

1. Einleitung

Proteine gehören neben Kohlenhydraten, Nukleinsäuren und Lipiden zu den zentralen Molekülen des Lebens. Sie haben bei fast allen Lebensvorgängen eine herausragende Bedeutung. So sind Proteine an der enzymatischen Katalyse, bei Transport und Speicherung anderer Moleküle, Immunabwehr, Bewegung und Stützfunktion, Weiterleitung von Signalen und bei der Kontrolle von Wachstum und Differenzierung beteiligt. Proteine sind aus einem Satz von 20 verschiedenen α -L-Aminosäuren aufgebaut. Die Zusammensetzung (Sequenz) und Länge des Proteins ist in den Genen codiert. Doch die Kenntnis der Sequenz reicht alleine nicht aus, um die Funktion eines Proteins vollständig zu verstehen. Erst die Kenntnis der dreidimensionalen Struktur des Proteins in Verbindung mit biochemischen Daten vermag ein vollständiges Bild über Wirkungsmechanismen und Spezifität zu geben. Daraus ergibt sich, daß die vollständige Kenntnis der Sequenz eines Genoms nicht ausreichend ist, die Lebensvorgänge zu erklären. Erst die Übersetzung der Informationen eines Genoms in Proteine und deren Analyse, sowohl funktionell als auch strukturell, vermag erste Erkenntnisse zu liefern. Es bedarf dann weiterer Analysen, um das komplexe Zusammenspiel mit anderen Proteinen und oder Molekülen, wie Cofaktoren, Nukleinsäuren usw., zu verstehen.

Der Zusammenhang von Funktion und Struktur ist von besonderem Interesse beim Design von Proteinen mit neuartigen Eigenschaften. Besonders bei medizinischen oder biotechnologischen Fragestellungen kommt dem Design von Proteinen eine immer größere Bedeutung zu. So kann durch genaue Untersuchung eines aktiven Bereichs oder einer Bindungsstelle versucht werden, die Aktivität oder Affinität eines Proteins durch zielgerichtete Mutation zu steigern. Weiterhin ist es prinzipiell möglich, einen Bindungspartner, der zum Beispiel einen Rezeptor effektiv(er) aktiviert oder blockiert, *de novo* zu entwerfen oder einen bekannten zu verändern. Von großem Interesse ist auch die Proteinstabilität. Besonders für biotechnologische Anwendungen kann die Steigerung der Thermostabilität eines Enzyms die Effektivität eines Prozesses erhöhen oder den Weg für weitere Anwendungen frei machen.

In der Natur beobachtet man Proteine mit ähnlichen Strukturen, die unterschiedliche Funktionen haben, als auch Proteine mit unterschiedlichen Strukturen gleicher oder ähnlicher Funktion. Des weiteren findet man Strukturen homologer Proteine, die sich nur gering in der Sequenz unterscheiden, aber erheblich in der Thermostabilität differieren.

Um die molekularen Ursachen solcher Phänomene aufzuklären, bedient man sich der Strukturanalyse.

Um Strukturen zu erhalten, werden im wesentlichen drei Methoden verwendet. Die älteste und universellste ist die Röntgenkristallstrukturanalyse. Für diese Methode ist das Vorhandensein von Kristallen der einschränkende Fakt. Auch für Neutronenbeugung konnte gezeigt werden,

1. Einleitung

daß eine Strukturbestimmung möglich ist. Die Strukturanalyse basiert auf dem gleichen Regelwerk, wie bei der Röntgenbeugung. Bei der Neutronenbeugung ist die Bestimmung von Wasserstoffpositionen experimentell möglich. Die Neutronenstrahlung tritt im Unterschied zur Röntgenstrahlung mit den Atomkernen in Wechselwirkung, und nicht wie bei der Röntgenstrahlung mit der Elektronenhülle. Die experimentelle Bestimmung der Wasserstoffpositionen ist für Röntgenbeugungsexperimente nur mit atomar aufgelösten Strukturen (besser als 1 Å) möglich. Neutronenstrahlung ist jedoch nur eingeschränkt verfügbar. Die zweite bedeutende Methode der Proteinstrukturanalyse ist die NMR-Spektroskopie. In den letzten Jahren konnte sie in ihrer Leistungsfähigkeit enorme Fortschritte erzielen. Sie ist aber hinsichtlich der Größe der untersuchten Moleküle beschränkt. Zwar ist die Elektronenmikroskopie als dritte Methode hinsichtlich ihrer Auflösung begrenzt, jedoch liefert sie bei großen Strukturen wie Multiproteinkomplexen Einblicke in die Anordnung der Proteine zueinander bis hin zum Erkennen möglicher Sekundärstrukturelemente.

1.1. Das Kälteschockprotein (cold shock protein - Csp)

Kälteschockproteine (Csps) findet man in vielen grampositiven und gramnegativen Bakterien in Familien mit drei (*B. subtilis*) bis zu neun (*E. coli*) Mitgliedern hoher Sequenzübereinstimmung zueinander. Die Sequenzidentität liegt bei 70% [Graumann & Marahiel 1998, Schindler et al., 1998]. Die weite Verbreitung der Csps zeigt, daß dieses Protein schon am Beginn der bakteriellen Evolution vorhanden waren und eine evolutionär alte Klasse von Proteinen darstellen. Dem Csp homologe Domänen werden auch in einigen eukaryotischen nukleinsäurebindenden Proteinen gefunden [Sommerville & Ladomery 1996], wo sie spezifisch an RNA binden [Matsumoto et al., 1996] oder verantwortlich für die DNA-Bindung an die Y-Box [Schnuchel et al., 1993] sind. Y-Box bindende Proteine agieren als Regulatoren der Transkription und Translation. Csps binden an einzelsträngige und doppelsträngige DNA sowie kooperativ mit geringer Spezifität an RNA [Jiang et al., 1997; Graumann et al., 1997; Mayr et al., 1996].

Ursprünglich wurden Csps als eine Antwort auf einen Kälteschock in *E. coli* entdeckt [Jones et al., 1987]. Da Csps bei einem Kälteschock stark induziert werden, wurde angenommen, daß sie für die Adaption an geringe Temperaturen notwendig sind. Für CspA von *E. coli* konnte in der Tat gezeigt werden, daß es die Synthese von einigen durch Kältestress induzierbaren Proteinen steigert, wenn die Temperatur abgesenkt wird. Verschiedene Mitglieder der Csp-Familien werden unterschiedlich reguliert und scheinen unterschiedliche Funktionen wahrzunehmen, sowohl bei der Zellteilung als auch in der stationären Phase. Für Csps von *B. subtilis* konnte gezeigt werden, daß sie sowohl bei niedrigen als auch bei optimalen Temperaturen essentiell für die Proteinsynthese sind [Graumann et al., 1997]. Für *Bs*-CspB wurden Modelle, die eine

1. Einleitung

Bindung an DNA und mRNA bei tiefen Temperaturen mit einhergehender Transkriptions- und Translationsförderung vorschlagen, diskutiert [Graumann & Marahiel, 1994]. Von CspA aus *E. coli* ist weiterhin bekannt, daß es *in vitro* die Sekundär- und Tertiärstruktur von RNA destabilisiert. Das führte zu der Annahme, daß Csp's die Initiation der Translation erleichtern und als eine Art "RNA-Chaperon" wirken, indem sie die Bildung stabiler nichtproduktiver Sekundärstrukturen der mRNA verhindern [Jiang et al., 1997].

Csp's sind kleine (~70 Reste) kompakte Proteine. Die dreidimensionalen Strukturen der Csp's aus *B. Subtilis* (*Bs-CspB*), *E. coli* (CspA) und aus *B. caldolyticus* (*Bc-Csp*) sind aus NMR-Studien [Schnuchel et al., 1993; Newkirk et al., 1994] und Kristallstrukturanalysen [Schindelin et al., 1994, Schindelin et al., 1993; Mueller et al., 2000] bekannt. Alle diese Strukturen sind sich sehr ähnlich. Csp's besitzen eine einfache Faltung. Sie ist aus fünf β -Strängen, die durch vier Schleifen miteinander verknüpft sind, aufgebaut (siehe Abb. 1). Die Stränge lagern sich zu einem antiparallelen Faltblatt, das die Form eines Fasses hat, zusammen, allgemein als " β -barrel" bezeichnet (Abb.1). Die Struktur läßt sich weiter in zwei Subdomänen einteilen, die je ein dreisträngiges antiparalleles Faltblatt aus den Strängen $\beta 1$ bis $\beta 3$ sowie $\beta 4$, $\beta 5$ und $\beta 1$

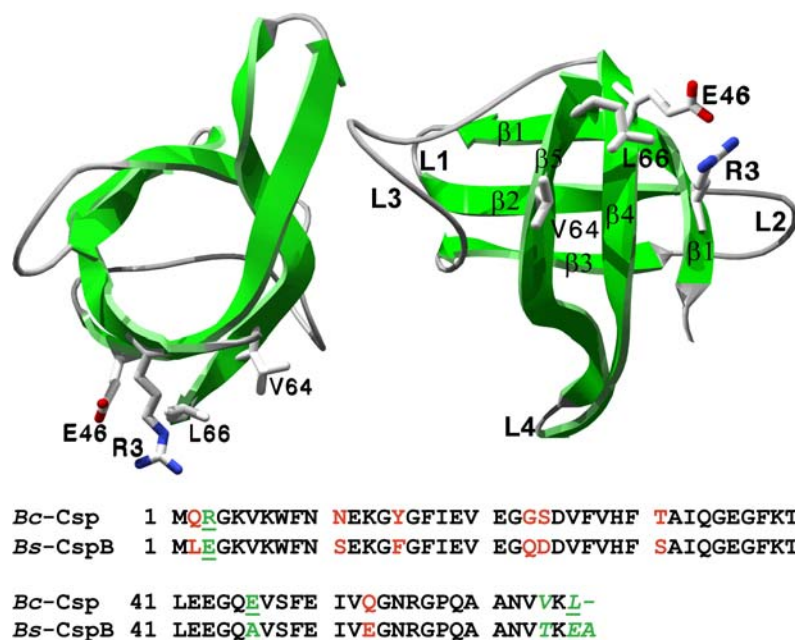


Abb. 1) Dreidimensionale Struktur von *Bc-Csp*. Die schematische Darstellung zeigt die Struktur von *Bc-Csp* in zwei Orientierungen. Die im Rahmen dieser Arbeit kristallographisch untersuchten Kälteschockproteine mit ortsgerichteter Mutagenese basieren auf den gekennzeichneten Resten. Sie liegen dicht benachbart auf der Oberfläche des Proteins. Im unteren Teil wird die Sequenz von *Bc-Csp* und *Bs-CspB* miteinander verglichen. Rot markierte Reste haben geringen Einfluß auf die Thermostabilität. Grün markierte Reste sind Gegenstand der Kristallstrukturen. Die unterstrichenen Reste sind als Einzel- und als Trippelmutante, die kursiv gesetzten als Trippelmutante vorhanden.

1. Einleitung

bilden. $\beta 1$ wird von beiden Faltblättern geteilt. Die beiden Faltblätter sind nahezu 90° zueinander angeordnet. Topologisch wird diese Supersekundärstruktur auch als "Greek key"-Motiv bezeichnet. Die Struktur der Kälteschockproteine gehört zu dem sogenannten "OB-Faltungstyp" (oligosaccharide/oligonucleotide binding fold) [Chothia & Murzin 1993, Murzin 1993]. Der "OB-fold" stellt ein ubiquitäres Faltungsmotiv dar, wie es von einigen nukleinsäurebindenden Proteinen bekannt ist. Er wird in folgenden Strukturen gefunden: der Antikodonbindungsdomäne der Aspartyl-tRNA-Synthetase im Komplex mit tRNA [Ruff et al., 1991], der Staphylokokkennuklease [Hynes & Fox, 1991] und dem Telomereenden bindenden Protein aus *Oxytricha nova* (OnTEBP)[Horvath et al., 1998].

Die Sequenzunterschiede zwischen den Csps sind gering, führen aber zu bemerkenswerten Unterschieden. So differieren *Bc*-Csp und *Bs*-CspB an 12 von insgesamt 67 Positionen bezogen auf *Bs*-CspB (Abb. 1), was zu signifikanten Unterschieden in der Thermostabilität führt. Ausgedrückt durch die Schmelztemperatur findet man für *Bs*-CspB einen Wert von $53,6^\circ\text{C}$ und für *Bc*-Csp von $76,9^\circ\text{C}$, wobei die Schmelztemperatur als Mittelpunkt der thermischen Entfaltung definiert ist. Die freie Energie der Entfaltung (ΔG_D) ist um $15,8 \text{ kJ}\cdot\text{mol}^{-1}$ erhöht [Mueller et al., 2000; Perl et al., 2000]. Diese Tatsache reflektiert auch die unterschiedliche Herkunft der Organismen. So kommt *Bs*-CspB aus einem mesophilen Organismus und *Bc*-Csp aus einem thermophilen Organismus. Es ist bekannt, daß sich Mikroorganismen an ein breites Spektrum verschiedener Umweltbedingungen anpassen können. So leben thermophile Organismen bei Temperaturen, bei denen die meisten Proteine von mesophilen Organismen bei *in vitro* Experimenten entfalten würden. Augenscheinlich hat die Evolution Organismen hervorgebracht, die ihre Proteinmoleküle so stabilisieren können, daß sie bei erhöhten Temperaturen nicht entfalten und funktionell aktiv sind.

Um die molekularen strukturellen Ursachen der Unterschiede in der Thermostabilität von Kälteschockproteinen zu untersuchen, wurde ein Paar von Csps aus dem mesophilen *B. subtilis* und dem thermophilen *B. caldolyticus* als Modell gewählt. Beide Proteine unterscheiden sich an nur zwölf Positionen und sind einer systematischen Mutationsanalyse zugänglich. Damit sollte es möglich sein, die einzelnen additiven Beiträge jeder Mutation zur freien Energie nach Gibbs zu bestimmen. Oft unterscheiden sich Proteine aus thermophilen Organismen von ihrem mesophilen Homolog an vielen Positionen. Dadurch ist es über Mutageneseansätze nicht möglich, die einzelnen Energiebeiträge, die für eine Stabilisierung verantwortlich sind, zu identifizieren [Jaenicke & Böhm, 1998]. Die kleinen Csps mit ihrer einfachen Struktur eignen sich besonders gut als Modellsystem für Faltungs- und Stabilitätsstudien. In der Struktur werden keine *cis*-Proline gefunden, und sie enthalten keine fest gebundenen Kofaktoren und Disulfidbrücken. Csps liegen in Lösung als Monomere vor [Perl et al. 1998]. Die beiden Csps denaturieren reversibel in einer monomolekularen Zweizustandsreaktion $N \leftrightarrow U$ und sind somit auch einer quantitativen thermodynamischen Studie zugänglich.

1. Einleitung

→

Bei der Mutation eines jeden Restes, ausgehend von *Bc*-Csp, in den korrespondierenden Rest von *Bs*-CspB wurde gezeigt [Perl et al., 2000], daß sich der Unterschied in der Thermostabilität im wesentlichen auf zwei Reste - Arg3 und Leu66 - zurückführen läßt. Die Summe beider Beiträge ergibt nahezu den gesamten Beitrag der Destabilisierung. Die beiden Reste, Arg3 in der N-terminalen Region und Leu66 direkt am C-Terminus, sind zwar in der Sequenz weit auseinander, aber in der Struktur eng benachbart und zum Lösungsmittel hin exponiert. Allein die Mutation Arg3 → Glu bringt einen Beitrag von rund 70% ($11,5 \text{ kJ}\cdot\text{mol}^{-1}$) in $\Delta\Delta G_D$ zwischen beiden Proteinen ein. Die Mutation Leu66 → Glu trägt mit einem Beitrag von $5,2 \text{ kJ}\cdot\text{mol}^{-1}$ bei. Umgekehrt läßt sich *Bs*-CspB durch die Mutationen Glu3 → Arg und Glu66 → Leu in ein thermostabiles Protein umwandeln. Der jeweilige Beitrag $\Delta\Delta G_D$ der Stabilisierung ist fast gleich wie der der Destabilisierung. Dies scheint auf eine relative lokale Unabhängigkeit der Mutation hinzudeuten, da die räumliche Umgebung in beiden Proteinen unterschiedlich ist. Die anderen zehn Differenzen tragen nur marginal zum Unterschied in der Thermostabilität bei. Der Hauptanteil der erhöhten Thermostabilität von *Bc*-Csp ist elektrostatischen Ursprungs [Perl et al., 2000]. Tatsächlich stabilisieren coulombsche Wechselwirkungen das thermophile *Bc*-Csp, aber destabilisieren das mesophile *Bs*-CspB. So wird bei Abschwächen dieser Wechselwirkungen durch Zugabe von 0,5 M Salz (NaCl) ΔG_D für *Bc*-Csp gesenkt und für *Bs*-CspB erhöht.

Oft wird angenommen, daß die Optimierung der ionischen Wechselwirkungen signifikant zur erhöhten Stabilität thermophiler und hyperthermophiler Proteine beiträgt [Grimsley et al., 1999; Karshikoff & Ladenstein, 1998; Loladze et al., 1999; Pace et al., 2000; Perutz & Raidt, 1975; Spector et al., 2000]. Es bleibt aber unklar, ob diese Beiträge von Ionenpaaren, von einem ausgedehnten Ionennetzwerk, von generellen Effekten, wie Änderungen in der Nettoladung des Proteins oder anderweitigen Ursprungs sind.

Um die strukturellen Determinanten der Thermostabilität auf atomarer Ebene zu studieren, sollen von den Schlüsselmutanten *Bc*-Csp R3E und *Bc*-Csp L66E die Strukturen gelöst werden. Zusätzlich ist die Struktur der Mutanten *Bc*-Csp E46A von Interesse, um potentielle Wechselwirkungen der Position 46 mit den Positionen 3 und 66 zu untersuchen, wie sie anhand der Struktur von *Bc*-Csp zu vermuten sind. Ebenfalls wurde Material von den Mutanten basierend auf *Bs*-CspB zur Verfügung gestellt. Sie stellen die entsprechende Mutationen von *Bs*-CspB kommend dar (*Bs*-CspB E3R, *Bs*-CspB A46E sowie *Bs*-CspB E66L). Um weitere Erkenntnisse über den Ursprung der Stabilisierung an Position 3 zu bekommen, wurde auch Material der Mutante *Bs*-CspB E3L bereitgestellt. Ferner wurde die Tripelmutante *Bc*-Csp R3E/E46A/L66E strukturell charakterisiert. Sie ist an den mutierten Positionen equivalent zu *Bs*-CspB besetzt und hat ähnliche thermodynamische Eigenschaften. Eine Wechselwirkung des C-Terminus mit dem Rest an Position 3 läßt sich nicht ausschließen. Zusätzlich steht die Frage,

1. Einleitung

ob der verlängerte C-Terminus in *Bs*-Csp einen Einfluß auf die Thermostabilität hat. In diesem Fall wird die Ladung des C-Terminus, der carboxyterminale Sauerstoff, verschoben. Weiterhin ist der starke hydrophobe Charakter des C-Terminus in *Bc*-Csp auffällig. Diese Einflüsse sollen anhand der Tripelmutante *Bc*-Csp V64T/L66E/67A studiert werden.

Die Untersuchung der molekularen Ursachen der Thermostabilität ist nicht nur von Interesse bezüglich der thermischen Adaption, sondern auch im Hinblick auf die Möglichkeiten des Designs von thermostabilen Proteinen in der Biotechnologie. Fundierte Kenntnisse über den Zusammenhang von Struktur und Stabilität sind dabei von größtem Interesse und Voraussetzung für ein effektives Design thermostabiler Proteine.

1.2. Das Repressorprotein KorB aus RP4 infizierten *Escherichia coli*

KorB wird in der zentralen Kontrollregion (ccr) des Plasmides RP4 codiert. Das Plasmid RP4 ist ein Mitglied der *E. coli* Inkompatibilitätsgruppe P (IncP-1 α) und ist nicht unterscheidbar von RP1, R68, R18 und RK2. Das IncP-1 α Plasmid RP4/RP1/RK2 war das erste, das im Detail untersucht worden ist [Pansegrau et al., 1994]. Das RP4-Plasmid ist ein sich selbst übertragendes Resistenzplasmid von ca 60 kb. IncP-1 α Plasmide sind wegen ihrer Fähigkeit, sich in fast allen gramnegativen Bakterien selbst zu erhalten und zu transferieren, von besonderem Interesse [Thomas & Smith, 1986]. Eine Hauptrolle beim Überleben spielen die globalen Plasmidregulatoren, die den Transfer, die Replikation und die Partitionsfunktion kontrollieren und koordinieren [Jagura-Burdzy & Thomas, 1994, 1997; Motallebi-Veshareh et al., 1992; Zatyka et al., 1994, 1997]. Diese Regulatoren sind KorA, KorB, TrbA and KorC.

Im Zusammenspiel mit anderen Proteinen spielt KorB eine direkte Rolle bei der Partitionierung des Plasmids, fungiert aber auch als globaler Regulator der Expression von RP4 Genen. KorB gehört zur Familie der ParB-Proteine, die in die Genompartitionierung involviert und auf Plasmide und bakteriellen Chromosomen codiert sind [Lobočka & Yarmolinsky, 1996; Motallebi-Veshareh et al., 1990; Williams & Thomas, 1992; Williams et al., 1998].

In Verbindung mit KorA inhibiert KorB die Transkription der *kilA*, *trfA* und *korAB* Operons. Sie sind involviert in die negative Kontrolle des *kilB* Operons. Die Wirkung von KorA und KorB ist stark kooperativ.

Gereinigtes KorB existiert als Dimer und als höheres Multimer in Lösung. Quervernetzungsstudien deuten auf ein Vorwiegen der dimeren Form hin [Balzer et al., 1992; Williams et al., 1993]. KorB hat eine Länge von 358 Aminosäuren. Daraus resultiert eine berechnete Masse von 39011 Da. Die Polypeptidkette hat eine vorhergesagte Netzladung von 21 negativen Ladungen bei einem neutralen pH und einem berechneten pI von 4,76. Dies ist ein ungewöhnlicher Wert für DNA-bindende Proteine, die oft über einen Überschuß an positiven

1. Einleitung

Ladungen verfügen. Ungeachtet dessen bindet KorB spezifisch an die symmetrische Operatorsequenz OB 5'TTTAGC^G/cGCTAAA3', die 12 mal in dem RP4-Plasmid vorhanden ist. Die Bindung von KorB an DNA wird möglicherweise über ein Helix-Turn-Helix Motiv vermittelt, dessen Lokalisation auf Sequenzvergleichen beruht [Balzer et al., 1992].

Die 12 Operatorsequenzen werden in drei generellen Positionen relativ zur Lage von RP4-Promotoren gefunden. Gemessen vom Mittelpunkt der Operatorsequenz wird 39/40 bp aufwärts vom Transkriptionsstartpunkt die Klasse I, weiter auf- oder abwärts im Bereich von 80 bis 190 bp des Transkriptionsstartpunkts die Klasse II und in einem Abstand von mehr als 1 kB von einem bekannten Promotor eine Klasse III [Jagura-Burdzy et al., 1999b] gefunden. KorB kann die RP4-Promotoren der Klasse I und II unterdrücken [Jagura-Burdzy et al., 1999b; Jagura-Burdzy & Thomas, 1997; Macartney et al., 1997; Motallebi-Veshareh et al., 1992; Thomson et al., 1993], aber die Rolle in den Regionen, wo Operatoren der Klasse III vorkommen, ist nicht klar. Die Konservierung der Positionen in IncP-1 RP4 und dem Plasmid R751 [Thorsted et al., 1998] deutet an, daß eben diese Klasse III Operatoren eine wichtige Rolle bei der Expression des Genoms spielen.

Detaillierte Analysen der Bindung von KorB an die zwölf Operatoren zeigen [Kostelidou et al., 2000], daß sie bezüglich ihrer Bindungsstärke in drei Gruppen A, B und C eingeteilt werden können. Die O_B10 mit der höchsten Aktivität kommt in der Promotorregion der Gene für Replikation, trfAp, vor. Gereinigtes IncC1 verstärkt die Affinität von KorB zu allen OBs mit Ausnahme von O_B3. Diese O_B3 ist in die Plasmidpartitionierung involviert. IncC1 scheint ohne selbst direkt an DNA zu binden, den KorB-DNA Komplex zu stabilisieren. Fünf flankierenden Basenpaare des O_Bs beeinflussen die KorB-Bindungsaktivität als auch die stabilisierende Wirkung von IncC1. Untersuchungen mit Hybridoperatoren zeigen, daß flankierende Basen auf nur einer Seite für den Unterschied der Eigenschaften von O_B3 und O_B10 verantwortlich sind. Diese sind auch für die Wirkung von IncC1 verantwortlich. Es ist möglich, daß die Kontakte zwischen KorB mit den flankierenden Sequenzen als auch der Kontakt mit IncC1 die Konformation von multimeren KorB ändert, so daß es für KorB günstige Kontakte zur DNA geben kann.

Um die strukturelle Basis der Bindungseigenschaften zu verstehen, ist es das Ziel, das Zusammenspiel von KorB und DNA strukturell zu charakterisieren, und die Hypothese des "Helix-Turn-Helix" Motivs zu verifizieren. Ebenfalls von Interesse ist die Aufklärung der Domänenstruktur von KorB, die Rückschlüsse über mögliche Interaktionsstellen zuläßt und die sowohl auf Homodimere und -multimere als auch auf Interaktionen mit anderen Proteinen hindeutet.