Structural studies on metal-containing enzymes:

T4 endonuclease VII
and
D. gigas formate dehydrogenase.

H.C.A Raaijmakers

Promotor:

dr. N.C.M. Laane, voormalig hoogleraar in de biochemie.

Co-promotor:

dr. D. Suck, Groepsleider, EMBL Heidelberg.

Samenstelling promotiecommissie:

prof. dr.ir. I.M.C.M. Rietjens (Wageningen Universiteit)
dr. H.B.C.M Haaker (Wageningen Universiteit)
dr. P. Gros (Universiteit Utrecht)
prof. dr. eng. M. J. Romão (Universidade Nova de Lisboa).

H.C.A. Raaijmakers

Structural studies on metal-containing enzymes:

T4 endonuclease VII
and
D. gigas formate dehydrogenase.

Proefschrift

ter verkrijging van de graad van doctor
op gezag van de rector magnificus,
van Wageningen Universiteit,
prof. dr. ir. L. Speelman
in het openbaar te verdedigen op 23 april 2001
des namiddags te vier uur in de Aula

ISBN 90-5808-412-4

Voorwoord

Mijn interesse voor eiwit-kristallografie ontstond in 1994. Voor de laatste fase van mijn studie was ik op zoek naar een stageplaats. Ik had toen vijf maanden met veel plezier bij Huub Haaker in het natte lab gewerkt, en het werd het tijd voor iets drogers.

Net als iedereen bij de vakgroep biochemie, waar elke rondleiding besloten werd met een drie dimensionale blik op een eiwit, was ik gebiologeerd door het gebruik van grafische computers. Michel Eppink gaf mij een vijf-daagse crash-course "unix" en "kristallografie", waarbij ik terloops de structuur van twee van zijn vele hydroxylase mutanten mocht ophelderen. Daarna werden mij de grondslagen van kristallografie bijgebracht door Geoff Ford, tijdens mijn stage in Sheffield. Dat beviel erg goed. Ik besloot dan ook dit werk voort te zetten als PhD student aan het EMBL in Heidelberg. Na anderhalf jaar voorbereiding (!) braken de drie magische dagen aan, waarin ik mijn eerste eiwit model gebouwd heb. Dat is te vergelijken met het oplossen van een legpuzzel zonder voorbeeld, soms zonder rand en hoek stukjes, waar stukjes kunnen ontbreken of zelfs van een andere puzzel blijken te zijn. Het is een langdurig proces, maar zodra alle stukjes in elkaar passen is het resultaat indrukwekkend: De eerdere resultaten van vele onderzoekers blijken precies te passen en er zijn even meer antwoorden dan vragen. Dat is trouwens een reëel gevaar voor lopend onderzoek: Soms valt het onderzoek stil zodra de oude, interessante problemen opgelost zijn, terwijl nieuwe vragen nog niet bedacht zijn. Gelukkig biedt kristallografie veel vaker een stevig fundament voor verder onderzoek met nieuwe uitdagingen.

Vele mensen hebben direct of indirect bijgedragen aan dit proefschrift:

At the EMBL, I had my first job interview ever, talking to twelve group leaders in one day and Dietrich offered me a project on DNA binding proteins. You were a great help writing the papers, though I couldn't have finished my experiments without help from the other members of the group: Gunther, for all the cloning, protein expression and Jazz music; Hiang, for about everything dat doesn't involve computers, and for teaching me never to say "I'm off" again on Friday afternoons; Olivier, who started the work on EndoVII, and promoted the "never look back approach". Imre's help on the EndoVII project ranged from the expression of protein to the data collection at the synchrotron. In between there was enough time for lengthy discussions of European history. I wouldn't have enjoyed the lab as much without the other members of the lab: Eva, Regis, Sylke, Claudine, Joachim, Tom and Christophe.

I also got lots of help, joy and hacking experience from Yorgo, John and Brian (Sidney?) from Rik Wierenga's group next door, and also from Guillermo, Hans, Chris and Peter.

The social life was enormously enriched by all my fellow PhD students, who organised many parties and "excursions" and those who convinced me to keep swimming once a week (Frank, Gaia, Janneke, Louise) or play poker (John, Yorgo, Paul, Brian, Simon).

After Heidelberg, I went to Caparica (Could one invent a more exotic name?) in Portugal, to work with Maria João Romão on redox proteins. I'd like to thank my current group leader, for being so patient with my thesis and language, (Mas naturalmente, devia ter escrito esta parte em Português...) my colleagues that organised many excursions in Portugal in exchange for tulips and stroopwafels: Susana, Carlos, Ana Luísa, João Miguel, Cecília and Jorge; and also Helena, for the tour through Braga, among many other things.

Ik wil ook mijn promotor Colja Laane bedanken, die, ondanks de afstand altijd zeer geïnteresseerd was in mijn onderzoek. De halfjaarlijkse bezoeken waren altijd erg motiverend en vooral in de latere fase van mijn promotie erg belangrijk. In Wageningen was er na afloop ook altijd gelegenheid om met oude bekenden van de vakgroep biochemie bij te praten.

In het buitenland heb ik altijd op bezoek en gezelligheid kunnen rekenen van studiegenoten en ex-hoevestein bewoners. Soms werd dat handig gecombineerd met een cursus of congress. Bedankt Cathelijne, ook voor je bijdrage aan paragraaf 1.2, Bram, Katinka (Iris is nog niet langs geweest), Joost, Connie, Christine, Annechien, André, David, Rachel, Gert, Carrie, Marcel en Lisi.

Verder ontving ik veel steun, bezoek en pakketjes van mijn ouders, broer, zus en "schoon" familie, die langzamerhand gewend geraakt zijn aan de gestage voortgang van de wetenschap.

En Paula, het viel niet altijd mee, een vriendje op zo'n 500 km afstand te hebben, of om zelf 2000 km van huis te zijn. Bedankt voor je steun, geduld en organisatievermogen.

Contents

1

General introduction

1.1 Metalloproteins

Many processes in biology require metal ions, including respiration, DNA repair, nitrogen fixation, photosynthesis, muscle contraction and signal transduction. They are commonly found as natural constituents of proteins. Metals allow proteins to react with oxygen, carry out radical reactions, facilitate rapid electron transfer and they are often excellent in promoting catalysis. In addition, metals such as zinc, magnesium and calcium are often crucial for maintaining the proper 3D-structures of proteins and nucleic acids.

A search in the Protein Data Bank (Berman, 2000) reveals the presence of metal ions and other electron-dense ions in protein structures, as determined by X-ray crystallography or NMR (table 1.1). It shows which metals have been studied most in structural biology: calcium, iron, magnesium and zinc. These belong to the 15 metals, out of 29 elements, that are evidently indispensable for some form of life (Wilkins & Wilkins, 1997). The other, non-essential elements may have been studied for their clinical value, their toxicity or their inhibitory effects on enzymes. In addition, these atoms are used to localise the active site of enzymes, to mimic a transition state, or are required to obtain protein crystals.

In X-ray crystallography, heavy atoms are better resolved than light atoms. Thus, the detection of light elements can be enhanced if they are substituted by heavy ones; lanthanides can often substitute for calcium or magnesium. Molybdate, phosphate or sulphate can sometimes be replaced by tungstate. Proteins often require metal ions for their stability, for catalysis, or for electron transport. Besides that, metals might be associated with proteins for their transport or storage.

I) Structurally important ions.

Magnesium, calcium and zinc ions are often crucial for maintaining the proper 3D-structures of proteins and nucleic acids. r-RNA and t-RNA have magnesium ions associated with the phosphate groups. Magnesium ions are also known to stabilise

H																	He
Li 2	Be 11											**B**	**C**	**N**	**O**	**F**	Ne
Na 397	**Mg** 867											Al 42	**Si**	**P**	**S**	**Cl**	Ar
K 115	**Ca** 1246	Sc 0	Ti 0	**V** 16	**Cr** 3	**Mn** 286	**Fe** 1150	**Co** 147	**Ni** 85	**Cu** 224	**Zn** 952	Ga 4	Ge 0	**As** 23	**Se** 130	**Br** 37	Kr 10
Rb 4	Sr 11	Y 1	Zr 0	Nb 0	**Mo** 35	Tc 0	Ru 2	Rh 2	Pd 0	Ag 3	Cd 90	In 1	**Sn** 0	Sb 3	Te 1	**I** 57	Xe 12
Cs 8	Ba 5	La 2	Hf 0	Ta 0	**W** 30	Re 2	Os 2	Ir 2	Pt 43	Au 5	Hg 121	Tl 3	Pb 8	Bi 0	Po 0	At 0	Rn 0

Ce 3	Pr 0	Nd 0	Pm 0	Sm 3	Eu 2	Gd 6	Tb 2	Dy 0	Ho 3	Er 0	Tm 0	Yb 6	Lu 2
Th 0	Pa 0	U 1											

Table 1.1: Abundance of the metals and other electron dense elements in the Protein Data Bank (10-10-2000). Counted are the elements and their ions, but not organic compounds of non-metals. The table is biased towards heavier elements as they are easier seen by crystallography, and can be used as markers. To distinguish Na^+, K^+, Mg^{2+}, Cl^- or F^- from water one needs clear electron density and/or liganding patterns. Still, some metals are more common than others: Ca, Fe, Zn, Mg, Na, Mn, and Cu are clearly more investigated than other metals. In bold are the essential elements according to (Wilkins & Wilkins, 1997).

protein/DNA complexes, but sometimes they can be replaced by manganese or calcium. Quite often, these divalent cations do not bind to the protein itself, but bind only upon complex formation. In proteins, structurally important magnesium can bind to six, calcium to six or seven carbonyl-oxygens simultaneously. Zinc is usually bound to cysteine of histidine side-chains. Especially when surrounded by four cysteines, the zinc is tightly bound to the protein in an obviously structural role. This is the case in endonuclease VII, whose fold depends on a zinc ion (chapter 2). This zinc is liganded to two groups of two cysteines and links helix-1 firmly to the four-helix cross, stabilizing the beta-finger in between.

Calcium plays a structural role in, e.g. thermolysin, where it enhances thermostability (Dahlquist *et al.,* 1976). Sometimes metal ions stabilise, reversibly, a certain conformation of a protein, depending on the concentration of the metal ion. This creates a regulatory role; e.g. in calmodulin (Babu *et al.,* 1988) calcium ions stabilise a particular conformation, if the concentration is high enough. In that conformation, calmodulin will activate other enzymes.

II) Catalytic ions.

Metal ions can have a higher charge-density than organic compounds and can help the enzyme to stabilise substrates or transition states with the opposite charge. For example, the phosphodiester bond is very stable, even in 1 M NaOH, due to the electrostatic repulsion of the attacking hydroxide by the negative charge of the phosphate group (Chin *et al.*, 1989; Westheimer, 1987). As a rule, proteins that interact with (poly-) nucleotides use a magnesium ion to counter the charge of the phosphate.

Both magnesium and zinc ions are commonly employed as Lewis acids in enzymatic hydrolysis reactions. Water co-ordinated to a metal atom is more acidic then "free" water, and hence a better nucleophile, though not as good as the hydroxyl ion. As a general rule Zn^{2+} is used where the substrates bear a carbonyl functional

group (esters, amides, CO_2) and Mg^{2+} is found as a cofactor of enzymes that catalyse the hydrolysis or formation of phosphate esters. (Cowan, 1993).

Proteins involved in redox reactions or electron transport must be able to "store" electrons in a redox-active group. In the latter, redox centres pass on the electrons, but are not themselves involved with substrates. Some redox enzymes contain an organic redox active group, like flavin, ubiquinone or topa quinone. Alternatively, redox proteins may contain transition metals as redox active cofactors. Transition elements have partially filled valency orbitals and usually exist in two or more redox states. Many transition metals are found in proteins involved in redox processes; vanadium, manganese, iron, cobalt, nickel, copper, molybdenum and tungsten. Iron, the most abundant of these, is also the most common metal in electron transport proteins, where it is present as a heme (cytochromes) or as iron-sulphur clusters.

Tungsto- and molybdoenzymes function primarily in oxygen atom transfer, in which the metal enzyme cycles between oxidation states VI and IV, with the exception of nitrogenase (Mo) acetylene hydratase (W) and formate dehydrogenase (Hille, 1996; Rosner & Schink, 1995; Khangulov *et.al,* 1998). In transferring oxygen, these enzymes catalyse a hydroxylation reaction, where water is used as the source of the oxygen atom (Hille, 1994):

$$RH + H_2O \leftrightarrow ROH + 2e^- + 2H^+$$

Tungsten-containing enzymes have been found in a variety of (mainly hyperthermophilic) archaea and two *Clostridium* species. Several of these tungsten enzymes are known to have a molybdenum-containing counterpart. For example, carboxylic acid reductase from *C. formicoaceticum* was shown to occur in a W and a Mo form. It was shown that these two forms are not iso-enzymes, but catalytically distinct enzymes (Huber, 1994). Tungsten also sometimes substitutes for molybdenum in case of Mo-deficiency (Andreesen *et al*., 1974), but often these enzymes are inactive, e.g. in nitrogenase from *Azotobacter vinelandii* (Nagatani & Brill, 1994), formate dehydrogenase and nitrate reductase from *Escherichia coli* (Enoch & Lester, 1972) and xantine oxidase (Johnson *et al.*, 1974). The formate dehydrogenase from *D. gigas* (chapter 4) is a rare example of a tungstoenzyme of a mesophile bacterium, which is expressed even though molybdenum is available.

1.2 A short introduction into crystallographic methods.

The knowledge of the 3 dimensional structure is vital to understand the function of enzymes. Two techniques are widely used for structure determination at atomic resolution, namely X-ray diffraction and nuclear magnetic resonance (NMR). The two techniques use a very different approach towards solving the structure. This starts already at the level of the samples. X-ray diffraction is applied to well ordered crystals of the protein, while NMR measures the proteins in solution. In X-ray diffraction, X-ray beams are cast at the crystals of the protein under investigation. The atoms of the protein are made visible via the interference pattern of the beam, which depends on the positions of all the atoms. In NMR, the protein sample is put in a magnetic field and radiated with waves at the radio frequency. The structure can be determined after performing several experiments, which will result in a distance map of atoms of less than 5 Å (10^{-10}m) apart and provide torsion angle restraints. Both

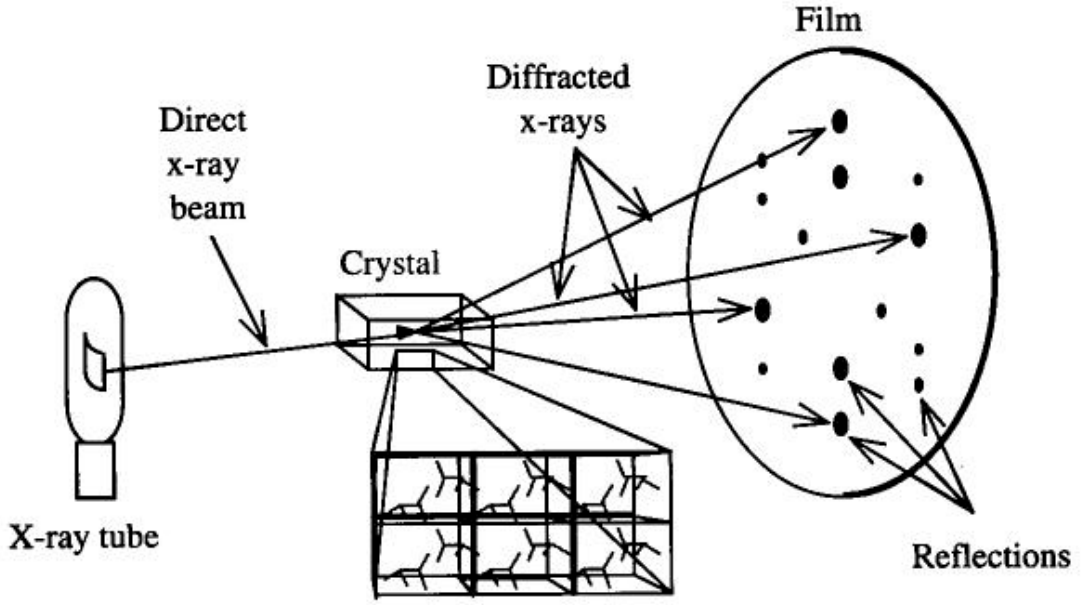

Figure 1.1: Crystallographic data collection. The crystal diffracts the source beam into many discrete beams, each of which produces a distinct spot (reflection) on the detector. The positions and intensities of these reflections contain the information needed to determine molecular structures (Rhodes, 1993).

techniques have their advantages and disadvantages. NMR can study both structure and dynamics of proteins, but only up to a molecular weight of ~30 kDa. X-ray diffraction studies molecules in conditions in which the proteins are rigid, but can solve the structure of very large proteins and protein complexes. The largest complex solved to date is the ribosome; its structure contains nearly 100,000 non-hydrogen atoms (Ban, 2000).

Diffraction

When X-rays interact with matter, they are bent (diffracted) by the individual electrons. The sum of all X-rays diffracted in all directions by all electrons in the molecule is called the structure factor $\mathbf{F}_{hkl}$. Technically it is impossible to measure diffraction of single molecules, the intensity of the diffraction pattern is far too low. One needs a huge collection of molecules that diffract parallel and additive to amplify the signal; i.e. they need to be aligned in a crystal, to be able to measure the diffraction.

The diffraction of a three dimensional crystal is described by the structure factor $\mathbf{F}_{hkl}$, which consists of an amplitude, $|\mathbf{F}_{hkl}|$, and a phase α_{hkl} for each reflection (h,k,l). The amplitude is obtained as the square root of the measured intensities, while the phase information is lost.

The structure factor can be seen as a Fourier transform {equation 2} of the electron density in the crystal, or alternatively {equation 3}, as the vector-sum of the scattering contributed by each atom j. Now, since each atom is at a particular position (x,y,z), the phase they contribute, but not their relative weight, depends solely on their position in the unit cell.

$$\{1\} \quad \mathbf{F}_{hkl} = |\mathbf{F}_{hkl}| \cdot e^{i\alpha_{hkl}}$$

$$\{2\} \quad \mathbf{F}_{hkl} = \int\int\int \rho(x,y,z) \cdot e^{2\pi i(hx+ky+lz)}\, dx\, dy\, dz$$

$$\{3\} \quad \mathbf{F}_{hkl} = \Sigma_j f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

$$\{4\} \quad \rho(x,y,z) = \Sigma_{hkl} |\mathbf{F}_{hkl}| \cdot e^{i\alpha_{hkl}} \cdot e^{-2\pi i (hx+ky+lz)}$$

$\mathbf{F}_{hkl}$ = Structure factor, consisting of an amplitude and a phase for each reflection h,k,l. It is also valid to consider the equations for individual reflections.
h,k,l = Miller indices for each reflection, vector in reciprocal space.
α_{hkl} = Phase of reflection h,k,l
$\rho(x,y,z)$ = Electron density as a function of real space.
f_j = Atomic structure factor for atom j.

Phasing

The Fourier summation {equation 4} can be used to calculate the electron density, which provides the necessary information to solve the molecule structure. Obviously one needs somehow to obtain (an approximation for) the phase of each reflection, to calculate the electron density. Three phasing techniques are commonly used in protein crystallography: Molecular Replacement (MR), Multiple Isomorphous Replacement with heavy metals (MIR) and Multiple-wavelength Anomalous Dispersion (MAD).

Molecular replacement

In Molecular Replacement, electron density is calculated {equation 4} from measured intensities, while the phases are derived from a model. It can be used if a reasonably good model is available, usually a mutant structure, a modelling result, or a protein with a similar fold, as indicated by sequence homology. Of critical importance are the similarity of the three dimensional protein structure (yet undetermined!) to the model, the software used and experience of the crystallographer.

Since the phases of the atomic structure factor depend solely on the location of the atoms in the unit cell, the phasing model must be superimposed on the unknown structure of the new protein. This can be done by calculating {equation 3} the structure factor of the model in a particular position and orientation in the unit cell. This can be repeated for every possible rotation and translation. The rotation and translation that give calculated structure factors with the highest correlation to the measured one, is deemed correct. In practice, one can separate the rotation and translation search. The rotation correlation function (Rossmann & Blow, 1962) will be calculated in steps of 5 to 10 degrees, in three dimensions. After proper orientation of the model, a translation search is carried out. Depending on the space group, up to three dimensions have to be searched in small steps, each roughly 1/3 of the resolution of the data. This can be very quick with modern computers. In clear cases, the right solution will give a much better correlation to the measured structure factors then the second best, or other solutions.

Isomorphous replacement

In Multiple Isomorphous Replacement, one, or a few electron-dense atoms are inserted at a specific position in the crystal, without changing the rest of the structure.

This changes the intensities of the measured reflections. Once the heavy atom positions have been determined, the differences can be used to estimate phases for each reflection.

Part of the trick stems from f_j in equation 3: The diffraction contributed by each atom, f_j is proportional to the square of the number of their electrons (assuming equal temperature factors). In the case of T4 endonuclease VII, described in chapter 2 of this thesis, there are 2552 non-hydrogen atoms in the dimer, which have about 15000 electrons. Each mercury contributes 80 electrons. The square contribution is about 7^2 x 2552 = 125048, while the mercury contributes 2 x 80^2 = 12800; more than 10% of the total scattering. Theoretically the "root-mean-square" difference for each individual reflection R_{iso} is then about $0.10^{½}$ or 32%, which can easily been measured: The estimated accuracy of the measurement of each reflection, R_{sym} is 6.8% for the mercury derivative and 6.0% for the native structure (chapter2). In practice R_{iso} is 29.4%, which also includes errors like non-isomorphism of the crystals, inaccuracy of the mercury position and a temperature factor.

According to equation 3, for a completely isomorphous derivative crystal, the structure factor $\mathbf{F}_{HP}$ (for heavy atom plus protein) equals $\mathbf{F}_P$ (protein alone) + $\mathbf{F}_H$ (contribution of the heavy atom). For small values of F_H this would also be true for the amplitudes alone: $|\mathbf{F}_{HP}| = |\mathbf{F}_P| \pm |\mathbf{F}_H|$. So $(|\mathbf{F}_{HP}| - |\mathbf{F}_P|)^2$ would give the diffraction pattern for the heavy metal alone. The heavy atom structure is usually very simple, and can be determined by, for example, Patterson methods (Patterson, 1935). Once we can determine this structure, the complete structure factor $\mathbf{F}_H$ is known: The phases depend only on the position of the atoms. Once $\mathbf{F}_H$ is known, only two possible phases remain for each reflection F_P (figure 1.2). This ambiguity can be resolved by using a second derivative, which has electron-dense atoms at different positions.

Anomalous scattering

A second means of obtaining phases from heavy atom derivatives results from their capacity to absorb X-rays. As a result, reflections h,k,l and -h, -k, -l (measured when the path of the direct X-ray beam is inversed) have different intensities. This inequality of symmetry-related reflections is called anomalous dispersion, and the effect is strongest near the absorption edge of the element. As a result two values for F_H (and F_{PH}) are obtained, F_H^+ and F_H^- and the phase ambiguity (figure 1.2) can be resolved with one derivative.

Not only the anomalous difference ($F_H^+ - F_H^-$) but also the average length of F_H, $(F_H^+ + F_H^-)/2$ is wavelength dependent. The latter becomes only apparent after a second wavelength has been measured and gives rise to a "dispersive difference". With a clever choice of wavelengths for the measurements, both anomalous and dispersive differences can be exploited to resolve the phase ambiguity, with the use of only one crystal that contains a suitable electron-dense element (MAD).

After an initial estimate for the phases has been obtained, the electron-density has to be interpreted. A major problem is the limited resolution of the data. An atomic model is being built, but usually, protein crystals do not diffract to atomic resolution (carbon-carbon distance). To overcome this, assumptions about the protein are put into the model in the form of restraints and constraints. This can be anything known about the protein and usually includes bond-distances and angles, torsion angles, chirality's, temperature-factor distributions, van der Waals radii and obviously, the protein sequence. Once an initial model is built, phases can be estimated more

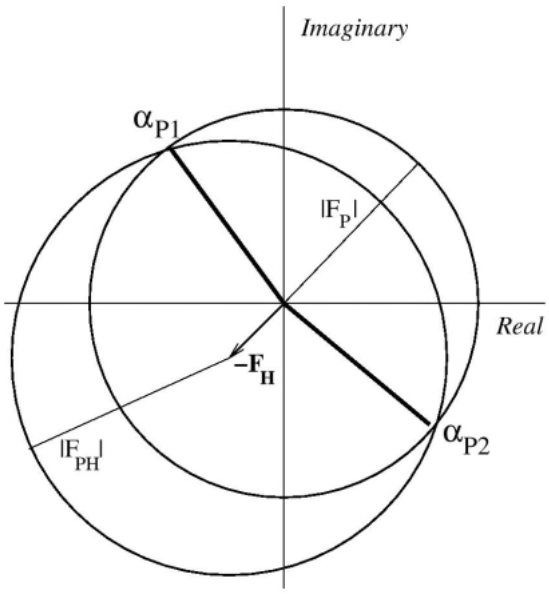

Figure 1.2: The structure factor for one reflection in the complex plane. The length of the vector represents the amplitude, while the angle to the horizontal axis represents the phase. The length and direction of F_H are known (arrow), while the length of $|F_P|$ and $|F_{PH}|$, and unknown direction, can be represented by a circle. Only two possible phases remain for F_P.

accurately, which are used to build a better model. This iterative procedure leads to the final crystal structure.

Protein dynamics

Proteins often require rearrangement of side-chains or of their (tertiary) structure for their functioning. In general, it is difficult to determine this dynamic behaviour by crystallography: Molecules in X-ray diffraction studies are fixed in a rigid lattice, their movement restricted by crystal contacts. Further more, the crystal structure is derived of electron density that has been averaged over all unit cells, during hours of data collection. Therefore, crystallography gives a static structure, representing one stable conformation of a dynamic molecule.

The most obvious indicators of dynamic behaviour in a refined structure are the temperature-, or B-factors for each atom. They are a measure of freedom of movement, proportional to the mean-square displacement of the atom. Unfortunately, this displacement is measured relative to the crystal lattice, not relative to other parts of the protein: A very rigid part of a domain could show a high temperature factor, just because the domain moves as a whole.

Another, infrequently used method to analyse dynamic behaviour, is the refinement of "TLS" parameters (Schomaker and Trueblood, 1968). It can point out the magnitude, directionality and origin of translational, rotational and screw vibrations. Instead of B-factors for each atom, 20 displacement parameters (Translation, Libration and Screw component) are refined for a rigid body, which might be a side chain, a helix or beta-sheet, a domain or even a whole protein, depending on interest and data resolution. This method has been applied to T4 endonuclease VII (chapter 3), but evidently, the usefulness as an analysis tool is

limited by the crystal contacts. However, the refinement statistics of the lower resolution structures improved significantly.

A third way to quantify the dynamic properties of the protein is to study different crystal forms. Since crystal contacts are different for different crystal forms, the limitations they impose on the conformation of the protein are also different. Usually there will be only a few crystal forms available, but they can give a rough idea of possible conformations. In chapter 3, three crystal forms are compared.

1.3 The role of endonuclease VII in phage T4

Phage T4

Bacteriophages, viruses that infect bacteria, became of scientific interest in 1939 when Max Delbrück set out to understand their multiplication. He restricted his work to a set of seven phages, T1 through T7, which are all active on *Escherichia coli* strain B. Within this set the closely related T2, T4 and T6 phages form the "T-even" group showed more then 85% sequence homology (Delbrück, 1946). These T-even phages were used in the early fifties to point out the genetic role of DNA. In 1961 Francis Crick and co-workers used T4 to show that the genetic code is read as triplets. (Crick *et al*, 1961)

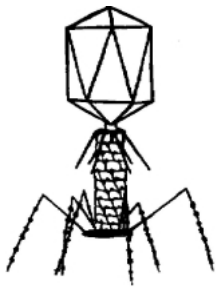

Figure 1.3 Phage T4

Like all phages, or viruses, T4 is dependent on a specific host for reproduction. When T4 infects *E. coli,* it attaches itself to the surface and injects its DNA into the cell. A few minutes after infection, the *E. coli* RNA polymerase starts to transcribe the first genes; these are the "early" enzymes. These proteins stop host transcription, reduce the host DNA to nucleotides and start the synthesis of T4 DNA. A special polymerase replicates the T4 DNA; it uses 5-hydroxymethylcytosine instead of cytosine (Allen *et al.,* 1983). Then, after synthesis, the DNA is glucosylated. These modifications protect the phage-DNA against restriction enzymes and also allow discrimination between viral and host DNA (Hermann *et al*., 1983).

Five minutes after infection starts the transcription of the "late"-genes. These code largely for the components of new viral capsids. The capsids assemble simultaneously: The phage heads are made in one pathway, the tail fibres in another, and the base-plate of the tail in yet two other pathways, which converge in a common pathway for tail assembly. The heads are then filled with DNA and joined to the tails, and then the tail-fibres are added. At the end of the multiplication cycle the host cell lyses, and about 200 new phages are released into the medium (Kutter *et al.*, 1983).

An unusual feature of phage T4 is that its genetic map is circular even though the DNA in each virion is linear (Streisinger *et al*., 1964). Each phage genome contains a short terminal redundancy, such that genes found on one end are repeated on the other end. The replicated DNA molecules can recombine with one another through their terminal redundancies, creating long molecules, many genomes in length. Then each virion is made by packaging a head full of DNA cut from such a molecule, each containing about 105% of the full genome size (Black & Showe, 1983).

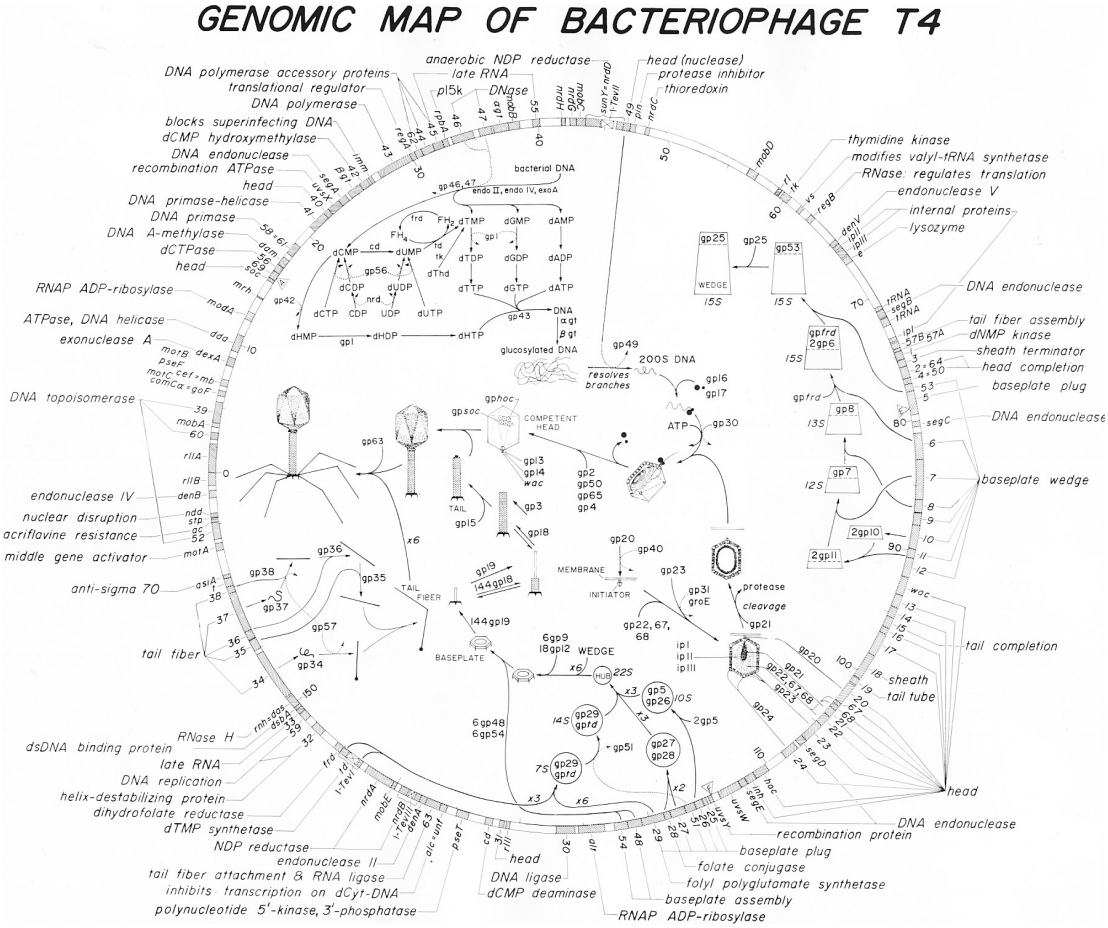

Figure 1.4 A genetic map of T4.

T4 genome

In 1964 Epstein *et. al*. mapped and numbered 50 genes, out of 87 protein coding genes known today, containing conditional lethal mutations, and classified them in groups with related phenotypes. Of these, gene product 49 (gp49) codes for endonuclease VII (EndoVII). Genes 16, 17 and 49 are classified as genes responsible for packaging newly synthesised phage DNA into newly synthesised phage heads. Mutants in either gene cause the accumulation of unpacked DNA, while the phage heads appear empty (Epstein *et al*. 1963). While the heads generated by gene 16- and 17-defective mutants are completely empty, the heads in gene 49- defective mutants are partly filled (Luftig & Ganz, 1972). This classifies mutations in genes 16 and 17 as packaging-initiation deficient and mutations in gene 49 as packaging-continuation deficient. The precursor DNA's in 16-/17- infections are markedly different from the DNA in 49- infections; the latter are unusual stable and compact as a result of an accumulation of uncleaved branches (Frankel, 1971). The product of gene 49 (gp49) proved to be a nuclease, specific for resolving these branches: 49- infection heads can be filled, but the branches in the DNA obstruct packing of the total genome into the phage-heads (Kemper & Brown 1976). This nuclease, now known as endonuclease VII, is the first "Holliday junction resolvase" discovered, i.e. an enzyme that specifically cleaves four-way DNA junctions (Holliday, 1962; Mizuuchi *et al*., 1982).

EndoVII

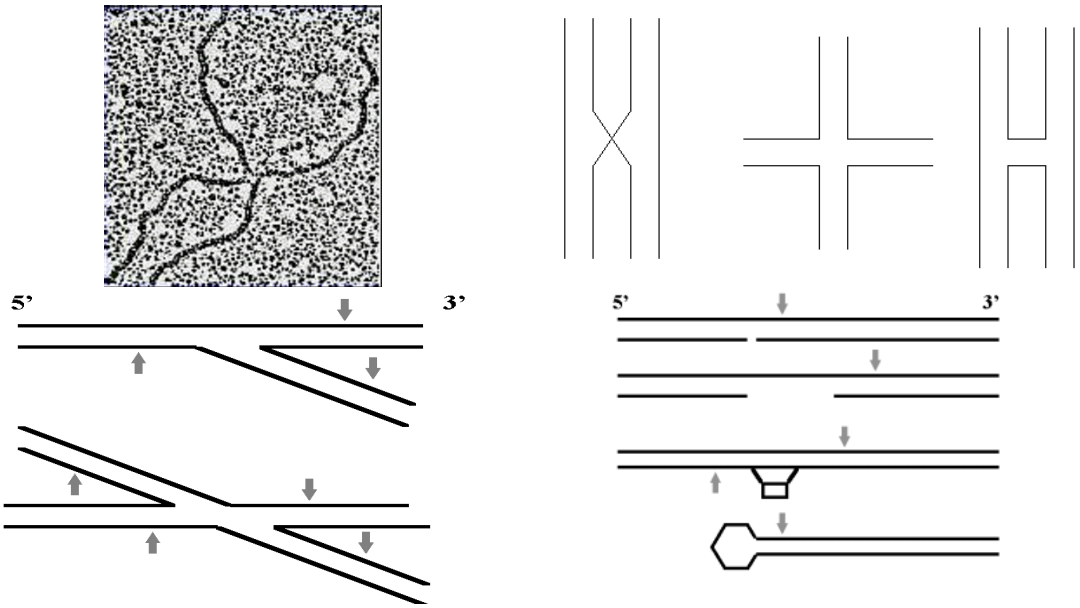

Figure 1.5: Top: EM picture of a Holliday junction. Top-right: Schematic diagram of the Holliday junction and its isomers. A distinction can be made between the continuous strand and the exchanging strand, for the isomers on the right and left.
Below: Bacteriophage T4 endonuclease VII is a nuclease that is selective for 3 and 4 way DNA junctions, base pair mismatches, nicked DNA, gaps, cis-platin adducts and loops. EndoVII cuts 3' of the irregularity (arrow).

The resolving of DNA junctions is an essential step in DNA recombination and repair. Proteins that bind and resolve these junctions have been found in many organisms. Crystal structures of complexes are available for *E. coli* RuvA, (Roe *et al*., 1998) and CCE1 (Gopaul *et al*., 1998). These nucleases are more specialised than EndoVII: They act only on junctions, while other enzymes exist for mismatch repair. The ability of EndoVII to recognise mismatches makes it very interesting for diagnosis of inherited diseases: It allows detection of mutations without sequencing of the DNA.

In recent years, several mutants of EndoVII have been isolated, which are totally defective in cleaving DNA, but on the other hand, show normal binding affinity to four-way junctions (Pöhler *et al*., 1996). In addition, it was found that the binding of endonuclease VII to branched DNA is rather a structure- and not a sequence-selective process (Bhattacharyya *et al*., 1991). The binding constants of endonuclease VII-branched DNA complexes are in the range of 10-100nM (Giraud-Panis & Lilley, 1996).

The resolution of a junction into double stranded species means cleavage of both strands, which would involve two active sites. Indeed, endonuclease VII acts as dimer on branched DNA (Pöhler, *et. al*., 1996) and also in solution, it forms a dimer, though the exchange of the subunits is a relatively fast process (Kosak & Kemper, 1990). Both strands are cleaved within the lifetime of the protein-DNA complex, and this time appears to be long enough for repair with a ligase (Giraud-Panis & Lilley, 1997).

Holliday junction

Two independent groups have recently solved crystal structures of the four-way DNA junction (Ortiz-lombardía *et al.,* 1999; Nowakowski *et al.,* 2000). The structures have been determined in different crystal forms and differ largely in conformation. The DNA junctions, crystallised in the presence of magnesium, fold into X-shaped structures, while the arms cross at an angle of 40º, 55º or 135º in the different structures. In the absence of cations, the four-way DNA junction adopts a structure in which the helices are unstacked, and extended towards the corners of a square (Ducket *et al*., 1988). Figure 1.5 sketches several isomers of this dynamic structure.

1.4 Biological background of D. gigas formate dehydrogenase.

Formate dehydrogenases (FDH) are a diverse group of enzymes, found in both eukaryotes and prokaryotes that catalyse the reversible two-electron oxidation of formate to carbon dioxide. It allows the assimilation of carbon dioxide or provides energy for growth through oxidative phosphorylation coupled to the reduction of p.e. oxygen, nitrate, sulphate or fumarate.

The formate dehydrogenases from aerobic bacteria and yeast reduce NAD^+, usually without using other cofactors or metals (Ferry, 1990). An exception is *P. oxalaticus* FDH, that contains a flavin to transfer electrons from one-electron iron-sulfur centers to the obligate two-electron accepting NAD^+ (Müller *et al*., 1978).

FDHs found in anaerobic organisms contain a molybdenum or a tungsten ion, which is bound to two molybdopterin cofactors. They also contain an abundance of iron-sulfur clusters; some contain additionally cytochromes and/or FAD (Ferry, 1990). Selenium is commonly found in these enzymes. If present, it is found as a selenocysteine, which ligands the molybdenum, like in *E. coli* FDH-H. Interestingly, the selenium independent FDH from *M. formicicum* reveals high overall homology and contains a cysteine in the corresponding position (Schuber *et al*., 1986).

So far, only one FDH has been discovered in *D. gigas*, but the related *D. vulgaris* has probably two homologue FDHs. *E. coli* has three related FDH isoenzymes that are expressed under different growth conditions: *E. coli* FDH-H is expressed during fermentative growth and is part of the hydrogen lyase complex, which decomposes formate into carbon dioxide and dihydrogen. FDH-O is expressed in the presence of oxygen and FDH-N is expressed when nitrate is available, but not oxygen. The latter two are part of a respiratory chain (Sawers, 1994).

Formate dehydrogenases from anaerobic bacteria and archae are usually extremely sensitive to oxygen. However, FDH from *Desulfovibrio gigas,* a sulphate reducing bacterium, is air stable. It looses activity upon exposure to air, but can be reactivated under anaerobic conditions with ß-mercapto-ethanol (Almendra, 1999).

The anaerobe FDHs belong to the DMSO reductase family, one of the four families that contain a mononuclear molybdenum or tungsten. The family trait is reflected in the Mo/W liganding pattern; they are coordinated by two molybdopterin guanosine dinucleotide (MGD) cofactors and one amino acid side chain. This can be serine, cysteine or selenocysteine, as found in resp. *R. capsulatus* DMSO reductase *D. gigas* periplasmic nitrate reductase and *E. coli* FDH-H (Schneider *et.al.,* 1996; Dias *et. al.,* 1999; Boyington *et. al.,* 1997). Interestingly, the FDH from *M. formicum* reveals a high sequence homology with FDH-H, but uses a cysteine instead of the selenocysteine (Schuber *et. al.,* 1986).

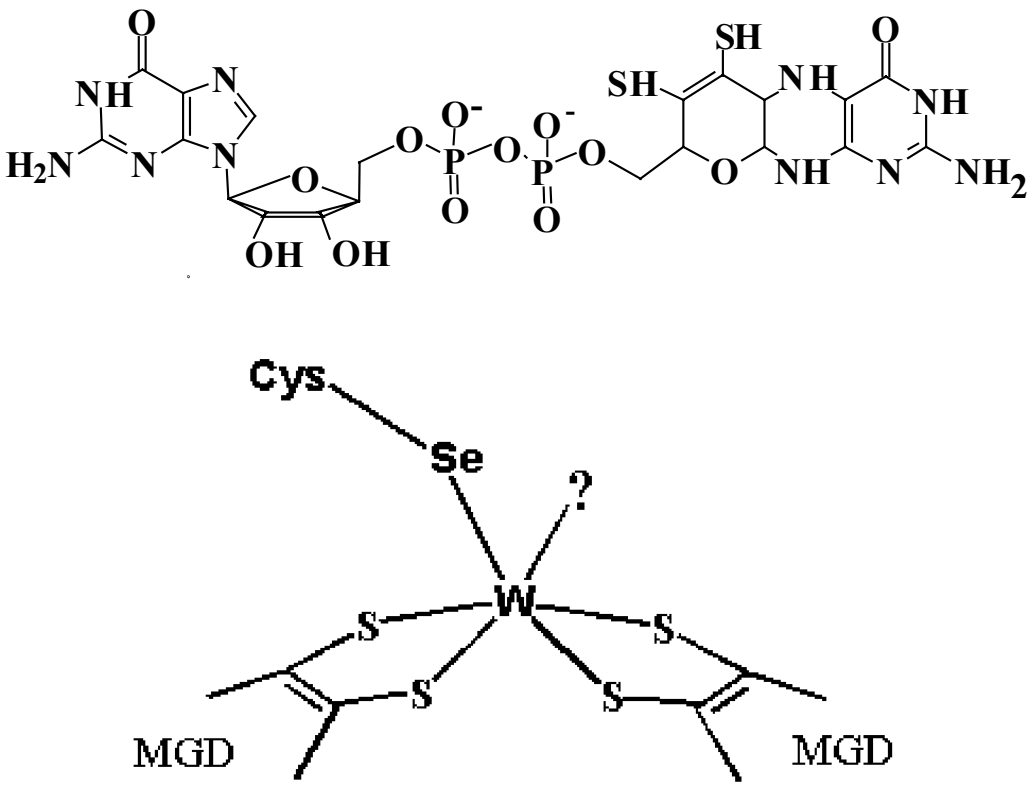

Figure 1.6 Top: Molybdopterin guanosine dinucleotide (MGD) Bottom: The tungsten in *D. gigas* FDH is liganded to two MGD cofactors and a selenocysteine, the sixth ligand could be water.

1.5 Outline of thesis

For many protein functions, metal ions are indispensable. The role of the metal ions can be quite diverse and varies from catalysis and electron transport to structural stability and signalling. In this thesis, two of such metalloproteins are investigated by crystallographic methods, to elucidate the action of the enzymes and the role of the metal therein.

This thesis starts with the structure determination of T4 endonuclease VII, a small nuclease that cleaves flexible, non-standard, DNA. Its structural stability depends on two tightly bound zinc ions, while two calcium ions are found as part of the active sites and could be used to model protein-DNA interactions. The substrate selectivity seemed to depend on the flexibility of the protein and the DNA and has been further studied in chapter 3, using three different crystal forms. A high-resolution structure (1.3 Å) revealed many details that were used to propose a reaction mechanism.

Chapter 4 describes a crystallographic study on the redox-enzyme formate dehydrogenase from *Desulfovibrio gigas*. This is the first tungsten-containing enzyme isolated from a mesophile, which has not been depleted from molybdenum. The preliminary data presented exploit the specificity of the X-ray absorption edges to locate and identify iron, tungsten and selenium atoms. A selenium atom and two out

of four [4Fe-4S] clusters had been overlooked with previously published spectroscopic and metal analysis methods. The active site of formate dehydrogenase contains a tungsten atom bound by two molybdopterin ligands and a selenocysteine, while four [4Fe-4S] clusters form an electron transfer pathway from the active site to the exterior of the protein.

References

Allen JR, Lasser GW, Goldman DA, Booth JW, Mathews CK.(1983) *J Biol Chem*. **258** 5746-53.

Almendra MJ, Brondino CD, Gavel O, Pereira AS, Tavares P, Bursakov S, Duarte R, Caldeira J, Moura JJG, Moura I (1999) *Biochemistry* **38**: 16366-16372

Andreesen JR, El Ghazzawi E and Gottschalk G (1974) *Arch. Microbiol*. **96**, 103-118.

Androphy EJ, Hubbert NL, Schiller JT, Lowy DR, *EMBO J.,* **6**, 989 (1987)

Babu YS, Bugg CE and Cook WJ (1988) *J.Mol.Biol*. **204**, p191

Ban N, Nissen P, Hansen J, Moore PB and Steitz TA. (2000) *Science*. **289**, 905-20.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. (2000) The Protein Data Bank. *Nucleic Acids Res*. **28,**235-42.

Bhattacharyya A, Murchie AIH, von Kitzing E, Diekmann S, Kemper B, Lilley DMJ, *J. Mol. Biol*., **221**, 1191 (1991)

Boyington JC, Gladyshev VN, Khangulov SV, Stadtman TC, Sun PD (1997) *Science* **275** 1305-1308.

Chin, J., Banaszczyk, F., Jubian, V. & Zou, X. (1989). *J. Am. Chem. Soc*. **111**, 186-190.

Cowan JA, (1997) Inorganic Biochemistry, VCH, New York.

Crick F, Barnett L, Brenner S and Watts-Tobin, R (1961) *Nature*, **192** 1227-1232.

Dahlquist FW, Long JW and Bigbee WL (1976) *Biochemistry*, **15**, 1103

Delbrück M (1946) *Biol. Rev.* **21**:30-40

Dias JM, Than ME, Humm A, Huber R, Bourenkov GP, Bartunik HD, Bursakov S, Calvette J, Caldeira J, Carneiro C, Moure J. Moura I and Romão MJ (1999) *Structure* **7,** 65-79.

Duckett D, Giraud Panis MJ and Lilley D. (1995) J. Mol. Biol. **246**, 95-107.

Duckett D, Murchie A, Diekmann S, von Kitzing E, Kemper B. & Lilley D (1988) *Cell*, **55**, 79-89.

Enoch HG and Lester RL (1972) *J.Bacteriology* **110,** 1032-1040.

Epstein R, Bolle A, Steinberg C, Kellenberger E, Boy de la Tour E, Chevalley R, Edgar R, Susman M, Denhardt C and Lielausis I (1964) *Cold Spring Harbour Symp. Quant. Biol*. **28**: 375-392.

Ferry JG (1990) *FEMS rev*. **87,** 377-382.

Frankel F, Batcheler M and Clark C (1971) *J. Mol. Biol*. **62**: 439-463.

Giraud-Panis MJE, Duckett DR, Lilley DMJ (1995) *J. Mol. Biol*., **252**, 596

Giraud-Panis MJE, Lilley DMJ. (1997) *EMBO J*., **16,** 2528-34

Giraud-Panis MJE, Lilley DMJ (1996) *J. Biol. Chem*., **271**, 33148

Golz S, Christoph A, Birkenkamp-Demtr?der K, Kemper B (1997) *Eur. J. Biochem*., **245**, 573

Gopaul DN, Guo F, Van Duyne GD. (1998) *EMBO J.* **17**:4175-87.

Herman RE, Haas N, Snustad DP. (1984). *Genetics* **108** 305-17.

Hatful GF, Sarkis GJ (1993) *Mol. Microbiol*., **7**, 395

Hensgens CM, Hagen WR, Hansen TA (1995) *J Bacteriol*. **177**, 6195-6200.

Hille R (1994) *Biochim. Biophys. acta,* **143**, 143-169.

Hille R (1996) *Chem. Rev.* **96**, 2757-2816

Holliday R (1964) *Genet. Res*. **5** 282-304.

Huber C, Caldeira J, Jongejan JA. and Simon H. (1994) *Arch. Microbiol*. **162**, 303-309.

Johnson JL, Cohen HJ and Rajagopalan KV (1974) *J. Biol. Chem*. **249**, 5056-5061.

Kemper B, Brown DT (1976) *J.Virol* **18**, 1000-15.

Khangulov SV, Gladyshev VN, Dismukes C and Stadtman C (1998) *Biochemistry,* **37**, 3518-3528.

Kosak HG, Kemper BW (1990) *Eur. J. Biochem*., **194**, 779-84

Kutter EM, Bradley D, Schenck R, Guttman BS, Laiken R. (1981) *J Virol*. **40**, 822-9.

Luftig R, and Ganz C (1972). *J. Virol*. 10:545-554.

Mizuuchi K, Kemper B, Hays J and Weisberg RA. (1982) *Cell*. **29**, 357-65.

Møllegaard N, Murchie A and Lilley D (1994) *EMBO J*., **13**, 1508-1513

Müller U, Willnow P, Ruschig U and Hopper T (1978) *Eur. J. Biochem*.

Nagatani H and Brill WJ (1974) *Biochem. Biophys. Acta* **362**, 160-166
Nowakowski J, Shim PJ, Stout CD and Joyce GF (2000) *J. Mol. Biol.* **300**, 93-102.
Ortiz-lombardía M, Gonzáles A, Eritja R, Aymamí J, Azorín F and Coll M (1999) *Nature Struct. Biol.* **10,** 913-917.
Patterson AL (1935) *Z. Krist*., **A 90** 517.
Pöhler JRG, Giraud-Panis MJE, Lilley DMJ, *J. Mol. Biol*., **260**, 678 (1996)
Roe SM, Barlow T, Brown T, Oram M, Keeley A, Tsaneva IR, Pearl LH. (1998) *Mol Cell.* **2**:361-72.
Rhodes, G. (1993) Crystallography Made Crystal Clear. Academic Press, San Diego
Rosner BM and Schink B (1995) *Arch. Microbiol*, **177**, 5767-5772.
Rossmann MG; Blow DM (1962*) Acta Cryst*., **15**, 24-31.
Sawers G (1994) *Antonie van Leeuwenhoek, **66**, 57-88*
Streisinger G, Edgar RS and Denhardt GH *Proc. Natl. Acad. Sci. U.S.A.* 51:775-79.
Schneider F, Löwe J, Huber R, Schindelin H, Kisker C and Knäblein J (1996) *J. Mol. Biol.* **263,** 53-69.
Schomaker V and Trueblood KN (1968) *Acta Cryst*., **B24**, 63 - 76.
Schuber A, Orr E, Recny M, Schendel P, May H, Schauer N and Ferry J (1986) *J. Biol. Chem*. **261**, 12942-12947
Westheimer FH (1987) *Science* **235**:1173-8.
White MF, and Lilley DMJ (1996) *J. Mol. Biol*. **257**, 330-341
Wilkins PC and Wilkins RG (1997) *Inorganic chemistry in biology*, Oxford university press, NY.

2

X-ray structure of T4 endonuclease VII - a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture.

Hans Raaijmakers, Olivier Vix, Imre Törõ, Stefan Golz, Börries Kemper and Dietrich Suck

EMBO Journal **18** (1999)1447-1458

Abstract

Phage T4 endonuclease VII (EndoVII), the first enzyme shown to resolve Holliday junctions, recognizes a broad spectrum of DNA substrates ranging from branched DNAs to single-base mismatches. We have determined the crystal structures of the Ca^{2+} -bound wild-type and the inactive N62D mutant enzymes at 2.4Å and 2.1Å, respectively. The EndoVII monomers, which do not have a stable fold on their own, form an elongated, highly intertwined molecular dimer exhibiting extreme domain-swapping. The major dimerization element is formed by two pairs of antiparallel helices, creating a novel '4-helix cross' motif. The unique monomer fold, almost completely lacking ß-sheet structure and containing a zinc ion tetrahedrally coordinated to four cysteines, does not resemble any of the known junction-resolving enzymes, including the *E. coli* RuvC and λ integrase-type recombinases. The S-shaped dimer has two 'binding-bays' separated by about 25Å which are lined by positively charged residues and contain near their base residues known to be essential for activity. These include Asp40 and Asn62, which both function as ligands for the bound calcium ions. A pronounced bipolar charge distribution suggests that branched DNA substrates bind to the positively charged face with the scissile phosphates located near the divalent cations. A model for the complex with a four-way DNA junction is presented.

2.1 Introduction

The recognition and resolution of DNA junctions is an essential step in DNA recombination and repair, and proteins binding to and resolving these junctions have been found in all kingdoms of life (West, 1993; Kemper, 1997; White *et al*., 1997). Endonuclease VII (EndoVII), a 157 amino acid product of gene 49 of bacteriophage T4, expressed at early and late stages of infection from different promoters, is the prototypic junction-resolving enzyme, since it was the first enzyme shown to resolve Holliday junctions (Mizuuchi, 1982). EndoVII has been shown to be involved in mismatch repair, however, its major function *in vivo* - at least in the late stages of phage infection - appears to be the resolution of branch points prior to packaging of the DNA into the phage head (Solaro *et al*., 1993; Grebenshchikova *et al*., 1994; Kemper & Brown, 1976). Phages mutated in gene 49 are defective in packaging their newly synthesized DNA and accumulate highly branched DNA molecules.

In contrast to other resolvases, like e.g. yeast CCE1 or *E. coli* RuvC, EndoVII has a broad substrate specificity and recognizes a variety of branched DNA structures and/or structural perturbations in DNA. Besides Holliday junctions and cruciform DNA, EndoVII will cleave Y-junctions, heteroduplex loops, single-strand overhangs, curved DNA, but also abasic sites and single-base mismatches (Kemper, 1997; Greger & Kemper 1998). Mismatch cleavage by EndoVII has been exploited successfully for screening for mutations (Youil *et al*., 1995).

It has been suggested that the inclination of the DNA helix segments on either side of the branch point is important for recognition by EndoVII, however, the fact that single-base mismatches and abasic sites are efficiently cleaved puts a question mark behind this hypothesis (Bhattacharyya *et al*., 1991). Furthermore, comparative gel mobility experiments indicate a global structural change of four-way junctions on binding to EndoVII (Pöhler *et al*., 1996). So, clearly structural studies are required to understand the mechanism underlying the structural selectivity of EndoVII.

EndoVII is active as a dimer and nicks both strands in a divalent cation - catalyzed reaction 2-6 base pairs 3' of the branch point in independent, however temporally closely correlated reactions (Pottmeyer & Kemper, 1992). For a supercoil-stabilized cruciform substrate it was shown, that the two strands of the junction are cleaved within the lifetime of the enzyme-junction complex (Giraud-Panis & Lilley, 1997). On the other hand, the time delay between nick and counternick of mismatches is such that it allows for the repair by DNA polymerase and ligase in vitro (Solaro *et al*., 1993). This has led to the conjecture that EndoVII primarily acts a repair enzyme and represents a member of an ancient class of broad-specificity repair enzymes (Kemper, 1997).

While EndoVII like other resolvases is active as a dimer, higher order complexes with cruciform DNA containing two or even three protein dimers per junction have been observed in band-shift assays, electron micrographs and gel chromatography experiments (Golz *et al*., 1997; Kupfer *et al*., 1998; Törõ & Suck, unpublished observations). At present, it is not known whether these complexes are in vitro artifacts or whether they have any biological meaning.

The cleavage pattern and the relative cleavage efficiencies of EndoVII are influenced by the local base sequence (Pottmeyer & Kemper, 1992), however other

resolvases display a more pronounced sequence specificty. E.g. the yeast CCE1 protein shows a strong preference for cutting 3' to a CT (Schofield *et al*., 1998).

EndoVII shows little sequence homology with other resolvases and as a unique feature contains one zinc ion per monomer, coordinated to four cysteines, which is essential for stabilizing the fold (Giraud-Panis *et al*., 1995). Not clear is the relevance of significant sequence homology of the N-terminal 62 residues (containing the cysteines) with gp59 from mycobacteriophage L5 and of a 30 amino acid stretch near the C-terminus with endonuclease V from phage T4.

We present here the high resolution crystal structures of Ca^{2+}-bound wild-type and the inactive N62D mutant of phage T4 EndoVII. Implications of its unique fold and highly unusual domain-swapped dimer architecture for the catalytic mechanism and the interaction with branched DNA are discussed, and a model of a complex with a four-way DNA junction is presented.

2.2 Materials and methods

Cloning and Expression

The Wild-type EndoVII was expressed and purified as described (Golz, S. *et al*. 1995), while the N62D mutant (Golz, S. *et al.* 1997) was recloned in vector pET24d to obtain better expression. The N62D mutant was expressed in *E. coli* BL21(DE3)pLys. Overnight pre-cultures on LB-plates (30 mg/l kanamycin, 20 mg/l chloramphenicol, 37°C) were used to inoculate one 2l Erlenmeyer each, filled with 800 ml LB medium (+ antibiotics). After three hours shaking at 200 rpm, 37°C, (OD_{600}□0.8) the culture was induced with 0.25 mM IPTG. After another three hours the cells are harvested, spun down and stored at -80□C.

For the seleno-labelling, each pre-culture was used to inoculate 800 ml of M9 medium, supplemented with 50 mg/l of each of the following amino acids: Arg, His, Ile, Leu, Lys, Phe, Thr, Tyr, Val; 20 mg/ml of methionine, 2 mg/l biotin and thiamine, 30 mg/l Kanamycin and 20 mg/l Chloramphenicol. At the required density (OD_{600}□0.8) the cells were spun down, washed twice and dissolved in the same medium, but without methionine. After starvation for eight hours, 40 mg/l L-seleno-methionine was added. After another two hours the cells were induced with IPTG for three hours, spun down and stored at -80□C.

Purification of EndoVII N62D

The frozen cells were thawed and suspended in lysis buffer consisting of 10% glycerol, 10 mM HEPES pH 7.5, 10 mM β-mercapto-ethanol, 10 mM EDTA, 2 mM Pefablock or PMSF, 450 mM NaCl. The lysate was centrifuged in an ultracentrifuge at 85000 g and the supernatant was loaded onto a Q-Sepharose FF column. After washing with 2 volumes of lysis buffer, the protein was eluted by a linear NaCl gradient (200 to 800 mM) in the following buffer: 10% glycerol, 10 mM HEPES pH 7.5, 10 mM β-mercapto-ethanol, 2 mM EDTA.

The peak fractions eluting at ~600 mM NaCl were adjusted to pH 6.5 with MOPS-buffer and loaded onto a Heparin HiTrap column. A linear gradient of 0.60 to 1.3 M NaCl in 10 mM MOPS buffer pH 6.5, 10% glycerol, 10 mM β-mercapto-ethanol, 2 mM EDTA, was applied and EndoVII eluted at 1.1 M NaCl. The purest

fractions were concentrated on a Filtron membrane with a 30 kDa cut-off to 12 mg/ml and stored in the presence of 5 mM $ZnCl_2$ and 15 mM $MgCl_2$, 10% glycerol, 10 mM

MOPS buffer and 150 mM of NaCl. The protein was >99% pure (no other bands visible on silver stained SDS-PAGE). The yield was about 20 mg per litre of LB-medium for the native protein and approximately 10-fold lower for the seleno-methionine derivative. Identity and full incorporation of seleno-methionine were confirmed by mass spectroscopy (MALDI-TOF).

Crystallization and data collection

Wild-type EndoVII was crystallized in hanging drops at 4°C. 1 µl of a 16 mg/ml protein solution was mixed with 1 µl of the reservoir solution containing 200 mM $CaCl_2$, 16.5-20% PEG 2k-MME, 100 mM Tris-HCl pH 8.2, 5 mM $ZnCl_2$, 5 mM $(NH_4)_2SO_4$ and 10 mM β-mercapto-ethanol. Monoclinic crystals (space group C2, a=144.99 Å b=39.45 Å c=75.748 Å □=106.229˚) containing a dimer in the asymmetric unit appeared overnight and these were used for seeding. The crystals kept growing for about five days.

The crystals of EndoVII N62D were also grown by the hanging drop vapour diffusion method at 4°C, using 12 mg/ml protein solution as described above and 15.5-16% PEG 5K-MME in 100 mM NaOAc pH 4.5, 160-250 mM $(NH_4)_2SO_4$, 20 mM $Mg(OAc)_2$, 10 mM β-mercapto-ethanol in the reservoir.

A drop equilibrated with a slightly (2% w/v) higher PEG concentration was used to microseed a second drop, which in turn provided macroseeds for the following drops.

Crystals up to 0.8x0.4x0.3 mm have been obtained, but crystals up to 0.3 mm in their longest dimension proved most useful, due to their lower mosaicity. The mutant crystals belong to space group $P2_1$ (a=57.96 Å b=35.90 Å c=92.043 Å □=103.95˚) containing a dimer in the asymmetric unit. Both mutant and wild-type crystals show strongly anisotropic diffraction and mosaicity (varying from 0.45˚ to 1.35˚ for the native mutant dataset and even higher for the others) which prompted us to choose a lower resolution cut-off then overall I/σ values alone seem to justify (table 2.1). The seleno methionine derivative crystals diffracted to 1.6 □ at the BW7B beam line, but due to time constraints only 2.0 □ data were collected.

All Cu K_α diffraction data were collected using a MAR345 area detector system mounted on an Enraf-Nonius rotating anode generator operating at 40 kV and 90 mA and equipped with nickel coated focusing mirrors (in-house design). The seleno-methionine derivative data were collected at the EMBL BW7B beam line at the DORIS storage ring, DESY, Hamburg. All data were collected at 100K using the mother liquor + 20% PEG400 as a cryoprotectant.

The EMTS and $KAu(CN)_2$ derivatives were obtained by soaking crystals for 1 hour in their mother liquor without β-mercapto-ethanol and additionally 10% PEG400 plus 1 mM of the heavy atom compound. In case of the wild-type protein, only crystals over two months old survived this treatment. Many of the tried heavy atom compounds caused serious physical damage and/or non-isomorphism. The space-groups and cell dimensions were determined and oscillation data were processed using HKL (Otwinowski, 1997) except for the wild-type EMTS derivative data, which were processed with XDS98 (Kabsch, 1988) and Scala (CCP4, 1994).

Table 2.1 Data collection and phasing statistics for the wild-type and N62D mutant crystals.

	Wt[a]		N62D[b]			
	Native	EMTS[g]	Native	EMTS[g]	$KAu(CN)_2$	seleno-methionine
Xray source (λ)	Cu K_α	Cu K_α	Cu K_α	Cu K_α	Cu K_α	BW7B[h] (0.8373 Å)
dmin used (Å)	2.33	2.33	2.1	2.48	3.2	2.0
No. reflections.	59911	60492	67931	24788	11272	57164
Unique reflections	15953	15652	21898	13736	3287	22633
Completeness	99.3	99.5	99.9	99.0	99.3	99.8
(outer shell) (%)	(98.2)	(98.2)	(99.0)	(100)	(100)	(99.0)
$\langle I/\sigma(I)\rangle$	11.8	10.2	7.8	9.4	5.8	10.4
(outer shell)	(3.0)	(2.8)	(2.8)	(7.4)	(2.4)	(4.5)
R_{sym}[c]	6.0	6.8	6.9	7.5	17.0	6.5
No. of sites [d]		2		2	2	9
R_{iso} [e]		29.4		28.0	27.0	28.0
Phasing power [f] centric / acentric / anomalous		1.9/2.1/1.5		2.9/3.4/1.7	2.2/2.4/0.76	2.3/3.1/1.3
Figure of merit (solvent flattened)	0.42 (0.87)				0.55 (0.97)	

[a] Space group C2 a = 144.17Å b = 39.31Å c= 75.54Å, β = 106.3°, Z = 8
[b] Space group $P2_1$ a = 57.96Å b = 35.90Å c= 92.04Å, β = 103.95°, Z = 4
[c] $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$
[d] excluding the protein bound Zn atoms
[e] $R_{iso} = \Sigma|F_{PH} - F_P|/\Sigma F_P$
[f] Phasing power $= \langle |F_H(calc)| / $ phase-integrated lack of closure $\rangle$
[g] EMTS = ethyl mercury thiosalicylate
[h] at EMBL outstation Hamburg, DESY

Phasing

The EMTS anomalous difference Patterson function and the EMTS-native isomorphous difference Patterson function were calculated using data from 10 to 3□ and interpreted manually to obtain the two mercury positions, both in the wild-type and N62D case. The gold and mercury were found to bind at the same site in the N62D crystals, sharing two (out of four) cysteine ligands with the protein bound zinc atom. The wild-type data were phased using the program SHARP (Fortelle & Bricogne, 1997) by the SIRAS method with anomalous scattering of the mercury dataset.

In case of the N62D mutant, the MIRAS phases from the gold and mercury derivatives were used to locate the selenium positions in the isomorphous and anomalous difference Fouriers calculated with the seleno-methionine dataset. Nine out of ten expected selenium positions were found. The Zn sites were also very clearly visible (8 and 11 σ peaks) in the anomalous difference map. The program SHARP was then used to calculate an electron density map of excellent quality from these four datasets, also using the Zn positions. A solvent content of 52% was assumed for solvent flattening. Twofold NCS-averaging based on the heavy metal positions, was deemed unnecessary. Attempts to solve the structure of the wild-type protein by molecular replacement using the structure of the N62D mutant were not successful.

Model building and Refinement

The electron density maps were displayed using the program 'O' (Jones *et al*, 1991). The SigmaA (CCP4, 1994) weighted 2m $|F_O|$ - D $|F_C|$ map of N62D was of excellent quality (Figure 2.5), allowing to construct a complete model of EndoVII

Table 2.2 Refinement statistics of the final model

	Wild-type	N62d
Resolution range (Å)	18-2.4	27-2.1
No. of reflections used in refinement	16 359	24 998
Reflections used in Rfree calculations (%)	5	5
No. of protein atoms	2 1272	2 1272
No. of zinc atoms	2	2
No. of water molecules	63	262
No. of sulfate molecules	-	12
No. of calcium atoms	3	-
R-factor (%)	25.3	21.3
Rfree (%)	30.9	26.2
R.m.s.d. from ideal stereochemistry		
Bond lengths (Å)	0.028	0.019
Bond angles (°)	3.20	2.13
Mean B-factor (Å2)	48.9	29.3
Main chain (Å2)	48.3	25.8
Side chain (Å2)	49.8	30.1
Solvent (Å2)	42.1	39.6
Ramachandran plot		
Residues in most favored regions (%)	93.6	95.7
Residues in additionally allowed regions (%)	6.4	4.3
Residues in generously allowed regions (%)	0.0	0.0
R.m.s.d. NCS-related aoms		
C trace (Å)	0.476	0.503
All (Å)	1.15	0.976

Side chains belonging to wild-type chain A: E123, L129, E137, K150;
and wild-type chain B: K6, E122, K150, K157 are invisible at 0.8 σ.

N62D, including side chains. The selenium positions were very helpful for tracing the chain. Two clusters containing three closely spaced methionines each (M124, M125, M128) were easily recognised. Another cluster of two selenium atoms near the NCS - axis was soon identified as M68, leaving the other selenium-site to correspond to one of the N-termini. Tracing the molecule from this N-terminus to the C-terminus was straightforward. The second molecule was traced by applying the NCS-operator and was manually adjusted.

The wild-type structure was traced by manually fitting whole helices or strands, from the N62D model, as rigid bodies into the SigmaA weighted 2m $|F_O|$ - D $|F_C|$ electron density map. A second round of fitting optimised the individual amino acids. The C_α-trace was clearly visible, with only a few chain breaks at the 1 σ level. Most side chains were visible too.

The structures were refined using the CCP4 suite of programs DM (Cowtan, 1994), REFMAC (Vagin & Dodson, 1997) and ARPP (Lamzin & Wilson, 1993) and validated with PROCHECK (Laskowski *et al*., 1993) and WHATIF (Vriend, 1990). 5% of the reflections have been selected to calculate R_{free}. The starting R-factor for the N62D mutant was 0.396 (- 2.1). Individual B-factors were restrained but no NCS restraints were applied, nevertheless the monomers differ only marginally (rmsd. = 0.476 Å for the C_α atoms). Arpp was run to find water sites automatically in the

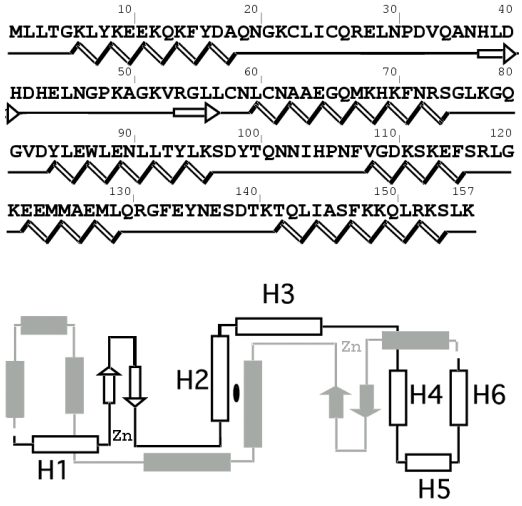

Figure 2.1 (top) Amino acid sequence and secondary structure assignment of EndoVII.
(bottom) Topology diagram of the dimer. The helices are numbered H1 to H6; the hairpin structure containing the two short ß-strands is referred to as "ß-finger" in the text.

later stages of refinement; 262 waters and 12 sulphates were included in the N62D structure.
The starting R-factor in the case of the wild-type structure was 0.395 (15-2.5 Å). Individual B-factors and NCS were restrained, but residues involved in crystal contacts that were clearly different in omit maps were excluded from NCS restraints. Three calcium ions were included and additionally 63 waters were found by Arpp. Seven side chains remained completely disordered. The coordinates have been submitted to the Brookhaven Protein Data Bank.

2.3 Results and Discussion

Experimental

The crystal structure of the N62D mutant was determined first by MIRAS techniques using a seleno-methionine labeled protein as the major source for phase determination at high resolution. The mercury and gold derivatives were used to locate 9 out of 10 selenium atoms in an anomalous difference Fourier. The resulting electron density map was of excellent quality (see figure 2.5) allowing chain tracing and model building of the complete EndoVII dimer present in the asymmetric unit. Hindering the X-ray analysis of the wild-type protein was severe non-isomorphism for many of the derivatives obtained by soaking, which often also lead to a change in space group. In contrast to the catalytically inactive N62D mutant also the expression of the seleno-methionine labeled wild-type EndoVII proved to be very difficult. Significant conformational changes in the mutant compared to the wild-type protein precluded the determination of the wild-type structure by simple molecular replacement and therefore SIRAS phasing based on a mercury derivative had to be applied. The wild-type and mutant EndoVII structures were refined to R-factors of 25.3 and 21.3% at 2.4 and 2.1Å, respectively. Refinement statistics of the final models are summarized in table 2.2.

EndoVII architecture: a domain-swapped dimer built from unstable monomers

The fold of EndoVII revealed by our X-ray analysis is novel and shows several unusual features not resembling any of the known structures of junction-resolving enzymes. The structure of the individual 157-amino acid EndoVII monomer does not represent a stable fold, as it exposes many hydrophobic residues in its central region and contains two completely separated, flexibly linked domains. They consist of a 60-residue N-terminal domain and a 50-residue C-terminal domain, which are separated by about 50Å and are connected by a central helix-loop-helix motif (residues 60-96) and an extended chain segment (residues 97-106) (Figures 2.1 and 2.2). The C-terminal domain comprising helices 4, 5 and 6 is a compact entity, while the N-terminal domain composed of helix 1, a loop region and a "β-finger" displays a less compact structure whose integrity critically depends on the presence of a zinc ion. The "ß-finger" formed by residues 38-56 is a hairpin structure containing a 4-residue stretch of antiparallel β-sheet, the only β-sheet structure found in EndoVII. The zinc ion is tetrahedrally coordinated to four cysteines located at the N-terminus of helix 2 (C58 and C61) and within the loop region (C23 and C26), which is thereby firmly tethered to the rest of the molecule (Figure 2.2). In keeping with the critical structural role of the zinc ions cysteine mutants interfering with zinc binding were found to be inactive (Giraud-Panis *et al*., 1995).

Two extensively intertwined EndoVII monomers related by a non-crystallographic two-fold axis form an elongated dimer of approximate dimensions 93Åx45Åx25Å (Figure 2.2). Within the dimer the C-terminal domain of one monomer interacts with the N-terminal domain of the other monomer resulting in an extreme case of domain-swapping. The C-terminal domain of the 'top' molecule (shown in red in figure 2.2A) is located at the 'bottom' end of the dimer and vice versa. N- and C-domains mutually stabilize each other and their orientation relative to the rest of the molecule. A lack of this interaction is presumably mainly responsible for the inactivity of N- and C-terminal deletion mutants (Birkenbihl & Kemper, 1998a; Golz *et al*., 1997; see discussion below). Only minor deviations from the local 2-fold symmetry are observed and Cα-positions of corresponding monomers can be superimposed with RMSD values of about 0.5Å.

When viewed along the molecular dyad the EndoVII dimer has an S-shaped structure with two "bays" separated by about 25Å (the distance between the bound Ca^{2+}-ions in wild-type EndoVII is 24Å; Figure 2.2A). These bays or channels, which are lined by positively charged residues on one face of the dimer, contain a number of residues shown to be essential for activity including N62, D40, H41 and E65 (Golz *et al*., 1997; Giraud-Panis & Lilley, 1996; Birkenbihl, unpublished results) as well as the bound calcium ions, clearly suggesting that they harbor the active sites (see discussion below).

In a view perpendicular to the molecular 2-fold axis EndoVII appears as a very elongated, slightly curved object with a highly bipolar charge distribution (Figure 2.2A). Predominantly positively charged residues are exposed on the concave face of the dimer, in particular on helices 2, 4 and 6.

The major dimerization element of the EndoVII dimer is formed by the central part of the monomers comprising helices 2 and 3 and the 10-residue loop connecting them. The arrangement of these helices relative to the molecular 2-fold axis gives rise to the formation of a "4-helix cross" with two pairs of antiparallel helices crossing each other at an angle of about 70 - 80^0 (Figure 2.2). This highly unusual dimerization

motif is reminiscent of the dimer interface in the recently determined structure of an *E. coli* RNA polymerase α subunit domain (Zhang & Darst, 1998). Numerous, mostly hydrophobic residues interacting with their symmetry-related counterparts form an extended hydrophobic core stabilizing the dimer (Figure 2.2B). The W87R mutant, hich directly interferes with these interhelical contacts, has lost its ability to dimerize and as a consequence neither shows DNA-binding nor cleavage activity (Birkenbihl & Kemper, 1998b).

Differences between the wild-type and N62D mutant structures

The structure of EndoVII was originally solved using crystals of the N62D mutant grown at acidic pH (4.5) in the presence of 20 mM Mg^{2+} ions (see Materials and Methods), while suitable crystals of the wild-type protein were obtained at pH 8.2 in the presence of high concentrations of Ca^{2+} (200 mM). Both magnesium and calcium ions catalyse the resolution activity of the enzyme (Pottmeyer & Kemper, unpublished results), however, under the crystallization conditions used no magnesium ions were bound at the active site of the mutant. A comparison of the two structures reveals significant conformational changes (Figure 2.3) explaining the failure of attempts to solve the wild-type structure by molecular replacement using the refined mutant coordinates. While the overall protein architecture is the same in both structures, there is a significant change in the orientation of the N- and C-terminal domains relative to the central dimerization domain. This leads to a widening of the binding clefts in the mutant and a concomitant increase in the long dimension of the mutant dimer to about 100Å.

As can be seen in figure 2.3B, it is possible to fairly closely superimpose the upper, and also the lower halves of the dimers including the N-terminal domain of one subunit and the C-terminal domain from the other subunit showing that the relative disposition of these domains is essentially conserved in the wild-type and the mutant protein. The position and orientation of helix 2 however differs and a closer inspection of the structures indicates, that this helix is bent in wild-type EndoVII around position E65/G66 towards the ß-finger and shifted roughly half a helical turn relative to its 2-fold related counterpart (Figure 2.3C). This narrows the binding cleft compared to the N62D mutant and also affects the packing of the helices without however disrupting the hydrophobic core of the dimerization interface. Responsible for these structural changes could be the Ca^{2+}-ion directly liganded to D40 and N62 in wild-type EndoVII, which neutralises the negative charges of the closely spaced carboxylate groups of D40 and E65. Another possible reason for the structural rearrangement is the difference of pH values in crystallization. The close approach of the D40 and E65 carboxylate groups (~3.0Å) in the N62D mutant suggests that at least one of them is protonated (Figure 2.3D). This may also account for the fact that no magnesium ions are bound at this site in the mutant. Finally, we cannot entirely exclude, that the different packing environments in the two crystal structures are causing conformational changes in the protein.

The structural changes observed in the mutant indicate a certain degree of conformational flexibility of the EndoVII dimer, which may be of importance for the interaction with substrates. On the other hand, a comparison of the two structures suggests that the N- and C-terminal domains are fairly rigid entities with a fixed spatial relationship.

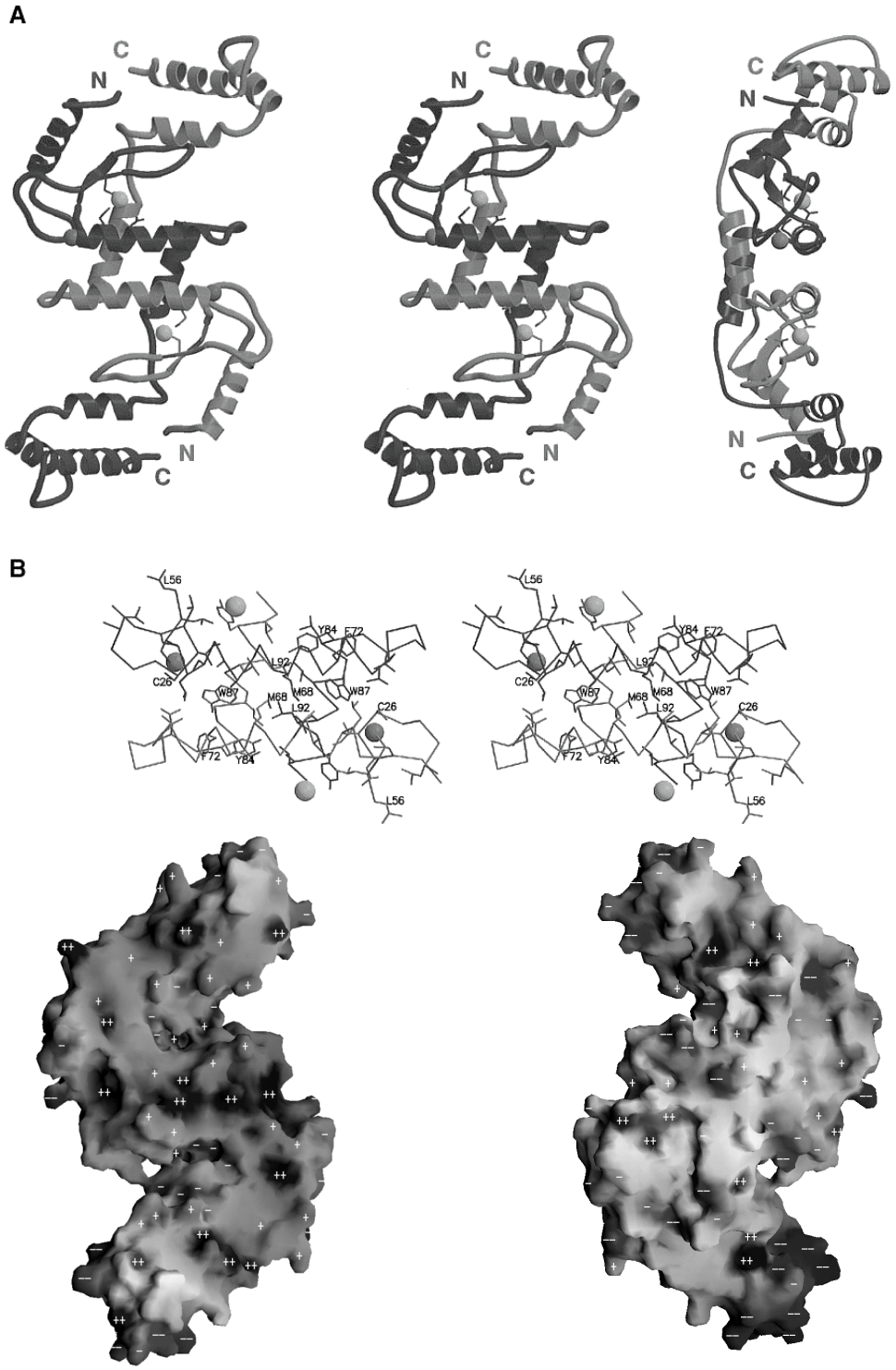

Figure 2.2 Structure of the wild-type EndoVII dimer. A) Ribbon plot representation in two perpendicular views (one of them in stereo) with the individual monomers coloured grey and black. The bound zinc and calcium ions are shown as dark and light spheres, respectively. Indicated in ball and stick representation are the side chains of D40 and N62 liganding the calcium as well as the nearby E65. B) Stereo representation of the 4-helix cross-region showing the hydrophobic residues. They form an extended hydrophobic core representing the major dimerization element of the EndoVII dimer. C) Electrostatic surface representation of the EndoVII dimer. The two opposite faces of the dimer, viewed approximately along the dyad, show a distinctly different charge distribution. The predominantly positively charged face shown on the left also contains the bound calcium ion and residues known to be essential for activity. The figure was produced with GRASP (Nicolls *et al*., 1991) with the scale ranging from –10 (--) to +17 (++), plus and minus signs indicate charge density, not individual charges.

Mutants affecting activity and the location of the active site

Extensive site-directed mutagenesis experiments have identified a number of residues affecting dimerization, DNA-binding or cleavage (for a review see White *et al*., 1997; Kemper *et al*. unpublished results). The effects of most of the mutants can be rationalized in terms of the 3D-structures and together with the Ca^{2+}-binding site found in the wild-type protein allow the identification of the putative active site.

Residues D40, H41, N62, and E65 located in the cleft between the β-finger and helix 2 are essential for activity, but not for DNA-binding (Giraud-Panis & Lilley, 1996; Golz *et al*., 1997; Birkenbihl, unpublished results). It is therefore highly likely, that the Ca^{2+}-ion directly coordinated to D40 and N62 and located close to E65 marks the position of the catalytic cation (Figure 2.2). A superposition of the putative active site regions in the wild-type and mutant proteins as shown in figure 2.3D does not provide a straightforward explanation for the inactivity of the N62D mutant, since no metal ion is bound in the mutant structure and D40 and/or E65 appear to be protonated. In principle, D62 could also function as a ligand for the catalytic cation; however, the charges in the active site cleft may not be properly balanced in the mutant. Another acidic residue, E86, located on the opposite face of the dimer and pointing into the solvent, has been reported to be catalytically essential (Pöhler *et al*., 1996). A reason for the apparent inactivity of an E86A mutant is not obvious from the 3D-structure (Figure 2.2), unless one assumes that the mutation somehow interferes with the proper folding of the protein.

Other critical residues including W87 and F72 are part of the hydrophobic 4-helix cross dimerization interface (Figure 2.2B). The W87R and F72S mutants will clearly disrupt this interface and as one would expect neither show dimer formation nor activity (Birkenbihl & Kemper, 1998b).

As mentioned above, zinc ions are essential for the integrity of the fold by tethering the loop between helix 1 and the β-finger to the N-terminus of helix 2 (Figure 2.2), and accordingly the C23S and C61S mutants which no longer bind zinc are inactive (Giraud-Panis *et al*., 1995). The inner two zinc ligands (C26 and C58) seem to be redundant to a certain extent since they still retain some zinc-binding ability, if one or the other cysteine is mutated into a serine.

Deletions at either the C-terminus or N-terminus severely affect DNA-binding and activity and mutants lacking residues 1-10 or 151-157 are inactive (Golz *et al*., 1997; Birkenbihl and Kemper, 1998a). The effect of these mutations will be twofold: firstly, a disruption of the stabilizing interaction between the N-terminal domain of one monomer with the C-terminal domain of the other monomer which will also affect their relative orientation to the rest of the molecule, and secondly, a loss of positively charged residues which may be important for DNA-binding. In agreement with the 3D-structure is the observation that the individually inactive N62D and the C-terminal Δ151-157 deletion mutant can complement each other to form an active heterodimer (Golz *et al*., 1997). One half of this heterodimer will have a wild-type active site with proper interactions of intact N- and C-terminal domains.

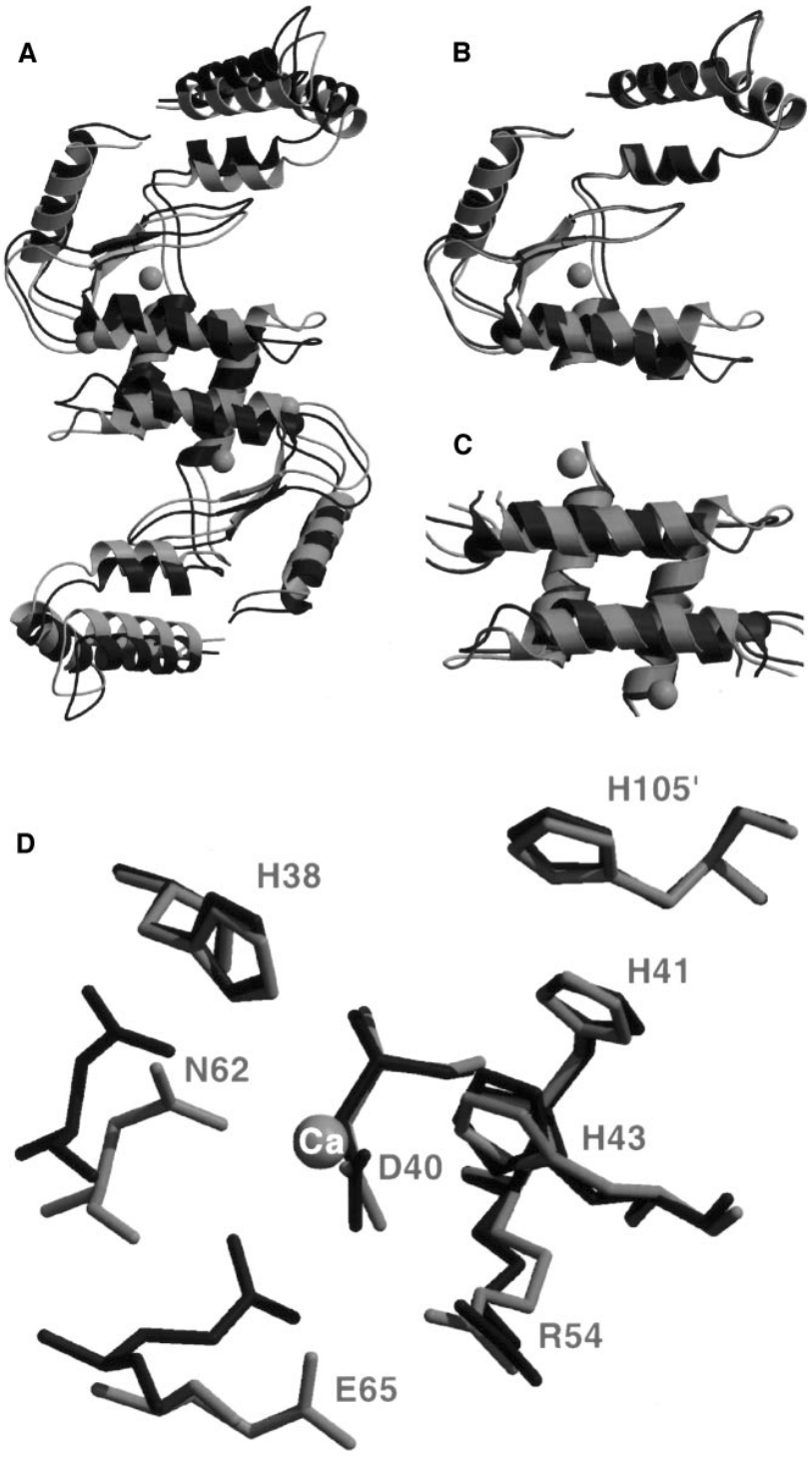

Figure 2.3 Superposition of the EndoVII wild-type and N62D mutant structures. A) The orientation of the central 4-helix cross relative to the N- and C-terminal domains has changed leading to a wider cleft between the ß-finger and helix 2 in the mutant (shown in black). B) In contrast, the conformation of the N- and C-terminal domains as well as their relative orientation remains essentially the same. C) Helix 2 in the mutant structure is shifted relative to its dyad-related counterpart by about half a helical turn compared to wild-type EndoVII. Cα-positions 1-157, 100-157, and 62-96 were used for superposition in A, B and C, respectively, corresponding to rmsd values of 2.8Å, 0.5Å and 1.0Å, respectively. D) Active site superposition of EndoVII wild-type and N62D mutant. The calcium ion bound in the wild-type protein (shown in grey) is liganded to N62 and D40. In the mutant (shown in black) E65 and D40 are in close contact (~3Å) suggesting that at least one of the carboxylates is protonated. H105' is contributed by the other monomer.

Comparison with RuvC and other junction-resolving enzymes

A number of junction-resolving enzymes from various organisms have been studied in detail by biochemical and various biophysical techniques (for reviews see e.g. White *et al*., 1997; Kemper, 1997), but structural data are available for only very few of these enzymes including the *E. coli* RuvC protein and some λ-integrase type recombinases which form and resolve a Holliday junction intermediate during site-specific recombination (Ariyoshi *et al*., 1994; Guo *et al*., 1997; Hickman *et al*., 1997; Subramanya *et al*., 1997; Kwon *et al*., 1997).

There is rarely any sequence homology detectable between EndoVII and other junction resolvases. As noted before by Giraud-Panis *et al*. (1995), a database search reveals significant homology only for the N-terminal section (residues 1-62) containing the zinc-binding cysteines with some proteins including gp59 from mycobacteriophage L5, and for a 30-residue section near the C-terminus (residues 115-145) with T4 endonuclease V, a UV-dimer excision repair enzyme. The significance of these limited sequence homologies is not clear, particularly since in the case of endonuclease V whose 3D structure is known (Vassylyev *et al*., 1995), no structural homology is present in this region. Likewise, the claimed moderate homology of a central 45 amino acid section with T7 endonuclease I, a functionally somewhat related resolvase that however cleaves 5' to the junction in the continuous strands, may not be meaningful.

E. coli RuvC resolves Holliday junctions, the central intermediates of genetic recombination, by cleaving the continuous strands to generate the recombinant DNA molecules (Iwasaki *et al*., 1991; Bennet & West, 1995). Like EndoVII, it requires divalent cations and is active as a dimer, but otherwise the two proteins show very little homology, both in terms of sequence and structure. As revealed by the X-ray structure determination by Ariyoshi *et al*. (1994) the slightly larger RuvC protein (172 residues) has a fold very different from that of EndoVII on the monomer level. It is a typical αβ-protein with a central 5-stranded β-sheet surrounded by helices displaying a topology closely related to that of *E. coli* RNaseH1 (Katayanagi *et al*., 1990). In contrast, EndoVII almost completely lacks ß-sheet structure and does not contain a stable fold on the monomer level.

The RuvC dimer with approximate dimensions 65Åx40Åx35Å is far less elongated than the EndoVII dimer and has a far more conventional dimerization interface consisting of a pair of parallel helices without intertwining of the monomers. Similar, however, is the relative disposition of the catalytic centres which are separated by about 30Å, and the bipolar charge distribution, clearly suggesting binding of the Holliday junction to the predominantly positively charged face of the dimer (see discussion below). At the active site clefts four acidic residues essential for activity are found, three of those in a disposition closely related to that of the putative active site residues D40, E65 and N62 in EndoVII and the D10, E48 and D70 carboxylates in RNaseH1. They presumably represent ligands of the catalytic metal ion and may suggest related catalytic mechanisms for these enzymes. Similar arrangements of acidic residues coordinating the catalytic cation have been found also in other Mg^{2+}-dependent nucleases.

Three-dimensional structures are available for several λ integrase-type recombinases including Cre, XerD, HP1, Int and the type I topoisomerase from vaccinia, which all are known to resolve Holliday junctions (Guo *et al*., 1997;

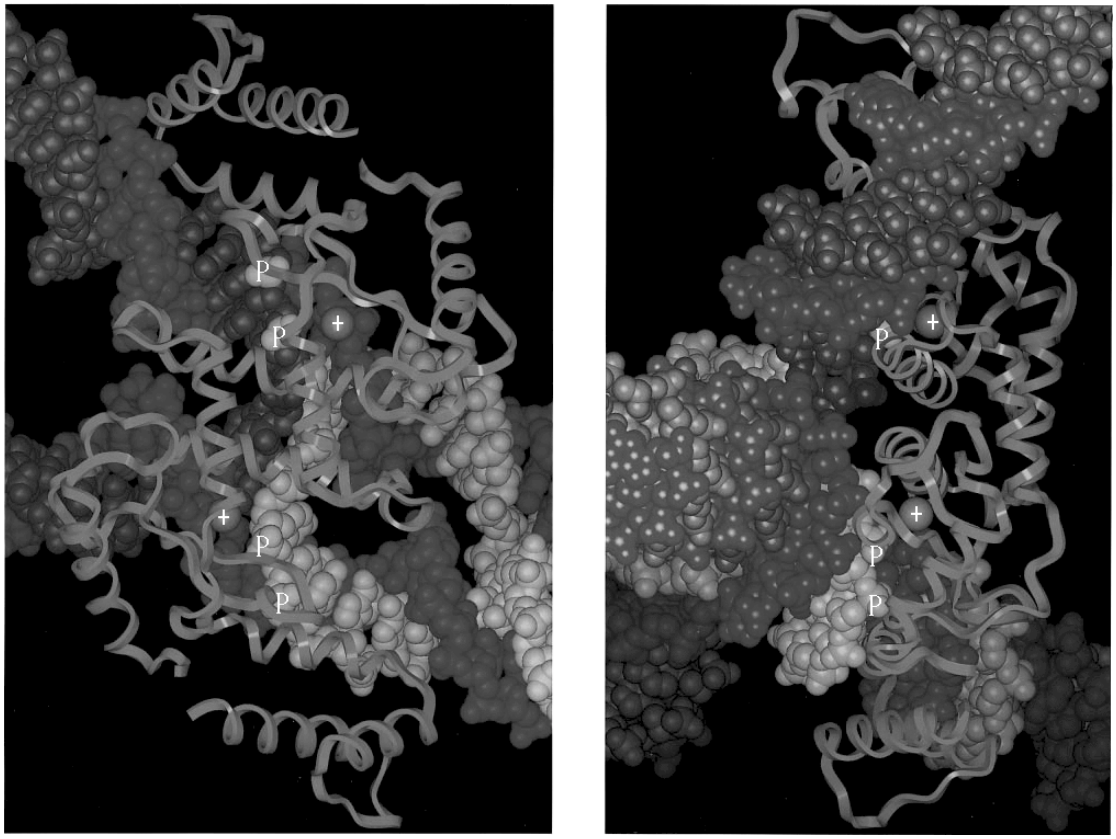

Figure 2.4 Model of a four-way DNA junction - EndoVII complex. Two perpendicular views are shown, one approximately along the molecular 2-fold axis of the EndoVII dimer (left), the other one rotated 90º around a vertical axis (right). The EndoVII dimer is docked onto the minor groove side of a four-way DNA junction, shown in a space-filled representation, corresponding to the anti-parallel stacked X-structure (von Kitzing *et al*., 1990). The continuous- and crossing- strands are coloured in different shades of grey. The scissile phosphates, 2 and 3 bases 3' of the point of strand exchange on the crossing strands are annotated with a P, the calcium ions at the active sites of EndoVII are marked with a +. Minor groove contacts with residues exposed on the central helices as well as major groove contacts with basic residues located in the C-terminal domains can be formed.

Hickman *et al*., 1997; Subramanya *et al*., 1997; Kwon *et al*., 1997). However, no significant structural similarity to EndoVII is detectable with any of these enzymes.

Model of an EndoVII - four-way DNA junction complex

EndoVII was the first enzyme shown to resolve Holliday junctions in vitro (Mizuuchi *et al*., 1982) and later found to accept as substrates a range of branched DNAs with different structural perturbations. A wealth of data derived from various biophysical and biochemical techniques has since been accumulated concerning the interaction of EndoVII with four-way junctions and other branched DNA molecules (for recent reviews see Kemper, 1997; White *et al*., 1997).

In brief, the following picture has emerged for the binding of EndoVII to a four-way junction, based mainly on comparative gel electrophoresis, protection against hydroxyl radical attack and the analysis of cleavage sites in synthetic cruciform structures with tethered arms. 1.) EndoVII binds as a dimer to the minor groove side of a four-way junction, cleaving the exchanging strands in an antiparallel stacked X-structure 3' to the point of strand exchange. 2.) The cleavage reactions of the two strands are independent, but temporally closely correlated and do occur within the lifetime of the junction-protein complex. 3.) Experiments with a protein A – EndoVII

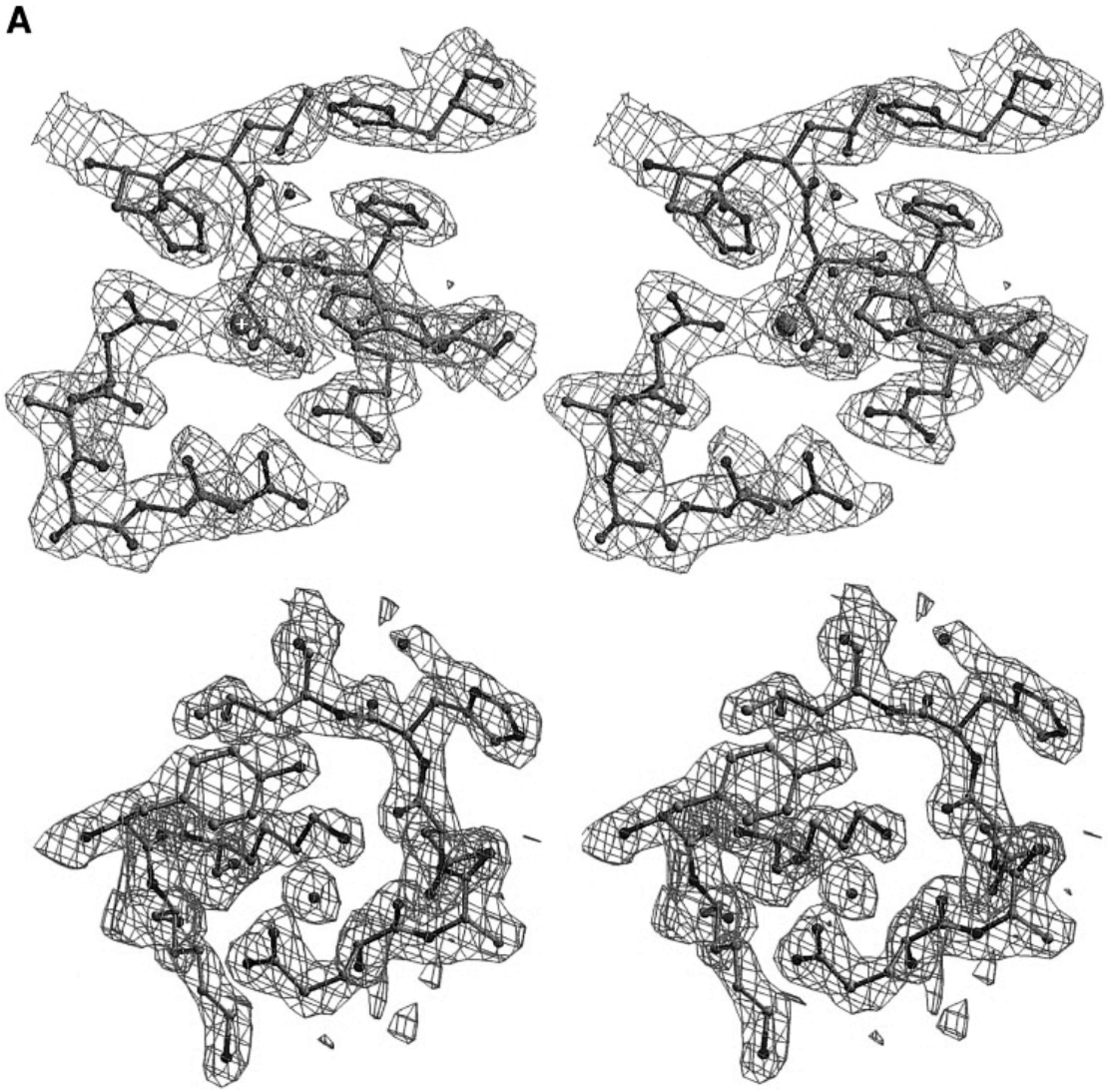

Figure 2.5 Portions of the final 2mFo-DFc map of the wild-type protein around the active site calcium ion, indicated by a larger sphere(A) and the experimental map of the N62D mutant near Lys12 (B). The contour level is 1.3σ. Note the excellent quality of the experimental map of the N62D mutant.

(E86A mutant) fusion construct seem to indicate a change in the global structure of the junction on binding. Other than in the free state, the presence or absence of Mg^{2+}-ions does not influence its structure in the bound state.

The pronounced bipolar charge distribution of the EndoVII dimer (Figure 2.2C) suggests that it binds with its predominantly positively charged face to the minor groove side of the four-way junction, such that the scissile phosphates reach into the "bays" harbouring the catalytic residues and the divalent metal ion, i.e. the Ca^{2+}-ion seen in the wild-type enzyme.

Given the 2-fold symmetry of the antiparallel stacked X-structure of the junction (in the presence of Mg^{2+}-ions) and the 2-fold symmetric distribution of the preferred cutting sites relative to the point of strand exchange, we are assuming that in the complex the molecular dyad of the junction roughly coincides with the molecular two-fold of the EndoVII dimer.

For docking we used the coordinates of the antiparallel stacked X-structure derived by von Kitzing *et al*. (1990) by computer modeling and extended the arms by

regular B-DNA. The RuvA and Cre Holliday junction complexes whose structures have been reported recently (Roe *et al*., 1998; Gopaul *et al*., 1998) cannot be used as models, since they exhibit exact or near square-planar conformations with unstacked bases at the cross-over point which are not compatible with the EndoVII structure and biochemical results. The resulting EndoVII dimer - DNA-junction complex is shown in figure 2.4. This model has not been refined and necessarily cannot provide a detailed picture of the interactions, particularly since both the DNA-junction and the protein may undergo conformational changes upon binding. Nevertheless, the overall features of the model are in agreement with the available experimental data.

Basic residues exposed on the central helices (residues H70, K71, R74 on helix 2) could interact with the phosphates of the exchanging strands and further contacts with the continuous strands are possible involving basic residues located on the helices in the C-terminal domain, in agreement with observed protection patterns (Parson *et al*., 1990). To avoid minor steric clashes with the C-terminal domains, the angle between the arms of the X-structure junction model derived by von Kitzing *et al*. may have to be slightly adjusted. In contrast to RuvC and CCE1, where an open structure with unstacked bases at the cross-over point is indicated by hypersensitivity towards permanganate and hydroxyl radicals upon complex formation (Bennet & West, 1995b; White & Lilley 1997), the footprinting patterns observed with EndoVII suggest a fully stacked conformation consistent with an antiparallel stacked X-structure of the junction.

The two putative active sites in the EndoVII dimer are about 24Å apart as measured by the distance between the two Ca^{2+}-ions bound in the wild-type structure. This has to be compared to a distance of ~23Å to ~35Å between the scissile phosphates (Figure 2.4) of the preferred cutting sites two or three base pairs on either side of the cross-over point. While there is a good match for cuts 2 base pairs from the cross-over point, a difference of more than 10Å is encountered in case of cutting 3 base pairs from the junction centre. The model suggests that even in the optimal case, due to steric constraints imposed mainly by the central helices (2-fold related, antiparallel helices 2), it may not be possible to simultaneously position both phosphates correctly with respect to the catalytic residues. This could possibly be achieved by a reorientation of the two arms of the junction contacting the protein, or by a corresponding conformational change in the EndoVII dimer. Alternatively, the junction would have to be reoriented somewhat after the first cut - possibly by a rocking motion - to bring the second phosphate into a productive binding position.

The latter scenario is attractive, since it would explain the asynchrony of bilateral cleavages and the time delay between nick and counter-nick allowing EndoVII to trigger mismatch repair (Solaro *et al*., 1993). It is also fully consistent with the notion, that the two cuts do occur within the lifetime of the protein-DNA complex, as has been shown for a supercoil-stabilized cruciform substrate (Giraud-Panis & Lilley, 1997).

The cleavage pattern and relative cleavage efficiencies of EndoVII are influenced by the local base sequence (Pottmeyer & Kemper, 1992), although other resolvases, like e.g. CCE1, display a more pronounced sequence selectivity (Schofield *et al*., 1998). This sequence selectivity could be caused by indirect effects of the base sequence on structural parameters of the DNA and/or by direct contacts to the bases likely to occur according to our model in the minor groove around the cross-over point, as well as in the major groove some 8 base pairs away (Figure 2.4).

What is the basis for the structural selectivity of EndoVII?

A remarkable feature of EndoVII is its broad substrate specificity already mentioned above, which distinguishes it clearly from cellular resolvases like RuvC or CCE1. Albeit with differing efficiencies, EndoVII cuts many different kinds of branched DNA species including Holliday junctions, but also DNA containing a single-base mismatch or an abasic site. What are the common features of these structural perturbations recognized by the enzyme?

It has been suggested, that EndoVII recognizes the inclination of DNA helices, i.e. the angle between the DNA segments on either side of a branch point or a structural perturbation (Bhattacharya *et al*., 1991). This angle is expected to be approximately 120^0 in the stacked X-conformation of a four-way junction, but also in a 3-way junction or bulged DNA containing two additional adenines. Bent DNA, either intrinsically through the presence of A-tracts, or by interaction with a compound, like in a cis-Pt adduct, also appears to be a substrate for EndoVII (Bhattacharya *et al*., 1991; Murchie & Lilley, 1993). Greger and Kemper have recently reported that an oligonucleotide with an abasic site, which was demonstrated to be extremely sensitive to cleavage by EndoVII, shows an unusual migration behaviour an polyacrylamid gels typical for bent or kinked DNA (Greger & Kemper, 1998).

Inspection of the EndoVII binding surface and the proposed model for the four-way junction complex (Figure 2.4) suggests that the X-ray structure is in general agreement with this hypothesis. However, it does not easily explain the recognition and cleavage of single-base mismatches, which - as verified in several X-ray structures - do not display any significant bending. These experimental results seem to indicate that flexibility may be an important parameter in the recognition process, in line with the observation, that the stability of a given mismatch (as measured by the melting temperature of the DNA) and cleavage efficiency are reciprocally correlated (Solaro *et al*., 1993). Possibly the lack of flexibility explains, why UV-cross-linked thymine dimers containing DNA is not a substrate for EndoVII (Kemper, unpublished results).

EndoVII appears to accept as substrates branched DNA with intrinsically inclined helices or DNA, which can easily adopt properly inclined configurations. This feature is somewhat reminiscent of the sequence-dependent cleavage by DNase I, which recognizes minor groove width and flexibility of the DNA and preferentially cleaves where either the DNA intrinsically has the proper groove geometry or where it needs the least energy to distort it for productive binding (Suck, 1994, 1998).

Acknowledgements

We would like to thank our colleagues Anja Christoph for providing the original clone for the N62D mutant, Hiang Dreher and Gunter Stier for help and advice with the cloning, expression and purification of the proteins, and Yorgo Modis and Joachim Meyer for advice in various computational aspects. We thank Ned Seeman for providing the model coordinates of the four-way DNA junction. We gratefully acknowledge the help of the staff of the EMBL outstation in Hamburg, in particular of Paul Tucker, with data collection at the BW7B beam line. This work was supported by the European Commission through an institutional fellowship to O.V. (contract no. ERBCHBGCT940535). Work in the laboratory of B. Kemper was

supported in part by the Deutsche Forschungsgemeinschaft (DFG Normalverfahren Ke188/12-3) and the Fonds der Chemischen Industrie.

References

Ariyoshi, M., Vassylyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H. and Morikawa, K. (1994) Atomic structure of the RuvC resolvase: a Holliday junction specific endonuclease from *E. coli*. *Cell* **78**, 1063 - 1072.

Bennett, R.J. & West, S.C. (1995a) RuvC protein resolves Holliday junctions via cleavage of the continuous (noncrossover) strands. *Proc. Natl. Acad. Sci. USA* **92**, 5635-5639.

Bennett, R.J. & West, S.C. (1995b) Structural analysis of the RuvC-Holliday junction reveals an unfolded junction. *J. Mol. Biol.* **252**, 213 - 226.

Bhattacharyya, A.A., Murchie, A.I.H., v. Kitzing, E., Diekmann, S., Kemper, B. & Lilley, D.M.J. (1991) Model for the interaction of DNA junctions and resolving enzymes. *J. Mol. Biol.* **221**, 1191 - 1207.

Birkenbihl, R.P. & Kemper, B. (1998a) Endonuclease VII has two DNA-binding sites each composed from one N- and one C-terminus provided from different subunits of the protein dimer. *EMBO J.* **17**, 4527 - 4534.

Birkenbihl, R.P. & Kemper, B. (1998b) Localization and characterization of the dimerization domain of Holliday-structure resolving endonuclease VII. J. Mol. Biol. **280**, 73 - 83.

Collaborative Computational Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760 -763.

Cowtan, K. (1994) Joint CCP4 and ESF-EACBM *Newsletter on Protein Crystallography*, **31**, 34 - 38.

Giraud-Panis, M.-J.E, Duckett, D.R. & Lilley, D.M.J. (1995) The modular character of a DNA junction resolving enzyme: a zinc binding motif in T4 endonuclease VII. *J. Mol.Biol.* **252**, 596 - 610.

Giraud-Panis, M.-J.E. & Lilley, D.M.J. (1996) T4 endonuclease VII: importance of a histidine-aspartate cluster within the zinc-binding domain. *J. Biol. Chem*. **271**, 33148 - 33155.

Giraud-Panis, M.-J.E. & Lilley, D.M.J. (1997) Near-simultaneous DNA cleavage by the subunits of the junction-resolving enzyme T4 endonuclease VII. *EMBO J.* **16**, 2528 - 2534.

Golz, S., Birkenbihl, R.P. & Kemper, B. (1995) Improved large-scale preparation of phage T4 endonuclease VII overexpressed in *E. coli*. *DNA Res.* **2**, 277 -284.

Golz, S., Christoph, A., Birkenkamp-Demtršder & Kemper, B. (1997) Identification of amino acids of endonuclease VII essential for binding and cleavage of cruciform DNA. *Eur. J. Biochem.* **245**, 573 - 580.

Gopaul, D.N., Guo, F. & Van Duyne, G.D. (1998) Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *EMBO J.* **17**, 4175 - 4187.

Guo, F., Gopaul, D.N. & Van Duyne, G.D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40 -46.

Grebenshchikova, S.M.L., Plugina, L.A. & Shcherbakov, V.P. (1994) The role of T4-bacteriophage endonuclease VII in correction of mismatched regions. *Genetika* **30**, 622 - 626.

Greger, B. & Kemper, B. (1998) An apyrimidinic site kinks DNA and triggers incision by endonuclease VII of phage T4. *Nucl. Acids Res*. **26**, 4432 - 4438.

Hickman, A.B., Waninger, S., Scocca, J.J. & Dyda, F. (1997) Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7□ resolution. *Cell* **89**, 227 - 237.

Hooft, R.W., Vriend,G.,Abola, E.E., (1996) Errors in protein structures. *Nature* **381**, 272-272

Iwasaki, H., Takahagi, M., Nakata, A. & Shinagawa, H. (1991) Escherichia coli RuvC protein is an endonuclease that resolves the Holliday structure. *EMBO J.* **10**, 4381 - 4389.

Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991) Improved methods for bulding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr*. **A47**, 110 - 119.

Kabsch, W. (1988) Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr*. **21**, 916 - 924.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, M. & Morikawa, K. (1990) Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* **347**, 306 - 309.

Kemper, B. (1997) Branched DNA resolving enzymes. In *DNA damage and Repair, Vol 1: DNA repair in Prokaryotes and Lower Eukaryotes*, Nickoloff, J.A. & Hoekstra, M.F., Eds., Humana Press Inc., Totowa, NJ, pp. 179 -204.

Kemper, B. & Brown, D.T. (1976) Function of gene 49 of bacteriophage T4. II. Analysis of intracellular development and the structure of very fast-sedimenting DNA. *J. Virol*. **18**, 1000 - 1015.

Kemper, B., Pottmeyer, S., Solaro, P. & Kosak, H. (1990) Resolution of DNA secondary structures by endonuclease VII (EndoVII) from phage T4. In *Structure and Methods, vol. 1. Humane Genome Initiative and DNA recombination*, Sarma, R.H. & Sarma, M.H., Eds., Adenine Press, Schenectady, NY, pp. 215 - 229.

Kosak, H.G. and Kemper, B. (1990) Large scale preparation of T4 endonuclease VII from over expressing bacteria. *Eur. J. Biochem* **194**, 779-784.

Kraulis, P.J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Cryst*. **24**, 946 - 950.

Kupfer, C., Lee, S. & Kemper, B. (1998) Binding of endonuclease VII to cruciform DNA. *J. Biol. Chem*. **273**, 31637 - 31639.

Kwon,H.J., Tirumalai, R., Landy, A. & Ellenberger, T. (1997) Flexibility in DNA recombination: structure of the lamda integrase catalytic core. Science 276, 126 - 131.

La Fortelle, E. de & Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Meth. Enzymolog*. **276**, 472 - 494.

Lamzin, V.S. & Wilson, K.S. (1993) Automated refinemant of protein models. *Acta Crystallogr*. **D49**, 129 - 147.

Laskowski, R.A., MacArthur, M.W., Moss, D. & Thornton, J.M. (1993) PROCHECK- a program to check the stereochemical quality of protein structures. *J. Appl. Cryst*. **26**, 283 - 293.

Lilley, D.M.J. & Clegg, R.M. (1993) The structure of branched DNA species. *Quat. Rev. Biophys*. **26**, 131 - 175.

Mizuuchi, K., Kemper, B., Hays, J. & Weisberg, R.A. (1982) T4 endonuclease VII cleaves Holliday structures. *Cell* **29**, 357 - 365.

Murchie, A.I.H. & Lilley, D.M.J. (1993) T4 endonuclease VII cleaves DNA containing a cisplatin adduct. *J. Mol. Biol*. **233**, 77 - 85.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr*. **D53**, 240 - 255.

Nicholls, A., Sharp, K.A. & Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *Struct. Function Gen*. **11**, 281 - 296.

Otwinowski, Z. & Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307 - 326.

Parson, C.A., Kemper, B. & West, S.C. (1990) Interaction of a four-way junction in DNA with T4 endonuclease VII. *J. Biol. Chem*. **265**, 9285 - 9289.

Pšhler, J.R.G., Giraud-Panis, M.-J. & Lilley, D.M.J. (1996) T4 endonuclease VII selects and alters the structure of the four-way DNA junction; binding of a resolution-defective mutant enzyme. *J.Mol.Biol*. **260**, 678 - 696.

Pottmeyer, S. & Kemper, B. (1992) T4 endonuclease VII resolves cruciform DNA with nick and counter-nick and its activity is directed by local nucleotide sequence. *J. Mol. Biol*. **223**, 607 - 615.

Roe, S.M., Barlow, T., Brown, T., Oram, M., Keeley, A., Tsaneva, I.R. & Pearl, L. (1998) Crystal structure of an octameric RuvA-Holliday junction complex. Mol. *Cell* **2**, 361 - 372.

Schofield, M.J., Lilley, D.M.J. & White, M.F. (1998) Dissection of the sequence specificity of the Holliday junction endonuclease CCE1. *Biochemistry* **37**,7733-7740.

Solaro, P.C., Birkenkamp, K., Pfeiffer, P. & Kemper, B. (1993) Endonuclease VII of phage T4 triggers mismatch correction in vitro. *J. Mol. Biol*. **230**, 868 - 877.

Subramania, H.S., Arciszewska, L.K., Baker, R.A., Bird, L.E., Sherratt, D.J. & Wigley, D.B. (1997) Crystal structure of the site-specific recombinase, XerD. *EMBO J*. **16**, 5151 - 5161.

Suck, D. (1994) DNA recognition by DNase I. *J.Mol.Recogntion* **7**, 65 - 70.

Suck, D. (1998) DNA recognition by structure-selective nucleases. *Biopolymers (Nucleic Acid Sciences)* **44**, 405 - 421.

Vassylyew, D.G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E. & Morikawa, K. (1995) Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* **83**, 773 - 782.

Von Kitzing, E., Lilley, D.M.J. & Diekmann, S. (1990) The stereochemistry of a four-way DNA junction: a theoretical study. *Nucl. Acids Res*. **18**, 2671 - 2683.

Vriend, G. (1990) What IF: a molecular modeling and drug design program. *J. Appl. Crystallogr*. **28**, 347 - 351.

West, S.C. (1993) The nucleases of genetic recombination. In *Nucleases*, 2nd Edition Linn, S.M., Lloyd, R.S. & Roberts, R.J., Eds., Cold Spring Harbor Laboratory Press, New York, pp. 145 - 170.

White, M.F., Giraud-Panis, M.-J., Pšhler, J.R.G. & Lilley, D.M.J. (1997) Recognition and manipulation of branched DNA structure by junction-resolving enzymes. *J. Mol. Biol.* **269**, 647 - 664.

White, M.F. & Lilley, D.M.J. (1997) The resolving enzyme CCE1 of yeast opens the structure of the four -way DNA junction. *J. Mol. biol*. **266**, 122 - 134.

Youil, R., Kemper, B. & Cotton, R.G.H. (1995) Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc. Natl. Acad. Sci. USA* **92**, 87 - 91.

Zhang,G. & Darst,S.A.(1998) Structure of the *E. coli* RNA polymerase a subunit amino- terminal domain. *Science* **281**, 262 - 266.

3

Conformational flexibility in T4 endonuclease VII revealed by crystallography: Implications for substrate binding and cleavage.

Hans Raaijmakers, Imre Törö, Rainer Birkenbihl,
Börries Kemper and Dietrich Suck

Journal of Molecular Biology (accepted)

Abstract

The structure of the N62D mutant of the junction resolving endonuclease VII from phage T4 has been refined at 1.3Å, and a second wild-type crystal form solved and refined at 2.8Å resolution. Comparison of the mutant with the wild-type protein structure in two different crystal environments reveals considerable conformational flexibility at the dimer level affecting the substrate binding cleft, the dimerization interface and the orientation of the C-terminal domains. The opening of the DNA binding cleft, the orientation of the C-terminal domains relative to the central dimerization domain as well as the relative positioning of helices in the dimerization interface appear to be sensitive to the crystal packing environment. The highly unexpected rearrangement within the extended hydrophobic interface does change the contact surface area but keeps the number of hydrophobic contacts about the same and will therefore not require significant energy input. The conformational flexibility most likely is of functional significance for the broad substrate specificity of EndoVII. Binding of sulphate ions in the mutant structure and their positions relative to the active site metal ions and residues known to be essential for catalysis allows us to propose a possible catalytic mechanism. A comparison with the active site geometries of other magnesium-dependent nucleases, among them the homing endonuclease I-PpoI and *Serratia* endonuclease, shows common features suggesting related catalytic mechanisms.

3.1 Introduction

Endonuclease VII (EndoVII) from phage T4 is a junction-resolving enzyme and was the first protein shown to resolve Holliday junctions *in vitro* (Mizuuchi, 1982). The 157 amino acid product of gene 49 of bacteriophage T4 is expressed at early and late stages of infection from different promoters and has been shown to be involved in mismatch repair, however, its major function *in vivo* - at least in the late stages of phage infection - appears to be the resolution of branch points prior to packaging of the DNA into the phage head (Solaro *et al*., 1993; Grebenshchikova *et al*., 1994; Kemper & Brown, 1976). Phages mutated in gene 49 are defective in packaging their newly synthesized DNA and accumulate highly branched DNA molecules.

A striking feature of EndoVII is its broad substrate specificity. Unlike other resolvases, including yeast mitochondrial CCE1 or *E. coli* RuvC, which specifically cleave four-way DNA junctions, EndoVII recognizes a variety of branched DNA structures or other structural perturbations in DNA. In addition to Holliday junctions and cruciform DNA, EndoVII will cleave Y-junctions, heteroduplex loops, single-strand overhangs, curved DNA, but also abasic sites and single-base mismatches (Kemper *et al*., 1990; Kemper, 1997; Greger & Kemper 1998). Mismatch cleavage by EndoVII has been exploited successfully for screening for point mutations (Youil *et al*., 1995). It has been suggested that the inclination of the DNA helix segments on either side of the branch point or structural perturbation is important for recognition by EndoVII (Bhattacharyya *et al*., 1991). The cleavage pattern and the relative cleavage efficiencies of EndoVII are influenced by the local base sequence (Pottmeyer & Kemper, 1992); however, the sequence specificity for cleavage is far less pronounced than for other resolvases, like e.g. CCE1 or RuvC (White & Lilley, 1996; Shah *et al*., 1997)

EndoVII is active as a dimer and nicks both strands in a divalent cation-catalysed reaction 2-6 base pairs 3' of the branch point in independent, however temporally closely correlated reactions (Pottmeyer & Kemper, 1992). For a super coil-stabilized cruciform substrate it was shown, that the two strands of the junction are cleaved within the lifetime of the enzyme-junction complex (Giraud-Panis & Lilley, 1997). On the other hand, the time delay between nick and counter nick of mismatches is such that it allows for the repair by DNA polymerase and ligase in vitro (Solaro *et al*., 1993). This has led to the conjecture that EndoVII primarily acts as a repair enzyme and represents a member of an ancient class of broad-specificity repair enzymes (Kemper, 1997).

We have recently determined the crystal structure of EndoVII, which revealed a novel fold not seen in any of the known junction resolvases (Raaijmakers *et al*., 1999). A striking feature of the structure is the domain-swapped dimer architecture, which stabilizes the open fold of the monomers. A novel "four-helix cross" motif with two pairs of antiparallel helices forms the major dimerization element containing an extended hydrophobic core. Surprisingly, the inactive single point mutant N62D showed a major rearrangement of this interface and also significant conformational changes affecting the DNA binding cleft. Since wild-type and mutant proteins were crystallized under different pH and buffer conditions, it was not clear, to what extent the observed changes were induced by the mutation, the crystal packing or whether they indicate an intrinsic conformational flexibility of the enzyme possibly important

for its broad-spectrum substrate specificity. We present here an analysis of these conformational changes based on the structure of the N62D mutant at 1.3Å resolution and two wild-type structures in different crystal packing environments obtained under identical crystallization conditions. In addition, we propose a catalytic mechanism based on the disposition of catalytically essential residues and bound sulphate and metal ions in the active sites of the inactive N62D mutant and the wild-type protein, respectively. A comparison with other magnesium-dependent nucleases reveals common features of the active site geometry, suggesting related mechanisms for cleavage possibly as a result of convergent evolution.

3.2 Materials and Methods

Preparation of protein

The wild-type and N62D mutant EndoVII were cloned into pET11a and pET24d respectively, expressed in *E. coli* BL21 (DE3), and purified as reported previously using basically two ion exchange columns, run at pH 7.5 and 6.5 respectively (Golz *et al.,* 1995; Raaijmakers *et al., 1999*).

Crystallization

The wild-type EndoVII was crystallized in hanging drops at 4ºC. 1 µl of 16 mg/ml protein solution was mixed with 1 µl of the reservoir solution containing 18% PEG 2K-MME in 100 mM Tris pH 8.2, 200 mM $CaCl_2$, 5 mM $(NH4)_2SO4$, 5 mM $ZnCl_2$ and 10 mM ß-mercapto-ethanol. Two crystal forms appeared, belonging to space group C2 (a = 145.0 Å b = 39.4 Å c = 75.7 Å, ß=106.2º) or $P2_1$ (a=60.92 Å b=37.38 Å c=74.02 Å, ß=103.9º), each containing a dimer in the asymmetric unit. These crystals were used for microseeding to improve both size and quality.

Crystals of EndoVII N62D were also grown by the hanging drop vapour diffusion method combined with microseeding techniques at 4^0C, using 12 mg/ml protein solution and 15.5-16% PEG 5K-MME in 100 mM NaOAc pH 4.5, 160 mM $(NH4)_2SO4$, 20 mM $Mg(OAc)_2$, 10 mM ß-mercapto-ethanol in the reservoir, and further improved through micro-seeding. The crystals belong to space group $P2_1$ (a=58.24 Å b=35.83 Å c=91.93 Å ß=103.9º) containing a dimer in the asymmetric unit.

Diffraction data collection

All wild-type EndoVII data were collected with CuKα (λ=1.54 Å) radiation at a Nonius rotating anode generator equipped with focussing mirrors and a Mar 345 detector.

High resolution data of EndoVII N62D were collected at the EMBL BW7B beam line at the DORIS storage ring, DESY, Hamburg in two passes with a maximum resolution of 1.32Å and 2.43Å respectively. All crystals were flash cooled in liquid nitrogen prior to data collection, using four parts mother liquor mixed with 1 part PEG400 as a cryoprotectant.

The diffraction data were processed using HKL (Otwinowski & Minor, 1997), and reduced to amplitudes with Truncate (CCP4, 1994). Data collection statistics are summarized in table 3.1.

Table 3.1. Data collection statistics of the EndoVII crystals.

	Wild-type $P2_1$[a]	Native N62D[b]	Wild-type C2[c]
X-ray source (λ)	Cu K_α	BW7B[e] (0.8345 Å)	Cu K_α
Outer shell (Å)	2.9-2.8	1.37-1.32	2.46-2.33
Observed reflections.	28458	243182	59911
Unique reflections	8966	83878	15953
Completeness (outer shell) (%)	99.8(99.7)	96.5 (88.9)	99.3 (98.2)
$<I/\sigma(I)>$ (outer shell)	7.2 (3.2)	9.9 (2.3)	11.8 (3.0)
R_{sym}[d] (outer shell)	0.096 (0.283)	0.075 (0.301)	0.060 (0.253)
V_M ($Å^3$/Da)	2.18	2.55	2.85
Estimated solvent content	43.5%	51.4%	56.5%

[a] Space group $P2_1$ a = 60.9Å b = 37.4Å c= 74.0Å, β = 108.2°, Z = 4

[b] Space group $P2_1$ a = 58.2Å b = 35.8Å c= 91.9Å, β = 103.9°, Z = 4

[c] Space group C2 a = 145.0Å b = 39.4Å c= 75.7Å, β = 106.2°, Z = 8

[d] Rsym $= \Sigma|I - <I>|/\Sigma I$

[e] at EMBL outstation Hamburg, DESY

[f] V_M = Matthews volume, Matthews, *J.Mol.Biol* **33**, 491-497 (1968).

Model building and Refinement

The mutant structure was refined using the CCP4 suite of programs Dm, Refmac and Arpp (CCP4, 1994; Cowtan, 1994; Murshodov *et al.,* 1999; Lamzin & Wilson, 1993) using 5% of the reflections to calculate R_{free} throughout the refinement. Starting with the previously described 2.1 Å structure and phases derived from the refined model (Raaijmakers *et al*, 1999), the resolution was increased from 2.3 to 1.32Å in five steps using Refmac, while Arpp was used to automatically add waters increasing their number from 275 to 430. The resulting electron density maps were displayed using the program 'O' (Jones *et al*, 1991), to manually adjust residues. Fourteen residues had multiple conformations. After rebuilding and refining twice, restrained individual anisotropic B-factors were refined for all atoms. This in itself initiated a 2.9% decrease of the R-factor and a 2.3% drop in R_{free}. After another round of refinement and rebuilding in the much clearer maps, hydrogen scattering contributions were calculated, but not refined, with Hgen and Sfall (CCP4, 1994) and included as partial structure factors, lowering the R-factor by another 1.2% and R_{free} by 1.1%.

The weight on X-ray terms was raised while carefully monitoring the geometry with Procheck (Laskowski *et al.*, 1993; CCP4, 1994) and the check options of Whatif (Hooft *et al.*, 1996). Then the geometry restraints on the zinc atoms (Zn-S distance and S-Zn-S angle) were released. Finally, the built-in solvent correction of Refmac (version 4.0) was replaced by a mask-based solvent correction, calculated with Ncsmask and Sfall (CCP4, 1994) and included as partial structure factors. A few protein atoms are poorly defined and are invisible at 0.5 sigma electron density ($2mF_O$-DF_C) contour including atoms C_δ, $O_{\varepsilon 1}$ and $O_{\varepsilon 2}$ from E89 chain A, and atoms C_ε and N_ζ from K114, K154, K157 in chain A and from K49, K114 and K154 in chain B.

The refinement of C2 wild-type EndoVII (C2wt) was described before (Raaijmakers *et al.,* 1999). The $P2_1$ wild-type EndoVII ($P2_1$wt) was solved by molecular replacement starting from the mutant structure using the program Amore (Navaza, 1994). The solution obtained was refined with the program Refmac. Initially the structure was divided into three rigid domains: the four-helix cross (residues 55-102 from both chains including the zinc atoms) and two NC terminal domains

Table 3.2 Refinement statistics

	Wt C2	Wt $P2_1$	N62D
Resolution range (Å)	10-2.4	27-2.8	12-1.32
No. of reflections used in refinement	15326	12364	83878
Reflections used in R_{free} calculations (%)	5	5	5
No. of protein atoms	2 x 1272	2 x 1272	2 x 1272
No. of zinc atoms	2	2	2
No. of water molecules	63	30	430
No. of calcium ions	3	2	
No. of sulphate molecules	-	-	14
R factor (%)	25.4	24.5	14.5
Free R-factor (%)	30.9	31.4	18.6
r.m.s. deviations from ideal stereochemistry:			
bond lengths (Å)	0.028	0.025	0.010
bond angles (□)	3.2	2.3	1.7
Mean B factor ($Å^2$)	48.9	38.1	30.3
main chain ($Å^2$)	48.3	37.5	24.3
side chain ($Å^2$)	49.8	38.5	31.2
solvent ($Å^2$)	42.1	28.9	39.3
Ramachandran plot[a] :			
residues in most favoured regions (%)	93.6	89.6	97.1
residues in additionally allowed regions (%)	6.4	8.9	2.9
residues in generously allowed regions (%)	0.0	0.7	0.0
rmsd NCS-related atoms			
C_α trace (Å)	0.48 [b]	1.77 [b]	0.49
all protein atoms(Å)	1.15 [b]	2.08 [b]	0.86

[a] Calculated with Procheck
[b] refined with NCS restraints.

(residues 1-55 from one chain and 103-157 from the other chain). These were refined during four rounds including data up to 5 Å, 4 Å, 3.5 Å and 3 Å. After manual adjustment using the program 'O' (Jones *et al*, 1991), the protein was refined to 2.8 Å with tight NCS restraints using one NCS operator for residues 8-103 and another for residues 1-7 and 105-157. Only clearly deviating residues were excluded from these restraints. No NCS restraints were applied for Cys23, Ile104, Ser117 and Arg118, and only loose NCS restraints for residues 27-29, 43-52 and 135-142. 30 waters and 2 calcium ions could be identified.

In an attempt to localize centers of rotational flexibility, we refined TLS parameters, i.e. translation, libration and screw-rotation tensors, (Schomaker & Trueblood, 1968) for the helices and ß-finger with an experimental version of Refmac5 (Winn *et al*, 2000). The version was only used for refinement of the TLS parameters, but not for further refinement, since it could not yet combine TLS parameters with other thermal restraints. The three EndoVII structures were divided in 2x7 TLS groups, defined as: helix 1 (residues 6-19), the ß-finger (res. 38-57), helix 2 (res. 58-74), helix 3 (res. 84-96), helix 4 (res. 107-117), helix 5 (res. 121-140) and helix 6 (res. 141-155) (see Figure 3.1). All B factors within each group were averaged before refinement. For comparison the same refinement was also done with the whole dimer as one TLS-group with all B-factors averaged. The results were analysed with TSLANL (Howlin *et al*, 1993)

Contact surfaces were calculated with the program Areaimol (Lee and Richards, 1971; CCP4, 1994) as the surface of each monomer, minus the surface of the complex, divided by two. Calculations were performed with a probe radius of 1.4 Å, but similar results were obtained using probe radii of 1.1 or 1.7 Å. Superpositions

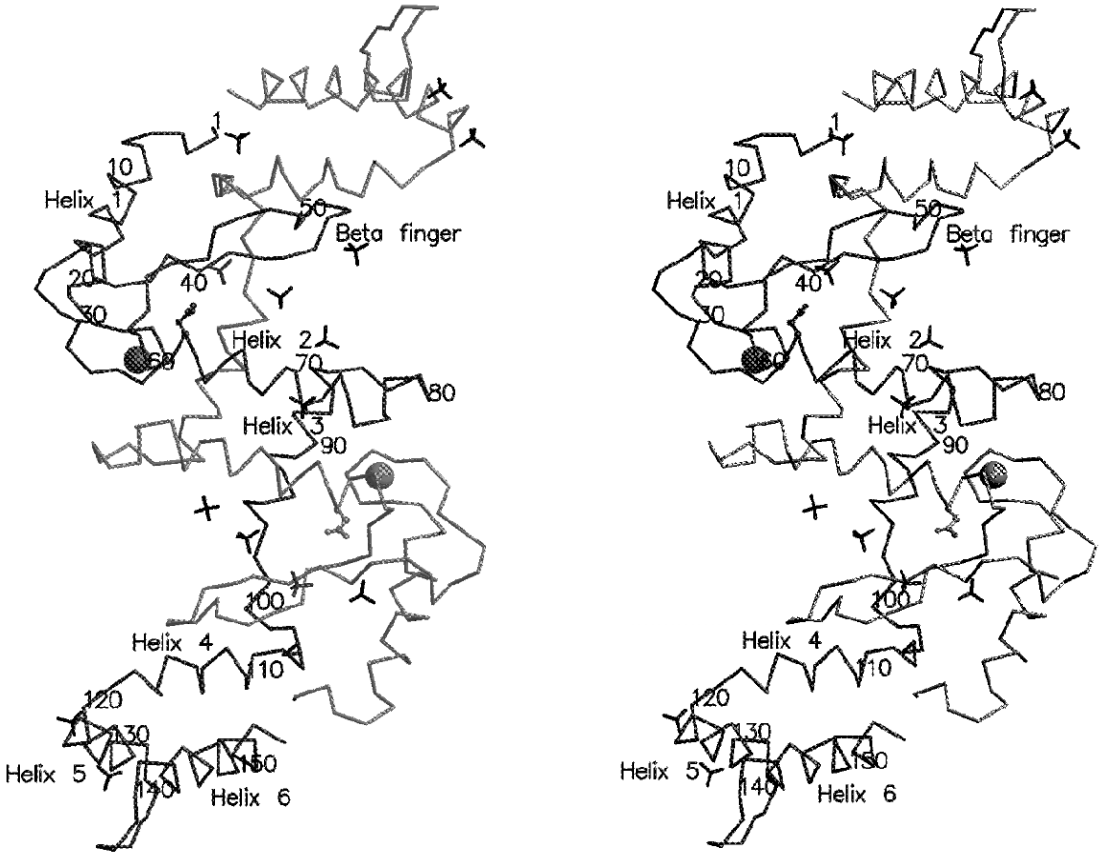

Figure 3.1 Stereo view of the Cα trace of the N62D mutant dimer. The D62 side chain is shown in ball-and-stick mode. Residues of chain A are numbered every 10 residues. The active site sulphate is near residue 40. The structural zinc ions are shown as spheres.

were calculated with Lsqkab (Kabsch, 1976). Figures were produced with the programs "O", Insight, Molscript (Kraulis, 1991) and Raster3d (Merritt & Bacon, 1997).

Protein Data Bank accession numbers

The atomic coordinates have been deposited in the Protein Data Bank as entries 1en7, 1e7d and 1e7l for the EndoVII wild-type C2, wild-type $P2_1$ and N62D mutant respectively.

3.3 Results and Discussion

Structure determination and refinement

The structure of the N62D mutant of EndoVII was refined to an R-factor of 14.5% using data to 1.32 Å resolution collected at the EMBL outstation at DESY (Hamburg). Data collection and refinement statistics are summarized in Table 3.1 and 3.2. The refined model confirms our previous results (Raaijmakers *et al*., 1999) and shows in addition the C-terminal Lys residues, 14 sulphate ions bound per dimer and a total of 429 water molecules (Figure 3.1). Fourteen residues exhibit alternative side chain conformations. The Ramachandran plot (Ramakrishnan & Ramachandran, 1965) shows no outliers, with 97% of the residues having ΦΨ-values corresponding to fully allowed regions according to the Procheck definition (CCP4, 1994), or 99%

according to (Kleywegt & Jones, 1996). A second crystal form of wild-type EndoVII belonging to space group $P2_1$ ($P2_1$wt) was obtained under identical conditions to those described previously for the C2 crystals (C2wt) (Raaijmakers *et al*. 1999). Like the mutant crystals both wild-type crystal forms contain one dimer in the asymmetric unit, however, the $P2_1$wt crystals are characterized by a significantly lower solvent content (Table 3.2). Using data to 2.8Å resolution the structure of the wild-type $P2_1$wt crystal structure was solved by molecular replacement using the coordinates of the N62D mutant. A clear solution was however only obtained after properly orienting the two N/C-terminal domains - i.e. the upper and lower half of the dimer - with respect to the central dimerization domain (see Materials & Methods).

TLS group refinement of EndoVII N62D resulted in a final R/Rfree of 18%/19.6%, not far above the 14.5%/18.6% of the full anisotropic refinement, while overall TLS refinement (one group) leads to an R/Rfree of 25.5%/26.4%. Thus anisotropic refinement of the 2x7 rigid groups (as defined in Materials & Methods) can model most of EndoVII's flexibility within the crystal.

C2wt refines with overall TLS to R/Rfree values of 24.9%/28.3%, and in the 14 groups case to 23.1%/26.3%, better than with restrained individual isotropic B-factor refinement (22.1%/27.6%). In wt$P2_1$, overall TLS-refinement lead to R/Rfree values of 25.5%/31.6%, 2x6 TLS groups to 24.6%/29.7% (refinement of helix 6 did not converge and had to be taken out of the refinement) while restrained individual B factors give 24.5/29.6%. Overall TLS refinement for C2wt, $P2_1$wt and the mutant gave a mean translation of 0.22, 0.11 and 0.03 $Å^2$ and a mean libration of 1.3, 0.6 and 0.6 deg^2 , respectively.

The superiority of the 14 TLS group over single TLS group refinement leads to the conclusion that EndoVII does not behave like a single rigid body, but rather can be modelled as flexibly connected, rigid elements.

Differences in the magnitude and directionality of the translation and libration parameters in the three structures make it hard to pinpoint possible motions in the protein. However, the maximum libration component of the groups runs always within 2-20 degrees of their long axis, for the mutant, and 5-25 degrees in case of the wild type structures, except for helix 5 (all structures) and helix 2 in C2wt. Helices 2 (21.0 and 20.5 deg^2 along their axis, in the other directions 4.1 deg^2 and 2.1 deg^2 for the first and 1.5 deg^2 and 1.0 deg^2 for the second chain) and 4 (24.6 and 14.5 deg^2 along their axis, 0.9 deg^2, 1.1 deg^2 and 0.4 deg^2 0.9 deg^2 in the other directions) have the highest libration in the mutant structure, while helices 4, 5 and 6 have the higher libration values in both wild-type structures. These latter helices have their libration axes parallel in all three structures, what indicates that they behave as a rigid body. There is no especially large translational disorder along helix 2 in any of the structures, suggesting that the four-helix-cross does not rearrange within one crystal.

In summary, we conclude from the TLS refinement that it is valid to model the EndoVII dimer as 2x7 or perhaps 2x4 (if one considers helix 4, 5 and 6 as one single rigid group) rigid bodies connected by flexible loops. Within a given crystal lattice most of the flexibility appears to originate from rotations along the helical axes or the ß-finger axis rather than from fluctuations of the angle between them.

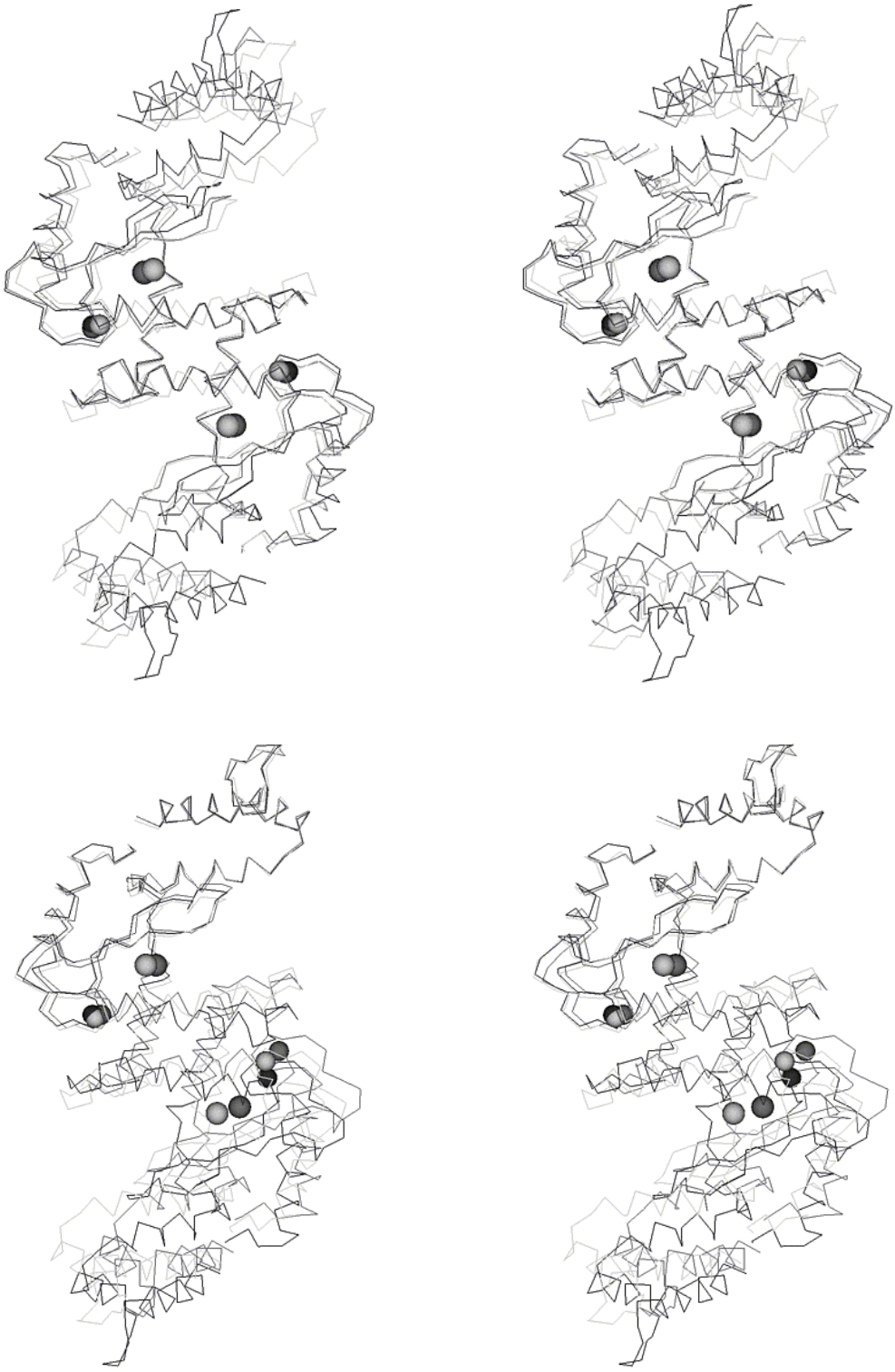

Figure 3.2 Superposition of the N62D mutant with the wild-type C2 and $P2_1$ structures. Stereo pairs of the Cα traces, with the N62D mutant depicted in black, wild-type C2 in light-grey, and wild-type $P2_1$ in medium-grey. Spheres indicate the active site Ca and Zn positions.
(A) Superposition of residues 55-103 shows that the wild-type $P2_1$ structure is very similar to the mutant structure in the helical-cross domain. (B) Superposition of residues 1-53 of chain A and 104-157 of chain B illustrates that the NC-terminal domain stays very similar in all three structures, but that the opening of the binding clefts is much larger in the mutant than in either of the wild-type structures.

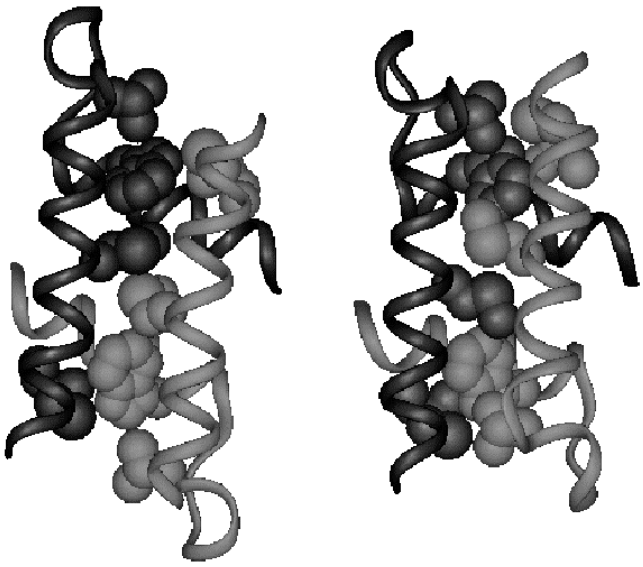

Figure 3.3 Changes in the dimerization interface. Helices 2 and 3 (residues 58-96) in the wild-type C2 (left) and mutant (right) structures are shown in ribbon representation with molecules A depicted in dark-grey, molecules B in light grey. Residues I25, C61, L77, M68 and W87 are shown in van der Waals representation. Note the change of positions of the dark and light M68 residues (center) and their W87 interaction partners in N62D (compared to wtC2) caused by a shift of almost a complete helical turn of the number 2 helices relative to each other. This shift does not interrupt the hydrophobic core and maintains the M68-W87 contacts.

Comparison of the N62D mutant with the two EndoVII wild-type structures.

The structures of the Endo VII mutant and the wild-type protein show significant conformational differences which prohibited the solving of the wild-type protein by molecular replacement using the mutant coordinates (Raaijmakers et al., 1999). Since the mutant protein had been crystallized at low pH (4.5) and in the presence of 20 mM magnesium ions, the wild-type protein however at pH 8.2 in the presence of 200 mM calcium ions (Raaijmakers et al., 1999), it was not clear to what extent the observed structural differences were due to the mutation or a consequence of the different crystallization conditions and crystal packing environments. With the structure of the wild-type protein solved from a second crystal form ($P2_1$wt) grown under identical conditions as the C2wt crystals, we are now in a position to address this question.

A comparison of the two wild-type structures reveals pronounced structural differences, obviously induced by different crystal packing environments and thereby clearly demonstrating the conformational flexibility of the Endo VII dimer. The extent of the conformational flexibility was not anticipated from the previous structures.

Changes in the dimerization interface.

Highly unexpected were in particular the quite dramatic changes in the dimerization interface involving the extended hydrophobic core formed by residues in the four-helix cross domain. The superposition of the four-helix cross region of all three structures shown in figure 3.2 reveals that this interface is very similar in wild-type $P2_1$ and the N62D mutant structures, while it is clearly different in the wild-type C2 crystal form. As can be seen in figure 3.2, the number 2 helices (see Figure 3.1 for numbering) in the mutant dimer and the wild-type $P2_1$dimer are shifted by almost a complete helical turn relative to each other compared to the wild-type C2 structure.

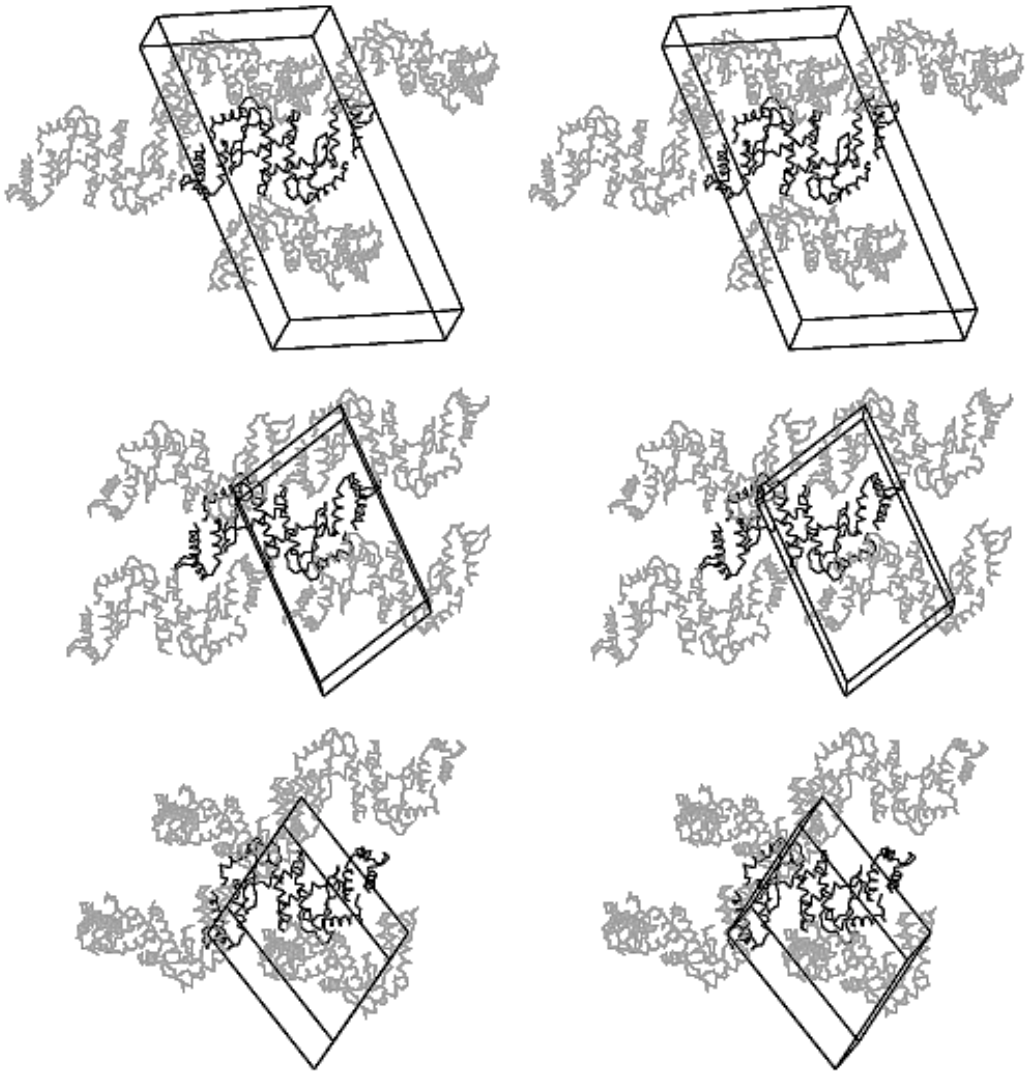

Figure 3.4. Crystal packing of respectively the wild-type C2, wild-type $P2_1$ and N62D EndoVII structures viewed in stereo, approximately along the molecular 2-fold axis of the unique dimer (dark). Noteworthy are the contacts inside the wide binding cleft of the mutant, while wild-type C2 has more contacts around the end of helices 2 and near the loop between helices 5 and 6. Wild-type $P2_1$ has many interactions around one of the NC-terminal domains.

This shift results e.g. in an exchange of the interaction partners of W87, a major contributor to the hydrophobic core and known from site-directed mutagenesis to be essential for dimerization (Birkenbihl & Kemper, 1998a). While in wild-type C2 crystals Trp87 interacts with Met68 of the same subunit, it interacts with Met68 of the other subunit in the mutant and wild-type $P2_1$ crystals, i.e. the methionines have switched positions, but still retain their interactions with the tryptophane (Figure 3.3). The packing density within the hydrophobic core appears to be somewhat tighter in the mutant.

While this shift, in what appears to be a compact hydrophobic interaction surface, is the major factor in reducing the intermolecular contact surface of the dimer from 1500 $Å^2$ in the $P2_1$ wild-type and 1430$Å^2$ in the mutant to 1286$Å^2$ in the wild-type C2 structure, it cannot be attributed to the N62D mutation and must therefore be

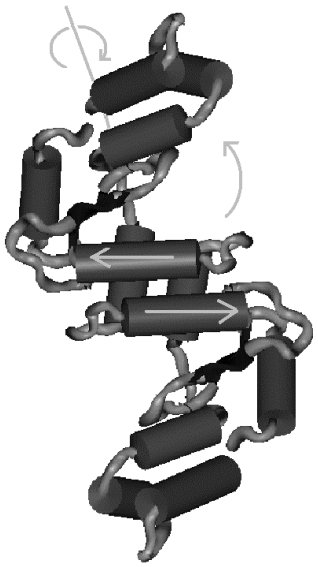

Figure 3.5 The three major elements of conformational flexibility of the EndoVII dimer. Indicated by arrows are the conformational changes revealed by the comparison of the wild-type and the mutant EndoVII structures involving the dimerization helices, the DNA binding cleft, and the orientation of the C-terminal domains with respect to the four-helix cross.

a consequence of the different packing environments. This in turn suggests that only small energy differences are involved in these conformational changes highlighting the intrinsic flexibility of the EndoVII dimer.

Different crystal contacts found at the loop region between helix 2 and 3 could account for the large differences seen in this area (Figures 3.2,3.4). In particular the presence of a Ca^{2+}-ion in the wtC2 structure, bridging two molecules and liganded to both Oδ-oxygens of the carboxylate group of E86 and the carbonyl oxygen of G81 in this loop, is probably responsible for these differences and may also trigger the shift of the dimerization helices.

Changes in the DNA binding cleft.

Another major difference between the mutant and the C2 wild-type structure is the orientation of the interacting N- and C-terminal domains relative to the four-helix cross, as noted previously (Raaijmakers *et al*., 1999). In this respect the two wild-type structures more closely resemble each other than the mutant (Figure 3.2). The larger opening of the binding cleft observed in the mutant could therefore indeed be, at least in part, a consequence of the different charge distribution at and around the active site caused by the presence of an additional charged residue, protonation of one or more of the acidic residues D40, D62 and E65, protonation of histidines 38, 41, 43, 70 and 105, and the lack of a calcium ion (which is only found in the wild-type structures; see below). However, somewhat smaller, but still significant differences in the relative orientation of these domains are also observed between the two wild-type structures (Figure 3.2). Note that the binding cleft in the EndoVII wild-type structures is still somewhat wider than the structurally similar active site cleft of *Serratia* nuclease or I-PpoI (see below; Figure 3.7).

A comparison of the crystal packing schemes in the various crystal forms reveals that the wider opening of the substrate binding cleft in the mutant is accompanied and probably partly induced by tighter contacts of neighbouring

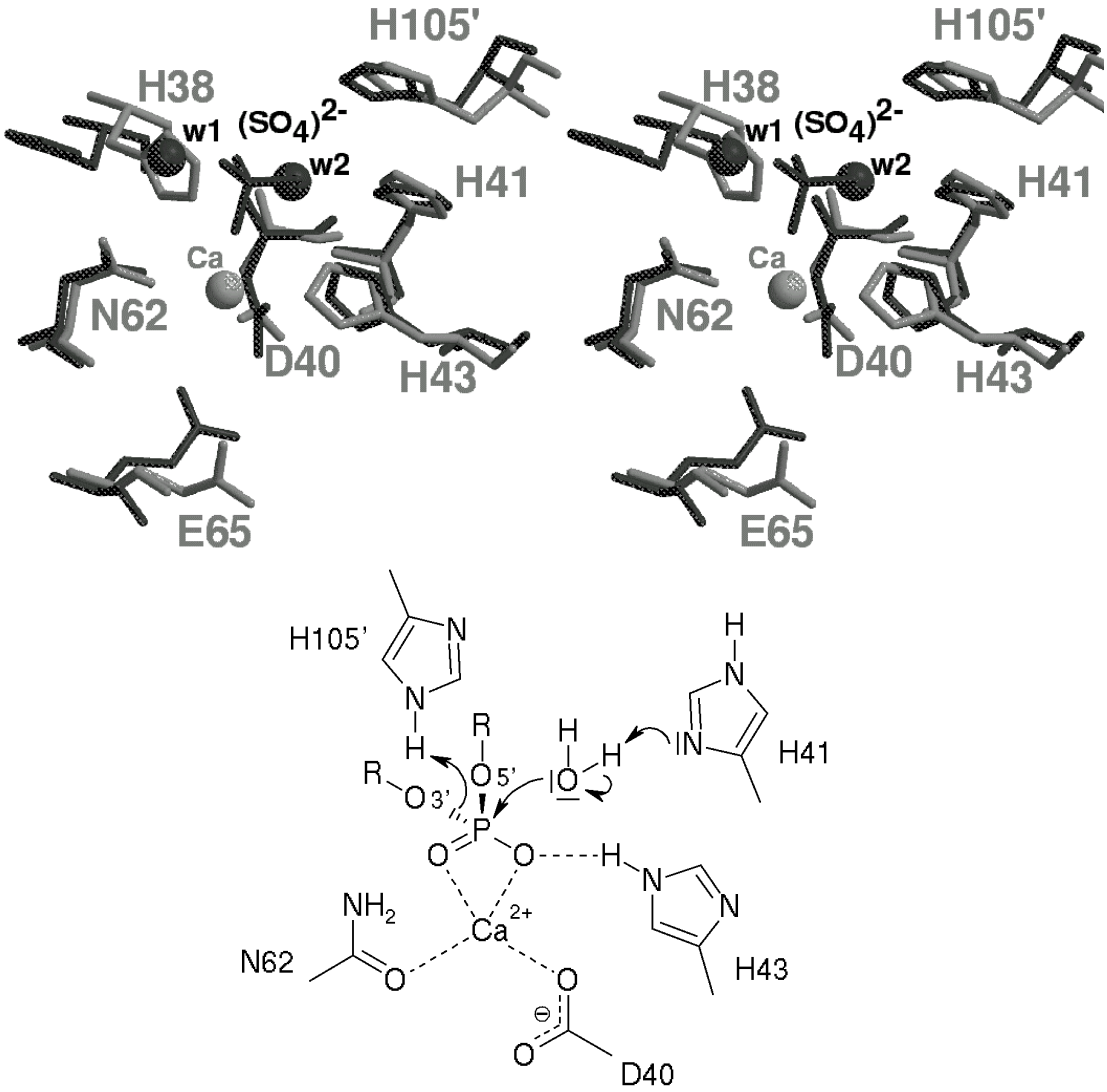

Figure 3.6. Stereo plot of the active sites of mutant EndoVII (black) and wild-type C2 (grey) superimposed on A-chain residues 40, 41 and 62. The calcium ion in the wild- type structure ligands to D40 Oδ1 (2.38Å) and N62 Oδ1 (2.42 Å) and is 3.55 Å from the sulphate (which belongs to the mutant structure). The sulphate itself is within hydrogen-bonding distance from H43 Nε1(2.94Å), water 1 (w1, 2.56Å) and water 2 (w2, 2.66Å). These waters in turn form hydrogen bonds to H38 (w1-Nδ1 = 2.79Å) and H41 (w2-Nδ1 = 2.68Å). D40 Oδ2 is expected to be protonated in the mutant structure considering the distances to E65 Oε1 (3.02Å) and Oε2 (2.93Å). Nε2 of H105' (belonging to the B-monomer of the EndoVII dimer) is 3.4Å from one of the sulphate oxygens.
A schematic diagram of the proposed reaction mechanism for EndoVII highlights the most important residues as identified by mutation studies.

molecules (Figure 3.4, middle). Helix 5 of a given monomer is partly inserted into the binding cleft of its neighbours. Instead, in wild-type C2 crystals only rather loose contacts involving residues of the loop region between helix 5 and 6 occur at this site.

Changes of the orientation of the C-terminal domain with respect to the four-helix cross.

Comparisons of the two monomers, either present in the asymmetric unit of the wild-type C2 crystals, or those in the N62D mutant, show no significant deviation from 2-fold non-crystallographic symmetry. The monomers can be superimposed with an rmsd of about 0.5Å for the Cα-positions in both cases (table 3.2). Differences within dimers are restricted to individual side chains (e.g. H38) and a few residues in the loop regions. Significant deviations from the molecular 2-fold (rmsd ~1.8Å for Cα atoms) are however observed within the dimer of the wild-type $P2_1$ structure. These deviations show that the relative orientation of the C-terminal domain consisting of helices 4, 5 and 6 relative to the four-helix cross is significantly different in the two molecules forming the dimer. Obviously, this reorientation is caused by different crystal contacts of the monomers. Figure 3.4 (bottom) illustrates that indeed the contacts involving the C-terminal domains are very different. It should be noted in this context, that model building studies suggest that the C-terminal domains form major-groove contacts with the DNA arms on either side of the junction or distortion (Raaijmakers *et al*., 1999).

The preceding analysis of the observed differences between the three structures reveals three major elements of conformational flexibility within the EndoVII dimer, which are indicated by arrows in figure 3.5: (1) The shift of the dimerization helices relative to each other; (2) the opening of the DNA binding cleft formed by the number 2 helices and the ß-finger; (3) the movement of the C-terminal domain relative to the four-helix bundle. While (1) and (2) will affect the interaction with the substrate at or near the active site, (3) will be important for binding to the DNA major groove some 6-9 bp away from the junction centre or the perturbation. In all three cases the conformational changes will influence substrate binding and/or cleavage and are therefore most probably of functional importance for the broad substrate specificity of the enzyme.

Ion binding at the active site suggests a possible catalytic mechanism

The putative active sites of the EndoVII dimer were revealed by the Ca^{2+} ion positions in the wild-type C2 structure (Raaijmakers *et al*.; 1999). They are located at the bottom of the cleft formed by the ß-finger and helix 2, a region which is lined by residues known to be essential for activity from site-directed mutagenesis experiments (reviewed in: White *et al*., 1997; Birkenbihl & Kemper, unpublished results; Golz *et al*., 1997). These catalytically essential residues include D40 and N62, which are direct ligands of the Ca^{2+}-ion, E65, and histidine 41. Additionally, histidines 105 and 43 appear to be important.

As discussed above, the relative orientation of the N- and C-terminal domains and the central four-helix cross differs significantly in the wild-type and the mutant structures resulting in a widening of the substrate binding cleft in the N62D mutant. These conformational differences also affect the disposition of some of the active site residues. Shown in figure 3.6 is a superposition of the active sites in the wild-type C2 structure (light grey) and the mutant protein (dark). Besides H38, which assumes different positions even within a dimer, E65 is clearly shifted and is located close to D40 in the mutant (distance ~3.0Å) indicating that at least one of the carboxylates is

Figure 3.7 Superposition of the conserved active site motifs of EndoVII (light-grey), I-PpoI (black) and *Serratia* nuclease (medium-grey). The Ca^{2+}-ion in EndoVII and the Mg^{2+}-ion in I-PpoI are represented as spheres. The conserved residues, H41 and N62 in EndoVII, H89 and N119 in *Serratia* nuclease and H98 and N119 in I-PpoI are shown in stick-mode. The PDB entries used are 1en7, 1a73 and 1smn.

probably protonated. While the 1.3Å density map of the mutant at this position does not allow an unambiguous location of the protons, it clearly reveals a sulphate-ion and two water molecules bound in the active site. The sulphate-ion is shown in figure 3.6 as a tetrahedron together with the two water molecules (w1 and w2) about 2.6Å away from the sulphate oxygens.
These water molecules are in turn in hydrogen bonding distance from the Nδ1 atoms of H41 and H38, respectively. Also in direct contact with the sulphate ion is H43 whose Nε1 atom is about 2.9Å away from a sulphate oxygen. The distance of the Ca^{2+}-position in the wild-type protein to one of the sulphate oxygens in the mutant is approximately 3.5Å if one superimposes the active sites, but would be shorter in the active substrate complex (see below).

Assuming that the sulphate-ion bound at the active site of the N62D mutant marks the position of the scissile phosphate in a substrate complex, the following catalytic mechanism is conceivable: H41 could act as a general base abstracting a proton from the attacking water (w2 in figure 3.6). The divalent metal-ion liganded to N62 and D40 and possibly also E65 would act as a Lewis acid activating the phosphate and positioning it with respect to the protein. H105 from the other monomer (labelled H105' in figure 3.6) or perhaps H43 might help stabilizing the positive charge of the pentacoordinate transition state and/or protonate the leaving O3'. This mechanism is consistent with mutagenesis experiments showing that mutations of D40, N62 or H41 completely destroy activity. A H105R mutant, while

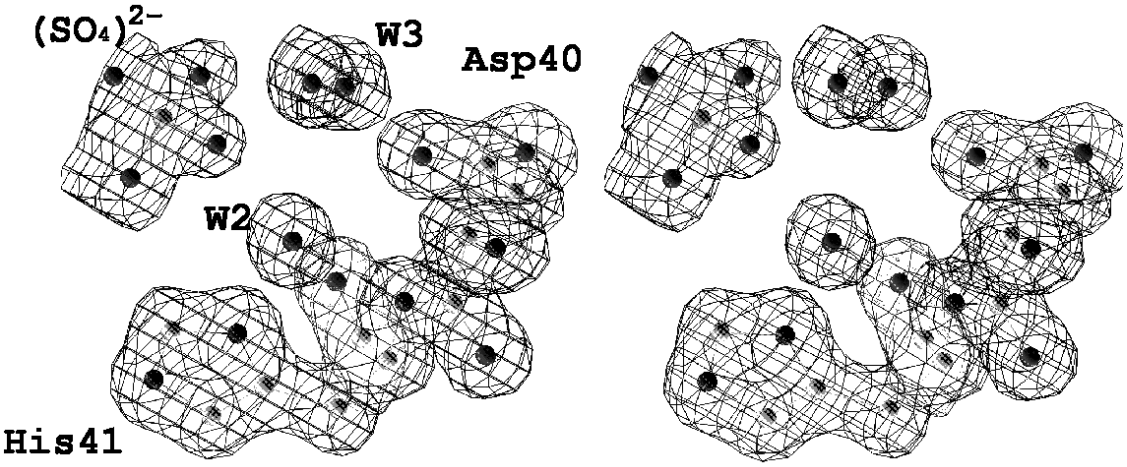

Figure 3.8 Electron density of the 1.32 Å $2mF_O$-DF_C map. (A) At the active site, depicting the sulphate, His41, proposed to activate water w2 (cf. Fig. 3.6) and water w3, which occupies a position similar to the Ca^{2+}-ion in the wild-type structures. Contour level is 1σ.

still binding DNA, is inactive (Birkenbihl & Kemper, unpublished results) and H43 mutants were reported to have almost no activity (H43T; Giraud-Panis & Lilley, 1996) or reduced activity (H43R; Birkenbihl & Kemper, 1998b).
H38, on the other hand, the other potential general base, is clearly not essential for activity (Birkenbihl & Kemper, unpublished results). A similar mechanism has been suggested for the homing endonuclease PpoI and *Serratia* endonuclease (Friedhoff *et al*., 1999)(see below).

Why is the N62D mutant inactive?

An aspartate residue at position 62 could also function as a ligand for the catalytic metal ion and therefore the mere exchange of an asparagine for an aspartate side chain in the N62D mutant is not expected a priori to lead to a complete loss of activity. The disturbance of the charge balance at the active site, the loss of a possible hydrogen bond between N62 Nδ1 and a phosphate group of the DNA, reorientation of E65 and secondary effects as a consequence of conformational changes may be responsible for the inactivity of the mutant. Since the mutant and the wild-type protein were crystallized at widely different pH values (4.5 and 8.2, respectively) and since no metal ion is bound at the active site of the mutant, a comparison of the structures does not directly provide an answer to the question, which of these factors is primarily responsible for the loss of activity. The wider opening of the binding cleft seen in the mutant will clearly affect substrate binding, however, as suggested by a comparison of the two wild-type structures, this may not be a consequence of the mutation, but rather an expression of the intrinsic flexibility of EndoVII.

A comparison of the two wild-type structures with the mutant indicates that the relative shift of the dimerization helices (helices 2 in figure 3.1) can be induced by the crystal packing (see discussion above). This shift does affect the distance of the active sites within the dimer, which differs by about 1.5Å in the two wild-type structures (as measured by the distance between the Ca^{2+} ions) and is between 3 to 5Å larger in the mutant. This difference may influence the enzymatic activity, but, more importantly, it highlights the internal flexibility enabling EndoVII to cleave substrates with varying distances between scissile phosphates in the two DNA strands.

Comparison with other resolvases and Mg^{2+}-dependent nucleases

A comparison with other junction-resolving enzymes did not reveal any overall structural homology with EndoVII, however, similar arrangements of certain active site residues and the divalent cation have been found in other Mg^{2+}-dependent nucleases. Carboxylates presumably liganding the catalytic metal ion are found in RuvC (D141, D7 and E66) and RNaseH (D70, D10 and E48) at positions corresponding to D40, N62 and E65 in EndoVII (Ariyoshi *et al*., 1994; Katayanagi *et al*., 1990). While this may be a simple consequence of the coordination geometry of the divalent cation, it still could indicate related catalytic mechanisms. In DNase I two histidines have been proposed to act as general base and acid respectively, however, their relative disposition is clearly different from that of H41 and H43 in EndoVII (Weston *et al*., 1992).

A structural motif reminiscent of the ß-finger and the α helix lining the active site cleft of EndoVII is present in a number of metal ion-dependent nucleases including *Serratia* endonuclease, the homing endonuclease PpoI and the endonuclease domain of colicin E9 (Miller *et al*., 1994; Flick *et al*., 1998; Kleanthous *et al*., 1999; Kühlmann *et al*., 1999). In case of *Serratia* endonuclease and the homing endonuclease PpoI part of the ß-finger and helix 2 lining the active site cleft in EndoVII can be superimposed with corresponding features in these proteins with rmsd values of 1.75 and 1.35Å for 14 and 15 Cα-positions, respectively (Figure 3.7). These architectural similarities together with the similar disposition of an essential asparagine (N62 in EndoVII; N119 in *Serratia* and PpoI) and histidine (H41 in EndoVII; H89 in *Serratia*; H98 in PpoI) relative to the divalent cation support the notion of a common mechanism for these nucleases despite their widely differing specificities and may be the result of convergent evolution.

The superposition of the active sites of EndoVII C2wt and PpoI in complex with DNA (PDB entry 1a73) reveals that the Ca^{2+}-ion is buried about 2.4 Å deeper in the binding cleft than the Mg^{2+}-ion in PpoI, while the gap between the beta-finger and helix 2 is smaller in PpoI, by about 2 Å near H41. An exact superimposition of the metal ions, while keeping D40 as a ligand, requires a ~20^{o} rotation of helix 2 of EndoVII about its axis and another rotation of ~10^{o}, in the opposite direction, along the length of the ß-finger. The refinement of TLS parameters discussed above indicate that such rotations are not unlikely.

Acknowledgements

We would like to thank our former colleague Olivier Vix for help with the purification of the wild-type EndoVII protein. We gratefully acknowledge the help of the staff of the EMBL outstation in Hamburg, in particular of Paul Tucker, with data collection at the BW7B beam line. This work was supported in part by grants from the Deutsche Forschungsgemeinschaft to D.S. (SU 61/6-1) and B.K. (Ke 188/16-1). Part of the work in the laboratory of B. Kemper was also supported by the Fonds der Chemischen Industrie.

References

Ariyoshi, M., Vassylyev, D., Iwasaki, H., Shinagawa, H. and Morikawa, K. (1994) Atomic structure of the RuvC resolvase: a Holliday junction specific endonuclease from *E. coli. Cell,* **78,** 1063-1072.

Bhattacharyya , A. Murchie, A., Von Kitzing , E., Diekmann, S., Kemper, B. And Lilley, D. (1991) Model for the interaction of the DNA junctions and resolving enzymes. *J. mol biol*. **221**, 1191-1207.

Birkenbihl, R. and Kemper, B. (1998a) Localization and characterization of the dimerization domain of Holliday-structure resolving endonuclease VII. *J. Mol. Biol.,* **280**, 73-83.

Birkenbihl, R. and Kemper, B. (1998b) Endonuclease VII has two DNA-binding sites each composed from one N- and one C-terminus provided by different subunits of the protein dimer. *EMBO J.* **17**, 4527-4534.

Computational Project, Number 4, (1994), "The CCP4 Suite: Programs for Protein Crystallography", *Acta Cryst*. **D50**, 760-763.

Cowtan, K. (1994) 'dm': an automated procedure for phase improvement by density modofication. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, **31**, p34-38.

Friedhoff, P., Franke, I., Krause, K.L. and Pingoud, A. (1999) Cleavage experiments with deoxythymidine 3',5'-bis-(p-nitrophenylphosphate) suggest that the homing endonuclease I-PpoI follows the same mechanism of phosphodiester bond hydrolysis as the non-specific *Serratia* nuclease. *FEBS Lett*. **443**, 209-214.

Flick, K.E., Jurica, M.S., Monnart Jr, R.J. and Stoddard, B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* **394**, 96-101.

Giraud-Panis M., and Lilley, D. (1997) Near simultaneous DNA cleavage by the subunits of the junction resolving enzyme T4 endonuclease VII. *EMBO J.* **16**, 2528-2534.

Golz, S., Christoph, A., Birkenkamp, K. and Kemper, B. (1997) Identification of amino acids of endonuclease VII essential for binding and cleavage of cruciform DNA. *Eur. J. Biochem.* **245,** 573-580.

Golz, S., Birkenbihl, R.. and Kemper, B. (1995) Improved large scale preparation of phage T4 endonuclease VII overexpressed in *E. coli*. *DNA Res*., **2**, 277-284.

Grebenshchikova S., Plugina, L. and Sccherbakov, V. (1994) The role of T4-bacteriophage endonuclease VII in autocorrection of mismatched regions. *Genetica*, **30**, 622-626.

Greger, B. and Kemper, B. (1998) An apyrimidinic site kinks DNA and triggers incision by endonuclease VII of phage T4. *Nucleic Acids res.,* **26**, 4432-4438

Howlin, B. Butler, S.A.,Moss, D.S., Harris, G.W. and Driessen, H.P.C. (1993) "TLSANL: TLS parameter analysis program for segmented anisotropic refinement of macromolecular structures." *J. Appl. Cryst*., **26**, 622-624.

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996) Errors in protein structures. Nature, **381**, 272.

Jones, T.A., Zou, J., Cowan, S. and Kjelgaard, M., (1991) Improved methods for building protein methods in electron density maps and the location of errors in these models. Acta Cryst. **A47**, 110-119.

Kabsch W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr.*. **A32** 922-923.

Katayanagi, Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, M. and Morikawa, K. (1990) Three dimensional structure of ribonuclease H. from *E. coli. Nature,* **347**, 306-309.

Kemper, B. and Brown, D. (1976) Function of gene 49 of bacteriophage T4 II. Analysis of intracellular development and the structure of very fast sedimenting DNA. *J. virol.* **18**, 1000-1015.

Kemper, B., Pottmeyer, S., Solaro, P. And Kosak, H. (1990) Resolution of DNA secondary structures by endonuclease VII (EndoVII) from phage T4. In Sarma,R.H. And Sarma, M.H. (eds), *Structure and methods. Vol. 1 Human genome initiative and DNA recombination.* Adenine press, Schenectady, NY, 215-229.

Kemper, B. (1997) Branched DNA resolving enzymes. In Nickoloff, J. and Hoekstra, M., (eds), *DNA damage and repair. Vol 1. DNA repair in prokaryotes and lower eukaryotes.* Humana press inc. Totowa, NJ, 179-204.

Kleanthous, C., Kühlmann, U.C., Pommer, A.J., Ferguson, N., Radford, S.E., Moore, G.R., James, R. & Hemmings, A.M. (1999) Structural and mechanistic basis of immunity toward endonuclease colicins. *Nature Struct. Biol*. **6**, 243-252.

Kleywegt, G. J. & Jones, T. A. (1996).Phi/psi-chology: Ramachandran revisited. *Structure*, **4**, 1395-1400.

Kraulis, P.J., (1991), Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst*., **24**, 946-950.

Kühlmann, U., Moore, G., James, R., Kleanthous, C. And Hemmings (1999) Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redifiend? *FEBS Lett*. **463**, 1-2

Laskowski, R., McArthur, M.W., Moss, D.S. & Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst*. 26 283-291.

Lamzin, V.S. & Wilson, K.S. (1993). Automated refinement of protein models. *Acta Crystallogr*. **D49**, 129-147.

Lee, B. and Richards, F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J.Mol.Biol.* **55**, 379 - 400.

Merritt, E.A. & Bacon, D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods in enzymology* **277**, 505-524.

Miller, M.D., Tanner, J., Alpaugh, M., Benedik, M.J. and Krause, K.L. (1994) 2.1Å structure of *Serratia* endonuclease suggests a mechanism for binding to double-stranded DNA. *Nature Struct. Biol*. **1**, 461-468.

Mizuuchi, K.,Kemper, B., Hays, J. and Weaisberg, R. (1982) T4 endonuclease VII cleaves Holliday structures, *Cell*, **29**, 357-365.

Murshudov,G., Lebedev, A., Vagin,A.,Wilson, K., and Dodson, E. (1999) "Efficient anisotropic refinement of Macromolecular structures using FFT", *Acta Cryst*. **D55**, 247-255.

Navaza, J. (1994) AmoRe: an automated package for molecular replacement. *Acta Crystallogr.* **A50**, 157-163.

Otwinowski, Z. & Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology* **276**, 307-326.

Pottmeyer , S. and Kemper, B. (1992) T4 endonuclease VII resolves cruciform DNA with nick and counter nick and its activity is directed by its local nucleotide sequence. *J. Mol. Biol.* **223**, 607-615.

Raaijmakers, H., Vix, O., Toro, I., Golz, S., Kemper, B. and Suck, D. *(*1999) X-ray structure of T4 endonuclease VII: a DNA junction resolvase with novel fold and unusual domain swapped architecture. *EMBO J*., **18**, 1447-1458.

Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J*. **5**, 909-933.

Schomaker, V. and Trueblood, K.N. (1968) On the rigid-body motion of molecules in crystals. *Acta Cryst*., **B24**, 63-76.

Shah, R., Cosstick, R. and West, S.C. (1997) The RuvC protein dimer resolves Holliday junctions by a dual incision mechanism that involves base-specific contacts. *EMBO J.* **16**, 1664-1672.

Solaro, P., Birkenkamp, K., Pfeiffer, P., and Kemper, B. (1993) Endonuclease VII of phage T4 triggers mismatch correction in vitro. *J. Mol. Biol.***230**, 868-877.

Weston, SA, Lahm A, Suck D. (1992) X-ray structure of the DNase-I d(GGTATACC)2 complex at 2.3 A resolution. *J Mol Biol.* **226**, 1237-1256.

White, M.F. and Lilley, D.M.J. (1996) The structure-selectivity and sequence-preference of the junction-resolving enzyme CCE1 of *Saccharomyces cerevisiae*. *J. Mol. Biol.* **257**, 330-341.

White, M., Giraud-Panis, M., Poehler, J. and Lilley, D. (1997) Recognition and manipulation of branched DNA structure by junction resolving enzymes. *J. Mol. Biol.,* **269,** 647-664.

Winn, M., Isupov, M., and Murshudov, G.N. (2000), "Use of TLS parameters to model anisotropic displacements in macromolecular refinement" *Acta Cryst*., submitted.

Youil, R., Kemper, B. and Cotton, R. (1995) Screening for mutations by enzyme mismatch cleavage by T4 endonuclease VII. *Proc. Natl. Acad. Sci. USA* **92,** 87-9

4

Tungsten-containing formate dehydrogenase from *Desulfovibrio gigas*: Metal identification and preliminary structural data by multi-wavelength crystallography

Hans Raaijmakers, Susana Teixeira, João Miguel Dias, Maria João Almendra, Carlos D. Brondino, Isabel Moura, José J. G. Moura and Maria João Romão

Journal of Biological Inorganic Chemistry (Accepted)

Abstract

The tungsten-containing formate dehydrogenase (W-FDH) isolated from *Desulfovibrio* (*D.*) *gigas* has been crystallized in space group $P2_1$, with cell parameters a=73.8 Å, b=111.3 Å, c=156.6 Å and $\beta = 93.7^{o}$. These crystals diffract to beyond 2.0 Å on a synchrotron radiation source. W-FDH is a heterodimer (92kDa and 29kDa subunits) and two W-FDH molecules are present in the asymmetric unit. Although a molecular replacement solution was found using the periplasmic nitrate reductase as a search model, additional phasing information was needed. A MAD dataset was collected at the W- and Fe-edges, at four different wavelengths. Anomalous and dispersive difference data allowed to unambiguously identify the metal atoms bound to W-FDH, as one W atom with a Se-cysteine ligand as well as one [4Fe-4S] cluster in the 92 kDa subunit, and three additional [4Fe-4S] centers in the smaller 29 kDa subunit.

The *D. gigas* W-FDH was previously characterized based on metal analysis and spectroscopic data. One W atom was predicted to be bound to two MGD pterin cofactors and two [4Fe-4S] centers were proposed to be present.

The crystallographic data now reported reveal a selenium atom (as a Se-cysteine) coordinating to the W site, as well as two extra [4Fe-4S] clusters not anticipated before. The EPR data were re-evaluated on the light of these new results.

4.1 Introduction

Desulfovibrio gigas formate dehydrogenase (W-FDH) is one of the first examples of a tungsten-containing formate dehydrogenase isolated from a mesophile, grown in a normal medium, not depleted of Mo [1]. W-FDH catalyses the oxidation of formate to CO_2, a reaction which involves the release of two electrons. The enzyme consists of two subunits, of 92 KDa and 29 KDa, (SDS-PAGE) and it belongs to the DMSO Reductase family of enzymes, one of the four classes into which molybdopterin-containing enzymes have been classified [2,3]. This family is considerably broad and, besides DMSOR [4,5], it includes enzymes such as dissimilatory nitrate reductase (NAP) [6] and formate dehydrogenase. The members of this family have two molybdopterins (MGD cofactor) in the coordination sphere of Mo (or W), which is also bound to an amino acid side chain, a serine residue in DMSOR, a cysteine in NAP and a Se-Cys in *E. coli* FDH-H [7].

Spectroscopic data as well as the presence of two MGD cofactors coordinating the W atom, and the enzyme function, suggested that the large subunit of W-FDH should be homologous to both the periplasmic nitrate reductase from *D. desulfuricans* ATCC 27774 (NAP) [6] and *E. coli* FDH-H [7], of which the 3D-structures have been determined, but also to the similar *E. coli* formate dehydrogenase-O (FDH-O) [8] or FDH-N [9]. In analogy, it is also likely that the small subunit of W-FDH is related to the smaller beta-subunits of the latter enzymes. Since the primary sequence of W-FDH is still not available, a similarity search in the Swiss-Prot database [13] was performed, starting from the small beta subunit of FDH-N. Based on the sequence, this search revealed other proteins with multiple potential [4Fe-4S] clusters. The beta subunit of FDH-O is predicted to have three [4Fe-4S] clusters [8]. Other homologue proteins like FDH-N can potentially bind four [4Fe-4S] clusters [9], as well as *hmc2* from *D. vulgaris* [10], FDH from *Haemophilus* (*H.*) *influenzae* [11] and the DMSO reductase iron-sulfur subunit from *H. influenzae* [11]. The *hyb* operon encoding *E. coli* hydrogenase 2 (*hybA*) can bind three [4Fe-4S] clusters and one [3Fe-4S] [12].

Metal analysis, EPR and Mössbauer spectroscopy predicted the existence of two [4Fe-4S] clusters besides the tungsten atom, coordinated by two MGD (molybdopterin guanine dinucleotide) cofactors [1]. Now, two additional [4Fe-4S] clusters have been identified by crystallography, corresponding to one cluster in the large subunit and three clusters in the smaller subunit in agreement with the predictions for related FDHs.

4.2 Materials and methods

Protein purification

Purification of *D. gigas* formate dehydrogenase was performed accordingly to the procedure described in reference [1]. The purification scheme included two first steps of anionic exchange chromatography (DEAE-52 and Source 15), a gel filtration step (Superdex 200) and another step on an anionic exchange column (Source 15). An extra purification step was introduced, where some remaining impurities were precipitated by adding isopropanol to a final concentration of 30% in an eppendorf reaction tube. This was left overnight at 4ºC, and centrifuged for ten minutes at

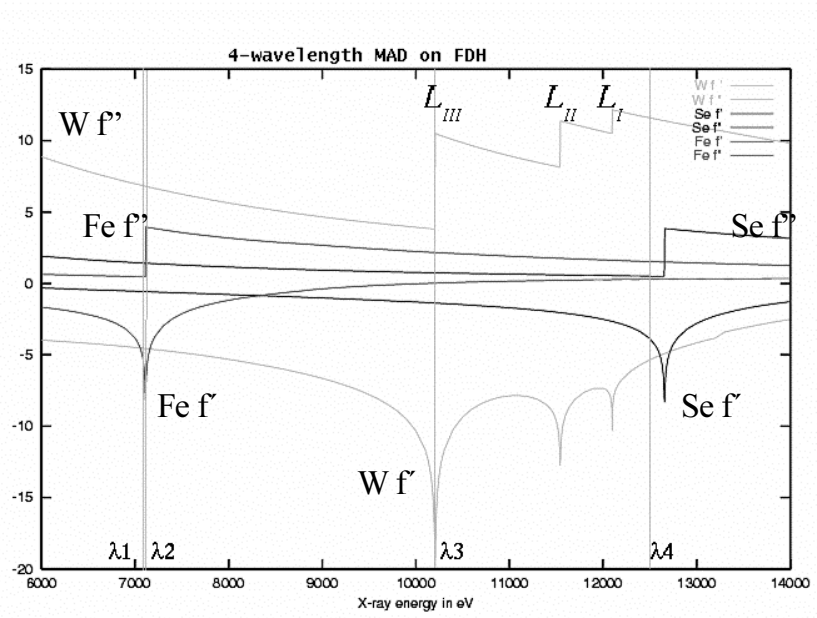

Figure 4.1 Calculated dispersion terms f′ and f" as a function of the photon energy for W, Fe and Se. Vertical lines indicate the energy at which the MAD datasets were collected. Also indicated are the tungsten L absorption edges. λ1 and λ2 are on the iron absorption edge, while λ4 is close to the selenium absorption edge. The plot was made with public available software [27,28].

maximum speed in an eppendorf table centrifuge. The supernatant was judged pure by SDS-PAGE with coomassie staining and was successfully used for crystallization.

Crystallization

Crystals from W-FDH were obtained using sitting drops with the vapor diffusion method. Red-brown needle-shaped crystals grew in 1-2 days at 4º C in drops containing 0.1 M HEPES (pH 7.5), 20% (w/v) PEG 3,350, 10% (v/v) isopropanol and 1% β–octylglucoside as an additive. The crystals were obtained in a very low yield, considering the number of crystallization trials that were set up. The few crystals obtained were first visible in the shape of very thin needles, which grew in thickness within 2 to 3 days. An optimum time interval had to be found as the original single crystals tend to form aggregates after sitting in the drop for some time.

Additional crystals were required to perform the MAD data collection. A new batch of protein was used for this purpose and again the optimum crystallization conditions had to be fine tuned, as the ones found for the previous batch failed to produce any good-quality crystals. A screening of conditions was made without any satisfactory results other than very small thin needles that grew in similar conditions as before after 1 to 2 weeks. Diffraction quality crystals were finally obtained through macro seeding of these needles, which grew bigger in 2 to 7 days.

Data collection

Cryo conditions, mother liquor + 10% PEG-400, were established and Synchrotron X-ray data were first collected at 100K in the BW6 beam line of the MPG-ASMB in DESY. 163901 unique reflections were collected between 22.6 and 2.0 Å resolution. Data were processed with Denzo and Scalepack [14] in space group $P2_1$ with cell dimensions a = 73.8 Å, b = 111.3 Å, c = 156.6 Å and β= 93.7°. An overall I/σ of 11.9 was obtained, with an R_{merge} of 7.1% (27% in the outer shell) and a

Table 4.1: Data collection and phasing statistics. Statistics are based on independent Bijvoet pairs.

	λ_1	λ_2	λ_3	λ_4*	**DESY**
λ (Å)	1.74149	1.73649	1.21417	0.992	1.06
# reflections	82298	82736	128934	163885	163901
(redundancy)	(1.9)	(1.9)	(1.8)	(2.3)	(4.5)
Resolution (Å)	20.0 - 2.50	20.0 - 2.50	40.0 - 2.15	40.0 - 2.00	25.0 – 2.0
Completeness (%)	93.0 (77.2)[1]	94.9 (84.9)[1]	95.4 (47.0)[1]	97.0 (91.0)[1]	99.2 (99.6)[1]
R_merge	0.10 (0.39)[1]	0.10 (0.40)[1]	0.07 (0.45)[1]	0.07 (0.30)[1]	0.07 (0.27)[1]
I/σ	5 (1.5)[1]	5 (2.0)[1]	5 (2.0)[1]	5 (2.0)[1]	12 (3.5)[1]
R_{cullis} (cen/acen)	0.72/0.85	0.76/0.82	0.74/0.74	- /0.87	-/0.82
Phasing power dispersive /anomalous	1.9/1.4	1.7/1.6	0.69/1.4	- / 1.4	-/0.80
Figure of merit λ_1–λ_4			0.45		
*reference dataset		MAD: a = 73.4 Å, b = 110.4 Å, c = 156.2 Å, ß = 93.5°			
[1] Last shell		DESY: a = 73.8 Å, b = 111.3 Å, c = 156.6 Å, ß = 93.7°			

completeness of 99.2% (99.6% in the outer shell) (Table 4.1). Calculation of the solvent content [15] suggested the presence of two heterodimers (92kDa x 29kDa) in the asymmetric unit, with the corresponding Matthew's ratio, V_M of 2.65 $Å^3$/Da. This is within the average range of V_M values found for protein crystals and indicates a solvent content of approximately 54%.

MAD data collection

The four wavelength MAD data sets were collected at BM-14 beam line at ESRF on a single crystal positioned in "mirror" orientation, so that the Bijvoet pairs were collected at the same image, and cooled to 100K with an Oxford cryostream system. Unfortunately, the observed fluorescence signal from the crystal was too weak to measure the Fe and W absorption edges and two separate metal foils of Fe and W were used. The crystals diffract up to 2.5Å around the iron K edge (Fe inflection point λ_1 at 1.74149 Å and Fe absorption peak λ_2 at 1.73649 Å) and up to 2.0Å at the tungsten L-III inflection point λ_3 (1.21417 Å) and near the tungsten L-I edge at λ_4 (0.992 Å) (Figure 4.1). The data sets were 93-97% complete (see Table 4.1).

Data were processed with Denzo & Scalepack [14] in space group $P2_1$ with cell dimensions (slightly different than those collected at DESY, Hamburg): a = 73.4 Å, b = 110.4 Å, c = 156.2 Å, β= 93.5°. R_{merge} was around 10% near the iron edge, and 7% at both shorter wavelengths.

Molecular replacement

Due to the high sequence homology (43%) and structural resemblance between *E. coli* FDH-H and *D. desulfuricans* 27774 NAP, both structures were tried as search models to solve the structure of *D. gigas* W-FDH by molecular replacement. These models correspond only to the large 92kDa subunit. The program AMoRe was used for this purpose [16,17]. Using FDH-H as a starting model an exhaustive number of changes were made, aiming at obtaining different models that would be more appropriate to input in AMoRe, based on both structural homology between NAP and FDH-H and sequence homology in the DMSO reductase family. However, no acceptable solutions were found in any of the resolution ranges scanned. The same

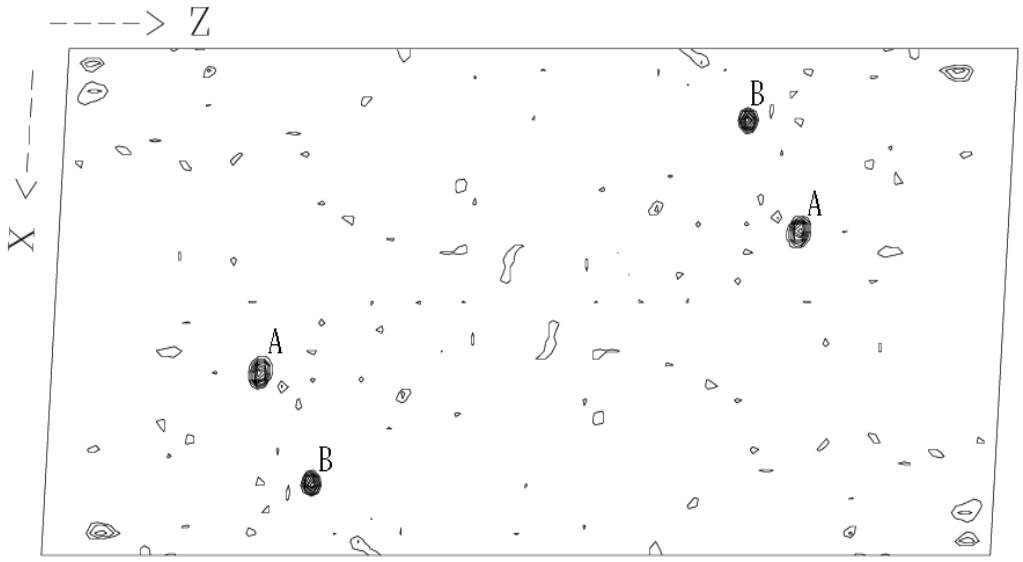

Figure 4.2 The v=1/2 Harker section of the anomalous difference Patterson function in $F_{\lambda 4}$, including data between 15 - 3 Å. Contouring is 2σ to 10σ in 1σ intervals. The overlapping tungsten peaks have a maximum of 11σ, or 6.5% of the origin peak. The 11σ tungsten cross-peaks (B) are also in the same Harker section because of the pseudo B2 symmetry of the crystal.

procedure was repeated using NAP. A solution was obtained with the sequence mutated to poly-alanines, and having removed the metals, the cofactors and the residues where the structures of NAP and FDH-H differ most. The optimal resolution range was 4-15 Å, and a correlation factor of 49.7% indicated an optimal molecular replacement solution corresponding to two W-FDH molecules per asymmetric unit.

SAD phasing

The 80% complete anomalous signal present in the native dataset (DESY data set: λ = 1.06 Å) allowed us to locate the tungsten atoms manually, using the anomalous difference Patterson function (shown in Figure 4.2, when recalculated with data from λ_4). The overlapping tungsten peaks (A) have a maximum of 11σ, or 6.5% of the origin peak. The 11σ tungsten cross-

peaks (B) are present in the same Harker section because of the pseudo B2 symmetry of the crystal. Initial phases (FOM = 0.16) were obtained with the program SHARP [18]. These phases were combined with the molecular replacement phases using the program SIGMAA from the CCP4 suite [17] and improved with solvent flattening and two fold averaging in DM [17,19]. Resulting maps were displayed in "O" [29] and it was attempted to rebuild the search model. The improved phases were strong enough to allow identification of one predicted [4Fe-4S] cluster as well as several secondary structure elements.

MAD Phasing

Directly after data collection it was tried to identify the iron atoms in a Fe specific dispersive difference map ($F_{\lambda 4}$ - $F_{\lambda 1}$) calculated with the previously obtained set of phases. This map showed the second predicted [4Fe-4S] cluster and hinted the existence of some extra Fe atoms, barely above the noise level. Four-wavelength refinement of the heavy metal structure in SHARP, which included the 2x1 W and 2x8 Fe positions known by then, did not converge. Then, only the two W sites were refined

in SHARP, using two datasets, the remote (λ_4) and the tungsten inflection point (λ_3). Phases thus obtained (FOM=0.22) were combined with previously obtained model phases and improved with solvent flattening and 2-fold averaging using DM [17,19] and produced a much better phased dispersive difference map ($F_{\lambda 4}$ - $F_{\lambda 1}$). It was now possible to unambiguously locate a total of four [4Fe-4S] clusters per heterodimer (Figure 4.3a and 4.3b). All 32 iron atom positions and 2 tungsten atom positions were included in the next, four wavelength refinement (FOM=0.42), which improved the phases so much that the selenium atom could now be located in coordination to the tungsten atom, as a selenium cysteine (as is found in *E. coli* FDH-H [7]). In the third round of refinement, including the selenium sites, the figure of merit rose to 0.45. To enhance convergence and decrease the number of refined parameters, refinement was done with isotropic B factors and fixed occupancy (1.0). All wavelength dependent scattering factors (f′ and f") were fixed to theoretical values [20], except those near absorption edges: Fe f′ and f" at λ_1 and λ_2, W f′ and f" at λ_3 and Se f′ at λ_4, which were refined in SHARP (Table 4.2) since the fluorescence signal observed was too weak to measure them directly. After combination with model phases, 2-fold averaging and solvent flattening, and improvement with ARP/wARP the FOM rose to 0.76. However, at this stage it was not yet possible to perform automatic model building of any polypeptide with ARP/wARP [21].

Table 4.2 Theoretical and refined wavelength dependent scattering factors in electrons. [20]

	λ_1	λ_2	λ_3	λ_4*
Fe f'	-6.11 (-6.22)^r	-5.46 (-4.58)^r	-0.0255	0.2417
Fe f''	2.2 (4.02)^r	3.93 (4.75)^r	2.1638	1.5339
W f'	-5.26	-5.275	-14.3 (-15.9)^r	-5.86
W f''	6.764	6.733	7.2 (16.4)^r	11.72
Se f'	-0.669	-0.674	-1.468	-3.988 (2.71)^r
Se f''	1.413	1.405	0.7394	0.5110

* Reference dataset ^{r} = refined value

Iron determination

Iron determination was performed by the TPTZ (tripyridyl-S-triazine) method [31].

EPR Spectroscopy

Variable temperature EPR measurements at X-band were performed on a Bruker EMX spectrometer equipped with a rectangular cavity (Model ER 4102ST) and an Oxford Instruments continuous flow cryostat. EPR spectra were simulated using the program WIN-EPR simfonia V.1.2 from Bruker Instruments, Inc. The signals associated with each paramagnetic were separately simulated and normalized. Then, they were added altogether in different proportions until a good agreement was obtained between experimental and simulated data. Spin quantitation was performed at the same temperature under non-saturating conditions using Cu-EDTA as standard.

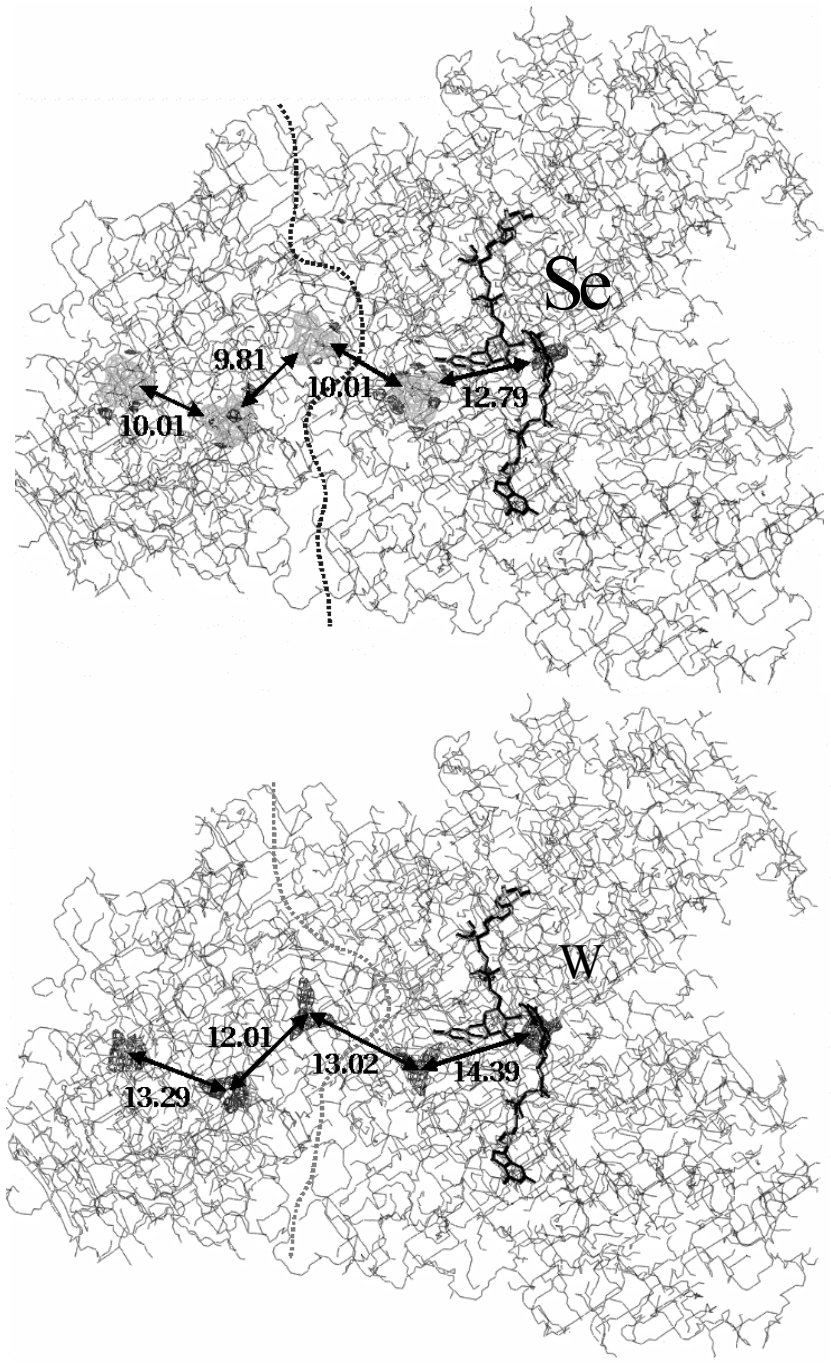

Figure 4.3 A) "Bones" representation of the skeletonized 2Fo-Fc electron density map calculated by ARP/wARP [21]. The dotted line has been drawn to define the approximate subunit interface. The right hand side corresponds to the larger subunit, homologous to NAP, while the left accounts for the new small subunit. The 2Fo-Fc electron density map is contoured at 6σ level. Clearly visible are the four [4Fe-4S] clusters, one present in the large subunit and three in the small. More to right of the molecule is the density corresponding to the W-Se, and some density is also clear for the phosphate groups of the MGD cofactors. Distances (Å) indicated are between the centers of mass of the iron clusters (and the tungsten atom).
B) Similar "bones" representation as that of figure 4.3a, but with the dispersive difference electron density maps density ($F_{\lambda 4}$ - $F_{\lambda 1}$), contoured at –4σ (dark grey, near Se) and +4σ (medium grey, near Fe). The four [4Fe-4S] clusters ($f'_{\lambda 4,Fe}$ - $f'_{\lambda 1,Fe}$ = + 6.5 e^- signal) are clearly defined. The deepest negative peak coincides with the Se position ($f'_{\lambda 4,Se}$ - $f'_{\lambda 1,Se}$ = -2.1 e^- signal), but the tungsten ($f'_{\lambda 4,W}$ - $f'_{\lambda 1,W}$ =-0.6 e^- signal) is also contoured at –4σ level. Distances indicated (Å) are between the nearest iron atoms and the tungsten. The nearest distance between the pterin (atom N1) and the first [4Fe-4S] cluster is only 6.9 Å.

4.3 Results and discussion

Structure solution

Since the *D. gigas* W-FDH primary sequence is not known, initial attempts to solve its structure were based on the assumption that there should be a structural relationship (of the large subunit) to other enzymes of the same family, which crystal structures are known, namely the formate dehydrogenase H (FDH-H) from *E. coli* [7] and the periplasmic nitrate reductase (NAP) from *D. desulfuricans* [6]. The clearest molecular replacement solution was obtained using NAP as a search model, which had been partially mutated and included in the calculations without the cofactors. In agreement with the packing density of 2.65 $Å^3$/Da two molecules were located in the asymmetric unit with a correlation coefficient of 49.7% (against a background of 29.1%) and an R factor of 53.3 % (against a background of 59.3%). First, the search was made for only one of the molecules, which was fixed as a solution and introduced in a translation-search for the second molecule. The correctly positioned molecules and their symmetry mates were examined in a graphics system. No clashes were observed and the molecules presented reasonable van der Waals contacts.

The absence $h + l = 2n + 1$ reflections of the X-ray data at low resolution suggests pseudo B2 symmetry. This is in agreement with a two-fold non-crystallographic symmetry axis (NCS), almost parallel to the b axis, which is confirmed by the presence of a strong peak in the native Patterson function at u=½, v=0, w=½. This fact limits the effectiveness of the two-fold averaging, since averaging mainly improves the systematically weak, and thus least contributing, reflections, therefore resulting in a poor map improvement by two-fold averaging.

Despite the fact that the cofactors and the metals had been omitted from the search model, positive electron density appeared in the regions of the maps calculated with XPLOR [22] where the cofactors were to be expected, according to the NAP model. These results confirmed the correct positioning of the model in the W-FDH crystal lattice. However, the quality of the electron density maps was still quite poor and additional phasing information was needed.

SAD phasing was performed using the anomalous signal present in the native data set collected at λ=1.06Å. This signal was clear enough to locate the two W atoms present in the asymmetric unit. The resulting phases were sufficient to find one predicted [4Fe-4S] cluster (the same as found in NAP and FDH-H) and several secondary structure elements. However, these phases were still not strong enough to improve the model without primary sequence information and a valuable contribution for phasing was obtained from MAD experiments

MAD experiment lead to metal analysis by crystallography

As the experiment was performed at different wavelengths, the wavelength dependent scattering properties could be used to locate and identify the heavy metals in the protein. The chosen wavelengths were optimized for Fe and W and the unanticipated Se atom could also be identified as a Se-cysteine coordinating the W atom.

The tungsten sites were localized with the Bijvoet anomalous difference Patterson function of λ_4 (Figure 4.2). Two overlapping peaks of 11σ, or 6.5% of the origin peak pinpointed the two tungsten sites, which are the strongest anomalous scatterers in the protein, at this wavelength. The metal identity was further supported

by its presence in phased ($F_{\lambda 4}$ - $F_{\lambda 3}$) dispersive difference maps, where the signal for tungsten ($f'_{\lambda 4,W}$ - $f'_{\lambda 3,W}$) is of about 10.0e^-. Additionally, in the maps ($F_{\lambda 4}$ - $F_{\lambda 1}$) and ($F_{\lambda 2}$ - $F_{\lambda 1}$), which are specific for locating iron, there is an absence of peaks for the W sites. In these maps, W should theoretically provide a signal of less than 0.6e^- and 0.01e^- respectively. Moreover, the determined tungsten sites coincide with the highest peaks present in 2Fo-Fc maps from ARP/wARP [21], which are 40σ above the average electron density.

The iron sites corresponding to 4x [4Fe-4S] clusters were located in the phased dispersive difference maps ($F_{\lambda 4}$ - $F_{\lambda 1}$) (Figure 4.3b). Their identity was confirmed in a weak but much more specific difference map ($F_{\lambda 2}$ - $F_{\lambda 1}$) ($f''_{\lambda 2,Fe}$ - $f''_{\lambda 1,Fe}$ gives only 1.64e^- signal, instead of 6.4e^- for $f''_{\lambda 4,Fe}$ - $f''_{\lambda 1,Fe}$). The wavelengths λ1 and λ2 are so similar that only Fe or Eu could give the difference signal observed. Additionally, the phased anomalous difference map $F_{\lambda 2}$, with about 4.7e^- signal ($f''_{\lambda 2,Fe}$), confirmed the identity of the iron atoms.

Selenium gives peaks in the ($F_{\lambda 4}$ - $F_{\lambda 1}$) map (Figure 4.3b), of about half the magnitude of the iron peaks but of opposite sign, which was indeed observed after the phases (based on W and Fe sites) were improved with ARP/wARP. The distance of Se to W refined to values between 2.49Å- 2.53Å in three different unrestrained ARP/wARP runs.

From these maps, the positions of 16 Fe, 1W and 1 Se atoms per heterodimer were identified and then refined with ARP/wARP.

The quality of the final electron density maps was not enough to allow building a reasonable model without the primary sequence data. However, after skeletonizing and visualizing the final 2Fo-Fc map in "O" [29] the boundaries of the protein were clear (Figure 4.3). The NAP model fitted well within the envelope, with $W(MGD)_2$ and one [4Fe-4S] cluster at almost the same position as in the W-FDH electron density map, thus defining the large subunit (92 kDa). In the skeletonized map there is enough space left to accommodate the small subunit (29 kDa) containing three more [4Fe-4S] clusters. In these preliminary maps it seems clear that the two clusters closest to the large subunit are enclosed by three β-strands and two α-helices, in a topology reminiscent of 2x[4Fe-4S] ferredoxin [30].

Reinterpretation of the metal analysis and spectroscopic data

In the previous report on W-FDH [1] it was concluded, based on iron determination, that the enzyme only contained two [4Fe-4S] clusters. It is now evident by the X-ray analysis that there is one [4Fe-4S] cluster in the large subunit and three [4Fe-4S] clusters in the small subunit.

Therefore, we have now repeated the iron determination and the value obtained is very similar to the one published before. To explain this discrepancy two hypotheses can be postulated: i) the protein determination by the Lowry method is erroneous or ii) the protein in solution contains molecules that have lost their metal centers and the crystallization procedure has selected the holoprotein form. We favor the second hypothesis for the following reasons. The number of tungsten atoms per protein is as expected (therefore, forms of the protein that are devoid of iron-sulfur centers are present, lowering the number of iron per tungsten). The EPR integration of the signals due to the iron sulfur centers performed on a partially reduced sample with formate and on a sample reduced with excess of dithionite, matches with the number determined by metal analysis. The ratio A_{410}/A_{280} is 0.14, which is a very low value

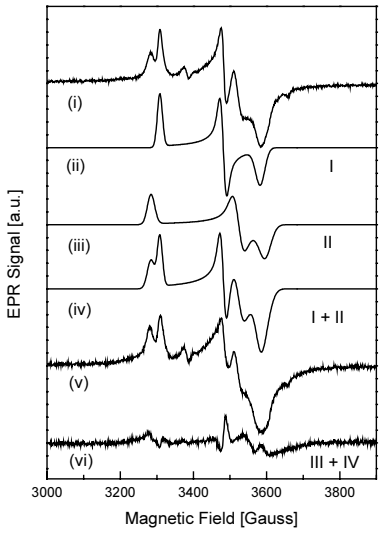

Figure 4.4 EPR spectra of a partially formate reduced state of *D. gigas* FDH at 20K (i) and at 10K (v), together with the spectral simulation of signal I (ii), signal II (iii), sum of signals I and II in a ratio 1:1 (iv). Difference spectrum between 20K and 10K (vi). Experimental conditions were: (i) temperature, 20 K; microwave power, 0.2 mW; modulation amplitude, 4 G_{pp}; microwave frequency, 9.49 GHz; (ii) as in (i) except temperature, 10 K; and microwave power, 0.2 mW. Simulations were performed with the parameters: [g_1= 1.892 (30), g_2= 1.947 (15), g_3= 2.049 (14)] (ii), and [g_1= 1.885 (32), g_2= 1.924 (26), g_3= 2.064 (20)] (iii) (line width between parentheses). Signal (i) was scaled to the experimental conditions of signal (v) in order to obtain signal (vi).

for a protein of 121 kDa with 4x[4Fe-4S] centers. For instance, the [NiFe] hydrogenase from *D. gigas* has a ratio A_{400}/A_{280}= 0.28 [26] and the proportion of iron-sulfur centers per polypeptide is similar (3 iron-sulfur centers per 90kDa).

We reanalyzed the EPR spectra according to the number of iron-sulfur centers. A selection of the most representative EPR spectra is shown in Fig. 4.4. We determined the iron concentration of the sample used in the EPR studies (1024 μM). If we assume that all the iron atoms are arranged as [4Fe-4S] centers, as observed by X-ray crystallography, each will have a concentration of 64 μM. In our previous EPR studies, four types of EPR signals were observed. Signals I and II were attributed to each of the [4Fe-4S] centers and signals III and IV were proposed to be two forms of tungsten or tungsten and pterin. Due to the low amount of iron/protein detected, our previous analysis led to the conclusion of the presence of only two [4Fe-4S] centers per protein, which correspond to signals I and II [1].

To elucidate whether all iron-sulfur centers are detectable by EPR, samples were reduced either with different amounts of formate or excess of dithionite. We learned by analyzing these samples that signals III and IV developed later at more negative redox potentials. Therefore, samples containing mainly signals I and II could be generated. Analysis of one of these EPR samples reduced with formate lead to a total spin concentration of 230 μM at 10 K (3.6 spin per holoprotein) and 211 μM at 20 K (3.3 spin per holoprotein), and that signals I and II are in ratio 1:1. In this sample the amount of signals III and IV are very weak and can be observable in the difference spectrum shown in Fig. 4.4. A similar spin concentration and ratio for signals I and II was obtained in the dithionite-reduced sample. Although this sample shows a higher number of spins (4.2 spin per protein), this is due to the increase of the intensity of signals III and IV (not shown). Therefore, we can conclude that in the formate-reduced sample at 20 K (Figure 4.4), when signals III and IV are negligible, we have the contribution of all the four iron-sulfur centers. This fact and the ratio 1:1 between signals I and II, suggest that signals I and II come from pairs of iron-sulfur clusters.

4.4 Conclusions

Although the crystallographic work is still in progress, the current state of the analysis has allowed to unambiguously identify two new [4Fe-4S] clusters and to identify a Se-cysteine residue. The data show that the large subunit from W-FDH contains the catalytic center with one W atom coordinated to two MGD cofactors and to a Se-Cys. Additional putative oxygen ligands cannot be identified at this stage of the analysis. One of the four [4Fe-4S] clusters is present in the same subunit and similarly placed when compared to the structures of *D. desulfuricans* NAP or *E. coli* FDH-H. The three [4Fe-4S] clusters are aligned in the small subunit *ca.* 10 Å away from each other, creating a feasible electron transfer pathway which connects the exterior of the protein to the [4Fe-4S] cluster present in the large subunit and close to the interface. The approximate distances are depicted on Figure 4.3b. The arrangement of the four [4Fe-4S] clusters has an obvious function as an electron transfer chain as found in many different biological systems [23]. It is for instance reminiscent of the pathway found in the Ni-Fe hydrogenase structure although the last three Fe/S clusters are almost linear in the hydrogenases, where the angle between the center of mass of the [4Fe-4S] clusters is 162° in *D. fructosovorans* Ni-Fe hydrogenase (pdb 1FRF, [24]), 157° in *Desulfomicrobium baculatum* Ni-Fe-Se hydrogenase (pdb 1CC1, [25]) as opposed to 112° in W-FDH. A more detailed structural comparison will have to wait until the refined crystal structure is available.

Acknowledgements

Work supported by EC-TMR/FMRX-CT980204, PhD grant PRAXISXXI/BD/13530/97 (JMD). Sergey Bursakov is acknowledged for support with the protein purification and Alice Pereira and Pedro Tavares for helpful discussions. We thank the EMBL Grenoble Outstation, beam line BM-14 support for MAD measurements at the ESRF under the European Union TMR/LSF Programme. Hans Bartunik and Gleb Bourenkov are acknowledged for help at the BW6 beam line of the MPG-ASMB in DESY, Hamburg, for collecting the first FDH native data set.

References

1. Almendra MJ, Brondino CD, Gavel O, Pereira AS, Tavares P, Bursakov S, Duarte R, Caldeira J, Moura JJG, Moura I (1999) *Biochemistry* **38**: 16366-16372
2. Hille R (1996) *Chem Rev* **96**: 2757-2816
3. Romão MJ, Knäblein J, Huber R, Moura JJG (1997) *Prog Biophys Mol Biol* **68**: 121-144
4. Schneider F, Löwe J, Huber R, Schindelin H, Kisker C, Knäblein J (1996) *J Mol Biol* **263**: 53-69
5. Schindelin H, Kisker C, Hilton J, Rajagopalan KV, Rees DC (1996) *Science* **272**: 1615-1621
6. Dias JM, Than ME, Humm A, Huber, R, Bourenkov, GP, Bartunik, HD, Bursakov, S, Calvete J, Caldeira J, Carneiro C, Moura JJG, Moura I, Romão MJ (1999) *Structure* **7**: 65-79
7. Boyington JC, Gladyshev VN, Khangulov SV, Stadtman TC, Sun PD (1997) *Science* **275**: 1305-1308
8. Abaibou H, Pommier J, Giordano G, Mandrand-Berthelot MA, (1995) *J Bacteriol* **177**: 7141-7149
9. Berg BL, Li J, Heider J, Stewart V (1991) *J Biol Chem* **266**: 22380-22385
10. Rossi M, Pollock WBR, Reij MW, Keon RG, Fu R, Voordouw G (1993) *J Bacteriol* **175**: 4699-4711

11. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J-F, Dougherty BA, Merrick JM, McKenney K, Sutton G, Fitzhugh W, Fields CA, Gocayne JD, Scott JD, Shirley R, Liu L-I, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC (1995) *Science* **269**: 496-512
12. Menon NK, Chatelus CY, Dervartanian M, Wendt JC, Shanmugam KT, Peck HD Jr, Przybyla AE (1994) *J Bacteriol* **176**: 4416-4423
13. Bairoch A, Apweiler R, The SWISS-PROT protein sequence database and its supplement TrEMBL (2000) *Nucleic Acids Res* **28**: 45-48
14. Otwinowski Z, Minor W (1997) *Methods in Enzymology* **276**: 307-326
15. Mathews BW (1968) *J Mol Biol* **33**: 491-497
16. Navaza J (1994) *Acta Crystallogr* **A50**: 157-163
17. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr* **D50**: 760-763
18. De La Fortelle E, Bricogne G (1997) *Methods in Enzymology* **276**: 472-494 New York: Academic Press
19. Cowtan K (1994), CCP4 and ESF-EACBM *Newsletter on Protein Crystallography* **31**: 34-38
20. Sasaki S (1989) *KEK report* 88-14, National Laboratory for high energy physics, Tsukuba, Japan
21. Perrakis A, Sixma TK, Wilson KS, Lamzin VS (1997) *Acta Crystallogr* **D53**: 448-455
22. Brünger AT (1992) X-PLOR, A system for X-Ray Crystallography and NMR, Yale University Press, New Haven
23. Beinert H (2000) *JBIC* **5**: 2-15
24. Rousset M, Montet Y, Guigliarelli B, Forget N, Asso M, Bertrand P, Fontecilla-Camps JC, Hatchikian EC (1998) *Proc Natl Acad Sci USA* **95**: 11625-11630
25. Garcin E, Vernede X, Hatchikian EC, Volbeda A, Frey M, Fontecilla-Camps JC (1999) *Structure Fold Des* **7**: 557-566
26. LeGall J, Ljungdahl PO, Moura I, Peck HD Jr, Xavier AV, Moura JJG, Teixeira M, Huynh BH, DerVartanian DV (1982) *BBRC* **106**: 610-616
27. Merritt E *http://brie.bmsc.washington.edu/scatter/AS_periodic.htm*
28. Brennan S, Cowan P *(1992) Rev Sci Instrum* **63**: 850
29. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M *(1991) Acta Crystallogr* **A47***: 110-119*
30. Moulis JM, Sieker LC, Wilson KS, Dauter Z (1996) *Protein Sci* **5**: 1765-1775
31. Fischer DS, Price DC (1964) *Clin. Chem*. **10**: 21-31

5

Summary and conclusions

Many biological processes require metal ions, and many of these metal-ion functions involve metalloproteins. The metal ions in metalloproteins are often critical to the protein's function, structure, or stability. This thesis focuses on two of these proteins, bacteriophage T4 endonuclease VII (EndoVII) and *D. gigas* formate dehydrogenase, which are studied by X-ray crystallography. The structure of EndoVII reveals how a magnesium or calcium ion is used to cleave several kinds of irregular but flexible DNA, while a zinc ion maintains the structural integrity of this DNase.

The formate dehydrogenase contains a tungsten ion and a seleno-cysteine at the active site, that catalyses the oxidation of formate to carbon dioxide. The two released electrons are transferred through four [4Fe-4S] clusters before they can be handed over to another protein. Two of the [4Fe-4S] and the selenium have been overlooked by other techniques, but could be localised and identified by crystallography.

Chapter 1 gives a general introduction on metals in biological systems, X-ray crystallography and also describes the biological background of both proteins.

Chapter 2 presents the structure of the four-way DNA-junction resolving enzyme T4 endonuclease VII, and that of the inactive N62D mutant. The better-expressed mutant was solved first, using seleno-methionine, mercury and gold derivatives. These mercury and gold derivatives bind to the sulphurs that also ligand the zinc. The wild-type was solved with help of a single mercury derivative since molecular replacement with the mutant structure failed.

On its own, the EndoVII monomer would not represent a stable fold, as it exposes many hydrophobic residues to the solvent. But two monomers intertwine to form a dimer without this problem. In this dimer, the monomers are aligned head-to-tail; the N-terminus of one monomer interacts with the C-terminus of the other monomer and *vice versa.* The major dimerization element, unique to EndoVII, is the "four-helix-cross" domain, which consists of helix-2 and helix-3 from each monomer. It contains an extended hydrophobic core.

Another feature is the "beta-finger", residues 38-56. Its stability depends critically on the zinc. This zinc ion is tetrahedrally co-ordinated to four cysteines, linking helix-1 through residues C23 and C26 firmly to the N-terminal part of helix-2 (C58, C61). Indeed, interfering mutations inactivate the protein. Finally, the calcium ion, which marks the active site, is liganded to aspartate-40 and asparagine-62. Mutation studies show that these amino acids are essential for activity: The N62D mutant is completely inactive.

The EndoVII structure has been docked to a "stacked-X" four-way DNA junction, one of its many substrates. This model is not refined, since both the DNA and the protein are known to be flexible and might undergo conformational changes. However, its overall features confirm experimental data: I) The EndoVII dimer binds to the minor groove side of the four-way junction; II) Basic residues on helix-2 can interact with phosphates on the exchanging strands and those on the C-terminal domain can interact with phosphates in the continuous strands, consistent with observed foot-printing patterns; III) The C-terminus binds up to nine base pairs away from the junction, confirming the minimal length of two arms of the substrate; IV). The active sites do not cleave both the scissile phosphates simultaneously.

Surprisingly, the N62D mutant shows a major rearrangement in the "four-helix-cross" domain, when compared to the wild-type: helices-2 are translated by half a turn each, in opposite direction and the opening of the "bays", between each helix-2 and beta-finger, is wider. These differences might be attributed to the point-mutation, which introduces an extra charge in the active site, to differences in crystallisation conditions, to the different pH employed, to crystal contacts or perhaps they are simply a sign of the intrinsic flexibility of EndoVII.

This dilemma is partly solved in **chapter 3**, which presents the crystal structure of wild-type EndoVII in a different space group, which contains less solvent. It crystallised in the same drop, so that differences observed between the two wild-type structures cannot be attributed to the mutation, pH or salt concentrations. Since the helical-cross region of this second structure is very similar to that of the mutant, rearrangements in this region must be seen as a consequence of intrinsic flexibility of EndoVII. The widening of the "bays", however, might still be a consequence of the mutation, different pH, absence of Ca^{2+} or crystal packing. An investigation of the flexibility of EndoVII with TLS- refinement, *i.e.* anisotropic refinement of rigid bodies, provides only limited insight. However, it confirms that rotations along the axes of the helices 2 and 4 and along the beta-finger are a main source of flexibility and also that the C-terminus, helix-4, 5 and 6, behave as a rigid body.

The high-resolution structure of the N62D mutant brings more clarity towards the reaction mechanism of the nuclease. This model contains important water molecules and reveals the position and orientation of 14 sulphate ions, which may indicate favoured phosphate (DNA) binding sites. Supported by new mutation data (Birkenbihl, unpublished), these sulphate and water positions, combined with the Ca^{2+} positions in the wild-type structures, suggest a reaction-mechanism similar to those proposed for some other magnesium dependent nucleases.

Asparagine-62, glutamate-65 and aspartate-40 are important to position Mg^{2+} or Ca^{2+} next to the scissile phosphate of the DNA substrate. Histidine-41 activates a water molecule, which in turn executes a nucleophilic attack on the phosphor atom. Histidine-43 stabilises this phosphate directly through a hydrogen bond. Unfortunately it is still unclear why the N62D shows no DNase activity at all; an aspartate would also be able to ligand/position a divalent cation. The extra charge that this mutation introduces in the active site might distort the geometry of the active site, and repel the DNA. A more attractive, albeit more speculative hypothesis, assumes that the amino group of asparagine-62 donates a hydrogen-bond to the phosphate, which would also stabilise the transition state.

At present, there are no known proteins with significant sequence homology to EndoVII, though nucleases with structural similarities do exist. One group consists of magnesium-dependent nucleases, which have a similar geometry of liganding side-chains around the magnesium (or calcium) ion in the active site; e.g. the *E. coli* proteins RuvC (Ariyosi *et al.*, 1994) and RNase H (Katayanagi *et al.*, 1990). However, these nucleases have no resembling fold. Most likely, this just shows that magnesium-dependent nucleases need a certain geometry to function.

A more interesting group shares a folding motif similar to the beta finger and helix-2: *Serratia* Nuclease (Miller *et al.*, 1994), PpoI (Flick *et al.*, 1998) and perhaps even Colicin E9 (Kleanthous *et al.*, 1999). Asparagine-62 and histidine-41 are conserved between *Serratia* nuclease, PpoI and T4 endonuclease VII. PpoI has also been crystallised in complex with DNA. If one superimposes this with the EndoVII structure, it turns out that the Ca^{2+} in EndoVII is buried deeper within the protein, but small rotations (10-20 degrees) along helix 2 and the beta-finger suffice to superimpose them. These two nucleases act on different substrates, and maybe the larger DNA junctions of EndoVII need a wider and deeper binding groove than the double stranded DNA of PpoI. However, it could also be the source of EndoVII's specificity; flexible DNA might impose this conformational change of EndoVII upon binding, readying the enzyme for cleavage, while the magnesium or calcium ion might be too far away if EndoVII approaches more rigid DNA. A structure of EndoVII in complex with DNA would solve these questions.

Chapter 4 presents the major part of the determination of the 3D structure of the tungsten-containing formate dehydrogenase (W-FDH) from *Desulfovibrio gigas*, one of the first tungsten-containing enzymes isolated from a mesophile. The large subunit (92 kDa) is structurally related to several tungsten- and molybdenum-containing enzymes and X-ray structures have been determined for two of them. One of these, the periplasmic nitrate reductase (Dias *et al*, 1999), could be used to obtain a molecular replacement solution. But the quality of phasing was not sufficient to generate a clear, interpretable electron density map. Furthermore, the amino acid sequence of W-FDH has not yet been determined, what makes model building complicated. Multiple wavelength diffraction (MAD) measurements were undertaken at the absorption edges of W and Fe to define unambiguously the number, positions and identity of these anomalous scatterers and to improve the X-ray phases. The MAD-analysis revealed one W-atom with a Se-cys ligand and one [4Fe-4S] cluster bound to the large subunit, and three [4Fe-4S] clusters in the small subunit. The four [4Fe-4S] clusters are *ca.* 10 Å apart, creating a feasible electron transfer pathway, which connects the exterior of the protein to the W/Se site in the large subunit. Two of the four iron-sulphur clusters had not been predicted before by spectroscopic techniques (Almendra *et al*., 1999). A reinvestigation of the spectroscopic data was performed, but gave the same results as before. If these data were correct, this means that the [4Fe-4S] clusters are instable, and that only protein with fully occupied clusters crystallises.

The formate dehydrogenase H (FDH-H) from *E. coli* catalyses the same reaction as W-FDH, but uses a molybdenum instead of tungsten. Both are liganded to two molybdopterin-cofactors and to a seleno-cysteine, so the question remains why W-FDH prefers tungsten to the more common molybdenum. The full structure will

allow a comparison of the two enzymes in atomic detail, and perhaps, it will shed some light on this phenomenon.

X-ray crystallography has been used to characterise the nature of metal-centres in proteins, their coordination geometry and even their identity. Sometimes, the way metal ions are bound to the protein already clarifies its role in the protein. In other cases it has to be supplemented with other studies before the role can be fully understood. Either way, crystallography provides a powerful tool for the study of metalloproteins.

References

Almendra, M.J., Brondino, C.D., Gavel, O., Pereira, A.S., Tavares, P., Bursakov, S., Duarte, R., Caldeira, J., Moura, J.J.G., Moura, I. (1999) *Biochemistry,* **38,** 16366-16372

Ariyosi, M., Vassylyev, D., Iwasaki, H., Shinagawa, H. and Morikawa, K. (1994) *Cell*, **78**, 1063-1072.

Flick, K.E., Jurica, M.S., Monnart Jr, R.J. and Stoddard, B.L. (1998), *Nature*, **394**, 96-101.

Katayanagi, M., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, M. and Morikawa, K. (1990) *Nature*, **347**, 306-309.

Kleanthous, C., Kühlmann, U.C., Pommer, A.J., Ferguson, N., Radford, S.E., Moore, G.R., James, R. and Hemmings, A.M. (1999) *Nature Struct. Biol.*, **6**, 243-252.

Miller, M.D., Tanner, J., Alpaugh, M., Benedik, M.J. and Krause, K.L. (1994) *Nature Struct. Biol.*, **1**, 461-468.

Samenvatting

Bijna alle biologische processen in levende wezens maken gebruik van chemische reacties. Op zichzelf verlopen deze reacties echter erg langzaam en ongecontroleerd. Daarom gebruiken levende wezens enzymen. Dit zijn eiwitten die bepaalde reacties kunnen versnellen. Ze bestaan uit lange ketens aminozuren, die driedimensionaal gevouwen zijn. Er zijn 20 verschillende aminozuren waarmee, door variatie van samenstelling en positie, een groot aantal verschillende enzymen gemaakt kunnen worden. Naast aminozuren komen soms ook andere moleculen (o.a. "vitamines") en metaalionen ("mineralen") voor in eiwitten. Zo geven ijzer-ionen in hemoglobine een rode kleur aan bloed. Ook andere metalen, zoals zink, calcium, koper en nikkel, zijn belangrijk voor allerlei chemische processen. Vaak blijken deze metaal direct betrokken te zijn bij de katalyse, en soms stabiliseren ze de driedimensionale structuur van het eiwit.

Twee van zulke eiwitten worden beschreven in deze thesis: T4 endonuclease VII (EndoVII) en *Desulfovibrio gigas* formiaat dehydrogenase (W-FDH). EndoVII gebruikt een magnesium (of calcium) ion voor het knippen van bepaalde soorten afwijkend DNA, terwijl een zink ion de driedimensionale structuur van het eiwit handhaaft.

De formiaat dehydrogenase voorziet de bacterie *D. gigas* van energie door mierenzuur om te zetten in kooldioxide. Hierbij komen twee electronen vrij, die vervolgens, via vier ijzer-zwavel clusters, doorstromen naar een ander eiwit, een cytochroom-C, en uiteindelijk naar een sulfaat ion. Bijzonder is dat dit enzym een wolfraam ion gebruikt om deze reactie te versnellen.

De inleiding in **Hoofdstuk 1** bevat een algemene introductie over metalen in biologische systemen en een beknopte inleiding over Röntgen diffractie. Daarna wordt de biologische rol van beide eiwitten aangegeven.

In **hoofdstuk 2** wordt de structuur van het enzym EndoVII beschreven. EndoVII werd, m.b.v. recombinant DNA technieken, geproduceerd in *E. coli* bacterien. Omdat EndoVII giftig is voor *E. coli* (Het knipt DNA!) werd eerst met een mutant gewerkt waarin een van de 157 aminozuren is gewijzigd. Na deze minimale ingreep, slechts een van de 3000 atomen is anders, knipt het geen DNA meer. Later werd ook de structuur van het natuurlijke enzym bepaald.

EndoVII blijkt als een dimeer (twee dezelfde eiwitten aan elkaar) voor te komen, waarbij het ene eiwit het andere stabiliseert. Beide eiwitten zitten zodanig dooreen gevlochten dat de "kop" van de ene, bij de "staart" van de andere ligt. De Helices-2 en 3 van beide delen vormen een kruisvormig geheel dat beide delen bij

elkaar houdt. (figuur 2.1 en 2.2) De beide helices-2 liggen naast elkaar en beide helices-3 liggen dwars op de helices-2, ook naast elkaar. Zulke binding is nog niet eerder gevonden in eiwitten.

Een ander kenmerk vormt de "*beta-finger*", een smalle lus van 19 aminozuren (N° 38 tot 56) waarvan er enkele direct betrokken zijn bij het knippen van DNA. Deze lus wordt in vorm gehouden door een zink ion, dat tegelijkertijd helix-1 en 2 aan elkaar bindt.

Het calcium ion tenslotte, markeert het actieve centrum van EndoVII, de plaats waar het DNA wordt geknipt. Het is gebonden aan het 40e en 62e aminozuur van EndoVII, aspartaat-40 en asparagine-62. Mutaties van deze aminuzuren verhinderen het knippen van DNA, maar niet het herkennen (binden) van DNA. In de N62D mutant is de asparagine gewijzigd in een aspartaat.

Een van de belangrijkste substraten van EndoVII is een kruispunt van DNA. Dat is erg flexibel en kan drie verschillende vormen aannemen, een daarvan wordt door EndoVII herkent. De structuur van dit kruispunt DNA was al bekend en blijkt maar op een manier te passen op de EndoVII structuur. (Maar EndoVII en het DNA zullen zich moeten aanpassen om precies te passen.) Het model van EndoVII, gebonden aan DNA, verklaart een heleboel experimentele resultaten; hoe het gebonden is, waar het knipt, en hoelang het daar tenminste over doet.

Nadat de structuur van de mutant opgelost was, werd ook de structuur van het natuurlijke enzym bepaald. De verschillen zijn ronduit verbazend: De twee robuust ogende helices-*nummer-2* (fig. 2.3c) zijn verschoven t.o.v. elkaar. Bovendien zijn de twee inhammen veel meer samengedrukt. Het is aantrekkelijk aan te nemen dat EndoVII net zo flexibel is als het substraat DNA. Maar de verschillen kunnen ook anders verklaard worden. De mutatie introduceert een extra, geladen groep in het eiwit, die vanalles kan verstoren. En de zuurgraad van de kristallen, de aanwezige mineralen en de manier waarop het eiwit in het kristal gestapeld is, zijn verschillend voor beide kristallen.

Dit dilemma wordt verhelderd in **hoofdstuk 3**, waar een tweede kristalvorm van het *natuurlijke* EndoVII wordt beschreven. De helices-2 zitten hier precies als in de mutant. De kristallisatie-mix was precies gelijk aan die van de eerste kristalvorm, dus moeten verschillen tussen de twee vormen een gevolg zijn van de plooibaarheid van het eiwit. De breedte van beide inhammen is echter vergelijkbaar, en kan nog steeds een gevolg zijn van de mutatie, de zuurgraad of van de aanwezige mineralen.

Het reactiemechanisme van EndoVII werd duidelijker door een nieuwe meting aan de N62D mutant. Het nieuwe model is gedetailleerder, bevat belangrijke water moleculen en toont ook duidelijk de positie en orientatie van 14 sulfaat moleculen. Deze laatsten kunnen in principe bindingsplaatsen van de fosfaat groepen van het DNA aangeven. Aangevuld met gegevens van mutatie studies, en de calcium posities in het natuurlijk enzym, kan hiermee een reactiemechanisme worden voorgesteld, dat lijkt op dat van andere nucleases. (Figuur 3.6)

Asparagine -62, glutamaat-65 en aspartaat-40 houden het magnesium-ion vast naast de te knippen fosfaat (van het DNA) om de lading te compenseren. Histidine-41 activeert een water molecuul, dat vervolgens als een nucleofiel aanvalt op het fosfor atoom. Histidine-43 stabiliseert dit fosfaat direct middels een waterstof brug. Helaas is het nog onduidelijk waarom de N62D mutant geheel inactief is: Een aspartaat kan ook een magnesium ion binden. Wel introduceert de mutatie een extra

lading in het actieve centrum, dat DNA kan afstoten, of de locale geometrie kan verstoren. Aantrekkelijk doch onbewezen is de veronderstelling dat de amino groep van asparagine-62 een waterstof brug slaat naar de fosfaat, wat de reactie vergemakkelijkt.

Hoewel er geen eiwitten bekend zijn met een vergelijkbare aminozuurvolgorde als die van EndoVII, zijn er wel degelijk nucleases met structurele overeenkomsten. Een groep bestaat uit nucleases met een magnesium ion en bijbehorende aminozuur-zijketens, zoals in de *E. coli* eiwitten RuvC en RNase H. De overeenkomsten zijn echter alleen lokaal, rond het actieve centrum. Waarschijnlijk betekent dit slechts dat er een zekere geometrie vereist is voor de functie van Mg^{2+} afhankelijke nucleases .

Interessanter is de groep die de structuur van de "beta finger" en helix-2 deelt: *Serratia* Nuclease, PpoI en misschien Colicin E9. Asparagine-62 and histidine-41 uit EndoVII komen qua locatie overeen met dezelfde aminozuren in *Serratia* nuclease en PpoI. In de kristalstructuur van PpoI is deze gebonden aan DNA. Vergeleken met EndoVII blijft het metaal ion meer aan de oppervlakte, maar kleine rotaties om de lengte-as van helix-2 en de "beta-finger" (10-20 graden) zijn al voldoende om beiden op dezelfde plaats te krijgen. Deze twee nucleases knippen verschillend substraat, en misschien vereisen de grotere DNA-kruispunten van EndoVII een bredere en diepere bindingsplaats. Maar dit zou ook de bron van EndoVII's specificiteit kunnen zijn: Gebonden, buigzaam DNA kan een structuur-verandering veroorzaken in EndoVII zodat het DNA kan knippen, terwijl het magnesium ion te ver weg blijft indien normaal of stijf DNA wordt gebonden.

Een structuur van EndoVII, in complex met DNA, zou dit kunnen bevestigen.

Hoofdstuk 4 beschrijft een belangrijk deel van de structuurbepaling van *Desulfovibrio gigas* Formate Dehydrogenase (W-FDH). Het bijzondere van dit enzym is dat het wolfraam bevat i.p.v. molybdeen. Dit enzym bestaat uit twee delen, 92 en 29 kDa groot. De 92-kDa eenheid behoort tot een familie van wolfraam- en molybdeen- bevattende eiwitten. Van twee leden hiervan was de 3D-structuur reeds bekend. Een hiervan gebruikt voor de structuurbepaling van W-FDH. Helaas is de kwaliteit van de gevonden oplossing onvoldoende, mede doordat de aminozuur-volgorde van het eiwit nog niet bekend was. Nieuwe metingen van de diffractie, met Röntgen straling van vier specifieke golflengtes (MAD), zijn gebruikt om het aantal, de plaats en de identiteit van de metalen in het eiwit te bepalen. De analyse toont aan dat het wolfraam atoom gebonden is aan een selenium atoom, dat zich, samen met een ijzer-zwavel cluster [4Fe-4S] in de 92-kDa eenheid bevind. De 29-kDa eenheid bevat nog drie [4Fe-4S] clusters. De vier [4Fe-4S] clusters bevinden zich op 1 nm afstand van elkaar; ideaal om electronen van het diepe wolfraam/selenium centrum naar de oppervlakte te geleiden.

Slechts twee van de vier ijzer-zwavel clusters werden al met andere technieken (uv-vis, Mössbauer, EPR, icp metaal analyse) voorspeld. De betreffende metingen zijn vervolgens opnieuw geinterpreteerd, en het ijzer gehalte opnieuw bepaald, maar gaven opnieuw hetzelfde resultaat. Indien de eerdere metingen correct waren, betekent dat dat [4Fe-4S] clusters in W-FDH instabiel zijn, terwijl alleen het enzym met alle clusters kristalliseert.

Het formiaat dehydrogenase H (FDH-H) uit *E. coli* katalyseert dezelfde reactie als W-FDH, maar maakt gebruik van molybdeen i.p.v. wolfraam. Beide zijn geligeerd

aan twee "molybdopterin" cofactoren, en aan een selenocysteine. Het blijft de vraag waarom W-FDH gebruikt maakt van wolfraam i.p.v het meer gebruikelijke en meer beschikbare molybdeen. De volledige structuur van W-FDH zal het mogelijk maken deze twee enzymen op atomair niveau te vergelijken, hopelijk voldoende om dit fenomeen te verklaren.

In deze thesis werd de aard van metaal-centra in eiwitten, hun coördinatie geometrie en zelfs de identiteit, gekarakteriseerd m.b.v. Röntgen diffractie. Soms blijkt de rol van de metaal-ionen al uit de manier waarop ze aan eiwitten gebonden zijn. In andere gevallen moet kristallografie aangevuld worden met andere technieken voordat de rol volledig begrepen kan worden. Hoe dan ook, kristallografie blijft een krachtig hulpmiddel voor onderzoek aan metallo-eiwitten.

Curriculum Vitae

Hans Cornelis Andreas Raaijmakers werd geboren op 13 februari 1971 te Eindhoven. Hij behaalde het ongedeeld VWO diploma aan het Eckart College te Eindhoven in 1989. Aansluitend startte hij met de studie Moleculaire Wetenschappen aan de Landbouw Universiteit te Wageningen. Tijdens de doctoraalfase verrichtte hij onderzoek bij de Vakgroep Biochemie aan de Landbouw Universiteit te Wageningen (Dr. H. Haaker) en bij het Krebs Instituut in Sheffield (Verenigd Koninkrijk) (Dr. G. Ford). In augustus 1995 behaalde hij zijn doctoraal diploma. In oktober 1995 startte hij een promotie onderzoek bij het Europees Moleculair Biologisch Laboratorium in Heidelberg (Duitsland) onder begeleiding van Dr. D. Suck en Prof. Dr. C. Laane. De resultaten van dit promotieonderzoek zijn beschreven in hoofdstuk 2 en 3 van dit proefschrift. Vanaf juni 1999 is hij werkzaam als onderzoeker in het laboratorium van Prof. Dr. Maria João Romão in Caparica, Portugal. De eerste resultaten zijn beschreven in hoofdstuk vier van dit proefschrift.

Hans C.A. Raaijmakers - 'Structural studies on metal-containing enzymes:
T4 endonuclease VII and *D. gigas* formate dehydrogenase.'
Dutch: 'Structureel onderzoek naar metaal-bevattende enzymen.'
Thesis Wageningen University - With summary in Dutch

ISBN 90-5808-412-4