDB-Code 1am2 (Klabunde *et al.*, 1998)) und 2.2 Å mit Hedgehog-Autoprocessing-Domäne,

17 kDa Fragment von *Drosophila melanogaster* (13 % Sequenzidentität, PDB-Code 1at0 (Hall *et al.*, 1997)). Diese drei Proteinen sind auch die einzigen Treffer in einer DALI-Suche (Holm und Sander, 1993) mit einem Z-Wert größer als 3 (Z Werte: 13.7 für 1am2, 13.6 für 1at0 und 11.8 für 1dq3\_A). Allerdings decken sie nur die Kern-Subdomäne von PI-SceI Domäne I ab. Die zangenförmige Subdomäne alleine ergibt nur Treffer mit einem Z-Wert von maximal 4.4. Keines dieser Proteine hat jedoch tatsächlich dieses zangenförmige Element sondern überlagert nur mit Helix **C** und den direkt benachbarten Teilen der Stränge **i**, **j** und **k**. Vergleichbares wird in der Strukturanalyse der Homing Endonuklease PI-PfuI beobachtet (Ichiyanagi *et al.*, 2000). Strukturell verwandte Proteine von PI-PfuI zeigen nur Übereinstimmungen im Kernbereich, während die sog. Stirrup-Domäne keine Entsprechungen besitzt. Sie ist im Gegensatz zur zangenartigen Subdomäne von PI-SceI-DI jedoch nicht in eine interne Schleife der Protein-Splicing-Domäne inseriert, sondern zwischen den beiden Domänen.



**Abbildung 5.20: Strukturüberlagerungen von PI-SceI Domäne I mit Proteinen der Superfamilie**

Strukturelle Überlagerung von (a) 1gpp und 1am2 (GyrA, Überlagerung von 122 Resten entsprechend der DALI-Suche). (b) 1gpp und 1at0 (17 kDa Hedgehog-Fragment, 145 Reste). (c) 1gpp und 1dq3 Monomer A (PI-PfuI, 118 Reste). Gezeigt sind nur die Peptid-Rückgrat-Atome, PI-SceI-DI (1gpp) ist jeweils in rot dargestellt.

### 5.3.5. Vergleich aller Strukturen von PI-SceI Domäne I

Zusammen mit der kürzlich von Mizutani *et al.* (2002) publizierten Struktur (PDB-Code 1jva, Koordinaten dankenswerterweise von Prof. Satow zur Verfügung gestellt) und der hier beschriebenen (PDB-Code 1gpp) gibt es nun fünf Strukturen der Domäne I von PI-SceI, die in Tabelle 5-3 zusammengefasst sind.

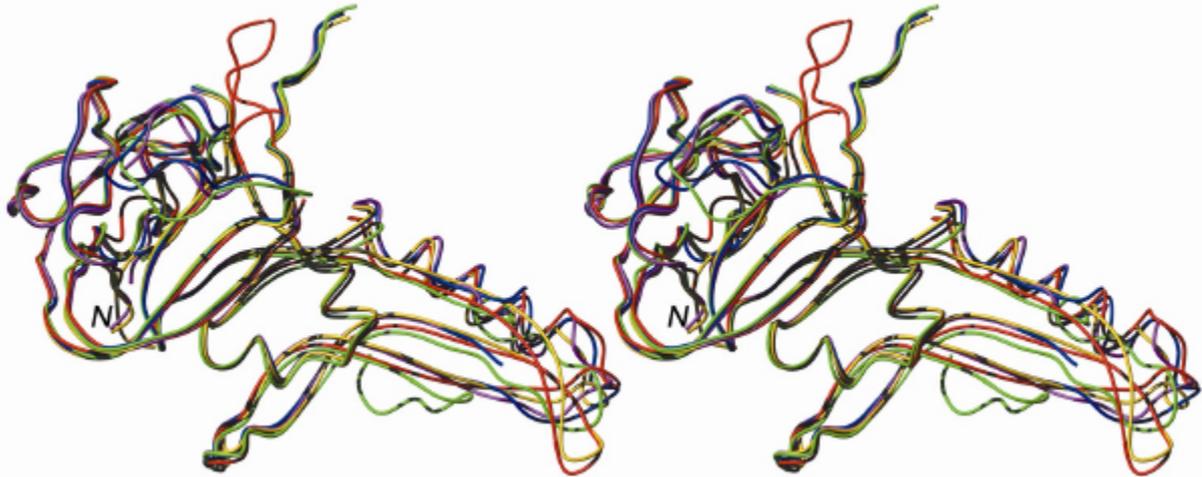
	1vde	1dfa	1ef0	1jva	1gpp
Referenz	(Duan <i>et al.</i> , 1997)	(Hu <i>et al.</i> , 2000)	(Poland <i>et al.</i> , 2000)	(Mizutani <i>et al.</i> , 2002)	<b>Werner (diese Arbeit)</b>
Methode	MAD (SeMet)	MR (1vde)	MR (1vde)	MR (1vde)	MR (1vde)
Auflösung[Å]	10 – 2.4	10 – 2.0	20 – 2.1	30 – 2.1	20 – 1.35
Raumgruppe	P2 <sub>1</sub>	C2	P2 <sub>1</sub>	P1	C2
Monomere pro AU	2	1	2	2	1
R <sub>sym</sub> [%]	6.0	6.0	11.0	3.2	2.9
Vollständigkeit [%]	96.7	95.0	95.1	89.4	94.3
R <sub>cryst</sub> [%]	19.2	21.2	23.1	19.8	15.0
R <sub>free</sub> [%]	23.8	28.0	28.1	24.0	18.9
r.m.s.d. Bindungslängen [Å]	0.14	0.01	0.007	0.006	0.018
r.m.s.d. Bindungswinkel [°]	1.9	0.18	1.3	1.2	1.83
Anzahl der Proteinatome (non H) pro Monomer	6768	6730	6840	6755	2763
Durchschnittlicher B-Faktor* [Å <sup>2</sup> ]	37.5	32.0	34.8	25.0	25.3
Anzahl der Solvensmoleküle	349	240	312	205	303
Fehlende Reste in Domäne I	Ser <sup>93</sup> – Phe <sup>102</sup>	Gly <sup>135</sup> – Asn <sup>151</sup>	A: Arg <sup>91</sup> – Glu <sup>103</sup> B: Gln <sup>65</sup> – Leu <sup>70</sup> , Thr <sup>95</sup> – Gly <sup>98</sup>	A: Gln <sup>55</sup> – Ser <sup>64</sup> B: Gln <sup>55</sup> – Glu <sup>66</sup>	Gln <sup>55</sup> - Arg <sup>66</sup> , Asn <sup>454</sup>
Mutationen im Vergleich zu P17255	--	--	Ile <sup>-4</sup> <-> Lys Tyr <sup>-3</sup> <-> Val Val <sup>-2</sup> <-> Glu Cys <sup>1</sup> <-> Ala Asn <sup>454</sup> <-> Ala	Cys <sup>1</sup> <-> Ser His <sup>79</sup> <-> Asn Asn <sup>454</sup> <-> Ser Cys <sup>455</sup> <-> Ser	Val <sup>-2</sup> <-> Ser Gly <sup>-1</sup> <-> Ala Arg <sup>44</sup> <-> Ser Val <sup>67</sup> <-> Met Ile <sup>132</sup> <-> Val Leu <sup>183</sup> <-> Gly
N-terminale Extein-Reste	--	--	Lys <sup>-4</sup> - Leu <sup>-3</sup> - Glu <sup>-2</sup> - Gly <sup>-1</sup> (B)	Ile <sup>-5</sup> - Ile <sup>-4</sup> - Tyr <sup>-3</sup> - Val <sup>-2</sup> - Gly <sup>-1</sup> (B)	Ser <sup>-2</sup> - Ala <sup>-1</sup>
C-terminale Extein-Reste	--	--	Cys <sup>455</sup> - Gly <sup>456</sup> - Glu <sup>457</sup> - Arg <sup>458</sup> (B)	Ser <sup>455</sup> - Gly <sup>456</sup> - Glu <sup>457</sup> - Arg <sup>458</sup>	--

Tabelle 5-3: Zusammenfassung der fünf Strukturen von PI-SceI (Domäne I)

\* berechnet mit BAVERAGE (Collaborative Computational Project, 1994).



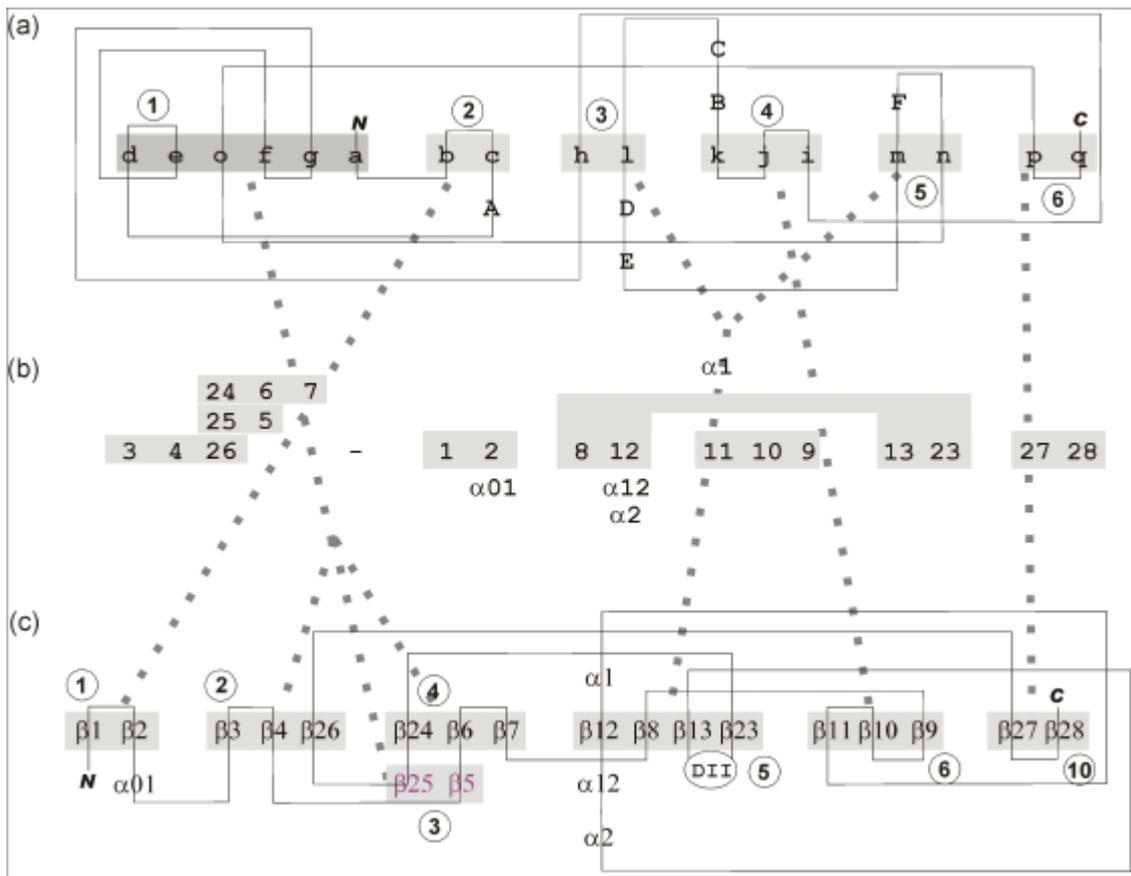
Abbildung 5.21 zeigt einen Sequenzvergleich aller in den Strukturen vorhandenen Aminosäureketten der Domäne I, Abbildung 5.22 eine strukturelle Überlagerung der jeweiligen A-Ketten.



**Abbildung 5.22: Strukturelle Überlagerung von PI-SceI-DI**

Stereo-Bänder-Darstellung der Domänen I (Reste C-Terminus bis 182 und 410 bis N-Terminus) von 1vde\_A (blau), 1dfa (grün), 1ef0\_A (magenta) und 1jva\_A (orange) nach struktureller Überlagerung aller Atome jeweils mit 1gpp (rot) (Programm LSQKAB (Collaborative Computational Project, 1994)).

Die Strukturen stimmen im Wesentlichen überein, weisen jedoch aufgrund verschiedener Proteinketten-Konstrukte und Kristallkontakte auch Unterschiede auf. Unterschiede in den Sekundärstrukturen sind marginal. Das oben erwähnte Fass in 1gpp schließt nur, da die zusätzlichen N-terminalen Reste ein zusätzliches  $\beta$ -Faltblatt (**a**) bilden. Dies wird in 1jva ebenfalls beobachtet. In 1ef0\_B bilden die 4 sichtbaren N-terminalen Extein-Reste zwar auch ein zusätzliches Faltblatt, ein Barrel wird jedoch nicht erkannt (Programm PROMOTIF (Hutchinson und Thornton, 1996)). Zum Barrel äquivalente Elemente in 1vde sind die Faltblätter 2 bis 4 (Stränge **b3** bis **b7** und **b24** bis **b26**). Ein weiterer geringer Unterschied besteht darin, dass das 1vde-Faltblatt 5 (**b8**, **b12**, **b13** und **b23**) in 1gpp zu Faltblättern 3 (Stränge **h** und **l**) und 5 (Stränge **m** und **n**) gespalten erscheint. Die zusätzliche Helix **F** (Reste 183 bis 413) ist artifiziell, da sie sich an der Stelle befindet, wo in allen anderen Strukturen Domäne II ist. Abbildung 5.23 fasst diese Unterschiede zusammen.



**Abbildung 5.23: Unterschiede in den Sekundärstrukturelementen**

(a) Sekundärstrukturelemente von 1gpp analog Abbildung 5.19. Kleinbuchstaben entsprechen  $\beta$ -Strängen, Großbuchstaben Helices. Grau unterlegt und mit Zahlen in Kreisen beziffert sind die Faltblätter, dunkelgrau das Barrel. (b) Entsprechende Sekundärstrukturelemente in 1vde. Übereinander liegende Elemente sind strukturelle Entsprechungen. Die Zahlen stellen den jeweiligen  $\beta$ -Strang dar. (c) Sekundärstrukturelemente von 1vde. Das Faltblatt 3 (Stränge **b25** und **b5**) ist in 1gpp im Faltblatt 1 (genauer: in den Strängen **o** und **f** enthalten). Die gestrichelten Linien markieren die strukturellen Entsprechungen von 1gpp und 1vde.

Interessanterweise ist in keiner der Strukturen die vollständige Sequenz sichtbar. Hier, genauso wie in 1jva, ist die Schleife Gln55 bis Glu66 ungeordnet, die in 1vde sichtbar ist. Andererseits sind die Regionen Ser93 bis Phe102 und Gly135 bis Asn151 gut definiert, die in 1vde und 1dfa nicht zu sehen sind. 1ef0 hat ein Monomer, in dem die Schleife Arg91 bis Glu103 fehlt (wie in 1vde) und eines, in dem Gln55 bis Leu70 (wie in 1gpp) und Thr95 bis Gly98 (wie in 1vde) fehlen. Genau diese Schleifen und deren Randbereiche haben die höchsten  $B$ -Faktoren und die größten r.m.s.d.-Werte nach struktureller Überlagerung (Abbildung 5.24) und können somit als die flexiblen Bereiche der Domäne I angesehen werden. Insbesondere die distalen Bereiche der Zangen-Subdomäne, etwa die Schleife zwischen den Strängen i und j (Reste 96 bis 99), zeigen große Abweichungen zwischen den Strukturen. Zwei Regionen mit geringeren aber deutlich vorhanden

Unterschieden sind die Reste 30 bis 32 und 438 bis 441. Beide sind in engem Kontakt zu Helix C des symmetrieverwandten Moleküls (-X, Y, -Z) in 1gpp, während die erste Region in keiner anderen Struktur einen Kristallkontakt hat und die zweite Kristallkontakte in 1vde und 1dfa mit der Schleife hat, die in 1gpp fehlt (um Rest 55), in 1jva mit einem Bereich in Domäne II oder keinen in 1ef0. Enge Kristallkontakte sind insbesondere für 1gpp charakteristisch, wo 96 von 217 Resten (44 %) eine Kontaktfläche mit symmetrieverwandten Molekülen haben. Dies entspricht etwa  $3200 \text{ \AA}^2$  oder 27 % der zugänglichen Oberfläche („solvent accessible surface“, berechnet mit NACCESS (Hubbard und Thornton, 1993)). Im Falle von 1dfa sind es immerhin auch 82 von 210 Domäne-I-Resten, die Kontakte mit symmetrieverwandten Strukturen oder Domäne II haben (ca.  $2500 \text{ \AA}^2$ ), bei 1vde 49 von 217 ( $1700 \text{ \AA}^2$ ), bei 1ef0 48 von 217 ( $1800 \text{ \AA}^2$ ) und bei 1jva 38 von 223 ( $1500 \text{ \AA}^2$ ). Die Kristallkontakte sind in Abbildung 5.24b symbolisiert und in Abbildung 5.25 dargestellt.

zu Abbildung 5.24:

Vergleich von 1gpp (rot), 1vde (blau), 1dfa (grün), 1ef0 (magenta/violett) und 1jva (orange). (a) In den Kristallstrukturen sichtbare Reste. Für 1ef0 und 1jva entsprechen die Quadrate der Kette A, die Rauten der Kette B. (b) Kristallkontakte. Ein Kreis symbolisiert einen Kontakte der entsprechenden Aminosäure mit einem Rest in Domäne II oder einem symmetrieverwandten Molekül. (c) R.m.s.-Abweichung der Hauptkettenatome (Dreieck) und der Seitenkettenatome (Kreis) nach struktureller Überlagerung mit 1gpp (Programm LSQKAB (Collaborative Computational Project, 1994)). (d) Durchschnittliche *B*-Faktoren der Aminosäuren, berechnet mit BAVERAGE (Collaborative Computational Project, 1994).

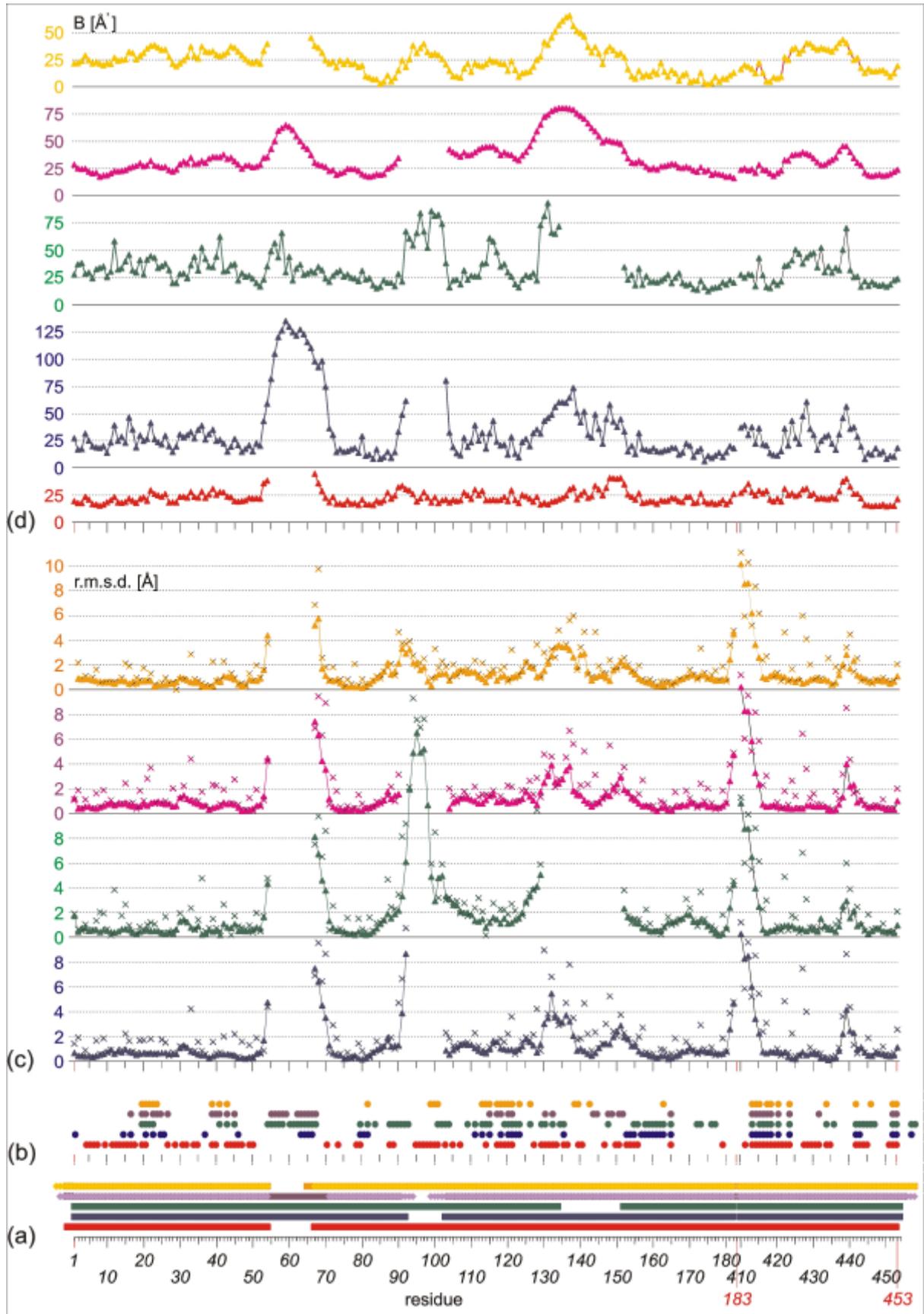
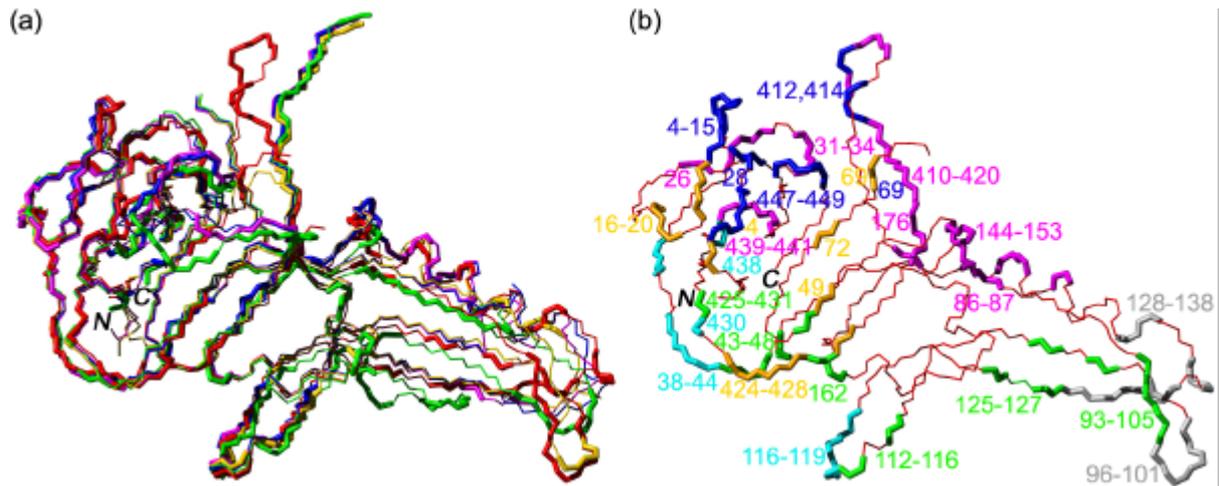


Abbildung 5.24: Vergleich der Domäne I-Strukturen



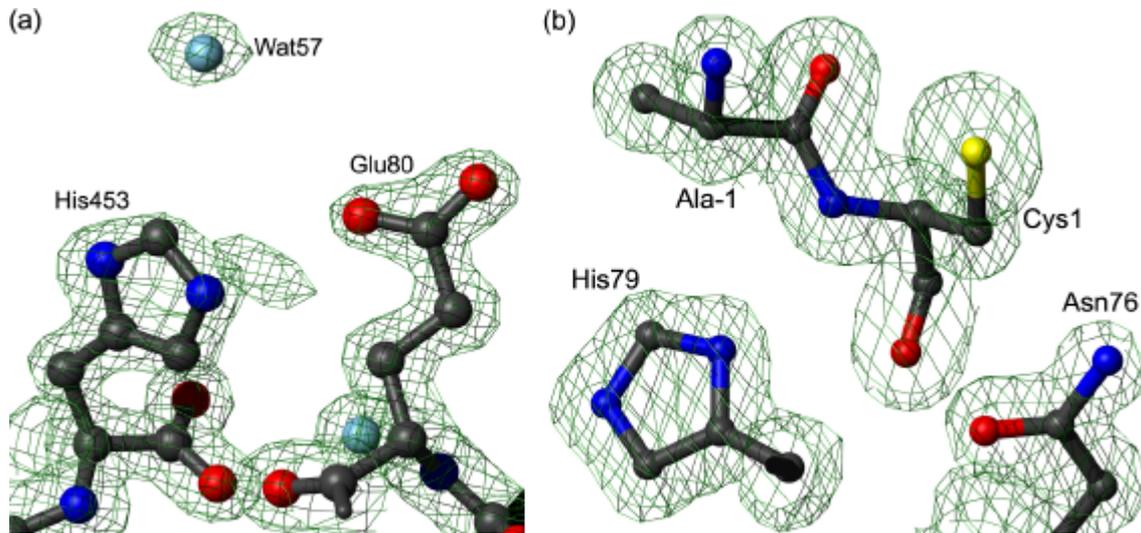
**Abbildung 5.25: Kristallkontakte von PI-SceI Domäne I**

(a) Strukturelle Überlagerung von 1gpp (rot), 1vde\_A (blau), 1dfa (grün), 1ef0\_A (magenta) und 1jva\_A (orange). Dargestellt ist jeweils das Proteinrückgrat. Bindungen mit vergrößertem Radius stellen Reste dar, die Kontakte mit symmetrieverwandten Molekülen oder Domäne II eingehen. (b) Kristallkontakte von 1gpp. Auch hier ist lediglich das Proteinrückgrat gezeigt. Farbige und verdickte Bindungen zeigen Kristallkontakte mit den symmetrieverwandten Molekülen nach der Symmetrieoperation  $(-X, Y, -Z)$  in magenta,  $(-\frac{1}{2}+X, \frac{1}{2}+Y, Z)$  in dunkelblau,  $(\frac{1}{2}-X, \frac{1}{2}+Y, -1-Z)$  in orange,  $(\frac{1}{2}-X, 1\frac{1}{2}+Y, -1-Z)$  in hellblau,  $(-X, Y, -1-Z)$  in grün und  $(\frac{1}{2}+Y, 1\frac{1}{2}+Y, Z)$  in grau.

### 5.3.6. Protein-Splicing-Stelle

Das aktive Zentrum von PI-SceI Domäne I, die Protein-Splicing-Stelle, ist nur partiell vorhanden in der hier dargestellten Struktur. Komplette abwesend ist das C-terminale Extein, genauso wie die C-terminale Intein-Aminosäure Asn454. Der N-Terminus des Inteins ist komplett (einschließlich Cys1) und Teile des Reinigungs-*tag* imitieren das N-terminalen Extein, wenn auch mit unterschiedlichen Seitenketten (Ala<sup>-1</sup> und Ser<sup>-2</sup> anstatt Gly<sup>-1</sup> und Val<sup>-2</sup>). Die Struktur des aktiven Zentrums kann am besten mit den Strukturen 1ef0 (Kette B) (Poland *et al.*, 2000) und 1jva (Mizutani *et al.*, 2002) verglichen werden, die jeweils Extein-Reste auf beiden Seiten enthalten (siehe Tabelle 5-3, Seite 90). In 1ef0 sind die terminalen Reste zu Alanin mutiert, in 1jva zu Serin, jeweils aus dem Grund, die Splicing-Reaktion zu verhindern. Weiterhin mutiert in 1jva sind der konservierte Rest His79 (zu Asparagin) und der erste Extein-Rest Cys455 (zu Serin). Das für 1ef0 beschriebene Zink-Atom, welches dort von His453, Cys455, Glu80 und Wasser53 koordiniert ist, wird hier nicht beobachtet (Abbildung 5.26a), unter anderem, da die Reste 453 und 454 fehlen. Die Seitenketten von His453 und Glu80 sind im Vergleich zu 1ef0 verdreht, was im Falle von His453

einzigartig für alle Strukturen ist. Das Zinkatom wird von Poland *et al.* (2000) als höchstwahrscheinlich strukturell und nicht als funktionell beschrieben.



**Abbildung 5.26: Aktives Zentrum von PI-SceI Domäne I**

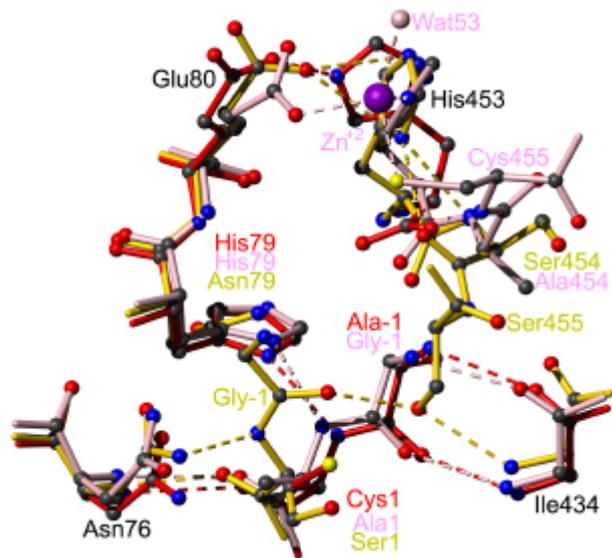
$2F_o-F_c$  Elektronendichte, konturiert mit  $1.5 \sigma$ . (a) His453 (der C-Terminus in 1gpp), Glu80 und Wat57. Elektronendichte für ein Metall-Ion wurde nicht beobachtet, Wasser57 ist nicht identisch mit Wasser53 von 1ef0. (b) His79, Asp76 und die Schnittstelle Ala<sup>-1</sup> und Cys1.

Die Lage der Exteine unterscheidet sich teilweise deutlich voneinander, wobei 1gpp und 1ef0 klare Ähnlichkeiten aufweisen. Offensichtlich beeinflussen die Exteine jedoch die Konformation des Inteins kaum, die für alle Strukturen in diesem Bereich sehr ähnlich ist (Abbildung 5.27). Der Rest -1 (Ala in 1gpp, Gly in 1ef0) ist hier eingebunden in das  $\beta$ -Faltblatt 1 und bildet die dafür typische Wasserstoffbrückenbindungen mit den Peptidrückgratatomen von Ile434 (Strang **o**) aus. In 1jva dagegen wird diese Aminosäure (Gly<sup>-1</sup>) und damit das N-terminale Extein durch eine Wasserstoffbrücke des angrenzenden Stickstoffes von Ser1 mit ND2 von Asp76 in der entgegen gesetzten Richtung stabilisiert. Gleichzeitig bildet sich eine Wasserstoffbrückenbindung zwischen Gly<sup>-1</sup>-O und Ser455-OG. Da es sich im realen Fall um Cys455 handelt, dem ersten C-terminalen Extein-Rest, befindet sich also das entsprechende Heteroatom schon in der Nähe von Gly<sup>-1</sup>-C (3.7 Å in der Struktur), dem Angriffspunkt während der Transesterifizierung. In 1gpp fehlen die Reste 454 und 455, weswegen hierzu keine Aussage gemacht werden kann. In 1ef0 dagegen ist das Heteroatom Cys455-SG von Zink koordiniert und daher mehr als 7 Å vom späteren Reaktionszentrum entfernt. Es kann angenommen werden, dass Zink auf diese

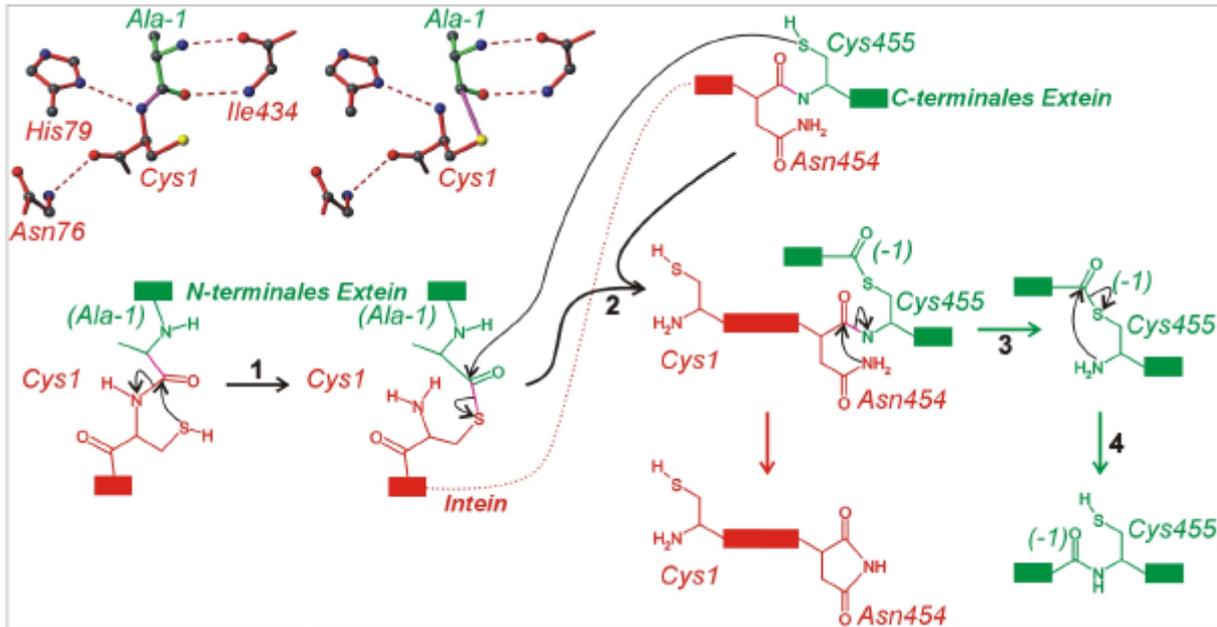
Weise hier eher einen inhibitorischen Effekt hat, wie das auch für RecA beschrieben ist (Mills und Paulus, 2001).

### Abbildung 5.27: Aktives Zentrum von PI-SceI Domäne I

Aktives Zentrum nach struktureller Überlagerung aller Atome der Domänen I von 1gpp (rot), 1ef0B (pink) und 1jvaA (gelb). Wasserstoffbrücken sind als gestrichelte Linien in der jeweiligen Farbe dargestellt. Bezeichnungen von Resten, die in allen Strukturen gleich sind, sind in schwarz, wenn sie unterschiedlich sind in Farbe dargestellt.



Der erste Schritt der Splicing-Reaktion, der N-S Acyl-Shift, kann dennoch übereinstimmend nachvollzogen werden (Abbildung 5.26b) und Abbildung 5.28, siehe auch Abbildung 5.4, Seite 70). Cys1-SG initiiert mit einem nukleophilen Angriff auf die Carbonylgruppe von Ala<sup>-1</sup> (Gly<sup>-1</sup> im realen Extein) die Umlagerung. Cys1-N, der spätere N-Terminus des Inteins, wird von ND1 des konservierten Rests His79 über eine Wasserstoffbrückenbindung stabilisiert. Es sei jedoch darauf hingewiesen, dass alle Strukturen in diesem entscheidenden Bereich Mutationen enthalten und daher im komplizierten Geflecht von Wechselwirkungen kleine Änderungen schon entscheidende Auswirkungen haben können. Die Bedeutung der konservierten Reste Cys1 und His79 wird jedoch deutlich durch ihre Rolle im Reaktionsmechanismus. Die weiteren Schritte entsprechend des zurzeit akzeptierten Mechanismus' (Noren *et al.*, 2000) sind in Abbildung 5.28 schematisch dargestellt.



**Abbildung 5.28: Protein-Splicing-Modell für den Reaktionsmechanismus**

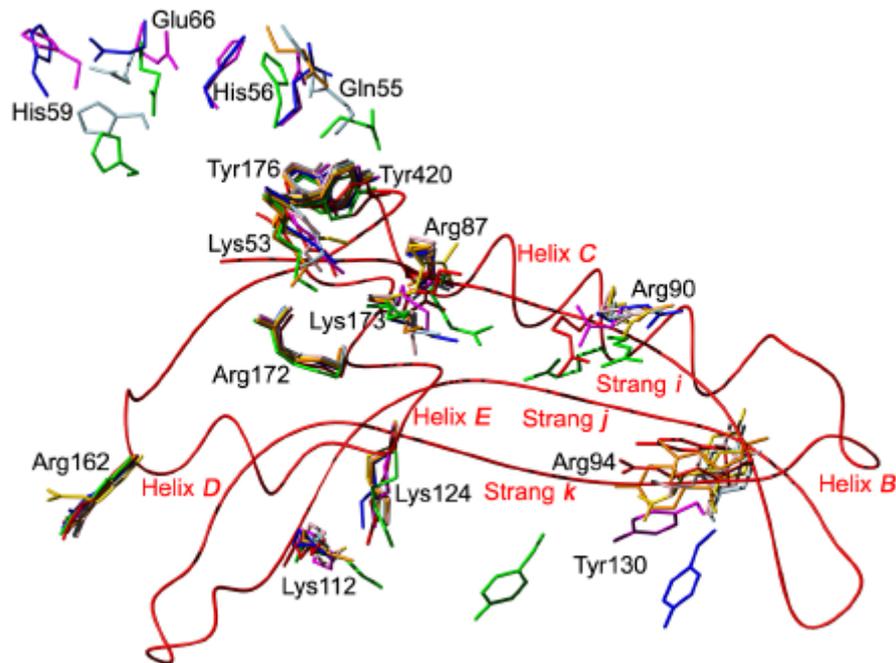
(1) N-S Acyl-Shift (2) Transesterifizierung (3) Succinimid-Bildung (4) S-N Acyl-Shift. Intein-Bestandteile (PI-SceI) sind in rot dargestellt, Extein-Bestandteile (die Untereinheit der ATPase) in grün, die Intein-Extein-Verbindungen in violett. Die hier präsentierte Struktur 1gpp gibt einen Einblick auf die Ausgangssituation (links oben).

### 5.3.7. DNA-Bindung

PI-SceI bindet seine Erkennungssequenz von 31 Basenpaaren spezifisch. Dabei wird der Bereich +5 bis +21, die minimale Bindungsregion, fest gebunden und verbleibt nach der endonukleolytischen Reaktion am Protein. Der Bereich, der die Schnittstellen enthält, die Basenpaare -10 bis +4, wird dagegen nur locker gebunden. Die DNA ist stark gebogen, um überhaupt vom Protein in dieser Länge gebunden zu werden (siehe Kapitel 5.1.5, Seite 70).

Ein genaues Verständnis der Art der DNA-Bindung könnte große Bedeutung im Bereich der Gentechnologie haben. Viele DNA-bindende Proteine erkennen nur wenige Basenpaare, diese jedoch hochspezifisch, so z.B. die meisten bekannten Restriktionsendonukleasen. Obwohl mehrere Gruppen versucht haben, Cokristalle von PI-SceI und DNA zu züchten, schlug dies bisher in jedem Fall fehl. So auch hier, wo versucht wurde, den Protein-DNA-Komplex auf die Region der festen Bindung zu reduzieren (siehe Kapitel 5.3.2, Seite 81). Offenbar sind die Wechselwirkungen der Proteine im Kristall stärker als die des Proteins mit den verschiedensten DNA-Fragmenten, so dass in allen Fällen das Protein alleine

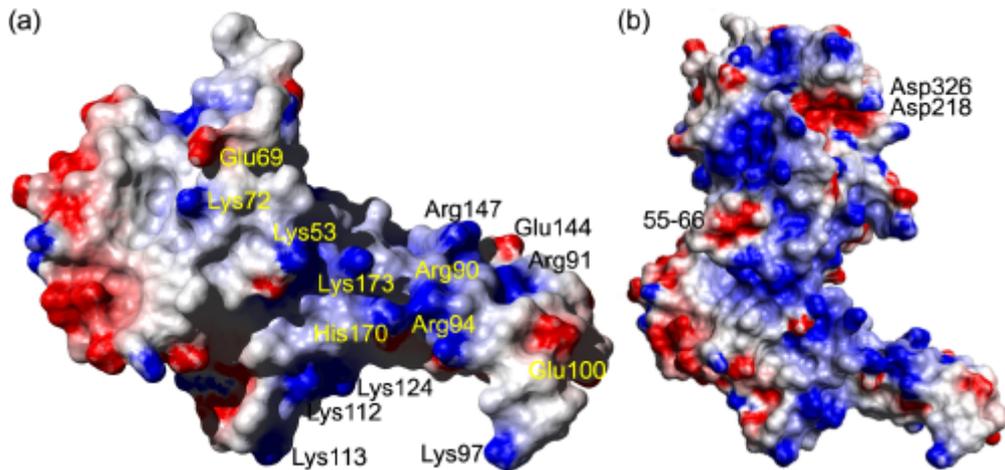
kristallisiert, lediglich etwas verlangsamt im Vergleich zu Lösungen, die nur Protein enthalten. Umfangreiche Kristallkontakte im Fall von Domäne I sind schon beschrieben worden (Kapitel 5.3.5, Seite 96). Sie sind ein Beleg für die zahlreichen Protein-Protein-Wechselwirkungen, genauso wie das sehr schnelle Kristallwachstum (teilweise innerhalb von Minuten) einen Vorteil dieser Art der Bindung nahe legt. Da also strukturelle Informationen zur DNA-Bindung fehlen, kann nur aufgrund von biochemischen Daten ein Modell erstellt werden. Diese existieren zahlreich und sind in Kapitel 5.2.2, ab Seite 77 und in Abbildung 5.8, Seite 74 zusammengefasst. Die 16 Aminosäuren von Domäne I, von denen die Bedeutung für die DNA-Bindung bekannt ist (sei es, dass ihre Mutation Bindung oder Schnitt der DNA behindert oder ein Quervernetzung nachgewiesen wurde (Tyrosine 176 und 420 sowie Lysine 112 und 124)), sind in Abbildung 5.29 für alle Proteinketten der fünf Strukturen dargestellt. Sie konzentrieren sich auf zwei Regionen, die „Vorderseite“ der zangenartigen Subdomäne (Stränge **i**, **j** und **k**) bzw. angrenzende Helices **D** und **E** sowie der Bereich der Schleife 55 bis 66, die in 1gpp fehlt. Zu letzterem Bereich an der Domänengrenze können im weiteren Sinne auch die Tyrosine 176 und 420 gezählt werden. Im Gegensatz zu den meisten anderen dieser 16 Aminosäuren sind gerade diese Tyrosine strukturell sehr konstant in allen Strukturen. Die Schleife 55 bis 66, vom dem eine Flip-Bewegung in Richtung der DNA vermutet wird (Wende *et al.*, 2000), dagegen zeigt besonders große Abweichungen, die in etwas geringerem Maße auch für die Arginine 87, 90 und 94 sowie Tyr130 und Lys173 beobachtet werden (siehe auch die r.m.s.-Abweichungen in Abbildung 5.24, Seite 95). Kann für die flexiblen Bereiche der Domäne I, also die Schleife 55 bis 66 und die Reste 90 bis 105 (Stränge **i** und **j**), angenommen werden, dass sie sich möglicherweise im Verlaufe der DNA-Bindung in Richtung des Bindungspartners bewegen, so könnten die strukturell unflexible Reste (insbesondere Tyr176, Tyr420, Arg172, Arg162, aber auch Lys112 und Lys124) an der initialen Positionierung der DNA beteiligt sein.



**Abbildung 5.29: Bedeutsame Reste für die DNA-Bindung**

Aminosäuren, die für die DNA-Bindung von PI-SceI Domäne I von Bedeutung sind nach struktureller Überlagerung der hier präsentierten Struktur 1gpp (rot) mit den Proteinketten der anderen Strukturen: 1vdeA (blau), 1vdeB (hell blau), 1dfa (grün), 1ef0A (magenta), 1ef0B (pink), 1jvaA (orange) und 1jvaB (hell orange). Das Proteinrückgrat von 1gpp in der Bänder-Darstellung ist als rote Linie gezeigt, die entsprechenden Sekundärstrukturelemente gekennzeichnet.

Die Ladungsverteilung auf der Oberfläche (Abbildung 5.30) der Domäne I verdeutlicht zudem, dass dieser Vorderbereich der zangenartigen Subdomäne aufgrund der Anhäufung von positiven Ladungen potenziell geeignet ist, die am Rückgrat negativ geladene DNA zu binden. Für die Reste Lys53, Lys173, Lys112, Lys124, Arg90 und Arg94 ist zumindest ein Einfluss auf die DNA Bindung nachgewiesen. Nichts bekannt ist über Lys113, His170 und Arg147, die demzufolge auch als potenziell beteiligt angesehen werden müssen. Von der Mutante K97A (He *et al.*, 1998; Wende *et al.*, 2000) ist jedoch bekannt, dass sie keine (oder keine wesentliche) Bedeutung für die Aktivität des Enzyms hat. Da an dieser Stelle jedoch Kristallkontakte in 1gpp (und auch in 1jva) festgestellt wurden, könnte dies die Konformation von Lys97 maßgeblich beeinflussen und die Orientierung in Richtung der positiv geladenen Region zufällig sein.



**Abbildung 5.30: Oberflächenpotenzial der PI-SceI Domäne I**

Oberflächenpotenzial von Domäne I/1gpp (a) und dem PI-SceI-GV-Konstrukt (siehe Kapitel 5.3.8, Seite 103). (b). Blaue Farbe repräsentiert positive, rote Farbe negative Ladung (berechnet im Programm MOLMOL (Koradi *et al.*, 1996)). Hervorstechende Reste sind bezeichnet, im ganzen Protein die in 1gpp fehlende Schleife 55 bis 66. Das aktive Zentrum der Domäne II hebt sich durch ein Cluster von negativen Ladungen hervor, wo die Aspartate 218 und 326 angesiedelt sind. Ihnen wird die direkte Beteiligung an der endonukleolytischen Reaktion zugeschrieben.

Generell ist jedoch anzumerken, dass bei einer so großen Erkennungssequenz auch eine große Anzahl von Kontakten zwischen Protein und DNA zu erwarten ist. Das Fehlen eines Kontaktes kann in vielen Fällen möglicherweise von anderen Resten kompensiert werden. Der Einfluss auf die Aktivität bliebe dann marginal. So ist auch zu verstehen, dass viele Mutationen in diesem Bereich keinen dramatischen Effekt hervorrufen (Wende *et al.*, 2000). Zudem kann der Effekt einer Mutation nicht unmittelbar als Beweis für einen direkten Kontakt mit der DNA angesehen werden. Vielmehr sind oft indirekte Effekte, etwa eine Kette von Konformationsänderungen, verantwortlich für die Bedeutung eines Restes, der auch weit entfernt vom aktiven Zentrum sein kann. Zweifellos ist dies der Fall, wenn Mutanten in Domäne I, etwa Y420A oder R172A (Wende *et al.*, 2000), eine Reduzierung der Endonuklease-Aktivität in Domäne II bewirken, gleichzeitig aber das Substrat mit unveränderter Affinität binden. Die Vielzahl der Möglichkeiten wird durch die große Anzahl der biochemischen Daten nur geringfügig verringert, teilweise tragen sie sogar eher zur Verwirrung bei.

Den bisher existierenden Bindungsmodellen der DNA an PI-SceI (siehe Abbildung 5.7, Seite 72) liegen solche Überlegungen zugrunde. Die Schnittstellen (zwischen C<sup>+2</sup> und G<sup>+3</sup> im oberen Strang, zwischen C<sup>3</sup> und C<sup>-2</sup> im unteren Strang) wurden in die Nähe des aktiven Zentrums gebracht und der Bereich der Basenpaare +5 bis +21 an die Vorderseite der zangenartigen Subdomäne. Wichtigen Resten

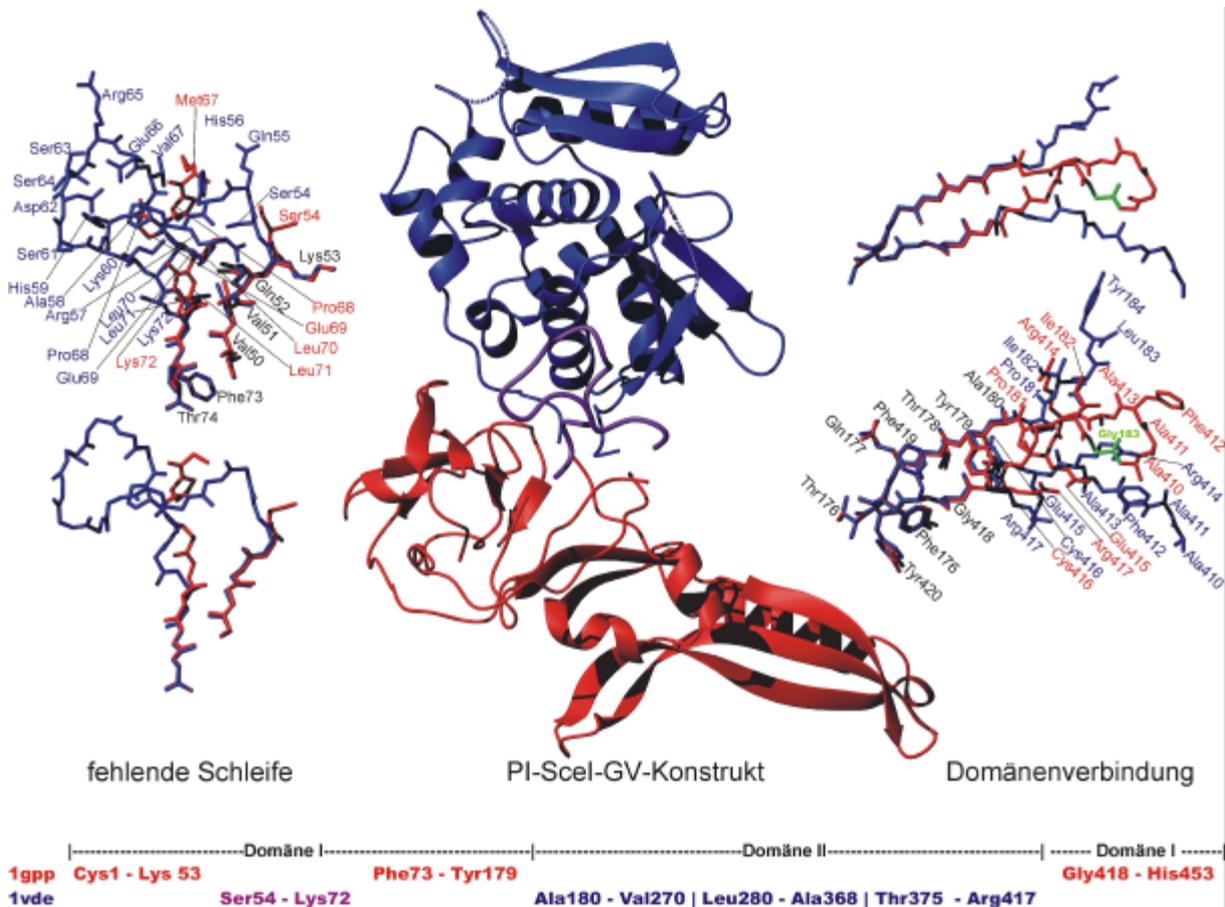
wurde eine Rolle im Bindungsvorgang zugewiesen, wobei insbesondere die nachgewiesenen Photo-Quervernetzungen (Christ *et al.*, 2000; Hu *et al.*, 2000; Pingoud *et al.*, 1999) von großer Bedeutung waren. Abgesehen von einem möglichen Umklappen der Schleife 55 bis 66 (Wende *et al.*, 2000) wurde jedoch stets von PI-SceI als einem festen Körper ausgegangen. Dabei gibt es durchaus Hinweise auf Konformationsänderungen oder Domänenbewegungen. So wurde z.B. beobachtet, dass PI-SceI nach Denaturierungs-/Renaturierungszyklen oder tryptischem Verdau an Position Arg227 (Pingoud *et al.*, 1998) zwar weiterhin DNA bindet, die endonukleolytische Aktivität aber verloren geht. Weiterhin können die isolierten Domänen I und II weder zu einem aktiven Komplex zusammengefügt werden noch einen Supershift mit DNA bewirken (Grindl *et al.*, 1998). Mit anderen Worten, die relative Anordnung der Domänen zueinander spielt eine bedeutende Rolle für die Aktivität. Erweitert man dies auf die schon angesprochene zangenartige Subdomäne, die sich auch aufgrund der strukturellen Überlagerung mit den Mitgliedern der Superfamilien abgrenzen lässt (Abbildung 5.20, Seite 89), kann man die relative Bewegung von drei Untereinheiten diskutieren. Dabei lassen sich unter Umständen auch die Helices **D** und **E** zur Einheit der zangenartigen Subdomäne zählen, da ihre Aminosäuren offenbar ebenfalls an der DNA-Bindung beteiligt sind (Abbildung 5.29).

Um diese Möglichkeit zu untersuchen wurde ein geometriebasiertes Docking durchgeführt, welches sich auf die sehr wenigen (vier) Kontakte zwischen Protein und DNA stützt, die als gesichert angesehen werden.

### **5.3.8. Geometriebasiertes Docking**

Trotz zahlreicher Hinweise auf Aminosäuren von PI-SceI, die an der Bindung der spezifischen Erkennungssequenz beteiligt sind, gibt es nur wenige gesicherte Kontakte, bei denen beide Bindungspartner bekannt sind. Sie befinden sich alle in Domäne II. Dies sind zum einen die Reste des aktiven endonukleolytischen Zentrums Asp218 und Asp326 sowie die Aminosäuren Tyr328 und His333, für die Photo-Quervernetzungen ohne distanzierenden Linker zu den Basen G4 (oberer Strang) bzw. T9 (unterer Strang) nachgewiesen wurden (Christ *et al.*, 2000; Pingoud *et al.*, 1999). Drei weitere Quervernetzungen zwischen Tyr384 und P von T7 (unterer Strang), Tyr420/Tyr176 und P von G13 (oberer Strang) sowie

zwischen Lys112/Lys124 und P von C18 (unterer Strang) wurden beobachtet (Christ *et al.*, 2000). Diese haben jedoch den 11 Å langen Quervernetzer p-Azidophenacylphosphorothioat (Yang and Nash, 1994) zwischen sich und können somit nicht als direkte Kontakte angesehen werden.



**Abbildung 5.31: GV-Konstrukt von PI-SceI**

Strukturelle Überlagerung von 1gpp (rot) und 1vde (Proteinkette A, blau/violett) zur Entscheidung, welche Reste in das PI-SceI-GV-Konstrukt eingehen. Jeweils dargestellt die Rückgratatome und die gesamten (beschrifteten) Aminosäuren für die in 1gpp fehlende Schleife (links) und die Verbindung der Domänen I und II (rechts). Die jeweils letzten Aminosäuren, die zwischen 1gpp und 1vde strukturell übereinstimmen sind Lys53 und Phe73 bzw. Tyr179 und Gly418. Somit kommen die Reste Ser54 bis Lys72 (in violett im Konstrukt dargestellt) und Ala180 bis Arg417 von 1vde. Gly183, welches in 1gpp die Domäne II ersetzt, ist in grün dargestellt.

Um eine mögliche Domänen- oder Subdomänenbewegung zu untersuchen, wurde folgender Ansatz gewählt: Ein Modell von PI-SceI mit möglichst vielen definierten Aminosäuren wurde durch strukturelle Überlagerung von 1gpp mit 1vde konstruiert. Die 1.35-Å-Struktur 1gpp erlaubt den bei weitem detailliertesten Blick auf Domäne I. Die Überlagerung ist mit einer r.m.s.-Abweichung von 2.0 Å (C $\alpha$ -Atome) bzw. 2.6 Å (alle Atome) möglich (berechnet mit LSQKAB

(Collaborative Computational Project, 1994)). Die fehlende Schleife sowie Domäne II wurden von 1vde übernommen. Die jeweils letzten strukturell übereinstimmenden Aminosäuren sowie der Rest von Domäne I entspricht der hier präsentierten Struktur 1gpp Abbildung 5.31. Das PI-SceI-Konstrukt („GV-Konstrukt“) enthält somit die Reste 1 bis 53, 73 bis 179 und 418 bis 453 von 1gpp sowie 54 bis 72 und 180 bis 417 von 1vde. Lediglich die Bereiche 271 bis 279 und 369 bis 374, fehlende Schleifen in 1vde Domäne II) sind überhaupt nicht enthalten.

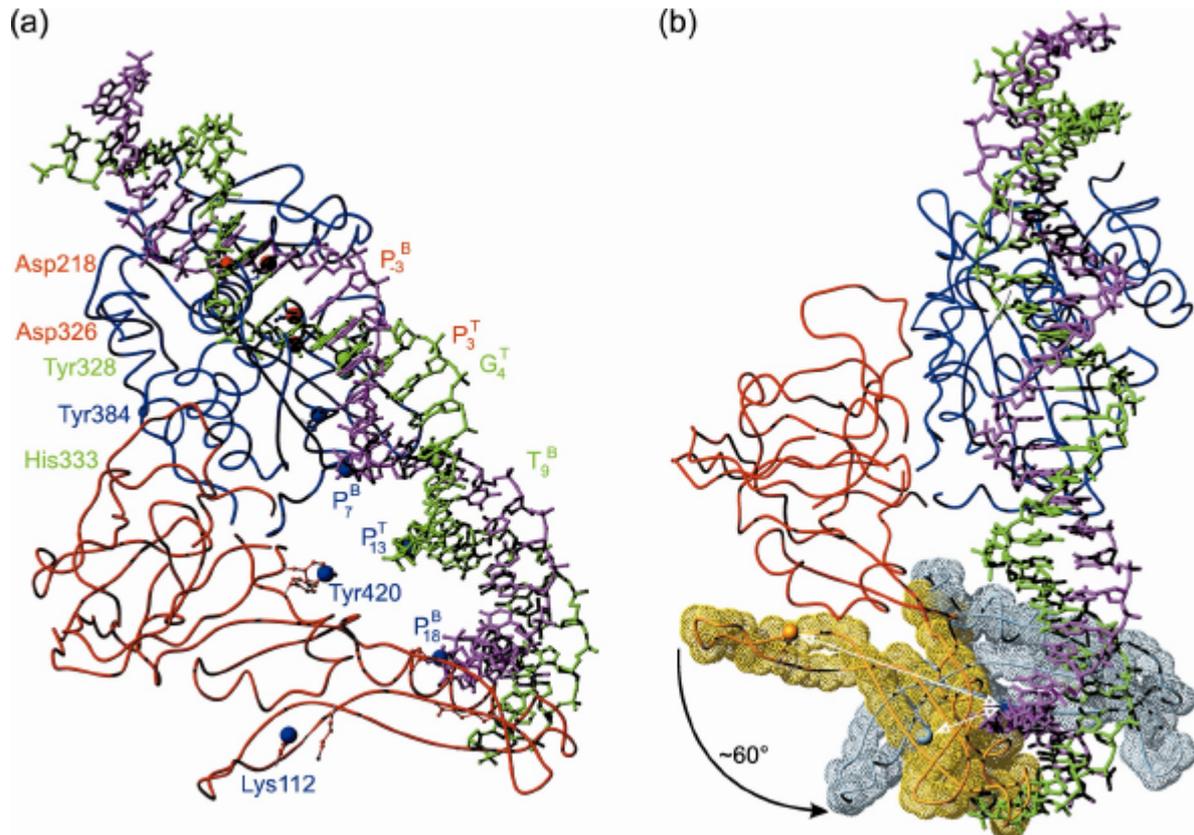
Des Weiteren wurde ein Modell der 31-bp-Erkennungssequenz von Basenpaar A:T<sup>-10</sup> bis T:A<sup>21</sup> erstellt, das eine Biegung von etwa 75° enthält. Dabei wurde ein regelmäßig gebogenes poly-A:T-DNA Model von Carter (Carter und Tung, 1996) benutzt (die Koordinaten „curve.pdb“), dessen Basen manuell ausgetauscht wurden. Die DNA ist somit komplett regelmäßig. Sie enthält kein Supercoiling und auch keine Verdrehung der Basen relativ zur Basenebene (propeller twist). Das 31-bp-Fragment hat eine Gesamt-Biegung von etwa 75°.

Die Atome mit den nachgewiesenen Verbindungen wurden mit dem Programm LSQKAB (Collaborative Computational Project, 1994) überlagert. Es sind

Asp218-OD2	und	P <sup>-3</sup> (unterer Strang),
Asp326-OD1	und	P <sup>+3</sup> (oberer Strang),
Tyr328-CZ	und	G <sup>+4</sup> -N7 (unterer Strang) sowie
His333-CE1	und	T <sup>+9</sup> -C5 (oberer Strang).

Hierzu ist anzumerken, dass zwar Asp218 am Schnitt des oberen Stranges und Asp326 am Schnitt des unteren Stranges beteiligt ist, die nächsten Nachbarn jedoch die hier angegebenen sind. Dies kann durch eine Überlagerung der aktiven Zentren von PI-SceI und I-CreI belegt werden (siehe Kapitel 5.1.3, Seite 67 und Abbildung 5.2, Seite 68). Die Photo-Quervernetzungen mit Tyr328 und His333 kamen durch Ersatz der Basen Guanin und Tyrosin mit 5-Iodo-Uracil (5-IdU) zustande (Christ *et al.*, 2000; Pingoud *et al.*, 1999). C5 von 5-IdU bildet nach UV-Bestrahlung eine kovalente Bindung mit einem ortho-Kohlenstoff (CE1 oder CE2) von Tyrosin (Norris *et al.*, 1996). Als Bezugspunkte wurden also CZ von Tyr328 (genau zwischen den CE-Atomen) und CE1 von His333 (entspricht etwa dieser Position) gewählt. Im Falle von Guanin ist die Position N7 diejenige, die der von C5 in Uridin am nächsten kommt. Diese jeweils vier Atome können mit einer r.m.s.-Abweichung von 2.3 Å (LSQKAB (Collaborative Computational Project, 1994) überlagert werden und ergeben so die Koordinaten eines Komplexes von PI-SceI (dem GV-Konstrukt) und der 75° gebogenen

Erkennungssequenz, ohne auf Überlappungen Rücksicht zu nehmen, Abbildung 5.32. Dieses Modell stimmt im Bereich der minimalen Bindungssequenz weder mit den bisher publizierten überein (Abbildung 5.7, Seite 72) noch mit den biochemischen Daten (Abbildung 5.8, Seite 74), die eine Bindung der DNA an der Vorderseite der zangenartigen Subdomäne nahe legt (mit der linken Seite der in Abbildung 5.8 gezeigten Orientierung). Hier befindet sich die DNA jedoch auf der „falschen“ Seite. Im Bereich der Domäne II kann das Modell mit den biochemischen Daten in etwa in Einklang gebracht werden. Für Domäne I wären jedoch deutliche konformationelle Änderungen nötig. Da sich die Aminosäuren Tyr420/Tyr176 jedoch immer noch in der geeigneten Entfernung für eine Quervernetzung mit p-Azidophenacylphosphorothioat befinden, wird hier eine Bewegung lediglich der zangenartigen Subdomäne vorgeschlagen, Abbildung 5.32b. Diese Bewegung würde in etwa  $60^\circ$  betragen und in der in Abbildung 5.32b gezeigten Orientierung leicht nach hinten gerichtet sein. Die Bewegung würde die Stränge **i**, **j** und **k** in Kontakt mit der großen Furche der Basenpaare 16 bis 18 bringen und Lys112 von einer Entfernung von  $35 \text{ \AA}$  zum Phosphat von C<sup>+18</sup> des unteren Stranges in Quervernetzungs-Entfernung. Da aufgrund der biochemischen Daten und der Oberflächenladungsverteilung (Abbildung 5.30, Seite 102) die Sekundärstrukturelemente Helices **D** und **E** und Strang **I** mit den Strängen **i**, **j** und **k** eine Einheit bilden, erscheint eine derartige Bewegung dieses gesamten DNA-Bindenden Bereiches wahrscheinlich (wird aus Gründen der Übersichtlichkeit jedoch nicht dargestellt). Aber auch die Bewegung einzelner Teile dieser Subdomäne erscheint möglich. Die Wechselwirkungen, die auf dieser Seite durch die Bewegung entfallen, könnten zum einen durch die Bindung der DNA, zum anderen durch mögliche neue Kontakte der Helix **C** mit den Resten 409 bis 413 sowie der Schleife 150 bis 152 mit den Aminosäuren 30/31 und 439 bis 441 energetisch kompensiert werden. Mögliche Kontakte zwischen Protein und DNA sind diejenigen von Arg90 mit den Nukleotiden A<sup>14</sup>/T<sup>15</sup> des oberen Stranges, von Arg94 mit benachbartem Nukleotid T<sup>16</sup>. Die Lysine 170 und 173 würden einen Kontakt mit der DNA im Bereich des oberen Stranges, Nukleotide 10 bis 12 eingehen.



**Abbildung 5.32: Geometriebasiertes Docking**

Modell eines Komplexes von PI-SceI (GV-Konstrukt) mit der 75° gebogenen Erkennungssequenz nach struktureller Überlagerung der vier nachgewiesenen Kontakte von Protein und DNA. Gezeigt sind die Bänder-Darstellung der Proteinrückgrate und die DNA in der Kugelstab-Darstellung. (a) Vorderansicht mit der PI-SceI Domäne I in rot, der Domäne II in blau, des oberen DNA-Stranges in grün und des unteren Stranges in violett. Große Kugeln repräsentieren die Kontaktatome, rot für die an der endonukleolytischen Reaktion beteiligten Atome, grün für die Quervernetzungen ohne zusätzlichen Linker, blau für die Quervernetzungen mit dem 11 Å Linker. (b) „Gelenk-Ansicht“ mit der zangenartigen Subdomäne (Reste 86 bis 154) vor der Oberflächendarstellung des Proteinrückgrates. In gelb die ursprüngliche Orientierung in 1gpp, in hellblau diese Subdomäne nach der vorgeschlagenen Bewegung in Richtung der großen Furche der DNA. Die großen Kugeln stellen die NZ-Atome von Lys112 dar, die durch die Bewegung in Quervernetzungs-Entfernung zum Phosphat von C<sup>+18</sup> des unteren Stranges kommen würde.

Ein anderer Weg, das Modell an die biochemischen Daten anzupassen, wäre eine zusätzliche Verdrillung der DNA im Mittelbereich, die die minimale Bindungssequenz in Kontakt mit der Vorderseite der zangenartigen Subdomäne bringt. Bei den oben angestellten Überlegungen muss außerdem noch berücksichtigt werden, dass mögliche Bewegungen einzelner Aminosäure-Seitenketten (zwischen 2.4 Å für Asp und 6.0 Å für Arg, gemessen von C $\beta$  bis zur Spitze) oder des Quervernetzlers p-Azidophenacylphosphorothioat (11 Å) ihre Fehlerintervalle von vorneherein schon mitbringen. Dennoch liefert das Modell

eine gewisse Plausibilität für eine Bewegung der Domänen/Subdomänen relativ zueinander, stellt aber natürlich keinen Beweis dar. Weitere strukturelle Experimente werden nötig sein, um die Bindung der DNA an PI-SceI im Detail zu verstehen. Bis heute sind keine Komplex-Kristalle verfügbar, weder für PI-SceI, noch für vergleichbare Proteine. Der DNA-Komplex von I-CreI (Jurica *et al.*, 1998) deckt lediglich Domäne II ab, ist hier jedoch ein gutes Modell (Christ *et al.*, 1999). Einen Hinweis für die DNA-Bindung von PI-SceI Domäne I liefert er jedoch nicht. Die andere Homing Endonuklease, für die ein DNA-Komplex bekannt ist, I-PpoI (Flick *et al.*, 1998) ist nicht strukturell verwandt mit PI-SceI. Zwar kann man optisch eine Ähnlichkeit mit einem Bereich von drei  $\beta$ -Strängen ausmachen, die bei I-PpoI in die große Furche „greifen“, eine strukturelle Überlagerung mit der zangenartigen Subdomäne von PI-SceI Domäne I hält diese Ähnlichkeit jedoch nicht stand.