Max-Planck-Institut für Biochemie
Abteilung Strukturforschung

Structure and function of the restriction endonucleases *Bse*634I and *Mun*I

Saulius Gražulis

Vollständiger Abdruck der von der für Fakultät Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. O. Nuyken
Prüfer der Dissertation:	
	apl. Prof. Dr. Dr.h.c. R. Huber
	Univ.-Prof. Dr. W. Hiller

Die Dissertation wurde am 30.10.2000 bei der Technischer Universität München eingereicht und durch die Fakultät für Chemie am 24.11.2000 angenommen.

Mano mielosioms Mamytei ir Senelei,
ir mano brangaus Tėvelio atminimui.

Acknowledgments

I am especially grateful to Prof. Dr. Robert Huber for the possibility to carry out this work in Max-Planck Institut für Biochemie, for his attention, encouragement and financial support.

The original idea of this work belongs to Dr. Virginijus Siksnys, and his crystallisation experiments of *Mun*I were the starting point of the whole project. I thank him for the invitation to join the restriction endonuclease structure team, and for the great help in all stages of the project.

I thank Dr. G. Burenkov and Dr. H. Bartunik for invaluable help with data collection at synchrotron DESY, DORIS/BW6 beamline. I am grateful to Dr. V. Repin for providing of the *Bacillus stearothermophilus* strain for our project and Dr. R. Skirgaila for the purification of the protein. Big thanks to Dr. I. Gutsche and E. Manakova for careful reading of the manuscript and numerous creative ideas and suggestions.

I thank very much Dr. Matthias Bochtler for the very interesting discussions of crystallography, physics and computer science during our stay in Martinsried.

Big thanks to Prof. Dr. Albrecht Messerschmitdt for the discussions of crystallography and computational methods, to Dr. Andreas Bergner and Dr. Rich Engh for the help with the computational work and data evaluation, and to Margit Bauer and Imgard Mayr for their help in the laboratory.

I am grateful to Dr. Markus Deibert for the interesting and fruitful collaboration in restriction endonuclease structure project. I also wish to thank Michael Blaesse for discussions and participation in the development of computational scripts used in this work.

Dr. Frank Schneider was an initiator of the design and maintenance house factorial set, which were successfully used in this project, as well as in many others.

This work has been supported by Volkswagen Stiftung and NATO linkage grant No. 960928.

Abbreviations

β-ME	β-Merkaptoethanol
AA	Acrylamide
ADP	Adenosindiphosphate
ATP	Adenosintriphosphate
AdoMet	S-Adenosylmethyonine
BisAA	bis-acrylamide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylendiamintetraacetic acid
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid
HPLC	High pressure liquid chromatography
IP	Image Plate
MALDI-MS	Matrix-assisted laser desorption-ionisation mass spectrometry
MES	2-(N-morpholino)-ethane sulphonic acid
MIR	Multiple isomorphous replacement
MOPS	3-(N-morpholino)-propane sulphonic acid
MW	Molecular weight
NCS	Non-crystallographic symmetry
PAAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PEG	Polyethylenglycol
PEG 400	Polyethylenglycol, average mol. weight 400 Da
PEG 8K	Polyethylenglycol, average mol. weight 8000 Da
RMS	Root-mean-square
RMSD	Root-mean-square deviation
RF	Reversed-phase (chromatography)
SANPAH	*N*-Succinimidyl 6-[4'-azido-2'-nitro-phenylamino] hexanoate
SIR	Single isomorphous replacement
Tris	Tris-oxymethyl aminomethane
UV	Ultra-violet (light)
WT	Wild-type
bp	base pair(s)
v/v	volume/volume
w/v	weight/volume

Contents

Abstract

The type II restriction endonucleases specifically recognise symmetric DNA sequence of 4 to 8 base pairs and cut the DNA strand within or close to their recognition site. Fidelity of these enzymes is remarkable: a change of just one base pair in a specific recognition site can reduce cleavage rates 10^5 to 10^9 times [99]. This property makes type II restriction endonucleases a useful tool for molecular biologists and an interesting model for studying DNA-protein interactions.

The basis of restriction endonuclease fidelity could not been understood from analysis of their sequences alone, because these sequences share in general little similarity. It was necessary to solve three dimensional structures of typical type II restriction enzymes to find out how they interact with DNA and to suggest hypotheses explaining recognition and catalysis mechanisms of these enzymes.

Among 11 restriction enzymes whose structures are known today we have several pairs of enzymes with recognition sequences differing only by the outer pair of nucleotides. Examples of such pairs are *Bam*HI and *Bgl*II or *Mun*I and *Eco*RI. One has expected that comparison of these pairs would give an answer how different nucleotides are discriminated. *Mun*I has been chosen for crystallographic and biochemical studies with this comparison in mind.

Surprisingly, the structures show that restriction enzymes can employ different strategies of DNA recognition even for their common inner base pairs [67]. Thus, a minimal difference of single base pair in recognition site may select enzymes with relatively large differences in their DNA recognition machinery.

We have therefore decided to investigate the variability of restriction enzymes with the same recognition sequence and chose a pair of isoshisomeric restriction endonucleases, *Bse*634I and *Cfr*10I, for our structural studies. Both enzymes recognise degenerate DNA sequence 5'-Pu$^{\downarrow}$CCGGPy, and cut as indicated by $^{\downarrow}$. Structure of *Cfr*10I is already known [15]; in this work we report the structure of *Bse*634I restriction enzyme from *Bacillus stearothermophilus*.

The two enzymes, *Bse*634I and *Cfr*10I, are the first pair of isoshisomeric restriction endonucleases where structures of both proteins in a DNA-free state are known. Comparison of the two structures shows how DNA recognition and catalysis are possibly coupled through the rearrangement of the mobile protein subdomains.

Chapter 1

Introduction

1.1 Restriction endonucleases

Certain microorganisms have capability to resist the attacks of phages by degrading the phage DNA with the help of endonucleases. To be able to distinguish between own DNA and that of the phage, these microorganisms modify own DNA by introducing methyl groups at certain positions; all unmodified DNA in such organisms is rapidly cleaved while modified molecules remain intact and can be replicated and transcribed as usual.

This phenomenon is called *restriction* (of a phage), and the microorganism that exhibits it is called *restricting host*. The nucleases involved in host protection are called *restriction endonucleases*, or shortly *restrictases*. They possess exclusively *endo*–activity, that is, cut the DNA strand in the middle of the strand, as opposed to *exo*–activity that consists in degrading of DNA strand from one end.

Host's own DNA is protected from restriction endonucleases by introduction of extra methyl group into adenosine or cytosine bases (giving 5'-methylcytosine, N^4-methylcytosine or N^6-methyladenosine). Enzymes that attach methyl groups to DNA are called *modification methyltransferases*, or *modification methylases*. Sometimes nuclease and methylase activities are possessed by one protein, but more often there are two distinct proteins – one with nuclease activity and one with methyltransferase activity. In both cases methylase and nuclease build a so called *restriction-modification* system of the host.

The restriction of phage was first observed in 1952 by Luria and Humman for T-even phages [68]. A year later, Bertani and Weigle observed the similar phenomenon with the P2 and λ-phages [9]. They showed that in *Escherichia coli* strain K12 λ-phages give $10^4 - 10^5$ times lower yields than in other *E. coli* strains [29, 5]. Phages that have been obtained after single passage through *E. coli* K strain, however, aquire capability to infect the *E. coli* K strain with nearly the same efficiency as the other non-restricting *E. coli* strains. The abil-

ity of λ-phages to grow efficiently on the restricting host is retained during the multiple passages through the restricting host, but disappears again after a passage through a non-restricting host [29]. It has been shown that the λ-phages get modified in a restricting strain and becomes indistinguishable from the host's "own" DNA. The non-restricting host lacking the restriction modification system does not modify the phage's DNA, and the phages from such host can no longer efficiently infect *E. coli* K.

It took, however, nine more years before Arber and Dussoix offered a molecular explanation of the phage restriction [5]. They proposed that there was a sequence–specific restriction endonuclease which cleaved that phage's DNA and in this way restricted its growth. Host's own DNA was protected from the same fate by some other enzyme that would modify the host's DNA and in this way make it distinguishable from that of the phage. After the passage through a restricting host, however, the surviving phage's DNA would also get modified, and the restriction of these phages would no longer take place. In 1967 Arber proposed that DNA modification was carried out by the sequence–specific methyltranspherase, and in 1972 this fact was proved by Kühnlein and Smith [62, 95].

The most remarkable property of the restriction endonucleases and the accompanying methylases is their extremely high site specificity. Each restriction-modification system recognises a single sequence of nucleotides, 4 to 8 base pairs in length and cuts DNA (or introduces the methyl group) in one strictly defined position within or close to their recognition sequence. The differences in the DNA cleavage rates of the restriction endonucleases between DNA possessing their cognate site and DNA without such site can reach $10^6 - 10^7$–fold. Such high fidelity is needed because, as pointed out by Jen-Jacobson [51], the DNA cleavage must happen in the presence of vast amounts of other, non-cognate DNA sequences, which are however very close chemically to the cognate sequence of the restriction endonuclease.

This high fidelity of the restriction endonucleases make them an invaluable tool for molecular biologists, who use them routinely to cut the genes out of plasmids, viruses or genomes, and for DNA mapping. The need to cut DNA at different positions stimulated intensive searches for new restriction-modification systems with different recognition sites. At the moment there are more that 3000 restriction-modification systems known, and the number is constantly growing.

Restriction endonucleases are also interesting objects to study DNA-protein interaction. The investigations of primary sequences of restriction enzymes have shown that they share, in general, little similarity, although the proteins catalyse the same reaction on a very similar substrates.

1.2 Classification of restriction-modification systems

On the basis of their cofactor requirements and substrate type restriction-modification systems can be divided into several types (see fig. 1.1). All restriction enzymes fall roughly into broad two groups — those that need ATP for DNA hydrolysis and those that do not. Originally, these two groups were denoted "type I" and "type II" restriction enzymes. The distinction between type I and type II restriction–modification systems was proposed in 1971 by Boyer [13]. As investigations continued, however, it became clear that "type I" group is not homogenous and should be subdivided further into smaller classes. In 1978 Kauc and Piekarovicz [54] showed that Boyer's type I restriction enzymes actually consists of two fundamentally different groups of type I and type III restriction endonucleases. According to his suggestion, the ATP dependent enzymes (original type I group) were named type I or type III depending on whether they hydrolyse ATP during DNA cleavage or just bind as a cofactor. The type II restriction enzymes retained their original name and contained all ATP-independent restriction-modification systems.

In addition to these three well-established types several groups of restriction enzymes have been discovered that have unusual properties which make them distinct from the bulk of other restriction endonucleases. Such enzymes are most often classified as belonging to the type with which they share most resemblance, and then grouping them in a subclass according to their distinctive properties.

In such fast changing field like restriction endonuclease research it is difficult to keep up to date for any paper printed database or catalog. Electronic databases are a great help in such cases. Probably the best known database of restriction endonucleases is REBASE, compiled by Roberts & Macelis [87]. Their database is updated daily and is available on the World Wide Web under URL `http://www.neb.com/rebase`.

1.3 ATP-dependent restriction endonucleases

1.3.1 Type I restriction-modification systems

Type I restriction enzymes cut DNA only in the presence of ATP and hydrolyse ATP in the process of DNA cleavage. They also require absolutely AdoMet[1] for both DNA hydrolysis (as an allosteric activator) and for methylation (as an allosteric cofactor and methyl group donor).

Each Type I restriction-modification system is encoded in three genes *hsdR*, *hsdM* and *hsdS*. The respective products of these genes must associate to build

[1] S-Adenosylmethyonine, see abbreviation list on p. 2

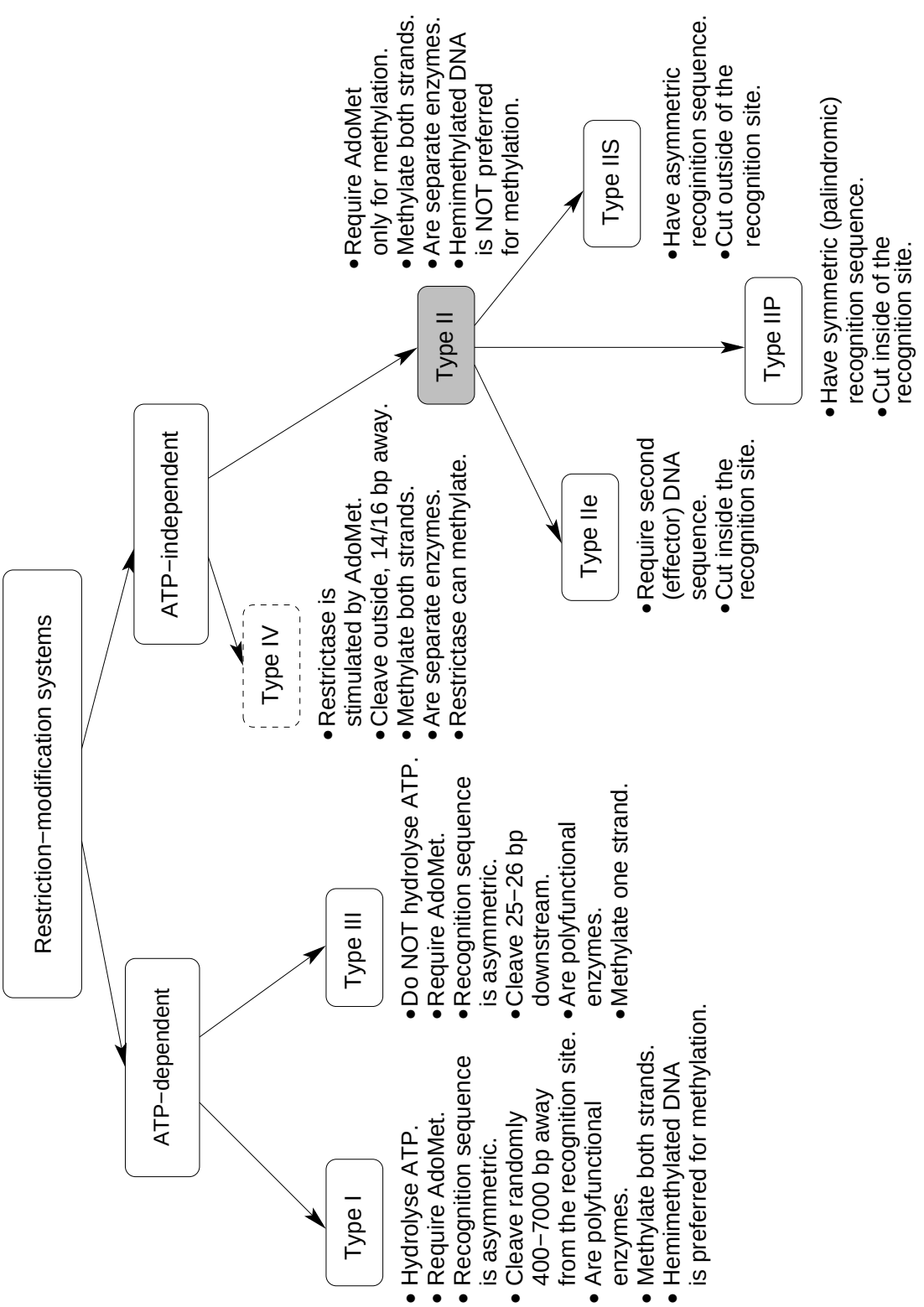


Figure 1.1: Classification of restriction endonucleases.

a functional enzyme. Two kinds of complexes have been isolated for all type I restriction-modification systems: a complex containing products of genes *hsdM* and *hsdS*, and a complex of all three gene products. The HsdM+HsdS complex has methylase activity, whereas the ternary complex has both methylase and endonuclease activity. The HsdS protein alone has no enzymatic activity, but there was demonstrated for *Eco*R124I that it can specifically bind cognate DNA sequence [63]. Thus the three protein chains seem to share their functional roles as follows: HsdS recognises the specific sequence and docks the methylase subunit HsdM to build modification enzyme, or both HsdM and nuclease subunit HsdR to produce restriction enzyme. Two chains HsdS and HsdR are not sufficient for building a functional restriction endonuclease, HsdM is also required.

Comparison of the *hsdS* genes from different species suggests that the HsdS protein consists of two DNA-binding domains joint by a short linker. The linker is supposed to have α-helical conformation and its length determines number of base pairs separating the specific moieties of the recognition sequence. In this way type I restriction modification systems can recognise their asymmetric DNA sites, interrupted by a short stretch (6–8 bases) of a non-cognate DNA.

Type I systems can be currently subdivided into three families named IA, IB and IC systems based on gene complementation analysis [14, 34, 43] and primary sequence comparison. Within the families genes *hsdR* and *hsdM* have much homology. For example, *hsdM* genes in average have 95% amino acid sequence identity [91]. There are no extensive homologies between the members of different families, but the weak homologies that are found indicate that all families had a common ancestor in the past.

Gene *hsdS* has two hypervariable regions that show little sequence similarity among the genes of one family, and several regions of high homology. Each hypervariable region encodes protein domain that recognises one half of the recognition sequence while conserved regions are thought to provide protein-protein interactions with HsdM and HsdR proteins. Interestingly, the hypervariable regions show much more homology between the families (e.g. 50% identical residues between the hypervariable regions of EcoAI and EcoEI families [24]), supporting the idea that these regions of the HsdS protein are responsible for the recognition of the DNA sequence.

The DNA cleavage mechanism of the type I restriction endonucleases is still not clearly understood. Type I restriction-modification enzymes requires AdoMet to form a complex with its target DNA. Upon forming such complex, hemimethylated DNA gets a methyl group on the second strand (fig. 1.2). Unmethylated DNA is methylated much slower in the absence of ATP, and in the presence of ATP it is cleaved. DNA cleavage is accompanied with ATP hydrolysis. Although type I restriction endonucleases remain bound to their recognition site throughout the reaction, the actual cut locations are random and can be hundreds to thousands base pairs away from the actual recognition sites [38, 2, 11, 88, 113]. At least three different models have been proposed to explain how restriction

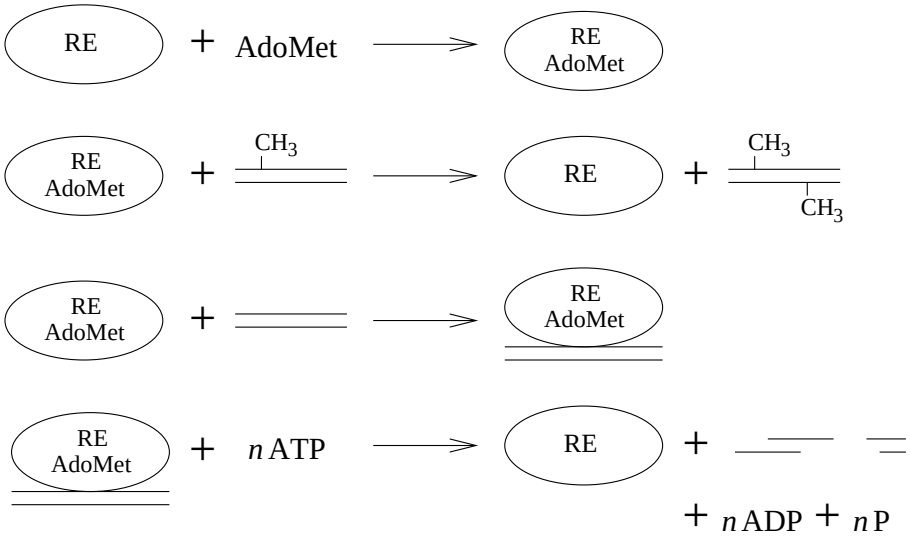


Figure 1.2: Reaction in type I restriction-modification systems. (Adapted from [10])

enzymes find their cleavage sites [88, 113, 96], but the unanimous answer is still to be found.

The DNA ends produced by the type I enzymes are poorly characterised. The only known facts are that they cannot be labelled by polynucleotide kinase, even after phosphatase treatment [30, 72] and that they have long 3' overhangs.

1.3.2 Type III restriction-modification systems

Type III restriction endonucleases also absolutely require ATP for cutting DNA, but unlike type I enzymes they do not hydrolyse it. ATP in this case acts as an allosteric activator of DNA cleavage. AdoMet has no detectable allosteric effects for methylation reaction, but acts as an allosteric activator of DNA hydrolysis [114]. It also acts as methyl group donor during methylation. In this situation DNA cleavage and methylation are competitive reactions.

Type III restriction-modification systems consist of two genes *res* and *mod* [45]. The *mod* gene product recognises the DNA sequence and in addition has methylase activity. The product of the *res* gene acts only in complex with the product of *mod* and is necessary for restriction.

Although the recognition sequence of the type III restriction endonucleases is asymmetric, there must always be two recognition sequences with opposite orientation for the DNA cleavage to take place [70]. These two recognition sites may be thought as one "composite" recognition site which is interrupted in the middle by a non-specific DNA sequence of arbitrary length. This view is further strengthened by the fact that only one half of a single recognition site is methylated by the type III modification enzyme to protect DNA from cleavage [71].

Thus the "composite" site of type III systems is symmetric and similar to the sites of type II systems, except that in type III systems it can be interrupted by a spacer of variable length, whereas in type II systems both halves of the recognition site are always adjacent to each other.

1.4 ATP-independent restriction endonucleases

1.4.1 Type II restriction-modification systems

Type II restriction endonucleases form the largest and the best studied family of restriction-modification systems. They require only Mg^{2+} as a cofactor and do not depend on ATP for their nuclease activity. AdoMet does not influence restriction and is needed only for methylation as a methyl group donor.

Type II restriction-modification systems consist of two genes that encode two separate enzymes. One enzyme has only endonuclease activity, and another is a sequence-specific methyltransferase. Sometimes a third gene is present which has been shown to encode a small protein which is supposed to play a regulatory role. The regulatory role of this protein has been demonstrated for *Bam*HI [98] and *Pvu*II [73] restriction-modification systems, and can be deduced for several others from the sequence similarity.

Most type II enzymes recognise DNA sequences that have a dyad symmetry axis (often such sequences are called palindromic, because the DNA duplex reads the same in both directions). This property of type II systems was explained by postulating that the enzymes function as dimers, each monomer recognising a corresponding half of the recognition site [55]. The dimeric composition of many type II restriction endonucleases can be easily proved by numerous biochemical and physical methods. Dimeric organisation, however, is not the only one compatible with the dyadic symmetry of the substrate. A tetrameric assembly built up as a dimer of dimers would also fulfil this symmetry requirement. A number of restriction enzymes have been shown to exist as tetramers in solution and require tetramerisation for their biological function. The first restriction enzyme which has been shown to function as a tetramer and cut two recognition sequences simultaneously was *Sfi*I [108]. Later, hypothesis that *Cfr*10I could build a tetramer was proposed from the calculations of the contact areas in the *Cfr*10I structure, and confirmed by numerous biochemical and biophysical methods [92]. Finally, a structure of *Ngo*MIV restriction endonuclease was the first structure of tetrameric restriction enzyme in complex with DNA and showed four molecules of the protein in an asymmetric unit bound to two DNA duplexes [28]

Unlike type II restriction endonucleases, the corresponding modification methylases function as momoners, and exist as monomers in solution. Some type II systems can have several modification methylases; for example type IIS systems (see below) may need two independent methylase activities to modify both strands of

their asymmetric recognition sequence.

With more than 2000 of type II restriction endonucleases available today, finer divisions within the family have been detected. There were several schemes that introduce finer classification of type II restriction-modification systems into subfamilies. The most comprehensive attempt was undertaken by Kessler and Manta [56], who suggested a subdivision of type II restriction endonucleases based solely on the properties of their recognition sequence. They have suggested to introduce subclasses IIP, IIW, IIN, IIS, IIT and a "catch-all" class IIU for still unclassified restriction endonucleases (table 1.1), and proposed the same subdivisions for methylases. This classification, although the most comprehensive among the ones suggested to date, still leaves out some important subclasses of type II restriction endonucleases that can be delimited using criteria other than DNA recognition site. Notably, an important subclass of type IIe restriction endonucleases is not reflected in this classification. Probably for this reason, the mentioned classification is not widely used, and only few types that it introduces (IIS and IIP) are cited more often.

Class	Definition	Examples	Rec. sequence
IIP	recognise palindromic uninterrupted DNA sequences of tetra-, hexa- and octanucleatides; cut inside of the recognition sequence	*Hae*III	5'-GG$^{\downarrow}$CC-3'
		*Eco*RI	5'-G$^{\downarrow}$AATTC-3'
		*Bse*634I	5'-$\frac{A}{G}^{\downarrow}$CCGG$\frac{T}{C}$-3'
		*Not*I	5'-GC$^{\downarrow}$GGCCGC-3'
IIW	recognise palindromic penta- or heptanucleotides with unspecified central pair; cut inside of the recognition sequence	*Bst*NI	5'-CC$^{\downarrow}\frac{A}{T}$GG-3'
		*Sau*I	5'-CC$^{\downarrow}$TNAGG-3'
IIN	recognise interrupted palindromes with unspecified central insert of more than one nucleotide; cut inside of the recognition sequence	*Sfi*I	5'-GGCCNNNN$^{\downarrow}$NGGCC-3'
		*Bgl*I	5'-GCCNNNN$^{\downarrow}$NGGC-3'
IIS	recognise largely asymmetric sequences and cut at a predefined distance outside recognition sequence	*Fok*I	5'-GGATGN$_9^{\downarrow}$N$_5$-3' 3'-CCTACN$_{13}^{\downarrow}$N-5'
IIT	very much like IIS enzymes, recognise asymmetric sequences but cut partially inside of the recognition sequence	*Bsm*I	5'-GAATGCN$^{\downarrow}$N-3' 3'-CTTAC$^{\downarrow}$GNN-5'
IIU	temporary subclass for enzymes with asymmetric recognition sequences for which cut position is unknown	*Ngo*BI	5'-GGTGA-3' 3'-CCACT-5'

Table 1.1: Classification of type II restriction-modification systems after Kessler & Manta [56].

Other classifications take into account not only the sequence of the DNA recognition site, but also biochemical properties of restriction-modification systems. So Yang and co-workers [111] mention type IIi restriction endonucleases that do not require Mg^{2+} to bind DNA sequence specifically (such as *Eco*RI) and type IId restriction endonucleases that bind both specific and non-specific DNA equally well and require Mg^{2+} do discriminate between them (examples are *Eco*RV and *Taq*I, see table 1.2).

Class	Definition	Examples	Rec. sequence	Ref.
IIi	Discriminate cognate DNA sequence at a binding stage; do not require Mg^{2+} for specific binding	*Eco*RI	5'-G$^{\downarrow}$AATTC-3'	[111]
IId	Bind both specific and non-specific DNA equally well; discriminate between them only at cleavage stage, in the presence of Mg^{2+}	*Eco*RV *Taq*I	5'-GAT$^{\downarrow}$ATC-3' 5'-T$^{\downarrow}$CGA-3'	[111]
IIe	Require a second DNA (effector) site to cleave DNA	*Eco*RII *Nae*I	5'-$^{\downarrow}$CC$\frac{A}{T}$*GG*-3' 5'-CGG$^{\downarrow}$CCG-3'	[84, 111][a]

[a]Pingoud & Jeltsch [84] also include type IIS and type IV into their table

Table 1.2: Functional classification of restriction endonucleases after Yang *et al.* [111] and Pingoud & Jeltsch [84].

1.4.2 Type IV restriction-modification systems

A small group of restriction-modification systems was discovered by Janulaitis and co-workers in 1992 that does not fall into any of the former classes [47]. Their properties make them similar to type III or type I restriction enzymes. This group includes restriction-modification systems *Eco*57I [47, 48] and *Gsu*I [83]. Because of their unusual properties, it has been suggested to call them "type IV" restriction endonucleases [48], although this terminology has not yet been widely accepted, and sometimes *Eco*57I is classified as type IIS enzyme [86].

Type IV restriction endonucleases do not require ATP for DNA hydrolysis (like type II enzymes). DNA cleavage can proceed without AdoMet, but the presence of AdoMet stimulates it (like for type III enzymes). The recognition sequence of these restriction endonucleases is asymmetric, and the cleavage occurs outside of it.

The type IV restriction-modification system consists of three genes, encoding methylase, nuclease and a small protein which probably plays regulatory role, like in type II systems. Methylase and restriction endonuclease act independently and, interestingly, endonuclease can also catalyse DNA methylation [48], although this methylation is not sufficient to protect DNA from cleavage. *Eco*57I [48] also has regions of homology between restriction endonuclease and "standalone" methylase genes that were not observed in other restriction-modification systems.

1.4.3 Unusual type II restriction endonucleases

Some restriction endonucleases that are currently classified as type II restriction enzymes have unusual properties. Restriction enzyme *Bcg*I, for example,

recognises asymmetric recognition sequence and cuts on both sides of it, thus excising a 34-bp fragment that is methylated [59]. Another example of restriction endonuclease with similar properties is *Sgr*20I (originally called *Sgr*II) [79].

Several type II restriction enzymes can bind two DNA strands simultaneously, and cannot cleave DNA efficiently and completely without prior binding of the second DNA sequence. The early indications of such behaviour were obtained for restriction endonucleases *Eco*RII [61] and *Nae*I [22]. The two DNA binding sites are not equivalent, as was shown for *Nae*I [112]. One of the sites is the actual catalytic site, and another acts as an allosteric activator (effector) of DNA cleavage. Without effector, some refractory DNA sites are cleaved by such enzymes poorly or not at all. Because of these properties it was suggested to call these enzymes type IIe (effector-dependent) restriction endonucleases [111].

It may be tempting to speculate that type IIe enzymes can be tetrameric, they are in fact not: dimeric organisation of the active species has been shown for *Eco*RII [82, 85] and *Nae*I [7]. Thus they are substantially different from the tetrameric restriction enzymes such as *Sfi*I, *Ngo*MIV and *Cfr*10I whose active form is a tetramer [108, 92].

Type IIS restriction endonuclease is defined as "an enzyme which cuts at precise distance away from its recognition site, without cleaving this site" [97]. Although it is not strictly required by this definition, all known type IIS restriction endonucleases have asymmetric recognition sites, and only in some cases there is a partial symmetry of the site. Type IIS restriction endonucleases exist as monomers and consist of two domains – one responsible for DNA recognition and one for catalysis. For *Fok*I restriction endonuclease (belonging to type IIS) it has been shown that a monomer can specifically bind to cognate DNA sequence [94]. However, when it comes to cleavage, *Fok*I needs to dimerise to cleave DNA efficiently [12]. Thus, at least some type IIS restriction endonucleases function as dimers, and it is not clear at the moment whether *Fok*I is a rule or an exception in this respect.

Closely related to type IIS are type IIT enzymes that have very similar properties as type IIS enzymes, but their cut position happens to be partially inside the recognition site.

1.5 Structures of restriction endonucleases

When a large variety of type II restriction endonucleases with different recognition sites has been discovered, it was expected that analysis and comparison of their sequences will show how different DNA sites are discriminated. Astonishingly, it has been found that the sequences of type II restriction endonucleases share, in general, little similarity. The clues for DNA recognition have to be search at the level of the three-dimensional protein structure. This stimulated crystallographic investigations of the restriction enzymes, and currently 11 structures of different

restriction endonucleases (table 1.3) and at least 4 DNA modification methylases from restriction-modification systems (table 1.4).

The X-ray structures of restriction endonucleases, from the first two, *Eco*RI [57] and *Eco*RV [109], to the ones obtained recently show that these enzymes, despite the differences in their overall fold, share significant similarities in the region of central core and active centre. Monomers of all enzymes have mixed α/β structure. The central core of restriction endonucleases has a mixed β-sheet which carries active site at one end.

Active centres of the solved restriction enzymes are also well conserved. Three active site residues, denoted A, B and C [3] have equivalents in all currently know structures of restriction endonucleases (table 1.5). Interestingly, residue A, one of Mg^{2+} binding residues, is always Asp and is strictly conserved in all restriction endonucleases. Other residues have wider variations. So, in *Bam*HI there is an Glu113 residue instead of conserved Lys in position C (table 1.5), and *Bgl*II has Gln95. Of note is that in *Bam*HI an opposite substitution, Glu→Lys, is found at Lys61. Therefore, the net distribution of positively and negatively charged residues remains unchanged in the catalytic centre of *Bam*HI. Since there is growing evidence that enzymes can tolerate "swaps" of important residues, provided that the spatial organisation remains undistorted, one can suggest that the two residues of *Bam*HI Lys61 and Glu113 have changed their positions in the active centre, but still retained the same functions as a similarly charged residues in the active centres of other restriction endonucleases. Two mutation experiments, E113K in *Bam*HI [110] and K92E in *Eco*RV [90] show that these mutants have strongly impaired activity, showing that the mutated residues are important for catalysis. To my knowledge, double mutants K92E-E45K of *Eco*RV or E113K-K61E have not been investigated. Other examples of such residue "swap" will be discussed later, in the discussion of *Bse*634I structure (see p. 74).

Active centre of nucleases seems to be conserved over even broader range of enzymes than restriction endonucleases. Venclovas and Siksnys have shown [102] that enzymes of different family of PNT[2] (HIV-1 integrase and RuvC, Holiday junction resolvase) can use the similar active cite arrangement as found in restriction endonucleases. Thus the active site architecture seen in restriction endonucleases might be well conserved over wide range of enzymes that catalyse chemically similar nucleotidyl-transfer reactions. Some variations in the active site arrangement might be due to "equivalence transformations" such as swap mutations that do not alter the spatial arrangement and charge distribution in the protein's active centre.

In contrast to the catalytic/metal ion binding site, DNA recognition elements show much more variability in restriction endonucleases. For example, the same nucleotide sequence GATC in the centre of the recognition site is recognised differently by *Bam*HI and *Bgl*II [67]. Another example of different base recognition

[2]polynucleotidyltransferases

Enzyme		Recognition sequence	Reference		
*Eco*RI	apo		????	Chandrasekhar, K.	[??]
*Eco*RI	DNA^+	5'-G$^{\downarrow}$AATTC	1990	Kim, Y.C.	[57]
*Eco*RV	apo	5'-GAT$^{\downarrow}$ATC	1993	Winkler, F.K.	[109]
*Eco*RV	DNA^+				
*Eco*RV	DNA^-				
*Eco*RV	DNA^-		1998	Horton, N.C.	[109, 41]
*Pvu*II	apo	5'-CAG$^{\downarrow}$CTG	1994	Athanasiadis, A.	[6]
*Pvu*II	DNA^+		1994	Cheng, X.	[18]
*Bam*HI	apo	5'-G$^{\downarrow}$GATCC	1994	Newman, M.	[76, 77]
*Bam*HI	DNA^+		1995	Newman, M.	[75]
*Bam*HI	DNA^-		2000	Viadiu, H.	[104]
*Cfr*10I	apo	5'-Pu$^{\downarrow}$CCGGPy	1996	Bozic, D.	[15]
*Fok*I	DNA^+	5'-GGATC-9N-13N	1997	Wah, D.A.	[106]
*Bgl*I	DNA^+	5'-GCCNNNN$^{\downarrow}$NGGC	1998	Newman, M.	[74]
*Mun*I	DNA^+	5'-C$^{\downarrow}$AATTG	1999	Deibert, M.	[27]
*Bgl*II	DNA^+	5'-A$^{\downarrow}$GATCT	2000	Lucas, C.M.	[67]
*Nae*I	apo	5'-GCC$^{\downarrow}$GGC	2000	Huai, Q.	[42]
*Ngo*MIV	DNA^+	5'-G$^{\downarrow}$CCGGC	2000	Deibert, M.	[28]

Table 1.3: High resolution three dimensional structures of restriction endonucleases.

Enzyme		Rec. sequence	Reference		
M·*Hha*I	apo	5'-GC^{me}GC	1993	Cheng, X.	[19]
M·*Hha*I	DNA^+		1994	Kilmasauskas, S.	[58]
M·*Hha*I	DNA^-		1999	O'Gara, M.	[78]
M·*Pvu*II			1997	Gong, W.	[35]
M·*Taq*I			1997	Schluckebier, G.	[89]
M·*Dpn*II			1998	Tran, P. H.	[101]

Table 1.4: Three dimensional structures of methylases from restriction-modification systems.

res.	*Bse*634I	*Cfr*10I	*Mun*I	*Ngo*MIV	*Eco*RI	*Eco*RV	*Bam*HI	*Pvu*II	*Bgl*I	*Bgl*II
	E80	E71	-	E70	D59[c]	E45	K61[c]	E55[c]	E87	N54
	P145	P133	P82	P139	P90	P73	I93	N57[c]	P115	I83
A	D146	D134	D83	D140	D91	D74	D94	D58	D116	D84
B	G196	S188[a]	E98	S185	E111	D90	E111	E68	D142	E93
C	K198	K190	K100	K187	K113	K92	E113	K70	K144	Q95
B	E212	E204[b]	L125	E201	N149	-	K156	-	Q161[c]	R108

[a]Not important for catalysis [93]

[b]Important for catalysis [93]

[c]Residues that overlap spatially but come from the non-equivalent secondary structure elements; their correspondence might be casual.

Table 1.5: Correspondence between Mg^{2+} binding and active site residues in the known structures of the restriction endonucleases (not included are *Nae*I (coordinates were not available at the moment of writing) and *Fok*I (type IIS restriction enzyme)). Correspondence of the residues has been taken from superpositions of the central β-sheets.

is *Mun*I as compared to *Eco*RV. The two enzymes have recognition sequences that differ by only one external base-pair. The four inner base pairs are recognised in a similar fashion, but the outer base pair discrimination is achieved in *Mun*I by completely different arrangement than in *Eco*RI: while in *Eco*RI three different residues from two separate structural elements are used, *Mun*I employs single Arg115 that is in a single polypeptide stretch as the inner base-pair recognition residues [27]. Yet another pair differing in recognition machinery but having identical active centre (*Bse*634I and *Ngo*MIV) is described in the discussion of *Bse*634I structure (p. 67).

1.6 Chemical mechanism

It is widely believed that most enzymes catalysing phosphodiester bond cleavage do so *via* pentacoordinate electron-rich phosphorane intermediate, exploiting S_N2 mechanism [33]. Restriction endonucleases are no exception. In restriction endonucleases, and in most other nucleases, hydrolysis proceeds without formation of a covalent intermediate[3]. It has been shown for restriction endonucleases *Eco*RI [21] and *Eco*RV [36] that DNA hydrolysis reaction proceeds with inversion of configuration at phosphorus, thus excluding possibility of pseudorotation. The simplest mechanism compatible with the inversion of the configuration at the phosphorus is an in-line attack of the phosphorus by a water molecule (nu-

[3]Although there are enzymes that catalyse their reactions with formation of the covalent enzyme-phosphate intermediate; notably, this has been proven for snake venom phosphodiesterase and spleen exonuclease

cleofile), through the pentacoordinate intermediate state, to the departure of the leaving pentose hydroxyl group.

A water molecule which attacks the scissile bond phosphate must be activated by some residue before it can do it efficiently. Also, some groups must be present to facilitate the departure of the leaving hydroxyl group and to stabilise the intermediate pentacoordinated phosphorus. The structures of the *Eco*RI and *Eco*RV, however, showed no clear candidate residue which could activate the attacking water molecule. It was therefore suggested that the neighbouring phosphate group next to the scissile bond does the job [49], and later experiments have been done that support this hypothesis [50]. While for *Eco*RI the substrate-assisted activation of the water fits well the experimental picture, for *Eco*RV further kinetic experiments suggested that the picture is different, because Ca^{2+} which itself does not support DNA cleavage by *Eco*RV was found to stimulate the cleavage in the presence of Mn^{2+} [105]. An alternative mechanism was therefore suggested for *Eco*RV that involves two metal ions bound to the active centre, one of which plays a role of water activator [105]. This mechanism was also supported by the finding of two metal ions bound to *Eco*RV-product DNA complex [60]. Recently, two other restriction endonucleases, *Bam*HI and *Pvu*II have been shown to bind two metal ions in their active sites [103, 39]. Thus, it seems that the two metal ion mechanism might be common among restriction endonucleases. Some other DNA processing enzymes that do not belong to restriction-modification systems, notably the exonuclease domain of *E. coli* DNA polymerase (Klenow fragment), have been also shown to employ two metals in catalysis [8], so this mechanism might be very general.

1.7 Scattering of the X-rays

A large amount of our knowledge about the world comes from the investigations of the interaction of radiation and matter. All microscopy, diffraction and spectral methods belong to this group. Since the basic equations describing wave phenomena are always linear, the same method of solving them can be applied in various fields: namely, the radiation waves are described as a sum of the flat sinusoidal waves, and each sinusoidal wave is investigated independently. The development of a function into sin waves is given by its Fourier transform, which provides us with a mathematical apparatus to be applied in a great variety of fields – from radioastronomy to neutron diffraction – that are at the first glance unrelated.

In the interpretation of X-ray diffraction experiments the Fourier transform plays a key role as well. Since in most crystallography textbooks the practical side of the X-ray analysis is more stressed, it was interesting to follow how the basic formulas of X-ray crystallography could be derived from the first principles. Below I have attempted to sketch, very briefly, an outline of such derivation. The

scattering of single electron has been derived after [64], and the treatment of the Fourier transform is after [17].

Notation

Notation	**Meaning**
a, x	real scalars
$\mathbf{A}, \mathbf{B}$	complex scalars, $\mathbf{A} = \lvert\mathbf{A}\rvert e^{i\phi} = \lvert\mathbf{A}\rvert \cos\phi + i\lvert\mathbf{A}\rvert \cos\phi$
ϕ	phase angle of the complex quantity
$\mathrm{Re}(\mathbf{A}), \mathrm{Im}(\mathbf{A})$	real and imaginary part of a complex number $\mathbf{A}$
$\boldsymbol{a}, \boldsymbol{x}$	vectors (in a 3-dimensional Euclidean space)
(a_x, a_y, a_z)	components of a vector, $\boldsymbol{a} = (a_x, a_y, a_z)$
$\dot{\boldsymbol{a}}, \ddot{\boldsymbol{a}}$	first and second derivatives of vector function $\boldsymbol{a}$, $\dot{\boldsymbol{a}} = \frac{d\boldsymbol{a}}{dt}$
$\lvert\boldsymbol{a}\rvert, a$	length of a vector $\boldsymbol{a}$, $a = \lvert\boldsymbol{a}\rvert = \sqrt{(\boldsymbol{a} \cdot \boldsymbol{a})}$
$\hat{\boldsymbol{a}}, \hat{\boldsymbol{x}}$	unit vector in the direction of vector $\boldsymbol{x}$, $\lvert\hat{\boldsymbol{x}}\rvert \underset{def}{=} 1$
$(\boldsymbol{a} \cdot \boldsymbol{b})$ or $(\boldsymbol{ab})$	scalar product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$
$[\boldsymbol{a} \times \boldsymbol{b}]$	vector product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$
$(\boldsymbol{abc})$	triple vector product, $(\boldsymbol{abc}) = (\boldsymbol{a} \cdot [\boldsymbol{b} \times \boldsymbol{c}])$
$\lvert\lvert R\rvert\rvert$	matrix
$\lvert\lvert R\rvert\rvert\boldsymbol{r}$	product of matrix $\lvert\lvert R\rvert\rvert$ and vector $\boldsymbol{r}$
m	mass of electron
c	speed of light in vacuum
ε_0	permittivity of free-space (dielectric constant)
μ_0	permeability of free-space (magnetic constant)
$\boldsymbol{E}$	electric field
$\boldsymbol{H}$	magnetic field
$\boldsymbol{E}_0$	electric field at the origin/scatterer
$\boldsymbol{E}_i$	electric field of the incident wave
$\boldsymbol{E}_o$	electric field of the outgoing wave
I	intensity of the electromagnetic wave
λ	wavelength
$\boldsymbol{k}$	wave vector, $\lvert\boldsymbol{k}\rvert = 2\pi/\lambda$
T	oscillation period
ω	frequency, $\omega = 2\pi/T$
Ω	spatial (spheric) angle
σ	1) scattering cross-section 2) standard deviation
$\boldsymbol{r}$	radius-vector of a point
$\boldsymbol{R}_0$	vector from the centre of mass of the charge system to an observation point
φ	phase of the X-ray reflection
$\boldsymbol{P}$	Poynting's vector

Notation	Meaning
$\rho(\boldsymbol{r})$	electron density at point $\boldsymbol{r}$
$\boldsymbol{S}$	scattering vector
$\mathcal{F}[g]$	Fourier transform of the function $g(\boldsymbol{r})$
$f * g$	convolution of two functions f and g
$\mathbf{F}_{\boldsymbol{h}}$	structure factor of the reflection with index $\boldsymbol{h} = (h, k, l)$
V	volume of a unit cell
V^*	volume of a reciprocal lattice unit cell

1.7.1 Flat waves

We can represent a flat monochromatic wave which runs in the direction of the vector $\boldsymbol{k}$ as sin function:

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E}_0 \sin((\boldsymbol{k} \cdot \boldsymbol{r}) - \omega t + \Delta\varphi) \tag{1.1}$$

For many calculations it is more convenient to represent the wave as an exponential with imaginary exponent, based on Euler's formula $e^{ix} = \cos x + i \sin x$:

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E}_0 e^{i((\boldsymbol{kr}) - \omega t + \Delta\varphi)} \tag{1.2}$$

Further on, we will see that intensity of an electromagnetic wave is proportional to a square of the field amplitude, $|\boldsymbol{E}|^2$. Since we express wave amplitudes as complex quantities, it is useful to express a square of the wave amplitude through its complex representation:

$$\text{let} \qquad E = |E| e^{i\phi} \tag{1.3}$$

$$\text{then} \qquad EE^+ = |E| e^{i\phi} \cdot |E| e^{-i\phi} = |E|^2 e^{i\phi - i\phi} = |E|^2 e^0 = |E|^2 \tag{1.4}$$

For a flat wave, an electrical field vector $\boldsymbol{E}$ is perpendicular to magnetic field vector $\boldsymbol{H}$, and both are perpendicular to the wave front normal $\hat{\boldsymbol{k}}$. Moreover, $\boldsymbol{E}$ and $\boldsymbol{H}$ are proportional to each other, and one can prove that

$$\boldsymbol{E} = \frac{1}{\varepsilon_0 c}[\boldsymbol{H} \times \hat{\boldsymbol{k}}] = \mu_0 c[\boldsymbol{H} \times \hat{\boldsymbol{k}}], \text{ and} \tag{1.5}$$

$$\boldsymbol{H} = \frac{1}{\mu_0 c}[\hat{\boldsymbol{k}} \times \boldsymbol{E}] = \varepsilon_0 c[\hat{\boldsymbol{k}} \times \boldsymbol{E}] \tag{1.6}$$

1.7.2 Scattering by a free electron

In a simplest case, an electromagnetic wave is scattered by a single point charge (electron). We assume that an electron is free, so that the only force acting on it results from the electric field of the incident electromagnetic wave (we can neglect the action of the magnetic field on the electron if velocity of the electron is small).

We write the motion equation of the electron in the electric field (Newton's second law):

$$m\ddot{\boldsymbol{r}} = e\boldsymbol{E_i} \tag{1.7}$$

The dipole moment of the electron is defined as $\boldsymbol{d} = e\boldsymbol{r}$, and we get:

$$\ddot{\boldsymbol{d}} = \frac{e^2}{m}\boldsymbol{E_i} \tag{1.8}$$

Now let's recollect that the field of the electrical dipole is:

$$\boldsymbol{E}_o = \frac{\mu_0}{4\pi R_0}[[\ddot{\boldsymbol{d}} \times \hat{\boldsymbol{k}}] \times \hat{\boldsymbol{k}}] \tag{1.9}$$

This formula can be obtained from the Maxwell's equations, through the *retarded potentials* that give electromagnetic field of moving charges, and under the approximation that 1) the distance R_0 to the observation point is much larger that the size of the charge system and 2) the wavelength of the incident wave is much larger that the size of the charge system [65].

After substituting 1.8 into 1.9, we get an expression for the scattered radiation field amplitude:

$$\boldsymbol{E}_o = \frac{e^2\mu_0}{4\pi m R_0}[[\boldsymbol{E}_i \times \hat{\boldsymbol{k}}] \times \hat{\boldsymbol{k}}] \tag{1.10}$$

$$E_o = \frac{e^2}{4\pi\varepsilon_0 m c^2 R_0} E_i \sin\vartheta \tag{1.11}$$

Let's introduce a shorthand notation

$$f_e = \frac{e^2}{4\pi\varepsilon_0 m c^2}, \tag{1.12}$$

and we can write the emitted field as

$$E_o = f_e \frac{\sin\vartheta}{R_0} E_i \tag{1.13}$$

Energy carried by an electromagnetic wave through a surface element in a unit time is given by the Poynting's vector. The general expression for the Poynting's vector is

$$\boldsymbol{P} = [\boldsymbol{E} \times \boldsymbol{H}] \tag{1.14}$$

Substitution of 1.6 into 1.14 gives

$$P = [\boldsymbol{E} \times \boldsymbol{H}] = \varepsilon_0 c[\boldsymbol{E} \times [\hat{\boldsymbol{k}} \times \boldsymbol{E}]] = \varepsilon_0 c E^2 \hat{\boldsymbol{k}} \tag{1.15}$$

Lets' now calculate the energy (Poynting's vector) of the outgoing wave. For this we substitute 1.9 into 1.15:

$$\boldsymbol{P}_o = \varepsilon_0 c E_o^2 \hat{\boldsymbol{k}} = \frac{\mu_0}{16\pi^2 c R_0^2} [\ddot{\boldsymbol{d}} \times \hat{\boldsymbol{k}}]^2 \hat{\boldsymbol{k}} \tag{1.16}$$

from this, the absolute values of Poynting's vectors for incident and outgoing wave can be written as

$$P_i = \varepsilon_0 c E_i^2 \tag{1.17}$$

$$P_o = \varepsilon_0 c E_o^2 \hat{\boldsymbol{k}} = \frac{\mu_0}{16\pi^2 c R_0^2} [\ddot{\boldsymbol{d}} \times \hat{\boldsymbol{k}}]^2 \tag{1.18}$$

$$= \frac{e^4 \mu_0}{16\pi^2 m^2 c R_0^2} E_i^2 \sin^2 \vartheta \tag{1.19}$$

$$= \frac{e^4}{16\pi^2 \varepsilon_0 m^2 c^3 R_0^2} E_i^2 \sin^2 \vartheta \tag{1.20}$$

The intensity I of electromagnetic wave is the energy carried by the wave through the unit surface in one second (*i.e.* power/area), and its absolute value is equal to an absolute value of the Poynting's vector. Let the $d\mathcal{I}$ be the energy carried by a wave in a spatial angle $d\Omega$. It is equal to the energy carried through the surface dS, where dS is a surface cut out from a sphere of radius R_0 by the spatial angle of the size $d\Omega$. The area dS is expressed through spatial angle as $dS = R_0^2 \, d\Omega$, and we have

$$d\mathcal{I} = I dS = P_o \, dS = P_o R_0^2 \, d\Omega \tag{1.21}$$

The intensity of the incident wave is P_i. The ratio

$$d\sigma \underset{def}{=} \frac{\overline{d\mathcal{I}_o}}{\overline{P_i}} = \frac{P_o}{P_i} R_0^2 \, d\Omega \tag{1.22}$$

is called *scattering cross-section*. It follows from the definition that it has a dimension of area: $[\sigma] = \mathrm{m}^2$

Recalling P_i and P_o from 1.20 and substituting into 1.21, we get a scattering cross-section for monochromatic wave on a free electron:

$$d\sigma = \left(\frac{e^2}{4\pi\varepsilon_0 \cdot mc^2} \right)^2 \sin^2 \vartheta \, d\Omega \tag{1.23}$$

To get the total scattered energy, we integrate over all directions. In spherical coordinates, $d\Omega = \sin\vartheta \, d\vartheta \, d\phi$, and to cover the whole sphere we integrate over ϑ from 0 to π and over ϕ from 0 to 2π. This yields

$$\sigma = \frac{8\pi}{3} \left(\frac{e^2}{4\pi\varepsilon_0 \cdot mc^2} \right)^2 \tag{1.24}$$

This is the scattering cross-section of the linearly polarised light by a free electron. For unpolarised light, we must average 1.23 over all possible orientations of the $\boldsymbol{E}$ field. The averaging gives:

$$\overline{\sin^2\vartheta} = 1 - \overline{(\hat{\boldsymbol{E}}_i \cdot \hat{\boldsymbol{k}})^2} = \frac{1}{2}(1+\cos^2 2\theta) \tag{1.25}$$

and finally

$$d\sigma = \frac{1}{2}\left(\frac{e^2}{4\pi\varepsilon_0 \cdot mc^2}\right)^2 (1+\cos^2 2\theta)\, d\Omega \tag{1.26}$$

Defining $I_{2\theta} = P_o(2\theta)$, $I_0 = P_i$ and remembering that $d\sigma = \overline{d\mathcal{I}_o}/\overline{P_i} = IdS/P_i = P_o dS/P_i = (P_o/P_i)R_0^2\, d\Omega = (I_{2\theta}/I_0)R_0^2\, d\Omega$, we get

$$I_{2\theta} = I_0 \frac{1}{2}\left(\frac{e^4}{4\pi\varepsilon_0 \cdot mc^2}\right)^2 \frac{1}{R_0^2}\,(1+\cos^2 2\theta) \tag{1.27}$$

This formula has been derived for the first time by Thomson in 1897 [100].

1.7.3 Scattering by a system of charges

When the the flat wave 1.2 irradiates the free electron at the origin, the electron starts emitting the electromagnetic wave

$$\boldsymbol{E}_o = \boldsymbol{E}_{o0} e^{i((\boldsymbol{k}\boldsymbol{R}_0)-\omega t+\varphi_0)} \tag{1.28}$$

The initial phase φ_0 reflects the fact that an electron can shift the phase of the re-emitted radiation. Since this shift is the same for all electrons, we can set it to zero by proper choice of the zero on the time axis.

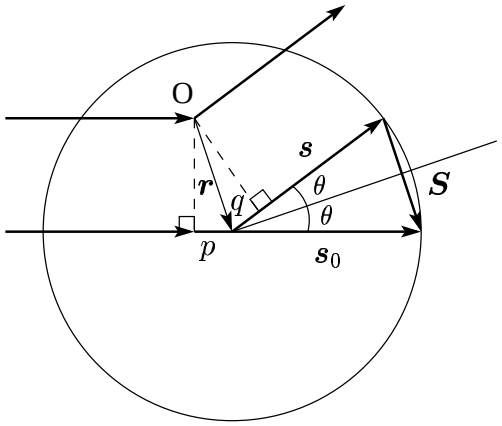


Figure 1.3: Scattering of the atom that is shifted with respect to the origin.

When an atom is shifted by a radius-vector $\boldsymbol{r}$ with respect to the origin, the emitted radiation acquires additional phase shift, namely, it is retarded while the incident wave travels extra distance p, and the emitted wave travels extra distance q. The phase difference compared to the atom at the origin is in this case $\frac{2\pi}{\lambda}(p+q)$. From the figure 1.3 one can derive that $p = (\boldsymbol{r}\cdot\hat{\boldsymbol{s}}_0)$ and $q = -(\boldsymbol{r}\cdot\hat{\boldsymbol{s}})$. The outgoing wave will be therefore given by expression:

$$E_i = E_{i0}e^{i((\boldsymbol{k}_0\boldsymbol{r})-\omega t)}, \tag{1.29}$$

$$E_o = E_i f_e \frac{\sin\vartheta}{R_0} e^{i((\boldsymbol{k}(\boldsymbol{R}_0-\boldsymbol{r}))-\omega t)} \tag{1.30}$$

$$= E_{i0} f_e \frac{\sin\vartheta}{R_0} e^{i((\boldsymbol{k}\boldsymbol{R}_0)-\omega t)} e^{i((\boldsymbol{k}_0-\boldsymbol{k})\boldsymbol{r})} \tag{1.31}$$

$$= E_{o0} f_e e^{i((\boldsymbol{k}_0-\boldsymbol{k})\boldsymbol{r})}, \tag{1.32}$$

$$\text{where} \quad E_{o0} = E_{i0}\frac{\sin\vartheta}{R_0} e^{i((\boldsymbol{k}\boldsymbol{R}_0)-\omega t)} \tag{1.33}$$

It is usual in crystallography to work with vectors $\boldsymbol{s} = \boldsymbol{k}/2\pi$, $\boldsymbol{s}_0 = \boldsymbol{k}_0/2\pi$, and to introduce a *scattering vector* $\boldsymbol{S} \underset{def}{=} \boldsymbol{s}_0 - \boldsymbol{s}$

$$p = (\boldsymbol{r}\cdot\boldsymbol{k}_0) = 2\pi(\boldsymbol{r}\cdot\boldsymbol{s}_0) \tag{1.34}$$

$$q = (\boldsymbol{r}\cdot\boldsymbol{k}) = 2\pi(\boldsymbol{r}\cdot\boldsymbol{s}) \tag{1.35}$$

$$\Delta\varphi = 2\pi(\boldsymbol{r}\cdot(\boldsymbol{s}_0-\boldsymbol{s})) = 2\pi(\boldsymbol{r}\cdot\boldsymbol{S}) \tag{1.36}$$

In this notation, we can write

$$E_o = E_{o0} f_e e^{2\pi i(\boldsymbol{r}\boldsymbol{S})} \tag{1.37}$$

All electrons in a small volume ΔV scatter almost in phase when $\Delta V \to 0$, so the waves of such electrons always add in phase, thus leading to a conclusion that the scattered field is proportional to the number of the electrons in a small volume ΔV. A number of electrons ΔN_e in a volume ΔV can be expressed through *electron density* ρ

$$\rho = \Delta N_e/\Delta V \tag{1.38}$$

When an ensemble of electrons scatters coherently, we must sum the electric field scattered from every small volume to get the total diffracted field in a given direction, taking into account the phase shifts 1.33. Thus we can write:

$$E_o(\boldsymbol{S}) = \sum_{n=0}^{N} E_{o0}\Delta N_{en} f_e e^{2\pi i(\boldsymbol{r}_n\boldsymbol{S})} \tag{1.39}$$

$$= \sum_{n=0}^{N} E_{o0}\rho_n\Delta V_n f_e e^{2\pi i(\boldsymbol{r}_n\boldsymbol{S})}, \tag{1.40}$$

going over to a limit where $N \to \infty$ and $\Delta V_n \to 0$, we get an integral with

$$E_o(\boldsymbol{S}) = \int_V E_{o0}\rho(\boldsymbol{r}) f_e e^{2\pi i(\boldsymbol{r}\boldsymbol{S})} dV \tag{1.41}$$

$$= E_{o0}\, f_e \int_V \rho(\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.42}$$

The integral on the right side depends only on the electron density of the scattering molecule, and is the Fourier-transform of the molecule electron density. It is denoted $F(\boldsymbol{S})$ and called *structure factor* of the molecule, or *molecular transformant*:

$$F(\boldsymbol{S}) = \int_V \rho(\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.43}$$

All possible scattering vectors $\boldsymbol{S}$ constitute a vector space which is often referred to as *reciprocal space*. The space of all laboratory frame vectors $\boldsymbol{r}$ is in contrast called *direct space*.

1.7.4 Properties of the Fourier-transform

As we see, the amplitude of the scattered wave is proportional to a Fourier transform of the electron density, and the intensity of the scattered wave will be proportional to the squared absolute value of the molecular transformant $I(\boldsymbol{S}) \sim |F(\boldsymbol{S})|^2$. We therefore want to examine closer the mathematical properties of the Fourier transform, because these properties allow to explain and predict an X-ray diffraction pattern from any electron density, *i.e.* any molecule.

We will denote Fourier-transform of the function $\rho(\boldsymbol{r})$ shortly as $\mathcal{F}$:

$$\mathcal{F}[\rho(\boldsymbol{r})](\boldsymbol{S}) \underset{def}{=} F(\boldsymbol{S}) = \int_V \rho(\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.44}$$

Linearity

The linearity of the Fourier-transform follows directly from the linearity of the integral (1.43).

$$\mathcal{F}[\alpha\rho_1 + \beta\rho_2] = \alpha\mathcal{F}[\rho_1] + \beta\mathcal{F}[\rho_2] \tag{1.45}$$

Shift property

Moving the electron density by a constant vector $\Delta\boldsymbol{r}$ adds a constant phase shift to a structure factor. More formally, assume $\rho'(\boldsymbol{r}) = \rho(\boldsymbol{r} - \Delta\boldsymbol{r})$, then

$$\mathcal{F}[\rho'] = \int_V \rho'(\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} = \int_V \rho(\boldsymbol{r} - \Delta\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.46}$$

and after variable substitution $\boldsymbol{r'} = \boldsymbol{r} - \Delta\boldsymbol{r}$, $d\boldsymbol{r'} = d\boldsymbol{r}$, we get

$$\mathcal{F}[\rho'] = \int_V \rho(\boldsymbol{r'}) e^{2\pi i((\boldsymbol{r'}+\Delta\boldsymbol{r})\boldsymbol{S})}\, d\boldsymbol{r'} \tag{1.47}$$

$$= e^{2\pi i(\Delta\boldsymbol{r}\boldsymbol{S})} \int_V \rho(\boldsymbol{r'}) e^{2\pi i(\boldsymbol{r'}\boldsymbol{S})}\, d\boldsymbol{r'} \tag{1.48}$$

$$= e^{2\pi i(\Delta\boldsymbol{r}\boldsymbol{S})} \mathcal{F}[\rho] \tag{1.49}$$

In short,

$$\mathcal{F}[\rho(\boldsymbol{r} - \Delta\boldsymbol{r})] = e^{2\pi i(\Delta\boldsymbol{r}\boldsymbol{S})}\, \mathcal{F}[\rho(\boldsymbol{r})] \tag{1.50}$$

Rotation of an electron density

An important question arises how does the scattering of the molecule changes when a molecule (or a whole diffracting specimen) is rotated. Let's denote rotated electron density as $\rho'(\boldsymbol{r})$, and let $||R||$ be a rotation matrix that transforms new vector to an old one (*i.e.* inverse rotation matrix of specimen/molecule rotation). Then

$$\boldsymbol{r'} = ||R||\boldsymbol{r} \tag{1.51}$$

$$\rho'(\boldsymbol{r}) \underset{def}{=} \rho(\boldsymbol{r'}) = \rho(||R||\boldsymbol{r}) \tag{1.52}$$

A molecular transformant of ρ' is

$$F'(\boldsymbol{S}) = \int_V \rho'(\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.53}$$

$$= \int_V \rho(\boldsymbol{r'}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.54}$$

$$= \int_V \rho(||R||\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.55}$$

$$\tag{1.56}$$

We change integration variable from $d\boldsymbol{r'}$ to $d\boldsymbol{r}$

$$\boldsymbol{r'} = ||R||\boldsymbol{r} \tag{1.57}$$

$$d\boldsymbol{r'} = \frac{\partial(x', y', z')}{\partial(x, y, z)}\, d\boldsymbol{r} = det||R||d\boldsymbol{r} = d\boldsymbol{r} \tag{1.58}$$

$$F'(\boldsymbol{S}) = \int_V \rho(||R||\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r'} \tag{1.59}$$

$$= \int_V \rho(\boldsymbol{r'}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r'} \tag{1.60}$$

Since rotation leaves a scalar product of two vectors unchanged, $(\boldsymbol{r} \cdot \boldsymbol{S}) = (||R||\boldsymbol{r} \cdot ||R||\boldsymbol{S})$, and

$$F'(\boldsymbol{S}) = \int_V \rho(\boldsymbol{r'}) e^{2\pi i(||R||\boldsymbol{r}\cdot||R||\boldsymbol{S})}\, d\boldsymbol{r'} \tag{1.61}$$

$$= \int_V \rho(\boldsymbol{r'}) e^{2\pi i(\boldsymbol{r'}\cdot||R||\boldsymbol{S})}\, d\boldsymbol{r'} \tag{1.62}$$

$$= F(||R||\boldsymbol{S}) \tag{1.63}$$

We see that the Fourier transform is rotated in the reciprocal space by the same matrix as a molecule is rotated in real space. In short,

$$\mathcal{F}[\rho(||R||\boldsymbol{r})] = F(||R||\boldsymbol{S}) \tag{1.64}$$

Scattering of a point

If all our electron density contains just one electron at point $\boldsymbol{r}_0$, we want to get back our scattering for a single electron as expressed by 1.37. In other words, then the following must hold:

$$E_o = E_{o0} f_e e^{2\pi i(\boldsymbol{r}_0\boldsymbol{S})} \tag{1.65}$$

$$= E_{o0} f_e \int_V \rho_e(\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.66}$$

We see that our "electron density" in this case must have a unusual property

$$\int_V \rho_e(\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} = e^{2\pi i(\boldsymbol{r}_0\boldsymbol{S})} \tag{1.67}$$

The equation 1.67 can not be fulfilled by any analytical function; we can fulfil it only if we postulate the existence of the special *delta function* such that

$$\int_V \delta(\boldsymbol{r} - \boldsymbol{r}_0) f(\boldsymbol{r})\, d\boldsymbol{r} \underset{def}{=} f(\boldsymbol{r}_0) \tag{1.68}$$

if the integration volume V contains point $\boldsymbol{r}_0$, and (1.69)

$$\int_V \delta(\boldsymbol{r} - \boldsymbol{r}_0) f(\boldsymbol{r})\, d\boldsymbol{r} \underset{def}{=} 0 \tag{1.70}$$

otherwise (1.71)

for any (integratable) function $f(\boldsymbol{r})$. Delta-function has been introduced by Dirac

Comparing 1.67 with 1.71 we can see that the electron density of a single scattering point can be expressed as

$$\rho_e(\boldsymbol{r}) = \delta(\boldsymbol{r} - \boldsymbol{r}_0) \tag{1.72}$$

Fourier-transform of a unity

The delta-function allows to formulate Fourier-transforms of functions that otherwise would have no defined Fourier-transform. Without prove, we note that it can be consistently defined

$$\delta(\boldsymbol{S}) = \int_V e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.73}$$

Convolution and Fourier transform of it

A *convolution* of two functions is defined as

$$(f * g)(u) \underset{def}{=} \int_{-\infty}^{+\infty} f(v)g(u-v)\, dv \tag{1.74}$$

or in three-dimensional case

$$(f * g)(\boldsymbol{u}) \underset{def}{=} \int_V f(\boldsymbol{v})g(\boldsymbol{u}-\boldsymbol{v})\, d\boldsymbol{v} \tag{1.75}$$

If we Fourier-transform a convolution of two functions, we get

$$\mathcal{F}[(f * g)(\boldsymbol{r})] = \int_V (f * g)(\boldsymbol{r}) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.76}$$

$$= \int_V \left(\int_V f(\boldsymbol{v})g(\boldsymbol{r}-\boldsymbol{v})\, d\boldsymbol{v} \right) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.77}$$

We can change the order of integration, and introduce variable substitution in the inner integral $\boldsymbol{r'} = \boldsymbol{r} - \boldsymbol{v}$, $d\boldsymbol{r'} = d\boldsymbol{r}$, $\exp(2\pi i(\boldsymbol{r}\boldsymbol{S})) = \exp(2\pi i((\boldsymbol{r'}+\boldsymbol{v})\boldsymbol{S})) = \exp(2\pi i(\boldsymbol{r'}\boldsymbol{S})) \cdot \exp(2\pi i(\boldsymbol{v}\boldsymbol{S}))$. Then

$$\mathcal{F}[(f * g)(\boldsymbol{r})] = \int_V f(\boldsymbol{v}) \int_V g(\boldsymbol{r'}) e^{2\pi i(\boldsymbol{r'}\boldsymbol{S})}\, d\boldsymbol{r'}\, e^{2\pi i(\boldsymbol{v}\boldsymbol{S})}\, d\boldsymbol{v} \tag{1.78}$$

$$= \int_V f(\boldsymbol{v}) e^{2\pi i(\boldsymbol{v}\boldsymbol{S})}\, d\boldsymbol{v} \int_V g(\boldsymbol{r'}) e^{2\pi i(\boldsymbol{r'}\boldsymbol{S})}\, d\boldsymbol{r'} \tag{1.79}$$

The last two integrals are Fourier transforms of the functions f and g, respectively. Thus

$$\mathcal{F}[f * g] = \mathcal{F}[f] \cdot \mathcal{F}[g] \tag{1.80}$$

Lattice function

A convolution of any function with a delta-function yields the original function shifted with respect to the origin:

$$\delta(\boldsymbol{r} - \boldsymbol{r}_0) * f(\boldsymbol{r}) = \int_V \delta(\boldsymbol{r'} - \boldsymbol{r}_0) f(\boldsymbol{r} - \boldsymbol{r'})\, d\boldsymbol{r'} \tag{1.81}$$

$$= f(\boldsymbol{r} - \boldsymbol{r}_0) \tag{1.82}$$

Let's now consider a *lattice function*

$$L(\boldsymbol{r}) \underset{def}{=} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \sum_{p=-\infty}^{+\infty} \delta(\boldsymbol{r} - \boldsymbol{a}m - \boldsymbol{b}n - \boldsymbol{c}p) \tag{1.83}$$

$$= \sum_{m,n,p=-\infty}^{+\infty} \delta(\boldsymbol{r} - \boldsymbol{a}m - \boldsymbol{b}n - \boldsymbol{c}p) \tag{1.84}$$

Three non-complanar vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are called *lattice vectors*, and their absolute values are *lattice constants* of a crystal.

Let's assume $\rho(\boldsymbol{r})$ is an electron density function of a single molecule. A convolution of ρ with a lattice function gives:

$$\rho_{cryst}(\boldsymbol{S}) = \rho(\boldsymbol{r}) * L(\boldsymbol{r}) \tag{1.85}$$

$$= \int_V \rho(\boldsymbol{r}') \sum_{m,n,p} \delta(\boldsymbol{r} - \boldsymbol{r}' - \boldsymbol{a}m - \boldsymbol{b}n - \boldsymbol{c}p)\, d\boldsymbol{r}' \tag{1.86}$$

$$= \sum_{m,n,p} \rho(\boldsymbol{r} - \boldsymbol{a}m - \boldsymbol{b}n - \boldsymbol{c}p) \tag{1.87}$$

Sum 1.87 describes infinite repetition of the same molecule in all three directions by translations $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ and can be regarded as a description of an infinite crystal.

The structure factor of the infinite crystal will be given a Fourier-transform of 1.87. Applying to it a convolution property 1.80, we get

$$\mathcal{F}[\rho_{cryst}] = \mathcal{F}[\rho * L] = \mathcal{F}[\rho] \cdot \mathcal{F}[L] \tag{1.88}$$

Reciprocal lattice

$\mathcal{F}[\rho]$ is a molecular transformant 1.43. We must now find the Fourier transform of a lattice function L. Denoting this Fourier transform L^* and using integral 1.43, we get

$$L^*(\boldsymbol{S}) = \mathcal{F}[L](\boldsymbol{S}) \tag{1.89}$$

$$= \int_V \sum_{m,n,p} \delta(\boldsymbol{r} - \boldsymbol{a}m - \boldsymbol{b}n - \boldsymbol{c}p) e^{2\pi i(\boldsymbol{r}\boldsymbol{S})}\, d\boldsymbol{r} \tag{1.90}$$

$$= \sum_{m,n,p} e^{2\pi i((\boldsymbol{a}m + \boldsymbol{b}n + \boldsymbol{c}p)\cdot \boldsymbol{S})} \tag{1.91}$$

We want to prove that the sum of exponentials in 1.91 is again a lattice function, with some different lattice vectors.

For this we choose arbitrary integratable function of the scattering vector, $f(\boldsymbol{S})$, and calculate its convolution with the $\mathcal{F}[L](\boldsymbol{S})$:

$$f(\boldsymbol{S}) * L^*(\boldsymbol{S}) = \int_{V^*} f(\boldsymbol{S}')L^*(\boldsymbol{S}-\boldsymbol{S}')\,d\boldsymbol{S}' \tag{1.92}$$

$$= \int_{V^*} f(\boldsymbol{S}') \sum_{m,n,p} e^{2\pi i((\boldsymbol{a}m+\boldsymbol{b}n+\boldsymbol{c}p)\cdot(\boldsymbol{S}-\boldsymbol{S}'))}\,d\boldsymbol{S}' \tag{1.93}$$

$$= \sum_{m,n,p} \int_{V^*} f(\boldsymbol{S}')e^{2\pi i((\boldsymbol{a}m+\boldsymbol{b}n+\boldsymbol{c}p)\cdot\boldsymbol{S}')}\,d\boldsymbol{S}'e^{-2\pi i((\boldsymbol{a}m+\boldsymbol{b}n+\boldsymbol{c}p)\cdot\boldsymbol{S})} \tag{1.94}$$

$$= \sum_{m,n,p} F(\boldsymbol{a}m+\boldsymbol{b}n+\boldsymbol{c}p)e^{-2\pi i((\boldsymbol{a}m+\boldsymbol{b}n+\boldsymbol{c}p)\cdot\boldsymbol{S})} \tag{1.95}$$

since $F(\boldsymbol{r}) = \mathcal{F}[f] = \int_{V^*} f(\boldsymbol{S}')e^{2\pi i(\boldsymbol{r}\boldsymbol{S}')}\,d\boldsymbol{S}'$ is a Fourier transform of f. Using periodicity of e^{ix} it is possible to show that the last sum is periodic function along three vectors $\boldsymbol{a}^*$, $\boldsymbol{b}^*$ and $\boldsymbol{c}^*$ such that

$$\boldsymbol{a}^* = \frac{[\boldsymbol{b}\times\boldsymbol{c}]}{V}, \qquad \boldsymbol{b}^* = \frac{[\boldsymbol{c}\times\boldsymbol{a}]}{V}, \qquad \boldsymbol{c}^* = \frac{[\boldsymbol{a}\times\boldsymbol{b}]}{V} \tag{1.96}$$

$$V = (\boldsymbol{abc}) = (\boldsymbol{a}\cdot[\boldsymbol{b}\times\boldsymbol{c}]) \tag{1.97}$$

If we now take a periodic function $\tilde{f}(\boldsymbol{S}) = f(\boldsymbol{S}) * \sum_{h,k,l=-\infty}^{+\infty} \delta(\boldsymbol{S}-h\boldsymbol{a}^*-k\boldsymbol{b}^*-l\boldsymbol{c}^*)$. Function $\tilde{f}$ can be developed into Fourier series:

$$\tilde{f}(\boldsymbol{S}) = \sum_{m,n,p=-\infty}^{+\infty} F_{mnp}e^{2\pi i(\xi m+\eta n+\zeta p)}, \tag{1.98}$$

where $\boldsymbol{S} = \xi\boldsymbol{a}^*+\eta\boldsymbol{b}^*+\zeta\boldsymbol{c}^*$. The Fourier coefficients F_{mnp} are given by the integral

$$F_{mnp} = \int_{V^*} \tilde{f}(\boldsymbol{S})e^{-2\pi i(\xi m+\eta n+\zeta p)}\,d\xi d\eta d\zeta \tag{1.99}$$

We change now integration variables from ξ,η,ζ to S_x, S_y, S_z. For this we note that $\xi m+\eta n+\zeta p = (\boldsymbol{S}\cdot\boldsymbol{r})$, since $\boldsymbol{S} = \xi\boldsymbol{a}^*+\eta\boldsymbol{b}^*+\zeta\boldsymbol{c}^*$, $\boldsymbol{r} = m\boldsymbol{a}+n\boldsymbol{b}+p\boldsymbol{c}$, and from definition 1.97 follows that $(\boldsymbol{aa}^*) = (\boldsymbol{bb}^*) = (\boldsymbol{cc}^*) = 1$ and $(\boldsymbol{ab}^*) = (\boldsymbol{ac}^*) = (\boldsymbol{bc}^*) = 0$. To change differential, we need a Jacobian of the transformation $(\xi,\eta,\zeta)\to(S_x,S_y,S_z)$

$$(S_x,S_y,S_z) = \xi\boldsymbol{a}^*+\eta\boldsymbol{b}^*+\zeta\boldsymbol{c}^* \tag{1.100}$$

$$\frac{\partial(S_x,S_y,S_z)}{\partial(\xi,\eta,\zeta)} = \begin{vmatrix} a_x^* & a_y^* & a_z^* \\ b_x^* & b_y^* & b_z^* \\ c_x^* & c_y^* & c_z^* \end{vmatrix} = V^* = \frac{1}{V} \tag{1.101}$$

$$d\xi d\eta d\zeta = \frac{\partial(\xi,\eta,\zeta)}{\partial(S_x,S_y,S_z)} dS_x dS_y dS_z \tag{1.102}$$

$$= V dS_x dS_y dS_z = V d\boldsymbol{S} \tag{1.103}$$

Substituting 1.103 into 1.99 yields:

$$F_{mnp} = \int_{V^*} \tilde{f} e^{-2\pi i(\boldsymbol{S}\boldsymbol{r})} V \, d\boldsymbol{S} \tag{1.104}$$

$$= \int_{V^*} f e^{-2\pi i(\boldsymbol{S}\boldsymbol{r})} V \, d\boldsymbol{S} \tag{1.105}$$

$$= VF(\boldsymbol{r}) = VF(m\boldsymbol{a} + n\boldsymbol{b} + p\boldsymbol{c}) \tag{1.106}$$

We notice that $F(m\boldsymbol{a} + n\boldsymbol{b} + p\boldsymbol{c}) = (1/V)F_{mnp}$, and substitute this into 1.95, which gives:

$$f(\boldsymbol{S}) * L^*(\boldsymbol{S}) = \sum_{m,n,p} F(\boldsymbol{a}m + \boldsymbol{b}n + \boldsymbol{c}p) e^{-2\pi i((\boldsymbol{a}m+\boldsymbol{b}n+\boldsymbol{c}p)\cdot\boldsymbol{S})} \tag{1.107}$$

$$= \frac{1}{V} \sum_{m,n,p} F_{mnp} e^{-2\pi i((\boldsymbol{a}m+\boldsymbol{b}n+\boldsymbol{c}p)\cdot\boldsymbol{S})} \tag{1.108}$$

$$= \frac{1}{V} \tilde{f}(\boldsymbol{S}) \tag{1.109}$$

$$= f(\boldsymbol{S}) * \frac{1}{V} \sum_{h,k,l=-\infty}^{+\infty} \delta(\boldsymbol{S} - h\boldsymbol{a}^* - k\boldsymbol{b}^* - l\boldsymbol{c}^*) \tag{1.110}$$

Since 1.110 must hold for any possible function $f(\boldsymbol{S})$, we conclude that

$$L^*(\boldsymbol{S}) = \mathcal{F}[L] = \frac{1}{V} \sum_{h,k,l=-\infty}^{+\infty} \delta(\boldsymbol{S} - h\boldsymbol{a}^* - k\boldsymbol{b}^* - l\boldsymbol{c}^*) \tag{1.111}$$

The lattice described by L^* is called a *reciprocal lattice*.

Laue conditions and Ewald sphere

Once again, scattering of the crystal is given by the Fourier transform of the crystal electron density 1.87, which we can continue with the use of 1.111

$$\mathcal{F}[\rho_{cryst}] = \mathcal{F}[\rho * L] = \mathcal{F}[\rho] \cdot \mathcal{F}[L] \tag{1.112}$$

$$= F(\boldsymbol{S}) \cdot L^\star(\boldsymbol{S}) \tag{1.113}$$

$$= F(\boldsymbol{S}) \cdot \frac{1}{V} \sum_{h,k,l=-\infty}^{+\infty} \delta(\boldsymbol{S} - h\boldsymbol{a}^* - k\boldsymbol{b}^* - l\boldsymbol{c}^*) \tag{1.114}$$

We see that the transformant of the crystal is non-vanishing only for a certain values of $\boldsymbol{S}$, namely for $\boldsymbol{S} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^*$, where h, k and l are all integers. Multiplying this by $\boldsymbol{a}$, $\boldsymbol{b}$ or $\boldsymbol{c}$ and taking into account 1.97, we get the equivalent formulation:

$$(\boldsymbol{S} \cdot \boldsymbol{a}) = h, \qquad (\boldsymbol{S} \cdot \boldsymbol{b}) = k, \qquad (\boldsymbol{S} \cdot \boldsymbol{c}) = l \tag{1.115}$$

$$h, k, l \in \mathbb{Z} \tag{1.116}$$

These conditions are called *Laue conditions*. Thus, X-ray beam will be diffracted only when $\boldsymbol{S}$ satisfies 1.115, since for other values of $\boldsymbol{S}$ the structure factor of the crystal will be null.

If we assume that the scattering elastic, then the wavelength of the scattered radiation does not change, and we can write $|\boldsymbol{k}| = |\boldsymbol{k}_0| = 2\pi/\lambda$, $|\boldsymbol{s}| = |\boldsymbol{s}_0| = 1/\lambda$. The termini of the scattering vector $\boldsymbol{S} = \boldsymbol{s}_0 - \boldsymbol{s}$ then lie on the sphere which has a radius $1/\lambda$ and centre $-\boldsymbol{s}_0$. This sphere is called *Ewald's sphere*, and has been introduced by Ewald in 1921 [31]. It gives us a very convenient method to visualise any diffraction experiment: an incident beam defines the Ewald sphere, our sample defines the molecular transformant, which is a Fourier-transform of the sample's electron density. For a crystal, transformant is a discrete reciprocal lattice. Whenever a node of the reciprocal lattice resides on the Ewald sphere, a beam is reflected by the crystal in the direction of vector $\boldsymbol{s}$, with the intensity proportional to $|F(\boldsymbol{S})|^2$

1.8 Phase determination

Currently, only time-averaged intensities $\overline{\mathcal{I}}$ can be measured directly in an X-ray diffraction experiment. From these intensities absolute values of structure factors $|\mathbf{F}(\boldsymbol{S})|$ can be deduced. For calculation of an electron density however a complex quantity $\mathbf{F} = |\mathbf{F}|e^{i\varphi}$ defined by its absolute $|\mathbf{F}|$ and phase φ for each reflection is required. The phases missing in the experimental data must thus be reconstructed by some other means, so that the *phase problem* is solved.

Phase problem can be solved in several ways:

- isomorphous replacement,
- molecular replacement,
- multiple anomalous diffraction experiment,
- direct methods.

Usually, a combination of approaches is used in structure determination. For example, in the isomorphous replacement method, heavy atom positions must be determined before phase calculation can take place, and this is usually done using some kind of direct methods.

1.8.1 Isomorphous replacement

In the isomorphous replacement method one attempts to collect X-ray reflection data from at least two crystals – one containing native (protein) molecules, and another which is exactly the same but for a few heavy atoms introduced into the molecule of the native compound. Such two crystals are called *isomorphous*, since

the electron density in the crystal is the same everywhere except in the regions where the heavy atoms are bound. If we can find the position of the heavy atoms by some means, then the phases of the native molecular transformant can also be calculated. For this, from the known positions and scattering factors of the heavy atoms, structure factors of the substructure consisting solely of the heavy atoms can be derived. Then, from the linearity of Fourier-transform, we get $\mathbf{F}_{pq} = \mathbf{F}_p + \mathbf{F}_q$. In this equation, $\mathbf{F}_q$, $|\mathbf{F}_{pq}|$ and $|\mathbf{F}_p|$ are know.

The equation can be visualised by plotting the complex quantities $\mathbf{F}_p$, $\mathbf{F}_q$ and $\mathbf{F}_{pq}$ on the complex plain ($\mathrm{Re}(\mathbf{F}) \times \mathrm{Im}(\mathbf{F})$) (fig. 1.4). Since only absolute values of native and derivative structure factors are known, their complex vectors can be anywhere on the circles with radii $|\mathbf{F}_p|$ and $|\mathbf{F}_{pq}|$, respectively. Assuming we can somehow obtain positions of the heavy atoms, then the vector $\mathbf{F}_q$ can be calculated – both its phase angle and length. Then if we choose the origin of the $\mathbf{F}_p$ vector at complex zero, the origin of $\mathbf{F}_{pq}$ vector will also be fixed at $-\mathbf{F}_q$ (fig. 1.4). The two circles with radii $|\mathbf{F}_p|$ and $|\mathbf{F}_{pq}|$ will have, in general, two intersections, corresponding to two possible phase angles of $|\mathbf{F}_p|$ that are compatible with the measured $|\mathbf{F}_p|$ and $|\mathbf{F}_{pq}|$ and with estimated heavy atom positions. The method described above is termed *single isomorphous replacement* (*SIR*).

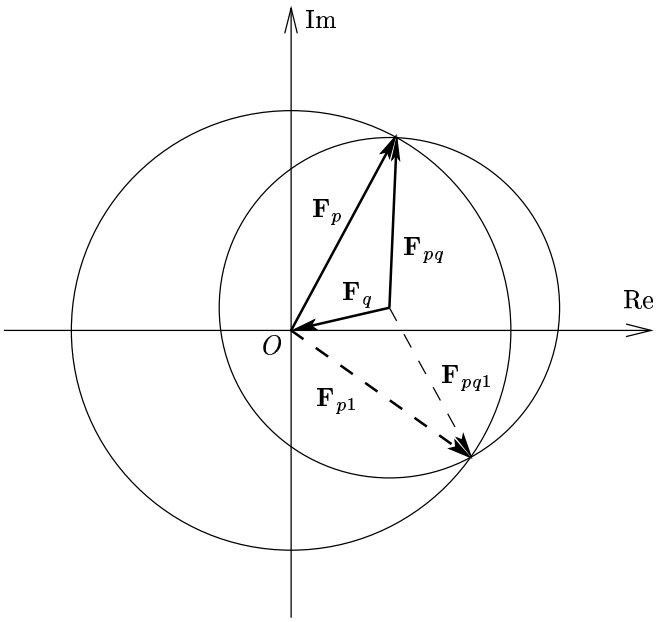


Figure 1.4: Harker's diagram of single isomorphous replacement.

As we see, the solution of the phase problem by means of SIR is ambiguous, giving two possible phase angles for every reflection. One can choose to ignore

this ambiguity and choose randomly one phase value. For the half of reflections the phases then will be correct; another half will be assigned phases that differ randomly from the "correct" value. This second "incorrect" half will add up to a noise in the Fourier synthesis, while the correctly phased reflections will produce electron density with the features of the real structure. In practice, SIR maps are very noisy and are seldom of use for determination of protein structure without some extra assumptions or measurements (such as non-crystallographic symmetry). Their primary use is to search for extra heavy atom sites in the same or other heavy atom derivatives, and to find correct *"hand"* and origin of the heavy atom coordinates.

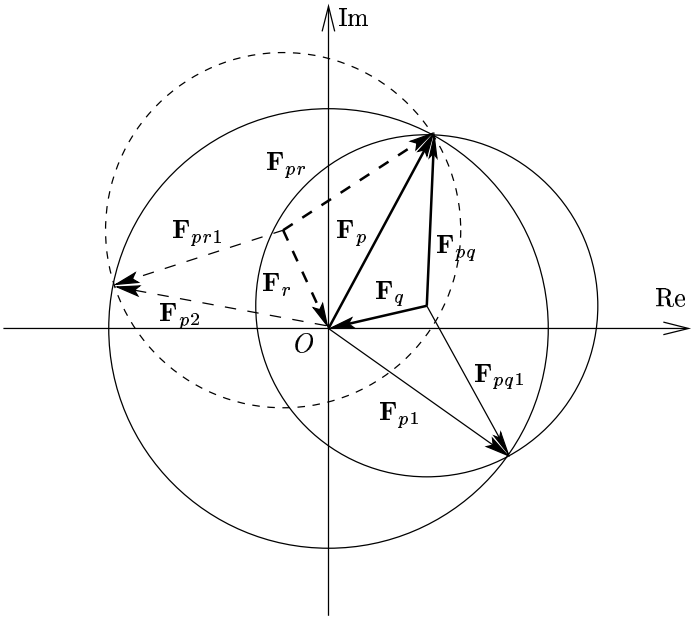


Figure 1.5: Harker's diagram of multiple isomorphous replacement.

If we can obtain second isomorphous derivative of the same native structure, then the ambiguity of SIR phasing can be resolved. Indeed, the second derivative will have, in general, different structure factor $\mathbf{F}_r$, and the origin and radius of the corresponding circle on the Harker diagram will be different. The three circles (describing native structure factor and two derivative structure factors, see fig. 1.5) must however intersect at one point, since both equations

$$\begin{aligned} \mathbf{F}_{pq} &= \mathbf{F}_p + \mathbf{F}_q \\ \mathbf{F}_{pr} &= \mathbf{F}_p + \mathbf{F}_r \end{aligned} \tag{1.117}$$

must be satisfied.

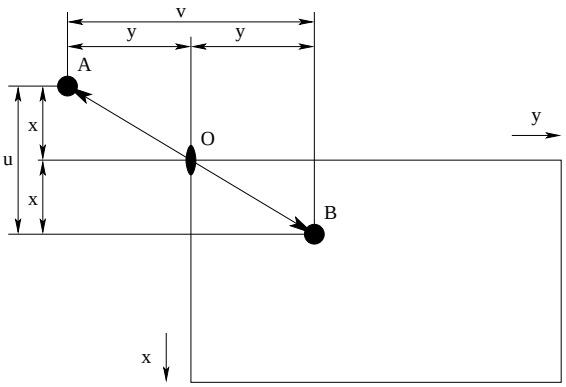


Figure 1.6: Building of the Harker peaks

In practice, the three circles never intersect at exactly the same point because of noise in $|\mathbf{F}|$ measurements and errors in the estimations of heavy atom positions. Therefore, one can only assume that one SIR solution is more probable that the other for every given reflection. Additional isomorphous derivatives can then be used to increase our confidence in the correctness of the chosen phase solution. The method is thus called *multiple isomorphous replacement* (*MIR*).

1.8.2 Harker section interpretation

Calculation of the heavy atom structure factors $\mathbf{F}_q$ in 1.117 requires a knowledge of the heavy atom positions. These positions can be found by inspecting *Harker sections* of the difference Patterson map.

Due to symmetry elements in the elementary cell of the crystal, Patterson vectors from the symmetry related (heavy) atoms fall into special planes, called *Harker sections* [37]. As shown in fig. 1.6 with two-fold crystallographic axis as example, an atom with a radius-vector $\overrightarrow{OA}$ from the origin will always have a symmetry equivalent $\overrightarrow{OB}$. Since a Patterson map always contains difference vectors from all atoms in the unit cell, vector $\overrightarrow{AB}$ will be present in the Patterson map. The two-fold axis at O maps the point A with coordinates (X, Y, Z) to the point $(-X, -Y, Z)$, and the components of the vector $\overrightarrow{AB}$ will be $\overrightarrow{AB} = (U, V, W) = (X, Y, Z) - (-X, -Y, Z) = (2X, 2Y, 0)$. We see that all Patterson vectors fall onto the plane $Z = 0$, which in this case is one of the Harker sections. If an atom at A is heavy enough, the there will be a prominent peak at the end of the Patterson-Harker vector $(U, V, W) = (2X, 2Y, 0)$ (see fig. 2.1), and the coordinates U and V can be easily found.

From the Harker peak positions, X and Y coordinates can be calculated. If the spacegroup has three perpendicular dyad axes, then it will have Harker sections $X = 0$ and $Y = 0$, from which coordinate Z of the heavy atom can be derived. The solution (X, Y, Z) is not unique – it is easy to see that a Harker peak $(U + 1, V, 0)$ will give a heavy atom position which is shifted half a cell with

respect to our original one. This solution relates heavy atom to another possible *origin* in the unit cell. Also, a solution $(-X, -Y, -Z)$ which is mirror-reflected at the origin will be equally compatible with the given Patterson map, leading to a so called *hand* ambiguity.

The origin of the first heavy atom can be chosen arbitrarily, but the rest of the heavy atoms must have coordinates expressed with respect to this chosen origin. This can be done by means of *difference Fourier* synthesis. Also, phasing statistics will become worse if we chose two heavy atoms with different origins. By inspecting difference Fourier maps and phasing statistics, the same origin for all heavy atoms can be chosen. The correct hand of the heavy atom solutions $(X, Y, Z$ or $-X, -Y, -Z)$ can be chosen later by inspecting the hand of the alpha-helices in the electron density (they should be right-handed), or by the analysis of the anomalous phasing statistics if the anomalous data are available.

Chapter 2

Materials and methods

2.1 Biochemical methods

2.1.1 Materials

All chemicals, unless specially stated, have been purchased from Aldrich (Steinheim), Merck (Darmstadt), Riedel-de-Häen (Seelze), Serva (Heidelberg), Sigma (Deisenhofen) or Fluka (Neu-Ulm), according to their yearly regular catalogs. Water for all buffers was Milli-Q (Millipore) deionised water.

Nr.	Item	Producer
1.	Centricon membrane concentrators, 10kDa cut-off	Amicon
2.	Centriprep membrane concentrators, 10kDa cut-off	Amicon
3.	Cellophane sheets	Novex
4.	Dialysis membranes, 12kDa cut-off	Roth
5.	Dialysis membranes, 12–16kDa cut-off, 25 Å pore size, type 8	Biomol
6.	Oligonucleotides	Pharmacia, MWG, Genzentrum
7.	Glass capillaries	W. Müller
8.	Harward Modeling Wax	Richter & Hoffmann Harward Dental GmbH
9.	Nylon loops and holders for cryo-crystallography	Hampton research
10.	Magnetic holders for cryo-loops	Hampton research

Nr.	Item	Producer
11.	Instruments for handling crystals in liquid N_2	Hampton research
12.	Goniometer head	Huber
13.	Cryschem crystallisation plates	Charles Super Company Inc., Nattic, MA, USA
14.	Linbro plates	ICN Biochemicals GmbH
15.	Cover glasses for Linbro plates, $\varnothing 21$ mm, $d = 0.2$ mm	Labor Schubert & Weiss GmbH
16.	Silicon grease	Roth, Sigma
17.	Parrafine oil, low density	Henry Lamotte
18.	Desalting gel filtration column NAP-25	Pharmacia
19.	Protogel acrylamide gel stock, 30% acrylamide, acrylamide:bis-acrylamide = 37.5:1	National Diagnostics

Nr.	Device	Type	Producer
1.	UV-spectrophotometer	Lambda 17	Perkin-Elmer
2.	UV/VIS diode array spectrophotometer	DU 7500	Beckman
3.	Image plate X-ray detector	Mar180	Mar Research
4.	Image plate X-ray detector	Mar300	Mar Research
5.	Image plate X-ray detector	Mar345	Mar Research
6.	CCD X-ray detector	MarCCD	Mar Research
7.	FPLC chromatography station	FPLC	Pharmacia
8.	HPLC chromatography station		Kontron Instruments
9.	Vacuum concentrator		Bachofer
10.	Rotating anode X-ray generator	RU-200	Rigaku
11.	Cryosystem		Oxford Cryosystems

Name	Length	Sequence	Source	Purif.	Notes
CAA	10	GC CAA TTGGC	MPI	MQ	
CAA	10	GC CAA TTG GC	Ph,GC	RF	
$CAAS_1$	10	GC CAA_S T_STG GC	Ph	RF	
$CAAS_2$	10	GC C_SAA TTG GC	Ph	RF	
$CAAI_1$	10	GC C_IAA TTG GC	MPI	MQ	does not bind?
$CAAI_2$	10	GC_I CAA TTG GC	MPI	MQ	
$CAAU_1$	10	GC CAA TU_IG GC	MPI	MQ	
$CAAU_2$	10	GC CAA U_ITG GC	MPI	MQ	
ATC10	10	AT CAA TTG AT	MWG	RF	
CGC12	12	CGC CAA TTG GCG	MWG	RF	
TAT12	12	TAT CAA TTG ATA	MWG	RF	
TCG13	13	*T* CGC CAA TTG GCG	MWG	RF	CGC12 with overhanging T
TGC11	11	*T* GC CAA TTG GC	MWG	RF	CAA with overhanging T
ATG12	12	*AT* GC CAA TTG GC	MWG	RF	CAA with overhanging AT
TTA13	13	*T* TAT CAA TTG ATA	MWG	RF	TAT12 with overhanging T
TAT20	20	TATC GGC CAA TTG GCC GATA	GC	RF	
GCT20	20	GCTC GGC CAA TTG GCC GAGC	GC	RF	
HPin-34	34	CGGC CAA TTG GCC G*TTC* GCCG GTT AAC CGG C*TTC*	GC	RF	Hairpin structure
HM-34	34	CGGC C*GA* TTG GCC G*TTC* GCCG GTT AAC CGG C*TTC*	GC	RF	HPin-34 with 1 base mismatch; does not bind
TTA13-U3I	13	*T* TAU_I CAA TTG ATA	MWG	RF	TTA13 with iodinated U
TTA13-U2I	13	*T* U_IAT CAA TTG ATA	MWG	RF	TTA13 with iodinated U
TTA13-U1I	13	U_I TAT CAA TTG ATA	MWG	RF	TTA13 with iodinated U
TTA13-C1I	13	*T* TAT C_IAA TTG ATA	MWG	RF	TTA13 with iodinated C
TTC13-C1I	13	*T* TC_IT CAA TTG AGA	MWG	RF	TTA13 with iodinated C instead of first A

Sequences are listed in 5'→3' order

Table 2.3: Tested oligonucleotides for *Mun*I

Unless specially stated, all percentages are given in weight/volume parts for solid substances and in volume/volume parts for liquids.

2.1.2 Protein storage

Proteins for crystallisation trials have been purified in the Institute of Biotechnology, Vilnius, Lithuania, and stored frozen at -20° in the storage buffer (table 2.4) in 1–2 ml aliquots. At these conditions proteins could be kept for more than a year without significant change in their activity or crystallisation behaviour.

Table 2.4. Protein storage buffer (4.09.1998)

Tris·HCl, pH 7.4	10 mM
KCl	100 mM
DTT	1 mM
EDTA	1 mM
Glycerol	50 % (v/v)

Table 2.5. Protein storage buffer, old composition (26.07.1994)

KH_2PO_4, pH 7.4	10 mM
NaCl	450 mM
β-ME	1 mM
EDTA	1 mM

2.1.3 Protein preparation

Prior to crystallisation experiments all proteins were dialysed against their respective dialysis buffers and then concentrated. Dialysis of 1.0 ml protein samples was performed against 2×0.5 l dialysis buffer (tables 2.6, 2.7) in Biomol dialysis membrane tubes (type 8, MW cutoff 12–16 kDa) at 4°. Buffer was changed after 3–4 hours, the dialysis was carried on overnight.

After dialysis, the protein was concentrated in Centricon concentrators at 4° to the final concentration of about 1 mM, as described in the Centricon user manual. The final concentration of the protein was monitored by UV absorbance at 280 nm. The concentrated protein was stored or at 4°C. Under these conditions, *Mun*I could be kept for about 4 days, and *Bse*634I for 2 weeks. After these periods, samples were no longer suitable for crystallisation, and fresh samples had to be prepared.

Table 2.6. *Mun*I dialysis buffer

Tris·HCl, pH 8.5	10 mM
NaCl	50 mM
EDTA	1 mM
NaN_3	0.02 %

Table 2.7. *Bse*634I dialysis buffer, low salt

Tris·HCl, pH 7.5	20 mM
NaCl	50 mM
EDTA	1 mM
NaN_3	0.02 %

2.1.4 Oligonucleotide purification

Oligonucleotides were ordered from MWG, Pharmacia, Gencentrum or MPI to be synthesised with a standard phosphoamidite chemistry. In cases where a crude oligonucleotide was delivered, the oligonucleotides were purified on the MonoQ column according to the method suggested by Pharmacia, and then desalted over a NAP-25 column. If the HPLC purified oligonucleotide was delivered, it was additionally precipitated from the Na Acetate to remove possible traces of Mg^{2+}.

For the MonoQ purification, a 3 ml MonoQ column was equilibrated against 3 column volumes of 10% buffer B, and then a piecewise-linear gradient was used (see table 2.8). A typical injection would contain 0.46 μmol (55 OU) of decamer CAA oligonucleotide. Middle fractions containing main product were collected, pooled and desalted over NAP-25 gel filtration column.

Table 2.8. Buffers for MonoQ purification of oligonucleotides

Buffer A:		
NaOH	10	mM
Buffer B:		
NaOH	10	mM
NaCl	1.5	M

Table 2.9. Gradient for oligonucleotide purification on MonoQ column

Gradient:			
Column Volumes	% B	Column Volumes	% B
0.0	10	20.0	100
1.7	10	21.7	100
6.7	45	23.3	0
18.3	75		

2.1.5 Concentration measurements

Concentrations of pure oligonucleotides were measured according to their absorption at 260 nm. Molar extinction coefficients were taken from handbook by Dawson *et al.* [26] (table 2.10). Total extinction coefficient was approximated as a sum of extinction coefficients for all monomers. The values provided assume that nucleotide forms duplex (*i.e.* hyperchromic effect is already accounted for in the monomer extinction coefficients). All molar oligonucleotide concentrations are quoted with respect to one DNA strand.

Protein concentration was measured by absorption at 280 nm. Protein molar extinction coefficient was approximated as a sum of tryptophane and tyrosine extinction coefficients [23] (table 2.11). All molar protein concentrations are quoted with respect to one polypeptide chain. Extinction coefficients of the macromolecules are summarised in table 2.12.

Table 2.10. Molar extinction coefficients of nucleotides

Base	ε_{260}, $\mathrm{mM}^{-1}\mathrm{cm}^{-1}$
A	15.3
G	11.8
C	7.4
T	9.3

Table 2.11. Molar extinction coefficients of aromatic amino acids

Amino Acid	ε_{280}, $\mathrm{mM}^{-1}\mathrm{cm}^{-1}$
Trp	5.6
Tyr	1.4

$\varepsilon_{280}^{Mun\mathrm{I}}$	=	45.7 $\mathrm{mM}^{-1}\mathrm{cm}^{-1}$
$\varepsilon_{280}^{Bse634\mathrm{I}}$	=	34.4 $\mathrm{mM}^{-1}\mathrm{cm}^{-1}$
$\varepsilon_{280}^{Cfr10\mathrm{I}}$	=	21.8 $\mathrm{mM}^{-1}\mathrm{cm}^{-1}$
$\varepsilon_{260}^{\mathtt{CAA}}$	=	118.6 $\mathrm{mM}^{-1}\mathrm{cm}^{-1}$

Table 2.12. Extinction coeffcients of proteins and oligonucleotides

2.1.6 Native electrophoresis

Native electrophoresis was performed in buffer systems described by McLellan [69]. The buffer systems used were histidine/MES (pH = 6.1–6.2), histidine/MOPS (pH = 6.5–6.6), imidasole/HEPES (pH = 7.5–7.6) and Tris/boric acid (pH = 8.5–8.7). Gels were polymerised with 5% acrylamide (AA:BisAA = 37.5:1) from Protogel acrylamide stock. Addition of 1.5 mM EDTA into the gel and electrode buffers would give sharper bands. Gels were 10 cm×10 cm×1.0 mm large, and were run at 100 V, constant voltage, at room temperature. A typical sample would contain 1 μg to 5 μg protein or protein-oligonucleotide complex in 2 μl volume (gel load 0.5 to 2.5 $\mu\mathrm{g}/\mathrm{mm}^2$), dissolved in the electrophoresis buffer with addition of 25% glycerol and 0.01% Bromphenol Blue dye. After the run gels were stained in Coomassie Brilliant Blue solution (0.04% Coomassie R-250, 40% methanol, 10% acetic acid) and destained with solution containing 20% ethanol and 10% acetic acid. Afterwards gels were rinsed with solution containing 30% glycerol and 10% ethanol and dried between cellophane sheets stretched on plexiglass frames.

Table 2.13. Buffer for native electrophoresis, Histidine/MES

Histidine base	30 mM
MES	30 mM
EDTA	1.5 mM

Table 2.14. Buffer for native electrophoresis, Histidine/MOPS

Histidine base	25 mM
MOPS	30 mM
EDTA	1.5 mM

Table 2.15. Buffer for native electrophoresis, Imidasole/HEPES

Imidasole	50 mM
HEPES	33 mM
EDTA	1.5 mM

Table 2.16. Buffer for native electrophoresis, Tris/boric acid

Tris base	50 mM
Boric acid	30 mM
EDTA	1.5 mM

2.2 Crystallographic methods

2.2.1 Crystallisation

Immediately before crystallisation, a protein was diluted with the protein dialysis buffer to the final concentration use for crystallisation. In a case of protein-oligonucleotide complexes, the protein was mixed with an oligonucleotide in a molar ratio protein:oligonucleotide = 1:1.2 (*i.e.* a small excess of oligonucleotide has been taken). The samples then were centrifuged in a table centrifuge at a maximum available speed for 15 min at 4°C to remove any precipitate or dust particles. The sample was then pipetted to the Cryschem® crystallisation plates or onto the glasses of Linbro® plates, mixed up with the reservoir (crystallisation) solution and immediately sealed. The protein droplet size used was 1 to 4 μl, and it was mixed with 1–4 μl of the reservoir solution. Plates were stored in crystallisation room at 18°C. Crystals would normally appear in 4–5 days. The largest crystals for *Bse*634I were obtained from the drops composed of 4 μl protein solution and 1 μl reservoir solution.

Table 2.17. *Mun*I crystallisation buffer

Na·MES, pH 6.0	100 mM
$CaCl_2$	50 mM
PEG 8K	5–9 %
NaN_3	0.02 %

Table 2.18. *Mun*I crystallisation buffer with Mg^{2+}

Na·MES, pH 6.0	100 mM
$CaCl_2$	50 mM
$MgCl_2$	3 mM
PEG 8K	1.0–3.0 %
Sucrose	10 %
NaN_3	0.02 %

Table 2.19. *Bse*634I crystallisation buffer

Na Acetate, pH 5.5	100 mM
composed as:	
Na Acetate	88.4 mM
Acetic acid	11.6 mM
PEG 8K	12–14 %
$CaCl_2$	100 mM
NaN_3	0.02 %

2.2.2 Crystal freezing

*Bse*634I crystals were frozen immediately before measurement in cold nitrogen stream (90K) at the measurement site. Before measurement, crystals were soaked

in the *Bse*634I cryoprotector buffer (see composition in table 2.20) for 1h 45 min. *Bse*634I crystals become disordered immediately after contact with the cryoprotector buffer, however, they regain their diffraction capability after soaking times >1h.

The *Mun*I crystals were soaked in the *Mun*I cryoprotector buffer (see composition in table 2.21) for 50 min. The crystals were then frozen in cold nitrogen stream (90K), tested for diffraction and stored in liquid nitrogen with the help of cryo-clamps (Hampton Research). Crystals were transported to the measuring site in liquid nitrogen.

Table 2.20. *Bse*634I cryoprotector buffer

Na Acetate, pH 5.5	100 mM
composed as:	
Na Acetate	88.4 mM
Acetic acid	11.6 mM
PEG 8K	14–16 %
$CaCl_2$	100 mM
Glycerol	25 %
NaN_3	0.02 %

Table 2.21. *Mun*I cryoprotector buffer

Na·MES, pH 6.0	100 mM
$CaCl_2$	50 mM
PEG 8K	9.5 %
Glycerol	35 %
Sucrose	10 %
NaN_3	0.02 %

2.2.3 Preparation of the heavy atom derivatives

Heavy atom derivatives of *Bse*634I were prepared by soaking *Bse*634I crystals in a crystallisation reservoir solution containing 1–10 mM of the heavy atom compound for highly soluble compounds, or a saturated solution for compounds with low solubility. Soaks took from several hours to several days. Crystals were measured at room temperature on the rotating anode source; images were recorded on Mar180 image plate detector. To exclude a possibility of systematic bias, a native data set has been also collected from *Bse*634I under the same conditions, and used to calculate isomorphous differences.

2.2.4 Data collection and processing

The data for the final refinement have been collected on the BW6 beamline at DESY. The data from the heavy atom derivatives have been collected on a Rigaku RU-200 rotating anode generator equipped with a MAR research Image plate detector. Optimal strategy for data collection was calculated using `STRATEGY` option of `MOSFLM` program. Oscillation images have been processed using `DENZO` program package [81] and scaled with `Scalepack`; the data collection statistics is shown in table 3.1.

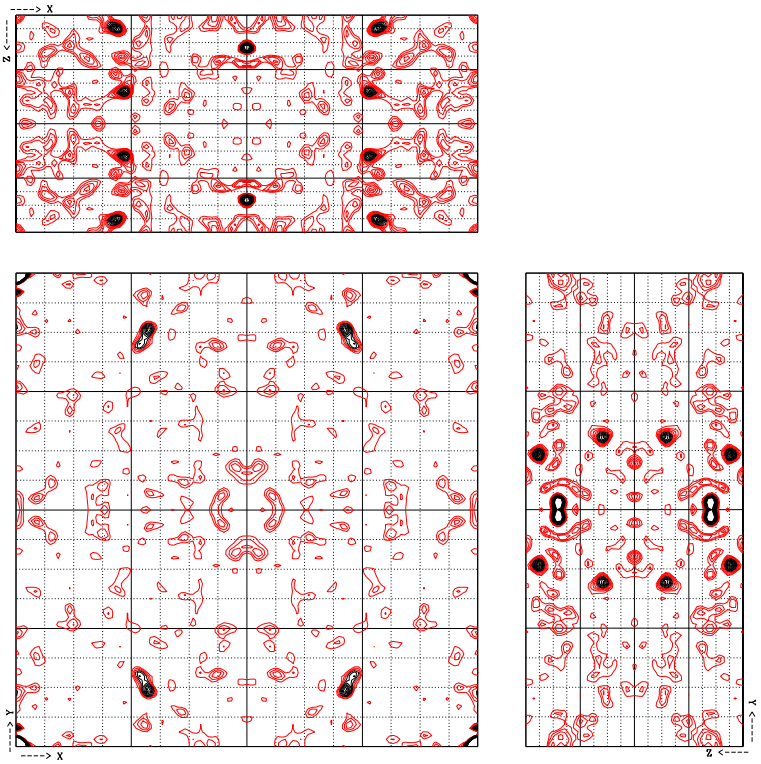


Figure 2.1: Harker sections of the *Bse*634I $HgCl_2$ derivative Patterson map.

2.2.5 Phase determination and refinement

Difference Patterson maps for the heavy atom derivatives have been calculated using `CCP4` [20] program suite. The Harker sections of the maps have been extracted and searched for possible heavy atom positions using `hara` program. The found locations of the heavy atom positions have been brought to a single origin and hand using difference Fourier synthesis, then the positions have been refined and the MIR phases have been calculated with the help of `mlphare` [80] program from the `CCP4` suite.

The `hara` program appeared to be advantageous for quick scanning of heavy atom derivatives for two reasons. A good difference Patterson map (for example see fig. 2.1) would normally yield more than 100 peaks, most of which are noise, when processed with available peak search programs. For bad maps, the ratio can be even worse. The "correct" heavy atom peaks may be at positions 50 to 75 in this list, so all peaks must be examined in order not to miss the correct ones. The peak search routine is sensitive to a threshold which must be selected individually for every Harker section. Since the programs usually do not have graphical output, there is no automated way to relate Patterson map peaks with the calculated heavy atom positions. An algorithm used in `hara` circumvents all these problems by using automatic threshold determination of all Harker sections and cross-checking all peaks in all available sections. In higher symmetry

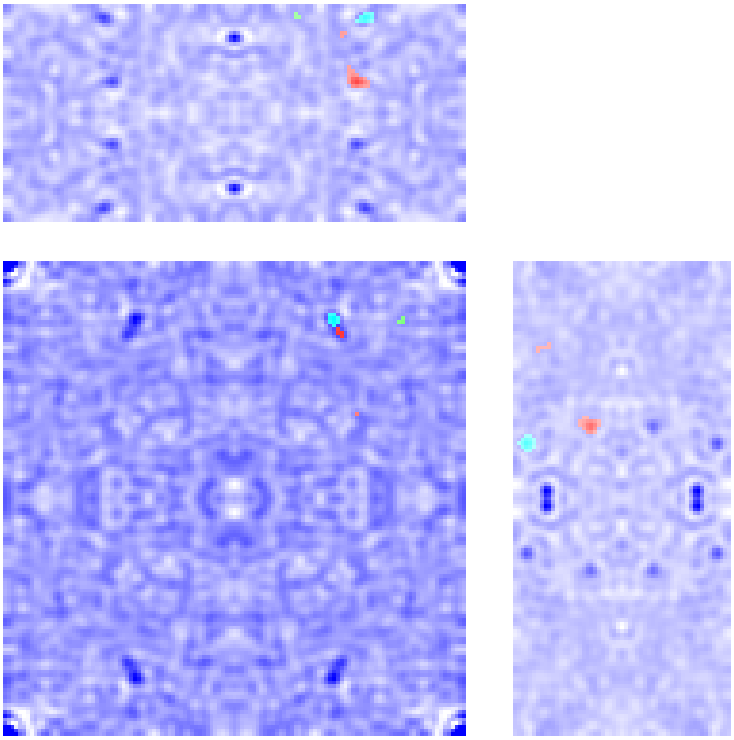


Figure 2.2: Harker sections of the *Bse*634I Patterson map, processed with `hara` program.

spacegroups this dramatically reduces the amount false positive peaks, leaving the correct answer on top of the list (fig. 2.3). Harker sections are written out as image files and immediately displayed on the X-window screen. This facility, although somewhat rudimentary, facilitates identification of peaks on the section and their corresponding Harker peaks (fig. 2.2).

2.2.6 Model building and refinement

The phases from four derivatives gave an interpretable map after the solvent flattening, into which 3/4 of the model could be build with the `O` molecular modelling program [52]. The model was then transferred to the *native 1* dataset (see table 3.1) using `CNS` [16] system. In the successing cycles of building and refinement, the rest of the model became visible in the σ_A weighted $2F_o - F_c$ electron density maps. The new parts of the model have been build into the model only when the density of these parts was well defined in the maps phased with the truncated model. Finally, ions and water molecules have been added.

```
# FILE T3D3-T10D1-3.0.map
# CELL 122.22 124.61  57.51  90.00  90.00  90.00
# GRID 120 128 56
# SPACEGROUP P21212
# PATTERSON  Pmmm
# AU                  0        1           0     0.25           0        1
# PATTERSON AU        0      0.5           0      0.5           0      0.5
# threshold for section 0: 25.996574 ( 3.7 sigma )
# threshold for section 1: 21.764631 ( 3.9 sigma )
# threshold for section 2: 31.813953 ( 5.6 sigma )
# number of hypotheses on section Z   105
# number of hypotheses on section Y    91
# number of hypotheses on section X     4
#
# confirmed hypotheses and their confirming peaks
#
#        x        y        z colour
#
 0.06802  0.06122  0.97619 green
 0.11667  0.16016  0.93455 red
 0.13539  0.07425  0.82702 magenta
 0.14217  0.05994  0.97176 cyan
```

Figure 2.3: A sample output of `hara` program.

2.2.7 Accessible surface calculations

All accessible surface areas have been calculated using `naccess` program [44]. For calculation of the buried surface between any two chains X and Y the `naccess` has been run 3 times – once for each individual chain in a separate file, and once for the file with the both chains. The buried area was calculated as $X+Y-XY$, where X and Y are surfaces of the separated chains X and Y, and XY is the area of the two-chain complex.

To compare the areas with the fit in [46], the buried surface area has been calculated for ribonuclease (PDB entry `8rsa`) using the same procedure. The procedure yielded contact surface 1812 Å^2, which differs less than 1 % from the calibration value 1795 Å^2 from [46]. The difference was assumed to be insignificant for the estimation of the contact probabilities and therefore no extra normalisation was applied to the values calculated by `naccess`.

2.2.8 Model analysis

To compare with another restriction endonucleases, the active centre residues of the *Bse*634I model have been superimposed as rigid bodies with the active centres of another restriction endonucleases using Kabsch's method [53]. For convenient access to the algorithm, it has been coded in `Perl` programming language [107].

Chapter 3

Results and discussion

3.1 Results on *Bse*634I

3.1.1 Crystals of *Bse*634I and diffraction data

*Bse*634I restriction endonuclease was crystallised using preliminary screening with house factorials or factorial set from Hampton Research. Both screenings led to the similar crystallisation conditions, with just MES buffer (from house factorials) substituted for cacodylate buffer (from Hampton Research factorials). The big crystals could be obtained, however, only after introduction of acetate buffer (fig. 3.1).

*Bse*634I crystals could also be obtained from basic conditions (fig. 3.1). Their diffraction, however, was much weaker than of crystals from acetate condition, so the crystals from acetate buffer were chosen for structure determination.

Attempts to screen for the crystals of *Bse*634I-oligonucleotide complex yielded only very small crystals under conditions similar to apo-protein crystallisation conditions. Crystals with cognate oligonucleotide could be refined to the size where X-ray measurement was possible, but they diffracted to a resolution below 8 Å and could not be indexed reliably.

The crystals of the apo-*Bse*634I were stable enough to transfer them to cryoprotectant buffer and freeze, and withstanded treatment with heavy atom compounds that enabled to screen for heavy atom derivatives. In the end, four isomorphous derivatives were found that made structure solution possible. The apo-*Bse*634I crystals diffracted to 3.0 Å on the rotating anode source/IP detector, and to 2.17 Å on the BW6/DESY (Hamburg) with IP detector, where a dataset for final refinement was collected. The resulting crystals belong to orthorhombic crystal class (space group $P2_12_12$)

Crystal	**native 1 (T10A2)**	**native 2 (T3D3)**	**$GdCl_2$ (T20A1)**	**$HgCl_2$ (T10D1)**	**AMMA[a] (T22C5)**	**cis-platin[b] (T20A3)**
Spacegroup, all datasets	$P2_12_12$					
Unit cell, Å	$a = 121.2$ $b = 122.3$ $c = 56.9$	$a = 122.2$, $b = 124.6$, $c = 57.5$, $\alpha = \beta = \gamma = 90°$				
Max. resolution, Å	2.17	3.30	3.17	3.30	4.00	2.56
R_{merge}[c]	0.055	0.101	0.255	0.150	0.135	0.096
R_{iso}[d]	-	-	0.194	0.247	0.268	0.203
Unique reflections	43316	13748	15017	15224	7385	27784
Completeness, %	95.0	99.7	96.5	97.9	93.1	93.0
Last resolution shell starts at, Å	2.23	3.40	3.26	3.40	4.12	2.64
Completeness, %, in the last resolution shell	87.2	99.3	77.9	78.9	92.9	46.8
R_{Cullis}[e], centric reflections	–	–	0.75	0.70	0.68	0.70
R_{Cullis}[e], acentric reflections	–	–	0.80	0.66	0.59	0.74
Phasing power[f], centric reflections	–	–	0.90	1.29	1.34	1.07
Phasing power[f], acentric reflections	–	–	1.16	1.89	2.17	1.28

[a] 4,6-*bis*-(acetoxymercury)-2-methyl-aniline

[b] *cis*-$[PtCl_2(NH_3)_2]$

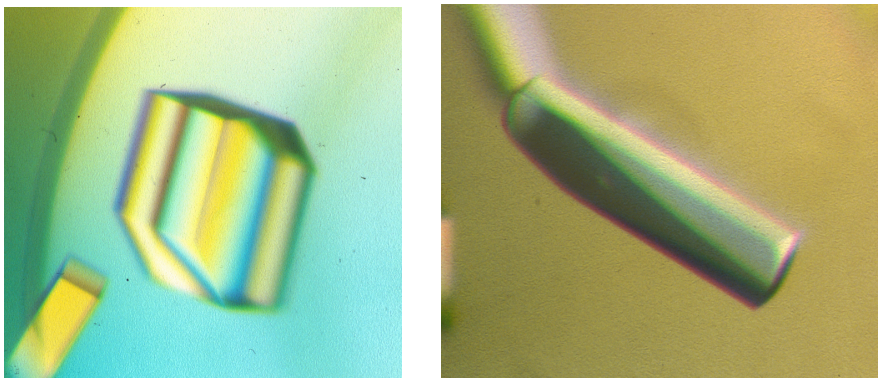
[c] $R_{merge} = \sum_{\vec{h}} \sum_{i=0}^{n_{\vec{h}}} (\langle I_{\vec{h}} \rangle - I_{\vec{h}i})^2 / \sum_{i=0}^{n_{\vec{h}}} I_{\vec{h}i}^2$, where $I_{\vec{h}i}$ is the intensity value of the i-th measurement of the reflection $h = (h, k, l)$, and $\langle I_{\vec{h}} \rangle$ is the average measured intensity of the reflection $\vec{h}$. $n_{\vec{h}}$ is the number of measurements of reflection $\vec{h}$.

[d] $R_{iso} = \sum_{\vec{h}} |F_{\vec{h}p} - F_{\vec{h}q}| / \sum_{\vec{h}} |F_{\vec{h}p}|$, where $F_{\vec{h}p}$ and $F_{\vec{h}q}$ are measured native and derivative structure factors, respectively.

[e] $R_{Cullis} = \sum_{\vec{h}} |F_{\vec{h}(obs)} - F_{\vec{h}(calc)}| / \sum_{\vec{h}} |F_{\vec{h}(obs)}|$

[f] Phasing power = $\langle |F_{\vec{h}(obs)}| \rangle$/r.m.s.d.$\epsilon$, where ϵ is lack of closure.

Table 3.1: Data collection and phasing statistics.



(a) from Na Acetate buffer (pH = 5.5)

(b) from Tris buffer (pH = 8.5)

Figure 3.1: apo-*Bse*634I crystals.

3.1.2 Phasing of *Bse*634I

Since *Bse*634I shares 30% identity (50% similarity) with *Cfr*10I restriction endonuclease, it was expected that the molecular replacement with the coordinates of *Cfr*10I [15] would give a good starting phases for the structure interpretation. Attempts to find a molecular replacement solution with several molecular replacement packages, however, did not bring clear signal of model orientation. In retrospect, the reason of this failure is clear – *Bse*634I protein appears to consist of two subdomains that can move with respect to each other. Two polypeptide chains in the asymmetric unit of *Bse*634I crystal differ from each other and from the known structure of *Cfr*10I by the rotation of the subdomains that leads to a poor overlap of unmodified *Cfr*10I model with *Bse*634I chain. Since there was only one chain in the asymmetric unit of the *Cfr*10I crystal, the position of the subdomain boundary was unknown.

The *Bse*634I diffraction data could be phased using MIR method. Initially, all heavy atom derivatives were used to calculate difference Patterson maps with the native dataset, and the Harker sections of the obtained maps were analysed for the possible peaks of heavy atoms.

From 26 measured heavy atom soaks, two mercury derivatives gave strong signals that allowed to identify four mercury sites in each derivative. The sites were refined and SIR phases from these sites were calculated. These phases were used to search for additional mercury sites from difference Fourier synthesis, and to re-evaluate all other derivatives. Two other derivatives, cis-platin and $GdCl_2$, contained heavy atom sites that could be used for phasing. The final refined heavy atom positions refined with `mlphare` [80] and used for phasing are given in

$GdCl_2$ derivative:

```
Gd1 0.748 0.076 0.361 0.675 0.642 BFAC 126.411
Gd2 0.000 0.000 0.165 0.446 0.326 BFAC 273.175
Gd3 0.261 0.043 0.792 0.660 0.524 BFAC 92.953
Gd4 0.000 0.000 0.984 0.418 0.394 BFAC 296.054
Gd5 0.079 0.096 0.649 0.176 0.082 BFAC 52.536
Gd6 0.931 0.110 0.503 0.247 0.083 BFAC 76.449
Gd7 0.193 0.156 0.560 0.314 0.266 BFAC 131.093
Gd8 0.282 0.244 0.667 0.235 0.103 BFAC 133.086
```

$HgCl_2$ derivative:

```
Hg1 0.866 0.075 0.177 0.852 0.562 BFAC 57.026
Hg2 0.142 0.058 0.972 0.888 0.542 BFAC 57.499
Hg3 0.419 0.098 0.027 0.654 0.677 BFAC 229.347
Hg4 0.593 0.142 0.135 0.487 0.486 BFAC 185.339
```

AMMA derivative:

```
Hg1 0.142 0.058 0.972 0.857 0.541 BFAC 44.991
Hg2 0.866 0.075 0.178 0.768 0.502 BFAC 38.884
Hg3 0.417 0.098 0.028 0.755 0.719 BFAC 248.225
Hg4 0.595 0.143 0.135 0.486 0.400 BFAC 180.057
```

cis-Pt derivative:

```
Pt1 0.123 0.029 0.210 0.668 0.545 BFAC 190.112
Pt2 0.880 0.043 0.938 0.608 0.512 BFAC 171.229
Pt3 0.405 0.175 0.081 0.825 0.760 BFAC 249.762
Pt4 0.615 0.208 0.066 0.408 0.327 BFAC 157.660
Pt5 0.306 0.058 0.095 0.342 0.387 BFAC 179.660
Pt6 0.091 0.127 0.544 0.188 0.178 BFAC 63.961
Pt7 0.125 0.267 0.916 0.587 0.565 BFAC 305.440
Pt8 0.703 0.094 0.047 0.481 0.297 BFAC 219.975
Pt9 0.923 0.135 0.597 0.194 0.166 BFAC 69.940
```

Figure 3.2: Heavy atom positions, estimated from difference Patterson and difference Fourier maps and refined with `mlphare` [80].

RMS bond length deviation	0.008 Å
RMS angle deviation	1.4°
R_{cryst}	0.218
R_{free}	0.252
Test set size	10% reflections, randomly selected
Number of solvent molecules	288

Table 3.2: Refinement statistics of the *Bse*634I model.

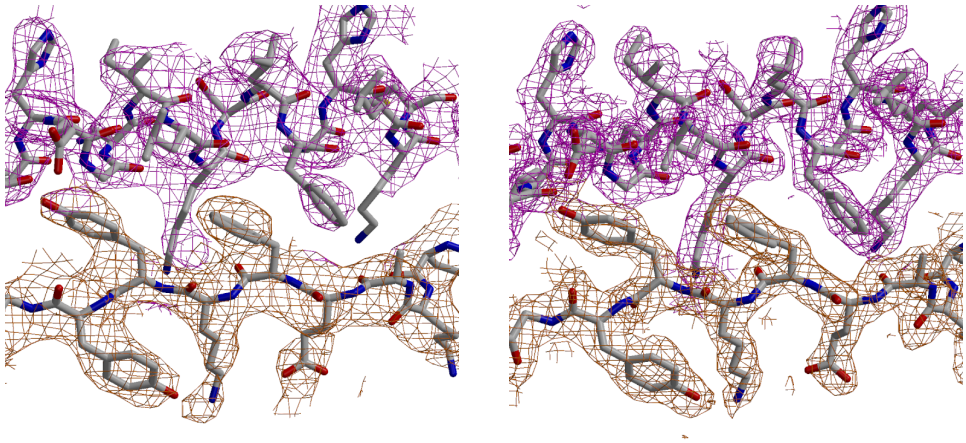
fig. 3.2. The MIR electron density calculated from these derivatives and solvent-flattened with `dm` [25] could be easily interpreted (fig. 3.3a).

3.1.3 Quality of the model

*Bse*634I sequence has 293 amino acids; the first methionine residue is not present in the expressed protein. 288 residues could be build into the final electron density map. The missing ones are three residues from the N-terminus (Met-Thr-Thr) and two C-terminal lysines.

Electron density in the region of the loop 15–20 in the subunit A is very poor, the main chain can hardly be traced. On the contrary, in the subunit B the same loop has a contact with the symmetry equivalent molecule $(X, Y, Z-1)$ and is rather well defined. The A15–A20 loop has been modelled after its NCS equivalent and included into the final model for refinement. This improved R–factors, although B–factor of the loop is very high, indicating increased mobility of the loop. Final refinement statistics is shown in table 3.2.

The final model has 288 water molecules, three acetate anions and 5 ions modelled as Cl^- because of their positively charged/hydrophobic environment.



(a) Experimental MIR electron density after solvent flattening with DM

(b) σ_A weighted final electron density

Figure 3.3: Electron density maps with *Bse*634I model.

3.1.4 Architecture

*Bse*634I chain is folded into a globule with an α/β structure of approximately 66×57×48 Å (fig. 3.4). Asymmetric unit of the crystal contains two chains A and B, related by a nearly twofold non-crystallographic axis (fig. 3.8). They have an extensive contact surface of 3103.1 Å^2 and build up an U-shaped dimer with a 30 Å wide cleft. This cleft that can accommodate a B-form DNA molecule, either undistorted or with a kind of distortions observed in the *Eco*RI–DNA or *Eco*RV–DNA complexes.

Two dimers in a unit cell related by a twofold crystallographic axis $(-X+1, -Y+1, Z)$ are arranged "back-to-back" with their DNA binding clefts facing the opposite directions (fig. 3.6). They form a second largest contact observed in the unit cell which might have a physiological significance.

3.1.5 Subdomains

A comparison of the two NCS related molecules shows that each monomer of the *Bse*634I consists of two subdomains: N-terminal subdomain (residues 1 to 89, helices α_1–α_3 and strands β_1 and β_2), and C-terminal subdomain (residues 90 to 293, helices $3_{10}1$–α_8 and strands β_3–β_7). The C-terminal domain contains the central β-sheet (β_3–β_7), which is flanked by the helices α_8 and the C-terminus of the helix α_3 on one side and with the short helix $3_{10}1$ and the helix α_6 on the other. The N-terminal domain contains the short two-stranded β-sheet and two

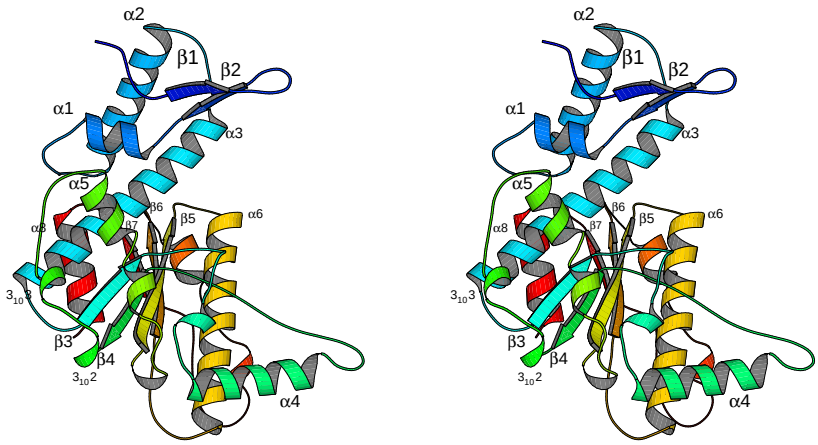


Figure 3.4: Stereo view of *Bse*634I monomer.

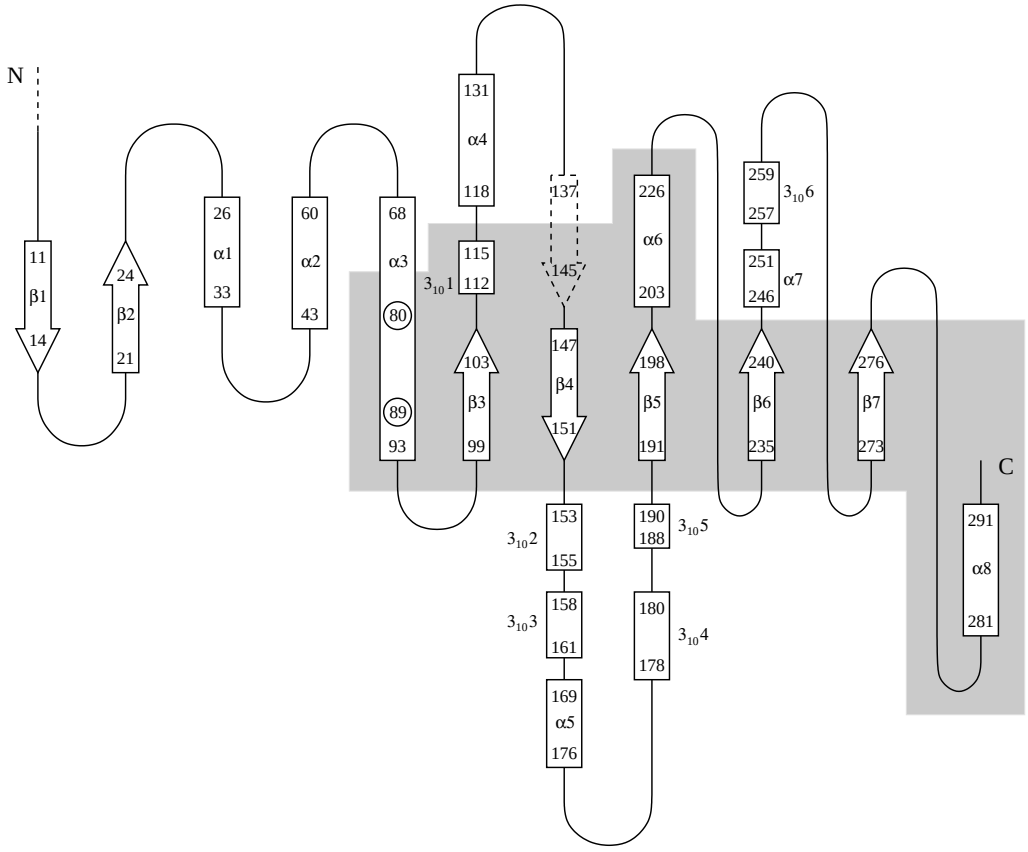


Figure 3.5: *Bse*634I restriction endonuclease topology diagram.

Figure 3.6: *Bse*634I tetramer.

helices $\alpha 1$ and $\alpha 2$. The corresponding domains of the different monomers can be overlayed fairly well (the best RMS deviations are 1.1 Å(all atoms)/0.61 Å(C_α) for C-terminal domains and 1.0 Å(all atoms)/0.61 Å(C_α) for N-terminal domains). However, the whole chains A and B of the *Bse*634I can be overlayed only with the best RMS deviations 1.7 Å(all atoms)/1.5 Å(C_α). Thus, the superposition of each domain individually can be performed with significantly better accuracy as of the monomers themselves.

The improvement of the domain superposition is not just an effect of the domain length. When we repeated the superposition of the *Bse*634I subchains of different length between the monomers A and B, we could see that any N-terminal subchain starting from the beginning of the protein and extending no longer than residue 90 down the sequence can be superimposed as good as the whole N-terminal domain (RMS of C_α 0.6 Å or better). When we take a longer N-terminal subchain, however, the RMSD of the best superposition rises sharply to 1.4–1.5 Å, the value close to that of the whole chain (fig. 3.7, curve "N-term"). The same behaviour is observed for the C-terminal subchains (fig. 3.7, curve "C-term"). It indicates that there is a hinge in the *Bse*634I structure located between residues 70 and 90 in the sequence.

Thus the N-terminal domains appear to be rotated for $\approx 10^\circ$ around an axis that passes through C_α atom of the residue Asn89 in the helix α_3, in good agreement with the hinge position estimated above. We suppose that the two rigid domains in the *Bse*634I are connected by a relatively flexible joint located at the residue Asn89. Although the domain movement in the *Bse*634I structure

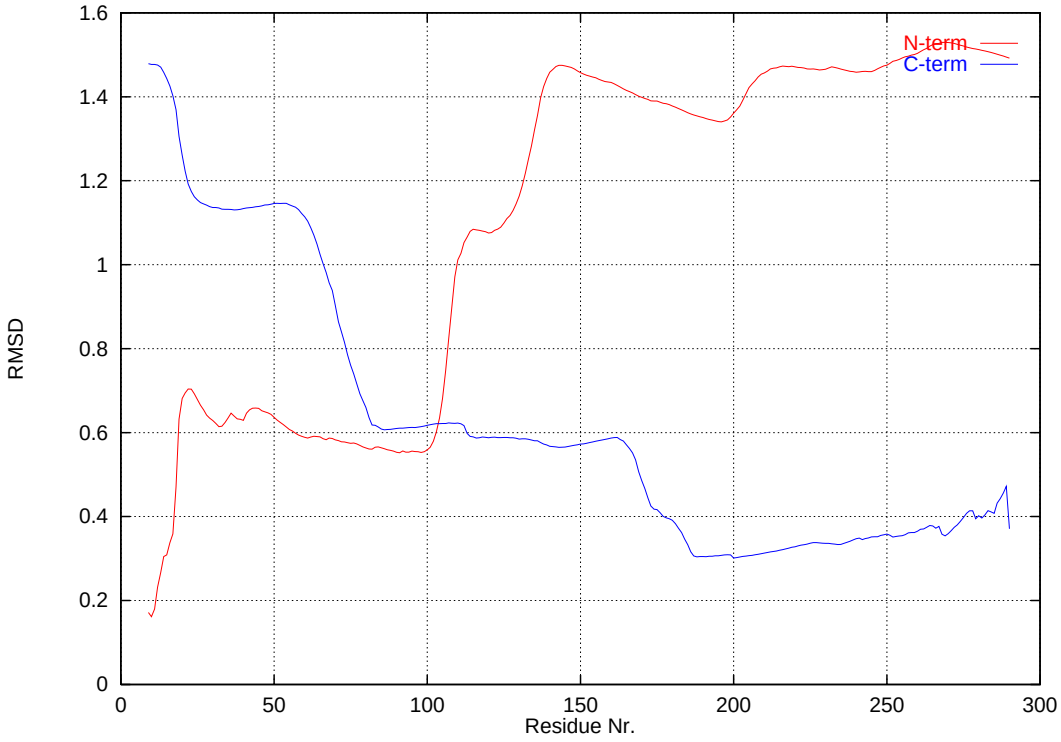


Figure 3.7: RMS deviations for C- and N-terminal subchain superpositions between the monomers A and B of the *Bse*634I restriction endonuclease.

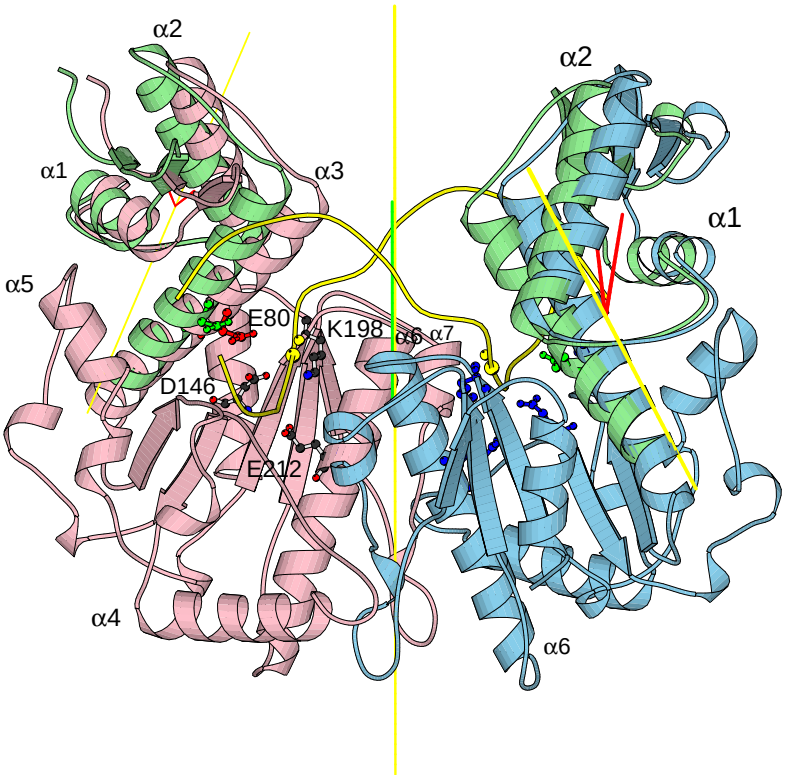


Figure 3.8: A *Bse*634I dimer with a predicted DNA position.

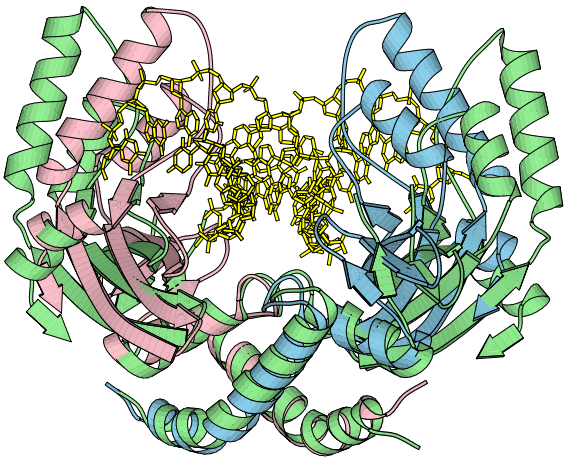


Figure 3.9: Domain movements in *Pvu*II restriction endonuclease.

without DNA is clearly induced by the crystal packing forces, it indicates that the two domains can be easily rotated with respect to each other, so that even weak lattice interactions are sufficient to move them, while the domains themselves preserve their conformation. One can anticipate that the specific interactions with the target DNA will induce even larger rotations around Asn89, and possibly some other changes which we do not see in the present crystal structure.

Glu80 in the helix α_3 of the N-domain is homologous to the Mg^{2+}-complexing residues Glu45 in *Eco*RV and Glu71 in *Cfr*10I (see table 1.5). The rest of the assumed Mg^{2+} binding residues and active centre residues are located on the C-terminal domain. In the observed motion of the domains in the *Bse*634I structure, the C_α atom of the Glu80 moves 2.3 Å, and C_δ moves about 3 Å. Upon binding of a specific DNA the movement could be even larger — we expect that specific interaction of the protein with its substrate have larger energies that lattice interactions. Such movement could reconstitute the Mg^{2+} binding site of the restriction endonuclease and trigger Mg^{2+} binding and subsequent DNA cleavage.

The conformational changes of the same kind are also observed in other restriction endonucleases. *Pvu*II restriction endonuclease, for example, exhibits

transition from an "open" conformation observed in apo-enzyme [6] to a "closed" DNA-bound form [40] upon DNA binding. The subdomains of the *Pvu*II rotate about 25° upon DNA binding. Comparing with the *Pvu*II we can say that the chain B in the *Bse*634I structure has an "open" conformation, while the chain A is on the way to the "closed" conformation (distance of the chain A Gly69 to the dimer axis is 7.5 Å, while distance of the chain B Gly69 to the dimer axis is 11.2 Å).

The comparison of the the *Eco*RV structures in its free and DNA-bound state show that there is a motion of two flexibly linked structural subdomains [109].

The changes of *Bse*634I and *Pvu*II are in great contrast with the structural rearrangements observed in the *Bam*HI and *Eco*RI structures. In the DNA bound structure of the *Bam*HI authors observe a 19° rotation of the whole subunits [75]. One can speculate that this rotation plays the same role as the subunit movement in *Bse*634I, in each case narrowing the DNA-binding cleft of a protein and enabling specific DNA–protein contacts that otherwise could not be formed. Comparing the structures of the *Eco*RI enzyme with and without DNA (PDB entries `1qc9` and `1qri`), we could not observe any significant subdomain movement. It can be that for *Eco*RI such movement is not necessary, because the enzyme distorts DNA strongly upon binding and that distortion alone could be sufficient for building all hydrogen bonds required for discrimination of the cognate DNA site.

The central core of the C-terminal subdomain in the *Bse*634I is well conserved over all restrictions endonucleases with known structures. On the contrary, N-terminal subdomain in some restriction endonucleases differs significantly.

3.1.6 Structural comparison of *Bse*634I and *Cfr*10I

The structural correspondence of the *Bse*634I and *Cfr*10I restriction endonuclease is depicted in the fig. 3.10.

*Bse*634I shares 30% sequence identity and 50% similarity with its isoshisomer *Cfr*10I. According to [1], this indicates that the two proteins must have the same fold. Indeed, their structures appear in overall very similar. However, there are also several regions in *Bse*634I that overlap poorly with the corresponding regions in *Cfr*10I. Notably, β–strands β_1 and β_2 in *Bse*634I each are one residue longer than the corresponding elements of *Cfr*10I, and the l_{15-20} that joins them in *Bse*634I extends past the corresponding *Cfr*10I loop l_{7-12}. This loop reaches the major groove of the DNA in the modelled *Bse*634I-DNA complex and could be involved in the DNA contacts. Another non-overlaping region in the N-terminal subdomain of *Bse*634I is a turn region 34–40 (residues 26–32 in *Cfr*10I).

In the C-terminal subdomain, a rather large region of poor overlap extends from residue 152 to 167 (the corresponding region in *Cfr*10I spans residues 139 to 159). Structural correspondence in this region of the two proteins is much worse than in the rest of C-terminal subdomains. The region has only two short 3_{10}

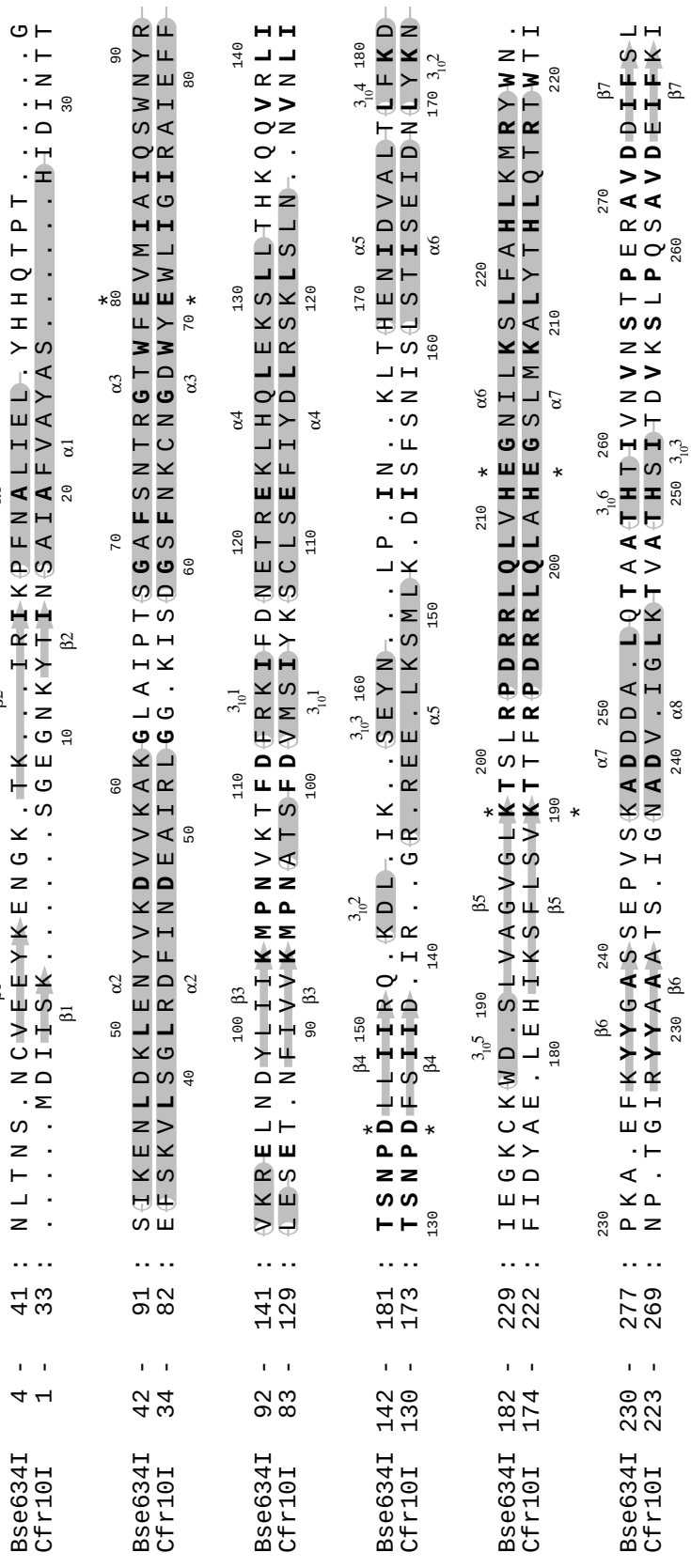


Figure 3.10: Correspondence between the sequences and secondary structure elements in *Bse*634I and *Cfr*10I restriction endonucleases.

helices and is otherwise in extended or loop conformation.

The two mentioned regions in the of the *Bse*634I restriction endonuclease are on the surface of the protein and form a contact between the C- and N-terminal subdomains of the protein. Both have increased mobility as indicated by the higher B-factors as compared to the other surface residues of the *Bse*634I. The explanation could be that these regions must accommodate the movement of the N-terminal subdomain of the *Bse*634I and therefore are more flexible and have less secondary structure than the rest of the *Bse*634I surface regions. It can also happen that these regions will become more ordered after DNA binding, as observed in the structure of *Bam*HI [75].

*Bse*634I sequence is 8 residues longer than *Cfr*10I (293 a.a. instead of 285), with 4 residue extension at N-terminus and several loops that are longer by one or two residues. The rest of the differences is accounted by 5 insertions each 1–2 residues long in *Bse*634I and 3 insertions in *Cfr*10I (fig. 3.10) that completely account for the difference in chain length.

From the similarity of the *Bse*634I and *Cfr*10I structures we can predict that the same kind of domain rotation could be found in *Cfr*10I. The N–terminal subdomain of the *Cfr*10I would extend from residues 1 to Glu80 — a structural counterpart of the *Bse*634I's Asn89.

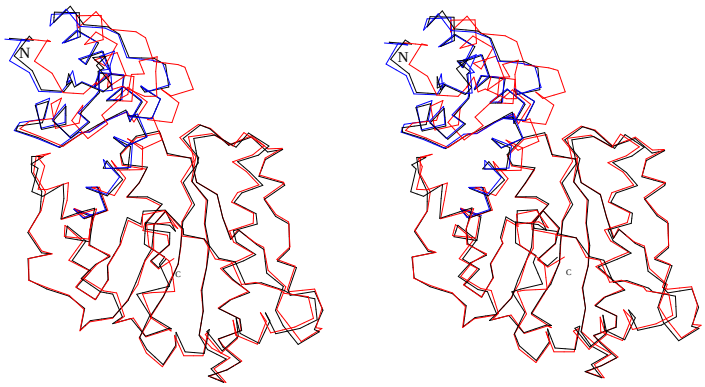
Indeed, the N-terminal subdomains of the *Bse*634I and *Cfr*10I can be superimposed only after rotation of the N-terminal subdomain of *Cfr*10I relative to its C-terminal subdomain. The necessary rotation is 9° to superimpose with subunit A of *Bse*634I and 13° to superimpose with subunit B; the rotation axis in both cases passes through the middle of the respective helices α_3.

When two different symmetry operators are applied to N- and C-terminal subdomains of *Cfr*10I, these subdomains can be overlayed onto the *Bse*634I domains with RMS deviations 1.1 Å for N-terminal domain and 1.3 Å for C-terminal domain[1]. If, however, the same residues are used for superpositions of the whole proteins, the RMS deviation is 2.0 Å. As in the case of the two subunits of *Bse*634I, this fact can be interpreted as a presence of two subdomains in *Cfr*10I protein that are structurally equivalent to the domains observed in *Bse*634I. This domain movement could not be observed in the structure of the *Cfr*10I, because there was only one molecule of *Cfr*10I in the asymmetric unit of the crystal.

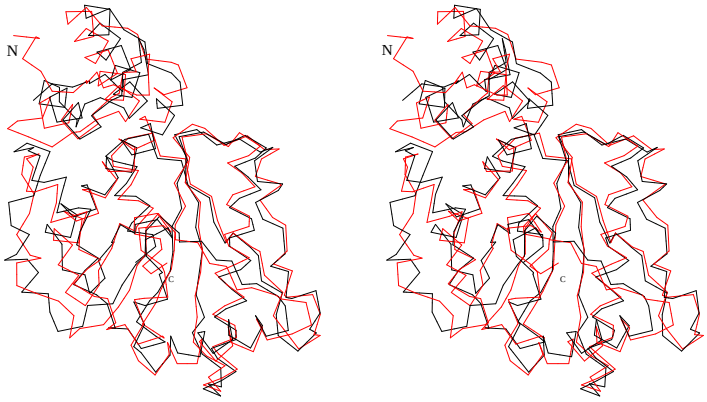
3.1.7 Oligomerisation state

Type II restriction endonucleases have been considered to be homodimers since the dimer model has been first proposed by Kelly & Smith [55]. While many of them indeed form dimers in solution and interact as dimers with DNA, there is a growing number of restriction enzymes that are tetrameric and require forma-

[1] Only the structurally equivalent residues which are shown as aligned in fig. 3.10 have been used for the superpositions of both proteins.



(a) Overlay of *Bse*634I subunits A and B. N-terminal subdomain is highlited in blue



(b) Overlay of *Bse*634I and *Cfr*10I

Figure 3.11: C_α chain overlays, *Bse*634I subunits A and B (upper) and *Bse*634I and *Cfr*10I (lower).

tion of the tetramer for their biological activity. One example of the tetrameric restriction endonuclease is the *Sfi*I [108]. For the *Cfr*10I enzyme, the evidence for tetramer formation has been obtained using accessible surface area calculation from the crystal structure, cleavage rate analysis, site-directed mutagenesis, equilibrium sedimentation, gel filtration and electron microscopy [92]. Because *Bse*634I enzyme has the structure close to that of the *Cfr*10I, one would expect that it's physiologically active form is also a tetramer. Indeed, the calculation of the accessible surfaces in the crystal structure also supports the hypothesis of the tetrameric state for *Bse*634I. As shown in the table 3.3, a contact between subunits A and B (B is NCS related to A) in the *Bse*634I dimer buries a surface of 3103 Å^2. The contacts between subunit A and it's symmetry related chains C and D buries surfaces 822.2 Å^2 and 1005.4 Å^2, respectively. According to [46], the size of the area indicates that first contact (A–B) is due to specific protein-protein interaction, and A–C the contact is most probably also specific. If we take into account that the chains C and D also form a dimer, like their crystallographic counterparts A and B, and therefore interact with AB dimer as a whole, we see that the supposed tetramer interface buries $822.2+1005.4 = 1827.6$ Å^2 per chain, which is highly probably specific protein contact and not a crystal packing artefact. Other contacts in the crystal of the *Bse*634I are by a factor of 2 or more smaller than the assumed tetramer contact. The largest of these contacts has 851.2 Å^2 and can be attributed to non-specific crystal packing contacts in accordance with observations in [46]. The equilibrium sedimentation data (not shown) are also compatible with the hypothesis that the *Bse*634I enzyme exists as a tetramer in solution.

3.1.8 Interfaces

The N- and C-terminal subdomains interface with each other along the helix α_3, which contacts the C-subdomain at β_3, β_4 and α_8. The C-terminal part of the α_3 is additionally embraced by the helix $3_{10}3$–loop$_{161-167}$–α_5 on one side, and by the helix α_8 on the other.

The dimerisation interface consists of the helices α_6 and $3_{10}6$ in the central part of the dimer, and the α_4 together with the loop 132–141 that flanks the dimer from the side (fig. 3.12). Of note is that structurally equivalent helices are located at the dimerisation interfaces of other restriction endonucleases.

The putative tetramerisation interface consists of the C-terminal end of the helices α_6 and the loop 260–264 from all four tetramer chains.

3.1.9 Active centre architecture

Correspondence between the Mg^{2+} binding and active site residues in the known structures of restriction endonucleases is given in table 1.5. Residues Pro145, Asp146, and Lys198 from the weak active site signature of restriction endonucle-

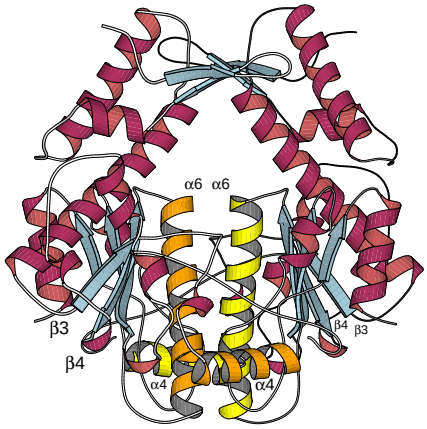


Figure 3.12: *Bse*634I dimer. Dimerisation elements are highlighted in yellow and orange.

ases [4] overlay well with the Mg^{2+} binding site residues of the other restriction endonucleases (RMSD from maximum 1.5 Å for *Eco*RV to minimum 0.5 Å for *Mun*I, see fig. 3.13 and 3.14). In addition, upon the superposition of Pro145, Asp146 and Lys198 the residue Glu80 overlaps spatially with the Glu71 in *Cfr*10I, Glu70 in *Ngo*MIV and Glu45 in *Eco*RV, although it has not been included in the calculation of the superposition operator. It has been shown [90] that Glu45 plays important role in catalysis in *Eco*RV, and the similar role can be suggested for the structurally equivalent *Bse*634I residue Glu80.

The *Bse*634I Glu212 residue overlaps with the Glu204 in *Cfr*10I (fig. 3.13) and Glu201 in *Ngo*MIV (fig. 3.14). In *Cfr*10I, Glu204 has been shown to be a structural counterpart of the Asp90 in *Eco*RV, although it comes from the different part of the sequence [93] – the "swap" mutation of the *Cfr*10I S188EE204S that puts Glu204 into the "correct" sequence location (as judged from the sequence alignment with *Eco*RV) retains significant *Cfr*10I activity. Thus, Glu212 from *Bse*634I might also be involved into catalysis.

The observed structural similarities allow one to conclude that residues Glu80, Asp146, Glu212 and Lys198 from *Bse*634I most likely form a Mg^{2+} active centre and binding site in the *Bse*634I restriction endonuclease.

Superposition of these residues with the corresponding residues in the structures of other restriction endonucleases brings helices α_3 and α_6, and strands β_3–β_7 from *Bse*634I to close proximity of the corresponding structural elements in another restriction endonucleases. The correspondence is not perfect — in some structures, some elements are missing or are displaced by 3–5 Å. In *Eco*RI, for example, the structural counterpart of the *Bse*634I helix α_3 is missing (α_2 in *Eco*RI that could be in the same spatial position is more than 6 Å away). In general, however, "common core motif" corresponding to the elements α_3, α_6 and

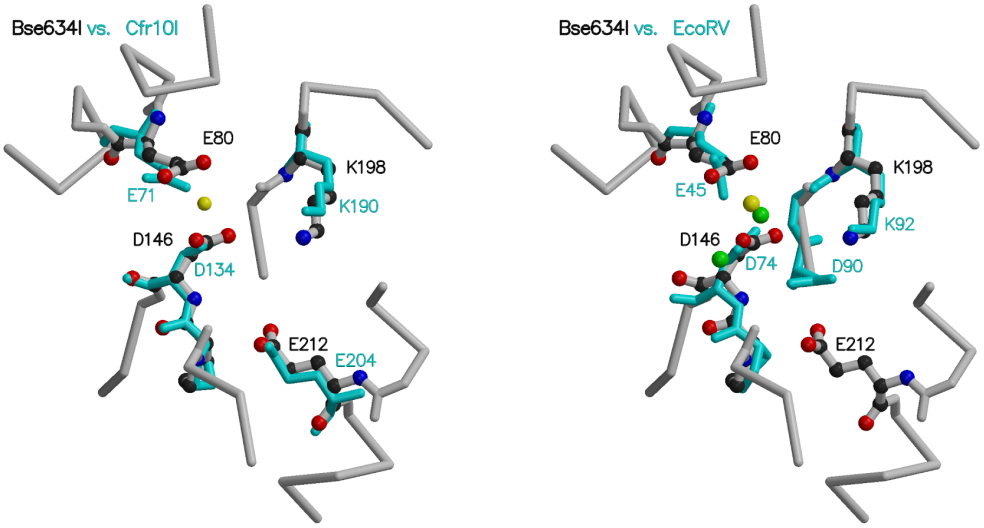


Figure 3.13: Active centre of the *Bse*634I restriction endonuclease compared with active centres of *Cfr*10I and *Eco*RV.

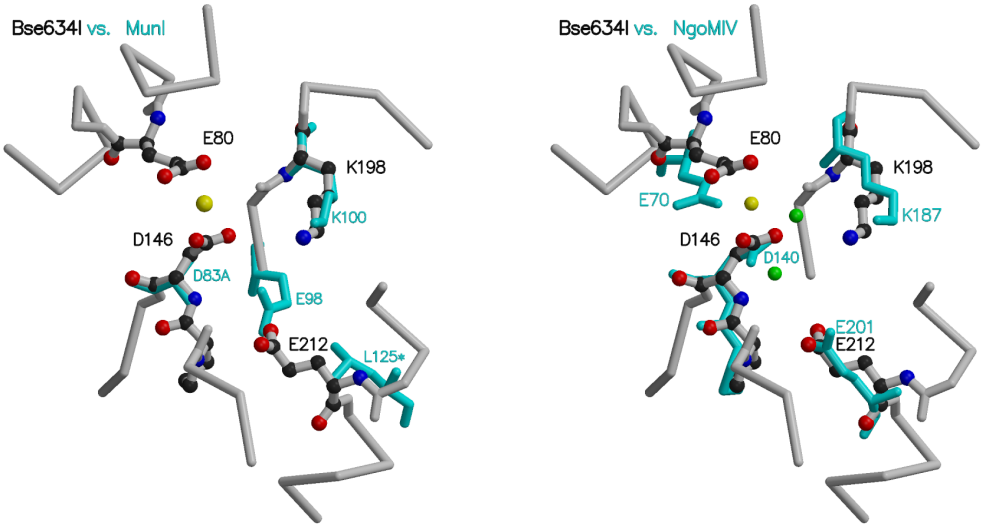


Figure 3.14: Active centre of the *Bse*634I restriction endonuclease compared with active centres of *Mun*I and *Ngo*MIV.

β_3–β_7 is present in all known structures of restriction endonucleases (fig. 3.5, on gray background).

There is no density that could be interpreted as metal ion in the active site region. In the subunit A, the two active centre residues Asp146, Lys198 are within the hydrogen bonding distance from the water molecule W45; Glu80 is 4.1 Å away. In the B subunit there is no equivalent for this water molecule, probably because the B subunit is found in an "open" conformation. If we superimpose the active site of the *Eco*RV restriction endonuclease with our protein, the Mg^{2+} ion MG2 of the *Eco*RV structure (PDB entry `1rvb`) is positioned close to the water W45 (within 2.2 Å). This location could be occupied by Mg^{2+} ion in the *Bse*634I–DNA complex.

In the gadolinium heavy atom derivative, the Gd^{2+} ion is complexed by the side chains of the Glu80 and Asp146. The position of the ion is spatially equivalent to the position of MG2 in the structure of the *Eco*RV restriction enzyme. This could be the position of the second Mg^{2+} ion. It should be noted, however, that the distance between the water molecule W45, the presumed site for the first Mg^{2+} ion, and the Gd^{2+} is 2.9 Å – nearly twice smaller than observed in the structure of the *Eco*RV (5.4 Å in PDB entry `1rvb`, 6.1 Å in PDB entry `1rvc`). Thus we can not tell from the presented structure whether the *Bse*634I can bind two metal ions, and whether the two ions are involved in catalysis. The structure of the *Bse*634I with DNA could be helpful to answer this question.

3.1.10 DNA binding model

The structure of the *Bse*634I does not contain the DNA molecule, so we can not derive the exact atomic coordinates of DNA bound it. Moreover, the previous structures show that upon binding to some restriction endonucleases DNA B-form structure is distorted [32], while some others preserve nearly unchanged the B-DNA helix [18, 75]. It is hard to say at present which is true for the *Bse*634I.

We can, however, predict an approximate position of the DNA molecule in the *Bse*634I dimer. If the cleavage of the DNA happens at the same position for *Bse*634I and for another restriction endonuclease, the DNA should maintain about the same position relative to the dimer's active centre residues, and therefore should be positioned reasonably in a cleft of the enzyme. In this case, the DNA dyad axis by necessity coincides with the axis of the protein. Indeed, when we superimpose the active centres from two dimer chains of 5'-overhang producing restriction endonucleases and apply the resulting operator to the foreign DNA molecule, it is positioned in the dimer cleft of *Bse*634I restriction endonuclease, with the secondary structure elements fitting quite well into the grooves of the DNA (see table 3.4, fig. 3.8).

From the restriction nucleases for which the 3D structure with DNA is known, three of them (*Eco*RI, *Mun*I and *Ngo*MIV) leave 5'-overhanging ends of four nucleotides like *Bse*634I. A superposition of their active centres with that of the

Chain	**Buried area, Å^2**	**Notes**
A	15143.0	} Surface of each monomer
B	15521.0	
$A + B$	30664.0	
AB	27560.9	Surface of the dimer
$A + B - AB$	3103.1	Buried surface in the dimer interface
C	15143.0	Same as A
$S = A + C - AC$	822.2	"Small" tetramer interface
D	15521.0	Same as B
$L = A + D - AD$	1005.4	"Large" tetramer interface
$S + L$	1827.6	
T	51723.6	Surface of the whole tetramer
$2(A + B)$	61328.0	
$2AB$	55121.8	Surface of the two dimers
$2AB - T$	3398.2	Buried surface in the tetramer interface
$AB - T/2$	1699.1	Buried surface in the tetramer interface, per chain
$2(A + B) - T$	9604.4	Total buried surface of the tetramer, buried by the dimer the and tetramer interfaces
$B_1 + B - B_1B$	851.2	Surface buried by crystal neighbour B_1 (obtained from the chain B the by the symmetry operator $-X, -Y + 1, Z$)

Table 3.3: Accessibilities of different chains in the *Bse*634I restriction endonuclease tetramer, as computed by the `naccess` program [44]. T stands for tetramer. $A + B + C + D = 2(A + B)$. In the tetramer, chain $A \equiv C$ and chain $B \equiv D$.

Structural element	**Possible contacts with DNA**
β_1–l_{15-20}–β_2	major groove
α_3	minor groove
α_6	major groove
β_5	major groove, P-backbone

Table 3.4: Secondary structure elements that possibly make contacts with DNA.

*Bse*634I position DNA in approximately the same fashion in the cleft of the *Bse*634I (fig. 3.8).

The minor groove of the positioned DNA faces the N-terminal part of the helix α_3 in the N-terminal subdomain of *Bse*634I, and the N-terminal end of the helix α_6 gets into its major groove. The loop β_1–loop$_{15-20}$–β_2 could fit into the major groove of the DNA next to the helix α_3, separated from the latter by a sugar-phosphate backbone. The helix α_7, although located pretty far from the DNA in the present model, could in principle make contacts with the major groove of the DNA if the DNA is distorted or if the protein subdomains undergo significant movements after binding of nucleic acid. Helix α_6 approaches DNA from the major groove side, together with helix α_7. The loop between β_3 and $3_{10}1$ might also reach the DNA from the minor groove side. In principle, any residue from the mentioned structural elements could build contacts in the DNA bases and/or backbone and take part in the DNA recognition through direct or indirect sequence readout.

The secondary structure elements α_3 and β_1–loop$_{15-20}$–β_2 belong to the N-terminal subdomain, while the helices $3_{10}1$, α_6 and α_7 and the loop between β_3–$3_{10}1$ come from the C-terminal subdomain and could be bridged by the bound DNA molecule. One can expect that upon binding of the DNA the two subdomains of each subunit in the dimer move even more than observed in the DNA-less structure of the *Bse*634I. The helix α_3 acts as one of the DNA recognition elements and fits into the minor groove of the DNA with its N-end; at the same time, α_3 carries one of the Mg^{2+} complexating residues (Glu80) in the middle, and a joint between C- and N- subdomains is situated at the C-end of α_3. Thus α_3 could act as a molecular cantilever that upon binding the cognate DNA sequence rotates the *Bse*634I N-domain, positions Glu80 so that it builds up a Mg^{2+} binding centre together with the rest of the active centre residues, and in this way triggers the phosphodiester bond cleavage.

Structural element	**Residue**	**Possible contacts with DNA**
β_1, β_2	15–21	
β_2–l_{24-25}–α_1	Lys21	base$_{-5}$, $(T/C)_3-p-$base$_4$, base$_4$
	Arg23	base$_4-p-$base$_5$, base$_5$, base$_5-p-$base$_6$
	Phe27	base$_5-p-$base$_6$
α_3	Thr67	$(T/C)_3$, $(T/C)_3-p-$base$_4$, base$_4$, base$_4-p-$base$_5$
	Ser68	$(T/C)_3-p-$base$_4$, base$_4$, $C_{-1}-p-G_1$
	Ser72	$C_{-2}-p-C_{-1}$, C_{-1}, $C_{-1}-p-G_1$, G_1, base$_4$, base$_4-p-$base$_5$
	Asn73	base$_{-5}$, C_{-1}, $(A/G)_{-3}$, $(A/G)_{-3}-p-C_{-2}$, C_{-2}, $C_{-2}-p-C_{-1}$, $(T/C)_3$, $(T/C)_3-p-$base$_4$, base$_4$, base$_4-p-$base$_5$, base$_5$
	Thr74	base$_4$, base$_4-p-$base$_5$, base$_5$
	Arg75	$C_{-2}-p-C_{-1}$
	Thr77	base$_4-p-$base$_5$, base$_5$, base$_5-p-$base$_6$
β_3–$l_{104-111}$–α_5	Asn106	base$_{-4}$, base$_{-4}-p-(A/G)_{-3}$
	Lys108	base$_{-5}$, base$_{-5}-p-$base$_{-4}$, base$_{-4}$, base$_{-4}-p-(A/G)_{-3}$
α_{11}–$l_{131-147}$–β_4	Ser143	base$_{-4}-p-(A/G)_{-3}$
	Asn144	base$_{-4}-p-(A/G)_{-3}$
	Asp146	$(A/G)_{-3}-p-C_{-2}$
α_{11}	Thr199	$C_{-2}-p-C_{-1}$, C_{-1} $C_{-1}-p-G_1$
	Ser200	$C_{-1}-p-G_1$
	Arg202	G_1, C_{-1}, $C_{-1}-p-G_1$
	Pro203	$(A/G)_{-3}$, G_1, G_2
	Asp204	C_{-1}
	Arg205	C_{-2}, $C_{-2}-p-C_{-1}$, C_{-1}, $C_{-1}-p-G_1$
	Gln208	$(A/G)_{-3}-p-C_{-2}$, C_{-2}, $C_{-2}-p-C_{-1}$, C_{-1}
β_5	Lys198[a]	base$_{-4}$, $(A/G)_{-3}-p-C_{-2}$, C_{-2}. $C_{-2}-p-C_{-1}$

[a] Active centre residue; essential for catalysis

Table 3.5: Possible DNA-protein contacts in the modelled *Bse*634I–DNA complex.

3.1.11 "Swap" mutations

In several places of the *Bse*634I:*Cfr*10I sequence alignment we find substitutions of a rather bulky residue with a relative small one (for example, Phe280 in the *Bse*634I has a corresponding Val272 in the *Cfr*10I sequence). When looking at the structure one can see that the same spatial position in *Cfr*10I is occupied by the Phe35, which aligns with the Ile43 in *Bse*634I sequence. If the residues Val272 and Phe35 in *Cfr*10I were be swapped (or, alternatively, Ile43 and Phe280 in *Bse*634I), they would occupy the same position in space and would align with their counterparts in the sequence alignment. Thus we observe a kind of "swap" mutations, where the two residues exchange their places in a sequence but retain the same position in the protein core. The observed swap mutations show yet another type of the possible protein sequence rearrangements that, along with the point mutations, deletions and insertions, preserve a 3D structure of a protein and maintain the compact protein core.

3.1.12 Ions

There are five peaks above 4.0σ (above 6.0σ for the position Z1) in the σ_A weighted $2F_o - F_c$ electron density of the final *Bse*634I structure. Inclusion of waters in these positions still leaves a positive $F_o - F_c$ density, which disappears only after inclusion of about 35–40 electrons. The density could be accounted either by a Cl^- or by a Ca^{2+} ion, both present in the crystallisation buffer. As shown in the table 3.6 and fig. 3.15, all five positions are mostly surrounded by nitrogen or carbon atoms. We think that finding Ca^{2+} in such environment is unlikely, and we have assigned Cl^- to those positions because there are no other suitable ions in the crystallisation buffer.

The four Cl^- ions (Z2–Z5) obey the non-crystallographic symmetry of the *Bse*634I. Strangely, one Cl^- (Z1) is present only in subunit B, but its NCS equivalent in the subunit A is missing. It is not clear at the moment whether this is related to the conformation differences of two subunits.

3.2 Results on *Mun*I

3.2.1 Crystals of *Mun*I

*Mun*I restriction endonuclease was crystallised using house factorial screening. Wild-type *Mun*I restriction endonuclease yielded three different crystal forms: crystals of the apo-enzyme (without a substrate nucleotide), crystals with cognate oligonucleotide and crystals with cognate oligonucleotide in the presence of Mg^{2+}.

The crystallisation behaviour of *Mun*I was very different depending on whether oligonucleotide was added to the crystallisation buffer and whether the oligonucleotide contained a *Mun*I recognition site. Under the crystallisation conditions

Pos.	Ion	Chain	Residue	Atom	Distance, Å
Z1	Cl^-	B	Arg113	N_ε	3.5
				$N_{\eta 2}$	3.5
		B	Arg121	N_ε	3.2
				$N_{\eta 2}$	3.7
		B	His125	$N_{\varepsilon 2}$	3.6
		B	Leu124	$C_{\delta 1}$	3.8
Z2	Cl^-	B	Lys25	C	3.5
				C_α	3.5
				C_β	3.4
				C_ε	3.5
		B	Phe27	C_β	3.5
				N	3.4
Z3	Cl^-	A	Lys25	C	3.2
				C_α	3.4
				C_β	3.4
				O	3.4
		A	Phe27	N	3.1
		A	Pro26	N	3.5
Z4	Cl^-	B	Arg226	O	3.6
		B	Lys123	C_ε	4.3
				N_ζ	4.3
		B	Tyr227	C_α	3.9
				$C_{\delta 2}$	3.8
				$C_{\varepsilon 2}$	4.4
Z5	Cl^-	A	Arg226	O	3.7
		A	Lys123	C_ε	4.1
				N_ζ	3.9
		A	Tyr227	C_α	4.1
				$C_{\delta 2}$	3.9
				$C_{\varepsilon 2}$	4.4

Table 3.6: Ions and their surrounding in the *Bse*634I structure.

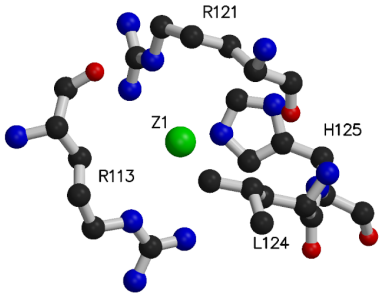


Figure 3.15: Ion binding in the *Bse*634I restriction endonuclease.

Crystal dimensions	$0.4 \times 0.3mm$
Space group	$C222_1$
Unit cell	$100.6 \times 109.5 \times 226.5$ Å
Resolution	3.0 Å

Table 3.7: *Mun*I crystal properties.

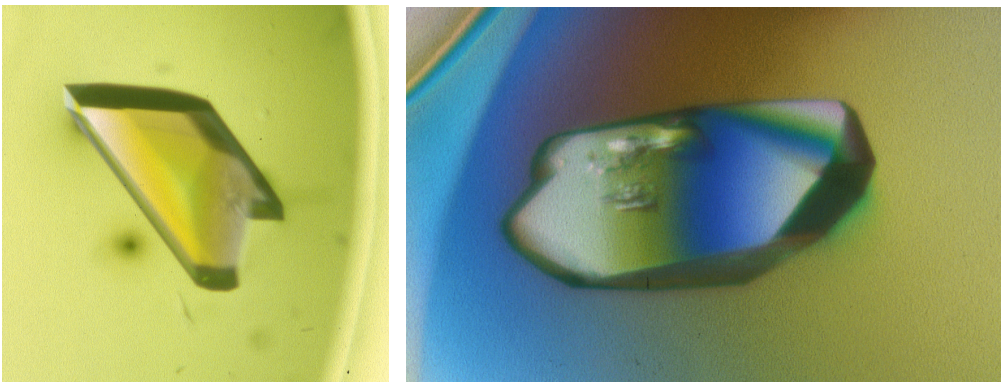


Figure 3.16: *Mun*I co-crystals with ion-exchange purified or Na-Acetate precipitated CAA oligonucleotide.

(table 2.17) *Mun*I-cognate oligonucleotide mixture remained clear and would produce crystals in a few days, while the *Mun*I protein alone immediately gave a heavy precipitate that did not lead to crystallisation. Reduction of the protein concentration in the drops containing *Mun*I alone did not prevent the formation of precipitate, but addition of the equimolar amount of cognate oligonucleotide would immediately dissolve it. This finding suggests that *Mun*I forms a complex with cognate oligonucleotide and that this complex most probably forms crystals.

The addition of non-cognate oligonucleotide did not prevent the formation of the precipitate in *Mun*I crystallisation drops. The crystallisation behaviour of the *Mun*I/non-cognate oligonucleotide mixtures was exactly the same as of the *Mun*I protein alone. This finding suggests that the *Mun*I complex with non-cognate oligonucleotide does not exist or is very weak in *Mun*I crystallisation buffer, or that the complex is much less soluble that the complex with cognate oligonucleotide.

Addition of 1.5 mM Mg^{2+} to *Mun*I-cognate oligonucleotide crystallisation drops would reduce solubility producing a visible precipitate. This precipitate unlike the precipitate of the apo-protein could be dissolved by reducing the concentration of precipitant PEG8K from 9% to 3%, and these modified conditions would yielded a new crystal form. Since the crystallisation buffer of *Mun*I contained 50 mM $CaCl_2$, the addition of 1.5 mM $MgCl_2$ would hardly affect the crystallisation through the change of ionic strength. Instead, Mg^{2+} most likely acts as a cofactor of *Mun*I restriction endonucleases and leads to immediate cleavage of the substrate oligonucleotide and crystallisation of the restriction endonuclease-product complex.

Crystals of the apo-enzyme did not diffract X-rays at all. Cocrystals with oligonucleotide `CAA` (see table 2.3) in the absence of Mg^{2+} diffracted to 3.3 Å on the rotating anode source and IP detector. The crystals suffered from radiation damage, and a complete data set could be only obtained by merging data

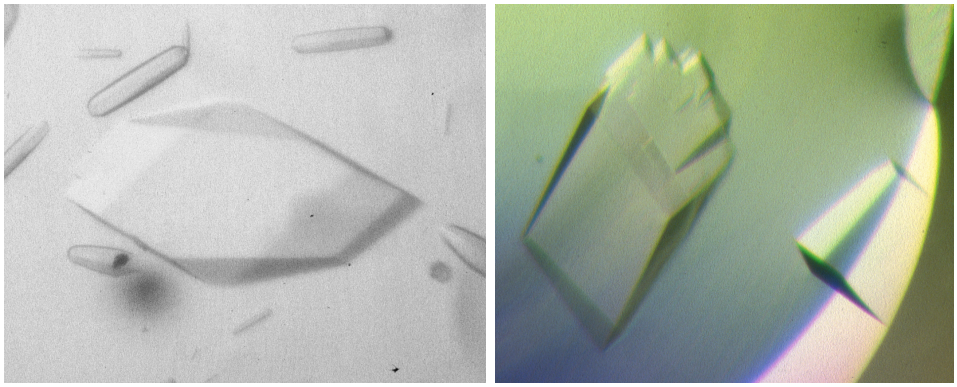


Figure 3.17: *Mun*I co-crystals with the CAA oligonucleotide, purified by RF chromatography. The same kind of crystals is produced in the presence of Mg^{2+} ions.

from the several crystals. The crystals did not survive any buffer change, and therefore could not be measured under cryo-conditions. The crystals belong to orthorhombic crystal form (space group $C222_1$).

Cocrystals with oligonucleotide `CAA` and Mg^{2+} diffracted to 5 Å on the rotating anode source and IP detector. These crystals could be however transfered into the cryoprotectant buffer and frozen in the 90K nitrogen stream without loss of diffraction. From a frozen crystal, a dataset was collected at BW6/DESY (Hamburg). Although mosaicity of the crystal increased substantially after freezing, a dataset of reasonable quality could be obtained. The crystals belong to the trigonal crystal form (space group R32).

Since magnesium is the only co-factor required by type II restriction endonucleases to cleave their substrate, it can be proposed that the crystals obtained with addition of Mg^{2+} contain a cleaved substrate. Indeed, HPLC and PAAGE analysis of the oligonucleotide that has been incubated at the conditions close to those in crystallisation drop with WT *Mun*I show that the oligonucleotide is completely cleaved under these conditions. Since *Mun*I without oligonucleotide has a completely different crystallisation behaviour as with oligonucleotide/Mg^{2+}, it is likely that the crystals contain a complex of *Mun*I and DNA cleavage product.

Of note is that sometimes the crystal form that was normally obtained after addition of Mg^{2+} would appear in the crystallisation trials where no Mg^{2+} was added. This behaviour was observed with the oligonucleotides that were purified by the producing companies by reversed-phase technique. Such oligonucleotides would yield the trigonal crystals in the conditions that normally produce tetragonal crystal form. These oligonucleotides could be however used to produce tetragonal crystals after additional purification over MonoQ column at high pH (11.0) or after their precipitation from Na acetate/ethanol. Addition of Mg^{2+} to such oligonucleotides would restore their ability to produce trigonal crystals.

Crystal:	**dataset 10**	**dataset 6**	**merged 6 and 10**	**127A1**	**149D3**
Spacegroup,	$C222_1$	$C222_1$	$C222_1$	P2	R32
Unit cell, Å	$a = 99.8$	$a = 100.0$	$a = 100.0$	$a = 70.3$	$a =$
	$b = 109.2$	$b = 109.6$	$b = 109.6$	$b = 112.7$	$b = 137.3$
	$c = 226.1$	$c = 226.5$	$c = 226.5$	$c = 78.7$	$c = 256.1$
	$\alpha = \beta = \gamma = 90°$	$\alpha = \beta = \gamma = 90°$	$\alpha = \beta = \gamma = 90°$	$\alpha = \gamma = 90°$, $\beta = 108.5°$	$\alpha = \beta = 90°$, $\gamma = 120°$
Max. resolution, Å	3.0	3.0	5.00	2.68	2.60
R_{merge}[a]	0.098	0.040	0.106	0.098	0.059
Unique reflections	20376	19226	5523	33575	27032
Completeness, %	82	75	98	94	94
Last resolution shell starts at, Å	3.1	3.1	5.17	2.76	2.68
R_{merge}, in the last resolution shell	0.254	0.249	0.140	0.565	0.237
Completeness, %, in the last resolution shell	56	78	97	94	30

[a] $R_{merge} = \sum_{\vec{h}} \sum_{i=0}^{n_{\vec{h}}} (\langle I_{\vec{h}} \rangle - I_{\vec{h}i})^2 / \sum_i I_{\vec{h}}^2$

Table 3.8: Data collection statistics of *Mun*I.

This behaviour is most probably caused by the trace amounts of Mg^{2+} in the RF purification but could be efficiently eliminated by precipitation or ion-exchange chromatography at low pH values.

3.2.2 Diffraction data

Data were indexed with REFIX program. Data were reduced and cell constants were refined using MOSFLM package. The obtained intensities were then scaled and merged using programs from CCP4 program package [20].

To get the completeness of data above 95%, it was necessary to merge data from at least two crystals, because the crystals were very sensitive to radiation damage and collecting of the full data set from one crystal was impossible.

3.2.3 Limited proteolysis

In course of search for optimal crystallisation conditions we observed that *Mun*I was subjected to limited proteolysis during storage and crystallisation. The proteolysis occurs in the fixed positions near N-terminus of the protein. The cleavage positions were verified by MALDI-MS and N-terminus sequencing. The produced inhomogeneity of the protein might be one of the reasons reason why it was not possible to produce the well-diffracting crystals for a long time.

```
1         10          20          30
|          |          |          |
MGKSELSGR-LNWQALAGLK-ASGAEQNLYNVFNAVFEGTKYVLY...
          ^          ^
```

Figure 3.18: Sites of the limited proteolysis of *Mun*I.

The identified cleavage position allowed us to suppose that the proteolysis of the *Mun*I is caused presumably by the trace amounts of the serine protease, which is present in the purified sample and cannot be neither detected nor avoided by usual purification techniques. The knowledge of the protease type, however, allowed to choose the inhibitor for the protease and to find the crystallisation conditions that allow to get reproducibly diffracting crystals.

We were also able to perform the same proteolysis under controlled conditions using small amounts of trypsin (V. Siksnus, unpublished). The cleavage occurs in the exactly the same sites as the unwanted "wild" proteolysis. The fragments obtained retained the catalytic activity. It was also shown that the addition of cognate oligonucleotide reduces the susceptibility of *Mun*I to the proteolysis. These findings allow to suggest that the N-terminus of the protein is less compact than the rest of the globule and that it undergoes structural rearrangements upon substrate binding.

3.2.4 Native electrophoresis

Native PAAG electrophoresis after McLellan [69] appeared suitable very well for testing restriction enzyme–oligonucleotide binding. Under the same conditions, protein-oligonucleotide moves much faster than the protein alone (compare lanes 2–3 and 4–7 in fig. 3.19). This can be explained by an extra negative charge brought by DNA phosphates. In accordance with this assumption, complexes with longer oligonucleotides move faster that those with shorter oligonucleotides (lanes 4 and 5 in fig. 3.19). Thus, the binding of the oligonucleotide to restriction endonucleases can be monitored under wide range of pH conditions in the buffers that allow electrophoresis. To prove that the higher mobility of the bands is really

caused by the oligonucleotide, electrophoresis with fluorescein (5FLU) labelled oligonucleotide was carried out, which showed that the fluorescence labels co-migrates with the protein band in both specific (lanes 8 in fig. 3.21 and fig. 3.22) and non-specific (lanes 4 in fig. 3.21 and fig. 3.22) complexes.

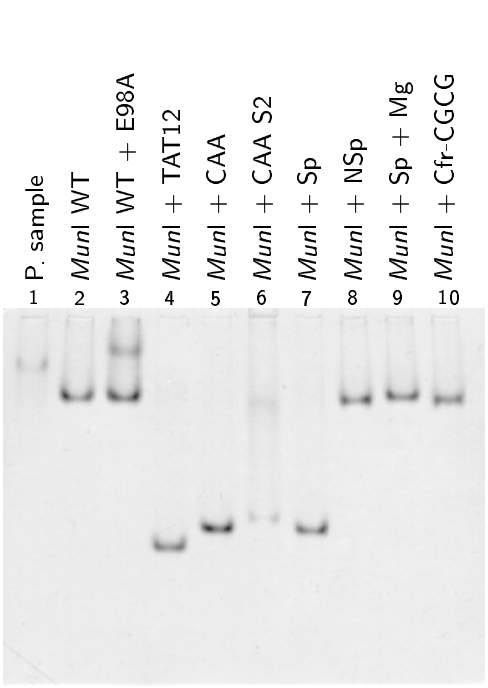


Figure 3.19: *Mun*I with specific and non-specific oligonucleotides, pH = 7.5

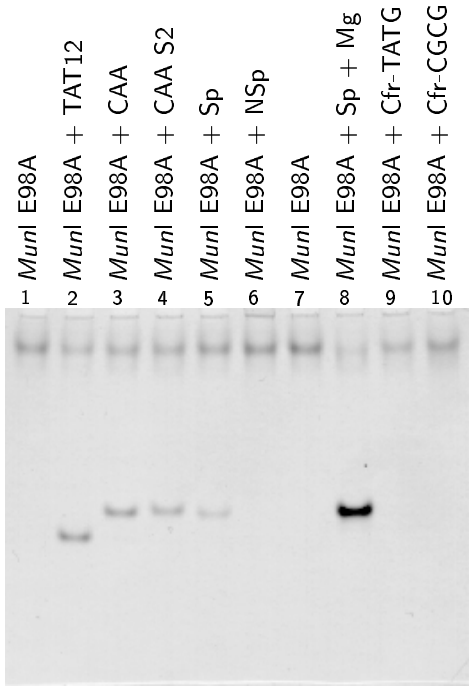


Figure 3.20: *Mun*I E98A mutant with specific and non-specific oligonucleotides, pH = 7.5

Two active centre mutants of *Mun*I, D83A and E98A, were also tested for nucleotide binding. D83A mutant behaves very much like the WT protein (fig. 3.24, lanes 3, 7, 10), except that it does not cleave DNA. In contrast, E98A mutant moves nearly twice slower as WT or D82A protein. Since both both E98A and D82A mutants lack the same amount of negative charge compared with wild-type *Mun*I, the differences in their mobility suggests that E98A is less tighter packed that WT protein (*i.e.* has higher particle radius). Unlike WT and D83A variants, E98A fails to form a specific complex at higher pH (compare fig. 3.24 lanes 8 and 9 (E98A) with lanes 5 (WT), 6 and 10 (D83A)). Only at lower pH values could a specific complex of E98A mutant be observed, like in fig. 3.20 lanes 2–5 and 8. Since E98A mutant is inactive, the presence of Mg^{2+} does not lead to cleavage of the specific complex (fig. 3.20 lane 8), while in the wild-type *Mun*I assay addition of Mg^{2+} leads do a disappearance of the specific complex band (fig. 3.19, lane 9). This also hints that the product complex with *Mun*I is unstable under the

electrophoresis conditions, since only a pure *Mun*I protein band can be observed after addition of Mg^{2+} to a specific complex.

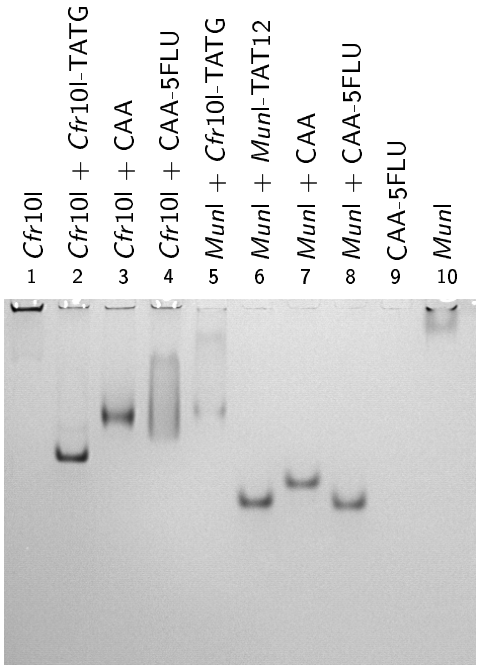


Figure 3.21: Coomassie stained native *Mun*I+CAA-FLU gel, pH = 6.6.

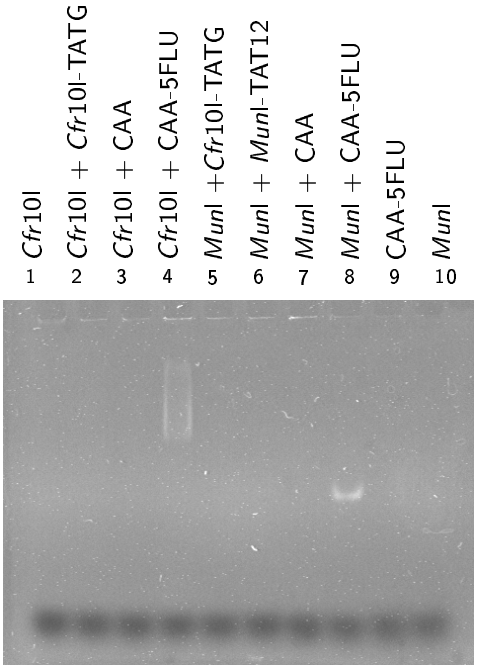


Figure 3.22: *Mun*I+CAA-FLU gel photographed in UV light (254nm) before staining, pH = 6.6.

In the presence of a non-specific oligonucleotide a wild-type *Mun*I builds a non-specific complex with the oligonucleotide at higher pH values (fig. 3.23, lanes 6 and 7). This complex appears to be less compact that a complex with specific oligonucleotide (showing lower electrophoretic mobility with the same protein and an oligonucleotide of the same length). As a rule, bands of non-specific restriction endonuclease–oligonucleotide complex come out smeared, suggesting that the non-specific complex is not stable and multiple association–dissociation events occur during migration in the gel. The same behaviour of specific and non-specific complexes was observed with other wild type restriction endonucleases (example of *Cfr*10I is shown in fig. 3.21 lanes 1–4 and fig. 3.22 lane 4). In contrast to WT *Mun*I, E98A mutant fails to bind non-specific oligonucleotides at any pH (fig. 3.24 lanes 8 and 9, fig. 3.20 lanes 9 and 10).

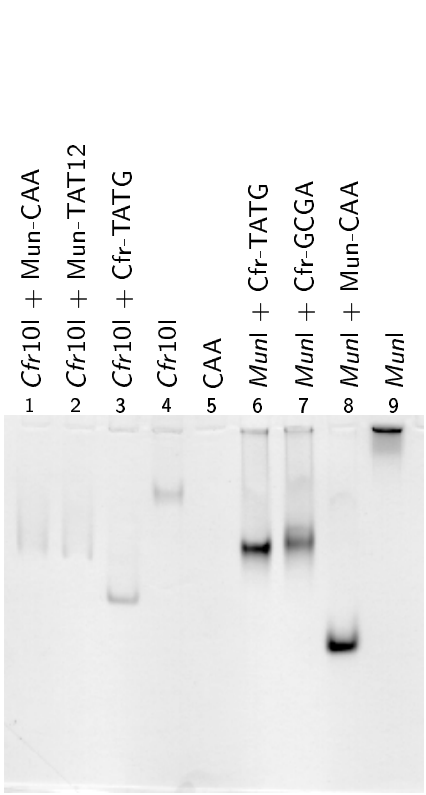


Figure 3.23: *Mun*I WT and *Cfr*10I with specific and non-specific oligonucleotides, pH = 6.6.

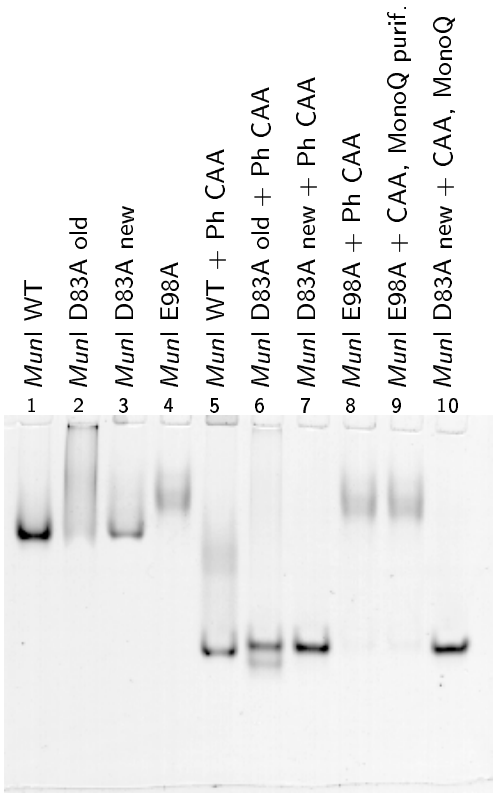


Figure 3.24: *Mun*I WT, E98A and D83A with and without oligonucleotide CAA, pH = 8.5

3.2.5 Studies of enzyme kinetics

The conditions under which the crystals for X-ray analysis were grown differ considerably from the optimal conditions for enzymatic reaction, as well as from the natural physiological environment of the enzyme. It is therefore necessary for understanding of the structure–function relations to know the activity of the enzyme under crystallisation conditions and to compare it with the activity under optimal conditions. It is also known that *Mun*I has "star" activity under certain unusual circumstances, *e.g.* in the presence of Mn^{2+} ions.

The kinetics of the enzymatic reaction is monitored by recording the change in UV absorption in the course of the reaction. Due to hyperchromic effect, the reaction products of oligonucleotide cleavage, that are short and therefore single-stranded at the reaction temperature, have a higher UV absorption than the double-stranded DNA substrate. A typical reaction course is shown in fig. 3.26.

Reaction speeds recorded at different substrate concentration allowed also to estimate a Michaelis reaction parameters of *Mun*I with short oligonucleotide substrates.

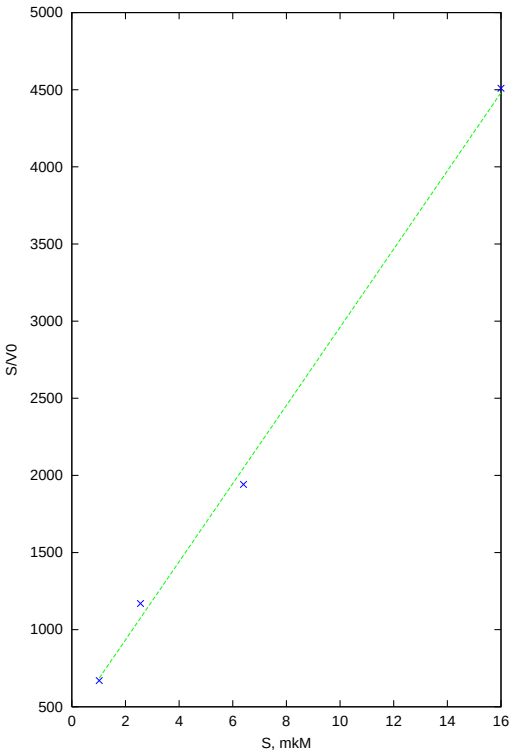


Figure 3.25:

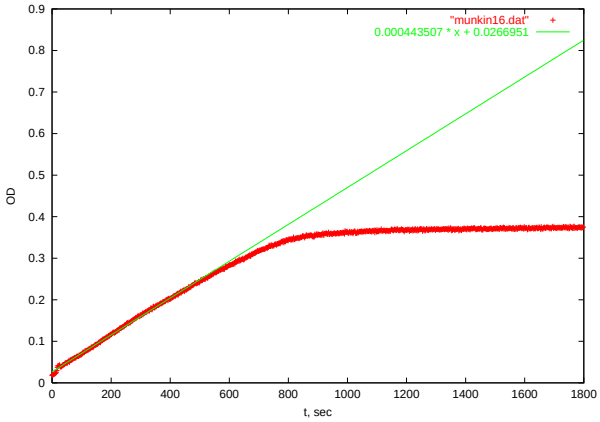


Figure 3.26:

$$K_M = 1.68638\,\mu\mathrm{M}$$
$$V_{max} = 0.00394725\,\frac{\mathrm{OU}}{\mathrm{s}} = 33.3\frac{\mathrm{nM}}{s}$$
$$\varepsilon_{260}^{\mathtt{CAA}} = 118.6\ \mathrm{mM}^{-1}\mathrm{cm}^{-1}$$

Chapter 4

Conclusions

The structure of the *Bse*634I restriction endonuclease has been solved using X-ray MIR technique. Together with *Cfr*10I, *Bse*634I makes a first pair of isoshisomeric restriction endonucleases with known DNA-free structures. Comparison of the different chains in *Bse*634I allowed to identify N- and C-terminal subdomains that are joined by a flexible link around which N-terminal subdomain can rotate. This rotation might be important for coupling DNA recognition and catalysis. Structurally equivalent subdomains have been postulated for *Cfr*10I from the comparison of the two structures.

Active site of *Bse*634I can be easily identified from the similarity with *Cfr*10I and other restriction endonucleases. Although there are no metal ion in the active centre of the DNA-free enzyme, possible Mg^{2+} binding sites have been identified from the comparison with other restriction endonucleases and from the binding position of Gd^{2+} ion. It is not possible to tell from the present structure whether two metal ions can be bound to *Bse*634I and whether two ions are required for catalysis.

Approximate DNA position was modelled from the overlays of *Bse*634I dimer active centres with other DNA-containing restriction endonuclease structures. The DNA model allows to identify structural elements of *Bse*634I that might be involved in DNA recognition.

Native electrophoresis and crystallisation experiments suggest that stable complexes of wild-type *Mun*I restriction endonuclease with cognate oligonucleotide can be obtained and crystallised. In the native gel electrophoresis, complexes of *Mun*I with oligonucleotide have been obtained under various pH and with both cognate and non-cognate oligonucleotides.

Several crystal forms of *Mun*I cocrystals with cognate oligonucleotide have been obtained. Two forms presumably contain enzyme-substrate complex, and one – enzyme-product complex. A dataset from the putative enzyme-product crystal was collected to 2.6 Å resolution, and molecular replacement solutions could be found that offer a possibility for the structure determination.

Protocols

Preparation of gels for PHAST system

Composition

Separating gel 14%

	Amount of solutions needed for		
	1 gel	2 gels	4 gels
Acrylamide 30% / Bis AA 0.8%	4.67 ml	9.33 ml	18.7 ml
Gel buffer	1.0 ml	2.0 ml	4.0 ml
Glycerin	2.5 ml	5.0 ml	10.0 ml
H_2O	1.83 ml	3.67 ml	7.3 ml
Total volume	10 ml	20 ml	40 ml
Polymerisation:			
12 % PS	15 μl	30 μl	60 μl
TEMED	7.5 μl	15 μl	30 μl

Stacking gel

	Amount of solutions needed for		
	1 gel	2 gels	4 gels
Acrylamide 30% / Bis AA 0.8%	0.83 ml	1.67 ml	3.33 ml
Gel buffer	0.37 ml	0.75 ml	1.5 ml
Glycerin	1.25 ml	2.5 ml	5.0 ml
H_2O	2.55 ml	5.08 ml	10.17 ml
Total volume	5 ml	10 ml	20 ml
Polymerisation:			
12 % PS	10 μl	20 μl	40 μl
TEMED	5 μl	10 μl	20 μl

Ammonium Persulfate 12%

Dissolve 0.3 g in 2.5 ml of H_2O.

Pre-mixed buffers for 15% SDS gels

Stock solutions, materials and equipment needed

1. AA/BisAA Protogel 30% (37.5:1)
2. Separating gel buffer ×2
3. Stacking gel buffer ×1.2
4. Ammonium Persulphate 10%
5. TEMED 100%

Stock solutions composition

Separating gel buffer, ×2, pH=8.8

Item	MW, g/mol	Amount for: 200 ml	500 ml	Final Conc. ×2	×1
Tris	121.14	**18.17 g**	**45.43 g**	0.75 M	0.375 M
SDS	288.38	**0.40 g**	**1.00 g**	0.2%	0.1%

Adjust pH to 8.8 with HCl.

Stacking gel buffer, ×1.2, pH=6.8

Item	MW, g/mol	Amount for: 200 ml	500 ml	Final Conc. ×1.2	×1
Tris	121.14	**3.63 g**	**9.09 g**	0.15 M	0.125 M
SDS	288.38	**0.24 g**	**0.60 g**	0.12%	0.1%

Adjust pH to 6.8 with HCl.

Gel composition

Separating gel (15%)

Item	Stock	Amount for:	
		1 gel	2 gels
Separating gel buffer	×2	2.5 ml	5.0 ml
Protogel	30% (37.5:1)	2.5 ml	5.0 ml
Total volume		5.0 ml	10.0 ml
Polymerisation:			
Ammonium Persulphate	10%	25 μl	50 μl
TEMED	100%	4 μl	8 μl

Stacking gel (5%)

Item	Stock	Amount for:	
		1–2 gels	3–2 gels
Stacking gel buffer	×1.2	4.167 ml	8.333 ml
Protogel	30% (37.5:1)	0.833 ml	1.667 ml
Total volume		5.0 ml	10.0 ml
Polymerisation:			
Ammonium Persulphate	10%	50 μl	100 μl
TEMED	100%	4 μl	8 μl

Precipitation of the Oligonucleotides Using Sodium Acetate

Stock solutions, materials and equipment needed

Nr.	Stock	Conc.	Amount	Notes
1.	Oligonucleotide solution	0.5 mM	100 μl	$\nu = 50$nmol
2.	Na Acetate	3 M	300 μl	pH = 5.0
3.	C_2H_5OH	100 %	800 μl	Absolute
4.	C_2H_5OH	70 %	2 ml	Pre-chilled to –20° C

Protocol

1. Add ≈50 nmol[1] of the oligonucleotide into the 1.5 ml Eppendorf tube. This will be 100 μl of the 0.5 mM solution.
2. Add 300 μl of the 3 M Na Acetate solution.
3. Add 800 μl of absolute ethanol.
4. Vortex the tube and leave overnight at –20° C.
5. Spin the test tube in the centrifuge at 23000 g for 10 minutes at –4° C.
6. Pour off carefully the supernatant. Wash the pellet with 1 ml of 70% ethanol, pre-chilled at –20° C.
7. Repeat the washing step 6.
8. Dry the pellet in the SpeedVac vacuum evaporator.
9. Dissolve the pellet in 100 μl Millipore H_2O and evaporate again.
10. Dissolve the pellet in 20 μl Millipore H_2O. Measure concentration by diluting a small aliquot of the product and measuring the absorbance at 260 nm.

Notes

1. This protocol works fine with the oligonucleotides of the length 10 — 12 bp, giving yields of 70%. Another salts may be used for precipitation, e.g. Cs Acetate. In the case of Cs Acetate, however, the yields drop to about 50% or less.

[1] the protocol works fine for the amounts of oligonucleotide from 10 to 50 nmol

2. Pellet after step 4 can be sometimes very weak or not visible at all. However, the final yield of the oligo measured by the UV absorbance is usually high.

3. The amount of pellet after washing step 6 is getting substantially lower. This is perhaps due to dissolving the co-precipitated salts, but it may happen that the oligo is dissolving too. Try to make this operation as fast as possible and not to heat the oligo.

4. In the step 5 the "23000 g" actually means "as fast as possible", and –4° C mean "as low temperature as your centrifuge can provide, down to –20° C".

5. The Eppendorf centrifuge has problems to cool the samples to –10° C when spinning at 23000 g. You may want to stop the centrifuge after some time and switch it to "fast cool" mode to drop the temperature and to prevent the dissolution of the oligonucleotide; after this you can go on spinning at the high speed. Always cool the centrifuge using the "fast cool" mode before putting your samples into it.

Incorporation of Sulfo-SANPAH into *Mun*I Restriction Endonuclease

Stock solutions, materials and equipment needed

1.	*Mun*I stock solution ∼1 mg/ml	1 ml
2.	oligonucleotide CAA[a] 0.29 mM	1 μl
3.	*Mun*I dialysis buffer	1.1 l
4.	Sulfo-SANPAH in H_2O, 20 mM	100 μl
5.	Glycin 100 mM	150 μl
6.	NAP-10 desalting column	1
7.	Centricon 10kD (green)	1
8.	UV (quarz) cuvetts	2

[a]sequence: 5'-GCCAATTGGC-3'

Protocol

Warning: All manipulations with Sulfo-SANPAH and labeled protein are performed in dark! (Illumination with red photo-lab lamp)

1. dialyse 1 ml of *Mun*I sample (∼1 mg/ml = 42.4 nM) overnight against 500 ml of *Mun*I dialysis buffer; change buffer after first few hours of dialysis (so you use 1 l of buffer total).

2. measure the protein concentration after dialysis.

3. transfer dialysed *Mun*I sample to 2 ml Eppendorf test tube; add 1 μl of oligonucleotide solution (this gives a 5-fold excess of oligonucleotide over protein, to ensure that all protein is bound).

4. add 70 μl of Sulfo-SANPAH solution (20 mM) to reach final concentration of 1.3 mM of the agent in the final solution (this gives approx. 30-fold excess of Sulfo-SANPAH over the protein).

5. let to react for 30 min at room temperature.

6. quench the unreacted Sulfo-SANPAH , adding 150 μl of glycine solution (100 mM, this gives the approx. 10-fold excess over the initial amount of the Sulfo-SANPAH)

7. separate the unreacted Sulfo-SANPAH and glycin on Pharmacia NAP-10 column, as described in the NAP-10 usage instruction.

8. measure the spectrum of the protein fraction. The extinction coefficients are:

*Mun*I	$\varepsilon_{280} = 45720\ M^{-1}cm^{-1}$
Sulfo-SANPAH	$\varepsilon_{458} = 5400\ M^{-1}cm^{-1}$ [66]

9. concentrate labeled *Mun*I in Centricon concentrator to approx. 15 mg/ml; this takes about 3 h at 5000 rpm in SORVALL centrifuge.

*Mun*I reaction kinetics

Materials and equipment

Spectrophotometer: Perkin-Elmer *Lambda 17* two beam spectrophotometer with thermostated cuvette holder.

Cuvettes: 2 quartz glass cuvettes for UV light (*e.g.* Hellma QX Suprasil 300)

Oligonucleotide: Synthetic oligonucleotide (name CAA, sequence `5'-GCCAATTGGC-3'`; stock concentration about 1 mM)

Protein: *Mun*I restriction endonuclease solution, stock concentration about 100 μM (2.3 mg/ml) in 10 mM Tris·HCl buffer (pH = 7.5)

Buffers: Optimal *Mun*I reaction buffer, $\times 2$:

Item:	Stock:	Volume:	Final:
Tris·OH	1 M	1.0 ml	20 mM
$MgCl_2$	2 M	0.5 ml	20 mM
Millipore H_2O	–	48.5 ml	–
Total Volume		50 ml	
pH adjusted to 7.5 with HCl			

Preparation of solutions

1. Prepare reaction mix, 700 μl:

Item:	Stock:	Volume:	Final:
*Mun*I Reaction Buffer	$\times 2$	350.0 μl	$\times 1$
CAA (6.06f3.13)	1.22 mM	9.18 μl	16 μM
Millipore H_2O	–	348.8 μl	–
Total Volume	–	700 μl	–

2. Prepare 50 μl of 100 μM *Mun*I stock solution
3. A half of reaction mix is always taken for reference.

Measurement

1. Switch on the spectrophotometer and the cuvette thermostate. Set the thermostate temperature 37° C. Allow the spectrophotometer to warm up for 20 – 30 min.

2. Take the reaction mix and divide it in two parts of 345 μl – one for reference cuvette, and the other for the reaction cuvette. Put both cuvettes into the holder and allow temperature to equilibrate for 5 min. Cuvettes should be closed with lids to prevent evaporation or accumulation of dust.

   One can monitor the baseline for the first few temperature equilibration runs and look if 5 min are enough. Baseline should have no drift when the equilibration is achieved.

3. Blank the spectrophotometer (auto zero) so that the current difference in absorbances is 0.

4. When the temperature of the reaction mix is equilibrated, take out the reaction cuvette, add quickly 1 μl of *Mun*I stock (final concentration of protein will be **0.167** μM), replace the lid of the cuvette and turn it upside down for the 5 – 7 times to mix the solutions properly. After this put the cuvette back into the cuvette holder and start measuring the optical density vs. time (on the Perkin-Elmer software this is called *time-drive* mode)

5. record the reaction for 30 min. In this time the linear increase of the OD should end due to the substrate depletion.

6. Take out the reference cuvette, dilute the solution 2.5 times with $\times 1$ *Mun*I reaction buffer:

Item:	Stock:	Volume:	Final:
Reference solution	16 μM	280 μl	6.4 μM
*Mun*I Reaction Buffer	$\times 1$	420 μl	$\times 1$
Total Volume		700 μl	

   In this way you will have again enough solution for reference and for the reaction cuvettes. Dividing by 2.5 each time gives you a rapid decrease of substrate concentration, which means that in a few shots you will reach K_M value. For fine determination of K_M and k_c, however, one should prepare a separate solutions for each substrate concentration and make a finer grid around the expected K_M value.

   When a substrate concentration becomes very low, the reaction ends too fast, and one can not monitor the initial velocity with the enough accuracy. To get around this, one can dilute a protein stock ≈ 5 times and proceed with the lower protein concentration. The resulting initial velocities must be then normalised to the protein concentration (as was done in this experiment).

7. Wash the reaction cuvette with Chrompick and with large amount of the Millipore water between runs to get rid of the protein and the substrate that might be adsorbed to the glass.

Bibliography

[1] R. A. Abagyan and S. Batalov. Do aligned sequences share the same fold? *Journal of Molecular Biology*, 273:355–368, 1997.

[2] S. P. Adler and D. Nathans. Studies of SV 40 DNA. V. Conversion of circular to linear SV 40 DNA by restriction endonuclease from *Escherichia coli* B. *Biochimica et Biophysica Acta*, 299(2):177–88, 1973.

[3] A. K. Aggarwal. Structure and function of restriction endonucleases. *Current Opinion in Structural Biology*, 5:11–19, 1995.

[4] J. E. Anderson. Restriction endonucleases and modification methylases. *Current Oppinion in Structural Biology*, 3:24–30, 1993.

[5] W. Arber and D. Dussoix. Host specificity of DNA produced by *Escherichia coli*: I. Host controlled modification of bacteriophage λ. *Journal of Molecular Biology*, 5:18–36, 1962.

[6] A. Athanasiadis, M. Vlassi, D. Kotsifaki, P. A. Tucker, K. S. Wilson, and M. Kokkinidis. Crystal structure of *Pvu*II endonuclease reveals extensive structural homologies to *Eco*RV. *Nature Structural Biology*, 1(7):469–75, 1994.

[7] B. K. Baxter and M. D. Topal. Formation of a cleavasome: enhancer DNA-2 stabilizes an active conformation of *Nae*I dimer. *Biochemistry*, 32(32):8291–8, 1993.

[8] L. S. Beese and T. A. Steitz. Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO Journal*, 10(1):25–33, 1991.

[9] G. Bertani and J. J. Weigle. Host controlled variation in bacterial viruses. *Journal of Bacteriology*, 65:113–121, 1953.

[10] T. A. Bickle. The ATP-dependent restriction enzymes. In S. M. Linn, R. S. Lloyd, and R. J. Roberts, editors, *Nucleases*, pages 89–109. Cold Spring Harbor Laboratory, second edition, 1993.

[11] T. A. Bickle, C. Barck, and R. Yuan. ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K 12. *PNAS*, 75:3009–3103, 1978.

[12] J. Bitinaite, D. A. Wah, A. K. Aggarwal, and I. Schildkraut. *Fok*I dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18):10570–10575, 1998.

[13] H. Boyer. DNA restriction and modification mechanisms in bacteria. *Annual Review in Microbiology*, 25:153–176, 1971.

[14] H. W. Boyer and D. Roulland-Dussoix. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *Journal of Molecular Biology*, 41(3):459–72, 1969.

[15] D. Bozic, S. Grazulis, V. Siksnys, and R. Huber. Crystal structure of *Citrobacter freundii* restriction endonuclease *Cfr*10I at 2.15 Å resolution. *Journal of Molecular Biology*, 255(1):176–86, 1996.

[16] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. Delano, P. Gros, R. W. Grosse Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D*, 54(5):905–921, 1998.

[17] C. R. Cantor and P. R. Schimmel. *Biophysical chemistry: Techniques for the study of biological structure and function*, volume 2. Freeman, San Francisko, 1980.

[18] X. Cheng, K. Balendiran, I. Schildkraut, and J. Anderson. Structure of the *Pvu*II endonuclease with cognate DNA. *The EMBO Journal*, 13(17):3927–3935, 1994.

[19] X. Cheng, S. Kumar, J. Posfai, J. W. Pflugrath, and R. J. Roberts. Crystal structure of the *Hha*I DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell*, 74:299, 1993.

[20] Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography*, 50(5):760–763, 1994.

[21] B. A. Connolly, F. Eckstein, and A. Pingoud. The stereochemical course of the restriction endonuclease *Eco*RI-catalysed reaction. *The Journal of Biological Chemistry*, 259(17):10760–10763, 1984.

[22] M. Conrad and M. D. Topal. DNA and spermidine provide a switch mechanism to regulate the activity of restriction enzyme *Nae*I. *Proceedings of the National Academy of Sciences of the United States of America*, 86(24):9707–11, 1989.

[23] R. A. Copeland. Methods for protein quantitation. In *Methods for protein analysis*, chapter 3, pages 39–58. Chapman & Hall, New York; London, 1994.

[24] G. M. Cowan, A. A. F. Gann, and N. E. Murray. Conservation of complex DNA recognition domains between familes of restriction enzymes. *Cell*, 56:103–109, 1989.

[25] K. Cowtan. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, 31:34–38, 1994.

[26] R. M. C. Dawson, D. C. Elliott, W. H. Elliott, and K. M. Jones. *Spravochnik biochimika [Russian]*. Mir, Moscow, 1991.

[27] M. Deibert, S. Grazulis, A. Janulaitis, V. Siksnys, and R. Huber. Crystal structure of *Mun*I restriction endonuclease in complex with cognate DNA at 1.7 Å resolution. *The EMBO Journal*, 18(21):5805–5816, 1999.

[28] M. Deibert, S. Grazulis, V. Siksnys, and R. Huber. Structure of the tetrameric restriction endonuclease *Ngo*MIV in complex with cleaved DNA. *Nature Structural Biology*, 7(9):792–799, 2000.

[29] D. Dussoix and W. Arber. Host specificity of DNA produced by *Escherichia Coli*: II. Control over acceptance of DNA from infecting phage λ. *Journal of Molecular Biology*, 5:37–49, 1962.

[30] B. Eskin and S. Linn. The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. II. Purification, subunit structure, and catalytic properties of the restriction endonuclease. *Journal of Biological Chemistry*, 247(19):6183–91, 1972.

[31] P. P. Ewald. Das "reziproke Gitter" in der Strukturtheorie. *Zeitschrift für Kristallographie*, 56:129–156, 1921.

[32] C. A. Frederick, J. Grable, M. Melia, C. Samudzi, L. Jen-Jacobson, B. C. Wang, P. Greene, H. W. Boyer, and J. M. Rosenberg. Kinked DNA in crystalline complex with *Eco*RI endonuclease. *Nature*, 309(5966):327–31, 1984.

[33] J. A. Gerlt. Mechanistic principles of enzyme-catalysed cleavage of phosphodiester bonds. In S. M. Linn, R. S. Lloyd, and R. J. Roberts, editors, *Nucleases*, pages 1–34. Cold Spring Harbor Laboratory, second edition, 1993.

[34] S. W. Glover and C. Colson. Genetics of host-controlled restriction and modification in *Escherichia coli*. *Genetical Research*, 13(2):227–40, 1969.

[35] W. Gong, M.O'Gara, R. M. Blumenthal, and X. Cheng. Structure of *Pvu*II DNA-(cytosine N^4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Research*, 25:2702, 1997.

[36] J. A. Grasby and B. A. Connolly. Stereochemical outcome of the hydrolysis reaction catalysed by the *Eco*RV restriction endonuclease. *Biochemistry*, 31:7855–7861, 1992.

[37] D. Harker. The application of the three-dimentional Patterson method and the crystal structures of proustite, Ag_3AsS_3, and pyrargyrite, Ag_3SbS_3. *Journal of chemical physics*, 4:381–390, 1936.

[38] K. Horiuchi and N. D. Zinder. Cleavage of bacteriophage fl DNA by the restriction enzyme of *Escherichia coli* B. *Proceedings of the National Academy of Sciences of the United States of America*, 69(11):3220–4, 1972.

[39] J. R. Horton and X. Cheng. *Pvu*II endonuclease contains two calcium ions in active sites. *Journal of Molecular Biology*, 300:1049–1056, 2000.

[40] J. R. Horton, H. G. Nastri, P. D. Riggs, and X. Cheng. Asp34 of *Pvu*II endonuclease is directly involved in DNA minor groove recognition and indirectly involved in catalysis. *Journal of Molecular Biology*, 284(5):1491–504, 1998.

[41] N. C. Horton and J. J. Perona. Role of protein-induced bending in the specificity of DNA recognition: crystal structure of *Eco*RV endonuclease complexed with d(AAAGAT)+s(ATCTT). *Journal of Molecular Biology*, 277:779–787, 1998.

[42] Q. Huai, J. D. Colandene, Y. Chen, F. Luo, and H. Ke. Crystal structure of *Nae*I — an evolutionary bridge between DNA endonuclease and topoisomerase. *The EMBO Journal*, 19(12):3110–3118, 2000.

[43] J. Hubacek and S. W. Glover. Complementation analysis of temperature-sensitive host specificity mutations in *Escherichia coli*. *Journal of Molecular Biology*, 50(1):111–27, 1970.

[44] S. J. Hubbard and J. M. Thornton. *'NACCESS' computer program*. University College London, 1993.

[45] S. Iida, J. Meyer, B. Bächi, M. Stålhammar-Carlemalm, S. Schrickel, T. A. Bickle, and W. Arber. DNA restriction–modification genes of phage P1 and plasmid p15B. Structure and in vitro transcription. *Journal of Molecular Biology*, 165(1):1–18, 1983.

[46] J. Janin. Specific versus non-specific contacts in protein crystals. *Nature Structural Biology*, 4(12):973–974, 1997.

[47] A. Janulaitis, M. Petrusyte, Z. Maneliene, S. Klimasauskas, and V. Butkus. Purification and properties of the *Eco*57I restriction endonuclease and methylase–prototypes of a new class (type IV). *Nucleic Acids Research*, 20(22):6043–9, 1992.

[48] A. Janulaitis, R. Vaisvila, A. Timinskas, S. Klimasauskas, and V. Butkus. Cloning and sequence analysis of the genes coding for *Eco*57I type IV restriction-modification enzymes. *Nucleic Acids Research*, 20(22):6051–6, 1992.

[49] A. Jeltsch, J. Alves, G. Maass, and A. Pingoud. On the catalytic mechanism of *Eco*RI and *Eco*RV: a detailed proposal based on biochemical results, structural data and molecular modeling. *FEBS Letters*, 304(1):4–8, 1992.

[50] A. Jeltsch, J. Alves, H. Wolfes, and A. Pingoud. Structure-assisted catalysis in the cleavage of DNA by the *Eco*RI and *Eco*RV restriction enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, 90:8499–8503, 1993.

[51] L. Jen-Jacobson. Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. [Review] [145 refs]. *Biopolymers*, 44(2):153–80, 1997.

[52] T. A. Jones, J. Y. Zou, S. W. Cowan, and M. Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica*, A47:110–119, 1991.

[53] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*, A32:922–923, 1976.

[54] L. Kauc and A. Piekarowicz. Purification and properties of a new restriction endonuclease from *Haemophilus influenzae* Rf. *European Journal of Biochemistry*, 92(2):417–26, 1978.

[55] T. J. Kelly, Jr and H. O. Smith. A restriction enzyme from *Hemophilus influenzae*. II. Base sequence of the recognition site. *Journal of Molecular Biology*, 51:393–409, 1970.

[56] C. Kessler and V. Manta. Specificity of restriction endonucleases and DNA modification methyltransferases – a review (edition 3). *Gene*, 92:1–248, 1990.

[57] Y. C. Kim, J. C. Grable, R. Love, P. J. Greene, and J. M. Rosenberg. Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, 249(4974):1307–9, 1990.

[58] S. Klimasauskas, S. Kumar, R. J. Roberts, and X. Cheng. *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell*, 76:357, 1994.

[59] H. Kong, R. D. Morgan, R. E. Maunus, and I. Schildkraut. A unique restriction endonuclease, *Bcg*I, from *Bacillus coagulans*. *Nucleic Acids Res.*, 21:987–991, 1993.

[60] D. Kostrewa and F. K. Winkler. Mg^{2+} binding to the active site of *Eco*RV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry*, 34:683–696, 1995.

[61] D. H. Krüger, G. J. Barcak, M. Reuter, and H. O. Smith. *Eco*RII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Research*, 16(9):3997–4008, 1988.

[62] U. Kühnlein and W. Arber. Host specificity of DNA produced by *Escherichia coli*: XV the role of nucleotide methylation in *in vitro* B-specific modification. *Journal of Molecular Biology*, 63:9–19, 1972.

[63] M. Kusiak, C. Price, D. Rice, and D. P. Hornby. The HsdS polypeptyde of the type IC restriction enzyme *Eco*R124 is a sequence-specific DNA-binding protein. *Molecular Microbiology*, 6:3251–3256, 1992.

[64] L. D. Landau and E. M. Lifschitz. *Lehrbuch der Theoretischen Physik. Band II. Klassische Feldtheorie.* Akademie-Verlag, Berlin, 6 edition, 1973.

[65] L. D. Landau and E. M. Lifschitz. *Lehrbuch der Theoretischen Physik. Band II. Klassische Feldtheorie*, chapter 8, pages 185–207. Akademie-Verlag, Berlin, 6 edition, 1973.

[66] A. Lomant and G. Fairbanks. Chemical probes of extended biological structures: synthesis and properties of the cleavable protein cross-linking reagent [^{35}S]dithiobis(succinimidyl propionate). *Journal of Molecular Biology*, 104:243–261, 1976.

[67] C. M. Lucas, R. Kucera, I. Schildkraut, and A. Aggarwal. Understanding the immutability of restriction enzymes: crystal structure of *Bgl*II and its DNA substrate at 1.5 Å resolution. *Nature Structural Biology*, 7(2):134–140, 2000.

[68] B. F. Luria and M. L. Human. A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology*, 64:557–569, 1952.

[69] T. McLellan. Electrophoresis buffers for polyacrylamide gels at various ph. *Analytical Biochemistry*, 126:94–99, 1982.

[70] A. Meisel, T. A. Bickle, D. H. Kruger, and C. Schroeder. Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*, 355(6359):467–9, 1992.

[71] A. Meisel, D. H. Krüger, and T. A. Bickle. M.*Eco*P15I methylates the second adenine in its recognition sequence. *Nucleic Acids Research*, 19:3997, 1991.

[72] N. E. Murray, P. L. Batten, and K. Murray. Restriction of bacteriophage lambda by *Escherichia coli* K. *Journal of Molecular Biology*, 81(3):395–407, 1973.

[73] P. D. Nathan and J. E. Brooks. Characterisation of clones of the *Bam*HI methyltransferase gene. *Gene*, 74:35–36, 1988.

[74] M. Newman, K. Lunnen, G. Wilson, J. Greci, I. Schildkraut, and S. E. Phillips. Crystal structure of restriction endonuclease *Bgl*I bound to its interrupted DNA recognition sequence. *EMBO Journal*, 17(18):5466–76, 1998.

[75] M. Newman, T. Strzelecka, L. F. Dorner, I. Schildkraut, and A. Aggarwal. Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, 269:656–663, 1995.

[76] M. Newman, T. Strzelecka, L. F. Dorner, I. Schildkraut, and A. K. Aggarwal. Structure of restriction endonuclease *Bam*HI phased at 1.95 Å resolution by MAD analysis. *Structure*, 2:439–452, 1994.

[77] M. Newman, T. Strzelecka, L. F. Dorner, I. Schildkraut, and A. K. Aggarwal. Structure of restriction endonuclease *Bam*HI and its relationship to *Eco*RI. *Nature*, 368(6472):660–4, 1994.

[78] M. O'Gara, X. Zhang, R. J. Roberts, and X.Cheng. Structure of a binary complex of *Hha*I methyltransferase with S-adenosyl-L-methionine formed in the presence of a short nonspecific DNA oligonucleotide. *Journal of Molecular Biology*, 287:201, 1999.

[79] A. V. Orekhov, B. A. Rebentish, and V. G. Debabov. A new site-specific endonuclease from *Streptomyces – Sgr*II. *Dokl. Akad. Nauk SSSR*, 263:217–220, 1982.

[80] Z. Otwinowski. Daresbury Study Weekend proceedings, 1991.

[81] Z. Otwinowski and W. Minor. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 276:307–326, 1996.

[82] O. V. Petrauskene, E. A. Karpova, E. S. Gromova, and W. Guschlbauer. Two subunits of *Eco*RII restriction endonuclease interact with two DNA recognition sites. *Biochemical & Biophysical Research Communications*, 198(3):885–90, 1994.

[83] M. P. Piatrushite, I. B. Bitinaite, D. R. Kershulite, S. I. Menkiavichius, and V. V. Butkus. [Restriction endonucleases of a new type]. [Russian]. *Doklady Akademii Nauk SSSR*, 295(5):1250–1253, 1987.

[84] A. Pingoud and A. Jeltsch. Recognition and cleavage of DNA by type-II restriction endonucleases. *European Journal of Biochemistry*, 246:1–22, 1997.

[85] M. Reuter, D. Kupper, A. Meisel, C. Schroeder, and D. H. Krüger. Co-operative binding properties of restriction endonuclease *Eco*RII with DNA recognition sites. *Journal of Biological Chemistry*, 273(14):8294–8300, 1998.

[86] M. Reuter, D. Kupper, C. D. Pein, M. Petrusyte, V. Siksnys, B. Frey, and D. H. Krüger. Use of specific oligonucleotide duplexes to stimulate cleavage of refractory DNA sites by restriction endonucleases. *Analytical Biochemistry*, 209(2):232–237, 1993.

[87] R. J. Roberts and D. Macelis. REBASE - restriction enzymes and methylases. *Nucleic Acids Research*, 28(1):306–7, 2000.

[88] J. Rosamond, B. Endlich, and S. Linn. Electron microscopic studies of the mechanism of action of the restriction endonuclease of *Escherichia coli* B. *Journal of Molecular Biology*, 129(4):619–35, 1979.

[89] G. Schluckebier, M. Kozak, N. Bleimling, E. Weinhold, and W. Saenger. Differential binding of S-adenosylmethionine, S-adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M.*Taq*I. *Journal of Molecular Biology*, 265(1):56–67, 1997.

[90] U. Selent, T. Ruter, E. Kohler, M. Liedtke, V. Thielking, J. Alves, T. Oelgeschlager, H. Wolfes, F. Peters, and A. Pingoud. A site-directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease *Eco*RV. *Biochemistry*, 31(20):4808–15, 1992.

[91] P. M. Sharp, J. E. Kelleher, A. S. Daniel, G. M. Cowan, and N. E. Murray. Roles of selection and recombination in the evolution of the type I restriction modification systems in enterobacteria. *PNAS*, 89:9836–9840, 1992.

[92] V. Siksnys, R. Skirgaila, G. Sasnauskas, C. Urbanke, D. Cherny, S. Grazulis, and R. Huber. The *Cfr*10I restriction enzyme is functional as a tetramer. *Journal of Molecular Biology*, 291(5):1105–18, 1999.

[93] R. Skirgaila, S. Grazulis, D. Bozic, R. Huber, and V. Siksnys. Structure-based redesign of the catalytic/metal binding site of *Cfr*10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. *Journal of Molecular Biology*, 279(2):473–81, 1998.

[94] P. Skowron, T. Kaczorowski, J. Tucholski, and A. J. Podhajska. Atypical DNA-binding properties of class-IIS restriction endonucleases: evidence for recognition of the cognate sequence by a *Fok*I monomer [published erratum appears in Gene 1994 Apr 8;141(1):151]. *Gene*, 125(1):1–10, 1993.

[95] J. D. Smith, W. Arber, and U. Kühnlein. Host specificity of DNA produced by *Escherichia coli*: XIV the role of nucleotide methylation in *in vivo* B-specific modification. *Journal of Molecular Biology*, 63:1–8, 1972.

[96] F. W. Studier and P. K. Bandyopadhyay. Model for how type I restriction enzymes select cleavage sites in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 85(13):4677–81, 1988.

[97] W. Szybalski, S. C. Kim, N. Hasan, and A. J. Podhajska. Class-IIS restriction enzymes–a review [published erratum appears in Gene 1991 Dec 20;109(1):169]. *Gene*, 100:13–26, 1991.

[98] T. Tao, J. C. Bourne, and R. M. Blumenthal. A family of regulatory genes associated with type II restriction-modification systems. *Journal of Bacteriology*, 173:1367–1375, 1990.

[99] V. Thielking, J. Alves, A. Fliess, G. Maass, and A. Pingoud. Accuracy of the *Eco*RI restriction endonuclease: binding and cleavage studies with oligodeoxynucleotide substrates containing degenerate recognition sequences. *Biochemistry*, 29(19):4682–91, 1990.

[100] W. Thomson. On the thermo-elastic, thermomagnetic and pyroelectric properties of matter. *Q. J. Math*, 1:57–77, 1897.

[101] P. H. Tran, Z. R. Korszun, S. Cerritelli, S. S. Springhorn, and S. A. Lacks. Crystal structure of the DpnM DNA adenine methyltransferase from the DpnII restriction system of *Streptococcus pneumoniae* bound to S-adenosylmethionine. *Structure*, 6:1563, 1998.

[102] Č. Venclovas and V. Siksnys. Different enzymes with similar structures involved in Mg^{2+}-mediated polynucleotidyl transfer [letter]. *Nature Structural Biology*, 2(10):838–41, 1995.

[103] H. Viadiu and A. K. Aggarwal. The role of metals in catalysis by the restriction endonuclease *Bam*HI. *Nature Structural Biology*, 5(10):910–916, 1998.

[104] H. Viadiu and A. K. Aggarwal. Structure of *Bam*HI bound to nonspecific DNA: a model for DNA sliding. *Molecular Cell*, 5:891–895, 2000.

[105] I. B. Vipond, G. S. Baldwin, and S. E. Halford. Divalent metal ions at the active sites of the *Eco*RV and *Eco*RI restriction endonucleases. *Biochemistry*, 34:697–704, 1995.

[106] D. A. Wah, J. A. Hirsch, L. F. Dorner, I. Schildkraut, and A. K. Aggarwal. Structure of the multimodular endonuclease *Fok*I bound to DNA. *Nature*, 388(6637):97–100, 1997.

[107] L. Wall. *Programming Perl, second edition.* O'Reilly & Associates, Cambridge, MA, 1996.

[108] L. M. Wentzell, T. J. Nobbs, and S. E. Halford. The *Sfi*I restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. *Journal of Molecular Biology*, 248(3):581–595, 1995.

[109] F. K. Winkler, D. W. Banner, C. Oefner, D. Tsernoglou, R. S. Brown, S. P. Heathman, R. K. Bryan, P. D. Martin, K. Petratos, and K. Wilson. The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *The EMBO Journal*, 12(5):1781–1795, 1993.

[110] S.-Y. Xu and I. Schildkraut. Isolation of *Bam*HI variants with reduced cleavage activities. *Journal of Biological Chemistry*, 266:4425–4429, 1991.

[111] C. C. Yang, B. K. Baxter, and M. D. Topal. DNA cleavage by *Nae*I: protein purification, rate-limiting step, and accuracy. *Biochemistry*, 33(49):14918–14925, 1994.

[112] C. C. Yang and M. D. Topal. Nonidentical DNA-binding sites of endonuclease *Nae*I recognize different families of sequences flanking the recognition site. *Biochemistry*, 31(40):9657–64, 1992.

[113] R. Yuan, D. L. Hamilton, and J. Burckhardt. DNA translocation by the restriction enzyme from *E. coli* K. *Cell*, 20(1):237–44, 1980.

[114] R. Yuan and J. Reiser. Steps in the reaction mechanism of the *Escherichia coli* plasmid P15-specific restriction endonuclease. *Journal of Molecular Biology*, 122(4):433–45, 1978.