Biophysical studies on Tn3 resolvase

Marcelo Nöllmann
Division of Molecular Genetics, IBLS
University of Glasgow

A thesis submitted for the degree of
Doctor of Philosophy

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

A mis viejos, Petra y Carlos.

Abstract

Site-specific recombination is a process employed by organisms in order to perform spatially and temporally defined rearrangement of DNA molecules, such as phage integration and excision, resolution of circular multimers, inversions for expression of alternate genes, and assembly of genes during development. Tn3 is the prototype of a family of closely related mobile genetic elements referred to as the class II or Tn3 family of transposons. Tn3 contains the genes encoding a transposase, β-lactamase, and a site-specific recombinase, Tn3 resolvase (Tn3R), that is responsible for the resolution of the cointegrate, an intermediate in the transposition reaction. Tn3R is able to resolve *in vitro* supercoiled plasmids containing two 114 bp *res* sites in direct orientation into two smaller circular plasmids, each of them with a single *res* site.

In this thesis, the solution properties of Tn3R were studied by sedimentation equilibrium (SE) and velocity (SV) analytical ultracentrifugation and small angle neutron scattering (SANS). Tn3R was found to be in a monomer-dimer self-association equilibrium, with a dissociation constant of $k_D^{1-2} = 50$ μM. SV and SANS demonstrated that the low-resolution conformation of dimeric Tn3R in solution is similar to that of $\gamma\delta$ resolvase in the co-crystal structure by Yang and Steitz (1995), but with the DNA-binding domains in a rather extended conformation. In addition, equilibrium binding of Tn3R to the individual binding sites in *res* (sites I, II and III) was investigated by employing fluorescence anisotropy (FA) measurements. This revealed that site IIL (site II left end) and site III have the highest affinity for Tn3R, followed by site I, and finally by site IIR (site II right end).

The specificity of binding of Tn3R for non-specific DNA was assayed by competition experiments, where it was shown that the affinity of binding of Tn3R to site I is 1000 times higher than to non-specific sites.

A new approach, involving a combination of rigid-body and *ab initio* modelling was developed for the study of the solution structure of macromolecules. At first, this approach was tested by applying it to the reconstruction of the low-resolution solution conformation of a DNA Holliday junction, based on small angle x-ray scattering and sedimentation velocity data. The scattering data were analysed in two independent ways: firstly, by rigid body modelling using previously suggested models for the Holliday junction (HJ), and secondly, by *ab initio* reconstruction methods. Sedimentation coefficients calculated for the models generated by both methods agreed with those determined experimentally and were compatible with the results of previous studies using different techniques, but provided a more direct and accurate determination of the solution conformation of the HJ. These results confirmed that addition of Mg^{2+} alters the conformation of the HJ from an extended to a stacked arrangement.

The solution conformation of a stable protein-DNA complex formed by a mutant of Tn3R and DNA was studied by a similar approach. Hyperactive mutants of resolvase form a complex (X-synapse) containing two site I DNA fragments and a resolvase tetramer. The low-resolution solution structure of the purified, catalytically competent X-synapse was solved from small angle neutron and x-ray scattering data, by fitting the models constructed by rigid-body transformations of a published crystallographic structure of a resolvase dimer bound to site I to the data. This analysis revealed that the two site I fragments are on the outside of a resolvase tetramer core, and provided some information on the quaternary structure of the tetramer.

Finally, the rigid-body modelling method was redesigned into a general systematic approach to retrieve the conformation of a macro-

molecule that simultaneously agrees with a range of experimental solution properties. In this method, generalised rigid-body modelling was combined with a Monte Carlo/simulated annealing optimisation method to search over a large range of possible conformations for the structure that best fits solution experimental properties derived from small angle scattering, fluorescence resonance energy transfer, and analytical ultracentrifugation datasets. This improved methodology was evaluated by applying it to two bulged DNA fragments with very different solution conformations.

Acknowledgements

Quiero agradecer sobre todo a mis viejos, sin cuyo apoyo durante estos últimos tiempos, y durante tantos años de sacrificios y esfuerzos que pusieron en mi educación. A las chicas, riel y sol, por ser quienes son, y soportarme a pesar de todo. Ojalá pudiéramos pasar más tiempo juntos.

Je veux te remercie, Marie, pour tous les temps bienheureux qu'on a déjà passé ensemble, dans tous ces voyages en Europe et en Amèrique, et tous ces endroits qu'on connaît sans connaître. Aussi, pour partager ta vie avec moi, chaque jour, de cette façon si intensive et passionnante.

Je veux aussi te remerci, Vincent, pour ton soutien, tes conseils et tes naïves petites conneries. Enfin, pour ton sincère amitié pendent ces années.

I should like to express heartfelt thanks to Dr Marshall Stark, for his continuing guidance, enthusiasm, and inspiration throughout the course of this project. It has been a pleasure working with you, Marshall. I would also like to thank Dr Olwyn Byron, for her enthusiasm, positiveness and advice. I really appreciated the great freedom you both gave me in choosing research projects and pursuing new directions.

I want to especially thank Sean Colloms and Martin Boocock for useful discussions and invaluable advice on many aspects of this work, and for critical reading of manuscripts.

I am really thankful to Mary Burke, for being always so nice and patient throughout the years, for always having time for a little question or a chat. We will miss you in the 6^{th} floor Mary! Also, I am really grateful to all the people in the 6^{th} floor past and present, Sally, Jie, Amy, Femi and Aram, for their patience and help with so many things throughout these years. Thanks to the secretaries of both I&I and Genetics and the prep room ladies for provisions of all kinds.

Thanks to Mischa and Alex, for the nice times and noisy movies in those long, endless nights spent collecting data.

I should also thank my assesor, Prof. Richard Cogdell, for his always useful and advice-full meetings, and his impartial assessment of the research in this project. Also, for his advices career-wise.

I particularly thank Günter Grossman of the Synchrotron Radiation Source, Warrington, UK, for his invaluable advice and kindness. I would also like

to acknowledge the help and guidance of Dmitri Svergun at beamline X33 of the European Molecular Biology Laboratory/Deutsches Elektronen Synchroton, Hamburg, Germany; Peter Timmins at beamline D11 of the Institut Laue-Langevin, France; Florian Meneau and Wim Bras at beamline BM26B of the European Synchrotron Radiation Facility, Grenoble, France; and Richard Heenan at beamline LOQ of the Rutherford Laboratory, ISIS, UK.

Abbreviations

$\bar{v}$	Partial specific volume
$\gamma\delta$R	$\gamma\delta$ resolvase
$p(r)$	Pair distance distribution function
AUC	Analytical ultracentrifugation
bp	Base pair
CV	Column volumes
D_{MAX}	Maximum intraparticle distance
FA	Fluorescence anisotropy
FRET	Fluorescence resonance energy transfer
GFP	Green fluorescent protein
HJ	Holliday junction or 4-way DNA junction
HPLC	High pressure liquid chromatography
krpm	10^3 rpm
MC/SA	Monte Carlo simulated annealing
NM-resolvase	See section 3.1.5
PAGE	Polyacrylamide gel electrophoresis
R_g	Radius of gyration

rpm Revolutions per minute

S Svedverg units, equivalent to 10^{-13} sec

$s^0_{w,20}$ Infinite dilution sedimentation coefficient at 20 °C in water

SANS Small angle neutron scattering

SAS Small angle scattering

SAXS Small angle x-ray scattering

SDS Sodium dodecyl sulphate

SE Sedimentation equilibrium

sec seconds

SV Sedimentation velocity

TE buffer 10 mM Tris-HCl (pH 8.4), 0.1 mM EDTA

TEGG90 20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10% v/v glycerol, 200 mM NaCl, 90% v/v D_2O

TEGGS 20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 20% v/v glycerol, 200 mM NaCl

TEGGS43 20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 20% v/v glycerol, 200 mM NaCl, 43% v/v D_2O

TEGGS65 20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 20% v/v glycerol, 200 mM NaCl, 65% v/v D_2O

TES100 buffer 20 mM Tris-Hcl (pH 7.5), 0.1 mM EDTA, plus 100 mM NaCl

TES1000 buffer 20 mM Tris-Hcl (pH 7.5), 0.1 mM EDTA, plus 1 M NaCl

TES50 buffer TE buffer, plus 50 mM NaCl

Tn3R Wild-type Tn3 resolvase

Tris	*tris* -(hydroximethyl)aminoethane
UV	Ultra-violet
VIS	Visible

Contents

List of Figures

Index

Chapter 1

Biological introduction

Summary

This chapter will introduce the biological problem under study. The different types of DNA recombination and transposition will be discussed in general terms. Later, a more detailed description of the Tn3 transposon, Tn3 resolvase, the models for synapsis formation and architecture, and finally the proposed mechanisms for strand exchange, will be presented.

1.1 DNA Recombination

The existence of a process leading to gene re-assortment, such as duplication, deletion, inversion and translocation was recognised before DNA was discovered to be the carrier of genetic material. This process is nowadays recognised as DNA recombination, and since its discovery its role has also been linked to other phenomena, such as DNA repair. DNA recombination mechanisms can be organised into four main categories:

- *Homologous recombination* is the exchange between homologous sequences placed anywhere in the genome. The process of homologous recombination is linked to myriad pathways. The RecA protein in *E.coli* (or its homologues,

e.g. Rad51 in yeast/humans) has been found to be fundamental to all these pathways, and is responsible for synapsing homologous sequences prior to 'strand exchange'. Other proteins, such as *E. coli* RecBCD, RecF, and RecE are probably responsible for generating the substrates used by RecA and for resolving its products. A Holliday junction is an intermediate in the strand exchange mechanism. The different strand exchange mechanisms in which RecA is involved are reviewed elsewhere (Kornberg and Baker, 1991).

- In *site-specific recombination*, exchange occurs when both exchanging partners possess specific sequences that are recognised by a specialised recombinase enzyme. A well known example of this category is the integration of λ phage into the *E.coli* genome.

- *Transposition* is the process by which a mobile genetic element called a *transposon* catalyses its own translocation to a new chromosomal location. In contrast with the site-specific recombination mechanism, transposition requires a specific sequence only on the DNA fragment belonging to the transposon, and typically has little specificity for the insertion site in the chromosomal DNA.

- *Illegitimate recombination* is the ensemble of recombination events that do not fall into any of the above categories. Nonhomologous integration of transfected DNA, as well as DNA translocations and deletions are some of the processes belonging to this group, and do not involve homology or site-specific recombinases.

In the next subsections, a more detailed description of some of these processes, with relevance to the problems studied in this thesis, will be presented.

1.1.1 Site-specific recombination

Site-specific recombination is a process employed by organisms in order to perform spatially and temporally defined rearrangement of DNA molecules, such as

phage integration and excision (e.g. λ integrase; Azaro and Landy, 2002), resolution of circular multimers (e.g. Xer, Cre and Flp recombinases; Barre and Sherratt, 2002; Rice, 2002; Van Duyne, 2002), inversions for expression of alternate genes (e.g. Hin and Gin invertases, Johnson, 2002) and assembly of genes during development. As opposed to homologous recombination, site-specific recombination events do not require extensive regions of homologous DNA or the homologous recombination machinery. In fact, they require only a small number of proteins, one of which interacts with a short recognition site. Recombination initiation depends on protein-DNA and protein-protein contacts rather than on DNA-DNA contacts.

Site-specific recombination reactions share a few features:

- DNA sequences are neither lost nor gained, and so DNA synthesis is not required.
- Exchange occurs at short DNA recombination sites.
- A small number of recombinase protein subunits recognise the recombination sites, and cleave and ligate the DNA. Nucleotide cofactors are not required for the reaction, which is often assisted by other DNA-binding proteins.
- Reactions can either be intermolecular, involving two DNA molecules with a single recombination site each, or intramolecular, involving a single DNA molecule containing two recombination sites.
- The relative orientation of the recombination sites defines whether intramolecular reactions bring about inversion (inversely repeated sites) or deletion (directly repeated sites) of the DNA fragment between the recombination sites.

Homology at the amino acid level, and type of catalytic mechanism employed allows for a classification of site-specific recombinases into two distinct groups:

1. Recombinases in the *tyrosine recombinase* (also integrase) family, including Cre, Flp, λ Int and other phage integrases, make staggered DNA breaks with 5' overhangs of 6 to 8 nucleotides, forming a covalent protein-DNA bond via a tyrosine to the 3'-phosphate. Their catalytic activity takes place at the C-terminal domains that present a relatively high degree of homology (with specific histidine, arginine, and tyrosine residues completely conserved). On the other hand, the N-terminal domains, involved in DNA binding, are not well conserved (see amino acid alignment in Fig. 1.1A). The process of strand exchange is a two-step mechanism that utilises a Holliday junction as a recombination intermediate.

2. In the *serine* family of recombinases, that includes Tn3 and $\gamma\delta$ resolvases and the Hin family of invertases, the protein cleaves the DNA and becomes attached to it by a serine-5'-phosphodiester linkage, the 3'-OH ends of the DNA at the cleavage site forming only a two nucleotide overhang. In this family, the C-terminal domains are involved in DNA binding, whereas the highly conserved N-terminal domains are involved in strand exchange (see amino acid alignment in Fig. 1.1B). As opposed to tyrosine recombinases, proteins in the serine recombinase family produce DNA double strand breaks, and exchange the strands in one step.

The functions of site-specific recombination involve the resolution of transposition intermediates, the switching of gene expression, the monomerisation of multimeric plasmids and chromosomes and the integration of bacteriophages.

1.2 DNA transposition

DNA mobile elements that catalyse their own movement to new chromosomal locations are called *transposons*, or *insertion sequences*. Transposons can be found in virtually all organisms, namely, bacteria, plants, fungi, insects and animals. Although some organisms can benefit from transposons genes (e.g. antibiotic resistance), they essentially act as 'selfish DNA'.

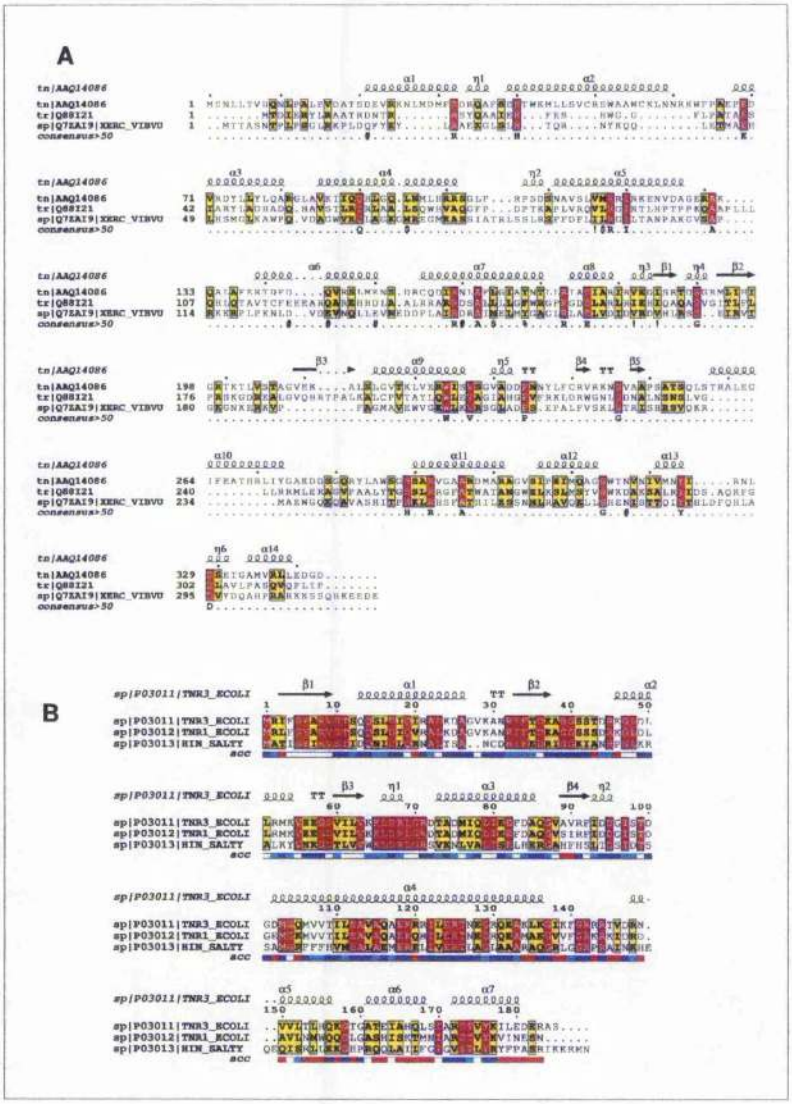

Figure 1.1: (A) Amino acid alignment of Cre (tnAAQ14086), a phage integrase (gene PP3181, *Pseudomonas putida*, strain KT2440, trQ88I21), and XerC (spQ7Z2I9), and secondary structure elements of Cre (Guo et al., 1997).(B) Amino acid alignment of Tn3 resolvase (spP03011), $\gamma\delta$ resolvase (spP03012) and Hin invertase (spP03013), and secondary structure elements of $\gamma\delta$ resolvase (Sanderson et al., 1990). Completely or partially conserved aminoacids are marked in red and yellow, respectively. The numbers above the sequences represent aminoacid position. α_i represent α-helices, β_j represent β-sheets and η_k represent unstructured regions.

Transposition events can be classified into three categories: (i) in *non-replicative transposition*, the transposon is transferred from one location of the chromosome to another; (ii) in *replicative transposition*, the transposon is copied to another site in the chromosome, leaving a copy of itself in the original site; (iii) in *retrotransposition*, an RNA intermediate is created, reverse-transcribed, and the resulting cDNA inserted into a new chromosomal site.

In both replicative and non-replicative transposition, the two ends of the transposon have nearly the same sequence, in inverted repeat. Transposition is achieved by a mechanism in which a protein, called a transposase, specifically binds and cleaves the end sequences.

Most bacterial genomes sequenced have been shown to have acquired a large proportion of their DNA from other sources (Normark and Normark, 2002). This horizontally acquired DNA provides the organism with selective advantages to antibiotic resistance, virulence and biodegradation pathways. Several types of mobile genetic elements have been shown to be responsible for the transfer of antibiotic resistance, such as transposons, integrons, resistance islands, and self-replicating plasmids. The study of the complete genome of the multidrug resistant *Salmonella enterica* serovar Typhi, recently sequenced by Parkhill et al. (2001), shows how the repeated insertion of DNA mobile elements contributes to the evolution of multiple antibiotic-resistant bacterial strains. At one locus, a tetracycline resistance transposon has integrated a chloramphenicol resistance cassette flanked by IS1, where a mercury resistance operon cassette, flanked by IS4321, was found. Lastly, a cassette encoding β-lactam, sulphonamide and streptomycin resistance, flanked by IS26, was located in the mercury resistance cassette. This example demonstrates that, although there is hope for novel classes of antibiotics to be developed, bacteria will certainly evolve resistance against them, and thus the study of how resistance evolves and spreads is paramount to prevent or at least delay the spread.

1.3 The Tn3 transposon

Tn3 is the prototype of a family of closely related mobile genetic elements referred to as the class II or Tn3 family of transposons. Transposons in this family encode a transposase protein, and have almost identical terminal inverted repeat sequences. By employing a replicative pathway, the transposase generates, from the donor replicon and the target DNA, a transposition intermediate referred to as a *cointegrate*, which contains two copies of the transposon in direct repeat orientation (see Fig. 1.2A). Tn3-family transposons almost always encode a site-specific recombinase that is responsible for the excision of the donor replicon from the cointegrate, leaving a simple transposon insertion in the target sequence. In Tn3 transposition, this is achieved by a serine recombinase of the resolvase/DNA invertase family, referred to as Tn3 resolvase (Tn3R). Tn3 is a single transposable unit, flanked by two 38 base pair (bp) inverted repeats, and contains the genes encoding transposase (*tnpA*), resolvase (*tnpR*), and β-lactamase (see Fig. 1.2B). A more detailed description of Tn3 transposition can be found in Grindley (2002).

1.4 The *res* site

res is located in the 170 bp intergenic region between the *tnpA* (transposase) and *tnpR* (resolvase) genes of Tn3. These genes are divergently transcribed from promoters within *res* (repressed by the binding of Tn3 resolvase). Each *res* site is composed of three individual binding sites for resolvase (see Fig.1.3A). Each binding site is made up of inverted copies of an imperfectly conserved common recognition sequence, separated by short spacers of slightly different sizes (Fig. 1.3A). In other natural *res* sites the distances between the centres of sites I and II vary, but are always close to an integral number of helical turns of B-DNA (Grindley, 1994). The most common separation between the centres of sites I and II is five turns (53 bp), but shorter (4 turns) or longer (7 turns) distances are found. The separation between the centres of sites II and III in Tn3 and $\gamma\delta$ is 34.5

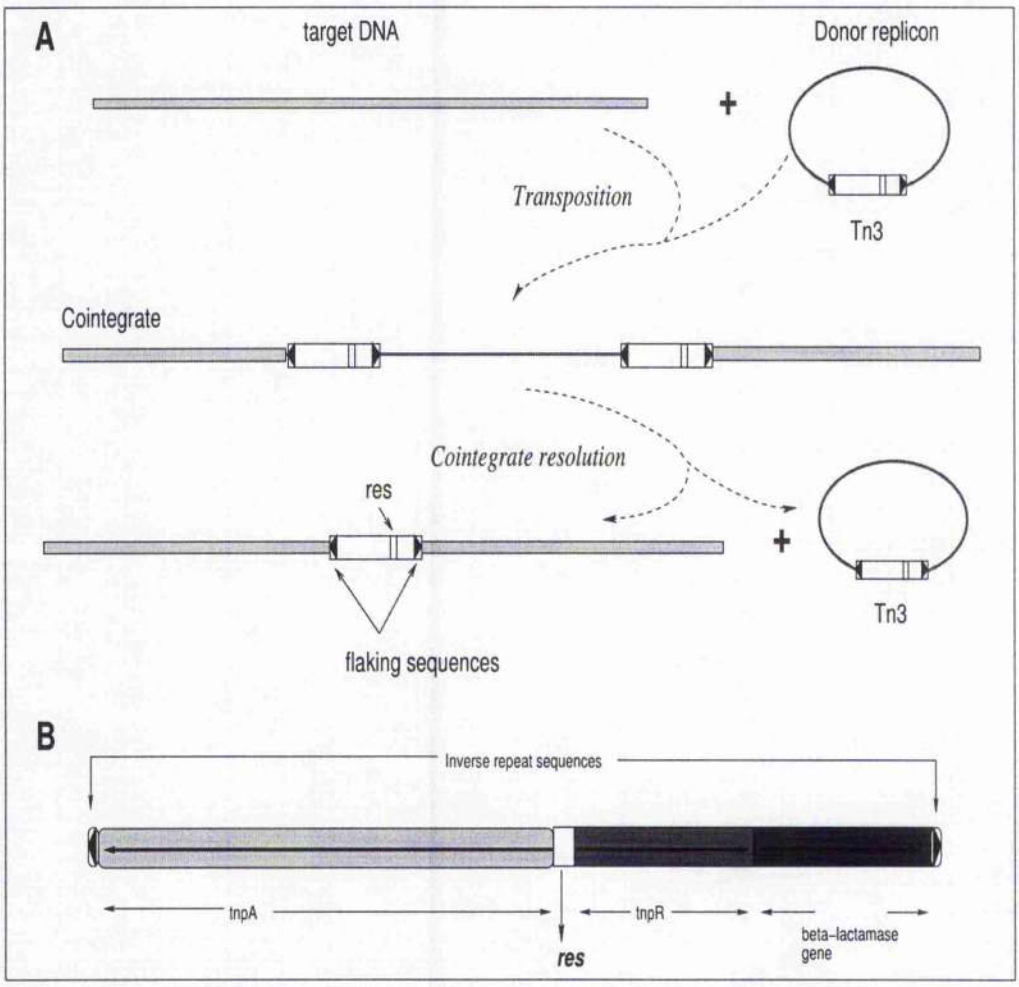

Figure 1.2: The Tn3 transposition mechanism. (A) The transposon-encoded transposase inserts the donor into the target by a replicative pathway, producing a cointegrate. The cointegrate is resolved by resolvase into a regenerated donor replicon and the target with a simple transposon insertion. (B) Genetic organisation of Tn3.

bp. These separations have been shown to be critical for recombination, and small variations have been shown to cause inhibition of the recombination reaction. Tn3 resolvase binds to the three sites in a similar manner (see Fig. 1.3B), so that six protein subunits bind to each *res* site. The geometrical differences between the individual sites might trigger structural differences in the way in which resolvase binds to them (Grindley, 1994). The crossover reaction occurs at the centre of site I (Reed and Grindley, 1981).

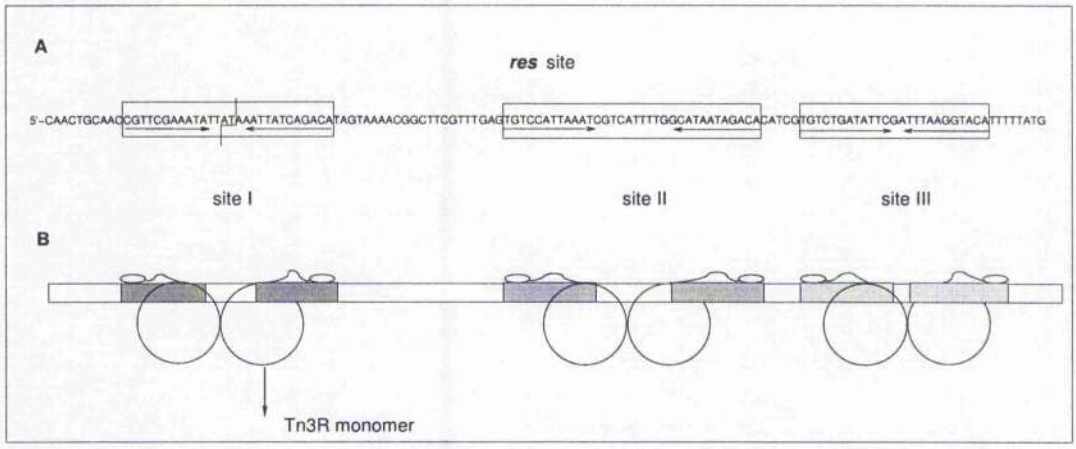

Figure 1.3: (A) The sequence of *res*. Individual sites are enclosed by boxes. Inverted repeat recognition sequences are signalled by arrows, and the crossover site is marked by a staggered line at site I. (B) Schematic view of a Tn3R dimer bound to each of the individual sites comprising *res*.

1.5 Tn3 resolvase

Tn3 resolvase (Tn3R) is a site-specific recombinase belonging to a family which also includes the $\gamma\delta$(Tn1000), Tn21 and Tn501 resolvases, and the DNA invertases Gin and Hin. Tn3 resolvase is able to resolve supercoiled plasmids containing directly repeated 114 bp *res* sites into two smaller circular plasmids, each of them with a single *res* site.

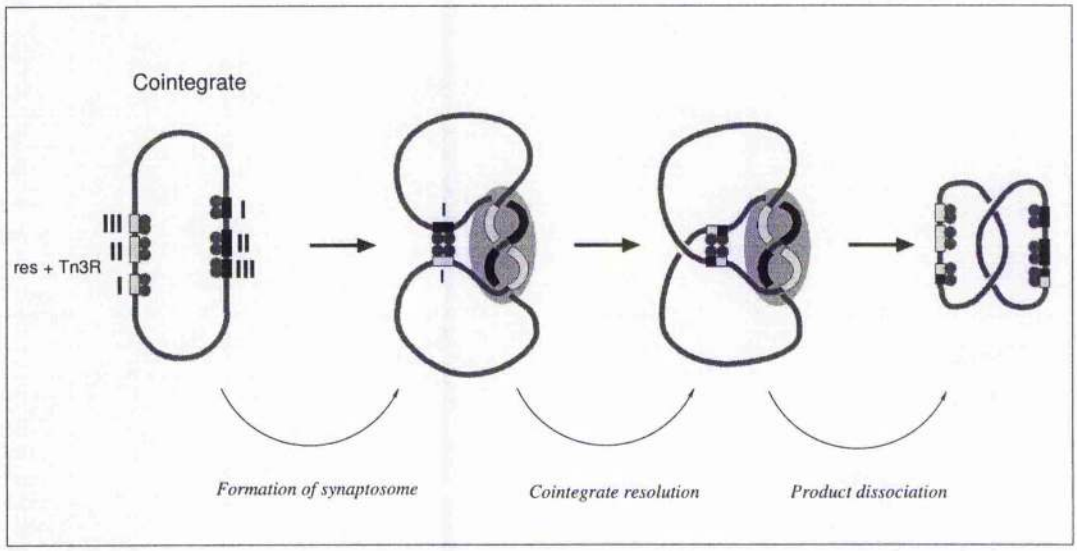

Figure 1.4: Resolution of the cointegrate by Tn3R. Tn3R binds a cointegrate containing two *res* sites in direct repeat orientation, giving rise to two resolvosomes. The resolvosomes come together and form a synaptosome, which is resolved into a 2-noded catenane by strand exchange. The relative positioning of the protein subunits and the DNA path shown in this scheme are only hypothetical.

The Tn3R *in vitro* resolution reaction requires a buffer, a supercoiled substrate with two head-to-tail *res* sites and wild-type Tn3 resolvase (Reed, 1981).

The site-specific intramolecular recombination reaction accomplished by Tn3R can be subdivided into 4 steps (Fig 1.4): (a) resolvase subunits bind to the two *res* sites present in a plasmid, forming two *res*-Tn3R complexes (resolvosomes); (b) the resolvosomes come together (see Section 1.7); (c) a synaptic complex (synaptosome) composed of two *res* sites and twelve Tn3R subunits is formed. This starting synapse traps three negative interdomainal superhelical nodes; (d) the DNA strands at sites I and I' are cleaved and exchanged[1]; finally, (e) the strands are rejoined into their new positions. Each round of recombination introduces one interdomainal positive node. Thus, one round of recombination gives rise to the product of cointegrate resolution, a (-2) noded catenane.

1.6 X-ray crystallography data

Tn3R and $\gamma\delta$ resolvase ($\gamma\delta$R) share 82% identity at the amino acid level, and therefore they are treated in this study as structurally identical (see Fig. 1.1). The structure of Tn3R can be divided into three regions. The globular N-terminus region (NT), from amino acid 1 to approximately 100, composed of an alternation of β-sheet ($\beta1 - \beta4$) and α helix ($\alpha1 - \alpha3$) secondary elements (see Fig. 1.1), is responsible for catalytic activity. The C-terminus region (CT), spanning from amino acid 139 to 185, composed of three α helices ($\alpha5 - \alpha7$ in Fig. 1.1) linked by unstructured loops, is responsible of DNA binding. Finally, the E-helix, a long α helix between amino acid 100 and 138 ($\alpha4$ in Fig. 1.1), links the two aforementioned regions (see Fig. 1.5).

Two distinct domains were identified by proteolytic cleavage (Abdel-Meguid et al., 1984). α-chymotrypsin cleaves resolvase at Phe-140, producing two domains, one of 140-residues (catalytic domain) and the other of 43 residues (DNA binding domain). Crystallographic and mutagenesis studies showed that the

[1]The prime being assigned to the second copy of the *res* site in the plasmid

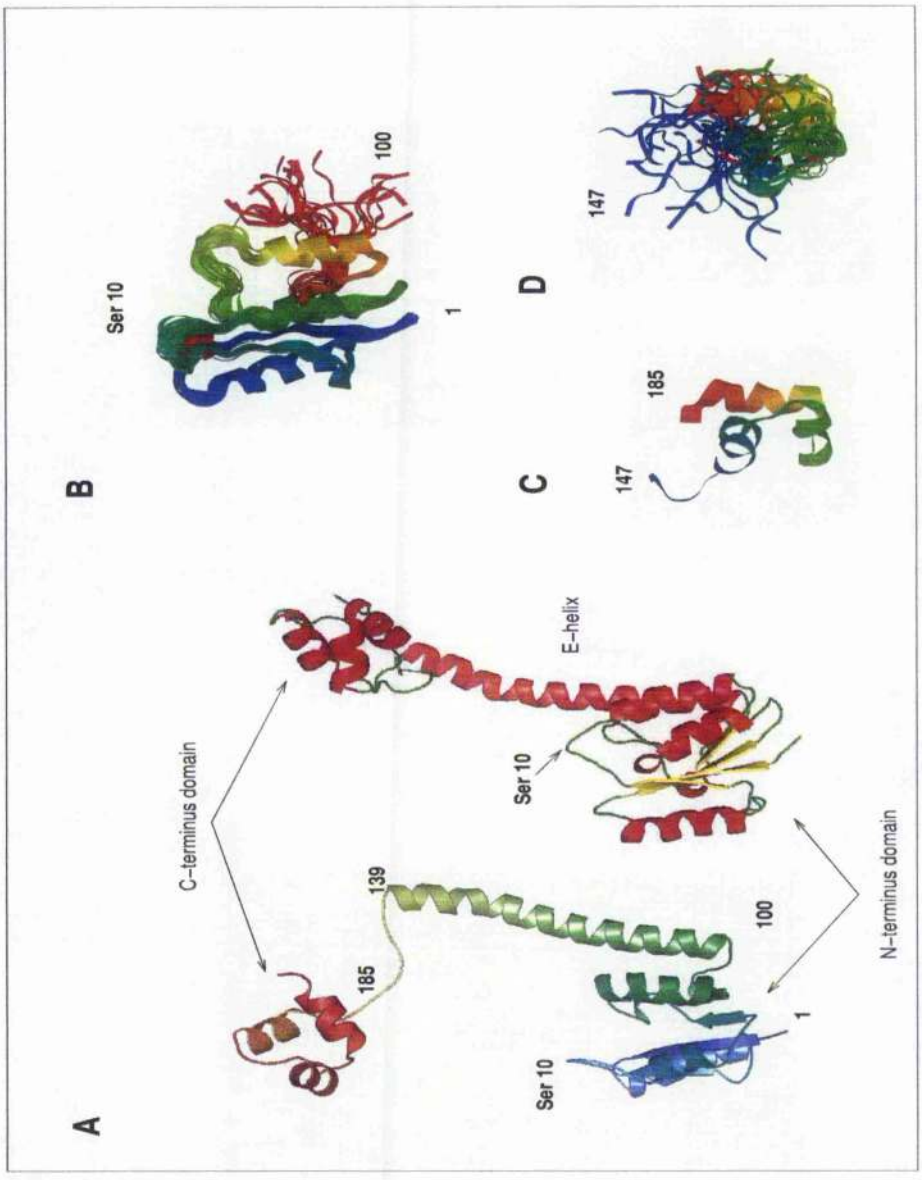

Figure 1.5: (A) Two orthogonal views of the domain structure of monomeric $\gamma\delta$R (based on Yang and Steitz, 1995). (B) Structure of the catalytic domain of $\gamma\delta$R determined by NMR spectroscopy (Pan et al., 2001). (C) Average structure (left) and (D) superposition of all NMR spectroscopy-determined structures of the DNA-binding domain of $\gamma\delta$R (Liu et al., 1994).

catalytic domain is not only responsible for cleaving the DNA, exchanging the strands and rejoining them, but also for dimerisation and for higher order interactions leading to the formation of multimeric protein assemblies, which play a fundamental role in resolvase activity. From a structural standpoint, this domain is an $\alpha - \beta$ structure consisting of a central β sheet made up of four parallel and one antiparallel β strands surrounded by four α helices (see Fig. 1.5). The active site lies mainly in a loop on a protruding part of the protein structure between the first β strand (residues 1-9) and an α helix (residues 14-32), Ser-10 being its most important residue (see Fig. 1.5A and B). Other residues (R8, D67, R68, etc) are also part of the active site.

The first crystal structure of the catalytic domain was determined by Sanderson *et al.* (PDB accession code 2RSL, 1GDR; Rice and Steitz, 1994b; Sanderson et al., 1990), and is shown in Fig. 1.6. The figure shows the crystallographic asymmetric unit composed of three protein subunits (only the N-terminal domain of $\gamma\delta$ resolvase). The subunits are arranged as two overlapping dimers (1-2 and 2-3) (Fig. 1.6C). The interface between subunits in the 1-2 dimer is well packed and is largely formed by two highly conserved E-helices interacting in a coiled-coil configuration. The distance between the active sites in each subunit (Ser-10) is more than 30 Å, and this dimer does not form a tetramer in the crystal. On the other hand, the contacts at the interface of the 2-3 dimer are not extensive and include amino acids highly variable amongst the resolvase family. The two Ser-10s are, however, only 19 Å apart and this 2-3 dimer participates in a tetramer of 222 symmetry (2-3/2'-3', see Fig. 1.6) in the crystal. In addition, there are experimental (site-directed mutagenesis) data implying that the interactions between the 2 and 3' or 2' and 3 subunits in this tetramer are relevant for recombination (Hatfull et al., 1989; Sanderson et al., 1990). This evidence led in the past to the belief that the 2-3 tetramer constituted the catalytic unit of resolvase (Sanderson et al., 1990).

Site-directed mutagenesis experiments (Hughes et al., 1993) and new crystallographic evidence (see below) have demonstrated that this hypothesis is false, and that the 1-2 dimer, even with its unexpectedly large distance between catalytic sites, is the one that apparently is implicated in catalysis. However, more

recent NMR spectroscopy studies on the catalytic domain of $\gamma\delta$ resolvase (Pan et al., 2001) in monomeric and dimeric forms revealed (see Fig. 1.5B) that the conformation of the active site region including Ser-10 changes as the protein dimerises and binds to the DNA (in comparison with a DNA-bound $\gamma\delta$R structure; see below). The different experimental structures for the 1-2 dimer (Pan et al., 2001; Rice and Steitz, 1994a; Sanderson et al., 1990; Yang and Steitz, 1995) show slightly different protein-protein interactions between the comprising subunits, and consequently different distances between the Ser-10 residues. This implies the existence of conformational changes induced in the active site region by protein-protein interactions.

The 2-3 interface is thus not involved in the formation of the catalytic subunit, even though it appears to mediate higher order inter-subunit interactions (Hughes et al., 1993). The L50C mutation (of a residue at the 2-3 interface) affects the efficiency of recombination but without achieving complete inhibition. Mutations in the 2-3' interface (see Fig. 1.6D) have been demonstrated to cause complete inhibition of recombination. Since each of the monomers taking part in the crystallographic tetramer (2, 3, 2' and 3') is attached to a second monomer through the 1-2 interface, the tetramer can be actually considered as an octamer, or tetramer of 1-2 dimers (see Fig. 1.6C) (Hughes et al., 1993). This octamer has been proposed to be the core of the synaptic complex (see below), linking the sites II and III in one *res* site to the ones in the other.

NMR structural studies on the binding domain (residues 143-183) of $\gamma\delta$R (Liu et al., 1994) found that this domain is composed of a helix-turn-helix motif (Fig. 1.5C), and that it has a high degree of flexibility when it is not interacting with DNA (Fig. 1.5D). Footprinting and DNA mutagenesis studies reveal that this domain interacts with the conserved inverted repeat sequence at site I primarily in the major groove, with some contacts in the minor groove located nearby. These results were confirmed by x-ray crystallography studies on $\gamma\delta$R (Yang and Steitz, 1995) and Hin (Feng et al., 1994).

Perhaps the most significant crystal structure of $\gamma\delta$ resolvase has been determined by Yang and Steitz (PDB accession code 1gdt; Yang and Steitz, 1995),

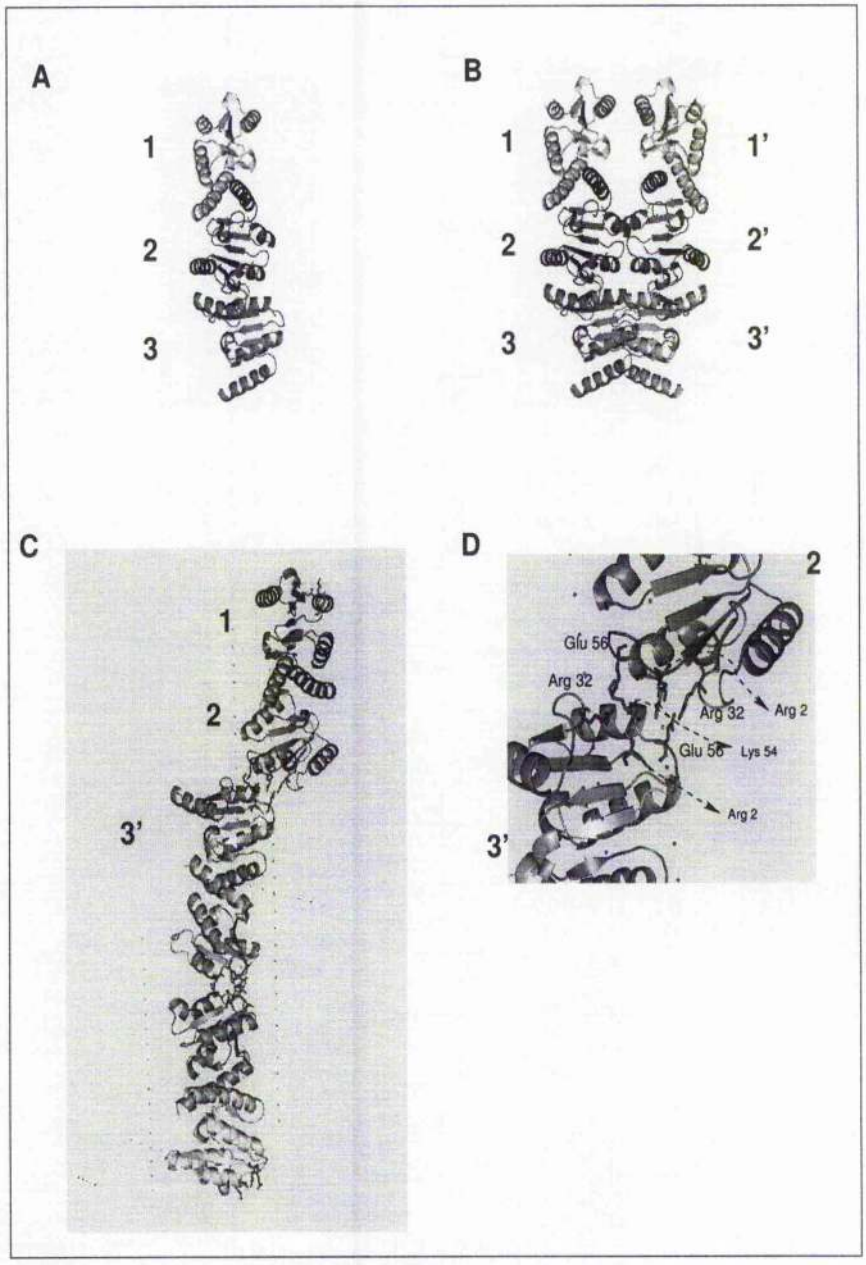

Figure 1.6: (A) Cartoon model of the three independent resolvase catalytic domains in the crystallographic asymmetric unit (1GDR, $C222_1$ symmetry, Rice and Steitz, 1994b; Sanderson et al., 1990). Subunit 1 is in yellow, subunit 2 in pink and subunit 3 in green. The non-crystallographic dyad axis relating subunits 2 and 3 is horizontal (in the plane of the page), whereas the non-crystallographic dyad axis relating subunits 1 and 2 is oblique to the page. (B) Subunits 1', 2' and 3' are generated by using a two-fold vertical crystallographic axis. (C) Arrangement of resolvase subunits in 1GDR ($C222_1$ symmetry, Rice and Steitz, 1994b; Sanderson et al., 1990). (D) Specific amino acids involved in the 2-3' interface (R2, R32, E54, E56).

showing a resolvase dimer bound to a symmetrized version of *res* site I (Fig.1.7). In this structure, the protein conformation is such that the nucleophilic residues are placed more than 20 Å from the crossover site, and it thus represents an inactive form of the protein. However, it is worth mentioning that a dimer of resolvase bound to site I in solution is actually 'designed' to be in a catalytically inactive state, and thus this structure is probably a very good approximation to how resolvase binds site I in solution. Additionally, it clearly shows the interactions taking place between the N-terminal domains of the two resolvase subunits bound to site I. The structure shows a 60° kink in the DNA at its centre which bends it away from the catalytic domains.

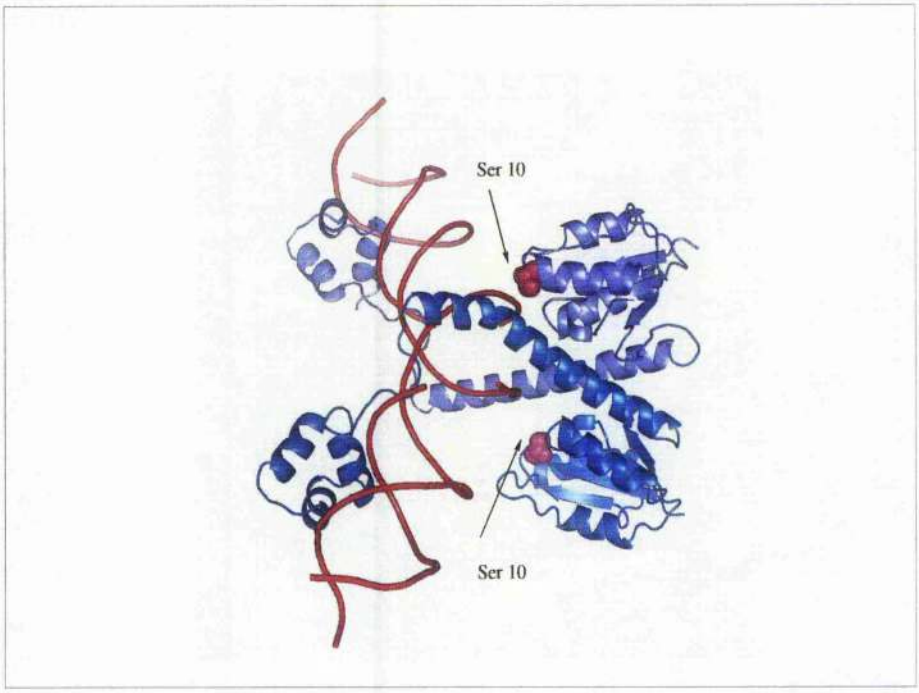

Figure 1.7: Co-crystal structure of a dimer of $\gamma\delta$-resolvase (backbone) bound to DNA (backbone) site I (Yang and Steitz, 1995).

1.7 Synapsis formation: slithering versus random walk

The first resolvase *in vitro* recombination reactions performed by Reed (Reed, 1981) and Cozzarelli (Cozzarelli et al., 1984) demonstrated that the reaction products were singly linked 2-noded catenanes. Subsequent rounds of recombination reactions on the same substrate were demonstrated to generate products with more complicated topologies, namely, a 4-noded knot, 5-noded catenane, 6-noded knot, etc (Cozzarelli et al., 1984; Wasserman and Cozzarelli, 1985). By studying the product topologies, it was concluded that the synaptic complex traps three negative interdomainal superhelical nodes regardless of the degree of supercoiling of the substrate and the distance between the *res* sites (Benjamin et al., 1985). Two models were then proposed to explain the formation of the synaptosome: Benjamin et al. (1985) suggested that synapsis occurs by slithering of the supercoiled DNA. Boocock et al. (Boocock et al., 1987, 1986) proposed instead that the synapse is formed when, by random collision, two *res* sites trap three nodes. The formation of synaptic complexes with other numbers of interdomainal supercoils was thus 'topologically filtered' by the architecture of the synapse. The existence of a synaptic complex also ensures that the recombination reaction is only intramolecular, as opposed to intermolecular, and that it occurs only when the *res* sites are in direct repeat orientation (head-to-tail).

1.8 Synapsis models

Before the recombination process takes place, a complicated mechanism must assemble a protein-DNA complex formed by twelve subunits of resolvase and two directly repeated copies of *res*, also referred to as a 'synapse' or 'synaptosome'. Various models for the architecture of the synaptosome, based on the reaction topology, structural information, and site-directed mutagenesis studies, have been hitherto proposed. However, there is so far little direct experimental structural

evidence to support these models, and they all present different kinds of inconsistencies with the available experimental data. In what follows, we shall present the most salient features of the proposed synaptosome architectures, thus providing a conceptual framework for the present study.

An early pre-crystallography synaptic model was proposed by Boocock et al. (1987) and was based on the topological information available at the time. The model proposed that site I interacts with site I' and that there are two additional quasi-equivalent dimer-dimer interactions mediating the interactions between sites II-III' and III-II' (Fig. 1.8A).

The solution of the first crystallographic structures drastically influenced the models that appeared thereafter. As noted above, a crystal structure (Sanderson et al., 1990) of the catalytic domain of $\gamma\delta$ resolvase led to the proposal of a tetramer of dimers being at the core of the synaptic complex at the accessory sites (II and III). This structure also permitted the identification of the 1-2 dimer as being the solution dimer, binding to the consensus repeat at the six sites within the synaptosome. These hypotheses gave rise to two models (Grindley, 1994; Rice and Steitz, 1994a). The most important difference between the models is in the orientation of the tetramer of dimers with respect to the complex at site I. Both models considered that sites I and III had to be spatially close to each other, that the DNA was bent at site II (based on footprinting data from Salvo and Grindley (1988)), and that the synaptic complex had to trap three negative superhelical nodes (which therefore would satisfy the reaction topology). These constraints determined the different paths taken by the DNA in the two models.

In the Rice and Steitz model (Rice and Steitz, 1994a), the direction of elongation of the tetramer of dimers is pointing towards site I, and the DNA is wrapped around the tetramer as shown in Fig. 1.8B. The two *res* sites are held together by the tetramer of dimers bound at sites II, II', III and III', and 2-3 (not 2-3') interactions between the DNA-binding domains of the subunits bound at sites I-I'. The tetramer of dimers is placed in such a way that the protein subunits bound to site II interact with the ones bound to site II'. Similarly, the dimers bound to site III interact with the ones bound to site III'. In the protein-DNA complex

formed at the site I/I' DNA, the protein catalytic domains are located on the outside and the DNA-binding domains, together with the DNA sites themselves, are on the inside. In this way, the site I-bound resolvases are in a position of 222 symmetry, but they are not in contact with one another. The major criticism of this model is, indeed, that it requires the breakage and reformation of any preexisting 2-3' interactions during synaptosome assembly. In the simplest case, two resolvase-bound *res* sites would have to break existing protein-protein and protein-DNA interactions in order to form the synaptosome. In a second scenario, two unbound *res* sites would come together and wrap around a preassembled resolvase core. None of these possibilities seems attractive (Grindley, 2002).

On the other hand, Grindley (1994) proposed an alternative model for the architecture of the synaptosome (see Fig. 1.8C). In this model, the direction of elongation of the tetramer of dimers is perpendicular to an imaginary axis joining the protein subunits at sites I-I' with the complex at the accessory sites. This constraint makes the DNA take a different path from that in the Rice and Steitz model. As before, the protein dimers bound to sites II and III in the same *res* site are proposed to interact via a 2-3' interaction. In this new model, however, the distance between sites II and III in the same *res* site is rather longer than expected (there are 34.5 bp between sites II and III, and this imposes a centre to centre maximum distance of 117 Å). Sites II and III' (similarly II' and III) are joined by protein tetramers, and the positions and interactions of the protein dimers at site I/I' are left undefined.

Until recently, there was no experimental information validating or refuting any of the structural models described above. By using mutant resolvases, Murley and Grindley (1998) recently proved that 2-3' proficient subunits are required at sites II-L and III-L, but not at site II-R and III-R (see Fig. 1.8E), in order to produce a competent synaptic complex at the accessory sites. This finding is compatible with both structural models. In addition, the authors also demonstrated that, for recombination, 2-3' proficient subunits are additionally required at sites III-R and I-R. This finding implies that, in the assembly of the final structure of the competent synaptosome, resolvase subunits at sites III-R and I-R interact

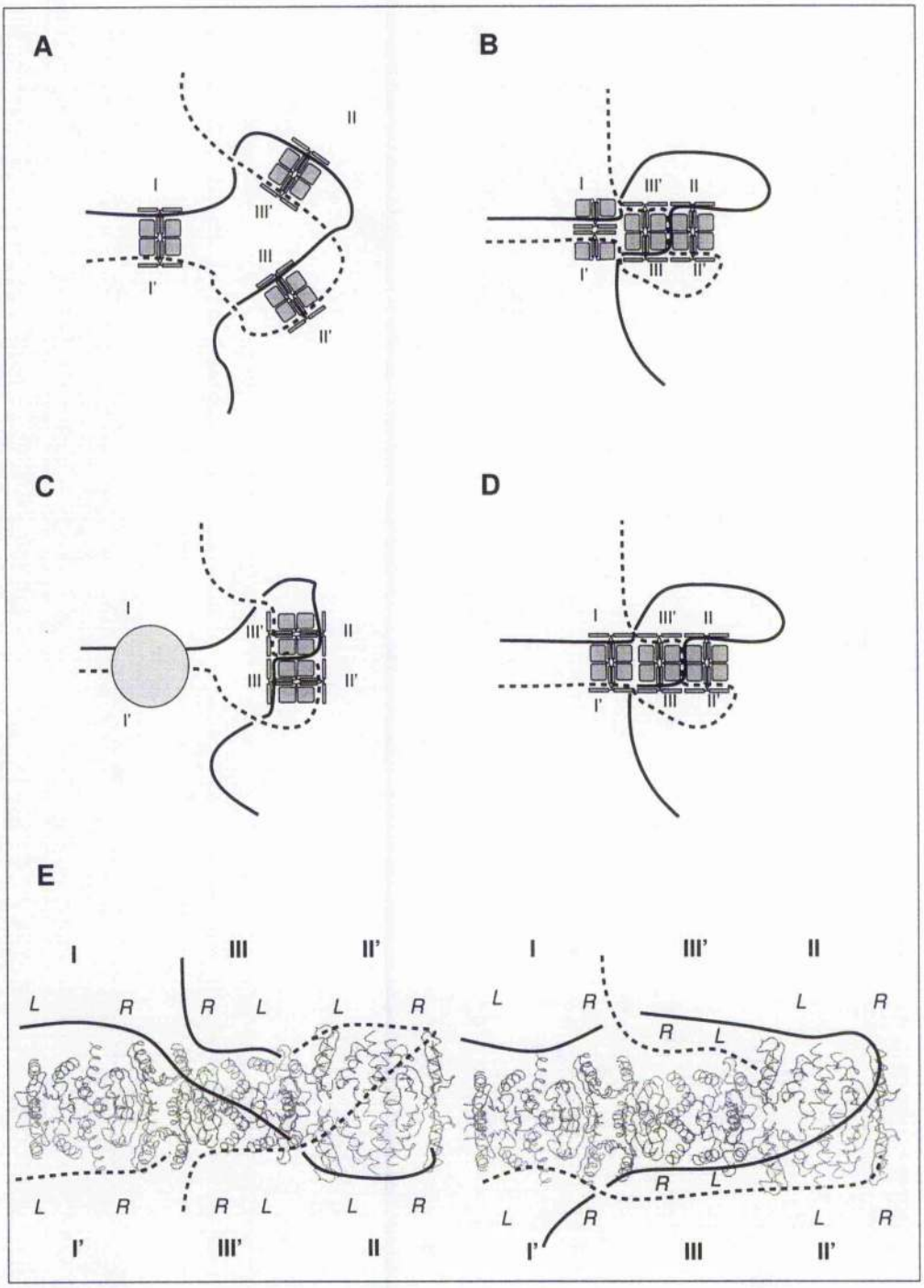

Figure 1.8: Schematic representations of the architectures of the synaptosome models proposed by (A) Boocock et al. (1987, 1986); (B) Rice and Steitz (1994a); (C) Grindley (1994) and (D) Sarkis et al. (2001). (E) Two models of the catalytic protein domains forming the synaptosome as proposed by Sarkis et al. (2001).

directly through a 2-3' interaction (Fig. 1.8E). This latter finding is, however, at odds with both the Rice and Steitz (1994a) and Grindley (1994) structural models.

In order to solve this problem, a new structural model was proposed by Sarkis et al. (2001) (see Fig. 1.8D and E). As in the previous models, a tetramer of dimers held by 2-3' interactions is at the core of the accessory site synapse. As a novel ingredient, this model proposes that each resolvase dimer bound at site I contacts a dimer at site III also through a 2-3' interaction. In this new model, the architecture of the synaptosome is formed by a fiber of twelve protein subunits around which the DNA is wrapped. Two different paths for the DNA have been proposed (see schemes in Fig. 1.8E), but no experimental evidence so far is able to discern between the two. In addition, this model is inconsistent with some experimental observations.

1.9 Strand exchange mechanisms

Early studies of the topology of resolvase-mediated recombination (Stark et al., 1989) showed that the motions of the DNA involved in strand exchange are equivalent to a $\pi/2$ right-hand rotation of one pair of half-sites relative to the other pair, in the cleaved-DNA intermediate. How this process is achieved is so far a mystery, although many years of experiments have imposed several restrictions on the possible mechanism. Stark et al. (1989) discovered that, upon strand exchange, there is a change in the linkage number of ΔLk=+4. Further studies (Stark and Boocock, 1994; Stark et al., 1991) showed that non-recombinant knotted products can be generated by multiple rounds of strand exchange, by using a single base substitution at the centre of the crossover site, which would result in a mismatch after one round of recombination. These studies demonstrated that the strand exchange mechanism can accommodate apparently controlled rotations of $2n\pi$ ($n = 0, 1, 2, ...$) without the synaptic complex falling apart, and without the need to form a transiently ligated recombinant product.

In addition to these constraints, McIlwraith et al. (1996) have shown that at least one subunit of the four taking part in the crossover site synaptic complex remains physically attached to the same half-site DNA during strand exchange. Further experiments have shown (McIlwraith et al., 1997) that two rounds of recombination can occur without any rejoining of the DNA strands in an intermediate recombinant form, eliminating a possible 'reset' mechanism[1] after one round.

Several models for the strand exchange mechanism have been proposed over the years, bearing a close relationship to the synaptosome architectural models.

The Rice and Steitz model for the synaptosome is tempting because the mechanism of strand exhange would involve a simple rotation of the DNA half-sites without the need to disrupt the protein architecture (see Fig. 1.9A). The mechanism would only require small movements of the DNA half-sites located in the inside of the complex. A mechanism similar to this has been unveiled for Cre and Flp recombinases by crystallography data (Guo et al., 1997; Rice, 2002). There are, however, a few caveats for this model to hold. The most reasonable movements for the DNA half-sites are not able to simultaneously satisfy the (+) node and (+) helical twist required to satisfy the reaction topology (Stark et al., 1989). To bypass this problem, Rice and Steitz proposed that the centre of each site might be unwound by half a turn, and that relaxation after strand exchange would produce the correct reaction topology. However, the crystal structure (Yang and Steitz, 1995) did not reveal evidence of the existence of such unwinding. Furthermore, the mechanism itself would not even allow a 2π rotation without resetting of the complex, thus failing to conform to some of the restrictions on the strand exchange mechanism found by the studies aforementioned.

By far the simplest mechanism for strand exchange, the *subunit rotation* model involves a right-handed $\pi/2$ rotation of the two protein subunits covalently bound to one pair of half sites relative to the other (Fig. 1.9B). This mechanism produces recombinants of the expected topology, and accounts for multiple rounds

[1] A mechanism by which the protein and DNA is reorganised into its original state before the initiation of the reaction.

of recombination. In one version of this model, DNA duplexes lie on the outside of the crossover site synapse and resolvase dimers interact via their catalytic domains. However, it is not so far clear what could hold the cleaved halves together while the protein subunits are rotating.

Finally, the *domain swap* model involves a right-handed $\pi/2$ rotation of the resolvase catalytic domains covalently attached to one pair of half sites relative to the other half-sites by using the loop at the base of the E-helix as a hinge (Fig. 1.9C). The DNA-binding domains and the E-helices remain in place and thus retain many of the protein-protein contacts giving rise to the crossover site synapse structure. This model satisfies the topology requirements for the reaction. However, multiple rounds of recombination would intertwine the rotational hinges, probably requiring some type of resetting mechanism. Both the domain swap and the subunit exchange models are compatible with the recently proposed synaptosome architecture (Sarkis et al., 2001), but not with the Rice and Steitz model.

The understanding of the synaptosome architecture and the mechanism of strand exchange are relevant to the study of many site-specific recombinases in the serine recombinase family. More importantly, it is also useful as a paradigm to comprehend, at the molecular level, fundamental mechanisms in protein-DNA interactions.

The aims of this thesis were

- to study the oligomerisation behaviour and low-resolution conformation of Tn3R in solution.
- to analyse its interaction with each of the three binding sites in *res*.
- to develop a rigid-body formalism to model small angle x-ray scattering datasets and test it on Holliday DNA junctions.
- to apply this formalism to solve the solution structure of the site I-Tn3R protein-DNA complex.

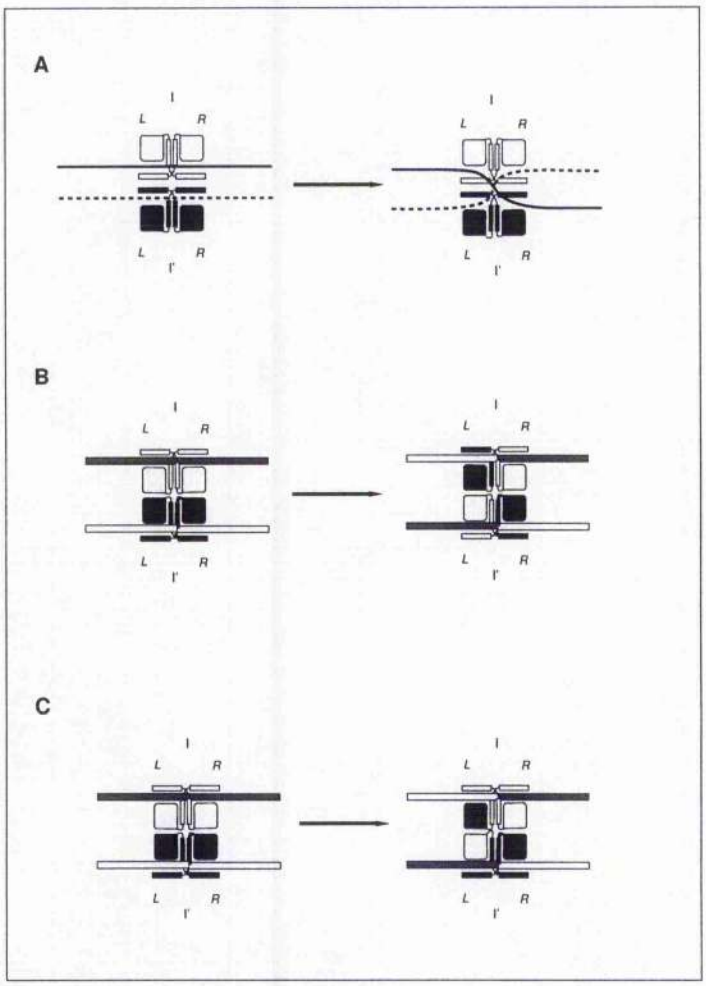

Figure 1.9: Three models for the strand exchange mechanism. (A) The DNA-mediated mechanism, with its origin in the Rice and Steitz model for the synaptosome architecture. (B) The subunit rotation mechanism, by which the two resolvase subunits each holding one half-site, perform a $\pi/2$ right-hand rotation in the presence of negative supercoiling. (C) The domain swapping mechanism, where the DNA binding domains of the left-hand half-sites dissociate, and the DNA-bound catalytic domains of the left-hand side protein subunits perform a $\pi/2$ rotation.

- finally, to extend this formalism and generalise it to also make use of fluorescence resonance energy transfer and sedimentation velocity data in the modelling procedure.

Chapter 2

Introduction to biophysical techniques

2.1 Summary

This chapter will provide an introduction to the biophysical experimental techniques employed in this study, namely, small angle x-ray/neutron scattering, sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation, and fluorescence anisotropy. In addition, it describes the recent applications of these techniques to problems of relevance to this thesis.

2.2 Small angle scattering

Small angle scattering (SAS) is a technique that provides structural information on characteristic dimensions and electron density inhomogeneities between ten and a few thousand angstroms. These distances are large compared to interatomic distances, and thus it is possible to describe the scatterers by using a continuous function of the electron density, as determined by the chemical composition of the particle under study. The scattering objects will be considered, in this introduction, as non-interacting, and therefore their spatial distribution will be uniform. Therefore, the scattering intensity will be only analysed in terms of

the particle scattering form factors.

A small angle x-ray scattering (SAXS) synchrotron radiation instrument comprises the following elements: A synchrotron radiation source provides the x-rays. A *monochromator* selects one well defined wavelength in the emission from the undulator. *Focusing optics*, mainly composed of either a toroidal mirror, or two orthogonal elliptically bent mirrors, are used to focus the beam. Tantalum or tungsten *collimating slits* are used to eliminate parasitic scattering. The sample position is usually fixed, and the sample-to-detector distance (d) can be adjusted by moving the detector, placed inside a vacuum chamber, away from the sample. As the beam is normally 10^4-10^5 times more intense than the total scattered intensity, a highly absorbing *beam stop* is placed between the sample and the detector to prevent the direct beam from damaging the detector. An *x-ray detector* measures the scattering intensity as a function of the scattering angle 2θ. Two-dimensional (Beamline 2.1 at Daresbury, Warrington, UK and beamline 26B at the ESRF, Grenoble, France) or linear (Beamline X33 at DORIS, Hamburg, Germany) detectors are often used. Data acquisition software and hardware are beamline-specific.

An introduction to x-ray scattering by matter is provided in Appendix 1. The main methods for SAXS and small angle neutron scattering (SANS) data analysis will be described in the following sections.

Ideality means in thermodynamic terms that the particles in solution do not have interactions between them. In this approximation, thermodynamic intensive quantities are proportional to the particle concentration. *Monodispersity* is a concept with two possible meanings. In *chemically* monodisperse solutions, the particles all have identical chemical composition. In solutions with *shape* monodispersity, the particles all have the same shape. In the rest of this introduction, I will strictly deal with ideal and monodisperse (in both senses) solutions. Therefore, all the particles in solution will scatter with the same scattering amplitude but in different directions. Moreover, the resulting intensity will be the sum of the intensities of the individual particles (i.e. no interference term). Thus,

the total scattering intensity will be proportional to the number of particles in the sample.

2.2.1 Global parameters and special cases

The scattered intensity $I(\vec{s})$ (see Appendix 1) can be written as

$$I(\vec{s}) = \int\int \Delta\rho(\vec{r_1})\,\Delta\rho(\vec{r_2})\,e^{i\,\vec{s}.(\vec{r_1}-\vec{r_2})}\,dV_1\,dV_2 \tag{2.1}$$

where the volume integral runs over the particle volume V, $\vec{r_1}$ and $\vec{r_1}$ are vectors inside the particle, $\vec{s}$ is the scattering vector, and $\Delta\rho$ are the fluctuations in the electron density of the particle with respect to the solution.

The isotropic scattering intensity $I(s)$ is often expressed in terms of the pair distance distribution function (PDDF or $p(r)$ function) as follows (Appendix 1)

$$I(s) = 4\pi \int p(r)\,e^{i\,sr}\,dr. \tag{2.2}$$

For homogeneous particles, the $p(r)$ function can be written as

$$p(r) = \frac{1}{4\pi}\sum_{i,j=0}^{N} \delta(|\vec{r_i} - \vec{r_j}| - r), \tag{2.3}$$

where N is the number of electrons in the particle, and $\delta(a-b)$ is Kronecker's δ-function that is 1 when $a = b$ and 0 otherwise (Appendix 1).

Some quantities can be directly derived from Eq. 2.1, for example, the scattering amplitude at zero angle is (Vachette and Svergun, 2000)

$$I(0) = \int\int \Delta\rho(r_1)\,\Delta\rho(r_2)\,dV_1\,dV_2 = \Delta m^2 = m^2 - m_0^2,$$

where m is the total number of electrons in the particle and m_0 is the number of electrons in the solvent displaced by the particle ($I(s)$ is thus proportional to the particle electron density and to the particle-specific partial volume).

Expansion of the scattering intensity in a Taylor series of s^2 around the origin, gives

$$I(s) = I(0)\left[1 - \frac{4\pi^2 R_g^2 s^2}{3} + \kappa s^4 + ...\right].$$

Retention of only the first order term, yields

$$I(s) \simeq I(0)\exp^{-\frac{4\pi^2 R_g^2 s^2}{3}}, \tag{2.4}$$

where R_g is the particle radius of gyration, defined by the following relation (Vachette and Svergun, 2000)

$$R_g^2 = \frac{\int \Delta\rho(\vec{r}) r^2\, dV}{\int \Delta\rho(\vec{r})\, dV}. \tag{2.5}$$

Eq. 2.4 represents the Guinier approximation (Guinier and Fournet, 1955), that states that at very small angles the scattering intensity behaves as a Gaussian, the width of which is *inversely proportional* to the radius of gyration (see Fig. 2.1). The *Guinier region* is the s-range in which the Guinier approximation is valid, usually defined as $0.2 < sR_g < 1.3$[1].

On the other hand, assuming that the particle has a uniform electron density and a sharp interface with the solvent, it is possible to deduce that the scattering intensity in the high-angle region will have the following behaviour (Porod, 1951)

$$8\pi^3 \lim_{s\to\infty}[s^4 I(s)] = S\bar{\Delta\rho}^2 + Bs^4, \tag{2.6}$$

[1]These approximate values vary for assymetric particles, such as prolates, oblates (Perkins, 1988)

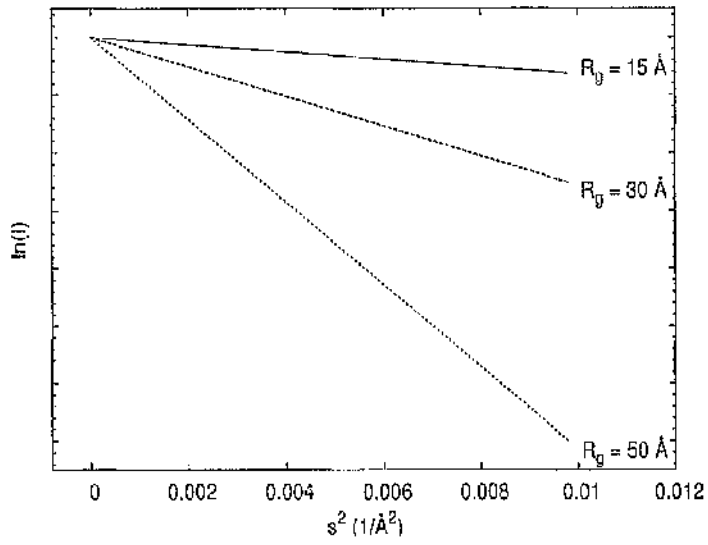

Figure 2.1: (A) The Guinier approximation for $R_g = 15$, 30 and 50 Å.

where S is the area of the interface between the particle and the solvent, and B is a correction factor. B can sometimes be quite large, due to the small values taken by $I(s)$ at high angles (at high enough angles the particle always scatters as the solvent, and so $\lim_{s\to 0} I(q) = 0$) and this often limits the usefulness of this relation.

The autocorrelation function $\gamma(r)$ is defined by

$$\gamma(r) = \int \rho(r_1)\rho(r_1 - r)\, d^3r1. \tag{2.7}$$

By applying the inverse Fourier transform to Eq. 2.2, the auto-correlation function in terms of the scattered intensity can be obtained, that is to say,

$$V\gamma(r) = \frac{1}{2\pi^2}\int_0^\infty s^2 I(s)\frac{\sin(sr)}{sr}\, ds. \tag{2.8}$$

By evaluating this expression at $r = 0$ it is possible to obtain

$$V\gamma(0) = \frac{1}{2\pi^2}\int_0^\infty s^2 I(s)\, ds,$$

that directly shows how the mean square fluctuation in the electron density is related to the integral of the intensity over the reciprocal space, without having anything to do with the particle structure. In this way, the system can suffer any conformation change, thus changing the intensity profile, but without modifying the value of the integral, upon which Porod's invariant Q is defined:

$$Q = \int_0^\infty s^2 I(s) ds. \tag{2.9}$$

By evaluating Eq. A.5 at $r = 0$ it is possible to conclude that Porod's invariant can also be expressed in terms of the mean square of the electron density contrast ($\Delta\bar{\rho} = \bar{\rho}_{particle} - \rho_{solvent}$) of the sample:

$$Q = V\Delta\bar{\rho}^2. \tag{2.10}$$

2.2.2 Simple geometrical examples

The $p(r)$ function of a homogeneous sphere can be calculated analytically by using Eq. 2.3 and was first calculated by Porod (Porod, 1948),

$$p(r) = 12x^2(2 - 3x + x^3),$$

with $x = r/D$, D being the sphere diameter. The function has its maximum at $r = 0.525D$, close to the sphere radius, and has a bell-function shape (see Fig. 2.2) with a maximum size of D_{MAX} (Glatter and Kratky, 1982).

Rod-like particles are another important class of simple shapes common in biological systems. These particles have a constant arbitrary cross-section (with

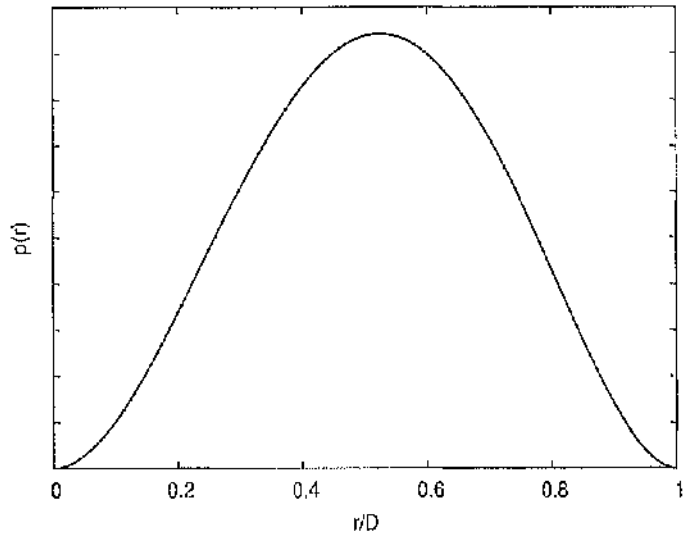

Figure 2.2: Particle distance distribution function $p(r)$ of a sphere of diameter D.

surface area A) whose characteristic dimension (d, e.g. radius for a cylinder) should be much smaller than its length L, i.e. $d << L$. For such particles, starting from the maximum size $r = L$, the $p(r)$ function will increase linearly with decreasing r-values (see Fig. 2.3). The $p(r)$ function is given by (Glatter and Kratky, 1982)

$$p(r) = \frac{2}{4\pi} \int_r^L \int_A \rho_e^2 \, df_1 dS \, dx = \frac{1}{2\pi} \rho_e^2 A^2 (L - r).$$

The slope of the linear region (long distances, close to $r \leq L$, see Fig. 2.3) is proportional to the square of the area of the cross section and the square of the particle electron density, i.e.

$$slope = -\frac{dp(r)}{dr} = \frac{A^2 \rho_e^2}{2\pi}.$$

This essentially implies that it is possible to estimate the cross-sectional area if the electron density ρ_e is known. However, since this linear region is very suscep-

tible to the values taken by the scattering profile at very low angles, special care must be taken with systematic errors (e.g. extrapolation to zero concentration must be made).

2.2.3 Data analysis

2.2.3.1 Direct modelling of the $p(r)$ function

The first class of direct modelling method consisted of building simple (sometimes more elaborate) models of the particle, whose scattering profile and $p(r)$ function could later be evaluated and compared with the experimental ones. The multibody and geometrical bodies approaches both apply this general methodology.

Multibody-type approaches were first introduced by Glatter in the early eighties (Glatter, 1980; Glatter and Kratky, 1982). The $p(r)$ function of a sphere is known analytically (see above), and thus the $p(r)$ function of any group of spherical particles can be analytically calculated (Glatter and Kratky, 1982). *Multibody* is a computer program designed by Glatter and successfully used for a number of applications over the years (Sano et al., 1994). It is mainly helpful for highly symmetric systems, where a simple algorithm for setting the positions of the spheres in the model can be found. However, less conventional shapes can be also generated with considerable effort from the user.

Many authors have addressed the problem of evaluating the scattering profile and $p(r)$ function of complex models made of multiple geometric building blocks (e.g. spheres, cylinders, ellipsoids), assembled into any excluded volume of interest, that would provide a rough model of the particle shape. For the sake of brevity, only a few of these approaches will be mentioned here.

The crudest and oldest version of this method was developed by Kratky and Pils (1972) in the early seventies, and produced scattering curves only for very simple, single geometric shapes (cylinders, spheres). Barnes and Zemb (1988) developed a method for calculating the scattering intensity by directly applying

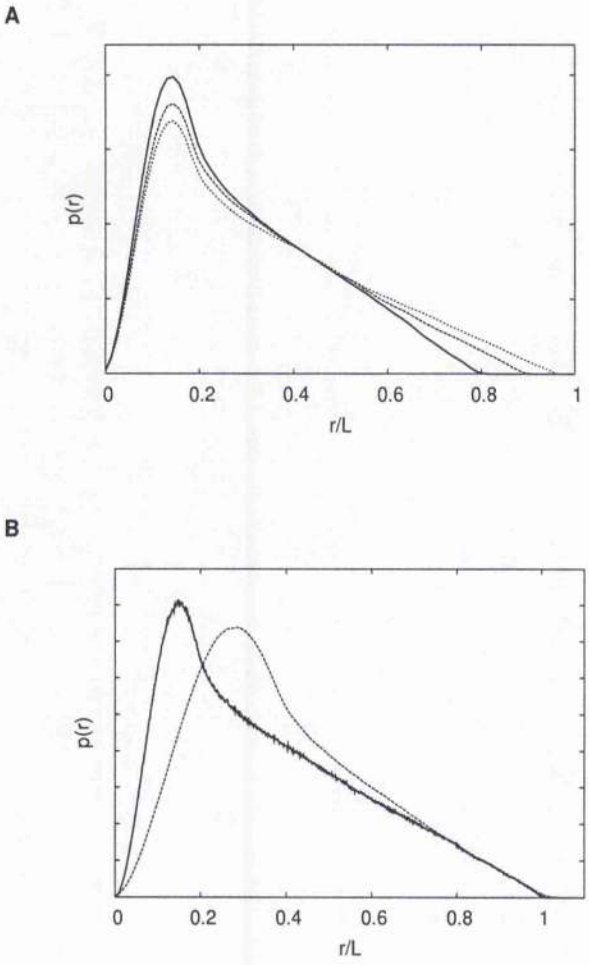

Figure 2.3: Particle distance distribution function $p(r)$ of cylinders with (A) different lengths (L, $r/L = 0.8$, solid line; $r/L = 0.9$, dashed line; $r/L = 0.96$, dotted line) and equal diameter (d), and (B) different diameters ($d = 0.15L$, solid line; $d = 0.3L$, dashed line) and equal lengths (L).

the Fourier transform to the electron density function sampled on a cubic lattice. Later on, Hansen (1990) developed a similar method based on Monte Carlo sampling. In this approach, the volume occupied by homogeneous geometric objects were randomly populated, and then the different objects were combined (applying specific translations and rotations to each of them) in order to produce the final desired model, from which the $p(r)$ function was calculated. From the latter, the scattering curve was obtained by Fourier transformation.

Henderson (1996) extended Hansen's algorithm by making more complicated geometrical objects available (ellipsoidal shells, hollow ellipsoidal cylinders, ellipsoidal helices, triangular prisms, rectangular prisms and semi-ellipsoidal shells). However, the objects were still modelled as if they were homogeneous. Müller et al. (1996) elaborated an algorithm that makes use of the maximum entropy method to find the parameters of single geometrical bodies (such as the diameter of a sphere or the axes of an ellipse) that best fitted the experimental scattering data available. In order to apply this method, it is compulsory to have some prior knowledge of the particle low-resolution shape. Then, the unknown parameters defining the geometrical body representing the particle can be calculated from real-space functions.

Zipper and Durchschlag (2000) developed a method by which an atomic structure was converted to a SAXS model by representing each atom or residue by a sphere of appropriate radius and electron density (similar to the dummy atom models commonly used in hydrodynamic modelling). The $p(r)$ function was then derived from the models, and the scattering curve was calculated by using the Fourier transform.

The second family of methods involves the direct calculation of the scattering intensity from atomic coordinates (derived from x-ray crystallography or NMR when available) of models that could then be compared to available experimental SAXS profiles.

The first approach in this family of methods used variations of what has been coined the 'cube method', which will be described hereafter. For calculating the

scattering intensity of a particle, it is necessary not only to take into account the contribution to the scattering from the atoms comprising the particle, but also the influence of the particle excluded volume. That is to say, the intensity can be written as

$$I(s) = < |F_{particle}(\vec{s}) - \rho_0 F_{buffer}(\vec{s})|^2 >_\Omega, \tag{2.11}$$

where ρ_0 is the buffer average electron density, and $<>_\Omega$ is the integral over the volume occupied by the particle. One of the main difficulties in calculating the scattering curve of macromolecules is the estimation of the second term in equation Eq. 2.11 (the excluded volume contribution) and this has led to several methods whose main difference is the way of tackling this problem. For compact particles with a small surface-to-volume ratio, the best way of dealing with the problem has been said to be the 'modified cube method' (Pavlov and Fedorov, 1983), initially developed in the seventies by Ninio and Fedorov (Fedorov, 1978; Fedorov and Ptitsyn, 1972; Ninio et al., 1972), according to which the particle occupied volume is modelled with densely packed homogeneous cubes (1.3 Å edge length in the original implementation). However, this approach falls short in accounting for the contributions made by the hydration layer, and does not provide satisfying results when dealing with rather open structures, such as nucleic acids.

An improvement of the method was introduced in the early eighties by Pavlov *et al.* (Pavlov and Fedorov, 1983), who basically concentrated on improving the latter approximations (i.e. hydration and openness). Their algorithm is based on the original cube method, but uses a smaller cube edge length and takes into account the different van der Waals radii of the atoms comprising the particle. In addition, it uses a technique developed by Lee and Richards (1971) for calculating the water-accessible particle surface by rolling a ball of 1.4 Å (that imitates a water molecule) on the particle van der Waals surface. However, the authors only considered the rolling ball technique for improving the description of the surface area, and did not attempt to include the full effects of hydration on the scattering curves (they only looked at the effect of bound waters seen in x-ray and NMR structures).

Lattman (1989) proposed a radically new method for estimating the scattering intensity of a given atomic structure. His approach was based on a decomposition into spherical harmonics of the Fourier integral needed to calculate the scattering intensity from the electron density. The main advantage of this method was that the spherical averaging that follows the Fourier transformation of the electron density is drastically simplified if all the terms are written as products of spherical harmonic functions. The author, however, failed to recognize the importance of the hydration layer, and restricted his method to particles with some bound water molecules predicted by x-ray crystallography or NMR. In addition, the excluded volume was accounted for by placing dummy atoms at the positions of the particle atoms, thus reducing the overall accuracy of the method to resolutions of up to 20 Å (that is to say, regions where the momentum transfer is $0 < s < 0.3$ Å^{-1}).

Svergun et al. (1995) extended Lattman's idea and included the effect of a hydration layer with a different electron density (ρ_{wl}) to that of the buffer (ρ_0). The inclusion of this effect modifies the scattered intensity shown in Eq. 2.11, which can be finally reduced to (Svergun et al., 1995)

$$I(s) =< |F_{particle}(\vec{s}) - \rho_0 F_{buffer}(\vec{s}) + \delta\rho_0 F_{hyd-layer}(\vec{s})|^2 >_\Omega, \tag{2.12}$$

where $F_{hyd-layer}(\vec{s})$ is now the scattering amplitude of the water layer and $\delta\rho_0 = \rho_{wl} - \rho_0$. Replacing the Cartesian atomic coordinates by their spherical counterparts, i.e. taking $\vec{r_i} = (r_i, \omega_i) = (r_i, \theta_i, \phi_i)$, the scattering amplitude *in vacuo* can be written as

$$F_{particle}(\vec{s}) = \sum_j^N f_j(s) \exp i\vec{s} \cdot \vec{r_j}, \tag{2.13}$$

where N is the number of particle atoms and f_j is the form factor of atom j. By rewriting the exponential function in 2.13 in terms of Bessel ($j_l(sr)$) and spherical harmonics ($Y_{lm}(\omega)$) functions (Edmonds, 1957; Svergun et al., 1995),

$$\exp i\vec{s}\cdot\vec{r_j} = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l} i^l j_l(sr)Y_{lm}^*(\omega)Y_{lm}(\Omega),$$

one can express the particle scattering amplitude *in vacuo* as

$$F_{particle}(\vec{s}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} F_{lm}^{particle}(s)Y_{lm}(\Omega), \tag{2.14}$$

where F_{lm} are the scattering partial amplitudes

$$F_{lm}^{particle}(s) = 4\pi i^l \sum_{j=1}^{N} f_j(s)j_l(sr_j)Y_{lm}^*(\omega_j). \tag{2.15}$$

In a similar fashion, by replacing the atoms in the structure by dummy atoms, it is possible to rewrite the excluded volume scattering amplitude $F_{buffer}(\vec{s})$, as in Eq. 2.14 but with partial scattering amplitudes given by

$$F_{lm}^{buffer}(s) = 4\pi i^l \sum_{j=1}^{N} g_j(s)j_l(sr_j)Y_{lm}^*(\omega_j), \tag{2.16}$$

where $g_j(s)$ are the dummy atom form factors.

Finally, it is necessary to write the scattering amplitude of the hydration layer in a similar manner. To do this, the border layer can be represented as a two-dimensional angular function $F(\omega)$ as first described by Stuhrmann (Stuhrmann, 1970),

$$\rho_{wt}(r) = \begin{cases} 1 & : \quad F(\omega) \leq r \leq F(\omega) + \Delta \\ 0 & : \quad 0 < r < F(\omega) \\ 0 & : \quad r > F(\omega) + \Delta \end{cases} \tag{2.17}$$

Then, as the partial scattering amplitudes are the Hankel transforms of the real-space radial functions (Svergun et al., 1995), it is possible to write the border layer scattering amplitude as in Eq. 2.14, with partial amplitudes given by

$$F_{wl}^{hyd-layer}(s) = \sqrt{\frac{2}{\pi}} i^l \int_{\omega} Y_{lm}^*(\omega) d\omega \int_{F(\omega)}^{F(\omega)+\Delta} j_l(sr) r^2 dr, \qquad (2.18)$$

By replacing each term in Eq. 2.12 by its form in spherical harmonics (as in Eq. 2.14) and by using Eqs. 2.15, 2.16 and 2.18, Eq. 2.12 can be rewritten in terms of spherical harmonics coefficients, that is to say:

$$I(s) = \sum_{l=0}^{L} \sum_{m=-l}^{l} < |F_{lm}^{particle}(s) - \rho_0 F_{lm}^{buffer}(s) + \delta\rho_0 F_{lm}^{hyd-layer}(s)|^2 >, \qquad (2.19)$$

where the truncation value L defines the resolution of the representation. By taking this approach, the authors have been able to produce a fast and accurate publicly available computer program (CRYSOL) that calculates the scattering curve from a given atomic structure, taking into account the hydration surrounding the particle. Both the hydration contrast (electron density of the hydration layer) and the average displaced solvent volume per atomic group were allowed to vary in order to fit experimental curves. The authors have shown that CRYSOL improves the fit to the experimental data with respect to the other proposed algorithms (based on the cube method) up to $s \leq 0.4$ Å^{-1}, where the inhomogeneously filled excluded volume and the finite number of multipoles start to produce deviations. In this region, cube methods are expected to outperform CRYSOL.

Mariani et al. (2000) developed a hybrid method for multi-subunit particles, for which individual subunit structural atomic information is available. By changing the arrangement of the subunits, a number of tentative atomic models were built and compared with the available experimental data. The overlap between different components of the created structure was accounted for by using the

method of Hansen (1990). The particle shape was represented by spherical harmonics functions (Lattman, 1989; Svergun et al., 1995) and the $p(r)$ function was calculated by Monte Carlo methods, as in Hansen (1990).

2.2.3.2 Particle reconstruction algorithms

The last decade has seen a very substantial increase in other methods to analyse small angle scattering data from biomolecules. As mentioned above, spherical harmonics decomposition was succesfully applied to calculate the scattering curves of structural models, based either on high-resolution atomic structures derived from x-ray diffraction methods and NMR, or on Multibody-like models (Lattman, 1989; Svergun et al., 1995).

Based on this decomposition, Svergun et al. (1996) developed a new algorithm that, instead of using the spherical harmonics coefficients from a model to fit the experimental data, introduced a global search algorithm in parameter-space in order to find the spherical harmonics coefficients that would generate the scattering curve most similar to the experimental one. After finding these coefficients, the program (SASIIA) could reconstruct the particle shape, thus providing for the very first time a way of generating an *ab initio* reconstruction from SAS data without requiring any previous structural information. Computer simulations (Svergun et al., 1996) have demonstrated that, for error-free data, this shape restoration procedure generates unique solutions, even when limited ranges were used in the simulated curves (only if the number of coefficients used is smaller than 1.5 times the number of Shannon channels). The reliability of the shape restoration produced has been shown to improve if information about the particle symmetry is available (Vachette and Svergun, 2000). This program has been employed in a number of applications (Svergun et al., 2000).

Other approaches, based on the same idea, have been implemented since then. In what follows, only one of the methods (DAMMIN) will be described in detail, and some of the different characteristics of the other methodologies available will be briefly mentioned. In DAMMIN, a dummy atom representation (on a

hexagonal lattice) of a sphere of diameter larger than the maximum particle size is created (Svergun, 1999). The dummy atoms can either belong to the particle or the solvent, thus having different scattering lengths. The scattering profile produced by this dummy atom model is evaluated by using the same procedure as in CRYSOL (see section 2.2.3.1). A simulated annealing algorithm is used to find the dummy atom model whose scattering profile is most similar to the experimental one. This program has been shown to generate good results in a number of test cases (Svergun, 1999; Volkov and Svergun, 2003).

DALALGA is very similar in concept to DAMMIN but, instead of using a simulated annealing algorithm, uses a genetic algorithm in order to search the configurational space of all possible conformations (defined again on a hexagonal lattice)(Chacón et al., 2000). The computer program has been tested with a number of different protein scattering profiles (with added noise) simulated from their atomic structures, adequately reproducing their shapes at low-resolution.

GASBOR is the latest of the algorithms developed by Svergun's group (Svergun et al., 2001). It introduces two types of beads, with different scattering lengths, corresponding to the hydration shell and the solvent. In addition, it adds more constraints to the connectivity of the beads so as to recreate a reconstructed particle with sequence-like residue connectivity. Other similar *ab initio* methodologies have been recently developed by Heller et al. (2003).

These reconstruction methods have been applied to several problems in recent years (Koch et al., 2003). However, there are a number of caveats to bear in mind. First, these methods often give different solutions for the same scattering curve (Nöllmann et al., 2004b; Volkov and Svergun, 2003). It is thought that improved data quality (i.e. higher signal to noise ratio) and extended experimental ranges increase the probability of the reconstruction method producing a unique solution. However, it has been recently shown that some shapes (particularly very assymetric or with voids) cannot be reconstructed by these methods (Volkov and Svergun, 2003).

2.2.4 DNA studied by small angle scattering

Müller (1983) developed an improved cube method for the analysis of DNA and RNA molecules. This method was used to test whether DNA structures deduced from x-ray crystallography data agree with the confomation of DNA in solution. By changing the model parameters the authors predicted a different twist angle from the one reported for atomic structures derived from x-ray crystallography. Grassian *et al.* (Grassian et al., 1983) measured scattering curves for small ds-DNA (double-stranded DNA) fragments and showed that they are consistent with those expected for a helix of pitch and radius corresponding to the Watson and Crick models. This study was designed to demonstrate that previously used cylindrical models of DNA are not correct, implying that estimations of, for instance, the thickness of the surrounding ionic cloud based on that approach are misleading. They also reported the existence of intermolecular interactions leading to DNA alignment at concentrations above 15 mg/ml. Hammermann et al. (2000) used small angle neutron scattering to study the changes of the DNA superhelix with salt concentration. They showed that the DNA superhelix diameter varies from 17 to 9 nm when going from 10 to 100 mM NaCl, respectively. The results found in this study contradict the report of a lateral collapse of the DNA superhelix in physiological salt concentrations, as shown by cryo-electron microscopy (Adrian et al., 1990).

2.3 Analytical ultracentrifugation

Under the force of gravity, a suspension of macroscopic particles (solute) with a density higher than the solvent would precipitate to the bottom of the cuvette containing it, creating a sedimentation boundary. In the case of macromolecules, this effect is so small that it is counteracted by diffusion. In ultracentrifugation, a solution containing the macromolecule under study is spun at high speeds ($\approx$ 5 – 60 krpm), and the role of gravity is thus replaced by the centrifugal force. Under these conditions, the macromolecules move to the bottom of the centrifuge

cell as a function of time, causing the depletion of the meniscus and the formation of a clearly defined sedimentation boundary between macromolecular solution and solvent. In analytical ultracentrifugation, this process can be monitored by measuring the concentration of solute (either by absorbance or interference optics) as a function of time and radial distance (r) (see Fig. 2.4A). The cell containing the sample is constructed so that the solute and the solvent solutions are each kept in individual cavities surrounded by quartz plates allowing the passage of light (UV/VIS) required to monitor the solute absorbance (see Fig. 2.4B).

2.3.1 Sedimentation velocity

In sedimentation velocity (SV) experiments, the macromolecule under study is placed in one of the chambers of a two-chamber cell (Fig. 2.4B) and is subjected to a high rotor speed (≈ 50 krpm). The sedimentation of the macromolecule over time is monitored by producing a series of absorbance/interference profiles at different fixed times. At time zero, the solute is homogeneously distributed over the whole r-range. Profiles at later times show the movement of the boundary towards the bottom of the cell (Fig. 2.4C). From these profiles, the sedimentation coefficient of the macromolecule can be deduced. The sedimentation coefficient is defined as (Lebowitz et al., 2002; Svedberg and Pedersen, 1940)

$$s = \frac{u}{\omega^2 r} = \frac{M(1-\bar{v}\rho)}{N_A f} = \frac{MD(1-\bar{v}\rho)}{RT}, \tag{2.20}$$

where u is the observed radial velocity of the boundary midpoint, ω is the radial velocity, M is the molar mass, $\bar{v}$ is the partial specific volume, ρ is the buffer density, f is the frictional coefficient, N_A is Avogadro's number, D is the diffusion coefficient, R is the gas constant and T is the temperature. The equation relating the diffusion and the frictional coefficients $D = RT/N_A f$ was employed to deduce the last term in Eq. 2.20. Sedimentation coefficients are expressed in Svedberg units (S), which are equivalent to 10^{-13} sec.

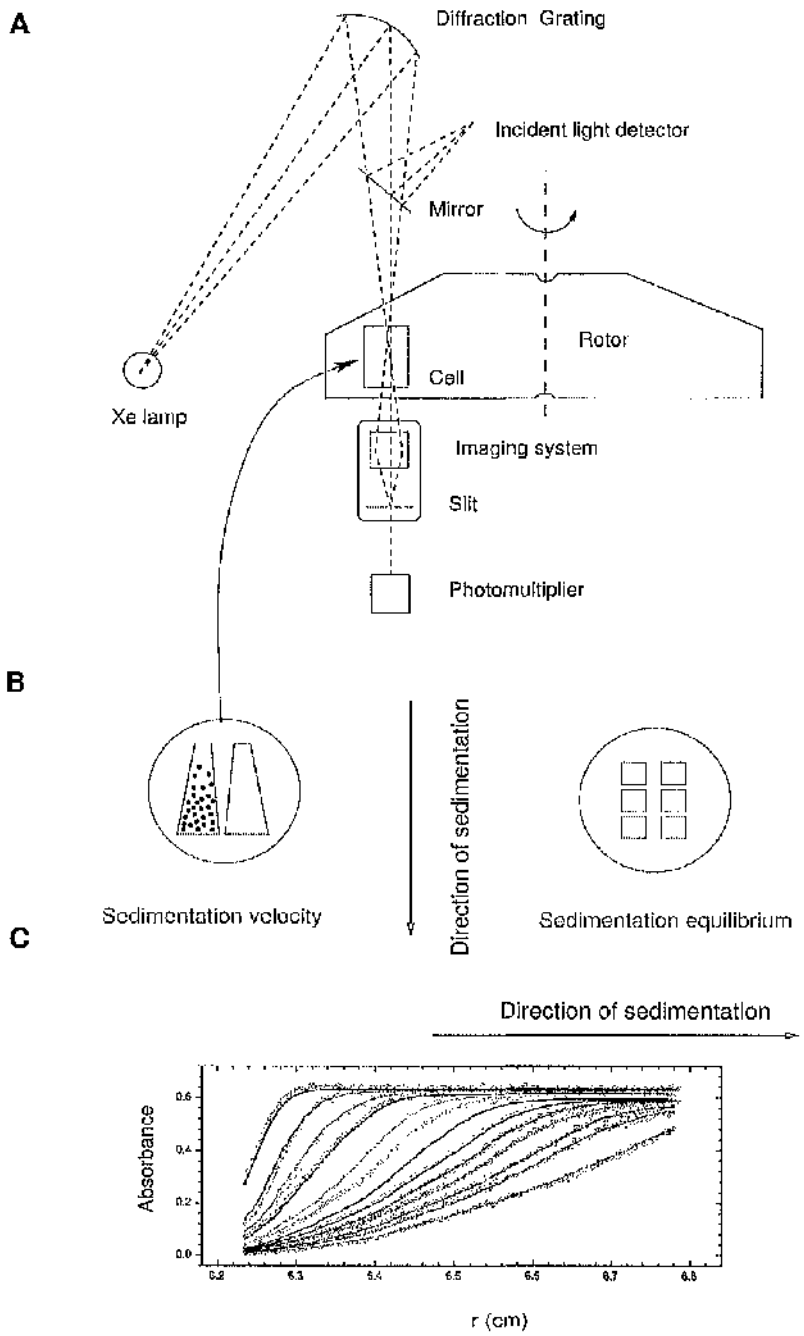

Figure 2.4: (A) Schematic diagram of the optical system in an XL-A analytical ultracentrifuge. (B) Cross-sections of two- and six- channel centrepieces employed for sedimentation velocity and equilibrium experiments. (C) Characteristic sedimentation velocity profiles $a(r,t)$ as a function of the radial position r at different times.

By supposing spherical symmetry, and by using Eq. 2.20 and the Stokes equation to model the frictional coefficient ($f_0 = 6\pi\eta R_0$, where η is the buffer viscosity and R_0 the sphere's radius), it is possible to express the sedimentation coefficient of a spherical particle in water at 20 °C as

$$s_{20,w}^{sphere} = 0.012\frac{M^{2/3}(1-\bar{v}\rho)}{\eta^{1/3}} \tag{2.21}$$

The standard correction to calculate the sedimentation coefficient in water at 20 °C from that at any temperature and buffer is

$$s_{20,w} = s_{T,B}\frac{\eta_{T,B}}{\eta_{20,w}}\frac{(1-\bar{v}\rho)_{20,w}}{(1-\bar{v}\rho)_{T,B}} \tag{2.22}$$

It is worth noting that the maximum value for the sedimentation coefficient ($s_{20,w}$) of a particle of a given mass M will be that of a sphere of the same mass ($s_{20,w}^{sphere}$). As a consequence, any asymmetry in the particle's shape will only decrease its sedimentation coefficient. Macromolecules in solution carry with them differents amounts of bound water mainly as a result of the interaction of polar and nonpolar residues on their surface with water molecules in the solvent. The presence of this 'water shell' (or hydration shell) affects not only the scattering properties of the macromolecule, as described above, but also its hydrodynamic parameters. Particularly, the sedimentation coefficient is considerably reduced by the effect of hydration. This influence will be taken into account when building hydrodynamic bead models (Section 3.3.1).

2.3.1.1 SV data analysis

Myriad methods exist in order to extract the sedimentation coefficient information from the experimental SV profiles. For the sake of brevity and simplicity, only the SV data analysis methods employed in this thesis will be described, namely 'sedimentation coefficient distribution' and 'finite element' analyses.

The general equation describing the sedimentation process (Lamm equation) can be deduced by imposing appropriate boundary conditions on the general transport equation. The Lamm equation is a partial differential equation describing the concentration profile $\chi(r,t)$ as a function of the radial position r and time t and can be written as follows

$$\frac{\partial \chi(r,t)}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[rD\frac{\partial \chi(r,t)}{\partial r} - s\omega^2 r^2 \chi(r,t)\right]. \tag{2.23}$$

The sedimentation coefficients of a group of several simultaneous sedimenting macromolecular species can be obtained by fitting the experimental SV profiles $a(r,t)$ with solutions of the Lamm equation of the form $\chi(r,t)$. As opposed to estimating the average sedimentation coefficient from the observed radial velocity of the boundary midpoint, this procedure makes use of the rich information present in the full sedimentation process. The computer program SEDFIT (Schuck, 2000) employs numerical finite element solutions of the Lamm equations in order to model the sedimentation experimental data. By using this approach, it is necessary to assume a model for the interaction of the macromolecular species present in the solution under study. Possible available models include non-interacting (e.g. n non-interacting species) and interacting (e.g. monomer-dimer association) species. This methodology generates, upon convergence, the true sedimentation coefficients and the relative loading concentrations of each of the species involved in the sedimentation process.

A more sophisticated method known as sedimentation coefficient distribution was developed by Schuck (2000). The SV profiles $a(r,t)$ can be thought of as being composed of a superposition of sub-populations, each with a different s-value (i.e. sedimentation coefficient) and diffusion coefficient D, which would then contribute to the radial- and time-dependent signal $a(r,t)$ with a term proportional to the Lamm equation solution $\chi(s,D,r,t)$ and to its loading concentration. In other words, $a(r,t)$ can be expressed as a linear summation of Lamm equation solutions $\chi(s,D,r,t)$

$$a(r_i, t_j) \approx \sum_{k=1}^{N} c_k \chi(s_l, D_k, r_i, t_j), \tag{2.24}$$

where c_k is the loading concentration of species k. A solution of this set of equations is thus the distribution of sedimentation coefficients c_k, which are usually expressed by a continuous function $c(s)$ (Schuck, 2000). This approach does not require any prior assumption of the number of species or their association behaviour, and it is thus considered to be model-free.

In this project, SV data were analysed analysed by using the sedimentation coefficient distribution approach, which provides an idea of the number of species present in the system, and their sedimentation coefficients and loading concentrations. Secondly, based on these findings, an interaction model was assumed, and the data were re-modelled by using the finite element solutions of the Lamm equation, which provides real and more precise values for the sedimentation coefficients of each of the species present in the system.

2.3.2 Sedimentation equilibrium

Sedimentation equilibrium (SE) analytical ultracentrifugation involves the centrifugation of a macromolecule at a relatively low angular velocity until a state of equilibrium is attained. In this situation, the net flow rate is zero throughout the cell, and the concentration profile is thus stationary due to the compensation of sedimentation and diffusion effects.

Methods for analysis of sedimentation equilibrium data can be divided into model-independent and model-dependent approaches. Model-independent methods are most useful at the initial stages of sample analysis, when the goal is to survey sample behavior, or for comparative analysis of samples that are too complex to be fit directly by model-dependent methods. The equation governing the sedimentation process for a single macromolecular species can be written as (van Holde, 1985)

$$\frac{\partial a(r,\omega,c)}{\partial r} = \frac{\omega^2 r M(1-\bar{v}\rho)}{RT} a(r,\omega,c), \tag{2.25}$$

where M is the molar mass of the macromolecule, $\bar{v}$ is the specific partial volume, R is the gas constant, T is the temperature, and $a(r,\omega,c)$ is the equilibrium concentration profile as a function of the distance to the centre of rotation r, the loading concentration c and the angular velocity ω. Integration of Eq. 2.25 between the meniscus and a point r inside the cell, gives

$$\frac{\partial \ln a(r,\omega,c)}{\partial r^2} = \frac{\omega^2 M(1-\bar{v}\rho)}{2RT}. \tag{2.26}$$

For a single, ideal species, a graph of $\ln a(r,\omega,c)$ *versus* r^2 gives a straight line, with a slope of $\frac{\omega^2 M(1-\bar{v}\rho)}{2RT}$. From this slope, the molar mass M can be determined, if the partial specific volume $\bar{v}$ is known.

A typical SE experiment involves the measurement of the equilibrium concentration gradients of a macromolecular solution at k different concentrations and j angular speeds $a(r,w_1,c_1),..,a(r,w_j,c_1),...,a(r,w_j,c_k)$. Usually, $k \approx 9$ and $j = 3$, which implies a total of 27 independent equilibrium profiles. Model-dependent analysis involve the global fitting of these kj concentration gradients $(a(r,w_1,c_1),..,a(r,w_j,c_1),...,a(r,w_j,c_k))$ to mathematical functions describing various physical models, such as single ideal species, a monomer $\rightarrow$ n-mer self-associating system, or an A + B $\longrightarrow$ C hetero-associating system, using nonlinear least squares algorithms. This is the method of choice for detailed quantitative analysis of sedimentation equilibrium data. This approach gives the best-fit values and the associated statistical uncertainties in the fitting parameters (e.g. molecular mass, oligomer stoichiometry, association constants) and thus provides a statistical basis to discriminate among different putative models. This global fitting approach helps to ensure that a unique solution is obtained and greatly reduces the statistical uncertainty in the parameters. In this project, an implementation of this approach (computer program sedphat; Schuck, 2003b) was employed to globally analyse the sedimentation equilibrium data. Further

information on analytical centrifugation can be found elsewhere (Lebowitz et al., 2002; Richards, 1980; Schuck, 2000, 2003a; van Holde, 1985).

2.4 Fluorescence anisotropy

Spectroscopic methods such as circular dichroism (CD), UV absorption spectroscopy, fluorescence emission and anisotropy provide many advantages for the study of protein-nucleic acid interactions. Steady-state fluorescence is a rapid, sensitive, non-radioactive and non-invasive method that can be applied at reasonably low concentrations (nM range) and that allows for a continuous monitoring of the binding process. The time scale of the fluorescence emission process (ns) allows for the determination of the hydrodynamic parameters of the fluorescent macromolecule, which is employed to study the association/dissociation of macromolecule-macromolecule interactions (Heyduk et al., 1996; Jameson and Sawyer, 1995; Lohman and Mascotti, 1992). It is also important to note that fluorescence based methods allow for the measurement of macromolecule-macromolecule binding in thermodynamic equilibrium, whereas most other methods (such as chromatography, ultrafiltration and gel electrophoresis) rely on separation of free and bound materials, and are thus limited to systems that do not re-equilibrate in the life-span of the experiment (which might be several hours).

In what follows, the physical basis for fluorescence anisotropy measurements will be described. The fluorescence emission polarisation of a fluorescent solution observed at right angles, both to the direction of propagation of the exciting beam and the direction of its electric field vector, can be expressed as

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}, \tag{2.27}$$

where $I_{\parallel}$ and $I_{\perp}$ are the parallel and perpendicular polarised intensities of the fluorescence emission. The fluorescence anisotropy is defined as

$$f = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} = \frac{2P}{3 - P} = \frac{2}{3(1/P - 1/3)}. \tag{2.28}$$

Perrin (1926) determined the mathematical relation existing between fluorescence anisotropy, fluorescence lifetime and rotational diffusion, which can be written as

$$f = \frac{r_0}{(1 + \frac{3\tau}{\rho})}, \tag{2.29}$$

where r_0 is the intrinsic fluorescence anisotropy in the absence of depolarising influences such as rotation or energy tranfer, τ is the excited state lifetime and ρ the Debye rotational relaxation time, which for spherical molecules can be rewritten as

$$\rho = 3\eta V/RT, \tag{2.30}$$

where V is the molar volume, η is the medium's viscosity, R is the gas constant and T is the absolute temperature.

Combining Eqs. 2.29 and 2.30, gives

$$f = \frac{r_0}{(1 + \frac{RT\tau}{V\eta})}, \tag{2.31}$$

which relates the fluorescence anisotropy to the particle volume. Even for non-spherical molecules, an increase in particle volume implies an increase in its fluorescence anisotropy. In addition to depolarisation due to global rotational motion of the macromolecule, one must also take into account the problem of local probe motion (understood as any additional motion of the fluorophore with respect to its point of attachment). In the case of covalently bound fluorophores, these local motions are usually restricted by structural barriers.

Weber (1952) demonstrated that in the case of mixtures, the fluorescence anisotropies of the components can be added:

$$f_{obs} = \sum \alpha_i f_i \tag{2.32}$$

where f_i is the fractional contribution of the ith component to the total fluorescence emission intensity, and r_i is the fluorescence anisotropy of the ith component.

In this project, steady-state fluorescence anisotropy measurements were used to monitor the changes in the DNA (acceptor) that accompany protein (ligand) binding. A titration was performed by adding certain amounts of ligand to a solution containing the labelled acceptor (DNA fragment). For each addition of protein, the FA signal (f_{obs}) was monitored.

2.4.0.1 Analysis of binding curves

For the determination of the binding stoichiometry, the titration is done so that the concentration of acceptor (C_A) is at least 10 times higher that of the dissociation constant of the reaction (K_d), i.e. $C_A \approx 10K_d$ (higher acceptor concentrations make the stoichiometry determination more accurate). Under these conditions, the titration curve takes the shape of the dashed curve in Fig. 2.5. The stoichiometry is then inferred by finding the point of intersection of the extrapolated linear segments (see Fig. 2.5) (Winzor and Sawyer, 1995).

In order to measure the binding constant of a given acceptor-ligand association, the titration is done so that the concentration of the acceptor is close to the dissociation constant (K_d) for the binding equilibrium, in which case the binding curve presents a rounded shape (see dotted curve in Fig. 2.5). Provided that each binding site in the acceptor has the same fractional contribution to the total fluorescence intensity, the fraction of binding sites occupied (f_a) can be written as (Winzor and Sawyer, 1995)

$$f_a = \frac{f_{obs} - f_f}{f_b - f_f}, \tag{2.33}$$

where f_b and f_f are the fluorescence signals when the fractional occupation of the acceptor sites is 1 or 0, respectively. This definition coincides with that of the normalised fluorescence anisotropy (f'_a). In the case of having more than one binding site, the concentrations of free and bound ligand can be written in terms of the fractional saturation (f_a) and the number of binding sites (p), by expressing the binding function (r) as

$$r = f_a p. \tag{2.34}$$

In this case, the concentration of free ligand (C_S) can be written as

$$C_S = \bar{C}_S - r\bar{C}_A, \tag{2.35}$$

where $\bar{C}_S$ is the total concentration of ligand (which changes as the acceptor is titrated) and $\bar{C}_A$ the total concentration of acceptor (which only varies during the titration due to dilution effects). The binding data can now be plotted as the binding function r versus the free ligand concentration C_S. Different models can be employed in order to fit the experimental data, and obtain the association constant K_a and the number of binding sites p.

In the case of a ligand binding to p sites on a single acceptor molecule, and provided that the interaction of acceptor and ligand is the same for all sites and occupancy-independent, the binding function r can be written as

$$r = \frac{pK_a C_S}{1 + K_a C_S}, \tag{2.36}$$

where K_a is the intrinsic association constant ($K_a \equiv 1/K_d$). In the case of two different classes of independent binding sites in the acceptor (with p and q binding sites each), the binding equation can be rewritten as

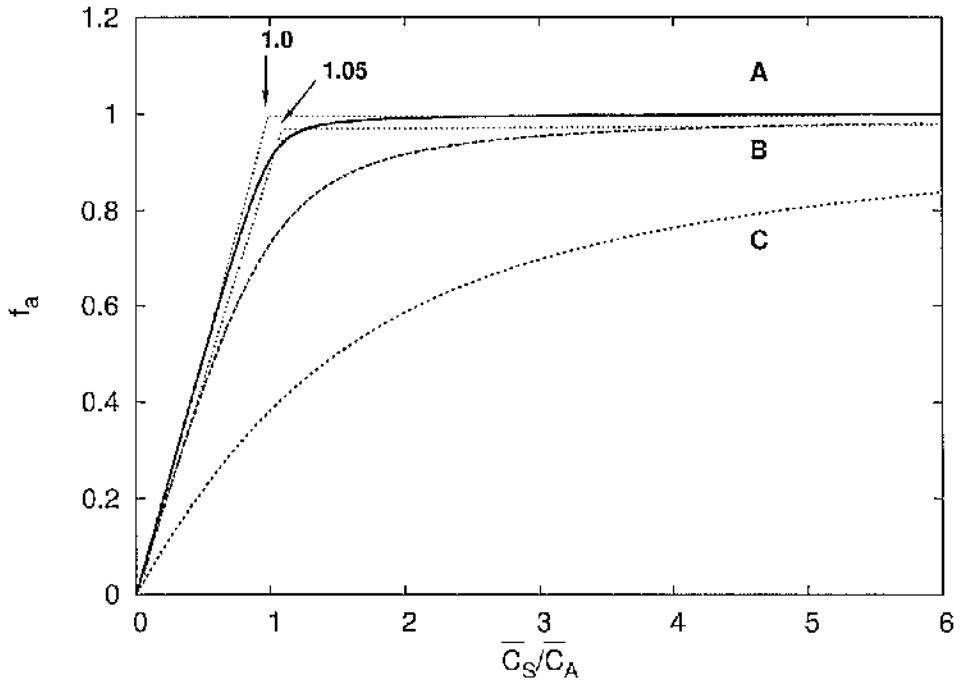

Figure 2.5: Fractional saturation (f_a) *versus* total ligand to acceptor molar ratio ($\bar{C}_S/C_A$). Binding curves were generated by using Eqs. 2.36 and 2.35, with $K_d = 10^{-6}$ M, and acceptor concentrations of (A) $\bar{C}_A = 10^{-4}$ M,(B) $\bar{C}_A = 10^{-5}$ M, and (C) $\bar{C}_A = 10^{-6}$ M. The binding stoichiometry can be ideally evaluated by extrapolating the two linear segments of each curve to their point of intersection, when the acceptor concentration ($\bar{C}_A$) is about one order of magnitude higher than the dissociation constant (K_d). On the other hand, the binding constant is ideally measured when $\bar{C}_A \approx K_d$.

$$r = \frac{pK_a^{(1)}C_S}{1 + K_a^{(1)}C_S} + \frac{qK_a^{(2)}C_S}{1 + K_a^{(2)}C_S}, \tag{2.37}$$

where $K_a^{(1)}$ and $K_a^{(2)}$ are the association constants of the class (1) and (2) binding sites, respectively.

Finally, for p occupancy-dependent but equivalent binding sites, cooperativity can be quantified by using the Hill equation, which can be written as

$$f_a = V\frac{C_S^n}{K_d^n + C_S^n}, \tag{2.38}$$

where V, n, and K_d are parameters of the fit. n is known as the Hill constant, and is a direct measure of cooperativity. At the upper limit n is equal to the number of binding sites. The cooperativity is positive for $n > 1$ and negative for $n < 1$. There is no cooperativity for $n = 1$. A more detailed description of binding curve analysis methods can be found elsewhere (Winzor and Sawyer, 1995).

Chapter 3

Materials and Methods

3.1 Biological experiments

3.1.1 Bacterial growth

The bacterial strains used for transformations, and overexpression of Tn3R and Tn3R mutants, are detailed in Table 3.1.

Bacteria were grown in either Luria broth (LB; 10 g bacto-tryptone, 5 g bacto-yeast extract, 5 g NaCl made up to 1 l with de-ionised water, pH adjusted to 7.5 with NaOH) or in Luria agar (LB with 15 g/l agar). For both liquid culture

Strain (*E.coli*)	Genotype	Source
BL21 (DE3) [pLysS]	*hsd*, *gal*, (λ cI *ts* 857, *ind*1, *sam*7, *nin*5, lac_{UV5}-T7 gene-1), T7 lysozyme expressing plasmid, pLysS)	W. Studier
DS941	AB1157, but *rec*F143, *sup*E44, *lac*ZΔM15 *lac*I^q, galK$^-$.	D.J. Sheratt

Table 3.1: Description of bacterial strains used

Antibiotic	Stock solution	Selective conditions
Kanamycin (Km)	100 mg/ml in H_2O	50 μg/ml
Ampicillin (Ap)	100 mg/ml in H_2O	100 μg/ml
Chloramphenicol (Cm)	100 mg/ml in ethanol	25 μg/ml

Table 3.2: Antibiotics used and selective concentrations

and agar plates, the temperature was kept constant at 37 °C. Agar plates were incubated for a minimum of 12 hours. For liquid cultures, agitation conditions depended on whether cells were being transformed (no agitation) or grown for either protein expression or large scale plasmid DNA preparation (vigorous shaking at $\approx$ 200 rpm). For long term storage, overnight liquid cultures were diluted (1:1 v/v) with 50% glycerol and kept at -70 °C.

The antibiotics employed and their selective conditions, for both liquid cultures and agar plates, are summarised in Table 3.2.

3.1.2 Various other protocols

Cells were made competent and later transformed by using the calcium chloride protocol as described by Sambrook and Russell (2001, vol 1, p. 1.116). Small scale plasmid DNA was extracted and purified by using 3 ml of an overnight culture using the QIAGEN Spin Purification Kit (Mini-prep) according to the manufacturer's instruction manual. Large scale purification of plasmid DNA was performed by equilibrium centrifugation in CsCl-ethidium bromide continuous gradients (Sambrook and Russell, 2001, vol 1, p. 1.65). DNA was purified by ethanol precipitation, as described by Sambrook and Russell (2001, vol 3, p. A8.14).

3.1.3 Oligonucleotides: manufacture, purification and annealing

Oligonucleotides were purchased from either MWG-Biotechnology (Germany) or Sigma-Genosys (UK), then resuspended in TE buffer (10 mM Tris-HCl (pH 8.4), 0.1 mM EDTA).

3.1.3.1 Purification for fluorescence experiments

For fluorescence experiments, a highly pure oligonucleotide solution is required. The protocol employed reflects this need. Oligonucleotide solutions were added to 1 volume of formamide buffer (80% formamide, 100 mM EDTA), heated at 95 ° C for 10 minutes, and finally purified by denaturing polyacrylamide gel electrophoresis (see Section 3.1.4). The bands were visualised by placing the gel on a phosphorescent screen and reflecting UV light from a hand-held UV lamp. Regions with high DNA density were seen as shadows, whereas regions with no DNA were green. This method is less sensitive than others (e.g. 'stains all' staining) but avoids the use of chromophoric markers, which would have to be removed later in the purification process. DNA bands were excised, crushed in 1 ml of TE buffer in a light-tight Nunc tube (Camlab, UK), and finally incubated at 30 °C overnight on a rotating wheel mixer. The supernatant was loaded into a 0.22 μm cellulose acetate filter (Spin-X, Costar) in order to get rid of excess polyacrylamide. The flow-through was then concentrated using a 'Speedivac' evaporator, and its buffer was exchanged for TE buffer with a G25 spin-column (Amersham, UK). Complementary oligonucleotides were annealed by mixing equal molar amounts in TES50 (TE buffer, 50 mM NaCl), and heating to 95 °C for 10 minutes, followed by slow cooling to room temperature.

3.1.3.2 Annealing and purification for SAXS/SANS experiments

Because of the large amounts of sample required (≈ 100 nmol) for SAXS/SASS experiments, the protocol for DNA purification was different from that described in the previous section. High pressure liquid chromatography (HPLC) or denaturing PAGE were employed to pre-purify the oligonucleotides when purchased. Oligonucleotides were then annealed as described before and re-purified by using size exclusion chromatography on two Superose-12 (Amersham Biosciences, Bucks, UK) columns connected in series, at a flow rate of 0.05 ml/min. Fractions were concentrated using 30 kDa or 10 kDa microconcentrators (Millipore, UK). Monodispersity of the sample was monitored by polyacrylamide gel electrophoresis and sedimentation velocity analytical ultracentrifugation (see Section 3.2.2).

3.1.4 Gel electrophoresis methods

Agarose gel electrophoresis, as described by Sambrook and Russell (2001, vol 1, p. 5.4), was employed for purification of long DNA fragments (> 200 base pairs), analysis of Tn3 resolution reactions, or synapse formation assays. Denaturing polyacrylamide gel electrophoresis (Sambrook and Russell, 2001, vol 2, p. 10.11) was employed for the purification of synthetic oligonucleotides. SDS-polyacrylamide gel electrophoresis (Sambrook and Russell, 2001, vol 3, p. A8.40) was employed to assay the purity and concentration of proteins.

3.1.5 Purification of Tn3R and Tn3R mutants

Wild-type and mutant Tn3 resolvases were purified by a procedure based on that described previously (Arnold et al., 1999; Wenwieser, 2001).

The purification procedure for His-tagged mutants was identical to that for wild-type resolvase, apart from the chromatography steps. In this case, a HiTrap

(Pharmacia, UK) Ni affinity column was employed for the purification. The chromatographic procedure was as follows:

- The column was initially washed with 5 column volumes (CV) of filtered distilled water at 1 ml/min.
- One CV of 100 mM $NiSO_4 \cdot 6H_2O$ (Sigma, UK) solution was injected, and the column was again equilibrated with filtered distilled water.
- The column was then equilibrated by flowing 5 CV of buffer A_{his} (50 mM sodium-phosphate buffer (pH 7.6), 8 M urea, 1 M NaCl).
- The sample was injected at 0.5 ml/min.
- The column was re-equilibrated in buffer A_{his} until the optical absorbance at both 260 nm and 280 nm recovered the original values before sample injection.
- A 0-12% gradient of buffer B_{his} (50 mM sodium-phosphate buffer (pH 7.6), 8 M urea, 1 M NaCl, 0.5 M imidazole) was used to wash off impurities and weakly bound species.
- Lastly, a 12-100% gradient of buffer B_{his} was used to elute the protein from the column.

Protein peaks were collected in 1 ml fractions, and analysed by SDS-polyacrylamide gel electrophoresis. Pure fractions were pooled together and dialysed in buffer B_{nat} (20 mM Tris-HCl (pH 7.5), 2 M NaCl, 0.1 mM EDTA, 1 mM DTT), concentrated to approximately 3.5 mg/ml, and finally stored at -80 °C.

Table 3.3 summarizes the different Tn3R mutants purified during this thesis. Tn3R refers to wild-type Tn3 resolvase. NM-resolvase has the following six mutations from the wild-type Tn3 resolvase sequence: R2A E56K G101S D102Y M103I Q105L (Burke et al., 2004). NMh-resolvase refers to NM-resolvase with a 6-His-tag attached to its C-terminus. NM-S10A resolvase has a further mutation S10A with respect to the sequence of NM-resolvase.

Resolvase mutant	Label	Chromatography method	Concentration (mg/ml)
Tn3R	Tn3R-291001	anion exchange	3 mg/ml
Tn3R	Tn3R-150402	anion exchange	3.7 mg/ml
NMh	NM-061202	metal chelating	2 mg/ml
NM	NM-080803	anion exchange	1.5 mg/ml
Tn3R	Tn3R-030204	anion exchange	2.2 mg/ml
NM-S10A	NM-S10A-030204	anion exchange	1 mg/ml
Tn3R	Tn3R-100404	anion exchange	3.5 mg/ml
NM-S10A	NM-S10A-100404	anion exchange	2.9 mg/ml

Table 3.3: Description of proteins purified.

3.1.6 *In vitro* recombination by resolvase

Tn3R recombination reactions were performed in buffer TEM (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10 mM $MgCl_2$). After incubation with resolvase (generally 1/10 volume), all reactions contained final concentrations of 100 mM NaCl and 5% v/v glycerol. A typical reaction (20 μl) contained 0.4 μg of plasmid DNA with 2 μl of diluted resolvase, and was carried out at 37 °C for 1 hour. The reaction was terminated by adding 5% of SDS loading buffer and heating at 70 °C for 5 minutes. Recombination products were analysed by agarose gel electrophoresis. More details can be found in Wenwieser (2001).

3.1.7 Purification of the X-synapse

NM-resolvase and site I DNA were mixed at a molar ratio of 30:1 in TEGGS buffer (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 20% v/v glycerol, 200 mM NaCl) plus 4% w/v ficoll, and the mixture was incubated at 37 °C for 30 minutes. The X-synapse was then purified by size exclusion chromatography in TEGGS buffer, using two Superose 12 columns (Amersham Biosciences, Bucks, UK) connected in series, at a flow rate of 0.05 ml/minute. SV analysis was employed to show that the purified X-synapses were monodisperse (more details in chapter 6). Bandshift assays were performed as described by Sarkis et al. (2001).

3.2 Biophysical methods

3.2.1 Measurement and calculation of densities, viscosities and protein concentrations

Resolvase concentration was estimated by absorbance measurements, by using an estimated extinction coefficient at 280 nm of $\epsilon = 2500$ M^{-1} cm^{-1}(Tn3R) or 3200 M^{-1} cm^{-1} (NM-resolvase), obtained by using the approximations in Pace et al. (1995). Buffer densities were measured by using a density/specific gravity meter (Model DA-510, Integrated Scientific Ltd, Rotherham, UK). Buffer viscosities were calculated from buffer composition, using the viscosity values reported by Matsunaga and Nagashima (1983) and the computer program SEDNTERP (Laue et al., 1992) (see Table 3.4).

Buffer	Temperature [°C]	Density [g/ml]	Viscosity [Poise]	Determination
H_2O	4	0.999	0.0157	computational
	4	0.999	-	experimental
	25	0.997	-	experimental
TE	4	1.001	0.0157	computational
	20	0.999	0.01	computational
TES50	4	1.002	0.0157	computational
	20	1.001	0.01	computational
TES100	4	1.004	0.0157	computational
	20	1.003	0.01	computational
TES1000	4	1.040	0.0157	computational
	20	1.039	0.01	computational
TEGGS	4	1.038	0.0316	computational
	4	1.041	-	experimental
	25	1.036	-	experimental
TEGGS43	4	1.113	0.046	computational
	4	1.114	-	experimental
	25	1.107	-	experimental
TEGGS65	4	1.136	0.048	computational
	4	1.14	-	experimental
	25	1.129	-	experimental

Table 3.4: Densities and viscosities of the buffers employed in this project.

3.2.2 Sedimentation velocity analytical ultracentrifugation

Sedimentation velocity experiments were performed using a Beckman Coulter (Palo Alto, CA, USA) Optima XL-I analytical ultracentrifuge and an AN-60 Ti rotor. The experiments were carried out at a constant temperature, and at a rotor speed of 45000-60000 rpm. A series of scans was obtained with a radial step size of 0.003 cm and interference and/or absorbance optics. Sample concentration depended on the specific experiment. The samples (380 μl) were loaded onto double-sector centrepieces. Sedimentation profiles were analysed with the computer program SEDFIT (Schuck, 2000) (http://www.analyticalultracentrifugation.com), which allows the user to subtract radial and time-independent noise, and to directly model boundary profiles as a continuous distribution of discrete non-interacting species (so-called c(s) analysis, see Section 2.3.1.1). The sedimentation coefficients were also evaluated by using the finite element method to find the sedimentation coefficient that best fitted the Lamm equations (also using SEDFIT).

3.2.3 Sedimentation equilibrium analytical ultracentrifugation

Sedimentation equilibrium experiments were carried out in a Beckman (Palo Alto, CA) Optima XL-A or XL-I analytical ultracentrifuge. The experiments were performed at 4 °C and at rotor speeds of 23000 and 36000 rpm. Scans were obtained at each speed until satisfactory overlay of traces separated by 5 h was obtained, thus indicating the attainment of equilibrium. Six samples of Tn3R in buffer TES1000 (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 1 M NaCl) at concentrations between 0.5 and 5.5 mg/ml were loaded into six-channel centrepieces. Scans were obtained in continuous mode with a 0.001 cm radial step size over the radial range 5.7-7.2 cm. The optical baseline was obtained by doing an overspeed run at 45000 rpm. The SE data were analysed as described elsewhere (Ackerman

et al., 2003), and values for the protein molar mass and dissociation constants were obtained as a result of the analysis.

3.2.4 Fluorescence anisotropy

Steady-state fluorescence anisotropy (FA) measurements were used to monitor the changes in the ligand (DNA) that accompany protein binding. Fluorescence anisotropy measurements were taken with a ThermoSpectronic series 2 luminescence spectroscope (Aminco-Bowman, Illinois, US). The temperature was regulated using a temperature-controlled water bath (Haake, Karlsruhe, Germany). FA signals from fluorescein and tetramethylrhodamine were measured by setting the excitation wavelength at 494 or 530 nm, and the emission wavelength at 518 or 550 nm, respectively. A titration was performed by adding certain amounts of ligand to a solution containing the labelled acceptor (DNA fragment), followed by thorough mixing. After each titration event, the FA signal was left to stabilise for at least 10 minutes, in order to be certain both that the equilibrium temperature was reached and that the binding process was fully completed. The quartz cuvette employed for the essays was thoroughly washed with methanol and a 4 M NaCl solution after each use. The FA signal was normalised by using Eq. 2.33 (Section 2.4.0.1).

3.2.5 Small angle neutron scattering

Small angle neutron scattering (SANS) data were obtained on beam line D11, at the Institut Laue-Langevin (ILL, Grenoble, France) and on the LOQ small-angle diffractometer at the ISIS Spallation Neutron Source (Rutherford Appleton Laboratory, Didcot, UK). Experiments were performed in 1 mm (samples in H_2O-based buffers) or 2 mm (samples with D_2O) optical pathlength quartz cuvettes. Samples were first prepared in normal buffer and then introduced into the D_2O-based buffer by extensive dialysis.

On D11, low-s data ($s = \frac{4\pi \sin\theta}{\lambda}$, where 2θ is the scattering angle and λ is the neutron wavelength) were obtained with a wavelength of 8 Å with the two-dimensional detector located 5 m from the sample, whereas a wavelength of 5 Å and a sample-to-detector distance of 2 m were used at higher s-values. The overall s-range measured was $0.01 < s < 0.15$ 1/Å. Data collection times were of the order of 8 hours per sample or buffer. At D11 neutrons scattered from the sample are detected on a 64×64 cm ^{3}He multi-detector mounted on a moveable trolley within an evacuated detector tube. The position of the beam centre was found by using the program *windet* on a transmission file. A mask of the beam was then produced with the program *rmask*. *rmils* was used for angular integration of the data, converting a 2D image into 1D scattering data. *xpolly* was employed to generate a detector response file from a vanadium sample transmission, and background and H_2O scattering data. Buffer and sample datasets were identically treated and then manually substracted using PRIMUS (Konarev et al., 2003). More details on the data treatment procedure can be found in Ghosh et al. (1989).

At LOQ, a fixed-geometry time-of-flight instrument that uses neutrons with wavelengths ranging from 2 to 10 Å provided an s-range of 0.02-0.2 1/Å. In this case, data collection times were of the order of 15 hours per sample or buffer. A mask file was produced and employed for data reduction by using the program COLLETE. The same program was then employed to calculate the transmission of the sample, to reduce the 2D scattering data to 1D by azimuthal averaging, and to introduce additional corrections. A detailed description of the procedure employed is described by King et al. (2000). Buffer and sample scattering curves were identically treated. The scattering of the buffer was then subtracted from that of the sample, and the difference curves were then scaled for concentration using the computer program PRIMUS (Konarev et al., 2003).

In all cases, the radius of gyration (R_g) was determined by using the Guinier approximation (Section 2.2.1, Guinier and Fournet, 1955).

3.2.6 Small angle x-ray scattering

On Beamline 2.1 at the Synchrotron Radiation Source (Daresbury, UK), beam currents were between 80 and 170 mA, with an electron energy of 2 GeV and a wavelength of 1.54 Å. The camera lengths used in the experiments were: 1.25 m to cover an s-range of $0.03 < s < 0.6$ 1/Å, and 3.25 m for $0.01 < s < 0.2$ 1/Å. The detector was calibrated using a sample of wet rat tail collagen. The experimental data were collected and averaged as 30 x 60 s frames. The data were then normalized to the intensity of the incident beam, and corrected for the detector response by using the computer program XOTOKO (Boulin et al., 1988). The scattering of the buffer was subtracted, and the difference curves were then scaled for concentration using the computer program PRIMUS (Konarev et al., 2003). The final scattering curve was obtained by merging the low-angle region of the low concentration curve with the high-angle region of the high concentration measurement (using PRIMUS), in order to eliminate possible interparticle interaction effects in the low-angle region of the high concentration data. The particle maximum dimension (D_{MAX}) and the particle distance distribution function ($p(r)$) were obtained by using the indirect Fourier transform program GNOM (Svergun, 1992). The radii of gyration (R_g) were determined by using both the Guinier approximation and GNOM.

Experiments on the X33 camera of the EMBL Hamburg outstation at the storage ring DORIS III of the Deutsches Elektronen Synchroton (DESY) were performed by employing procedures similar to those used for analysing the SAXS data at SRS. The sample-to-detector distance used was 2.2 m, at which the momentum transfer range was $0.02 < s < 0.35$ nm^{-1}. In this case, however, the raw data were processed by using the computer program SAPOKO, and similar procedures to those described in the previous paragraph. This beamline contains a 1D linear detector, and so it was not necessary to perform a circular average of the scattering data. A detailed description of the methodologies employed for data analysis specific to this beamline can be found elsewhere (Boulin et al., 1988; Koch and Bordas, 1983).

3.3 Computational methods

3.3.1 Hydrodynamic bead modelling

High-resolution structures, when available, are of utmost utility in producing models with which to interpret experimental hydrodynamic measurements. However, hydrodynamic computer simulations are computationally expensive. Even a hydrodynamic calculation for a small structure (e.g. a 36 base pair DNA fragment) would take of the order of 1 h with a modern Pentium III processor. A simplified approach is thus necessary, if these calculations are to be done on a repetitive basis. The AtoB program (Byron, 1997) can be used to grid a structure of a macromolecule on a cubic lattice, thus producing a so-called bead model. *NewAtoB*, a more extended version of this algorithm, was developed for this thesis (see chapter 7). This computer program was systematically employed to reduce the resolution of high-resolution models when performing hydrodynamic calculations.

In this implementation of *newAtoB*, a cubic lattice of parameter dx was generated, and the centre-of-mass (CM) of the structure was placed at the origin of coordinates. The coordinates of each atom t ($\vec{r}_t$) were used to determine which voxel (i, j, k) in the lattice it occupied. The resolution of the model was reduced by placing a single bead for each occupied voxel (i, j, k). The coordinates of the bead were calculated from the CM of the atoms that it replaced. Similarly, each bead radius was estimated so that its volume was identical to the total volume of the atoms in the voxel. In order to take into account hydration effects, the coordinates of the beads were updated by using the rule $\vec{r}_t = \vec{r}_t \cdot \beta$, where $\beta > 1$ is the so-called outward translation coefficient (see below). Finally, the radii of the beads were updated in order to eliminate bead overlaps and voids in an asynchronous manner (see *PDB2overlap*, Appendix B).

The output of *newAtoB* is in both BEAMS (Spotorno et al., 1997) and HYDRO (García de la Torre et al., 1994) formats. In this study, only HYDRO was employed to calculate the hydrodynamic parameters of reduced models. The

value of the outward translation coefficient β was estimated as follows: (i) the sedimentation coefficient of the original domain structure was calculated by using the computer program HYDROPRO (García de la Torre et al., 2000), which hydrates the macromolecule; (ii) the computer program HYDRO was then used to calculate the sedimentation coefficient of the bead model produced with *newAtoB*; (iii) The value of β was then modified and the process from (ii) restarted, until the sedimentation coefficient calculated from the bead model equalled that of the original high-resolution structure as calculated with HYDROPRO. A further description of this computer program can be found in Appendix B.

The amount of hydration δ of a macromolecule is defined by (van Holde, 1985)

$$\delta = \frac{1}{\bar{v}^0}(V_h \frac{N_A}{M} - \bar{v}), \tag{3.1}$$

where V_h is the volume of the hydrated bead model, $\bar{v}$ is its partial specific volume of the anhydrous macromolecule, M is its molecular weight, $\bar{v}^0$ is the specific volume of the solvent, and N_A is Avogadro's number. As a last step in the generation of a bead model, it was verified that its value of δ was within acceptable limits for protein and DNA (0.3-0.4 g water/g macromolecule; García de la Torre, 2001; Tanford, 1961).

3.3.2 Hydrodynamic calculations

The program HYDROPRO (García de la Torre et al., 2000) was employed to calculate the hydrodynamic parameters of high-resolution models. HYDROPRO computes the hydrodynamic properties of rigid particles from their atomic structure (specified in a protein data bank (PDB) format file). Firstly, HYDROPRO produces a shell model for the surface of the macromolecule by representing it by a collection of minibeads of equal size. The hydrodynamic properties (translational diffusion coefficient, sedimentation coefficient, intrinsic viscosity, and relaxation times) are calculated for shell models with different minibead sizes. The apparent hydrodynamic properties are estimated by extrapolation to zero minibead

size. The effective radius of the atomic elements (AER) is set at the beginning of the calculation and kept constant throughout it. HYDROPRO hydrates an input high-resolution model by adjusting the values of the AER of the different molecules (e.g. aminoacid sidechains).

In this project, the effective radius of the atomic elements was fixed at 2.8 Å for DNA, 3.2 Å for proteins and 3.0 Å for protein-DNA complexes, as suggested by the authors (García de la Torre et al., 2000). All runs were performed using five minibead radii, producing models with 400 to 2000 minibeads. In all calculations, it was verified that the amount of hydration was within acceptable limits for protein and DNA (0.3-0.4 g water/g macromolecule; García de la Torre, 2001; Tanford, 1961).

Similarly, HYDRO (García de la Torre et al., 1994) calculates the hydrodynamic parameters and other solution properties of rigid macromolecules from reduced bead models given the coordinates and radii of the beads comprising the model. HYDRO does not hydrate the model, and thus a properly hydrated model has to be built before using this program (see Section 3.3.1).

3.3.3 Computation of scattering curves from structural models

The simulation of scattering curves, and their fits to experimental data were calculated by using the computer program CRYSOL (for SAXS) or CRYSON (for SANS) (Svergun et al., 1998, 1995). The maximum hydration shell scattering contrast was allowed to vary between 0 and 150 e nm^{-3}, and the maximum excluded volume allowed was increased in order to allow for higher hydration of the DNA samples (compared with protein). The χ-value produced when fitting the scattering intensity of the models ($I(s_i)$) to the experimental data ($I_e(s_i)$) was calculated by:

$$\chi = \sqrt{\frac{1}{N_p - 1} \sum_{i=1}^{N_p} \left[\frac{I_e(s_i) - cI(s_i)}{\delta(s_i)} \right]^2}, \tag{3.2}$$

where N_p is the number of experimental points, $\delta(s_i)$ are the experimental errors, and c is a scale factor (more details are in Section 2.2.3.1 or in Svergun et al. (1995)).

3.3.4 Construction of DNA structures

The computer programs NAMOT (Tung and Carter, 1994) and NAB (Macke and Case, 1998) were employed to build nucleic acid structures from sequence. NAMOT was employed to construct the Holliday junctions, whereas NAB was used to generate straight DNA fragments. Kinked DNA was generated by using the program NAB, which inputs a set of points placed manually along the presumed kinked DNA and interpolates the positions of the base pair origins using a cubic spline function. This approach was similar to that used by Strater et al. (1999).

3.3.5 Superposition of reconstructed models

Superposition of both *ab initio* models and high-resolution structures was performed using the computer program SITUS (Wriggers et al., 1998). Situs is a package for combining multi-resolution data from a variety of biophysical sources, including cEM, electron tomography, SAXS, and x-ray crystallography, using vector quantization techniques. A minimum of 4 codebook vectors were employed for any superposition. Visualisation of results was done by using the program VMD (Humphrey et al., 1996).

Chapter 4

Tn3 resolvase in solution and its interaction with *res*

Summary

The solution properties of Tn3 resolvase (Tn3R) were studied by sedimentation equilibrium (SE) and sedimentation velocity (SV) analytical ultracentrifugation, and small angle neutron scattering (SANS). Tn3R was found to be in a monomer-dimer self-association equilibrium, with a dissociation constant of $K_D^{1-2} = 50$ μM. SV and SANS demonstrated that the low-resolution conformation of dimeric Tn3R in solution is similar to that of $\gamma\delta$ resolvase in the co-crystal structure determined by Yang and Steitz (1995), but with the DNA-binding domains in a rather extended conformation. In addition, the equilibrium binding properties of Tn3R to the individual binding sites in *res* were investigated in solution by employing fluorescence anisotropy (FA) measurements. It was found that site IIL and site III have the highest affinity for Tn3R, followed by site I, and finally by site IIR. Finally, the affinity of Tn3R for non-specific DNA was assayed by competition experiments.

4.1 Introduction

The Tn3 *res* site is composed of three individual subsites made up of inverted copies of an imperfectly conserved common recognition sequence (12 bp for each resolvase), separated by short spacers of slightly different sizes (4 bp in site I, 10 bp in site II and 1 bp in site III). The crossover reaction has been shown to occur at the centre of subsite I (Reed and Grindley, 1981). The distances between the three binding sites vary (22 bp between sites I and II, and 5 bp between sites II and III). In gel-retardation studies, $\gamma\delta$ resolvase has been shown to form three separable protein-DNA complexes with $\gamma\delta$ *res*, implying the binding of one, two or three $\gamma\delta$ resolvase dimers. Under similar conditions, Tn3R was found to form six complexes, corresponding to the binding of individual monomers of Tn3R. The unusual differences between the individual sites in *res* at the sequence level were proposed to trigger structural differences in the way in which Tn3R binds to them (Blake et al., 1995; Grindley, 1994). These different binding modes have been proposed to be due to protein conformational flexibility or to DNA distortion at the centres of the binding sites (this has not been proven yet). The DNA bending caused by the binding of Tn3R to the different binding sites in *res* has also been supposed to differ. In the case of site I, a $\gamma\delta$ resolvase-site I co-crystal structure (Yang and Steitz, 1995) showed that, in the crystal, site I is kinked by 60° towards the major groove. Gel mobility studies on $\gamma\delta$ resolvase have shown that at sites II and III the bend is even more pronounced (Grindley, 2002). DNase cleavage and cyclisation studies on $\gamma\delta$ resolvase showed that the bend in site II is directed towards the minor groove (Grindley, 2002).

Blake et al. (1995) studied the binding of Tn3R to *res* binding site II. It was shown that monomeric Tn3R binds to site II in a two-step process. A first monomer was found to bind to the left end of site II, followed by the cooperative binding of a second monomer to the right end of site II, as opposed to $\gamma\delta$ resolvase which has been found to bind to site II as a dimer. The authors also reported that the Tn3R-site II protein-DNA complex was highly structurally asymmetric, and that this asymmetry was functionally essential.

In this chapter, the oligomerisation behaviour and solution conformation of Tn3R will be studied. Also, the properties of binding of Tn3R to sites I, II, III and non-specific DNA will be investigated in equilibrium and in solution. The chapter is organised as follows. First, the study of the oligomerisation properties of Tn3R in solution with sedimentation equilibrium analytical ultracentrifugation will be presented. Later, the investigation of Tn3R dimeric conformation in solution employing sedimentation velocity analytical ultracentrifugation and small angle neutron scattering will be described. Finally, the study of the binding properties of Tn3R to the individual binding sites in *res* by using fluorescence anisotropy will be introduced.

4.2 Materials and methods

4.2.1 Tn3R preparation

The methods employed for protein overexpression, purification and purity, and concentration assessment can be found in Section 3.1.5.

4.2.2 DNA synthesis and purification

The synthesis and purification methods for the oligonucleotides employed in this study are reviewed in Section 3.1.3. The specific sequences of the oligonucleotides employed are as in Fig. 4.1. Plasmid DNA was purified by the CsCl method as described in Section 3.1.2.

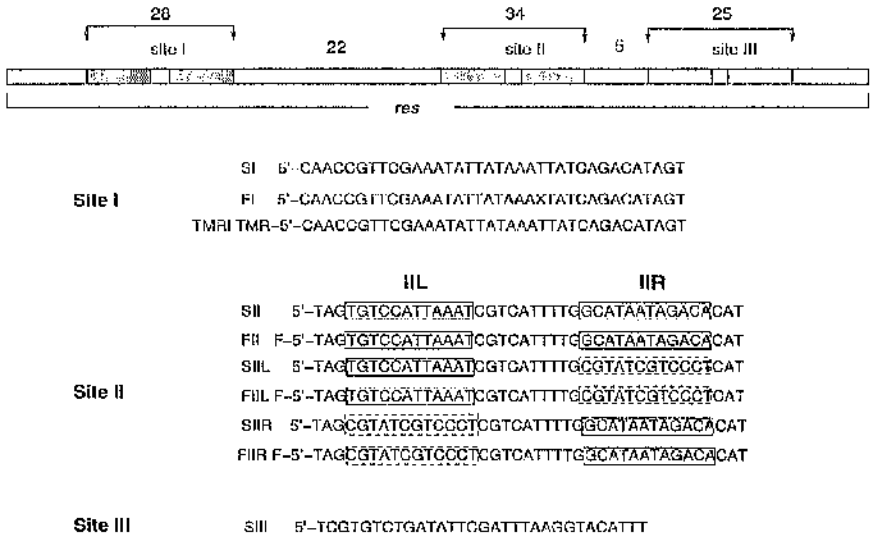

Figure 4.1: The *res* site and the sequences of the top strands of the oligonucleotides employed in this study. F and TMR refer to fluorescein and tetra-methyl-rhodamine, respectively. X refers to a thymine base modified with a C_6 aminolinker labelled with fluorescein at its end. The sequences of the bottom strands were designed to be complementary to the top strands, so as to obtain a final DNA fragment with blunt ends. The left-hand boxes in site II correspond to wild-type site II sequence (SII, FII, SIIL, FIIL, solid line) or to a random sequence (SIIR, FIIR, dashed line). Similarly, the right-hand boxes correspond to wild-type site II sequence (SII, FII, SIIR, FIIR, solid line) or to the same random sequence as employed for SIIL (SIIR, FIIR, dashed line).

4.2.3 Fluorescence anisotropy methods and binding data analysis

Details on the experimental methods employed to make the fluorescence anisotropy (FA) measurements can be found in the Materials and Methods chapter. Solutions of fluorescently labelled oligonucleotide solutions were initially prepared at concentrations ranging from 20 to 50 nM, in buffer TES50 (20 mM Tris-HCl (pH 8.4), 0.1 mM EDTA, 50 mM NaCl). Protein was added from a stock at $\approx$ 1 mg/ml. A thorough description of the methods employed in the data analysis can be found in Section 2.4.0.1.

4.2.4 Small angle neutron scattering

SANS datasets were collected at the D11 station at ILL, France. All experiments were performed at 4 °C, at a protein concetration of 6 mg/ml, in buffer TES1000 (20 mM Tris-HCl (pH 8.4), 0.1 mM EDTA, 1000 mM NaCl). Details on the data collection procedures and analysis of SANS datasets can be found in the Materials and Methods chapter.

4.3 Results and analysis

4.3.1 Behaviour of Tn3R in solution

The solution behaviour of Tn3R was investigated by sedimentation equilibrium (SE) and velocity (SV) analytical ultracentrifugation.

Five Tn3R concentrations, ranging from 5.5 to 0.5 mg/ml, and three rotor speeds, 16000, 23000 and 33000 rpm, were utilised for the SE experiment. A global fit of the 15 datasets was performed using a series of possible models, namely one-species, monomer-dimer equilibrium (1-2 model), and monomer-

dimer-tetramer equilibrium (1-2-4 model). The model that best globally represented the experimental datasets was the 1-2 model, with $K_D^{1-2} = 50$ μM (monomer-dimer dissociation constant). The 1-2-4 model fitted the data as well as the 1-2 model but the predicted total concentration of tetramer, even at the highest concentrations, was negligible (<1%)). A typical and representative fit to the SE data with a one-species and a 1-2 model are shown in Fig. 4.2B. The 1-2 model produces a considerably better fit to the data with smaller and more randomly distributed residuals (Fig. 4.2A). From the dissociation constant obtained from the fit to a 1-2 model, it was possible to calculate the relative distribution of each species assuming monomer-dimer self-association equilibrium (see Fig 4.3). At the highest Tn3R concentration (5.5 mg/ml), the relative amount of dimer would be expected to represent more than 80% of the species present. At the concentrations used for binding experiments (< 1 μM, see section 4.3.2), however, Tn3R would be expected to be $> 95\%$ monomeric.

In order to confirm these findings, SV experiments were performed. At a Tn3R concentration of 5 mg/ml, and a rotor speed of 50000 rpm, the sample behaved as a single species according to c(s) analysis (Fig 4.2C), with an apparent sedimentation coefficient of $s_{w,20} = 2.3$ S. The SV data were also fitted with a one-species non-interacting model by using the finite element method implemented in SEDFIT (Schuck, 2000). In this case, the resulting true sedimentation coefficient was $s^E_{w,20} = 2.4$ S.

The computer program HYDROPRO (García de la Torre et al., 2000) was employed to simulate the sedimentation coefficient of a monomer and a dimer of $\gamma\delta$ resolvase, extracted from the co-crystal structure of the $\gamma\delta$-site I complex (PDB code 1gdt Yang and Steitz, 1995) (Fig. 4.4B). These models were expected to represent the low-resolution structures of monomeric and dimeric Tn3R in solution, at least for the catalytic domains. The values for the simulated sedimentation coefficients for the monomer and dimer species were $s^{\mathrm{MON}}_{w,20} = 1.7$ and $s^{\mathrm{DIM}}_{w,20} = 2.7$ S (with a hydration of 0.4 and 0.36 g water/g protein), respectively. When no DNA is present, the DNA-binding domains of Tn3R are expected to be highly flexible and partially disordered in solution (Rice and Steitz, 1994b), thus

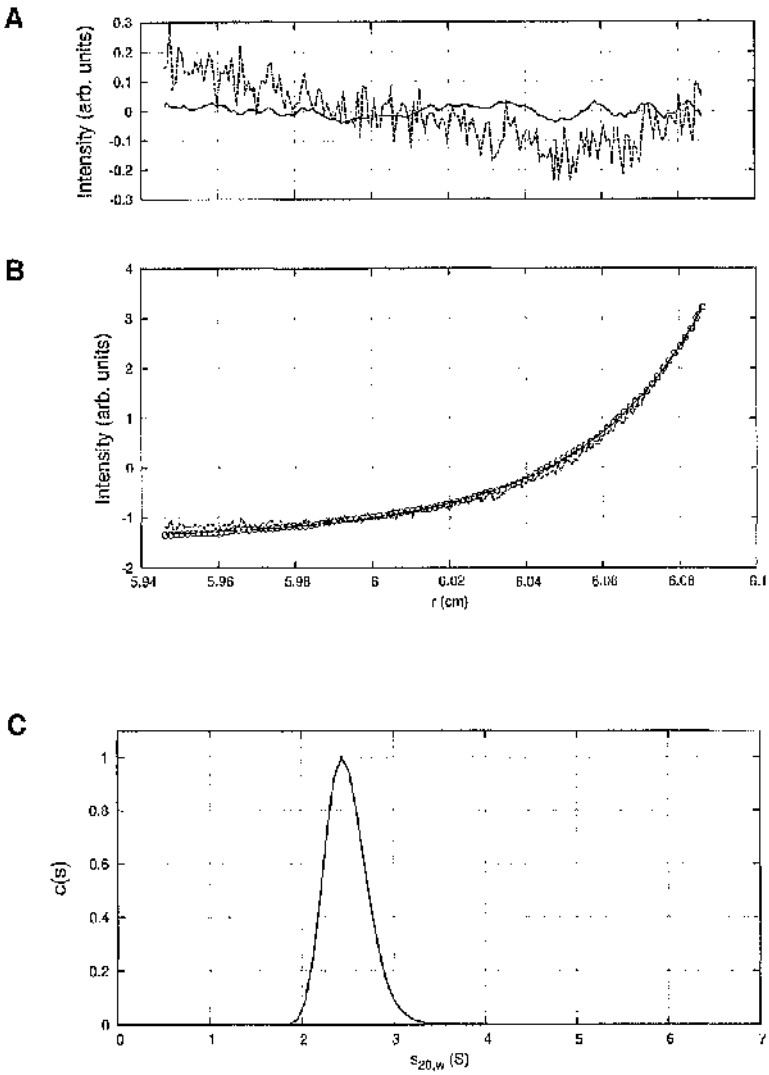

Figure 4.2: Experimental sedimentation equilibrium and velocity data for Tn3R. (B) A representative experimental SE profile for Tn3R at 33000 rpm (open circles) with a fit produced by a 1-2 model (solid line) and a one-species model (dashed line). (A) Residuals of the fits. (C) Size distribution c(s) profiles from SV data analysed with SEDFIT (Schuck, 2000) for Tn3R (solid line).

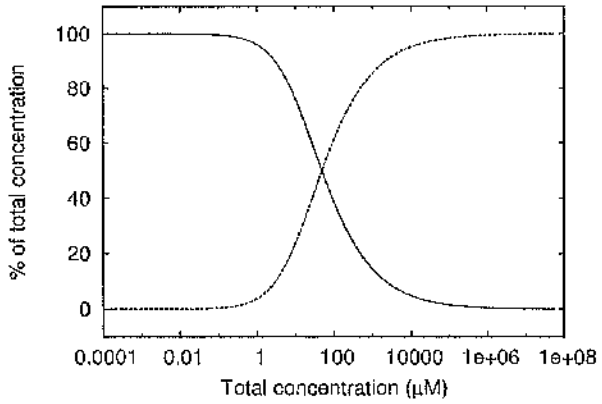

Figure 4.3: Simulated equilibrium monomer (solid line) and dimer (dashed line) fractional concentrations versus total protein concentration, assuming a monomer-dimer self-association equilibrium model with $K_D^{1,2} = 50\ \mu$M.

decreasing the observed sedimentation coefficient. A second model for the conformation of Tn3R in solution (Tn3R-ext, Fig. 4.4C) was generated by extracting the protein component of 1gdt and extending the DNA-binding domains so that they would have a more expanded conformation. The sedimentation coefficient of Tn3R-ext was simulated as above, resulting in a sedimentation coefficient of $s_{w,20}^{\text{ext-DIM}} = 2.31$ S (with a hydration of 0.45 g water/g protein). These findings confirm that Tn3R is essentially dimeric at concentrations higher than 5 mg/ml, and would suggest that the DNA-binding domains of Tn3R in solution are rather elongated.

SANS experiments were performed on Tn3R in solution, at a concentration of 6 mg/ml (Fig. 4.4A, circles). The radius of gyration (R_g) of Tn3R was firstly estimated by using the Guinier approximation to be: 34.7 ± 2 Å. The particle distance distribution function ($p(r)$) was calculated by using the computer program GNOM (Svergun, 1992). The R_g obtained from the $p(r)$ function is 35 ±1.3 Å, in agreement with that from the Guinier analysis. The low-resolution structure of dimeric Tn3R was modelled using the protein content of 1gdt, as described above (Fig. 4.4B). The SANS scattering curve from such a model was simulated by using CRYSON (Svergun et al., 1998), producing a good agreement with the

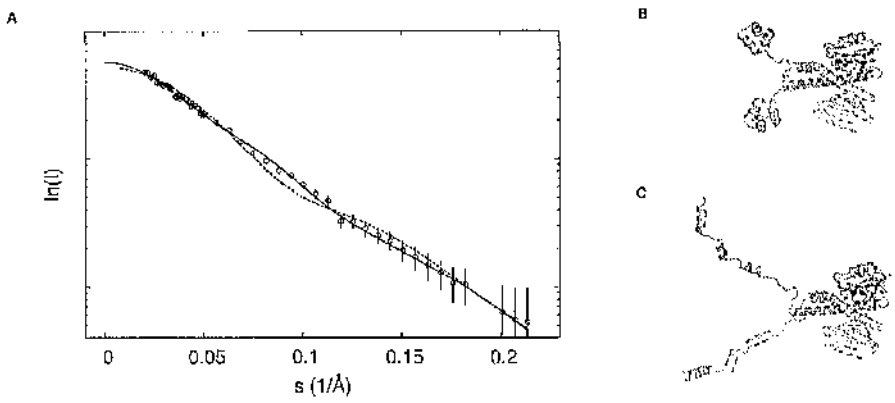

Figure 4.4: SANS of Tn3R in solution. Experimental (circles) and simulated (dashed and solid lines) SANS curves. Structural models for dimeric Tn3R in solution were generated by using the protein component of the co-crystal structure 1gdt (dashed line, Yang and Steitz, 1995) (B) and a modified version in which the DNA-binding domains of Tn3R were extended in an open conformation (solid line, Tn3R-ext, panel C).

experimentally obtained SANS curve (Fig. 4.4A, dashed line, R_g = 30.1 Å). This agreement was further improved by using Tn3R-ext, a structure in which the DNA-binding domains of Tn3R are extended in an elongated configuration (Tn3R-ext, structural model in Fig. 4.4C and fit of the data in Fig. 4.4A, solid line, R_g = 35.7 Å). These findings support the SV data in that Tn3R appears to be dimeric in solution at high concentrations. In addition, they are, in general terms, consistent with the conformation of $\gamma\delta$ resolvase in its co-crystal structure with DNA (1gdt, Yang and Steitz, 1995), and with the solution NMR structure of the catalytic domain of $\gamma\delta$ resolvase (Pan et al., 2001). Finally, they are consistent with the proposal put forward by Rice and Steitz (1994b), by which the DNA-binding domains of resolvase would be expected to be highly flexible and partially disordered in solution. Because of the low resolution of the SV and SANS data presented, other possible structural alterations can not be discarded.

4.3.2 Binding to site I

In order to investigate the equilibrium binding of Tn3R to site I in solution, the fluorescence anisotropy of a fluorescein-labelled site I (FI) was monitored as the Tn3R concentration was increased. The total fluorescence intensity was also monitored as FI was titrated, but there was no systematic change with Tn3R concentration (data not shown). From this, it is assumed that the quantum yield of FI does not change upon complex formation.

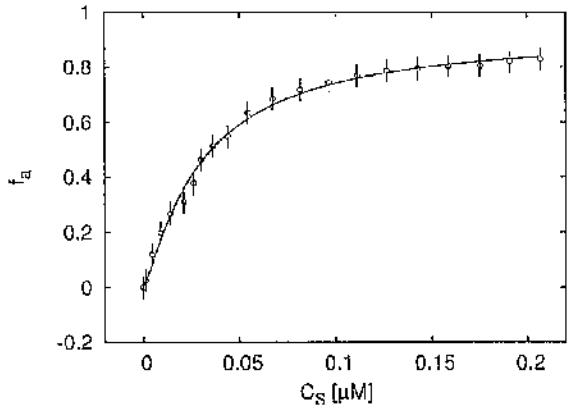

Figure 4.5: Titration of FI ($\approx$ 30 nM) with Tn3R at 20 °C (open circles) and Hill model best fit (solid curve), with $K_D^{FI} = 40$ nM and $n = 1.2$. r represents the binding function and C_S the free Tn3R concentration (see Section 2.4.0.1 and Winzor and Sawyer (1995)). FA signal varied from 0.015 to $\approx$ 0.1.

A series of titrations of FI was performed at different temperatures, ranging from 5 to 35 °C. The resulting binding curves were analysed by using the Hill model (see Section 2.4.0.1). At 20 °C, the equilibrium dissociation binding constant (K_D^{FI}) was 40 nM, and there were signs of weak cooperativity in the binding process ($n = 1.2$, Fig. 4.5). As the temperature increased or decreased from 20 °C, the dissociation binding constant increased, suggesting a decrease in the binding affinity (Fig. 4.6A). The degree of cooperativity (n) also varied with temperature, having its maximum (1.8) at approximately 25 °C (Fig. 4.6B). From these experiments it is possible to conclude that the optimum temperature for

Tn3R binding to site I is approximately 20 °C. The following experiments were performed at a fixed temperature of 20 °C.

The influence of the buffer conditions in the binding of site I to Tn3R was investigated by performing competition measurements. Tn3R was added to a solution containing FI at 20 nM until the fluorescence anisotropy (FA) signal increased to about half its saturation value. This solution was titrated with NaCl, and the FA signal was monitored as a function of NaCl concentration (Fig. 4.7A). From an initial concentration of 50 mM, the FA signal monotonically decreased with NaCl concentration, with half the amount of initial Tn3R-FI complex at $\approx$ 220 mM NaCl. A similar experiment showed a sharp decrease in the amount of Tn3R-FI complex as a function of $MgCl_2$ concentration (Fig. 4.7B), with half the amount of initial Tn3R-FI complex at $\approx$ 20 mM $MgCl_2$. In order to rule out a possible change in the FA of FI due to the interaction of fluorescein with the added ions, the experiments were repeated under the same conditions but without protein, observing no change in the FA signal (data not shown). Moreover, similar experiments using tetramethylrhodamine-site I (TMRI) as acceptor reproduced identical results (data not shown). These data are in agreement with the hypothesis that ionic interactions are important in the binding of Tn3R to site I.

4.3.3 Binding to site II

A similar approach to that employed to measure the equilibrium binding constant of site I was employed for site II. A fluorescein-site II (FII) DNA fragment was synthesized and purified as described in Section 3.1.3. Titrations of FII with Tn3R were performed under conditions identical to those used with FI (Fig. 4.8, open circles). A fit of the binding data with a Hill model produced a dissociation binding constant of $\overline{K_D^{FII}} = 20 \pm 1$ nM, with $n = 1.5$, suggesting a reasonable cooperativity between binding sites IIL and IIR in site II (see Fig. 4.1).

The binding of Tn3R to these sites has been previously reported to be highly asymmetric (Blake et al., 1995). In order to investigate this possibility, two

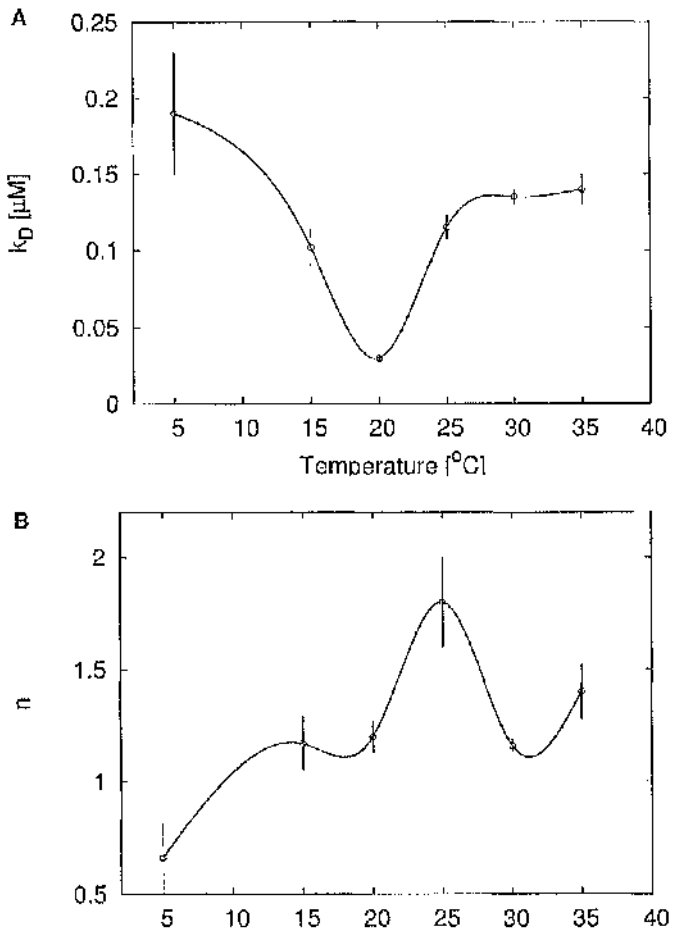

Figure 4.6: (A) Temperature dependence of the dissociation binding constant K_D^{FI} and (B) the degree of cooperativity n for the binding of Tn3R to site I DNA in solution, as observed with FA measurements. K_D^{FI} and n were calculated by fitting a Hill model to the binding curves produced at different temperatures (data not shown). Solid lines represent spline approximations, and serve as a 'guide to the eye'.

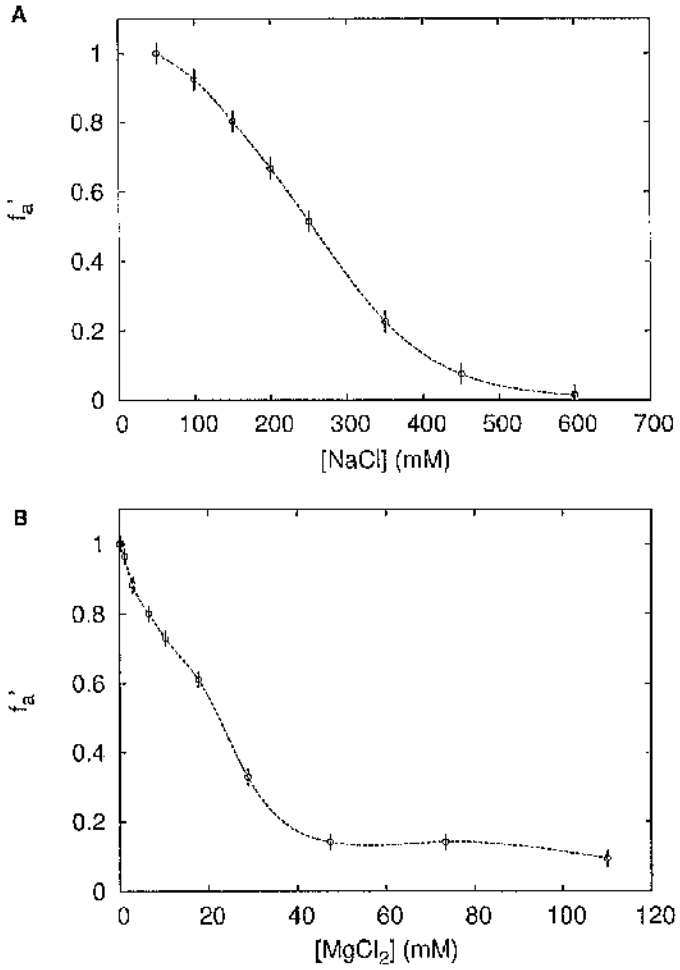

Figure 4.7: Competition titration of the Tn3R-FI complex with (A) NaCl and (B) $MgCl_2$ (see text for details). Plots show the normalised fluorescence anisotropy signal (f_a') as a function of NaCl or $MgCl_2$ concentrations. Dashed curves represent spline approximations to the datapoints and serve as 'guides to the eye'. The FA signal takes values between 0.1 and 0.02.

additional fluorescein-labelled DNA fragments were synthesized. FIIL had the wild-type sequence of the left end of site II and a random sequence replacing its right side end. Similarly, FIIR had the wild-type sequence of the right end of site II and a random sequence replacing its left end side (Fig. 4.1). FIIL and FIIR were titrated with Tn3R as before (Fig. 4.8, filled circles and crosses, respectively). FIIL shows a much higher affinity for Tn3R than FIIR. The FIIL binding data were fitted to a Hill model, resulting in $K_D^{\mathrm{FIIL}} = 50 \pm 3$ nM and $n = 1.8$ (Fig. 4.8, dashed line). A one-binding-site model could not fit this dataset. The FIIR binding data were fitted to a one-binding-site model, with $K_D^{\mathrm{FIIR}} = 140 \pm 10$ nM (Fig. 4.8, dotted line). These results show that the binding of Tn3R to site IIL is about three times stronger than to site IIR.

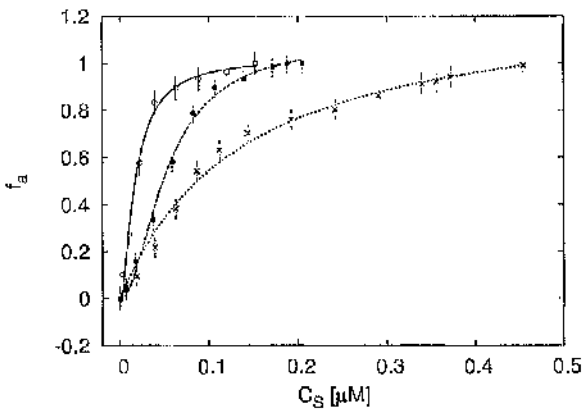

Figure 4.8: Titration of FII (17 nM, open circles), FIIL (18 nM, filled circles) and FIIR (18 nM, crosses) with Tn3R at 20 °C. Binding data were fitted with a Hill model (FII and FIIL) or a one-species model (FIIR), resulting in $K_D^{\mathrm{FII}} = 20 \pm 1$ nM (solid line), $K_D^{\mathrm{FIIL}} = 50 \pm 3$ nM (dashed line) and $K_D^{\mathrm{FIIR}} = 140 \pm 10$ nM (dotted line). FA signal varied from 0.030 to ≈ 0.20 for FII, from 0.050 to ≈ 0.19 for FII and from 0.050 to ≈ 0.19 for FIIR.

Competition experiments were performed to confirm the differences in specificity between sites IIL and IIR. Tn3R was added to a 20 nM solution of FII so that the FA signal increased to half the saturation value. The solution was titrated with unlabelled site II (SII), site IIL (SIIL) and site IIR (SIIR). The decrease in FA signal with increasing added amounts of competitor DNA was due to competition for Tn3R bound to FII. The IC_{50} was defined as the FII to

competitor DNA molar ratio at half the FA starting signal. When SII was used as a competitor, the IC_{50} was 1 ± 0.1 (Fig. 4.9). This showed that the specificities of FII and SII are equal, so the presence of fluorescein in FII does not affect the binding of Tn3R. The IC_{50}s for SIIL and SIIR were 2.0 ± 0.2 and 10 ± 1, respectively. These findings show that the binding of Tn3R to site IIL is stronger than to site IIR, in full agreement with the binding experiments reported above. From the values of the binding constants K_D^{FIIL} and K_D^{FIIR}, the IC_{50}s for SIIL and SIIR would be expected to be $IC_{50}^{\mathrm{pred}} = 2.5$, and $IC_{50}^{\mathrm{pred}} = 7$, in good agreement with the values determined by competition experiments.

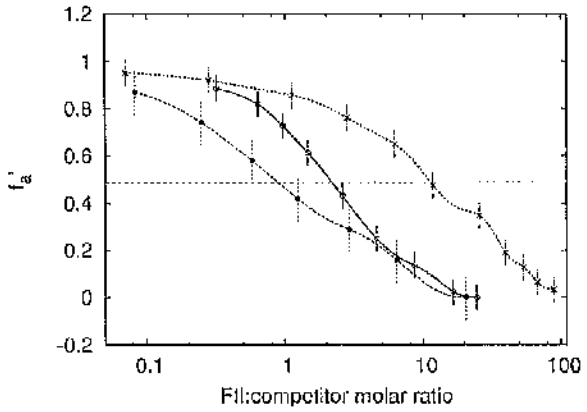

Figure 4.9: Experiments to assess the specificities of binding of SII (filled circles, dashed line, $IC_{50} = 1.0 \pm 0.1$), SIIL (open circles, solid line, $IC_{50} = 2.0 \pm 0.2$) and SIIR (crosses, dotted line, $IC_{50} = 10 \pm 1$) to Tn3R in competition with FII (see text for details). The lines represent spline approximations to the datapoints and serve as 'guides to the eye'.

4.3.4 Relative site specificities

Competition experiments were performed to assess the relative specificities of binding of sites I, II and III. Tn3R was added to a 20 nM solution of FII so that the FA signal increased to half the saturation value. This solution was then titrated with unlabelled site II, site I (SI) and site III (SIII). As above, the decrease in FA signal with increasing added amounts of competitor DNA was due to competition for Tn3R bound to FII. From these experiments, it was determined that SIII is

as specific in binding Tn3R as SII, with both sites having an $IC_{50} = 1.0$ (Fig. 4.10, solid and dashed curves). On the contrary, the specificity of binding of SI ($IC_{50} = 2.0$) was half that of SII or SIII (Fig. 4.10, dotted curve). These results agree well with those reported above for the binding constants of resolvase to each of these sites.

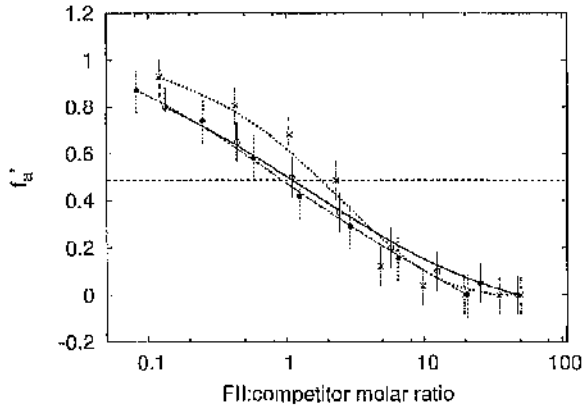

Figure 4.10: Experiments showing the specificities of binding of SI (crosses, dotted line, $IC_{50} = 2.0$), SII (filled circles, dashed line, $IC_{50} = 1.0$) and SIII (open circles, solid line, $IC_{50} = 1.0$) to Tn3R in competition with FII (see text for details). The lines represent spline approximations to the datapoints and serve as 'guides to the eye'.

4.3.5 Binding of Tn3R to non-specific DNA

Competition experiments were performed under conditions similar to those reported above to assess the specificity of binding of Tn3R to non-specific sites in closed supercoiled DNA (plasmid DNA). In this case, however, Tn3R was added to a solution of fluorescein-labelled site I (FI) until half saturation of the FA signal was achieved. This solution was further titrated with three different plasmids: pUC71K, a 3900 bp plasmid with no Tn3R binding site; pCO1, a 2530 bp plasmid with only one binding site for Tn3R (binding site I); and pMM1, a 4600 bp plasmid with six binding sites for Tn3R (two complete *res* sites). The FA signal was monitored as the solution was titrated with competitor DNA (Fig.

4.11). The molar ratio of non-specific binding sites in the plasmid DNA to the total number of specific binding sites was calculated as

$$x = \frac{[C] \times (N_{BP} - N_{bs}) - [C] \times N_{SS}}{[FI] + [C] \times N_{SS}},$$

where $[C]$ and N_{BP} are the molar concentration and number of base pairs in the plasmid DNA, N_{bs} is the length in bp of the specific binding site, $[FI]$ is the molar concentration of FI, and N_{SS} is the number of specific binding sites in the plasmid DNA. x is the molar ratio of non-specific binding sites in the plasmid DNA to labelled sites, assuming that the specific binding sites in the plasmid DNA bind resolvase with the same affinity as FI. When a plasmid containing no specific binding site for Tn3R is employed (pUC71K, $N_{SS} = 0$), the FA signal drops to approximately half its original value when $x = 1000$ (this defines x_{50}), implying that one specific Tn3R binding site in FI had the same binding specificity as 1000 non-specific binding sites in pUC71K (Fig. 4.11, solid line). On the other hand, the existence of specific Tn3R binding sites in the plasmid DNA reduced this number to approximately $x_{50} \approx 100$ (Fig. 4.11, dashed line). These findings imply that when a plasmid DNA contains binding sites specific for Tn3R, the affinity of the non-specific binding sites is enhanced by about ten fold.

4.4 Discussion

In this chapter, the behaviour of Tn3R in solution and its interaction with the different binding sites in *res* has been explored.

Sedimentation equilibrium analytical ultracentrifugation was employed to study Tn3R oligomerisation in solution. It was found that the protein is in a monomer-dimer self-association equilibrium, with a dissociation constant of 50 μM. Accordingly, the protein was found to be mostly dimeric at high concentrations (> 200 μM) and monomeric at low concentrations (< 10 μM).

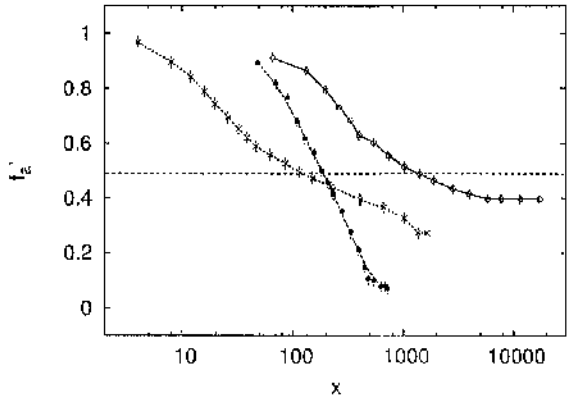

Figure 4.11: Competition experiments to determine the specificities of binding of Tn3R to specific and non-specific sites in plasmid DNA, by using pUC71K (open circles, solid curve, $x_{50} = 1000$), pCOI (crosses, dashed curve, $x_{50} = 100$), and pMM1 (filled circles, dashed curve, $x_{50} = 200$). x is defined in the text.

SV and SANS experiments are consistent with Tn3R being dimeric at high concentrations. The solution conformation of dimeric Tn3R was modelled from the crystallographic and NMR high-resolution data of the catalytic domains of $\gamma\delta$ resolvase (Pan et al., 2001; Sanderson et al., 1990) and a $\gamma\delta$ resolvase-DNA co-crystal structure (Yang and Steitz, 1995). SV and SANS experiments were, in general terms, in agreement with this model. A second model for dimeric Tn3R in solution was built by modifying the torsion angles at the hinges between the three α-helices comprising the DNA-binding domains so that they would acquire a rather extended conformation. The existence of such conformations has been previously proposed by Rice and Steitz (1994b). The sedimentation coefficient and SANS profile simulated from this new model improved the fits to both SV and SANS experimental datasets. Other models, where the DNA-binding domains were modified in a similar fashion but to different extents, were produced. The data simulated from these models fitted the experimental datasets equally well. In accordance, the experiments presented in this chapter are consistent with an extension of the DNA-binding domains with respect to their positions in 1gdt (Yang and Steitz, 1995), but are not of high enough resolution as to clearly define their final configurations. This extended conformation could help the protein in

the initial stages of binding to a prospective DNA binding site. Moreover, it could be important for the protein to be able to adopt different configurations when binding to binding sites with very different geometries, such as those in *res*. This ability of the DNA-binding domains in Tn3R to adopt different conformations might have a role in the assembly of the synaptosome, as previously proposed by Blake et al. (1995).

Fluorescence anisotropy was employed to study the binding of resolvase to the three binding sites in *res*, namely sites I, II and III. It was shown that the highest binding affinity of Tn3R to site I is at 20 °C, resulting in a dissociation binding constant of $K_D^{FI} = 40$ nM. At this temperature, there was only a weak degree of cooperativity ($n \sim 1.2$) so that, in practical terms, the binding of protein monomers to the two binding sites in site I can be considered as independent events. Previous gel-shift assays studying the binding of Tn3R to site I revealed a high cooperativity of binding of Tn3R to site I (Bednarz et al., 1990; Grindley et al., 1982; Watson et al., 1996). This conclusion was reached from the observation that, in such experiments, the site I-dimeric Tn3R complex is much more abundant than site I-monomeric Tn3R complexes. However, gel-shift assays can artefactually stabilise some of the complexes in a mixture, and thus not reflect their real equilibrium concentrations in solution.

The presence of cations in the buffer (Na^+ or Mg^{2+}) was shown to greatly affect the binding of Tn3R to site I, reflecting the existence of an electrostatic component in their interaction.

The equilibrium binding of Tn3R to site II in solution was investigated in a similar fashion. In this case, Tn3R was shown to have very different affinities and specificities for the two binding sites in site II. The binding of Tn3R to site IIL was shown to be at least three times stronger and more specific than to site IIR. These results agree well with those reported by Blake et al. (1995), who reported an asymmetric binding of Tn3R to site II from gel-shift assay data. The authors hypothesised that the higher affinity of Tn3R for site IIL might be related to its need to assemble at *res* in an organised manner, to prepare the complex for its role in the recombination reaction.

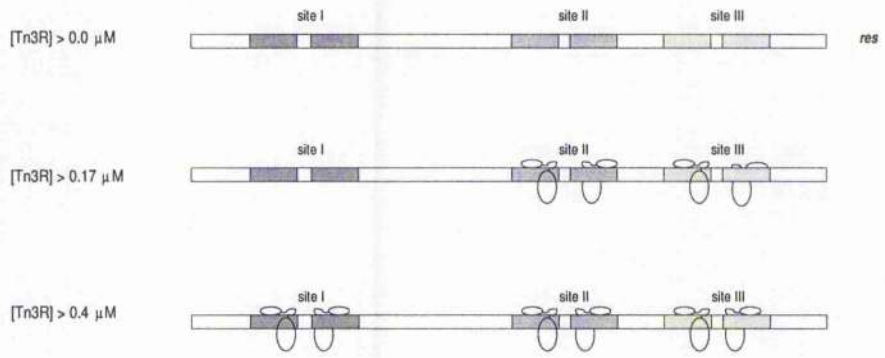

Figure 4.12: Model for the equilibrium binding of Tn3R to the different binding sites in *res*. Binding site saturation is assumed at a protein concentration ten times higher than the corresponding dissociation binding constant and a very low DNA concentration with respect to that of Tn3R.

Binding of Tn3R to site III was shown to be of the same affinity as that to site II, but twice that to site I. Based on this result, assuming binding saturation when the protein concentration is $10 \times K_D$, it is possible to design a simplified model for the occupancy of the sites in *res* at different Tn3R concentrations (Fig. 4.12). According to this model, at the lowest concentration Tn3R is mainly bound to site III and site II. Because of the high cooperativity of binding of site II, both sites IIL and IIR are occupied at this first stage. At larger concentrations, site I also becomes occupied. In the latest, and arguably most accepted, structural model for the architecture of the synaptosome (Section 1.8) (Sarkis et al., 2001), it has been proposed that the resolvase monomer bound to site IIR is at the centre of the accessory sites synapse. It could be proposed that, because of the weaker interaction between Tn3R and site IIR, the monomer bound to this site would be more likely to unbind *res* should any structural rearrengement is required in the assembly of the synaptosome at the centre of the accessory sites synapse. This model, although vague, agrees with all the binding data for Tn3R presented in this chapter and previously reported in the literature (Blake et al., 1995; Grindley et al., 1982). The different affinities of Tn3R to sites I, II and III in *res* might be also associated with the role of *res* as a regulatory sequence in the expression of the correct *in vivo* levels of Tn3R and Tn3 transposase.

Last but not least, the specificity of binding of Tn3R to non-specific supercoiled DNA was assayed. It was found that the specificity of binding of site I is equivalent to that of approximately 1000 non-specific binding sites. From this figure, it can be estimated that the dissociation binding constant of Tn3R to non-specific plasmid DNA is ≈ 40 μM. Hence, in genomic DNA, a monomer of Tn3R would be expected to be bound in equilibrium to a region of 1000 non-specific bp per monomer bound to a specific binding site. When plasmid DNA containing binding sites specific for Tn3R were employed, this figure showed a 10 fold reduction, even though the specific binding sequences in the plasmid DNA were considered as having an affinity to Tn3R equal than that of site I DNA. Two possible mechanisms could account for this difference. As a first possibility, multiple Tn3R monomers could be self-oligomerising at the specific binding sites in plasmid DNA, thus increasing the effective observed affinity of non-specific sites.

Alternatively, the specific binding sites could be functioning as nucleation or landing sites for Tn3R, thus indirectly increasing the observed specificity of non-specific sites. It is so far unknown whether this could be a general phenomenon in the interaction of many DNA-binding proteins with DNA. In any case, both models are able to explain these findings, and more experiments, outwith the scope of this thesis, would be required to refute or confirm them.

Chapter 5

Low-resolution reconstruction of a synthetic DNA Holliday junction

5.1 Summary

The low-resolution solution conformation of a Holliday (or 4-way) DNA junction was studied by using small angle x-ray scattering (SAXS), sedimentation velocity (SV), and computational modelling techniques. The scattering data were analysed in two independent ways: firstly, by rigid body modelling of the scattering data using previously suggested models for the Holliday junction (HJ), and secondly, by *ab initio* reconstruction methods. The models found by both methods agree with experimentally determined sedimentation coefficients and are compatible with the results of previous studies using different techniques, but provide a more direct and accurate determination of the solution conformation of the HJ. These results show that addition of Mg^{2+} alters the conformation of the HJ from an extended to a stacked arrangement. The contents of this chapter are based on published material (Nöllmann et al., 2004b).

5.2 Introduction

DNA recombination was originally thought of as a means for creating genetic diversity. However, more recently, its fundamental importance in the repair of double-strand breaks and for replication through DNA lesions has been established (Cox et al., 2000; Haber, 2000). Holliday (Holliday, 1964) first proposed that a 4-way junction (so-called Holliday junction, HJ) is generated as an intermediate in recombination when DNA strands cross over between homologous duplexes. HJs are substrates for proteins that induce structure-specific cleavage (White et al., 1997) or branch migration (Rafferty et al., 1996) and are also intermediates in site-specific recombination by the integrase family of recombinases (Kitts and Nash, 1987; McCulloch et al., 1994). Proteins involved in homologous and site-specific recombination bind specifically to their substrate HJs (Kitts and Nash, 1987; McCulloch et al., 1994; Schwacha and Kleckner, 1995) in a manner dependent not only on the HJ sequence but also on its tertiary structure. This makes the study of the three-dimensional structure of HJs essential to pave the way towards a deeper insight into these fundamental biological processes.

The first 4-way junctions studied were cruciform structures formed by inverted repeats in supercoiled DNA (Gellert et al., 1979; Lilley, 1980; Panayotatos and Wells, 1981). These are, however, impractical for most structural studies since the cruciforms are stable only when within topologically constrained circular DNA. Cloned or synthesised short DNA segments were later used to build more stable 4-way junctions by hybridising sequences that are incapable of extensive branch migration (Bell and Byers, 1979; Bianchi et al., 1989; Kallenbach et al., 1983). Chemical probing (Furlong and Lilley, 1986; Gough et al., 1986) and NMR (Wemmer et al., 1985) showed that these 4-way junctions were stable, and that full base-pairing was preserved.

Early attempts to model the structure of the HJ involved pairwise coaxial stacking of pairs of helical 'arms' (Calascibetta et al., 1984; Sigal and Alberts, 1972; Sobell, 1972), and most placed the double-stranded helices side by side, with parallel alignment of the continuous strands. In the mid-eighties, it was noted

(Gough and Lilley, 1985; Kosak and Kemper, 1990) that a HJ had an anomalously low mobility in polyacrylamide gel shift assays, suggesting the existence of a sharp kink or bend at the junction. The mobility was also shown to depend on the added counterion concentration, leading to the conclusion that, in the presence of counterions (eg Mg^{2+}), the HJ has an 'X' shape; the helical arms are stacked in coaxial pairs which are rotated with respect to each other. This arrangement maximizes base pair stacking interactions, and reduces steric and electrostatic interactions between the stacked arms. In the model, the continuous strands show an almost anti-parallel arrangement (Duckett et al., 1988; Lilley and Clegg, 1993). In contrast, in the absence of metal ions, data from gel electrophoresis and other techniques (Duckett et al., 1988) clearly suggest an extended square planar configuration of the arms. Chamberlain et al. (1998) used SANS to study the structure of a synthetic HJ sandwiched between two tetrameric *Mycobacterium leprae* RuvA complexes. The authors reported no change in the radius of gyration of the HJ with Mg^{2+} concentration.

In this chapter, the low-resolution structure of a HJ in solution in the presence and absence of Mg^{2+} will be determined by using small angle x-ray scattering (SAXS) and sedimentation velocity (SV). Two independent methods were implemented in order to analyse the scattering data, and hydrodynamic modelling was used to validate the models on the basis of their sedimentation coefficients. This study supports an extended square planar conformation in the absence of Mg^{2+}, but a stacked structure in the presence of Mg^{2+}.

5.3 Materials and methods

5.3.1 Sample preparation

The oligonucleotides forming the 4-way DNA junction were purified and annealed in TE buffer (10 mM Tris-HCl (pH 8.4), 0.1 mM EDTA) as described in Section

3.1.3.2. The sequences of the oligonucleotides (the same as those used by Fogg et al. (2001) in their FRET studies) are shown in Fig. 5.1.

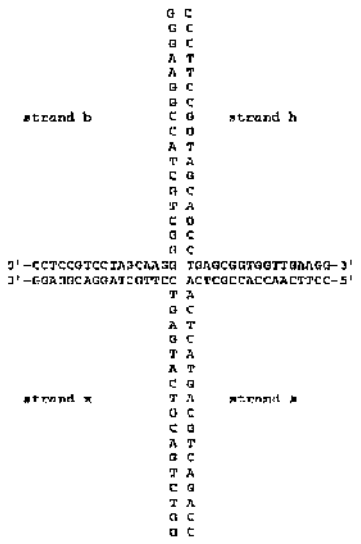

Figure 5.1: Sequence of the synthetic HJ oligonucleotides used in this study.

5.3.2 Sedimentation velocity and hydrodynamic modelling

Sedimentation velocity experiments were performed as described in Section 3.2.2. The experiments were carried out at 10 °C in TE buffer at a rotor speed of 60000 rpm. Both absorbance and interference optics were used in separate experiments. For the experiments performed with absorbance optics, a series of 40 scans, 6 minutes apart, was obtained for each sample, in continuous mode with a 0.003 cm radial step size and three averages. For the experiments carried out with interference optics, a series of 600 scans, 1 minute apart, was obtained with the same radial step size as before. All experiments were carried out with a HJ concentration of 800 nM (unless stated otherwise). The samples (380 μl) were

loaded into double-sector centrepieces. Sedimentation profiles were analysed with the computer program SEDFIT (Schuck, 2000) (see Section 2.3.1.1).

The program HYDROPRO (García de la Torre et al., 2000) was employed to calculate the hydrodynamic parameters of high-resolution models constructed with the program NAMOT (Tung and Carter, 1994) (see Sections 3.3.2 and 3.3.4 for details).

5.3.3 Small angle x-ray scattering

Small angle x-ray scattering (SAXS) experiments were performed on the X33 camera of the EMBL Hamburg outstation at the storage ring DORIS III of the Deutsches Elektronen Synchroton (DESY). The scattering curves were measured at two sample concentrations, 25 and 100 μM (see more details in Section 3.2.6).

The program GASBOR was employed for the reconstructions presented in this chapter (see more details in Section 2.2.3.2). The HJ investigated in this study contains 17 bases in each arm, or 68 Watson-Crick bases in the whole particle. Models with 280 or 840 beads were generated by using the *ab initio* reconstruction program GASBOR (Svergun et al., 2001). These models contained 2 or 6 beads per base.

5.4 Results and analysis

5.4.1 Sedimentation velocity

SV was employed in order to assess the monodispersity of the purified sample and to complement the low-resolution SAXS-based modelling with an independent technique. Solutions of HJs over a range of concentrations (800 nM to 8 μM) in a variety of buffers were examined with SV. Absorbance versus radius distributions at different times of the sedimentation at 60000 rpm for the 0 mM

Mg^{2+} HJ sample and fits to the experimental data produced using c(s) analysis and residuals of the fits are shown in Fig. 5.2. When studied at low concentration (800 nM), the HJ displays a qualitatively different behaviour depending on the presence or absence of Mg^{2+} (see Fig. 5.3 A). At 0 mM Mg^{2+}, the c(s) analysis clearly shows two peaks, corresponding to two distinct species, with apparent sedimentation coefficients of 2.2 and 3.9 S. In contrast, the same sample behaves as a single species when 50 mM NaCl is added to the solution, or when the sample concentration is increased by ten fold (Fig. 5.3 B). It is possible to conclude from these observations that, in the absence of any metal ions from the buffer, and at low concentrations, the HJ structure is not stable and partially dissociates into a secondary species whose sedimentation coefficient is consistent with that of two partially annealed oligonucleotides. The HJ structure is stabilised by the addition of either Mg^{2+} or Na^{+} to the buffer, as shown in Fig. 5.3 A and B. An increase in HJ concentration, without counterions, also produces a single peak in the size distribution analysis (Fig. 5.3 A, dashed curve), by shifting the equilibrium between the different species towards that with higher sedimentation coefficient (and thus probably higher mass, presumed to be the HJ). In the presence of 10 mM Mg^{2+} at 800 nM HJ concentration, a single species (presumed to be the stacked form of the HJ; see below) is observed. At this Mg^{2+} concentration, c(s) analysis yields identical results to the measurements at low or high sample concentrations, in the presence or absence of NaCl (data not shown).

The apparent sedimentation coefficients for the 0 mM and 10 mM Mg^{2+} samples are 3.9 S and 4.3 S (the centres of the peaks in Fig. 5.3 A). Fits using finite element analysis yield true sedimentation coefficients of 3.9 and 4.4 S, and masses of 41 and 40 kDa respectively (the mass of HJ calculated from its composition is 41.5 kDa, without taking counterions into account), for the same samples (fits not shown). These experiments confirm that the method employed for sample purification produces a highly homogeneous HJ sample at concentrations higher than 8 μM (for SAXS experiments the sample concentration was 100 μM). In addition, these results confirm that the HJ undergoes a conformational change to a less extended structure (ie with a larger sedimentation coefficient) upon addition of Mg^{2+}.

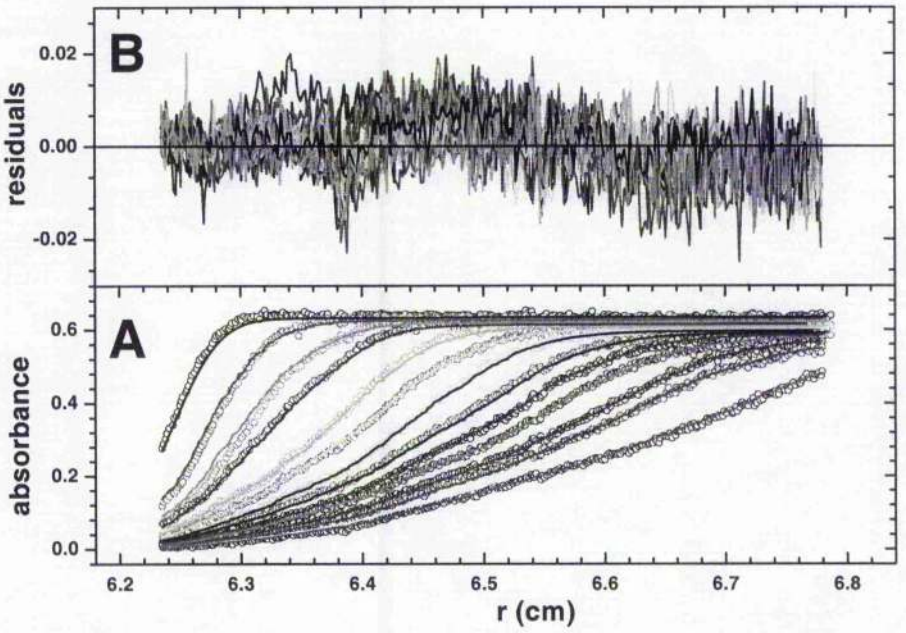

Figure 5.2: Absorbance profiles of the sedimentation of HJ at 10 mM Mg^{2+}. The absorbance distributions were scanned at 260 nm, at time intervals of 6 minutes, for a total time of 300 minutes. (A) Experimental (*circles*) and best-fit (*solid lines*) distributions constraining the sedimentation coefficients to $s_{w,20}(1) = 3.9$ S and $s_{w,20}(2) = 2.2$ S, with relative concentrations of 0.7 and 0.3, respectively. (B) Residuals of the fit with a rmsd of 0.005.

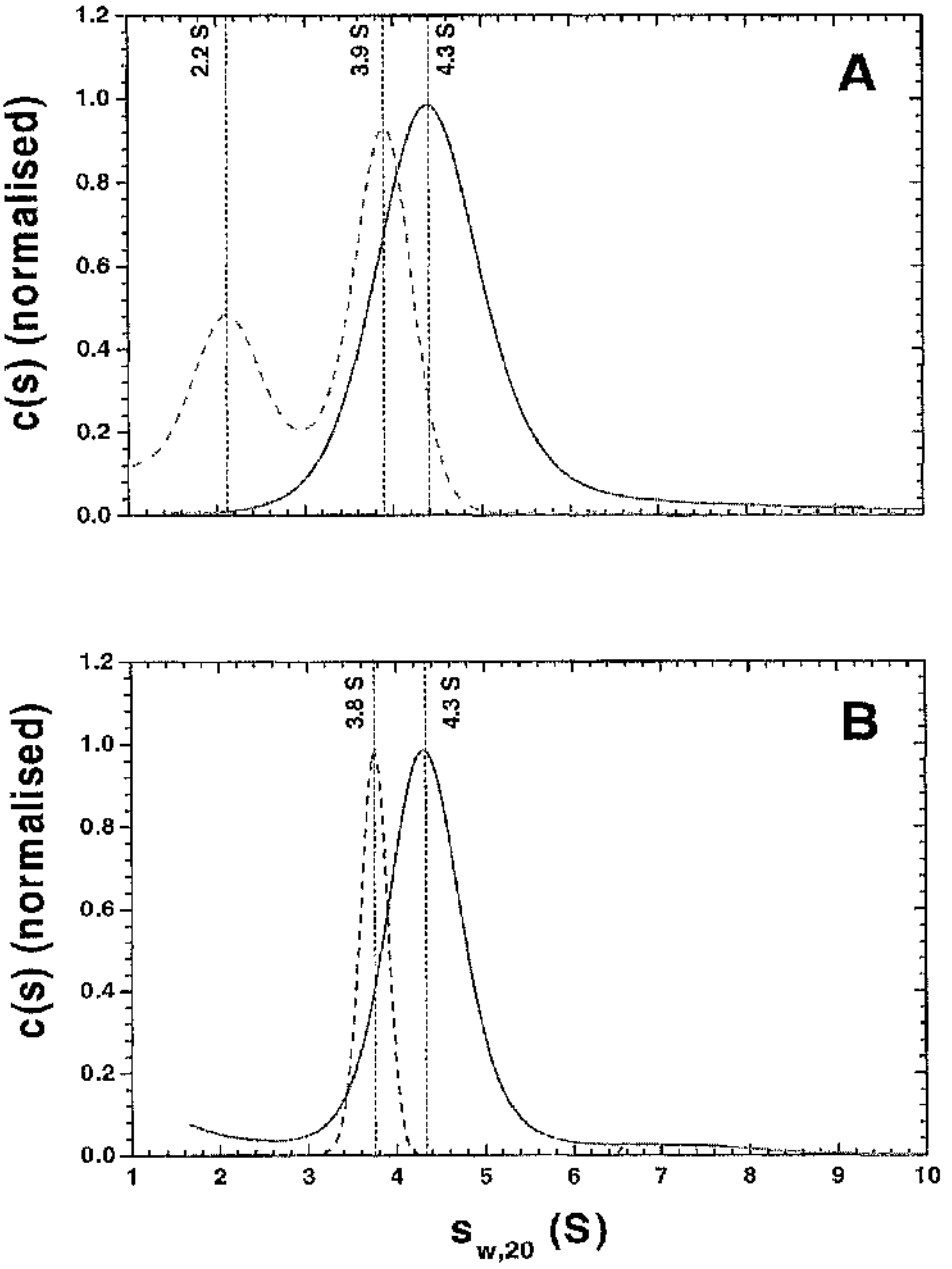

Figure 5.3: (A) Size distribution (c(s)) analysis obtained with SEDFIT (Schuck, 2000) for the HJ (800 nM) in the absence of Mg^{2+} (*dashed curve*) and in 10 mM Mg^{2+} (*solid curve*); (B) c(s) analysis for the HJ (8 μM) in the absence of Mg^{2+} or Na^{+} ions (*dashed curve*) or HJ (800 nM) in 50 mM NaCl (*solid curve*).

These results also show the applicability of SV measurements to the study of DNA stability under different experimental conditions. The relative mass concentrations of the fully and partially assembled HJ can be quantified at different temperatures and buffer conditions, allowing for the investigation of the thermodynamic stability of these molecules in solution.

5.4.2 Small angle x-ray scattering

Scattering curves for the HJs were obtained in TE buffer supplemented with different concentrations of Mg^{2+} (0, 10, 40 and 100 mM). All the analyses in the presence of Mg^{2+} yielded identical results. Hence, only the measurements at 0 and 10 mM $MgCl_2$ are shown here. Additionally, scattering curves were acquired at three different HJ concentrations for each buffer in order to account for effects of interparticle interactions. At each Mg^{2+} concentration, the scattering curves at the two lowest HJ concentrations (25 and 50 μM) were indistinguishable, indicating that at these concentrations the interparticle interactions are negligible. The processed curves are shown in Fig. 5.4 A.

The radius of gyration (R_g) of the HJs was firstly estimated by using the Guinier approximation to be: 39±1 and 38±1 Å for the 0 mM and the 10 mM $MgCl_2$ samples, respectively. The values of $I(0)/c$ (where $I(0)$ is the scattering intensity at zero angle and c is the sample concentration) for these samples were indistinguishable within experimental error, as expected for two species with identical molecular mass. The particle distance distribution functions ($p(r)$) for these samples are shown in Fig. 5.4 B. The D_{MAX} for the 0 mM $MgCl_2$ sample is 142±7 Å and that for the 10 mM $MgCl_2$ sample is 148±5 Å, indistinguishable within experimental error. The R_g obtained from the $p(r)$ functions are 40±1 and 41±2 Å, in agreement with those from the Guinier analysis. In the following sections, analysis of these data both by direct modelling and by *ab initio* reconstruction methods will be presented.

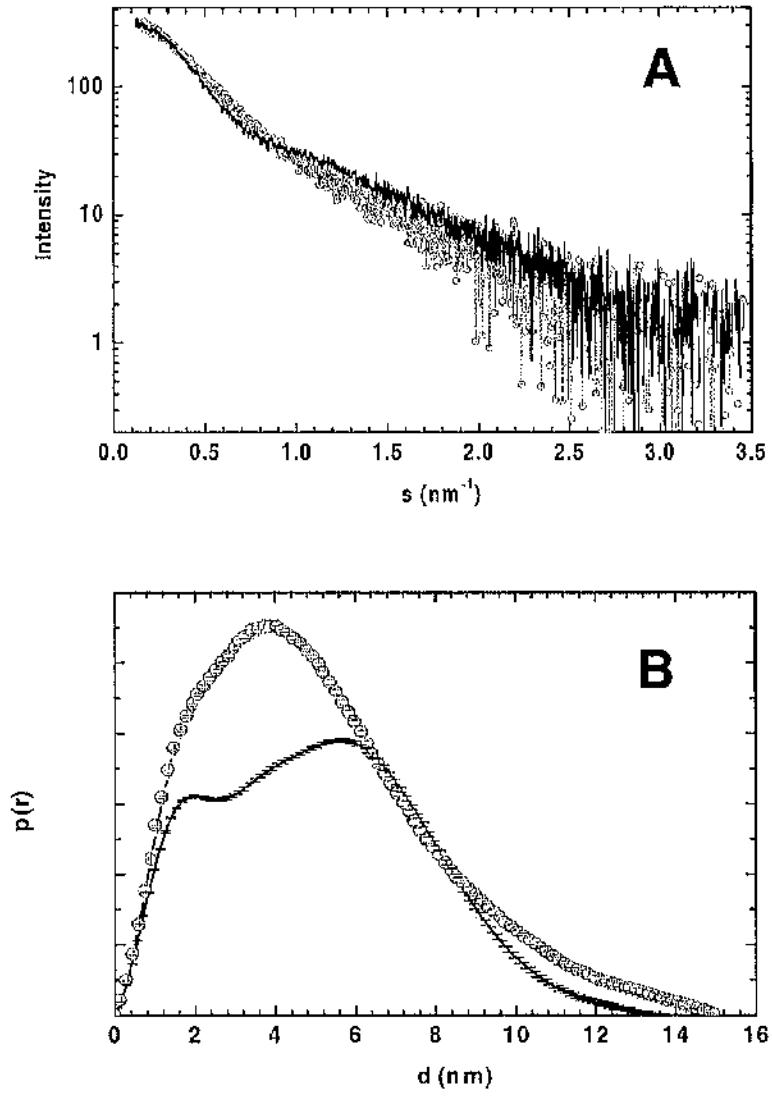

Figure 5.4: (A) X-ray scattering data from the HJs at 0 (*solid line*) and 10 mM (*circles*) Mg^{2+}. (B) Particle distance distribution functions from the scattering of the HJs at 0 mM Mg^{2+} (*solid line*) and 10 mM Mg^{2+} (*open circles*) calculated using the program GNOM (Svergun, 1992).

5.4.2.1 Rigid-body modelling of scattering data

Distinct structures have previously been proposed for the conformation of the HJ in solution at low and high Mg^{2+} concentration (Clegg et al., 1994; Duckett et al., 1988). In the absence of metal ions, the HJ has an extended conformation presenting an open central region (the square planar model). Upon addition of Mg^{2+} (or other metal ions), the junction changes its structure into a pairwise stacking of helices (Lilley and Clegg, 1993; Seeman and Kallenbach, 1994) (the stacked model). These models were employed as a starting point in order to find the structure that best fitted the experimental solution scattering data. The models were parametrised as shown in Fig. 5.5. To parametrise the stacked model the following variables were used: angle between arms as seen from the axis that passes through the points of strand exchange and that is perpendicular to the C_2 symmetry axis (φ^s), distance between the centres of the two stacked helices at the point of strand exchange (d), and the kink angle of the helices (ψ^s). Variation of these three variables within reasonable ranges allowed for the mapping of the stacked model parameter space. In this model, the angle between arms 1 and 2 (see Fig. 5.5 A) is the same as that between arms 3 and 4. In addition, when introducing a kink, both stacked arms are symmetrically deviated from the helical axis by the same amount. As a convention, when ψ is positive, the arms are kinked away from each other, whereas they get closer when ψ is negative.

Similarly, the square planar model was parametrised using the angle between arms as seen from the axis perpendicular to the plane containing the centres of the first four innermost basepairs of each arm (φ^p), and the kink angle between opposite arms (ψ^p) at the strand exchange point (see Fig. 5.5, C and D). In this instance, there were only two parameters as the out-of-plane distance between the helices at the point of strand exchange in this model is always zero. As for the stacked model, the angle between arms 1 and 2 (see Fig. 5.5 D) is the same as that between arms 3 and 4. However, in this case, when ψ is positive the two arms 1 and 3 are kinked in the opposite direction to arms 2 and 4. The kink has the same direction in both arms when ψ is negative.

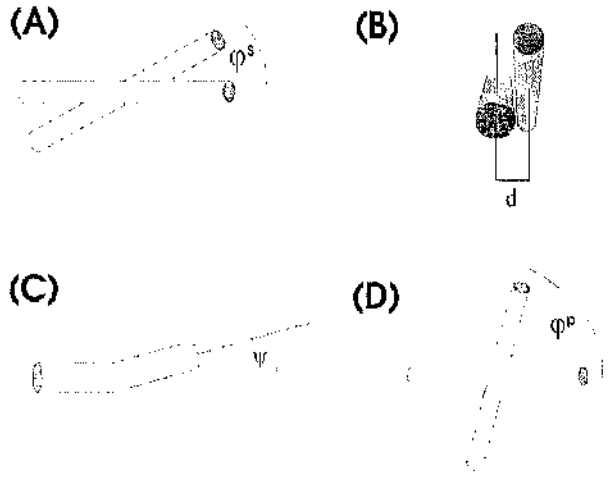

Figure 5.5: Parametrisation of HJ models. (A) Angle between arms (φ^s) in the stacked model; (B) distance between centres (d) of strand exchange in the stacked model; (C) kink angle for one arm (ψ^s and ψ^p) in either the stacked or the square planar model; (D) angle between arms (φ^p) in the square planar model.

The scripting language of the computer program NAMOT (Tung and Carter, 1994) was used to produce high-resolution models for the HJs. The four individual basepairs at the point of strand exchange were firstly positioned. Their coordinates, twist, roll and tilt angles were calculated from the parameters of the model to be produced. The four arms were subsequently built from these bases assuming B-form DNA. In order to speed up the generation of high resolution models, a computer program was developed (*namot2script*). The user defines the model employed, the range of the parameters that is to be explored, and the number of structures in the chosen parameter range. The program then generates scripts in NAMOT scripting language that are used to produce structural files in PDB format. The scattering curves for each of these structures, and their fits to the experimental data were then calculated by using the computer program CRYSOL (Svergun et al., 1995). The maximum hydration shell scattering contrast was allowed to vary between 0 and 150 e nm^{-3}, and the maximum excluded volume allowed was increased from 44 nm^3 to 65 nm^3 in order to allow for higher hydration of the DNA (compared with protein). Finally, the steric clashes (κ) within each structure were estimated by using the program *PDB2overlap*, also developed by me (see Appendix B).

With this procedure, the parameter space defining each possible structure for a given model was mapped onto a scalar field representing the goodness of fit, or in other words, each set of given parameters was assigned a corresponding goodness of fit ($\chi(r_0, \delta\rho, \psi, \varphi, d)$). In summary, the parameters were mapped onto a scalar field (Ξ) that takes into account how well the structural model generated from those parameters fits the experimental data, and penalises steric clashes ($\kappa(\psi, \varphi, d)$) within each structure,

$$\Xi(r_0, \delta\rho, \psi, \varphi, d) = \chi(r_0, \delta\rho, \psi, \varphi, d) + 10\theta(\kappa(\psi, \varphi, d) - \tau), \qquad (5.1)$$

where $\theta(x)$ is a step-like function that has a value of 0 when the argument $x < 0$ and 1 when $x > 0$, and τ is a steric clash threshold (defined so that the overlap volume is smaller than 5% of the total model volume). The introduction of

this penalty term did not modify the ultimate results but served to discard the structures containing steric clashes as possible models.

In order to find the minimum in Ξ with respect to the model parameters, a simple exploration of the parameter space was performed to identify the best fit by visual inspection. Figure 5.6 A shows the Ξ field as a function of φ^p and ψ^p for the 0 mM Mg^{2+} sample data analysed with the square planar model. A well-defined unique minimum at $\varphi_0^p = 90 \pm 20^o$, $\psi_0^p = 0 \pm 10^o$ defines the best fit to the experimental data (see Fig. 5.6 A). A criterion to decide how to estimate the uncertainties in the parameters had to be established. By visual inspection of the experimental scattering curves and the fits produced by parameters at the global minimum and nearby, it was found that fits with $\chi \leq \chi(\varphi_0^p, \psi_0^p) + 0.2$ (where φ_0^p and ψ_0^p define the position of the minimum in parameter space) are indistinguishable from the best fit. Accordingly, the contour levels are defined so that the difference between Ξ at the global minimum and Ξ at the second contour level is 0.2, and so the width of the lower contour level in every parameter direction defines its uncertainty. An identical approach was used to determine the uncertainty in the parameters for the stacked model. The values for the parameters in the square planar model that best fit the experimental data for the HJ at 0 mM Mg^{2+} are in full agreement with all the previous experimental data, which suggests a square planar conformation ($\psi^p = 0$) with an angle of 90° between adjacent helical arms ($\varphi^p = 90$).

A similar approach was used to find the set of parameters for the stacked model that best fitted the experimental data for the 0 mM Mg^{2+} sample (see Fig. 5.6 B). There is also a minimum at $\varphi_0^s = 90 \pm 20^o$, $d^s = 22$ Å, but it is much shallower than that found for the square planar model, representing a poor fit to the data. Similar poor fits were found when the distance between arms was fixed at $d^s = 22$ Å and the parameter space in the directions φ^s and ψ^s was explored (data not shown). Therefore, the data do not support a stacked structure at 0 mM Mg^{2+}.

In the same fashion, the mapping procedure was used to analyse the 10 mM Mg^{2+} sample scattering data with the square planar and the stacked models.

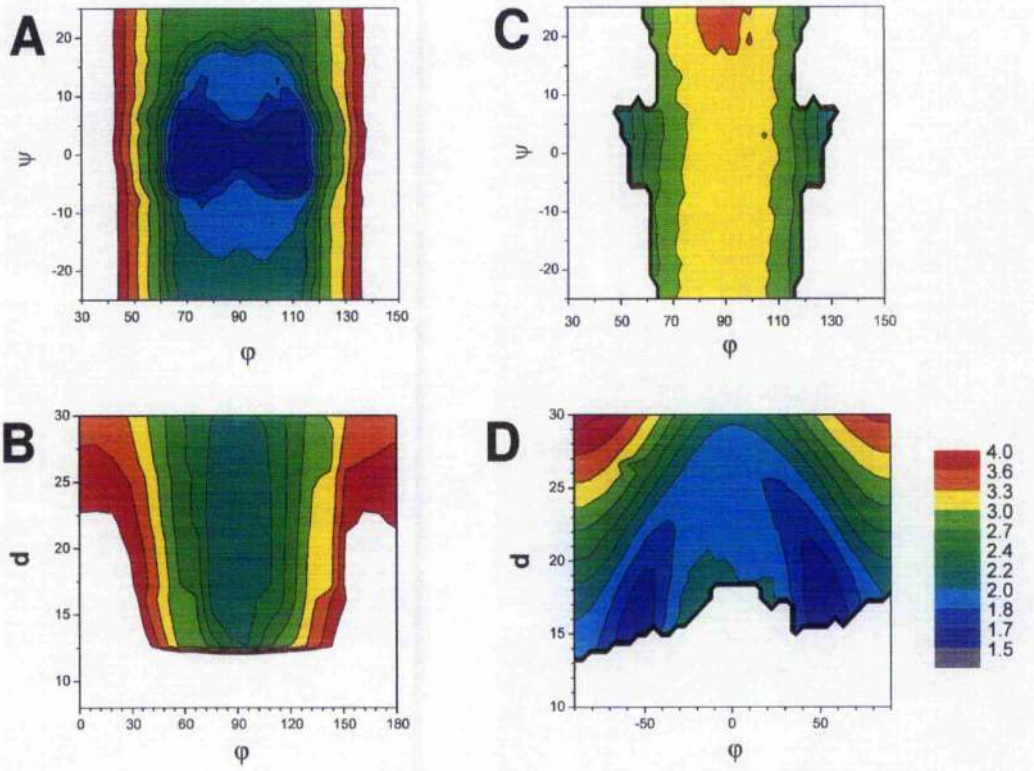

Figure 5.6: (A) Contour plots of Ξ as a function of the kink angle (ψ^p) and the angle between arms (φ^p) for the 0 mM Mg^{2+} HJ sample data fitted with the square planar model. The white areas in the plots specify structures with overlap; (B) Contour plots of Ξ as a function of the distance (d^s) and the angle between arms (φ^s) for the 0 mM Mg^{2+} HJ sample data fitted with the stacked model; (C) Contour plots of Ξ as a function of the kink angle (ψ^p) and the angle between arms (φ^p) for the 10 mM Mg^{2+} HJ sample data analysed with the square planar model; (D) Contour plots of Ξ as a function of the distance (d^s) and the angle between arms (φ^s) for the 10 mM Mg^{2+} HJ sample data analysed with the stacked model. d is in units of Å.

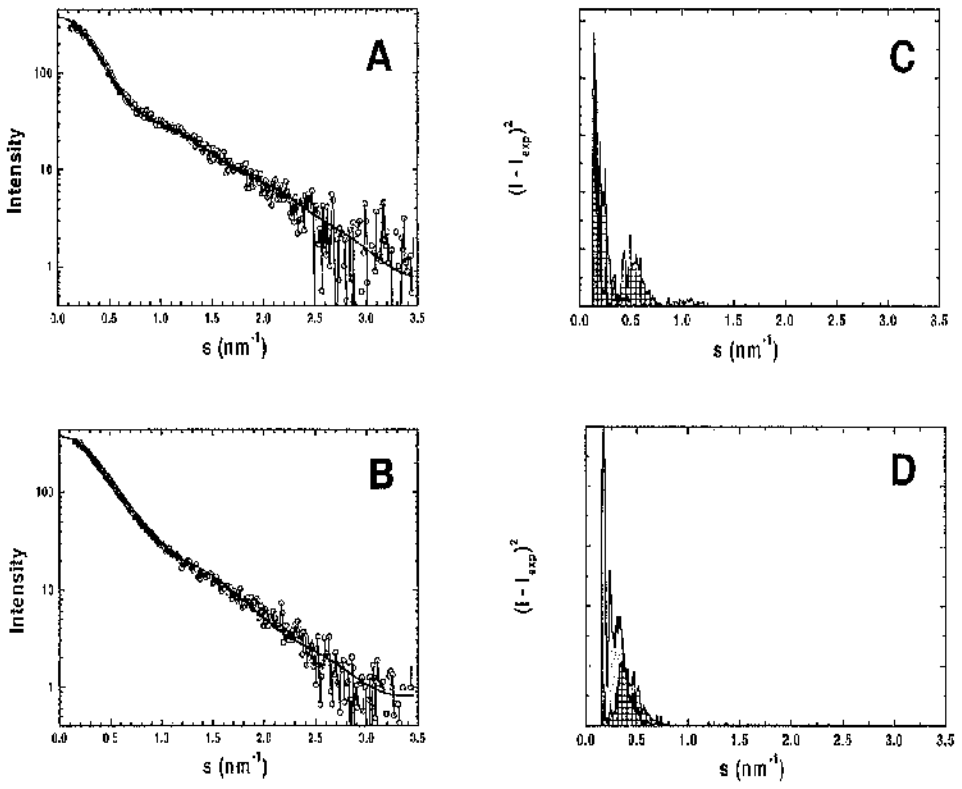

Figure 5.7: Comparison between best-fits for the 0 and 10 mM Mg^{2+} HJ samples. (A) 0 mM Mg^{2+} experimental data (*circles*) and best fit using the stacked model (*solid line*). (B) 10 mM Mg^{2+} experimental data (*circles*) and best-fit using the square planar model (*solid line*). (C) Square difference between best fit and experimental scattering curves for the 0 mM Mg^{2+} HJ sample using stacked (grid) and square planar (shaded) models. (D) Square difference between best fit and experimental scattering curves for the 10 mM Mg^{2+} HJ sample using stacked (grid) and square planar (shaded) models.

Firstly, the Ξ space defined by varying φ^p and ψ^p was explored assuming a square planar model (see Fig. 5.6 C). There is one extended minimum in the overlap-free region at $\varphi_0^p = 90^o$ (almost independent of ψ^p), but with an absolute value ($\Xi = 3.2$) that does not represent an acceptable fit to the scattering data, implying that a square planar model cannot correctly represent the data for any reasonable value of φ^p and ψ^p.

Secondly, Ξ was mapped as a function of φ^s and d using the stacked model (see Fig. 5.6 D). Now, there are two narrow global minima located at $\varphi_0^p = \pm 55 \pm 10^o$, $d_0 = 18 \pm 3$ Å. These minima represent two identical low-resolution structures that are related by a symmetry operation (reflection). Even though there are differences between these alternative structures at high resolution (due to DNA chirality and differences in the arm sequences), this technique is not able to distinguish between them. In order to find whether changes in the kink angle improve the fit to the experimental data, ψ^s was varied between -30^o and 30^o, and φ^s from -90^o to 90^o, keeping the distance between the arms constant at $d_0 = 18$ Å (the position on the d-axis of the global minimum found when varying d and φ). In this case, a well-defined narrow minimum was found at $\varphi_0^p = \pm 55 \pm 10^o$ and $\psi_0^p = 0 \pm 5^o$, confirming that, in the structure that best fits the experimental data, the arms are not significantly kinked.

The actual fits to the experimental data produced by the models are presented in Figure 5.7 A (for 0 mM Mg^{2+}) and B (for 10 mM Mg^{2+}). The values for the model coefficients at the minima are in full agreement with those previously measured by Duckett *et al.* (in the absence of Mg^{2+}, the HJ has a square planar conformation with $\varphi = 90^o$; in the presence of at least 100 μM Mg^{2+}, the HJ has a stacked conformation with $\varphi = 60^o$) (Duckett et al., 1988; Lilley and Clegg, 1993).

5.4.2.2 *Ab initio* reconstructions

The computer programs GASBOR (Svergun et al., 2001) and DAMMIN (Svergun, 1999) were employed to restore *ab initio* models of the HJ from the SAXS

data; however, only results obtained from GASBOR will be reported here (since the results obtained with DAMMIN were virtually identical). Ten independent *ab initio* reconstructions were generated from the 0 mM Mg^{2+} scattering data, starting from different random configurations and using different numbers of beads (leading to different packing radii). In all the runs, C_4 symmetry was imposed. The average R_g of the generated models was 39.5±0.5 Å. The fits of the models to the experimental data were excellent and almost identical for all the models, yielding in all cases a goodness of fit smaller than 1.4 (arb. units).

Superimposition of the restored models was performed using the computer program SITUS (Wriggers et al., 1998). Once superimposed, the models were averaged using a computer program I developed as part of this project (*pdb2average*) to remove voids in the structure by the combined use of a modified version of the improved cube method introduced by Pavlov and Fedorov (1983) with cycles of expansions and contractions as implemented by Lee and Richards (1971) in their algorithm for calculating the accessible surface area of a protein. The program recursively loads the input structures onto the same cubic lattice, calculates the cumulative Gaussian electron density, performs a series of expansions/contractions to reduce the appearance of voids, and finally samples the protein structure using a Monte-Carlo algorithm. This procedure produces the average bead model that contains the structural features that all the individual restored models have in common, and eliminates the traits that are not shared between them. A similar procedure has been used by Petoukhov et al. (2002).

Figure 5.8 shows side and front views of three of the reconstructions for the HJ at 0 mM Mg^{2+} in the same orientations in which they were averaged. The same running parameters were used to generate the second and third structures, whereas a much smaller number of beads was used for the first one. However, the models are remarkably similar. A mesh model of the average reconstruction, and the best square planar model (in CPK colour code) for the HJ at 0 mM Mg^{2+} found in the previous section are shown in Fig. 5.9. The two models are clearly in agreement, showing that two independent methods to interpret the experimental SAXS data result in essentially the same model for the HJ at 0 mM Mg^{2+}.

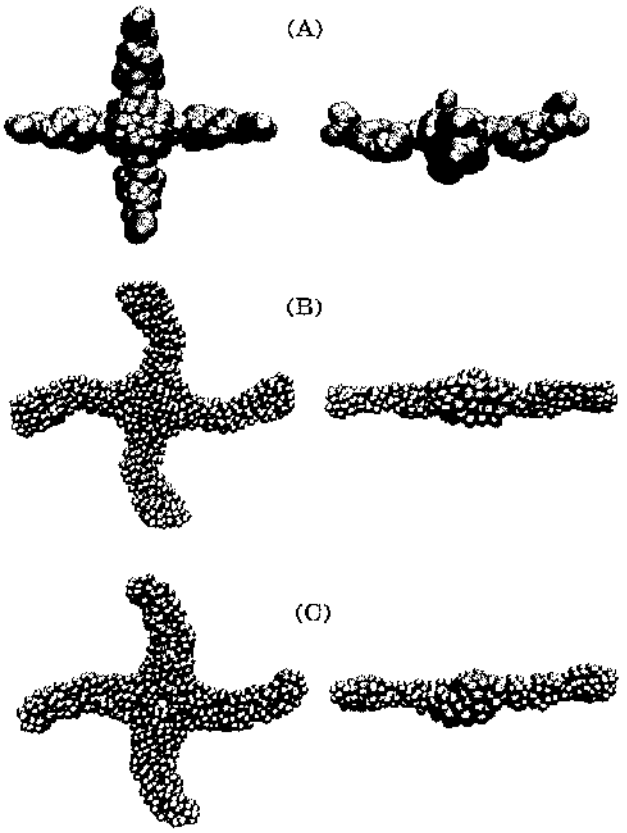

Figure 5.8: Front and side views of three models restored from SAXS data for the 0 mM Mg^{2+} HJ using GASBOR (Svergun et al., 2001). 280 beads were used for reconstructing model (A) and 840 beads for models (B) and (C). The three reconstructions are significantly similar.

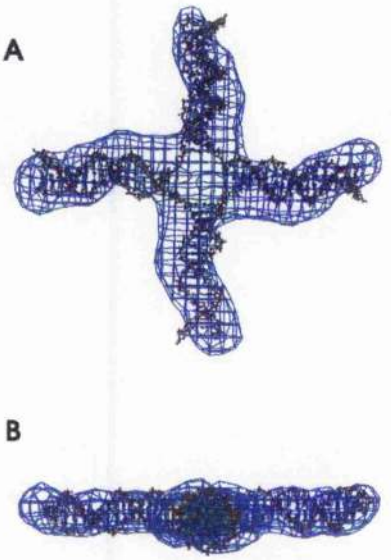

Figure 5.9: (A) Front and (B) side views of a mesh model of the average reconstruction superimposed onto the high resolution NAMOT (Tung and Carter, 1994) model for the HJ at 0 mM Mg^{2+} found using rigid body modelling.

In addition to these C_4 constrained simulations, models with other symmetry constraints, such as C_2 and C_1 were generated. The models made in this way produced fits to the experimental data that were equally as good as those of the reconstructions presented (data not shown), but that in all cases had physically unrealistic conformations. In order to restrict the number of possible reconstructions and discard unrealistic ones, a series of criteria that the reconstructions had to fulfil in order to be acceptable was defined: (a) because of the sequence complementarity of the oligonucleotides comprising the junction, the structure is expected to be formed by a combination of four arms representing the B-DNA fragments composing the HJ; (b) the models can not be disconnected; and finally, (c) the sedimentation coefficients of the models simulated with HYDRO (García de la Torre et al., 2000) have to agree with the experimentally determined sedimentation coefficient. All the reconstructions were tested according to those criteria, and only those generated with specified C_4 symmetry were able to satisfy all the requirements.

A similar procedure was applied to the *ab initio* reconstruction of the HJ at 10 mM Mg^{2+}, but in this case C_2 symmetry was imposed (other symmetry conditions will be mentioned later). The reconstructions were superimposed and averaged as for the 0 mM Mg^{2+} scattering data. Figure 5.10 shows a mesh model of the average reconstructed structure superimposed on a high resolution model that was generated by finding the best docking of two independent non-overlapping 34 bp straight DNA fragments. The angle between arms in the resulting model is 60° and their centre-to-centre distance is 17 Å, similar to the model proposed in the rigid body modelling analysis. However, in this new structure the point of strand exchange appears slightly displaced. Even though the arm lengths are symmetric in some individual reconstructions, the process of averaging independent reconstructions will inevitably accentuate the asymmetry of the arms by docking together reconstructions where the crossover point is displaced from the centre. Ultimately, these results emphasize the inherent variability of the reconstruction process. As before, a series of simulations was performed with different symmetry constraints, namely C_1 and C_4, and the same criteria as before were

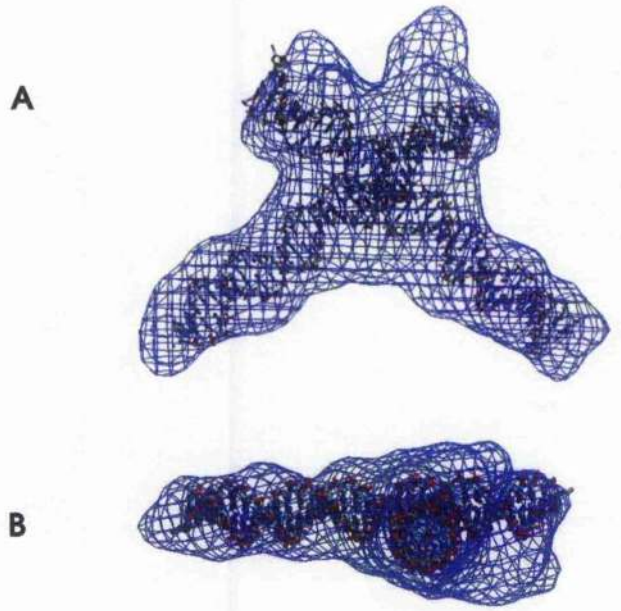

Figure 5.10: (A) Front and (B) side views of a mesh model of the average reconstruction superimposed onto the high resolution NAMOT (Tung and Carter, 1994) model for the HJ at 10 mM Mg^{2+} found using rigid modelling.

used to define their acceptability. In this case, only the models with C_2 symmetry satisfied these criteria while at the same time fitting the scattering data.

5.4.3 Hydrodynamic modelling

The computer program HYDROPRO (García de la Torre et al., 2000) was employed to calculate the hydrodynamic parameters of high-resolution models (see Section 3.3.2 for details). The calculation for the square planar structure that best fitted the 0 mM Mg^{2+} sample scattering curve gave a sedimentation coefficient of 3.96 S, which is in full agreement with the sedimentation coefficient determined experimentally (3.9 S). The calculation for the stacked structure that best fitted the 10 mM Mg^{2+} scattering data yielded a sedimentation coefficient of 4.38 S, which is again in complete accordance with the experimentally determined value (4.4 S). Thus, the hydrodynamic properties of the SAXS-based models are consistent with the hydrodynamic parameters measured experimentally.

5.5 Discussion

In this chapter, the changes in the low-resolution solution conformation of a Holliday junction with Mg^{2+} concentration were studied by using small angle x-ray scattering, sedimentation velocity, and various computational modelling techniques. Two independent analyses of the scattering data were implemented.

The first method was based on previous models for the HJ (Lilley and Clegg, 1993) in the absence and presence of Mg^{2+}, and was therefore model-dependent. The square planar model successfully fitted the scattering data for the 0 mM Mg^{2+} HJ sample, with $\psi = 0 \pm 10^0$ (kink angle) and $\varphi = 90 \pm 20^0$ (angle between arms), whereas the stacked model could not fit the scattering data. In contrast, the stacked model was shown to fit the 10 mM Mg^{2+} HJ sample scattering data, with a distance between arms of 18±3 Å, a kink angle of 0±5°, and an angle between arms of 55±10°. In this case, a square planar model could not fit the scat-

tering data. This rigid body modelling approach allows for the determination of the model parameters for each of the conformations adopted by the HJ in solution, providing higher accuracy than other experimental techniques, such as FRET or gel-shift electrophoresis assays (Lilley and Clegg, 1993). However, prior additional information on the possible conformation of the macromolecule under investigation is required in order to define the model and its parameters. The SANS study of Chamberlain et al. (1998) on a HJ and its complex with RuvA reported no variation in $I(0)$ or R_g of the HJ sample with Mg^{2-} concentration. The results presented in this chapter confirm these findings and show that the difference in scattering between HJ samples in the absence and presence of Mg^{2+} appears only at scattering angles $s > 0.4\,\mathrm{nm}^{-1}$.

The same experimental scattering data were analysed using model-independent *ab initio* reconstruction methods. These methods have recently been shown to be able to recreate the low-resolution structures of macromolecules (protein, DNA) in solution (Koch et al., 2003). The resolution of such models is not defined by the radii of the beads used but by the range of the scattering data fitted (in this study $2\pi/s_{max} \sim 17$ Å). The reconstruction process is ambiguous, in that a variety of models is obtained which fit the experimental data equally well. Some of these models can be ruled out by specifying a set of rules based on predictable properties of the structure and on other experimental measurements (such as the sedimentation coefficient). The models that both fitted the experimental data and satisfied these rules were shown to differ only in fine details. This methodology again showed a striking agreement with both the solutions found by rigid body modelling and previously reported results using other techniques (Lilley and Clegg, 1993). The models found by both methods not only fitted the scattering data but also agreed with the sedimentation coefficients found experimentally.

These results provide a direct determination of the low-resolution conformation of Holliday junctions in solution. The methods applied in this chapter could be used for the study of the low-resolution structures of other biologically significant nucleic acid structures such as DNA hairpins, bulges, bent sequences, tetraplexes, and RNA as well as their complexes with proteins. This methodology combined with small angle neutron scattering could also be successfully

applied to the assignment of the individual DNA and protein components in protein-DNA complexes by using contrast matching.

Chapter 6

Solution structure of the Tn3 resolvase-crossover site synaptic complex

Summary

Tn3 resolvase is a site-specific DNA recombinase which catalyses strand exchange in a synaptic complex containing twelve resolvase subunits and two *res* sites. Hyperactive mutants of resolvase can form a simpler complex (X-synapse) containing two shorter DNA segments at which strand exchange takes place ('crossover site' or site I), and a resolvase tetramer. The low-resolution solution structure of the purified, catalytically competent X-synapse was solved from small angle neutron and x-ray scattering data, using novel methods in which the data are fitted with models constructed by rigid-body transformations of a published crystallographic structure of a resolvase dimer bound to site I. This analysis reveals that the two site I fragments are on the outside of a resolvase tetramer core, and provides some information on the quaternary structure of the tetramer. Implications of this structure for the architecture of the natural synaptic complex and the mechanism of strand exchange are discussed. The contents of this chapter have been published elsewhere (Nöllmann et al., 2004a).

6.1 Introduction

The large synaptic complex formed by Tn3 resolvase and two *res* sites prior to strand cleavage and exchange (the synaptosome, see Section 1.8) can be regarded as being composed of two parts. One part, that includes a resolvase-binding site (site I) from each *res* and four resolvase subunits, is primarily responsible for catalysis (the 'X-synapse') . A second 'accessory' part, that includes binding sites II and III of each *res* along with a further eight resolvase subunits, is implicated in the regulation of activity (Grindley, 2002). Despite the solution of a number of crystal and NMR structures of $\gamma\delta$ resolvase, a close relative of Tn3 resolvase (Pan et al., 2001; Rice and Steitz, 1994a,b; Sanderson et al., 1990), and a co-crystal structure of a dimer of $\gamma\delta$ resolvase bound to site I (Yang and Steitz, 1995, Fig. 6.1A), the architectures of these synaptic complexes are currently unknown.

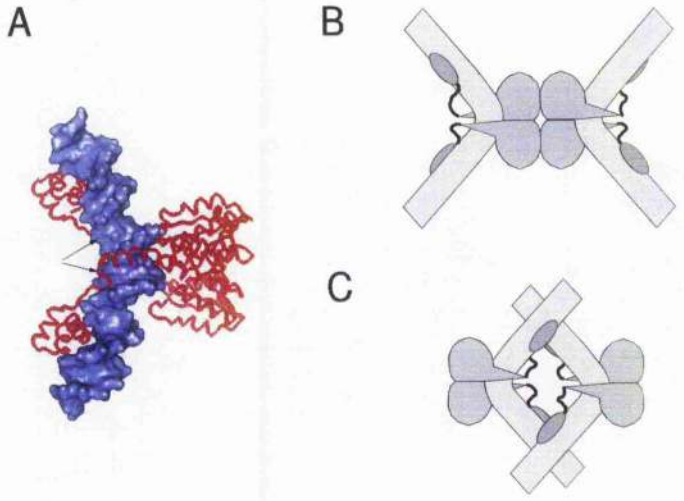

Figure 6.1: Structural models for the X-synapse. (A) Representation of the co-crystal structure of $\gamma\delta$ resolvase bound to site I (PDB code 1gdt) (Yang and Steitz, 1995). Protein is represented as ribbons and DNA as surface contour. Arrows represent the approximate position of residue V139 in both resolvase subunits. (B) and (C) Schematic representations of the DNA-out and DNA-in models, respectively. DNA is shown as a grey bent cylinder, with protein subunits in darker grey.

Tyrosine site-specific recombinases use a different mechanism for strand exchange. After formation of a synaptic complex, only one DNA strand in each site is cleaved; strand exchange then gives rise to a Holliday junction intermediate (HJ). Exchange of the two other DNA strands of the HJ completes recombination. Crystal structures of Cre and Flp synaptic complexes have provided important insights on the strand exchange mechanisms employed by the tyrosine recombinases (Chen and Rice, 2003; Van Duyne, 2001). Other, non-crystallographic evidence suggests that the solution structures of the Flp and Cre synaptic complexes might differ in some details from the crystal structures (Crisona et al., 1999; Huffman and Levene, 1999).

Wild-type Tn3 resolvase forms a specific dimer complex with *res* binding site I, but a 2 × site I X-synapse has not been observed (Bednarz et al., 1990; Grindley et al., 1982; Watson et al., 1996). 'Hyperactive' mutants of Tn3 resolvase have been selected that promote recombination at site I in the absence of the *res* accessory binding sites (Arnold et al., 1999; Burke et al., 2004). Some of these mutants can form a stable X-synapse, which has been shown to contain two site Is and four resolvase subunits (Sarkis et al., 2001; J. He and W. M. Stark, manuscript in preparation).

Two very different configurations for the X-synapse have been proposed, both of which have approximate D_2 symmetry, consistent with the topology and inferred rotational mechanism of strand exchange (Stark et al., 1989). In the 'DNA-out' configuration, the N-terminal, catalytic domains of two resolvase dimers come together to form a protein tetramer that binds the DNA fragments on the outside (Fig. 6.1B). In the 'DNA-in' configuration, the two DNA duplexes come close together at the centre of the synapse, and the catalytic domains are on the outside (Fig. 6.1C). Models for the full 2 × *res* synapse have been proposed in which the X-synapse component has either a DNA-out (Grindley, 1994; Sarkis et al., 2001; Stark et al., 1989) or a DNA-in (Rice and Steitz, 1994a; Yang and Steitz, 1995) configuration. The architecture of the X-synapse has important implications for the mechanism of strand exchange. In the DNA-in configuration, the DNA ends created by resolvase-mediated cleavage would have to move only a small distance to effect strand exchange (Rice and Steitz, 1994a). In the DNA-out configuration,

the sites of DNA cleavage are far apart. It is as yet unclear how resolvase could transport the DNA ends over the requisite large distance to bring about strand exchange, whilst at the same time holding the synapse together. Nevertheless, current experimental data lend support to the DNA-out model (Leschziner and Grindley, 2003; Sarkis et al., 2001), and activating mutations that stabilise the X-synapse are in the part of the catalytic domain predicted to be involved in dimer-dimer interactions to form a DNA-out synaptic tetramer (Burke et al., 2004).

In this chapter, the solution structure of the X-synapse is derived from small angle x-ray scattering (SAXS), small angle neutron scattering (SANS) and analytical ultracentrifugation sedimentation velocity (SV) data. It is concluded that the X-synapse has a DNA-out structure. The analysis presented uses novel computational methods that should be generally applicable to the structural characterisation of other large protein-nucleic acid complexes.

6.2 Materials and Methods

6.2.1 Sample preparation

The oligonucleotides used to make X-synapses for subsequent sedimentation velocity or scattering studies contained a phosphorothioate modification of the scissile phosphodiesters, to avoid resolvase-mediated cleavage. Oligonucleotides were synthesised and purified as described in Section 3.1.3.2. The sequences of the top strand for the 50-mer and 36-mer oligonucleotides containing binding site I were respectively: CGTGACTCAACCGTTCGAAATATTATaAATTATCAGACATAGTGGGGCGG and CAACCGTTCGAAATATTATaAATTATCAGACATAGT (the nucleotide with the 5' phosphorothioate modification being in lower case). Tn3R and NM-resolvase were purified by a procedure described in Section 3.1.5. NM-GFP comprises NM-resolvase and GFPE (Brejc et al., 1997) linked by the peptide sequence SGGGGS between resolvase residue 185 and GFP

residue 1. The purification method for NM-GFP was similar to that employed for NM-resolvase. Protein purity was assessed using SDS-PAGE gels, and function by standard *in vitro* recombination assays (He et al., 2002).

6.2.2 X-synapse purification

X-synapse purification details can be found in Section 3.1.7. X-synapse samples for SANS experiments were purified in TEGGS buffer and then extensively dialysed against the appropiate D_2O-containing buffer.

6.2.3 Sedimentation velocity and hydrodynamic modelling

Sedimentation velocity experiments were performed as described in chapter 3.2.2. The experiments were carried out at 4 °C in TEGGS buffer at a rotor speed of 45000 rpm. A series of 75 scans, 5 minutes apart was obtained, with a radial step size of 0.003 cm using interference optics. All experiments were carried out at an approximate sample concentration of 5 μM. The samples (380 μl) were loaded into double-sector centrepieces. Sedimentation profiles were analysed with SEDFIT (Schuck, 2000) as described in Section 2.3.1.1. Buffer density and viscosity measurements, and hydrodynamic simulations were as described in Section 3.3.2.

The percentage difference between experimental ($s^E_{20,w}$) and modeled ($s^M_{20,w}$) sedimentation coefficients $f(\bar{v}, \eta, \rho, M_0, T) \equiv 100[s^E_{20,w} - s^M_{20,w}]/[(s^E_{20,w} + s^M_{20,w})/2]$ was calculated as a function of the specific partial volume ($\bar{v}$) of the protein-DNA complexes in the range $0.6 - 0.73$ ml/mg for both the DNA-in^{HS} and X-exp structural models (see Results section), and was used to evaluate the degree of discrepancy between experimental and modelled sedimentation coefficients.

6.2.4 Small angle x-ray and neutron scattering

SAXS and SANS experiments were performed using solutions of the X-synapse at concentrations >10 μM, at a temperature of 6-7 °C, using the methods described in Sections 3.2.6 and 3.2.5. SANS90 data (see below) were obtained on beam line D11, at the Institut Laue-Langevin (ILL, Grenoble, France) in buffer TEGS90 (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10% v/v glycerol, 200 mM NaCl, 90% v/v D_2O). Datasets SANS43 and SANS65 were obtained on LOQ, a fixed geometry time-of-flight small angle diffractometer at the ISIS Spallation Neutron Source (Rutherford Appleton Laboratory, Didcot, U.K.) in buffers TEGGS43 (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 20% v/v glycerol, 200 mM NaCl, 43% v/v D_2O) and TEGGS65 (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 20% v/v glycerol, 200 mM NaCl, 65% v/v D_2O), respectively. For these datasets, data collection times were of the order of 15 hours per sample or buffer.

The SAXS dataset was obtained on the Beamline 2.1 at the Synchrotron Radiation Source (Daresbury, UK), and analysed by using the procedures described in Section 3.2.6 in buffer TEGGS (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 20% v/v glycerol, 200 mM NaCl). Scattering curves were measured at two sample concentrations, 1 μM and 10 μM. The scattering profiles of all structural models produced were calculated by using either CRYSOL or CRYSON (x-rays and neutrons, respectively) (Svergun et al., 1995), as described in Sections 2.2.3.1 and 3.2.6. The χ-value produced when fitting the scattering intensity was used as defined in Section 3.3.3.

I developed the computer program *domain-rot* in order to generate the different structures for given values of the parameters d, φ and σ, and to calculate from them (as described above) the χ-values that the simulated scattering curves produced when fitted to the experimental data.

6.2.5 Construction of structural models

The Cartesian axes in all structures derived from the $\gamma\delta$ resolvase/site I co-crystal structure (PDB code 1gdt, Yang and Steitz (1995)) were defined as follows (Fig. 6.4E): the z-axis is parallel to the line passing through the atom O4* in residue T19, DNA chain E (atom 4806) and atom O in residue Met 106, chain B (atom 3124) on the same 1gdt unit; the x-axis is parallel to the line passing through the atom CZ in residue Phe 4, chain A (atom 916) and atom CG in residue Leu 3, chain B (atom 2328) on the same 1gdt unit. The y-axis was constructed by using the vector product between the z-axis and the x-axis versors in order to obtain a right-handed Cartesian coordinate system. The site I DNA axis was defined as the line joining atom C6 in residue G 1, chain C (atom 2863) and atom C2 in residue G 35, chain D (atom 3561) in 1gdt.

$DNA50_0$ was built as a 50 bp B-DNA fragment using NAMOT (Tung and Carter, 1994). Site I B-DNA fragments with different kink angles ($DNA50_{20}$, $DNA50_{40}$, $DNA50_{60}$ and $DNA50_{70}$) were built by using the computer program NAB (Macke and Case, 1998) with a similar procedure to that employed by Strater et al. (1999). 1gdt50 was built using protein and DNA coordinates from 1gdt, and the DNA was extended by adding two 7 bp B-DNA fragments to its ends. To model variations in the angle σ (see Results section), the 1gdt structure was cut after the protein residues V139 (see arrows in Fig. 6.1 A) to produce two substructures: 1) $1gdt_{NTD}$ containing the two catalytic domains of resolvase up to residue V139, and 2) the site I DNA fragment together with the two resolvase DNA binding domains ($1gdt_{DNA+CTD}$). $1gdt_{\sigma}$ was produced by fixing $1gdt_{NTD}$ and rotating $1gdt_{DNA+CTD}$ by an angle σ in the z-axis. The angle σ is measured so that 1gdt is recovered at $\sigma = 0$. σ can be determined from a particular X-synapse structure as the angle between the site I DNA axis and the x-axis minus 59.4°.

NM36 X-synapse models were assembled from a 'fixed' 1gdt unit in the xyz coordinate system as shown in Fig. 6.4E, and a second 'movable' unit added by a 180° rotation around the y-axis from the position of the fixed unit. The second unit was then subjected to subsequent translations and/or rotations about the

z-axis. Further models were built by translating one 1gdt unit along the y-axis (data not shown). The distance d between the two 1gdt units is defined as the distance between atom O5* in residue A20, DNA chain E (atom 4831) on the fixed 1gdt unit and atom C4 in residue A20, DNA chain E (atom 4817) on the movable 1gdt unit. φ is defined as the angle of rotation of the movable 1gdt unit in the z-axis direction, and can be determined from a particular X-synapse structure by measuring the angle between x-axes of each 1gdt. The angle between the two site I DNA fragments in any given X-synapse structure can be determined by measuring the obtuse angle between their axes. NM50 and NM-GFP36 X-synapses were assembled from 1gdt50 and 1gdt-GFP in an identical manner. The atomic coordinates of GFP were obtained from the Protein Data Bank, code 1emb (Brejc et al., 1997).

The DNA-out 'back-to-back' model of Sarkis et al. (2001) (DNA-out^{BB}) can be built by using the procedure described above, using the parameters $(d, \varphi) = (63\,\text{Å}, 0^o)$. Similarly, the DNA-in 'hand-shake' model (DNA-in^{HS}) as proposed by Yang and Steitz (1995) can be built by using $(d, \varphi) = (-31\,\text{Å}, -75^o)$.

6.3 Results

6.3.1 Formation of stable X-synapse

Wild-type Tn3 resolvase forms a complex containing two protein subunits and one site I DNA fragment, which can be observed in a polyacrylamide gel 'band-shift' assay (Bednarz et al., 1990; Blake et al., 1995) (band II in lane 2, Fig. 6.2B). Hyperactive mutants, such as NM-resolvase (which has the six mutations R2A E56K G101S D102Y M103I Q105L; Burke et al., 2004) make, instead, an X-synapse with lower electrophoretic mobility, which contains four protein subunits and two site Is (Sarkis et al., 2001; J. He and W. M. Stark, manuscript in preparation) (band III, lanes 3-7 in Fig. 6.2B). The X-synapse made by NM-resolvase

with unmodified site I is unstable, being lost following resolvase-mediated cleavage of its DNA (lane 4 in Fig. 6.2B). However, a stable X-synapse is observed if the site I DNA contains a phosphorothioate modification at both scissile phosphodiesters (lane 6), or if an inactive mutant of NM-resolvase (NM-S10A) is used with unmodified site I (lane 3). This, and other evidence (J. He and W. M. Stark, manuscript in preparation) that NM-resolvase cleaves the DNA in the X-synapse confirms that it is in an active configuration, and is not an unproductive 'dead end'.

6.3.2 Size exclusion chromatography and analytical ultracentrifugation

Three versions of the X-synapse were purified, all with phosphorothioate-modified site I oligonucleotides, to block resolvase-mediated DNA cleavage (see above). NM36 contains NM-resolvase and two 36 bp blunt-ended site I DNA fragments, the same as were used to generate the complexes shown in Fig. 6.2B; NM50 contains NM-resolvase and two 50 bp site I DNA fragments; and NM-GFP36 contains NM-GFP, which has a 23 kDa green fluorescent protein (GFP) domain fused to the C-terminus of NM-resolvase, complexed with two 36 bp site I DNA fragments (see Section 6.2.1).

X-synapses were purified by size exclusion chromatography (see Section 3.1.7). Traces in Fig. 6.2A show UV absorbance at 260 nm and 280 nm of the eluate from the size exclusion columns during NM36 purification. Fractions were concentrated and assayed by native polyacrylamide gel electrophoresis, showing that fraction 2 (lane 7 in Fig. 6.2B) corresponds to X-synapse.

Solutions of the three X-synapses were analysed by sedimentation velocity (SV) analytical ultracentrifugation. SV profiles were fitted using c(s) and finite elements analysis with the computer program SEDFIT (Schuck, 2000). Figure 6.2C shows c(s) profiles of a solution of the purified NM36 X-synapse (corresponding to band 2 in Fig. 6.2A and band III, lane 7 in Fig. 6.2B) compared

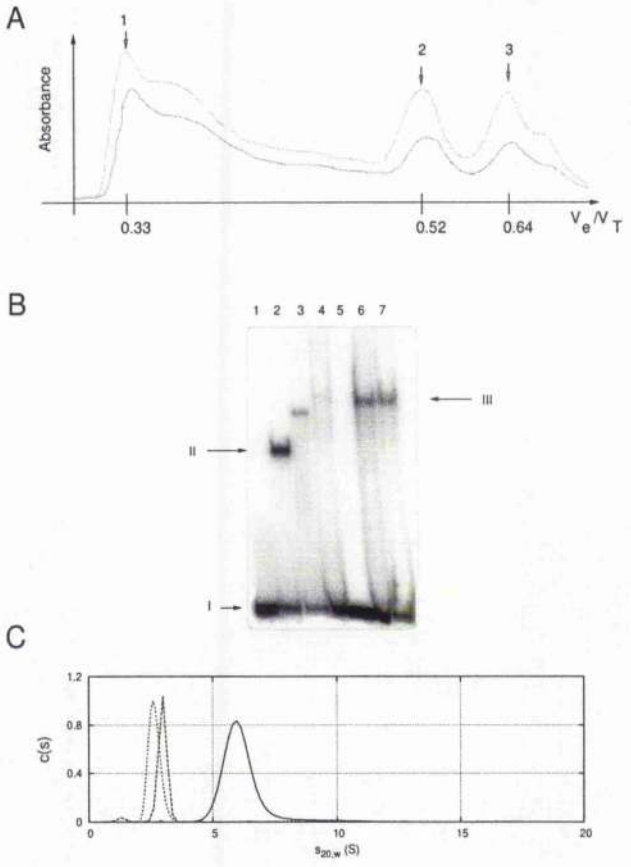

Figure 6.2: X-synapse purification and characterisation. (A) Size exclusion chromatography optical absorbance traces versus eluted/total column volume: (1) peak corresponding to ficoll (400 kDa), (2) NM36 X-synapse, and (3) site I (36 bp) DNA. V_e/V_T values refer to the upper curve (260 nm); the lower curve (280 nm) is slightly shifted to the right. The extinction coefficient of NM-resolvase at 260 nm (2500 M^{-1} cm^{-1})) is negligible with respect to that of site I DNA (467000 M^{-1} cm^{-1}), and so the absorbance of the X-synapse is essentially due to its DNA component. Re-separation of the X-synapse fraction (peak 2) by size exclusion chromatography indicated that it was at least 85% pure (data not shown). (B) Native polyacrylamide 'band-shift' assay of resolvase-site I complexes. Lane 1, 36 bp site I; lane 2, site I incubated with Tn3 resolvase; lane 3, natural site I incubated with NM-resolvase-S10A; lane 4, natural site I incubated with NM-resolvase; lane 5, phosphorothioate-modified 36 bp site I; lane 6, phosphorothioate-modified site I incubated with NM-resolvase; lane 7, peak (2) fraction from the size exclusion purification in 6.2A. Bands I, II and III represent uncomplexed site I DNA, a dimer of resolvase bound to site I DNA, and NM36 X-synapse, respectively. The X-synapse is not completely stable during a band-shift experiment, hence some uncomplexed site I DNA is observed in lane 7. Because the NM-resolvase is His-tagged and NM-resolvase-S10A is not, complex III in lane 3 runs slightly faster than in lanes 4-7. (C) Size distribution c(s) profiles from SV analytical ultracentrifugation data analysed with SEDFIT (Schuck, 2000) for NM36 X-synapse (*solid line*, as extracted from peak 2 in panel (A)), 36 bp site I DNA fragment (*dashed line*) and NM-resolvase (*dotted line*).

with solutions of NM-resolvase or 36 bp site I DNA.

The c(s) profiles of the three purified X-synapses show single peaks, with apparent sedimentation coefficients of 5.9 S (NM36) (Fig. 6.2C), 6.4 S (NM50), and 7.0 S (NM-GFP36) (data not shown). Finite element analysis using a two-species non-interacting model yields true sedimentation coefficients of 6.0 S, 6.5 S, and 7.0 S respectively, for the main species (>98%), which are at concentrations of >1 μM.

The program HYDROPRO (García de la Torre et al., 2000) was employed to simulate the sedimentation coefficients ($s_{20,w}$) of the structures deduced from the scattering data (see below) and previously proposed models for the X-synapse. A series of simulations of the sedimentation coefficients for the X-exp[SANS43], X-exp[SANS65] and X-exp[SAXS] models (see below) and the DNA-in^{HS} configuration of the NM50, NM-GFP36 and NM36 X-synapses were performed (see Table 6.1). For the DNA-out models of the X-synapses, the differences between experimental and modelled sedimentation coefficients (see Section 6.2.3) were $f^{out} \leq 18\%$, whereas these differences were $f^{in} \geq 28\%$ when the DNA-in^{HS} configuration was employed. These significant differences are the result of uncertainties in the calculations of the partial specific volumes ($\bar{v}$) for the protein-DNA complexes (Chalikian, 2003; Maluf and Lohman, 2003; Petry et al., 2001) that ultimately lead to the appearance of systematic errors in the modelled sedimentation coefficients. Also, other phenomena, such as the presence of non-homogeneous viscosity gradients during centrifugation (Schuck, 2003a) can cause systematic errors in the experimental determination of the sedimentation coefficient. A further investigation into these systematic discrepancies is outwith the scope of this study. All in all, the modelled sedimentation coefficients agreed better with DNA-out models of the X-synapses than with their DNA-in counterparts.

6.3.3 Small angle scattering

Four sets of small angle scattering data were collected.

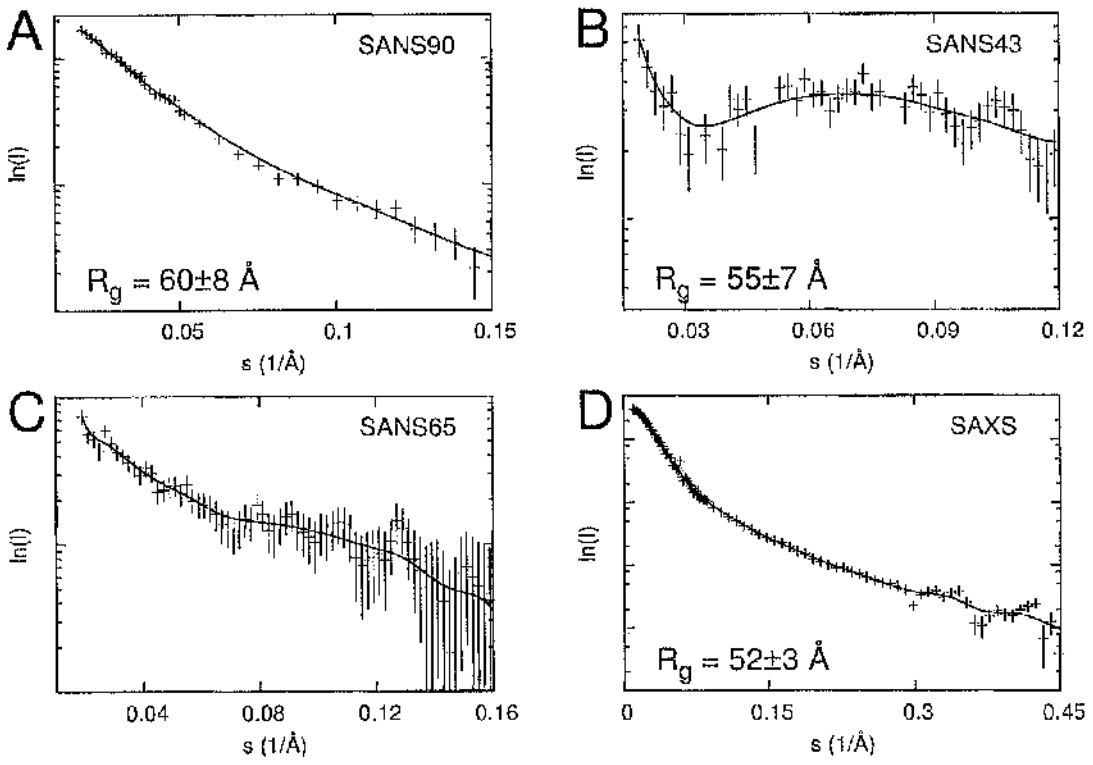

Figure 6.3: Experimental SANS and SAXS intensity data (*crosses*), with error bars (*vertical bars*) and a Bezier approximation as a 'guide to the eye' (*solid line*). Radii of gyration (R_g) were estimated by using the Guinier approximation (Guinier and Fournet, 1955). (A) Dataset SANS90: SANS data acquired at 90% (v/v) D_2O for NM36 X-synapse. (B) Dataset SANS43: SANS data acquired at 43% (v/v) D_2O for NM50 X-synapse. (C) Dataset SANS65: SANS data acquired at 65% (v/v) D_2O for NM-GFP36 X-synapse. R_g is not quoted as the experimental range did not fully extend over the Guinier region. (D) Dataset SAXS: SAXS data for NM36 X-synapse.

X-synapse sample	Experimental	X-exp		DNA-in^{IIS}	
	$s_{20,w}$ (S)	$s_{20,w}$ (S)	f^{out}	$s_{20,w}$ (S)	f^{in}
NM50	6.5	7.8	18%	8.9	31%
NM-GFP36	7.0	7.2	3%	9.9	34%
NM36	6.0	6.8	12%	7.9	28%

Table 6.1: Experimental and modelled sedimentation coefficients for the three X-synapses. f-values (see Section 6.2.3) represent the percentage discrepancies between experimental and modelled sedimentation coefficients.

(1) Dataset SANS90: SANS was used to study the low-resolution solution structure of the NM36 X-synapse, in a buffer with 90% (v/v) D_2O (Fig. 6.3A). Under these conditions, the scattering curves predicted from the DNA-in and the DNA-out models are markedly different only in the Guinier region ($0.01 < s < 0.03$ Å^{-1}).

(2) Dataset SANS43: The protein component of the NM50 X-synapse was contrast-matched by using SANS in a buffer with 43% (v/v) D_2O (Fig. 6.3B). At this D_2O concentration, most of the observed scattering is due to the DNA component in the X-synapse. A 50 bp DNA fragment was employed in this experiment in order to enhance the differences between the neutron scattering profiles predicted by the DNA-out and DNA-in models, as well as to increase the signal-to-noise ratio of the SANS profile.

(3) Dataset SANS65: The DNA component of the NM-GFP36 X-synapse was contrast-matched by using SANS in a buffer with 65% (v/v) D_2O (Fig. 6.3C). At this D_2O concentration the main contribution to the observed scattering is due to the protein component. When the DNA component is contrast-matched, models for the DNA-in and DNA-out configurations of the X-synapse containing NM-resolvase (not GFP-tagged) produce very similar simulated SANS curves (and accordingly similar R_gs). The use of NM-GFP, with its large GFP domain attached to the NM-resolvase C-terminus, greatly increases the structural differences between the protein components of the two configurations. A DNA-out

model incorporating NM-GFP is more elongated and less globular than a DNA-in one, regardless of the exact positions of the GFP domains (see below).

(4) Dataset SAXS: SAXS was used to analyse the low-resolution solution structure of the NM36 X-synapse (Fig. 6.3D).

The direct modelling procedure for analysis of the scattering data was similar to that used in Section 5.4.2.1. The object under study (in this case the X-synapse) is parameterised by regarding it as being composed of a small number of sub-structures for which structural information is available (e.g. crystal structures). Model structures are then made by transformations (translations and rotations) of these substructures. Each structural model is thus defined by a small number of parameters, such as the distances and angles between the substructures. A computer program *domain-rot* (see Section 6.2.4) is used to produce a large number of structures ($\geq$ 2000) that map a wide range of the parameter space. The scattering profile and its χ (fit to the experimental data) are calculated for each structure by using either CRYSOL (for SAXS) or CRYSON (for SANS) (Svergun et al., 1995). The set of parameters that generates the scattering profile with the best fit to the experimental data is identified by visual inspection of the contour plots of χ. The criterion used here for estimation of the uncertainties in the best-fit parameters was identical to that employed in chapter 5.4.2.1. Tn3 resolvase and $\gamma\delta$ resolvase have 82% identity at the amino acid level, and are treated in this study as identical. The structural differences between $\gamma\delta$ resolvase and Tn3 resolvase, and between Tn3 resolvase and NM-resolvase, are likely to be insignificant at the resolution of the experiments presented here. The structural models that best represent the experimental datasets are defined as X-exp[SANS43], X-exp[SANS65] and X-exp[SAXS] (see below).

6.3.3.1 The X-synapse has a DNA-out configuration, with kinked DNA

First, the configuration of the DNA component of the X-synapse will be analysed. In dataset SANS43, the measured scattering is primarily from the DNA

component of the NM50 X-synapse (see above). Therefore, initial models were built with two 50 bp site I DNA fragments. In order to test whether alteration of the kink angle at the centre of site I would affect the results, a series of five alternative site I conformations were constructed, with kink angles of $\psi = 0°$ ($DNA50_0$), $\psi = 20°$ ($DNA50_{20}$), $\psi = 40°$ ($DNA50_{40}$), $\psi = 60°$ ($DNA50_{60}$) and $\psi = 70°$ ($DNA50_{70}$). For each simulation, models containing two copies of $DNA50_\psi$ were created by antiparallel alignment of the DNA fragments, followed by variation of distance and angular parameters (d, φ), as shown in Fig. 6.4A: d is the distance between the DNA fragments in the complex, whereas φ is the angle of rotation of one $DNA50_\psi$ with respect to the other in the z-axis direction ($\varphi = 0°$ when the obtuse angle between the $DNA50_\psi$ fragments is about 120°, becoming larger for positive φ-values; see definitions in Section 6.2.5). Parameter values of $(d, \varphi) = (-31\,\text{Å}, -75°)$ generate a 'hand-shake' DNA-in configuration (DNA-in^{HS}) (Yang and Steitz, 1995; see Fig. 6.1C), whereas $(d, \varphi) = (63\,\text{Å}, 0°)$ generate a 'back-to-back' DNA-out configuration (DNA-out^{BB}) as proposed by Sarkis et al. (2001) (see Fig. 6.1B). Several extended explorations of the parameter space in the ranges $-180° < \varphi < 180°$ and $-200 < d < 200$ Å were performed. An example, showing contour levels of χ versus d and φ for $DNA50_{20}$ (with $\psi = 20°$) is presented in Fig. 6.5A. The global χ minimum is located at $(d, \varphi) = (75 \pm 15\ \text{Å}, 60 \pm 90°)$ (Fig. 6.5A; note that the contour plot is periodic in φ, with period 180°), consistent with a DNA-out configuration. A secondary local minimum at $(d, \varphi) = (-100 \pm 20\,\text{Å}, 60 \pm 90°)$ would represent a DNA-out model where the DNA fragments are kinked towards each other. However, the fit to the experimental data at this minimum is much poorer, and the structure is physically unrealistic, having a d-value so large that the two Tn3 resolvase subunits would not meet each other. Note that χ is sensitive to variation in the angular parameter φ only near $d = 0\,\text{Å}$. The global minimum at $d = 75$ Å is thus virtually independent of φ.

Simulations using site I DNA fragments with other kink angles gave similar results, in that they all yielded a global minimum in χ at $d = 75 \pm 15$ Å with little sensitivity to φ. Figure 6.5B shows plots of χ versus d for $DNA50_0$, $DNA50_{20}$ and $DNA50_{70}$ at $\varphi = 60°$ (other values of φ are less informative, as seen in Fig. 6.5A).

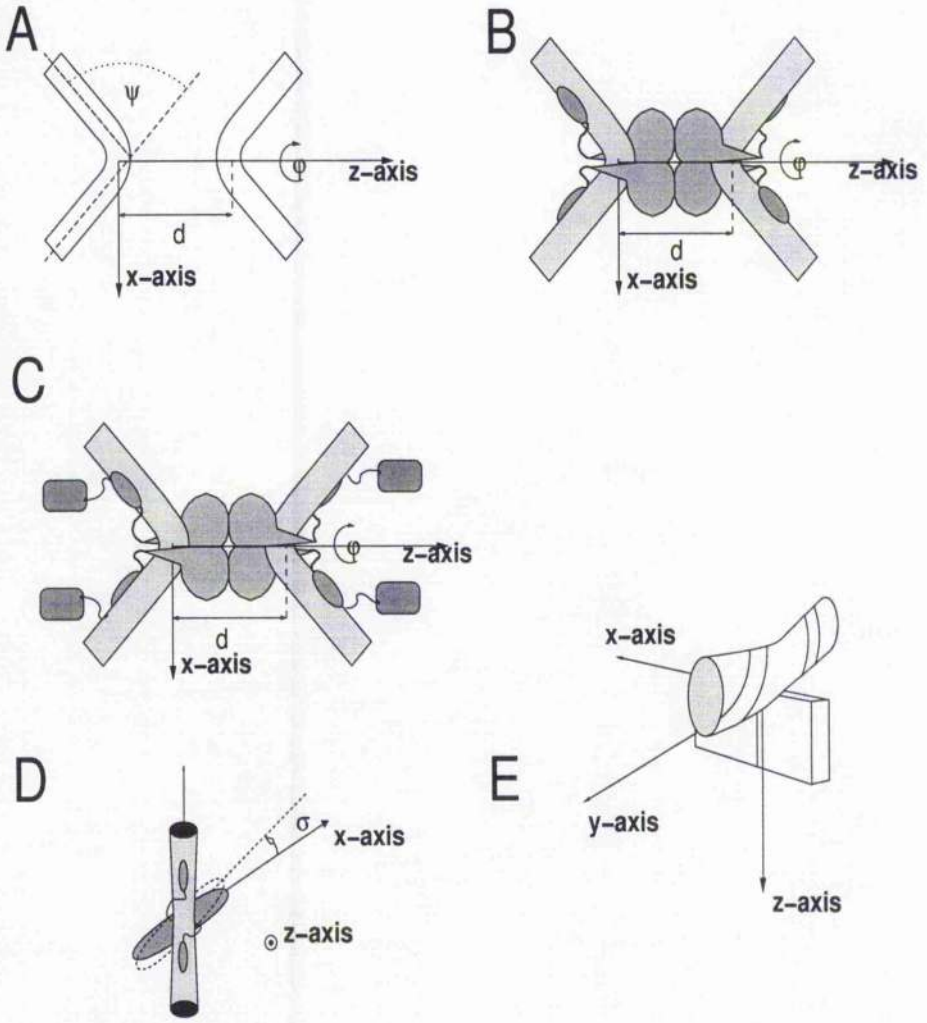

Figure 6.4: Diagrams showing the variables used in direct modelling of the SAXS/SANS data. (A) Two DNA fragments, kinked by an angle of ψ, and separated by a distance d, are used to model the NM50 X-synapse SANS data, in which the protein is contrast-matched (dataset SANS43); (B) Two 1gdt structures separated by a distance d and rotated relative to each other by an angle φ in the z-axis are employed to model the SAXS data from NM36 X-synapse (dataset SAXS) (see panel D for definition of σ, also used in modelling the SAXS data); (C) Two 1gdt-GFP36 structures are used to model the NM-GFP36 X-synapse SANS data (dataset SANS65); (D) σ represents the angle of rotation of the DNA and resolvase CTDs with respect to the resolvase NTDs about the z-axis. (E) Definition of coordinate axes, as used in all simulations, with respect to a representation of 1gdt in which the catalytic domain dimer is shown as a 'domino', and the DNA as a bent cylinder.

The χ-value at the global minimum varies with the kink angle ψ (Figs. 6.5B and C), having a minimum at $\psi = 30 \pm 20^o$. None of the DNA-in models produces acceptable fits to the experimental data (Fig. 6.5A, and data not shown).

NM50 X-synapse models that include the protein were built similarly, using both protein and DNA coordinates of a modified structure made by adding two straight 7 bp B-DNA fragments to the ends of the site I DNA in the co-crystal structure 1gdt (this object is called 1gdt50). The site I DNA fragment in 1gdt is kinked by about 60^o (Yang and Steitz, 1995). Predicted scattering curves for the models were compared with the NM50 experimental dataset SANS43, as described above for the DNA-only analysis (Fig. 6.5D). The global minimum is located at $(d, \varphi) = (55 \pm 10$ Å$, 50 \pm 50^o)$. This particular model is referred to as X-exp[SANS43] (see the fit to the experimental data in Fig. 6.6A). Models that include the protein component of the X-synapse produce better fits to the experimental data ($\chi = 0.85$) than those that do not ($\chi = 1.7$) (compare Figs. 6.5A and 6.5D). X-exp[SANS43] resembles the DNA-out^{BB} model. Again, DNA-in models do not produce acceptable fits to the experimental data (Figs. 6.5D and 6.6A).

6.3.3.2 Refinement of the model: SAXS data analysis

SAXS data were acquired for the X-synapse NM36. The procedures for comparing the experimental data with scattering curves generated from models, based on two copies of the $\gamma\delta$ resolvase-site I co-crystal structure 1gdt, were essentially as described above. The global minimum in χ is located at $(d, \varphi) = (64 \pm 3$ Å$, 25 \pm 10^o)$, and is sensitive to variation in both d and φ (data not shown). The minimum ($\chi \approx 4.0$) corresponds to a DNA-out model, in which the alignment of the two resolvase dimers is twisted about the z-axis by an angle $\varphi = 25^o$ relative to that of the DNA-out^{BB} model.

The N-terminal (catalytic) and C-terminal (DNA-binding) domains of resolvase in 1gdt are joined by a short sequence of amino acids (V138-I146) in an extended conformation, which might be flexible enough to allow considerable

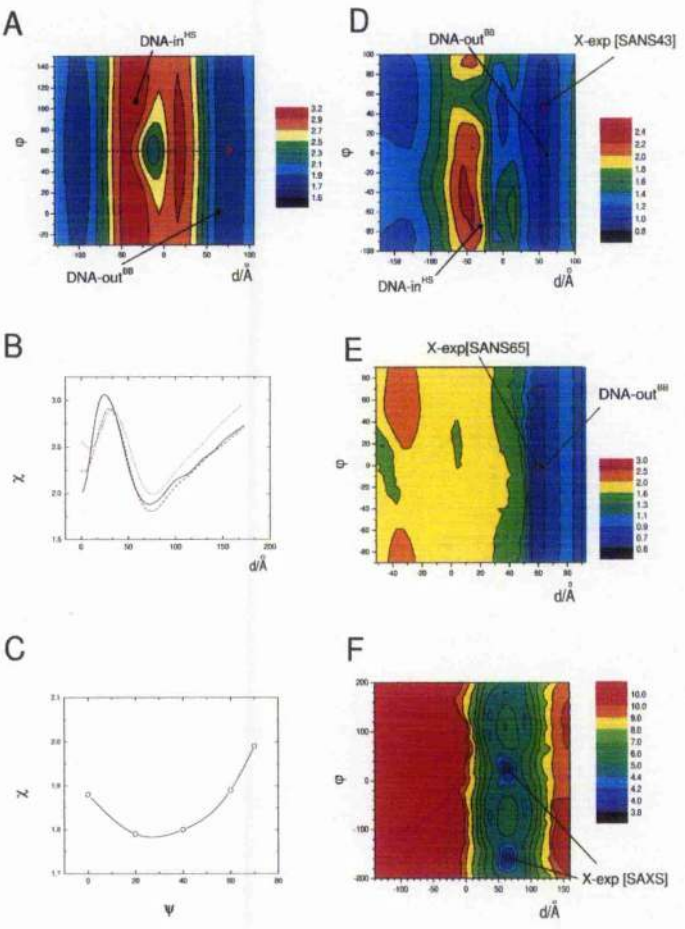

Figure 6.5: Direct modelling of small angle scattering data for NM50 (A to D), NM-GFP36 (E) and NM36 (F) X-synapses. In A, D, E and F black circles identify the position of the DNA-out^{BB} and DNA-in^{HS} configurations and red circles indicate the position of the structural models that best represent the experimental datasets (X-exp). (A) Contour plots of χ as a function of d and φ for NM50 X-synapse SANS data (dataset SANS43) modeled with two DNA_{20} fragments (see text for more details). (B) Variation of χ with d at $\varphi = 60^o$ for $DNA50_0$ (black solid line), $DNA50_{40}$ (red dashed line), and $DNA50_{70}$ (blue dotted line). (C) χ-value at the global minimum plotted versus the kink angle ψ. (D) NM50 X-synapse SANS data modeled with two 1gdt50 structures. The structure with minimum χ is defined as X-exp[SANS43]. (E) NM-GFP36 X-synapse SANS data (dataset SANS65) modeled by employing two 1gdt-GFP units with angles $(\theta, \phi) = (-35^o, 5^o)$ (see text for more details). The structure with minimum χ is defined as X-exp[SANS65]. (F) NM36 X-synapse SAXS data (dataset SAXS) was modeled by using two $1gdt_{\sigma=8}$ units. The two minima in χ (with φ differing by 180^o) correspond to the same low-resolution structure, defined as X-exp[SAXS].

freedom of movement of the DNA (and the small C-terminal domains bound to it) relative to the catalytic domains (see also Leschziner and Grindley, 2003). Two sub-structures were generated from 1gdt by 'cutting' the protein subunits after residue V139 (the positions marked by arrows in Fig. 6.1A), thereby separating the dimer of N-terminal catalytic domains (NTDs) and the C-terminal domains (CTDs) bound to the site I DNA. New $1gdt_{\sigma}$ structures were then generated by rotation of the DNA and CTDs with respect to the NTDs about the z-axis by an angle σ (see Section 6.2.5 for more details).

Simulations were performed by variation of σ between -90^o and 90^o in initial X-synapse models with fixed values of d and φ, ranging from 50 to 90 Å and from 0 to 90^o, respectively (data not shown). The minimum χ (3.6) was found when $(d, \sigma, \varphi) = (64 \pm 3$ Å$, 8 \pm 3^o, 20 \pm 10)$. Therefore, the angle $\varphi = 25^o$ found when 1gdt was used for the modelling may be regarded as being the result of two types of structural distortion: a twisting about the z-axis of the resolvase catalytic domain dimers by $\varphi = 20^o$, relative to their positions in the DNA-out^{BB} model, and a rotation of the two site I DNA-CTD substructures by $\sigma = 8^o$ relative to the NTDs in 1gdt. Together, these contributions change the angle of alignment of the two DNA fragments from 120^o in DNA-out^{BB} to 124^o; that is, they become slightly closer to antiparallel (or parallel, in low-resolution terms).

Finally, X-synapse models were generated using the structural unit $1gdt_{\sigma=8}$, in which the site I DNA is twisted relative to the catalytic domains as indicated in the previous analysis, instead of 1gdt. The global minimum in χ was found at $(d, \varphi) = (64 \pm 3$ Å$, 20 \pm 10^o)$ (Fig. 6.5F). This refined model, with parameters $(d, \sigma, \varphi) = (64$ Å$, 8^o, 20^o)$ is hereafter referred to as X-exp[SAXS]. The fits to the SAXS experimental dataset predicted from this model, and, for comparison, to a 'hand-shake' DNA-in model (DNA-in^{HS}), are presented in Fig. 6.6C.

Simulations were also performed to explore less symmetric models for the X-synapse, where the pseudo-dyad axes of the two resolvase dimers do not coincide. However, no good fits to the experimental data were found (data not shown).

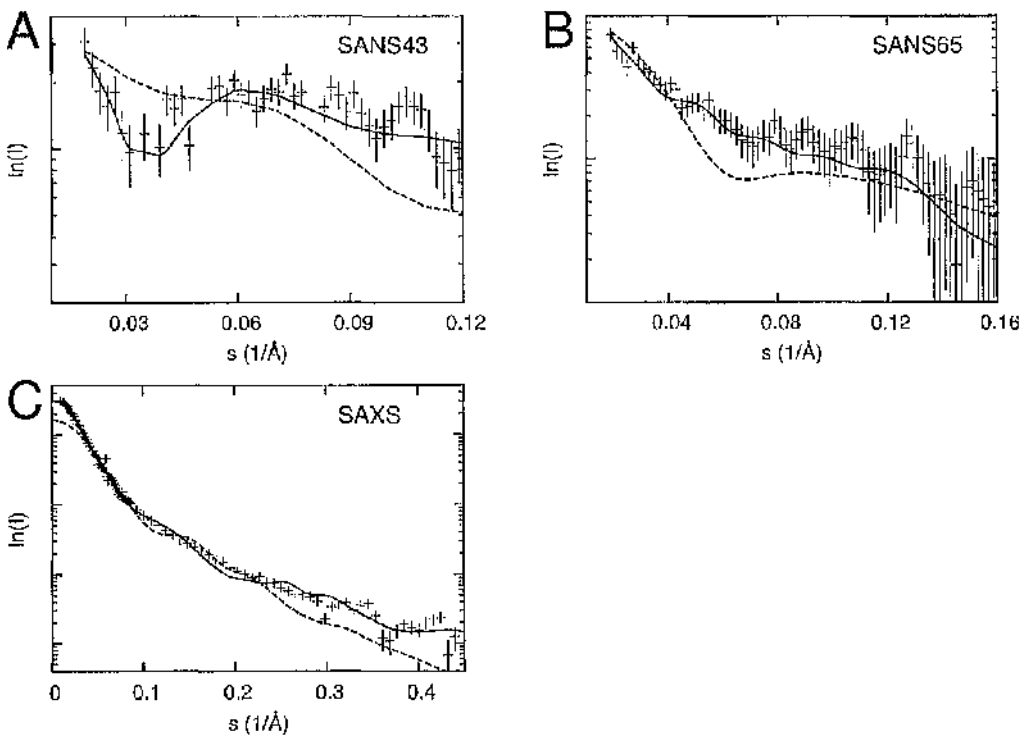

Figure 6.6: Small angle scattering experimental data (*crosses*) with error bars (*vertical bars*) and best fits with the models. Best fits of simulated curves from the X-exp experimentally derived models are shown as solid lines, and fits with the DNA-in^{HS} model are shown as dashed lines. (A) Dataset SANS43; model X-exp[SANS43]. (B) Dataset SANS65; model X-exp[SANS65]. (C) Dataset SAXS; model X-exp[SAXS].

6.3.4 Modelling of the X-synapse structure containing GFP domains

The NM-resolvase-GFP fusion protein NM-GFP was used to make it easier to detect differences between DNA-out and DNA-in configurations of the protein component of the X-synapse (see above) The direct modelling approach was employed to analyse the SANS curve of the NM-GFP36 X-synapse (dataset SANS65, 65% D_2O, contrast-matching the DNA). In this case, the unknowns were not only the relative positions of the 1gdt structural units in the X-synapse but also the relative positions and orientations of the GFPs. A two-step procedure was thus used to analyse this dataset.

First, the positions of the GFPs were varied (whilst constrained by their flexible linkage to the resolvase C-terminus), while two 1gdt structural units were locked into either a X-exp[SAXS] or a DNA-in^{HS} conformation. In this way, the position of the GFPs that best fitted the experimental data for each of these two possible X-synapse architectures was found. Calculation of the fit to the experimental data included a term to penalise structures in which there was overlap between the electron density of the GFP domains and that of other parts of the structure (see Section 5.4.2.1 and Appendix B). Later, having found the best-fit positions of the GFPs for the X-exp[SAXS] and DNA-in^{HS} configurations, the GFPs were then fixed at these positions in each 1gdt-GFP unit, and the relationship of the two 1gdt-GFP units was varied (see below).

In the first approach, the structure of NM-GFP36 X-synapse was assembled using either X-exp[SAXS] or DNA-in^{HS}, and four GFP (Brejc et al., 1997) subunits. Initially, one GFP was positioned with its first residue at 15 Å from N183 (last residue in $\gamma\delta$ resolvase) and its main axis of inertia in an orientation perpendicular to the plane defined by the resolvase-site I DNA molecular surface close to N183 (z'-axis) (see Fig. 4D). The x' and y' axis are arbitrarily defined perpendicular to z'. A large number of positions for this GFP subunit was generated by applying discrete rotations (ϕ and θ) about the x' and y' axes, respectively (Fig. 6.7A). The other GFPs were generated by 180° rotations around the x-, y-

and z-axis of the X-synapse, respectively. Structures in which there was overlap between the electron density of the GFP domains and that of other parts of the structure were penalised as described in Chapter 5 and in Appendix B. The parameters ϕ and θ were varied between -180 and 180°, and a large number of structures with different combinations of these parameters was calculated.

The optimal positions for the GFP subunits in a NM-GFP36 X-synapse built from DNA-in^{HS} are $(\theta, \phi) = (30 \pm 6^o, -35 \pm 10^o)$ (Fig. 6.7B, shaded regions implying heavy overlaps). This model has a high value of χ; regardless of the position of the GFPs, a DNA-in model for the NM-GFP36 X-synapse cannot fit the experimental data acceptably (Fig. 6.6B).

Similar simulations based on X-exp[SAXS] produced a single global minimum at $(\theta, \phi) = (-35 \pm 10^o, 5 \pm 5^o)$ and two other local minima with higher χ-values (Fig. 6.7C). The scattering curve simulated from the model at the global minimum fits the experimental SANS data very well (Fig. 6.6B).

In summary, the direct modelling presented in this section shows that a DNA-out model based on X-exp[SAXS] is able to fit the SANS experimental data for the NM-GFP36 X-synapse, whereas a model based on DNA-in^{HS} does not fit the data.

In the second approach, the spatial relationship of the resolvase dimers was considered. Having found the best-fit positions of the GFPs for the X-exp[SAXS] configuration, the GFPs were then fixed at these positions in each 1gdt-GFP unit, and the relationship of the two 1gdt-GFP units was varied. The distance between the resolvase dimer units in the model with optimal fit to the experimental data (defining the parameters for X-exp[SANS65]) (d) was 58 ± 8 Å, corresponding to a DNA-out model very similar to that deduced in the previous sections (Fig. 6.5E, fit to experimental data in Fig. 6.6B). However, the fit at the minimum was insensitive to variations in the angle of alignment on the z-axis (φ). No models based on a DNA-in configuration could be fitted acceptably to the experimental data (Fig. 6.6B, and Fig. 6.7). Equivalent results were obtained when only the protein content of 1gdt and the corresponding GFP subunits were used for the

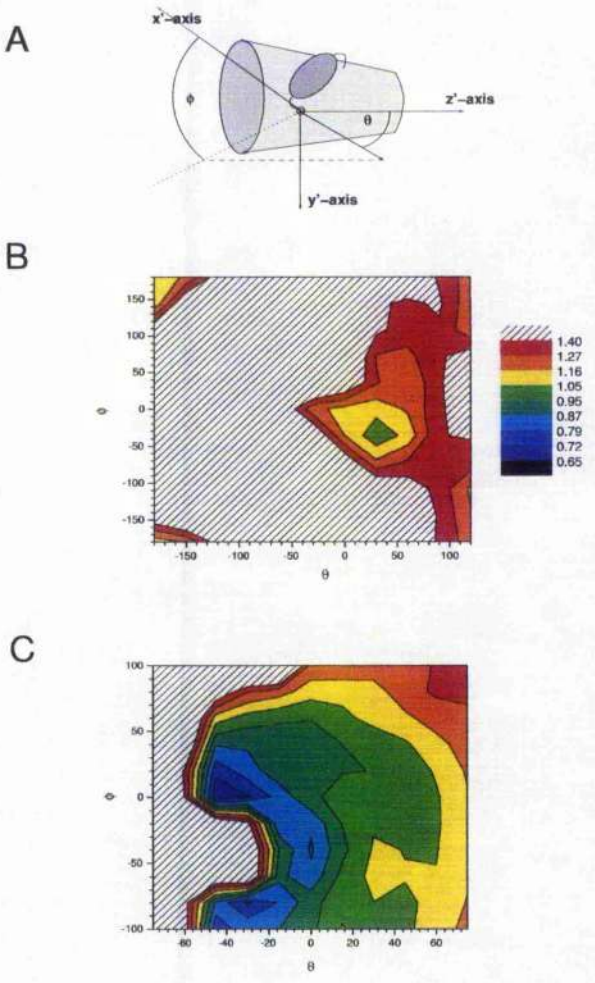

Figure 6.7: Optimisation of the position and orientation of the GFP domains in an X-synapse model. (A) GFP subunits are modelled into the NM-GFP36 X-synapse structure by adding one GFP to each C-terminus of NM-resolvase ('oval shape'; in the x'y'z' coordinate system) using the spherical coordinates ϕ and θ (see text for more details). The site I DNA fragment is represented as a 'cylinder'. (B and C) Contour plots of χ as a function of ϕ and θ for dataset SANS65 with the GFP domains attached to the DNA-in^{HS} (B) or X-exp[SAXS] (C) model structure (see text for details). Areas corresponding to structures having substantial steric clashes are shaded.

modelling (data not shown).

6.4 Final Model

The final model, illustrated in Fig. 6.8B, is built by combining the parameters from the best models X-exp[SANS43], X-exp[SANS65] and X-exp[SAXS] (the new d-value was calculated as an error-weighted average of the different d-values), resulting in a structure with $(d, \sigma, \varphi) = (62\,\text{Å}, 8^o, 20^o)$.

6.5 Discussion

The solution structure of the X-synapse has been modeled from small angle x-ray/neutron scattering data. In summary (see Fig. 6.8A), the data indicate that:

1. The X-synapse has a DNA-out structure, with the site I DNA segments separated by about 62 Å .
2. The site I DNA is kinked or bent by about 30°, away from the resolvase catalytic domains.
3. The angle σ between the dimerised resolvase catalytic domains and the site I DNA differs slightly from that in the crystal structure (1gdt) of a site I-resolvase dimer complex.
4. The two resolvase catalytic dimers are docked in a configuration distinct from that proposed by Sarkis et al. (2001).
5. Models with the DNA on the inside of a resolvase tetramer, or other arrangements such as a side-by-side alignment of the two dimers, are not supported by the data presented in this chapter.

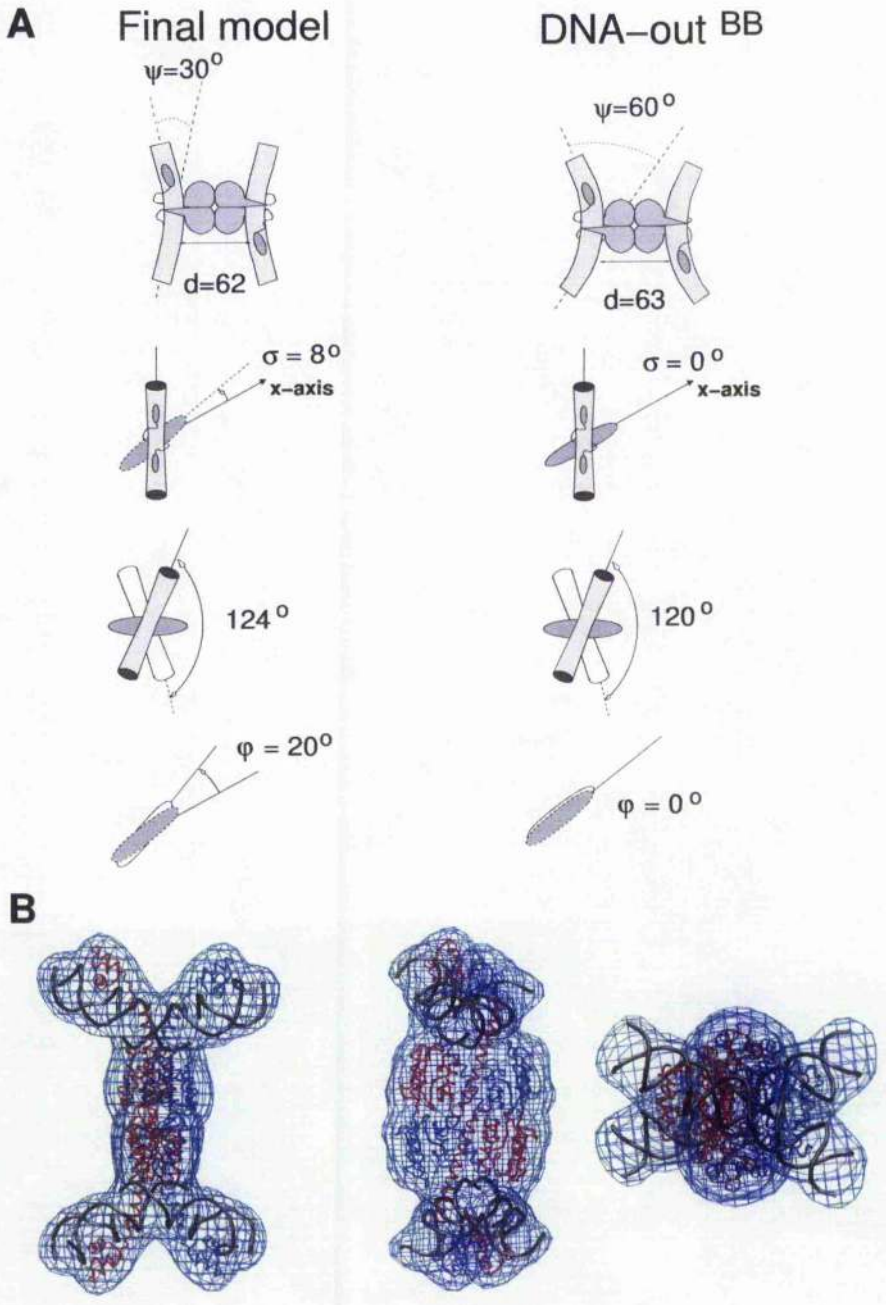

Figure 6.8: The final model for the X-synapse solution structure. (A) Schematic representations comparing the different parameters obtained for the final model and those for DNA-out^{BB}. (B) Mesh representation of three views of the final model for the X-synapse produced by docking two 1gdt structures (ribbons) using the parameters $d = 62\mathring{A}$, $\sigma = 8^o$, and $\varphi = 20^o$.

Mutation studies (Burke et al., 2004), *in vitro* experiments (Sarkis et al., 2001), and crystallography-based modelling (Sarkis et al., 2001) have previously been used to support a DNA-out model for the X-synapse. From *in vitro* studies with conformationally restricted DNA substrates, Leschziner and Grindley (2003) proposed that the X-synapse has a DNA-out architecture, with an angle between the site I DNA segments of approximately 150°, in reasonable agreement with the results presented in this chapter, which indicate an angle of about 124° (Fig. 6.8A). However, this study also suggested that the angular relationship between the dimerised catalytic domains and the site I DNA in the X-synapse is $\sigma \geq 70°$, a value inconsistent with the results presented here, which indicate $\sigma = 8°$ (Fig. 6.8A). A possible explanation for these discrepancies is that the experiments of Leschziner and Grindley might report on a later synaptic intermediate which has undergone further conformational changes after formation of the X-synapse, in order to initiate catalysis of strand exchange.

The method used here for structure determination, which involves fitting scattering data to models built from known structural units, has implications for the strength of certain aspects of the conclusions made here. The approximate shape and relationship of the two DNA segments in the X-synapse can be assigned with high confidence (dataset SANS43), because it is virtually certain that the structure of the DNA will show only minor variations (e.g. bends) from normal A/B form; it can therefore be modeled reliably. Likewise, It is very likely that the resolvase tetramer is formed by interactions of the catalytic domains of the two dimers, with the C-terminal domains on the outside of the tetramer, because the scattering data with NM-GFP resolvase (dataset SANS65) are definitely incompatible with alternative models (Fig. 6.5E and 6.7). Therefore, the analysis of datasets SANS43 and SANS65 strongly supports a DNA-out structure for the X-synapse. The predictions regarding the detailed shape of the tetramer of resolvase N-terminal domains are less certain, because some protein conformational changes upon synapsis might be expected, whereas the structure has been modeled by juxtaposing dimers folded as in the crystal structure of Yang and Steitz (1995) (1gdt), at various distances and angles. Published high-resolution structures do not reveal what structural changes might be expected upon synapsis,

and therefore more detailed modelling would have been very speculative. Current methods for *ab initio* computational determination of biomolecular structures from scattering data are still too primitive to make useful predictions of the shape of the protein in the X-synapse (see Section 2.2.3.2).

6.5.1 Implications for synapsis and strand exchange

In order to obtain an X-synapse that was stable enough for solution structural analysis, a mutant version of Tn3 resolvase had to be used. The mutations in NM-resolvase were identified as allowing resolvase to recombine at site I, in the absence of the accessory resolvase-binding sites II and III (Burke et al., 2004). The natural X-synapse might be thermodynamically unstable in the absence of the remainder of the natural Tn3 synaptic complex. Alternatively, the mutations might remove a natural kinetic barrier to stable synapsis of two site I-resolvase dimer complexes. Interestingly, the wild-type Hin invertase, a recombinase related to resolvase, does make relatively stable synapses of two dimer-bound sites prior to catalysis (Heichman and Johnson, 1990). This synapse is predicted to be structurally very similar to the X-synapse reported here. A barrier to interaction of the invertase dimer-bound sites is thus apparently not an important regulatory feature of this system, whereas it may be so for resolvase.

The X-synapse structure deduced in this chapter, and other results (Burke et al., 2004; Sarkis et al., 2001) are consistent with formation of a DNA-out X-synapse following an initial interaction of two dimers at a surface close to the mutated residues S101, Y102, I103, and L105 in NM-resolvase. The analysis presented suggests that the tetramer thus formed is not greatly distorted from its predicted form at the initial contact of two 1gdt units. In contrast, new crystallographic structures of resolvase synaptic tetramers, whilst confirming the DNA-out architecture, show substantial alterations of the protein tertiary structure, as well as cleavage of both strands in each site I (N.D.F. Grindley, personal communication). The species studied here is an intermediate prior to DNA cleavage, which is

presumably yet to undergo further transformations associated with the catalytic steps of the reaction.

A DNA-out architecture for the X-synapse is consistent with 'subunit rotation' or 'domain-swapping' mechanisms for resolvase-mediated strand exchange, whereas mechanisms that require a DNA-in architecture can be eliminated (see also Burke et al., 2004). The centres of the two recombining sites are about 62 Å away from each other in the final X-synapse structure. Nevertheless, during catalysis the cleavages and ligations of strands at the two sites are apparently highly coordinated; cleaved substrate DNA is not observed under normal *in vitro* reaction conditions (McIlwraith et al., 1997). This suggests that there must be an ingenious mechanism for long-distance coupling of the cleavage and rejoining events. The X-synapse structure presented also confirms that the mechanism of strand exchange by serine recombinases must be fundamentally different from that of the tyrosine recombinases (see Introduction); an especially obvious distinction being the large movements of DNA segments that will be required by serine recombinases to reach the recombinant configuration.

The X-synapse is the catalytic moiety of the larger complex ('the synapse') involved in natural cointegrate resolution (see Introduction). The synapse also includes a regulatory part, comprising the intertwined accessory binding sites (II and III) of the two participating *res* sites, together with the presumed eight resolvase subunits bound to them. The X-synapse structure is broadly compatible with the recent model of Sarkis et al. (2001) for the full synapse (see Section 1.8), which proposes that the intertwined *res* DNA is wrapped around the outside of a core filament of three 'DNA-out' resolvase tetramers, contacting each other via a previously characterised 2-3' interface.

Chapter 7

A global multi-technique approach to study low-resolution solution structures

7.1 Summary

Finding the conformation of large macromolecular complexes has become an important problem in structural biology, not always solved by high-resolution techniques such as x-ray crystallography and NMR spectroscopy. Solution biophysical properties, on the other hand, provide direct or indirect structural information on the whole complex. A general systematic approach to construct a structural model of the macromolecule that agrees with all the experimental solution properties is currently lacking. In this chapter, we present such an approach, where generalised rigid-body modelling is combined with a Monte Carlo/simulated annealing optimisation method, to search over a large range of possible conformations for the structure that best fits solution experimental properties derived from small angle scattering, fluorescence resonance energy transfer, and analytical ultracentrifugation.

7.2 Introduction

Finding the conformation of large multi-protein complexes, DNA-protein complexes and multi-domain proteins is of fundamental importance in understanding a large number of biologically significant problems. High-resolution techniques, such as x-ray crystallography and nuclear magnetic resonance (NMR), had proven to be paramount in unveiling the structures of many of these macromolecular complexes. In some cases, however, these techniques fall short in providing such information. Often, crystals of large macromolecular complexes are extremely difficult to obtain, or they provide information only on segments of the complexes. Crystal packing effects, especially in these large systems, are always a concern since they might substantially affect the architecture of the complex with respect to that in solution. Crystals of a protein in all its relevant conformations, particularly when one or more states in the functional pathway exhibit flexibility, are very difficult to manufacture. Similarly, NMR-derived distance information on large complexes, specifically distances between subunits, is very demanding and scarce. In addition, flexible or disordered regions may appear in many cases to be absent in the final structure. Such regions may involve receptor binding motifs, loops involved in the active site or antigenic epitopes, to cite but a few possibilities, and might have a significant relevance for understanding the biological processes themselves.

Solution techniques, such as fluorescence resonance energy transfer (FRET) (Stühmeier et al., 2000), small angle x-ray and neutron scattering (SAXS/SANS) (Feigin and Svergun, 1987) and analytical ultracentrifugation sedimentation velocity (Lebowitz et al., 2002) provide independent direct or indirect structural information on particulate systems. These techniques are well established and have been successful in retrieving structural information on large macromolecular complexes. The structure of the ribosome was studied by Svergun and Nierhaus (2000) using small-angle scattering (SAS), and more recently by Gilbert et al. (2004) using c-EM. Reconstruction of biologically significant DNA structures, such as Holliday junctions and DNA bulges, have been derived from FRET (Lilley and Clegg, 1993) and SAXS (Nöllmann et al., 2004b) data. Hydrodynamic

methods and computational modelling have been employed to study the low resolution conformations of human IgG subclasses by investigating the relative spatial orientation of their Fab' and Fc domains (Carrasco et al., 2001).

Ab initio methods can be used to restore the low resolution conformation of a particle (Chacón et al., 1998; Heller et al., 2003; Stuhrmann, 1970; Svergun, 1999; Svergun et al., 2001) from the small angle scattering (SAS) profile. Recently, a computational method has been developed to add missing loops and domains to protein models (Petoukhov et al., 2002). These restoration methods have been successfully applied to a large variety of problems (see review in Koch et al., 2003). When an *ab initio* reconstruction algorithm is run several times using the same initial conditions, the outcome is, at best, a family of reconstructed models with minor/moderate conformational differences. In this case, the reconstructions can be overlapped and averaged, and the averaged model can be used as a seed for a new reconstruction (Kozin and Svergun, 2001; Svergun et al., 2001). In another approach, the family of reconstructed models can be used to find a consensus model that captures the essential features of the individual models (Heller et al., 2003). Alternatively, SV data have been employed in order to filter *ab initio* reconstructed models that fit the SAXS data but fail to agree with the experimental sedimentation coefficient (Nöllmann et al., 2004b) or to gain a greater level of confidence in the retrieved *ab initio* models (Ackerman et al., 2003; Scott et al., 2002). Nonetheless, when applied to macromolecules with anisometric or hollow shapes, *ab initio* reconstruction methods produce a variety of considerably different reconstructions that fit the SAS data equally well (false positive reconstructions) (Heller et al., 2003; Rosenzweig et al., 1993; Volkov and Svergun, 2003).

Provided that the high- or low-resolution structures of the subunits comprising the quaternary complex are known or can be modelled, their arrangement can be found by searching for the quaternary conformations that best fit the SAS experimental data. Variations of this rigid-body modelling approach have been successfully applied to model SAS data (Aslam et al., 2003; Feil et al., 2001). The use of available structural information has a major advantage, with respect to *ab initio* reconstruction methods, in that the number of false positive reconstructions

can be greatly reduced. NMR has been combined with SAXS data to determine the relative positionings of calmodulin (CaM) and trifluoperazine (TFP) in their complex (Mattinen et al., 2002). Docking algorithms were used by de Azevedo et al. (2003) to generate geometrically compatible quaternary structures of purine nucleoside phosphorylase based on available crystallographic information. The best structural models were selected by finding the highest correlation between modelled and experimental SAXS data. These studies, however, did not take into account excluded volume and hydration effects in simulating the scattering data from models, demonstrated to have large systematic influences in the evaluation of SAXS profiles and consequently the radius of gyration (R_g) (Svergun et al., 1998).

Rigid body refinement was employed by Petoukhov et al. (2003) to improve models for *Azospirillum brasilense* glutamate synthase holoenzyme based on *ab initio* reconstructions. SAXS and SANS were used by Aslam and co-workers (Aslam et al., 2003; Aslam and Perkins, 2001) to study the solution structure of Factor H, composed of 20 short consensus repeat domains, for which there were high-resolution data available. Molecular dynamics simulations (MD) were employed to predict the possible conformations of the linkers in solution, and these were subsequentially used to produce models for the whole particle. A large number of structures were chosen by fitting, in the Guinier region, the simulated scattering profiles of the models to the experimental data, and discarding structures whose sedimentation coefficients were considerably different from the experimental one. This procedure, however, produced conformations for the whole macromolecule that were fundamentally restricted by the structures of the linkers. Most importantly, this method is not of general application and has been shown to generate multiple conformations equally fitting the available experimental data.

In the method presented in this chapter, the macromolecule is firstly divided into any number of structural domains. These domains are then combined by using a small number of parameters, such as interdomain distances and angles, in order to generate a large number of possible conformations of the macromolecule. The solution properties, such as SAS profiles, FRET distances and sedimentation

coefficient, are then predicted *in silico* for each conformation of the macromolecule and compared to the experimental datasets. Finally, a general systematic method that searches for the parameters producing the model whose solution properties best fit the available experimental datasets is described.

In the methodology presented in this chapter, the different structural domains comprising the macromolecule are combined, using a rigid-body approach, to generate a large number of possible conformations for the macromolecule. The solution properties of each conformation are simulated and compared to the experimental values. The conformation that best globally satisfies all the experimental datasets is found by using a search algorithm based on Monte Carlo/simulated annealing. As a model system to validate this methodology, a series of DNA molecules was chosen composed of three double-stranded DNA (dsDNA) helical fragments (H_{16}, H_n and H_{17}, with n being either 9 or 14) with a single stranded loop of five adenosines (A_5 bulge) between each dsDNA fragment, namely $H_{16}A_5H_9A_5H_{17}$ and $H_{16}A_5H_{14}A_5H_{17}$. DNA bulges may arise in natural DNA from recombination between imperfectly homologous DNA sequences or from errors in DNA replication. They play an important role in frame-shift mutagenesis (Stassinopoulos et al., 1996) and in specific interactions with RNA-binding proteins (Weeks and Crothers, 1991). Amongst other techniques, FRET (Gohlke et al., 1994) and NMR (Dornberger et al., 1999) have established that a single 5-adenosine (A_5) bulge introduces a defined kink into the DNA helical axis of about $\approx 90 \pm 15^o$. Stuhmeier et al. (2000) employed FRET on DNA structures containing two A_5 bulges ($H_{16}A_5H_nA_5H_{17}$, with $6 < n < 11$), similar to those used in this study, and showed that the distance between DNA ends is the shortest in the sample with $n = 9$, for which the dsDNA fragments H_{16} and H_{17} were proposed to be coplanar. In this chapter, the developed methodology will be validated by restoring, from simulated SAXS, SV, and FRET datasets, the low-resolution structures of two bulged DNA samples with very distinct conformations.

The chapter is organised as follows. The Computational Methodology section describes in some detail the computational algorithms and approximations employed. The Results section is devoted to the evaluation of the algorithm

in restoring low-resolution structures from simulated and experimental datasets. Finally, a few conclusions are presented in the Discussion section.

7.3 Computational Methodology

The macromolecular complex or macromolecule is firstly divided into a number of structural domains. These domains are then combined by using a small number of parameters, such as interdomain distances and angles, in order to generate a large number of possible conformations of the macromolecule. The solution properties, such as SAS profiles, FRET distances and sedimentation coefficient, are then predicted *in silico* for each conformation of the macromolecule and compared to the experimental datasets. A function that measures the discrepancy between simulated and experimental datasets is thus calculated for each set of values of the parameters defining each conformation of the macromolecule. Finally, a general systematic method that searches for the parameters producing the model whose solution properties best globally fit the available experimental datasets is used.

7.3.1 Construction of the generalised rigid-body model

A general method for generating possible conformations for a macromolecule from the structures of individual domains, applicable to a large number of problems, was devised. The conformation of the macromolecule is generated from the structural data available, experimentally derived or computationally modelled high- or low-resolution structures, for each of the N_D individual domains comprising the macromolecule. The possible conformations that the macromolecule can take are constrained by the definition of the movements of each domain. The allowed movement of each domain has two components: rotations with respect to its centre of mass (CM), and translations and rotations of its CM.

The transformation from any given Cartesian coordinate system to another can be carried out by three successive rotations in a specific sequence, defined by

the Euler angles the choice of which is, within limits, arbitrary. The convention employed here is that used in celestial and applied mechanics, and molecular and solid state physics (Goldstein, 1980). The sequence of rotations is started by rotating the initial coordinate system xyz by an angle φ counterclockwise about the z-axis (Fig. 7.1A). Secondly, a counterclockwise rotation, by an angle ϕ, about the intermediate x axis is performed (Fig. 7.1B). The transformation is finished by a counterclockwise rotation by an angle θ about the z'-axis, resulting in the new system of coordinates $x'y'z'$ (Fig. 7.1C). The three Euler angles φ, ϕ, and θ represent the three required generalised coordinates specifying the orientation of the $x'y'z'$ system relative to xyz.

The rotations of a domain i around its CM in a system of coordinates $x_i'y_i'z_i'$ are defined by using the Euler angles φ_i^r, ϕ_i^r and θ_i^r. The movement of the CM of domain i is specified in a similar manner. A new coordinate system $x_i''y_i''z_i''$ is defined by the rotations specified by a given set of Euler angles. The CM is then translated in the direction of y_i'' a distance r_i, and arbitrarily rotated in the new coordinate system by using the Euler angles ϕ_i and θ_i (rotations around x_i'' and z_i''). The assignment of the ranges of variation of φ_i^r, ϕ_i^r, θ_i^r, r_i, ϕ_i and θ_i complete the definition of the allowed movements for domain i.

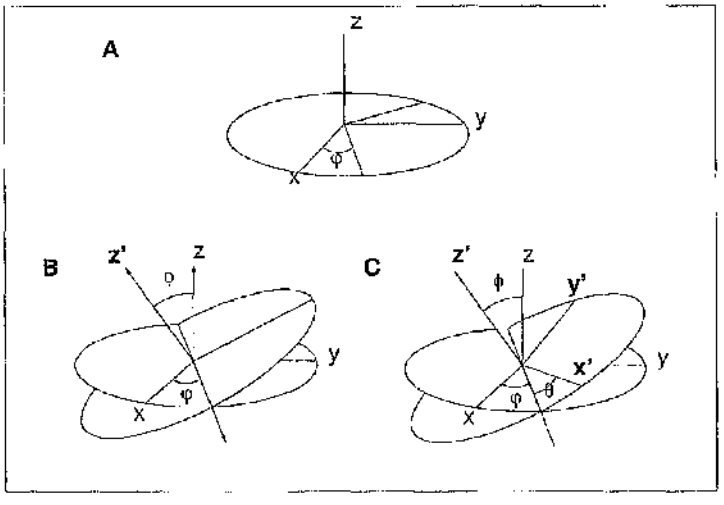

Figure 7.1: The rotations defining the Eulerian angles

This process is repeated for each domain $i = 1...N_D$. Therefore, this set of $6 \times N_D$ parameters $\vec{r} \equiv (\varphi_1^r, \phi_1^r, \theta_1^r, r_1, \phi_1, \theta_1, ..., r_{N_D}, \phi_{N_D}, \theta_{N_D})$ defines a particular configuration of the macromolecule, for which a point in *configurational space* can be assigned.

7.3.2 Computation of SAXS/SANS intensities

SAXS intensities were calculated from the atomic coordinate files of the structural model of a particular configuration by using the computer program CRYSOL (Svergun et al., 1995). CRYSOL evaluates the solution scattering of the given structure ($I(s)$) taking into account the scattering of the particle *in vacuo*, and additionally, the scattering from the excluded volume and the hydration layer around the particle. CRYSOL also fits the experimental scattering ($I_e(s_i)$) curve to the simulated one ($I(s)$) using the average displaced solvent volume and the contrast of the hydration layer as free parameters. The function χ_X provides a measure of the discrepancy between simulated and experimental scattering curves, and is defined as

$$\chi_X^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} \left[\frac{I_e(s_i) - cI(s_i)}{\delta(s_i)} \right]^2 , \tag{7.1}$$

where N_p is the number of experimental points, $\delta(s_i)$ are the experimental errors and c is a scale factor (more details in Svergun et al. (1995)). SANS intensities can be calculated in an identical manner, by employing the computer program CRYSON (Svergun et al., 1998).

7.3.3 Computation of FRET distances

In a FRET experiment, the positions of donor and acceptor molecules are usually not unique. At least one, or sometimes both, dyes show a distribution of positions with respect to their attachment point. Molecular dynamics (MD) simulations

can be employed to find these distributions (Stühmeier et al., 2000), and to estimate from them an average dye position, referred to as the *singular position*. The experimentally determined FRET efficiency is employed to calculate the apparent singular distance between donor and acceptor which, combined with the MD simulations, can be used to estimate the distance between the attachment points of both dye molecules (referred to as d). For each given configuration of the macromolecule, the distance d between attachment points is calculated and compared to the experimental value.

7.3.4 Computation of the sedimentation coefficient

The sedimentation coefficient of the macromolecule was calculated by using the computer program HYDRO (García de la Torre et al., 1994). Firstly, a low-resolution bead model of the structure of each domain comprising the macromolecule was produced by using a modified version of the AtoB algorithm (Byron, 1997), implemented in the computer program *newAtoB*. More details on the generation of bead models from high-resolution structures can be found in Section 3.3.1.

7.3.5 Estimation of steric clashes

Configurations containing steric clashes between the different domains were discarded. The maximum total volume (V_{max}) of the macromolecule was estimated as the sum of the volumes of each domain. The volume of a particular configuration (V) was calculated using the same algorithm as that used for the estimation of the volumes of each of the domains (see Section B.5). The configuration was accepted if $V \geq \gamma V_{max}$, where γ is the overlap threshold coefficient, usually set at a value of 0.95 (for more information, see Appendix B).

7.3.6 Search for the best configuration

The Monte Carlo simulated annealing method (MC/SA) (Kirkpatrick et al., 1983) has been widely used in statistical physics (Landau and Binder, 2000) to locate a global mimimum in a rugged landscape containing many local minima. The method can be summarised as follows: (1) a random point $\vec{r_0}$ in configurational space (space defined by the parameters) is chosen, and a quantity $E(\vec{r_0})$ (energy or scoring function) is calculated from the values of the parameters; (2) a random modification in one random parameter is introduced and the new energy $E(\vec{r_1})$ is calculated; (3) the change is accepted with a Boltzman probability factor $e^{-[E(\vec{r_1})-E(\vec{r_0})]/T}$, where T is the 'temperature'. Note that the change is always accepted if $E(\vec{r_1}) < E(\vec{r_0})$, but it might still be accepted even if $E_1 > E_0$, depending on the value of the temperature T; (4) the process is repeated by restarting from step 3) for a large number of steps N_{steps}. After N_{steps} have been performed, the temperature of the system is decreased by a factor α, so that $T_{new} = \alpha * T_{old}$ (with $0 < \alpha < 1$), and the whole process is restarted from (1). The first configuration at this new temperature T_{new} is taken from the best configuration found at the previous temperature T_{old}.

The run is started at an initial temperature T_0 and stopped when the system reached a predefined minimum temperature T_F, where no further decrease in energy is registered. It is worth noting that T_0, T_F and α are related by the relation $\alpha^{N_a} = T_F/T_0$, where N_a is the total number of temperature updates in a given run.

The energy E is a function of all the parameters that define a given architecture and decreases as the fit to the available experimental solution properties improves. The functions χ_X and χ_N (Eq. 7.1) provide a measure of discrepancy between simulated and experimental SAXS/SANS curves. For an experimental variable with only one value, such as the sedimentation coefficient $s_{w,20}$, the measure of discrepancy between experimental ($s^E_{w,20}$) and simulated ($s^S_{w,20}$) values can be defined as

$$\chi_s = \left| \frac{s^E_{w,20} - s^S_{w,20}}{\delta_s} \right|, \tag{7.2}$$

where δ_s is the experimental error in $s_{w,20}$. These measures of discrepancy between experimental and simulated data can be simultaneously combined in a single expression defining the total energy, as follows

$$E(\vec{r}) = \eta_X \chi_X(\vec{r}) + \eta_N \chi_N(\vec{r}) + \eta_F \chi_d + \eta_{SV} \chi_{s(w,20)}, \tag{7.3}$$

where η_X, η_N, η_F, η_{SV} are user-defined penalties for each technique (SAXS, SANS, FRET and SV, respectively), the determination of which is described below. In this way, the configuration that globally satisfies all the available experimental datasets can be found by minimising the total energy $E(\vec{r})$ as a function of the parameters $\vec{r}$.

At high temperatures, the configurational space is effectively explored in the ranges available to each variable. A large number of different configurations for the quaternary structure is investigated in this regime, even if they do not produce acceptable fits to the experimental data. At intermediate temperatures T, the system will still be able to climb energy barriers smaller than T, but will tend to be localised in regions of low energy. Only configurations with reasonable fits to the experimental data are possible, but the system is still exploring all the accessible regions of low energy. At the lowest temperatures, only moves that reduce the energy would be accepted, and so the system can only descend on the energy landscape. Provided that T_F is low enough so that no change in $E(\vec{r})$ is observed, and that the annealing is sufficiently slow, the configuration with best fit to the experimental data is found.

Ideally, each different technique should be equally important in determining the final configuration. In other words, each separate term contributing to the total energy in Eq. 7.3 should take, on average, similar numerical values when the algorithm explores the phase space at the highest temperature T_0. In practical terms, this implies that the η_i can be determined so that, for all i, $\eta_i \times max(|\chi_i|) \approx$

1, where $max()$ is a function that calculates the maximum value taken by $|\chi_i|$ at temperature T_0. The values of the penalties η_i were further normalised so that, when simultaneously employing several datasets, the fluctuations in the values of each variable (see below) had a similar temperature evolution. This thwarted the domination of a single dataset in driving the convergence to a minimum that only satisfied itself, and not the other datasets.

7.3.7 Parameter and configuration likelihood estimators

For any individual simulation i, the optimum value of any given parameter r_j is determined by its value $r_{j,i}(T_F)$ at the lowest temperature T_F. A family of simulations is a set of N_r simulations performed under the same conditions but with random starting values for the N_p parameters. From each family of simulations, the average value of each parameter r_j was calculated from the final values of each individual run as

$$< r_j^F >= \frac{1}{N_r} \sum_{i=1}^{N_r} r_{j,i}(T_F).$$

In addition, for each family of simulations, the uncertainty σ_{r_j} in the value of a parameter r_j was calculated as

$$\sigma_{r_j} = \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} (r_{j,i}(T_F) - < r_j^F >)^2}.$$

A family of simulations produces a set of N_r final configurations $\vec{r}_k^F$ $(k = 1...N_r)$. In order to measure the dispersion of these configurations around a given fixed configuration $\vec{R}$, a function F was defined as follows.

$$F(\vec{r}^F) = \frac{1}{N_r * N_p} \sum_{i=1}^{N_r} \sum_{j=1}^{N_p} (r_{i,j}^F - R_j)^2, \tag{7.4}$$

where N_p is the total number of parameters, $r_{i,j}^F$ is the final value of parameter j in run i, and R_j is the value of parameter j in configuration $\vec{R}$. The lower the F-value, the closer the N_r final configurations are to configuration $\vec{R}$. If the datasets used to run the simulations were generated from a given configuration $\vec{R}$, the success of a family of simulations in retrieving that configuration can be measured by calculating its F-value.

The mean value ($< r_{j,i}(T) >$) and standard deviation ($\sigma_{r_{j,i}}(T)$) of the set of values taken by each parameter r_j in accepted configurations during a particular run i were calculated as a function of temperature. The temperature evolution of the standard deviation of the parameters was used to analyse the transition from configurational space exploration to localisation (see below).

7.4 Results

7.4.0.1 Simulated datasets

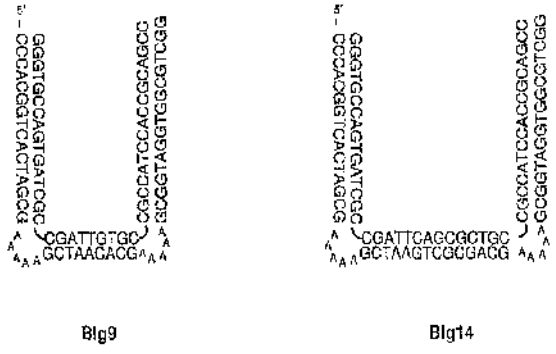

Figure 7.2: Scheme of the bulged DNA samples employed in this study.

The methodology developed in this chapter was tested on two bulged DNA structures, $H_{16}A_5H_9A_5H_{17}$ (otherwise called blg9) and $H_{16}A_5H_{14}A_5H_{17}$ (also referred to as blg14) (Fig. 7.2), where H_x refers to dsDNA with x base-pairs, A_5 is

a single stranded loop of five nucleotides, and n is the number of base-pairs in the central dsDNA fragment between the bulges (the sequence of blg9 was identical to that used by Stuhmeier et al. (2000)). The possible conformations of the bulges were modelled by combining three dsDNA fragments of appropriate sizes (16, 9 and 17 bp for blg9 and 16, 14 and 17 bp for blg14) as shown in Fig. 7.3. These fragments are hereafter referred to as domains 1, 2 and 3. The Euler angles ϕ and θ for domains 1 and 3 were employed to model all the possible conformations of the bulges, while domain 2 was kept fixed aligned with the "z_1" axis (see Fig. 7.3A). The *in silico* structures of blg9 and blg14 were generated by using the angles $(\phi_1, \theta_1, \phi_3, \theta_3) = (0^o, 0^o, 0^o, 0^o)$ and $(\phi_1, \theta_1, \phi_3, \theta_3) = (0^o, 0^o, 180^o, 0^o)$, respectively (see Fig. 7.3B and C). The use of these parameters was based on the NMR-spectroscopy and FRET data available for DNA fragments containing one (Dornberger et al., 1999; Gohlke et al., 1994) or two (Stuhmeier et al., 2000) A_5 bulges. The SV, SAXS and FRET data were simulated from the structures of blg9 and blg14 DNA produced *in silico*.

Sedimentation coefficients were simulated from the *in silico* structures by using the computer program HYDROPRO (García de la Torre et al., 2000). The results for the blg9 and blg14 structures were 3.36 and 3.32 S (datasets SV9 and SV14, respectively).

The SAXS intensity profiles were similarly predicted from the *in silico* structures of blg9 and blg14 by using the computer program CRYSOL (Svergun et al., 1995) (datasets SAXS9 and SAXS14, Fig. 7.4A). Different levels of white noise (noise whose frequency spectrum is constant) were added to the original simulated curve for blg9 (SAXS9), in order to evaluate the robustness of the method (datasets $SAXS9_{20}$ and $SAXS9_{40}$, Fig. 7.4B).

Finally, FRET distances were calculated from *in silico* structures of blg9 and blg14 by measuring the distance between atom 480 in domain 1 and atom 558 in domain 3, both placed at the ends of the DNA fragments near the axis of the double helix. The average distances measured were 48 ± 5 Å for blg9 (dataset FRET9) and 135 ± 10 Å for blg14 (dataset FRET14). In the case of blg9, the

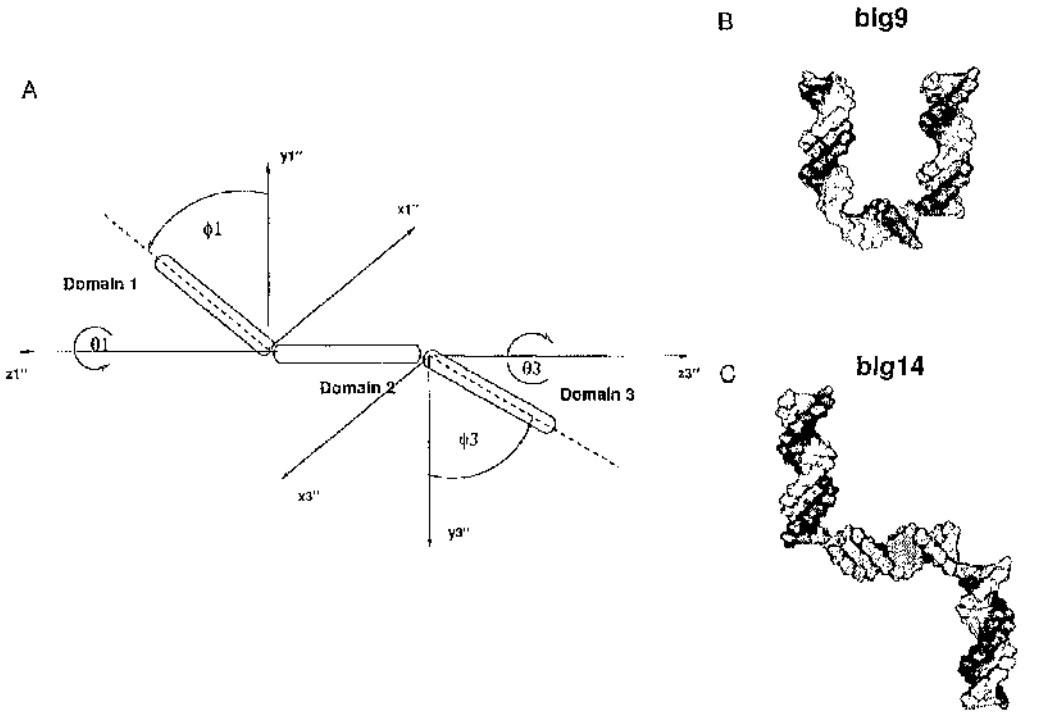

Figure 7.3: (A) Schematic representation of the parameters employed in the simulations. The helix axes are used as reference lines to define the angles between domains. (B) and (C) Surface representations of the blg9 and blg14 *in silico* generated structures.

proposed distance between the ends of domains 1 and 3 agrees with that reported by Stuhmeier et al. (2000).

7.4.1 Method validation

In order to validate the methodology proposed in this chapter, the structures of both blg9 and blg14 were reconstructed by using different combinations of the datasets. In all simulations, the ranges of variations of the angular parameters were as follows: $0^o < \phi_1 < 180^o$, $0^o < \theta_1 < 360^o$, $180^o < \phi_3 < 360^o$ and $0^o < \theta_3 < 360^o$. Different combinations of the parameters give rise, however, to effectively the same low-resolution structure (for instance $(\phi_1, \theta_1, \phi_3, \theta_3) = (0^o, 0^o, 0^o, 0^o)$ and $(0^o, 90^o, 0^o, 90^o)$). For this reason, in order to compare results from different runs, the four aforementioned parameters were reduced to only three parameters: $\psi_{12} = -sin(\phi_1)$ representing the angle between domains 1 and 2, $\psi_{23} = -sin(\phi_3)$ the angle between domains 2 and 3, and finally

$$\begin{aligned} \psi_{13} \;=\; & acos[cos(\phi_1)cos(\theta_1)cos(\phi_3)cos(\theta_3) \\ & +cos(\phi_1)sin(\theta_1)cos(\phi_3)sin(\theta_3) \\ & +sin(\phi_1)sin(\phi_3)] \end{aligned}$$

the angle between domains 1 and 3, in a plane containing their helical axes. Using this convention, blg9 is defined by the angles $(\psi_{12}^9, \psi_{23}^9, \psi_{13}^9) = (90^o, 90^o, 0^o)$, whereas blg14 is defined by the angles $(\psi_{12}^{14}, \psi_{23}^{14}, \psi_{13}^{14}) = (90^o, 90^o, 180^o)$ (see Fig. 7.3B and C).

Different families of simulations using only one of the possible datasets were performed. In addition, other families of simulations employing both SAXS and SV, SAXS and FRET, or SV, FRET and SAXS data were made. The settings utilised for the simulations are shown in Table 1. Comparison of these simulations was used to evaluate the advantages of using this multi-technique global modelling approach.

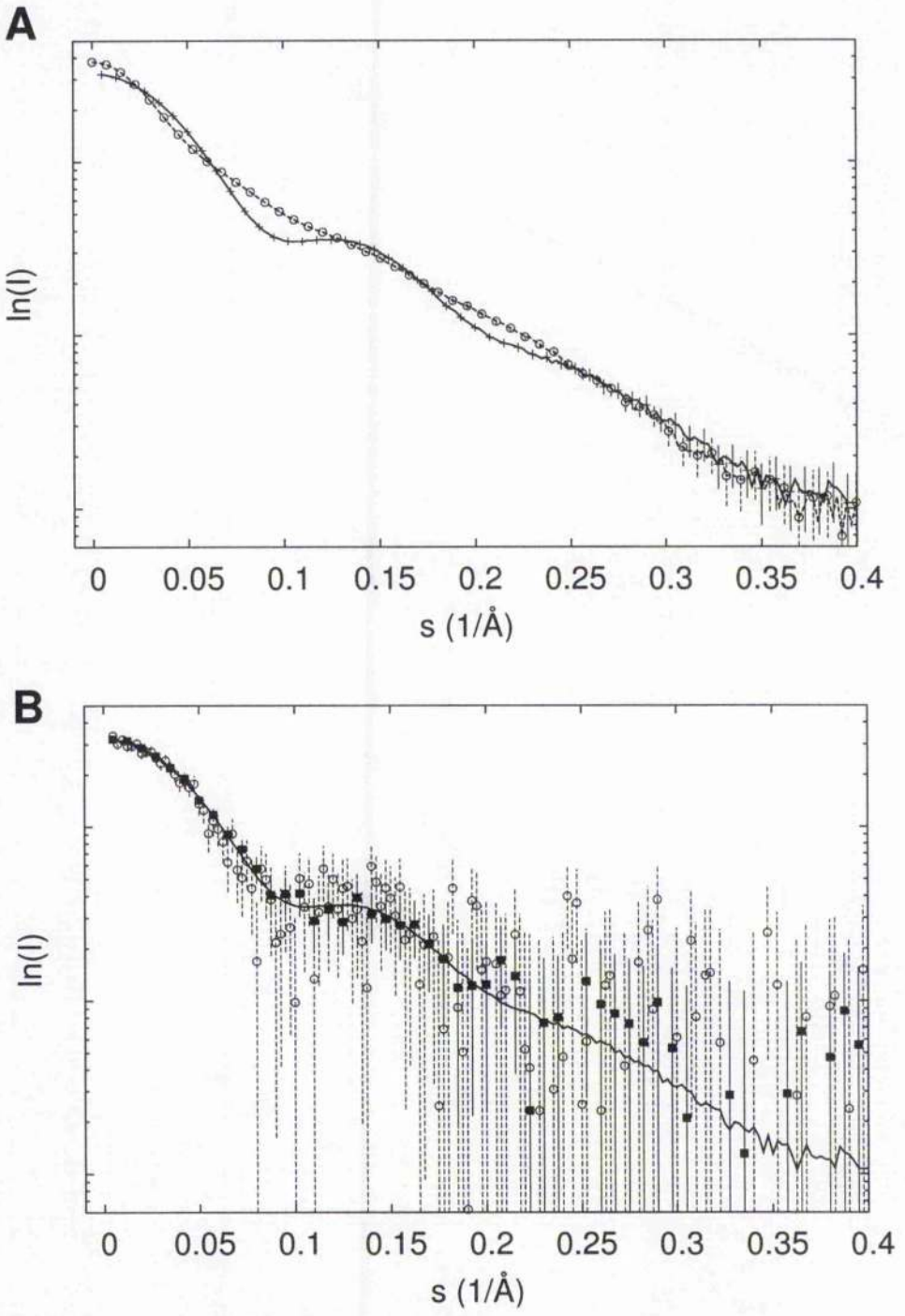

Figure 7.4: (A) Simulated SAXS data for blg9 (dataset SAXS9, crosses, solid line) and blg14(dataset SAXS14, open circles, dashed line). (B) Simulated SAXS data for blg9 with 0 (dataset SAXS9, solid line), 2000% (dataset $SAXS9_{20}$, filled boxes) and 4000% (dataset $SAXS9_{40}$, open circles) added noise. Solid/dashed vertical lines represent error bars.

Sample	Identifier	Datasets used	N_r	T_0	T_F	η_{SV}	η_F	η_X
blg9	S9-H	SV9	12	2.0	0.004	1	0	0
blg9	S9-F	FRET9	10	2.0	0.004	0	1	0
blg9	S9-X	SAXS9	10	2.0	0.004	0	0	1
blg9	S9-X-20	$SAXS_{20}$	10	2.0	0.004	0	0	1
blg9	S9-X-40	$SAXS_{40}$	10	2.0	0.004	0	0	1
blg9	S9-HX-0	SV9, SAXS9	10	6.0	0.0001	1	0	6
blg9	S9-HXF-0	SV9, SAXS9,FRET	10	6.0	0.01	1	0.1	10
blg9	S9-HXF-20	SV9, $SAXS_{20}$,FRET9	10	2.0	0.004	1	1	6
blg14	S14-H	SV14	10	2.0	0.004	1	0	0
blg14	S14-F	FRET14	10	2.0	0.004	0	1	0
blg14	S14-X	SAXS14	10	2.0	0.004	0	0	1
blg14	S14-HXF	SV14, SAXS14, FRET14	10	6.0	0.004	1	0.1	10

Table 1. Settings employed for the different simulation families.

In all simulations, the annealing was performed by using at least 10 temperature update cycles ($N_a = 10$). The corresponding temperature update factor α was calculated from T_0, T_F and N_a as described before (Section 7.3.6). At each temperature, $N_{steps} = 1500$ configurations were evaluated.

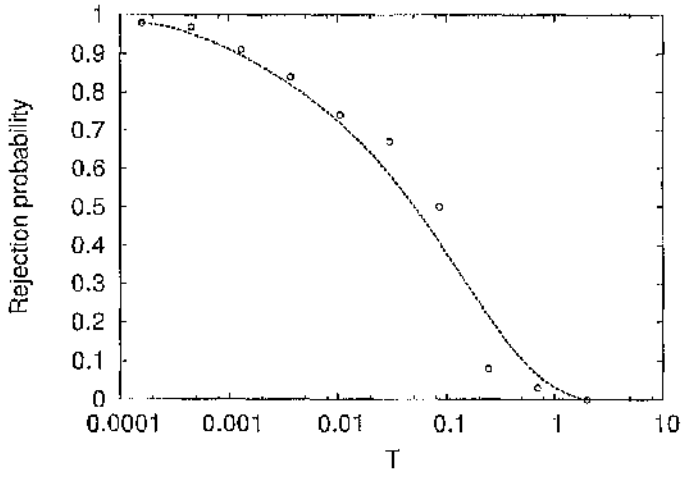

Figure 7.5: Rejection probability versus temperature T for one run in the family of simulations S9-H.

The rejection probability factor was defined as the number of rejected configurations over the total number of configurations evaluated at each temperature. As the annealing process evolves, and the temperature decreases, the rejection probability increased from 0, at the highest temperature, to 1 at the lowest (Fig. 7.5). This evolution reflects the fact that, as the temperature decreases, the chance of a configuration with high energy being accepted decreases.

The evolution of the values taken by the parameters ϕ_1, θ_1, ϕ_3 and θ_3 also exhibited this trend. A plot of the values taken by ϕ_1 *versus* those of θ_1 at high temperatures ($T = 2.0$, Fig. 7.6A) shows a uniform distribution. At this temperature, the system explored the energy landscape, and thus the values taken by the parameters randomly span their whole variation ranges. A similar plot at a lower temperature ($T = 0.17$, Fig. 7.6B), shows that the parameter values are now somewhat confined, reflecting the localisation of the system in configurational space. Finally, at the lowest temperatures ($T = 0.01$), the system is stranded in the global minimum, and the parameter values are completely localised (Fig. 7.6C).

Parameter localisation was also observed by looking at the histograms of the values taken by each parameter *in accepted configurations* as the temperature decreases. At high temperatures ($T = 2.0$, Fig. 7.7A-D, dotted lines), the accepted configurations have parameters randomly distributed over their variation ranges. At lower temperatures ($T = 0.17$, Fig. 7.7A-D, dashed lines), the values taken by the parameters in accepted configurations have localised distributions. These distributions become even narrower for lower temperatures ($T = 0.01$, Fig. 7.7A-D, solid lines), for which a total localisation of the parameters at the configuration with minimum energy is apparent.

Meanwhile, the values of the variables, such as χ_X, d and $s_{w,20}$ also evolve with temperature. At high temperatures, the variables explore a wide range of values, limited only by the allowed configurations. For instance, at $T = 2.0$, the sedimentation coefficient $s_{w,20}$ varies between 2.9 S and 3.35 S, and thus takes all accessible values given the ranges of variation of the different parameters defining a configuration (Fig. 7.8B). The fluctuations in the variables, and thus their

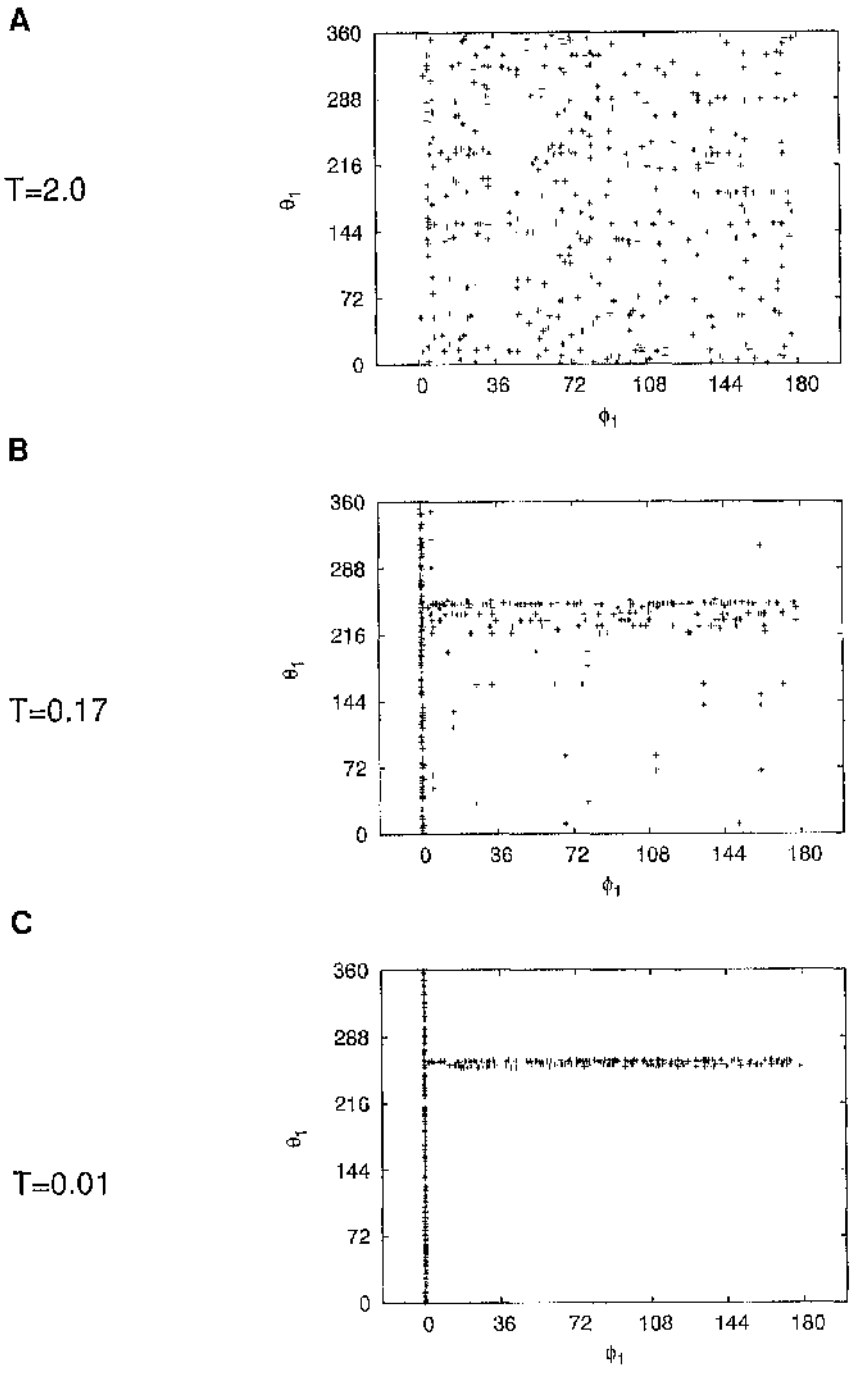

Figure 7.6: Demonstration of parameter localisation. Plots of the values taken by ϕ_1 *versus* those of θ_1 at temperatures (A) 2.0, (B) 0.17, and (C) 0.01, for a single run in the S9-X family of simulations.

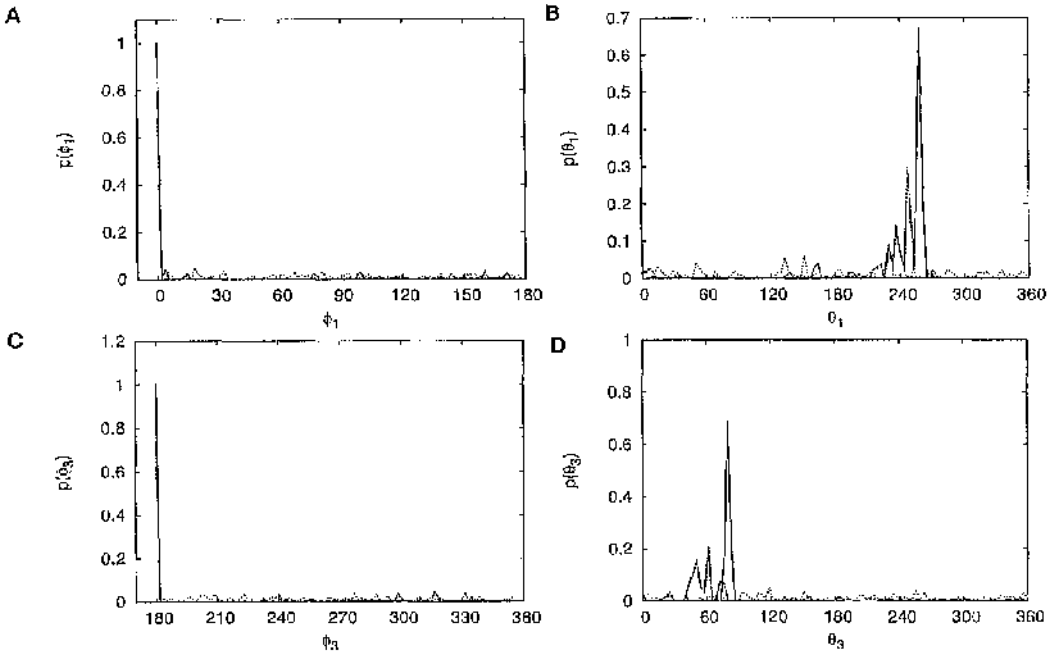

Figure 7.7: Histograms of the values taken by the parameters (A) ϕ_1, (B) θ_1, (C) ϕ_3 and (D) θ_3 at temperatures $T = 2.0$ (dotted lines), $T = 0.17$ (dashed lines) and $T = 0.01$ (solid lines), for a single run in the S9-X family of simulations.

standard deviations, diminishes with decreasing temperatures, and takes their final values once the configuration having minimum energy has been found. This behaviour in the temperature evolution is observed for all variables (Fig. 7.8A-C).

Identifier	ψ_{12}^F	ψ_{23}^F	ψ_{13}^F	F
in silico	90^o	90^o	0^o	-
S9-H	91.4 ± 1.8^o	88.8 ± 1.8^o	7.0 ± 7^o	5.8
S9-F	90.2 ± 0.3^o	89.7 ± 0.45^o	4.3 ± 3.2^o	2.7
S9-X	90.2 ± 0.14^o	89.7 ± 0.23^o	2.5 ± 1.9^o	1.85
S9-HX	90.1 ± 0.1^o	89.8 ± 0.15^o	1.4 ± 0.8^o	0.93
S9-HXF	90.09 ± 0.09^o	89.8 ± 0.15^o	0.89 ± 0.5^o	0.63

Table 2. Final parameter statistics for S9 runs using the blg9 structure.

The effectiveness of each individual technique to retrieve the original parameters giving rise to the two bulged DNA structures, blg9 and blg14, was first evaluated. The F-value (defined in section 7.3.7, the $\vec{R}$ configuration being here the configuration defined by the parameters generating the *in silico* structures of blg9 and blg14) provides a measure of the effectiveness of a family of simulations in retrieving the original configuration. In all cases, the structure of blg9 was properly reconstructed (Table 2). Simulations employing only the SV9 dataset was sufficient to obtain structures that resembled, at low-resolution, the structure of blg9 generated *in silico* (Fig. 7.9A-C, dotted lines, and S9-H in Table 2). This occurred also when only the FRET9 or the SAXS9 datasets were employed separately. In the case of the FRET9 dataset, the final parameters also agreed with the parameters of blg9, but with lower F-value and uncertainties (7.9A-C, solid lines, and S9-F in Table 2). Finally, the final parameters obtained by using the SAXS9 dataset had the lowest F-value and uncertainties (7.9A-C, dashed lines, and S9-X in Table 2).

The combination of two techniques not only reduced the uncertainties of the final parameters but also decreased the F-values (Fig. 7.9D-F, dashed and dotted lines, and S9-HX in Table 2). This trend was even more pronounced when three datasets, SV9, SAXS9 and FRET9, were employed in the reconstruction process (S9-HXF in Table 2, and Figs. 7.9D-F, solid lines, and 7.12A). When combining

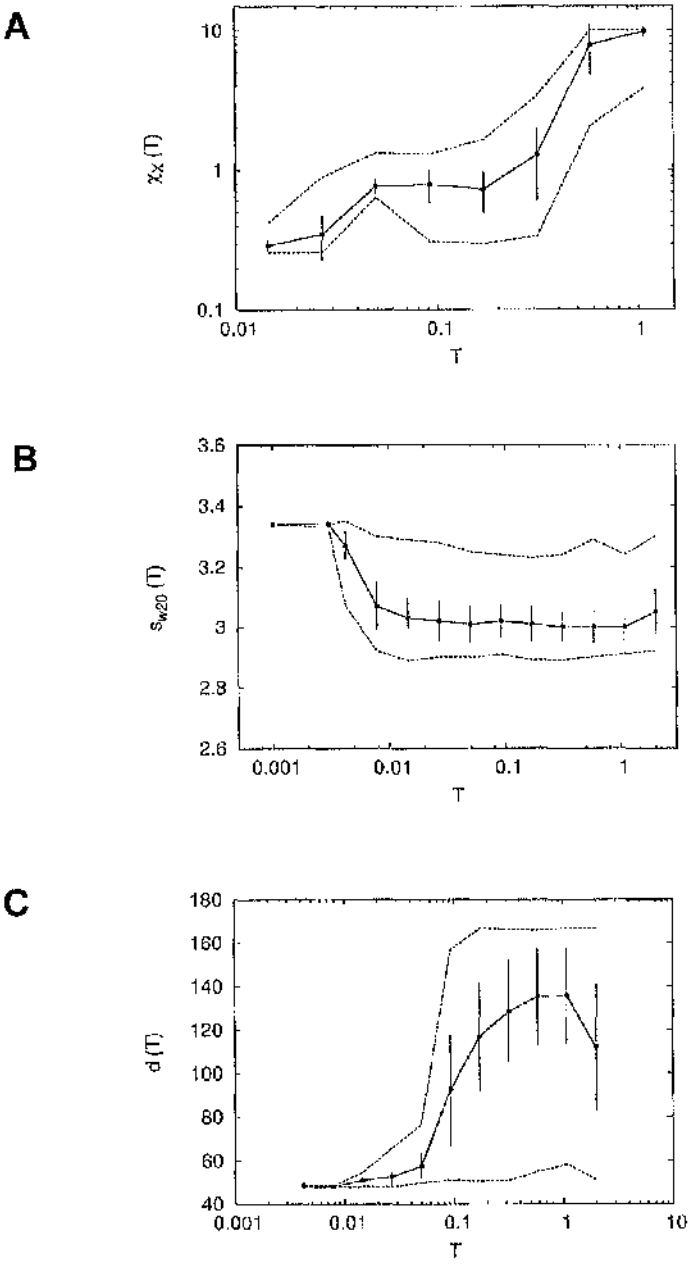

Figure 7.8: Values taken by (A) χ_X, (B) $s_{w,20}$, and (C) d as a function of the annealing temperature for a single run in the (A) S9-X, (B) S9-H, and (C) S9-F family of simulations. Vertical bars represent the standard deviation of the values taken by each variable at a fixed temperature, whereas dashed lines indicate their limits of variation.

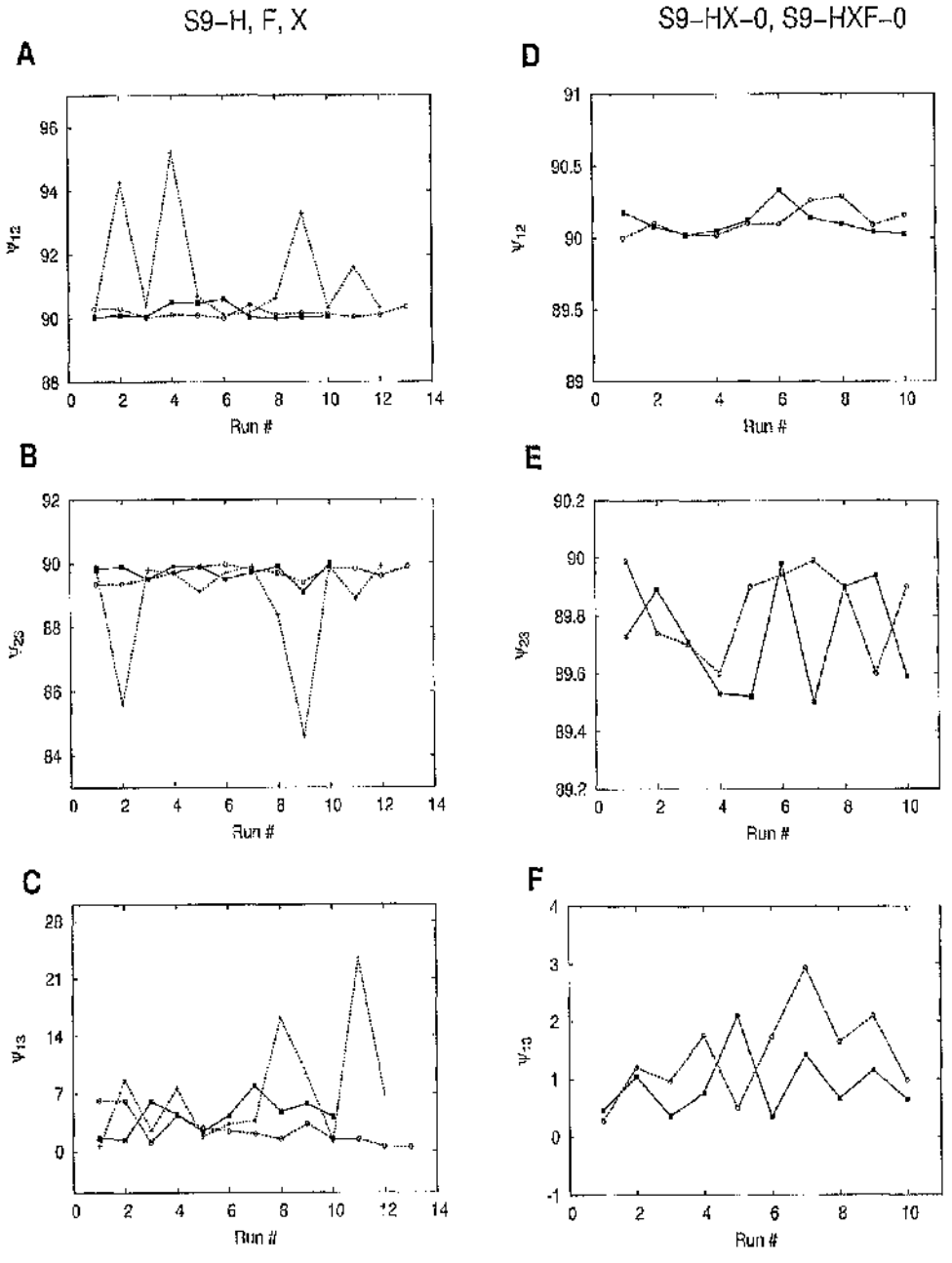

Figure 7.9: Final values of the parameters in different runs of simulations using blg9. (A)-(C) Angles ψ_{12}, ψ_{23} and ψ_{13} for S9-II (dotted line, crosses), S9-F (solid line, filled boxes) and S9-X (dashed line, open circles). (D)-(F) Angles ψ_{12}, ψ_{23} and ψ_{13} for S9-HX-0 (dotted line, crosses), S9-HXF-0 (solid line, filled boxes).

techniques, the configuration with minimum energy was also found in a smaller number of annealing cycles (data not shown).

The robustness of the reconstruction method in retrieving the original parameters was evaluated by applying different levels of noise to the simulated scattering data. Two additional SAXS simulated datasets, with noise added at 2000 and 4000% of the simulated experimental uncertainty, were produced ($SAXS9_{20}$ and $SAXS9_{40}$, Fig. 7.4B). Families of runs, using these SAXS datasets, either alone (S9-X-20 and S9-X-40) or in combination with the SV9 and FRET9 datasets (S9-HX-20 and S9-HXF-20), were performed. In both S9-X-20 and S9-X-40, the original parameters were retrieved in all the performed runs (Table 3, Figs. 7.10). As expected, not only the difference between retrieved and original parameters (resulting in higher F-values) but also the parameter uncertainties augmented with increasing noise levels. As seen before, the inclusion of more datasets improved the reconstruction process. When the $SAXS9_{20}$, SV9 and FRET9 datasets were combined, the F-value decreased and the uncertainties in the parameters were reduced, with respect to using only the $SAXS_{20}$ dataset (S9-HXF-20, Table 3, Fig. 7.10, solid line). Even with large noise levels, the reconstruction process was successful in retrieving the *in silico* structures of blg9 (see superpositions of *in silico* and reconstructed structures in Fig. 7.12B-C). This demonstrated the robustness of the method with respect to the introduction of high levels of noise in the datasets.

Identifier	ψ_{12}^F	ψ_{23}^F	ψ_{13}^F	F
in silico	90^o	90^o	0^o	
S9-X	90.2 ± 0.14^o	89.7 ± 0.23^o	2.5 ± 1.9^o	1.85
S9-X-20	90.9 ± 0.8^o	89.4 ± 0.6^o	5.5 ± 3.9^o	3.76
S9-X-40	91.2 ± 1.3^o	89.1 ± 1.5^o	7.9 ± 4.2^o	5.3
S9-HXF-20	90.3 ± 0.2^o	89.9 ± 0.14^o	2.6 ± 1.6^o	1.74

Table 3. Final parameter statistics for simulations using blg9 with different noise levels.

Several families of simulations were performed by using different combinations of the blg14 simulated and experimental datasets. The reconstructions performed using only the FRET14 dataset could not retrieve the original parameters (S14-F, Table 4, Fig. 7.11, dashed lines). A similar failure was observed when the SV14 dataset was used alone (S14-H, Table 4, Fig. 7.11, thin dotted lines). These failures are reflected in large F-values. Without any other constraint, there were far too many conformations for blg14 having an end-to-end distance of 135 Å or a sedimentation coefficient of 3.32 S, resulting in different runs producing very different final values for the parameters. This represented a typical case where the energy space defined by the datasets is degenerate (contains several minima), resulting in many conformations fitting the datasets equally well.

On the contrary, when using only the SAXS14 dataset, the retrieved parameters converged towards the original ones. This is reflected by a dramatic reduction in the F-value (S14-X, Table 4, Fig. 7.11). When simultaneously using the SAXS14, FRET14 and SV14 datasets, there was only a marginal improvement in the F-value and in the parameter uncertainties (S14-HXF, Table 4, Figs. 7.11). This, again, manifested the existence of many conformations of blg14 that equally fit the FRET14 and SV14 datasets. Apart from a systematic deviation in the final value of ψ_{13} with respect to that in the *in silico* structure, the reconstruction process of blg14 was successful (Fig. 7.12D).

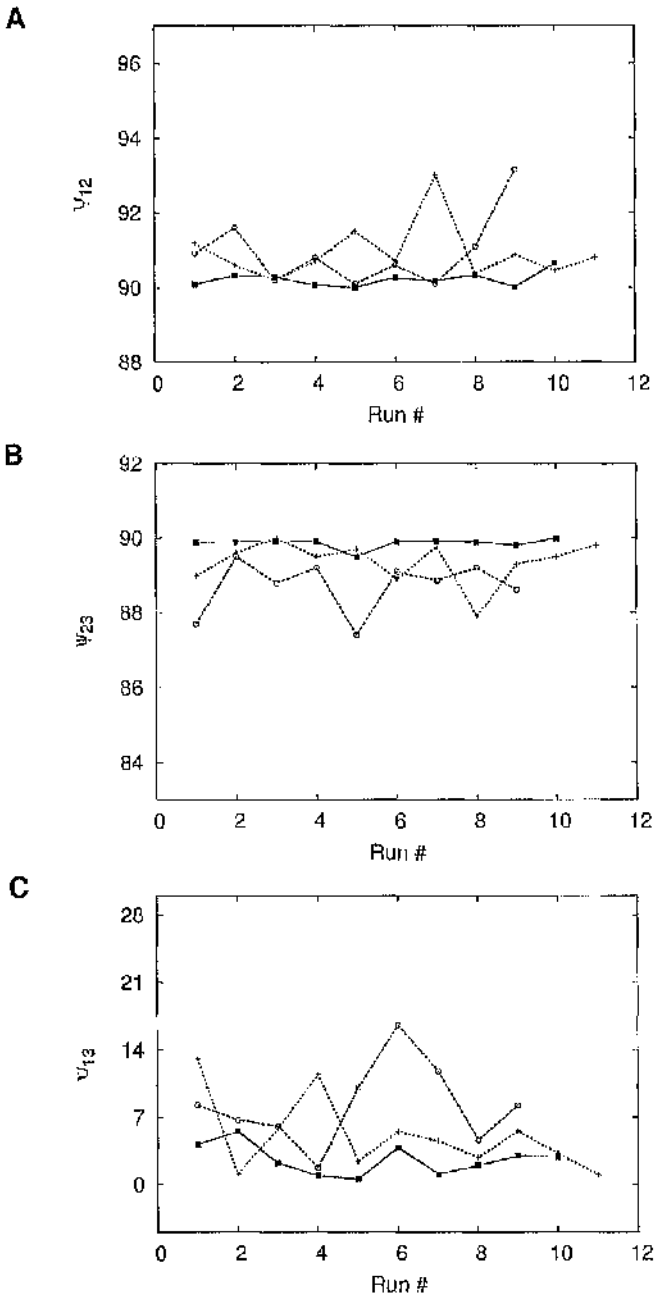

Figure 7.10: Final results for simulations using blg9 with different noise levels. (A)-(C) Angles ψ_{12}, ψ_{23} and ψ_{13} for S9-X-20 (dotted line,crosses), S9-X-40 (dashed line, open circles), and S9-HXF-20 (solid line, filled boxes).

Identifier	ψ_{12}^F	ψ_{23}^F	ψ_{13}^F	F
in silico	90^o	90^o	180^o	
S14-H	70 ± 40^o	96 ± 6^o	81 ± 7^o	66
S14-F	123 ± 29^o	56 ± 23^o	85 ± 9^o	65
S14-X	91.7 ± 0.6^o	88.8 ± 0.7^o	158.9 ± 1.3^o	12.4
S14-HXF	91.6 ± 0.7^o	88.5 ± 1.3^o	159.7 ± 2.7^o	11

Table 4. Final parameter statistics for simulations using blg14 .

7.5 Discussion

In this chapter, a general methodology that can be used to reconstruct the low-resolution solution structure of a macromolecular complex from several sources of experimental data has been presented. The macromolecule is firstly divided into domains for which structural data is available. The domains are combined by using a small number of parameters to produce a conformation of the macromolecule. The algorithm reconstructs the low-resolution shape of the macromolecule by finding the relative positioning of each domain so that a number of solution properties (SAXS/SANS profiles, SV and FRET data) are simultaneously satisfied. The assembly of the structure of the macromolecule in terms of its domains has been implemented in a generalised manner, so that the methodology can be applied to a large variety of problems. The MC/SA algorithm employed to search for the best conformation is easily scalable to problems with large numbers of domains. The procedure was validated against two DNA bulged samples with very different overall conformations. The conformations of the two samples were generated *in silico* from previous data (Dornberger et al., 1999; Gohlke et al., 1994; Stuhmeier et al., 2000) by using two sets of values for the parameters. SAXS, SV, and FRET data were simulated from these *in silico* structures. The methodology was tested by using different combinations of the SAXS, FRET and SV datasets to restore the original *in silico* structures of the macromolecules.

In the majority of cases, each individual dataset, when used separately, was able to restore the original parameters. The combination of more than one dataset

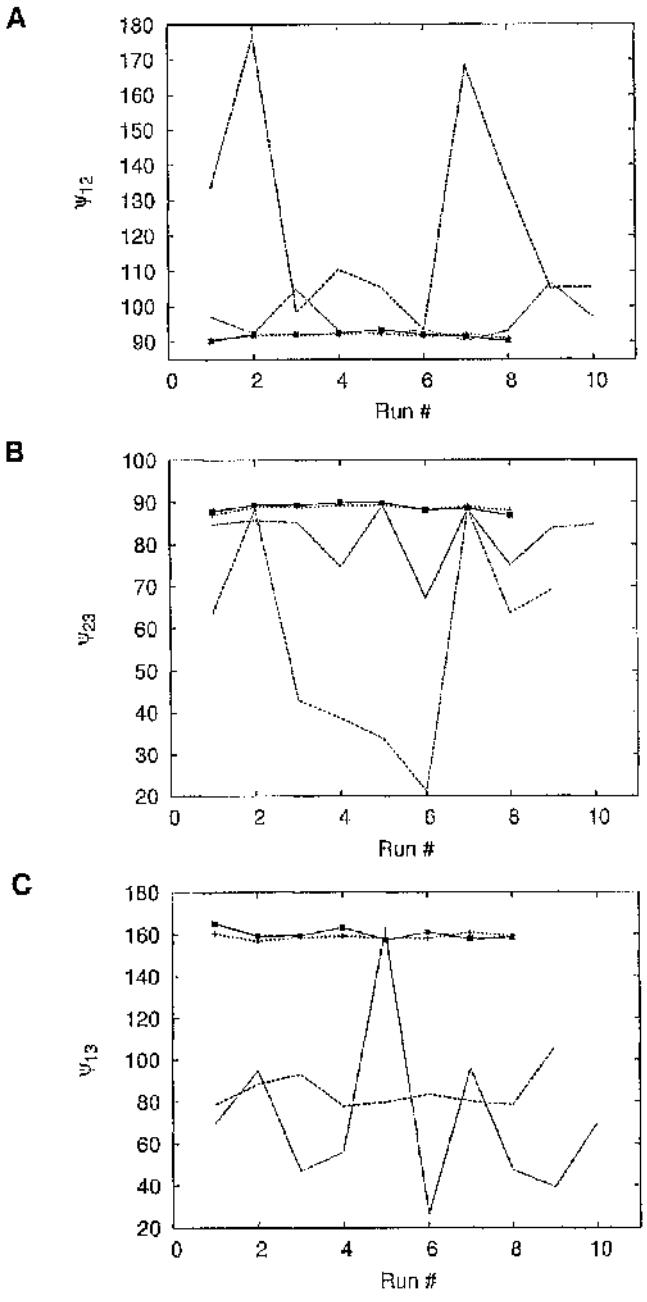

Figure 7.11: Final results for simulations using blg14. (A)-(C) Angles ψ_{12}, ψ_{23} and ψ_{13} for S14-F (dotted line), S14-H (thin dotted line), S14-X (solid line, filled boxes), and S14-HXF (dashed line, crosses).

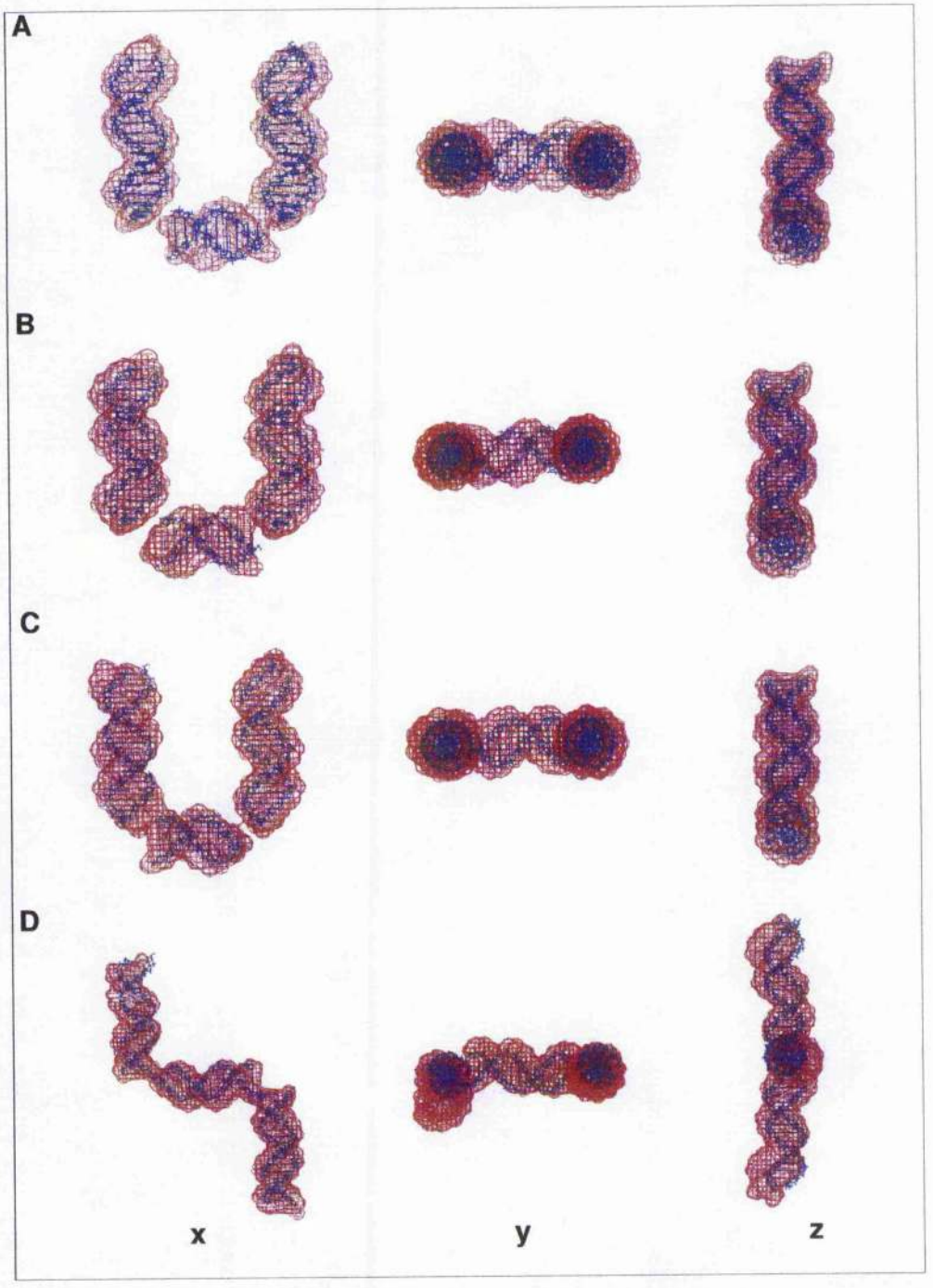

Figure 7.12: Final reconstructions of blg9 and blg14. Each panel contains the high-resolution *in silico* model of blg9 (panels A-C) or blg14 (panel D) as blue sticks and a reconstruction with the final average parameters from simulations (A) S9-HFX, (B) S9-HXF-20, (C) S9-X-40 and (D) S14-HXF (in red).

was shown to produce better restorations, in that the final restored parameters were more similar to those used for the *in silico* simulation of the original structures of the bulges (lower F-values). The uncertainties in the values of the parameters in the final reconstructions were shown to diminish as more datasets were used for the restoration process. The introduction of noise in the original datasets produced similar restored parameters, but with higher F-values and uncertainties. All in all, the method was able to restore the original *in silico* structures even when a considerable amount of noise was introduced into the datasets.

In most, but not all cases, the use of multiple datasets considerably improved the restored parameters. The ability to simultaneously fit multiple datasets generally resulted in a reduction of the parameter uncertainties and a decrease in the F-values (improved restored parameters). In any case, the implemented methodology provides a single integrated framework for finding the solution conformations that would potentially fit any individual dataset. In cases where the available datasets, when fitted individually, predict several possible models for the conformation of the macromolecule, the global modelling approach implemented in this algorithm allows one to find the model that simultaneously satisfies all the datasets. The fact that a single model fits, at the same time, a range of datasets increases its likelihood to represent the real conformation of the macromolecule in solution. In some circumstances, the values for the parameters predicted by different datasets might be contradictory. In such cases, it is very important to be able to fit the datasets individually and to manually compare the results provided by each dataset.

SAXS *ab initio* restoration methods are uniquely suited to producing low-resolution reconstructions of macromolecules whose structure is unknown. For simple shapes, different *ab initio* restorations usually differ only in minor details. In these cases, the reconstructed models can be superimposed and averaged in order to find the consensus reconstruction (Heller et al., 2003; Kozin and Svergun, 2001). In some cases, however, the restoration process produces very different reconstructions that fit the SAS data equally well (Heller et al., 2003; Rosenzweig et al., 1993; Volkov and Svergun, 2003). Some of these models would in fact be false positives. Volkov and Svergun (2003) recently showed that *ab initio*

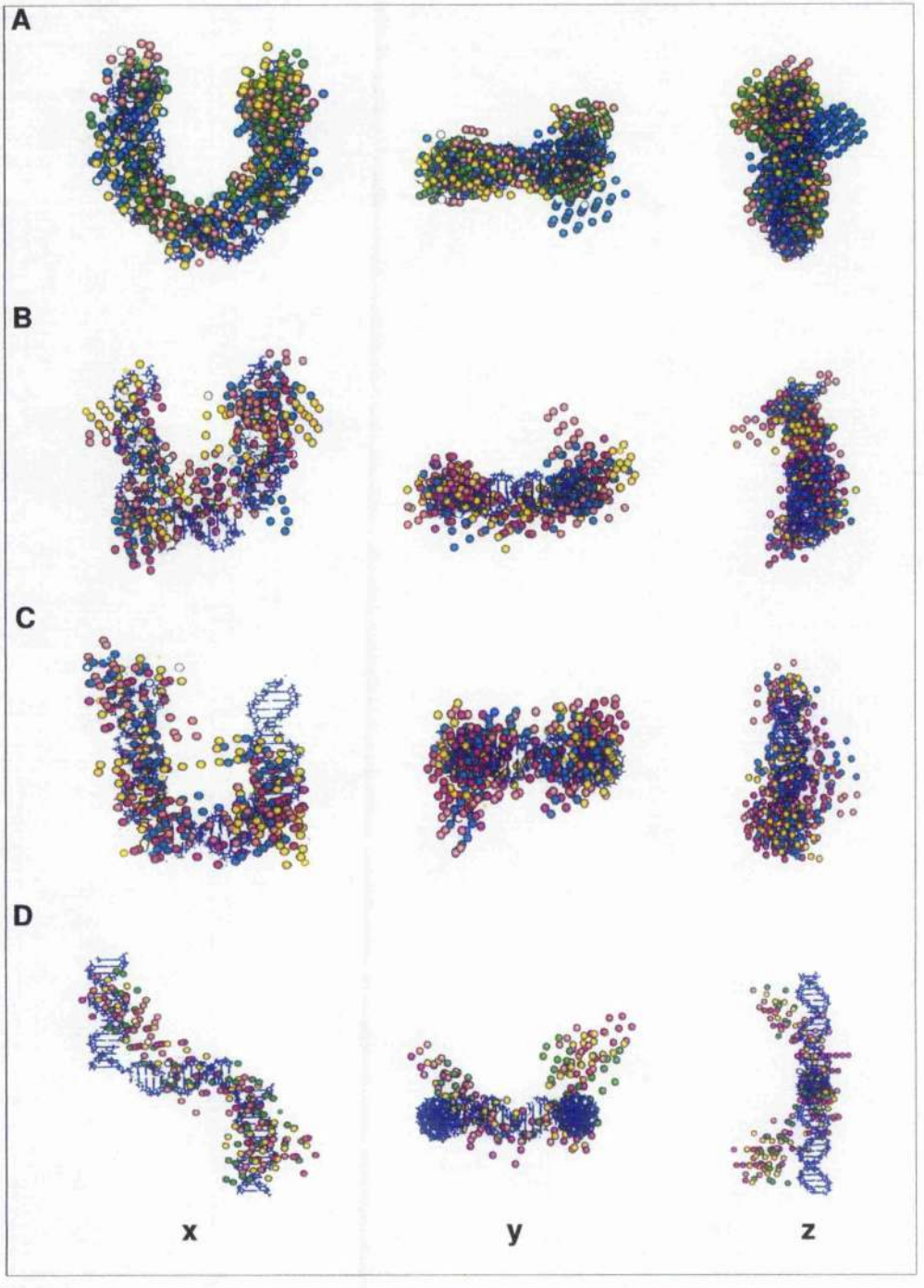

Figure 7.13: Three views of various superimposed DAMMIN reconstructions of blg9 and blg14. Each panel contains the high-resolution *in silico* model of blg9 (panels A-C, blue lines) or blg14 (panel D, blue lines) and four DAMMIN reconstructions (represented as yellow, cyan, orange and green beads) from the datasets (A) SAXS9, (B) $SAXS9_{20}$, (C) $SAXS9_{40}$ and (D) SAXS14.

reconstruction methods are able to reconstruct shapes with small anisometries and, sometimes, with voids. Shapes with larger anisometries or smaller voids often cannot be reconstructed at all, even if the reconstruction process is stable (i.e. always reproduces a similar shape). For this situation, there is no method currently available that would systematically sort the reconstructed models into families of similar models, or that would provide a reliable measure of the probability of certain models/families of models being false positive reconstructions.

DAMMIN (Svergun, 1999), arguably the most recognised *ab initio* reconstruction method, was applied to the SAXS9, $SAXS9_{20}$, $SAXS9_{40}$ and SAXS14 datasets in order to test its ability to reconstruct the original *in silico* models. DAMMIN was able to reconstruct the blg9 model when using the noiseless SAXS9 dataset (Fig. 7.13A). However, the reconstructed models differed greatly from the structure of blg9 when the noisy datasets were used (Fig. 7.13B-C). Similarly, DAMMIN was unable to reconstruct the structure of blg14 (Fig. 7.13D). These simulations demonstrate the inability of this *ab initio* reconstruction method to deal with relatively complex macromolecular conformations or with data with considerable noise levels. Based on the study by Volkov and Svergun (2003), we expect similar findings with other ab initio approaches. The methodology proposed in this chapter is restricted to cases where previous structural data for the subunits comprising the macromolecule are available or can be modelled on the basis of available experimental data. However, it makes full use of these structural constraints and of several sources of solution properties to reduce the number of false positives in the reconstruction process, and increase its reliability. The method shown here is more suitable than *ab initio* reconstruction approaches for reconstructing macromolecular shapes of complex nature.

The generalised rigid-body algorithm presented here is unsuitable for to the reconstruction of macromolecules whose crystal structures differ significantly from their solution conformations. In such cases, *ab initio* shape determination procedures are expected to produce more reliable results.

Ab initio reconstruction methods are able only to reconstruct particles with homogeneous electron density (atomic density for SANS). Accordingly, they can-

not be used for the reconstruction of macromolecular complexes whose parts are inhomogeneous, such as protein-DNA or protein-polysaccharide complexes. A simplified version of the methodology presented here has been recently shown to produce excellent results when applied to the reconstruction of the solution conformation of a protein-DNA complex (Nöllmann et al., 2004a). The ability to reconstruct inhomogenous macromolecular complexes represents another advantage of this approach with respect to *ab initio* retrieval methods.

Appendix A

X-ray scattering by matter

In this section, the principles of x-ray scattering and the equations for small angle x-ray scattering (SAXS) shall be reviewed. A more thorough introduction can be found both in Kratky (1982) and Guinier (1963).

X-rays are transverse electromagnetic waves with wavelengths typically ranging from 0.5 to 2.5 Å, between the ultraviolet (2000 Å) and the gamma (0.3 Å) regions of the spectrum. In the de Broglie interpretation, electromagnetic radiation can behave both as waves or particles (photons); however, as SAXS is concerned with interference processes, only the wave nature of electromagnetic radiation will be manifested.

The geometrical optical properties of x-rays are rather different from those of visible light. The index of refraction of most substances shows variations smaller than 10^{-4} in the x-ray region, and the x-rays are almost not refracted. This essentially implies that x-rays cannot be focused by means of typical lenses, but only by using other methods (e.g. mechanically bent Si monocrystals that actually act as x-ray lenses). However, x-rays can be reflected at the total reflection angle at grazing incidence on a solid medium, provided that its index of refraction is smaller than unity.

The *intensity* of a light beam is defined as the flux of energy that crosses a unit surface area normal to the direction of propagation per unit time. In the

case of a point source, the intensity in a given direction is defined as the flux of energy emitted by the source in that direction per unit solid angle per unit time.

A number of processes are responsible for the decrease in beam intensity that occurs when x-rays travel through matter. The x-rays can be either scattered (ie dispersed) or absorbed. In addition, the scattering can be elastic or inelastic. In the first case, the photon is scattered by the electronic cloud surrounding the atom, and changes its direction, but not its energy. Inelastic scattering can only be produced through x-rays interacting with matter by Compton processes. In this case, the photon interacts with an electron, changing its energy and momentum. As a result, the photon changes its own energy and direction of propagation, and thus the process is incoherent and inelastic. It is worth mentioning that photons scattered by Compton processes cannot interfere (they have different wavelengths and phases), and so their contribution to the final intensity is simply additive.

Last but not least, the photon can be absorbed by the sample. In this case, an atom in the sample absorbs the photon and gets excited (*photoelectric effect*). Upon de-excitation the atom emits either an electron (*Auger effect*) or an x-ray with lower energy than the absorbed one (i.e. x-ray *fluorescence*).

A.0.1 Interference and scattering profile

Since SAXS is concerned with waves elastically scattered at very small angles with respect to the incident radiation, only elastic coherent dispersion will be considered (i.e. Compton processes will not be taken into account). This implies that the final scattered amplitude can be calculated as a sum over all the photons scattered by different parts of the sample (Guinier, 1963). The intensity is then the complex module of the scattering amplitude.

If $A_0\, e^{i\omega t}$ is the amplitude of the incident electromagnetic field at the origin of coordinates (**O**) (Fig. A.0.1), the amplitude of the scattered wave at a distance r at time t can be expressed as

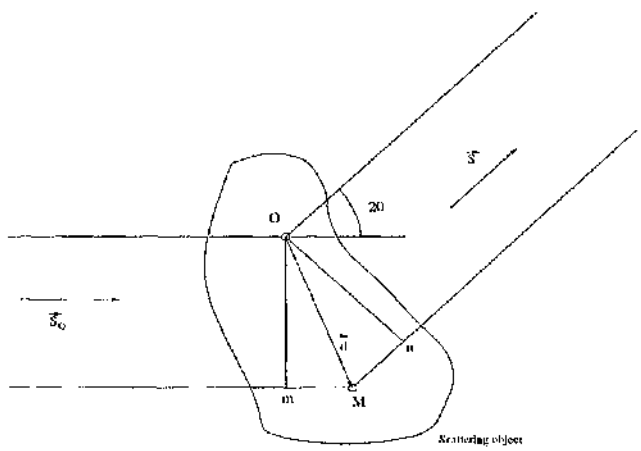

Figure A.1: Scheme of a scattering process by two point centres.

$$F(\vec{r},t) = f A_0\, e^{i(\omega t - \vec{k}.\vec{r}) - \psi},$$

where ω is the frequency of the incident wave, $\vec{k}$ the Poynting vector of the scattered light, ψ the scattering phase shift, and f the scattering factor, which generally is a function of the angle between the incident and the scattered waves.

The interference term between two waves that were scattered by the sample at **O** and **M** will be now considered. The directions of the incident and scattered waves will be $\vec{S}_0$ and $\vec{S}$ (Fig. A.0.1), respectively (Guinier, 1963). If we assume that the phase shift ψ is the same for both scatterers[1], the interference term will be proportional to the product of the scattering amplitude coming from **O** and the complex conjugate of the amplitude of the wave scattered at **M**, i.e.

[1]This assumption is valid even if the scattering atoms are different, and breaks down only in exceptional cases.

$$interference \propto F_1 F_2^* \propto e^{i(\omega t_1 - \omega t_2) - i(\vec{k_1}\vec{r_1} - \vec{k_2}\vec{r_2})}. \quad (A.1)$$

The incident wave needs $t_2 - t_1$ more time to get to **M** than to get to **O**. Taking into account that $\omega = kc$ (dispersion relation), this term can be rewritten as $t_2 - t_1 = k(mM)$, where mM is the distance between **m** and **M** (Fig. A.0.1). If we define the vector joining **O** and **M** as $\vec{d}$, the first phase term in Eq. A.1 can be expressed as

$$\omega(t_1 - t_2) = -k\vec{d}.\vec{S_0}.$$

Meanwhile, the second phase difference arises from the increased distance (i.e. Mn) that the wave scattered at **M** has to travel with respect to the one scattered at **O**. Considering that $\vec{k_1} = \vec{k_2} = k\vec{S}$ and $\vec{r_1} - \vec{r_2} = \vec{d}$, the spatial phase difference can be rewritten as

$$(\vec{k_1}\vec{r_1} - \vec{k_2}\vec{r_2}) = -k\,\vec{S}.\vec{d}$$

Finally, the total phase difference can be written as

$$\phi = -2\pi\vec{d}.\frac{(\vec{S} - \vec{S_0})}{\lambda}$$

where $\lambda = \frac{2\pi}{|\vec{k}|}$ is the wavelength of the incident radiation. If we define $\vec{s} = 2\pi(\vec{S} - \vec{S_0})/\lambda$, then the interference term can be expressed as

$$fF_0 e^{i\phi} = fA_0 e^{i\vec{s}.\vec{d}}.$$

The vector $\vec{s}$ plays a very important role in scattering theory. If we define the angle between the incident and the scattered radiation as 2θ, and by taking into account that $cos(2\theta) = cos^2(\theta) - sin^2(\theta)$, the modulus of $\vec{s}$ can be shown to be

$$s = \frac{4\pi}{\lambda} \sin\theta$$

By using the same reasoning, it is possible to deduce the interference term from N scatterers

$$F_N(\vec{s}) = F_0 \sum_{i=1}^{N} f_i e^{i\,\vec{s}.\vec{r_i}}$$

It will now be convenient to consider the continuous limit, because of the enormous number of scattering electrons present in a sample and the fact that electrons are not localized in space (Kratky, 1982). If we define $\rho(\vec{r})$ as the number of electrons per unit volume, the volume element dV at position $\vec{r}$ will now contain $\rho(\vec{r})\,dV$ electrons, and the previous equation can be written as

$$F(\vec{s}) = \int \rho(\vec{r})\, e^{i\,\vec{s}.\vec{r}}\, dV \tag{A.2}$$

where the volume integral runs over all the particle volume. Now, the intensity is the complex square of the scattering amplitude and can be thus be written as (Kratky, 1982)

$$I(\vec{s}) = F.F^{*} = \int\int \rho(\vec{r_1})\,\rho(\vec{r_2})\, e^{i\,\vec{s}.(\vec{r_1}-\vec{r_2})}\, dV_1\, dV_2$$

Expressing the electron density $\rho(\vec{r})$ by its mean $\bar{\rho}$ and its fluctuations $\Delta\rho$

$$\rho(\vec{r}) = \bar{\rho}(\vec{r}) + \Delta\rho(\vec{r}),$$

with

$$\bar{\rho}(\vec{r}) = \frac{1}{V} \int \rho(\vec{r})\, dV, \tag{A.3}$$

and taking into account the large dimension of the scattering volume, the scattered intensity can be rewritten as

$$I(\vec{s}) = \int \int \Delta\rho(\vec{r_1})\,\Delta\rho(\vec{r_2})\,e^{i\,\vec{s}.(\vec{r_1}-\vec{r_2})}\,dV_1\,dV_2 \tag{A.4}$$

This Fourier integral involves the relative distance $\vec{r} = (\vec{r_1} - \vec{r_2})$ for every pair of points, and it is then convenient to divide the double integral into two parts: first integrating over all pairs with equal relative distance, and then integrating over all relative distances, including then the phase factor. In order to do that, it is necessary to introduce the auto-correlation function

$$\gamma(\vec{r}) = \frac{1}{V} \int \Delta\rho(\vec{r_1})\,\Delta\rho(\vec{r_1} - \vec{r})\,dV_1, \tag{A.5}$$

where V is the particle volume. The auto-correlation function $\gamma(\vec{r})$ can be interpreted as the average of the product of two fluctuations at a distance $r = |\vec{r_1} - \vec{r_2}|$=*constant* (Debye and Bueche, 1949), that is to say

$$\gamma(\vec{r}) = < \eta(\vec{r_1})\eta(\vec{r_2}) > .$$

In terms of γ the intensity is now

$$I(\vec{s}) = V \int \gamma(\vec{r})\,e^{i\,\vec{s}.\vec{r}}\,dV. \tag{A.6}$$

As small angle scattering only measures the photons scattered at very small angles ($\sim 3{-}5^o$) with respect to the incident beam, and as the particles in solution have no preferential orientation, further simplifications can be introduced.

From the isotropic particle distribution it follows that the autocorrelation function will only depend on r, and not on its direction. In addition, the isotropic average of the phase term in A.6 will be reduced to (Debye, 1915)

$$< e^{i\vec{s}.\vec{r}} >_{all\ directions} = \frac{\sin(s\,r)}{s\,r};$$

by means of which Eq. A.6 can be rewritten as

$$I(s) = \frac{4\pi}{V} \int r^2 \gamma(r)\, e^{i\,sr}\, dr = 4\pi \int p(r)\, e^{i\,sr}\, dr. \tag{A.7}$$

where $p(r) = r^2 V\, \gamma(r)$ is the pair distance distribution function. For inhomogeneous particles, the $p(r)$ function is proportional to the product of the different scattering lengths $n_i n_j$ (with $n_i = \Delta\rho(\vec{r_i})\, dV_i$) of two volume elements i and j with a centre to centre distance between r and $r + dr$, summing over all pairs at that distance. For the discrete case, this reduces to

$$p(r) = \frac{1}{4\pi} \sum_{i,j-0}^{N} \Delta\rho(\vec{r_i}) \Delta\rho(\vec{r_j}) \delta(|\vec{r_i} - \vec{r_j}| - r), \tag{A.8}$$

where N is the number of electrons in the particle, and $\delta(a - b)$ is Kronecker's δ-function that is one when $a = b$ and zero otherwise.

For homogeneous particles things are less complicated, and the $p(r)$ function reduces to

$$p(r) = \frac{1}{4\pi} \sum_{i,j=0}^{N} \delta(|\vec{r_i} - \vec{r_j}| - r), \tag{A.9}$$

that is to say, a histogram of distances between pairs of points in the particle.

A further description of SAXS and its applications can be found in Section 2.2.

Appendix B

Developed computer programs

B.1 Summary

This chapter will give a brief description, when this has not been done in any of the chapters above, of the computer programs developed throughout this project. All the programs have been developed in the C language and run under the Linux platform. Working versions can be found in the CD attached to this thesis, and further information will be available on request.

B.2 *PDB2Origin*

This program translates a structure so that its centre of mass (CM) is at the origin of coordinates. This routine also determines the maximum inter-atomic distance (D_{MAX}) in the structure.

B.3 *PDB2MainAxis*

This program rotates a structure so that its main axes of inertia are located in the directions defined by the coordinate axes (xyz). Several routines from (Press et al., 1995) were employed to develop this program.

B.4 *gaussian-shape*

The PDB structure is loaded and *PDB2Origin* and *PDB2MainAxis* are invoked. The structure is then enclosed in a box of size $1.3D_{MAX}$, which is divided into a 3D array of smaller cubes (voxels) of side dx. The total number of voxels in the box is defined as N^3_{POINTS}, where $N_{POINTS} = 1.3D_{MAX}/dx$. The coordinates of each atom t ($\vec{r}_t$) in the structure are used to determine which voxel (i, j, k) in the lattice it occupies. The contribution made by an atom t to the Gaussian electron density (ρ) at a voxel (i, j, k) is calculated by

$$\rho^t_{i,j,k} = \frac{A\rho^0_t}{\sigma^3_t} \exp\left[-\pi|\vec{r}_{i,j,k} - \vec{r}_t|^2/\sigma^2_t\right], \tag{B.1}$$

where ρ^0_t is the electron density of atom t, σ_t is the van der Walls radius of atom t, A is a constant, $\vec{r}_{i,j,k}$ is the coordinate vector pointing to the centre of voxel (i, j, k), and $\vec{r}_t$ is the coordinate vector of atom t. After looping over all the atoms, the total Gaussian electron density ρ for each voxel (i, j, k) is obtained by

$$\rho_{i,j,k} = \sum_{t=1}^{N} \rho^t_{i,j,k},$$

where N is the total number of atoms in the original PDB structure. Then, $\rho_{i,j,k}$ is thresholded so that

$$\rho_{i,j,k} = \begin{cases} 0 \text{ if } \rho_{i,j,k} < 0_t \\ \rho_{i,j,k} \text{ otherwise} \end{cases} \tag{B.2}$$

where θ_t is a user-defined threshold. Finally, several expansion-contraction cycles $(6-10)$ of the ρ matrix were performed as in Gerstein (1992); Lee and Richards (1971) in order to fill holes and cavities in the gaussian electron density.

B.5 *volumePDB*

The program loads a bead model or high-resolution structure in PDB format, calls *PDB2Origin* to translate it to the origin of coordinates and *PDB2MainAxis* to align its main axes of inertia with the coordinate versors. Then, it invokes *gaussian-shape* to obtain the Gaussian electron density $\rho_{i,j,k}$, and calculates the total number of occupied voxels (N_o, i. e. voxels with nonzero electron density). The volume of the macromolecule can be estimated as

$$V = N_o \times dx^3.$$

B.6 *PDB2average*

The program PDB2average is designed to load a number n_0 of structures (either bead or high-resolution models) and obtain from them a consensus average electron density. The program assumes that the models have been already superimposed. At first, all the different structures are loaded and their Gaussian electron densities calculated, as described above. Then, the global Gaussian electron density is obtained by averaging of the individual electron densities. Finally, the global electron density is thresholded by a value chosen by the user, and one atom is placed for every node containing a non-zero electron density. Alternatevely, the gloabl Gaussian electron density is populated with a user-defined number of atoms by using a Monte Carlo strategy.

B.7 Overlap removal routines

A few routines were developed for the quantification and removal of overlaps and voids from bead models. These routines are employed in the program *PDB2overlap* but are also used by other programs.

B.7.1 Quantifying overlaps

The overlap between any two beads in the structure (i, j) is estimated as

$$\zeta_{i,j} = R_j + R_i - |\vec{r}_i - \vec{r}_j|, \tag{B.3}$$

where R_j is the radius of bead j, R_i is the radius of bead i, $\vec{r}_j$ are the coordinates of bead j and $\vec{r}_i$ those of bead i. By looping over all pairs of beads in the structure, the total number of overlaps is counted as

$$N_{overlaps} = \sum_{i=1}^{N} \sum_{j=1}^{N} \theta(\zeta_{i,j} - \Theta),$$

where N is the number of beads, $\theta()$ is the Heaviside function, and Θ is a threshold usually set at 0.01 Å,

B.7.2 Eliminating overlaps

This routine reduces the overlaps over the structure in an asynchronous manner. The steps are as follows. (1) The matrix $\zeta_{i,j}$ is calculated. (2) The overlaps are sorted in ascending order and the radii of those atoms i, j with highest overlap are reduced by

$$R_i = R_i \times \Gamma, \; R_j = R_j \times \Gamma$$

where Γ is an overlap reduction factor, usually 0.99. This reduction is only applied to beads with a radius higher than a predefined minimum radius, usually 0.5 Å. (3) If the number of overlaps is nonzero, the procedure restarts at step (1).

Finally, after there are no more overlaps in the bead model, a tangency criterium is applied. The radius of each bead i is incremented until its overlap with any of its neightbours is exactly zero (as defined in Eq. B.3). This procedure is repeated for each bead i in the structure in order to reduce voids.

B.7.3 *PDB2overlap*

This program integrates the routines above. *PDB2overlap* loads a bead model, which consists of N bead coordinates ($\vec{r}_i$) and radii (R_i), estimates the overlaps and eliminates them, if necessary.

B.8 *newAtoB*

NewAtoB was developed in order to grid a structure of a macromolecule on a cubic lattice, thus producing a so-called bead model with reduced resolution. This computer program was systematically employed to reduce the resolution of high-resolution models when performing hydrodynamic calculations. A detailed description can be found in Section 3.3.1.

B.9 *namot2script*

The program generates scripts in NAMOT (Tung and Carter, 1994) scripting language, each coding for a particular DNA structure. The inputs of the program are: the choosen model, the range of the parameters that is to be explored, and the number of structures in the chosen parameter range. The output file can be

directly run in NAMOT, which generates a series of PDB models that are then processed through CRYSOL, CRYSON or HYDROPRO in order to calculate their scattering and hydrodynamic properties.

B.10 *domain-rot*

The hypothetical structure of a macromolecule is divided into n domains, for which structural data is available. *Domain-rot* produces myriad instances of the structure of the whole macromolecule from the structures of its composing domains by using a small number of parameters. The permitted translations and rotations of these domains, as well as the ranges of variation of the parameters are specified by the user. For every given conformation of the macromolecule, *domain-rot* predicts its scattering profile (using CRYSOL or CRYSON) and calculates the χ-values that the simulated scattering curves produced when fitted to the experimental data. The minima are found by visual inspection of the contour plots of χ versus the varied parameters. These minima represent the structures of the macromolecule that best fit the experimental scattering data.

B.11 *rayuela*

This program uses generalised rigid-body modelling, combined with a Monte Carlo/simulated annealing optimisation method, to search over a large range of possible conformations for the structure of the macromolecule that best fits solution experimental properties derived from experimental or simulated small angle scattering, fluorescence resonance energy transfer, and analytical ultracentrifugation datasets. A detailed description of the algorithms involved in making this program can be found in Chapter 7.

Bibliography

Abdel-Meguid, S. S., Grindley, N. D., Templeton, N. S. and Steitz, T. A. (1984). Cleavage of the site-specific recombination protein $\gamma\delta$ resolvase: the smaller of two fragments binds DNA specifically. Proc. Natl. Acad. Sci. USA *81*, 2001–2005. 1.6

Ackerman, C. J., Harnett, M. M., Harnett, W., Kelly, S. M., Svergun, D. I. and Byron, O. (2003). 19 Å solution structure of the filarial nematode immunomodulatory protein, ES-62. Biophys. J. *84*, 489–500. 3.2.3, 7.2

Adrian, M., Heggeler-Bordier, B., Wahli, W., Stasiak, A. Z., Stasiak, A. and Dubochet, J. (1990). Direct visualization of supercoiled DNA molecules in solution. EMBO J. *9*, 4551–4554. 2.2.4

Arnold, P. H., Blake, D. G., Grindley, N. D. F., Boocock, M. R. and Stark, W. M. (1999). Mutants of Tn3 resolvase which do not require accessory sites for recombination activity. EMBO J. *18*, 1407–1414. 3.1.5, 6.1

Aslam, M., Guthridge, J. M., Hack, B. K., Quigg, R. J., Holers, V. M. and Perkins, S. J. (2003). The extended multidomain solution structures of the complement protein Crry and its chimeric conjugate Crry-Ig by scattering, analytical ultracentrifugation and constrained modelling: implications for function and therapy. J. Mol. Biol. *329*, 525–550. 7.2

Aslam, M. and Perkins, S. (2001). Folded-back solution structure of monomeric factor h of human complement by synchroton x-ray and neutron scattering,

analytical ultracentrifugation and constrained molecular modelling. J. Mol. Biol. *309*, 1117–1138. 7.2

Azaro, M. and Landy, A. (2002). λ integrase and the λ Int family. In Mobile DNA II, Craig, N., Craigie, R., Gellert, M. and Lambowitz, A., eds. (Washington, DC: American Society of Microbiology Press), pp. 118–148. 1.1.1

Barnes, I. and Zemb, T. (1988). Calculation of small-angle scattering from models for surfactant systems. J. Appl. Cryst. *21*, 373–379. 2.2.3.1

Barre, F.-X. and Sherratt, D. (2002). Xer site-specific recombination: promoting chromosome segregation. In Mobile DNA II, Craig, N., Craigie, R., Gellert, M. and Lambowitz, A., eds. (Washington, DC: American Society of Microbiology Press), pp. 149–161. 1.1.1

Bednarz, A. L., Boocock, M. R. and Sherratt, D. J. (1990). Determinants of correct *res* site alignment in site-specific recombination by Tn3 resolvase. Genes Dev. *4*, 2366–2375. 4.4, 6.1, 6.3.1

Bell, L. and Byers, B. (1979). Occurrence of crossed strand-exchange forms in yeast DNA during meiosis. Proc. Natl. Acad. Sci. USA *76*, 3445–3449. 5.2

Benjamin, H. W., Matzuk, M. M., Krasnow, M. A. and Cozzarelli, N. R. (1985). Recombination site selection by Tn3 resolvase: topological tests of a tracking mechanism. Cell *40*, 147–158. 1.7

Bianchi, M., Beltrame, M. and Paonessa, G. (1989). Specific recognition of cruciform DNA by nuclear-protein HMG1. Science *243*, 1056–1059. 5.2

Blake, D. G., Boocock, M. R., Sherratt, D. J. and Stark, W. M. (1995). Cooperative binding of Tn3 resolvase monomers to a functionally asymmetric binding site. Curr. Biol. *5*, 1036–1046. 4.1, 4.3.3, 4.4, 4.4, 6.3.1

Boocock, M., Brown, J. and Sherratt, D. (1987). Topological specificity in Tn3 resolvase catalysis. In DNA Replication and Recombination, Kelly, T. and McMacken, R., eds. (New York: Alna R. Liss), pp. 703–718. 1.7, 1.8, 1.8

Boocock, M. R., Brown, J. L. and Sherratt, D. J. (1986). Structural and catalytic properties of specific complexes between Tn3 resolvase and the recombination site *res*. Biochem. Soc. Trans. *14*, 214–216. 1.7, 1.8

Boulin, C. J., Kempf, R., Gabriel, A. and Koch, M. H. J. (1988). Data acquisition systems for linear and area x-ray detectors using delay-line readout. Nucl. Instr. Meth. in Phys. Res. *269*, 312–320. 3.2.6

Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormo, M. and Remington, S. J. (1997). Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. Proc. Nat. Acad. Sci. USA *94*, 2306–2311. 6.2.1, 6.2.5, 6.3.4

Burke, M. E., Arnold, P. H., He, J., Wenwieser, S. V. C. T., Rowland, S.-J., Boocock, M. R. and Stark, W. M. (2004). Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. Mol. Microbiol. *51*, 937–948. 3.1.5, 6.1, 6.3.1, 6.5, 6.5.1

Byron, O. (1997). Construction of hydrodynamic bead models from high-resolution x-ray crystallographic or nuclear magnetic resonance data. Biophys. J. *72*, 408–415. 3.3.1, 7.3.4

Calascibetta, F., de Santis, P., Morosetti, S., Palleschi, A. and Savino, M. (1984). Modelling of the DNA cruciform core. Gazz. Chim. Ital. *114*, 437–441. 5.2

Carrasco, B., García de. la. Torre., J., Davis, K. G., Jones, S., Athwal, D., Walters, C., Burton, D. R. and Harding, S. E. (2001). Crystallohydrodynamics for solving the hydration problem for multi-domain proteins: open physiological conformations for human IgG. Biophys. Chem. *93*, 181–196. 7.2

Chacón, P., Diaz, J. F., Moran, F. and Andreu, J. M. (2000). Reconstruction of protein form with x-ray solution scattering and a genetic algorithm. J. Mol. Biol. *299*, 1289–1302. 2.2.3.2

Chacón, P., Moran, F., Diaz, J. F., Pantos, E. and Andreu, J. M. (1998). Low-resolution structures of proteins in solution retrieved from x-ray scattering with a genetic algorithm. Biophys. J. *74*, 2760–2775. 7.2

Chalikian, T. V. (2003). Volumetric properties of proteins. Annu. Rev. Biophys. Biomol. Struct. *32*, 207–235. 6.3.2

Chamberlain, D., Keeley, A., Aslam, M., Arenas-Licea, J., Brown, T., Tsaneva, I. and Perkins, S. (1998). A synthetic Holliday junction is sandwiched between two tetrameric *mycobacterium leprae* RuvA structures in solution: new insights from neutron scattering contrast variation and modelling. J. Mol. Biol. *284*, 385–400. 5.2, 5.5

Chen, Y. and Rice, P. A. (2003). New insight into site-specific recombination from Flp recombinase-DNA structures. Annu. Rev. Biophy. Biomol. Struct. *32*, 135–159. 6.1

Clegg, R., Murchie, A. and Lilley, D. (1994). The solution structure of the four-way DNA junction at low-salt conditions: a fluorescence resonance energy transfer analysis. Biophys. J. *66*, 99–109. 5.4.2.1

Cox, M., Goodman, M., Kreuzer, K., Sherratt, D., Sandler, S. and Marians, K. (2000). The importance of repairing stalled replication forks. Nature *404*, 37–41. 5.2

Cozzarelli, N. R., Krasnow, M. A., Gerrard, S. P. and White, J. H. (1984). A topological treatment of recombination and topoisomerases. Cold Spring Harb. Symp. Quant. Biol. *49*, 383–400. 1.7

Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W. and Cozzarelli, N. R. (1999). The topological mechanism of phage lambda integrase. J. Mol. Biol. *289*, 747–775. 6.1

de Azevedo, W. F., dos Santos, G. C., dos Santos, D. M., Olivieri, J. R., Canduri, F., Silva, R. G., Basso, L. A., Renard, G., da Fonseca, I. O., Mendes, M. A., S., P. M. and Santos, D. S. (2003). Docking and small angle x-ray scattering studies of purine nucleoside phosphorilase. Biochem. Biophys. Res. Comm. *309*, 923–928. 7.2

Debye, P. (1915). Zerstreuung von Röntgenstrahlen. Ann. Physik *46*, 809. 4

Debye, P. and Bueche, A. (1949). Appl. Phys. *20*, 518. 4

Dornberger, U., Hillisch, A., Gollmick, F. A., Fritzsche, H. and Diekmann, S. (1999). Solution structure of a five-adenine bulge loop within a DNA duplex. Biochemistry *38*, 12860–12868. 7.2, 7.4.0.1, 7.5

Duckett, D., Murchie, A., Diekmann, S., von Kitzing, E., Kemper, B. and Lilley, D. (1988). The structure of the Holliday junction, and its resolution. Cell *55*, 79–89. 5.2, 5.4.2.1, 5.4.2.1

Edmonds, A. (1957). Angular Momentum in Quantum Mechanics. Princeton Univ. Press. 2.2.3.1

Fedorov, B.A. Danesyuk, A. (1978). Large-angle x-ray diffuse scattering, a new method for investigating changes in the conformation of globular proteins in solutions. J. Appl. Crystallogr. *11*, 473–477. 2.2.3.1

Fedorov, B. and Ptitsyn, O.B. Voronin, L. (1972). x-ray diffuse scattering of globular protein solutions: consideration of the solvent influence. FEBS Lett. *28*, 188–190. 2.2.3.1

Feigin, L. A. and Svergun, D. I. (1987). Structure analysis by small-angle x-ray and neutron scattering. Plenum Press, New York. 7.2

Feil, I. K., Malfois, M., Hendle, J., van der Zandt, H. and Svergun, D. I. (2001). A novel quaternary structure of the dimeric α-crystallin domain with chaperone-like activity. J. Biol. Chem. *276*, 12024–12029. 7.2

Feng, J. A., Johnson, R. C. and Dickerson, R. E. (1994). Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. Science *263*, 348–355. 1.6

Fogg, J., Kvaratskhelia, M., White, M. and Lilley, D. (2001). Distortion of DNA junctions imposed by the binding of resolving enzymes: a fluorescence study. J. Mol. Biol. *313*, 751–764. 5.3.1

Furlong, J. and Lilley, D. (1986). Highly selective chemical modification of cruciform loops by diethyl pyrocarbonate. Nucleic Acids Res. *14*, 3995–4007. 5.2

García de la Torre, J. (2001). Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins. Biophys. Chem. *93*, 159–170. 3.3.1, 3.3.2

García de la Torre, J., Huertas, M. L. and Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys. J. *78*, 719–730. 3.3.1, 3.3.2, 4.3.1, 5.3.2, 5.4.2.2, 5.4.3, 6.3.2, 7.4.0.1

García de la Torre, J., Navarro, S., López Martínez, M. C., Diaz, F. G. and López Cascales, J. (1994). A computer software for the prediction of hydrodynamic properties of macromolecules. Biophys. J. *67*, 530–531. 3.3.1, 3.3.2, 7.3.4

Gellert, M., Mizuuchi, K., O'Dea, M., Ohmori, H. and Tomizawa, J. (1979). DNA gyrase and DNA supercoiling. Cold Spring Harb. Symp. Quant. Biol. *43*, 35 40. 5.2

Gerstein, M. (1992). A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. Acta Cryst. *A48*, 271–276. B.4

Ghosh, R. E., Egelhaaf, S. U. and Rennie, A. R. (1989). A computing guide for small-angle scattering experiments, report no. ILL98GII14T. Technical report Institut Laue-Langevin, Grenoble, France. 3.2.5

Gilbert, R. J. C., Fucini, P., Connell, S., Fuller, S. D., Nierhaus, K. H., Robinson, C. V., Dobson, C. M. and Stuart, D. I. (2004). Three-Dimensional Structures of Translating Ribosomes by Cryo-EM. Mol. Cell *14*, 57–66. 7.2

Glatter, O. (1980). Acta Phys. Austriaca *52*, 243–256. 2.2.3.1

Glatter, O. and Kratky, O. (1982). Small Angle X-ray Scattering. Academic Press, London. 2.2.2, 2.2.2, 2.2.3.1

Gohlke, C., Murchie, A. I., Lilley, D. M. and Clegg, R. M. (1994). Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy transfer. Proc. Natl. Acad. Sci. U S A *91*, 11660–11664. 7.2, 7.4.0.1, 7.5

Goldstein, H. (1980). Classical Mechanics. Addison-Wesley Publishing Company. 7.3.1

Gough, G. and Lilley, D. (1985). DNA bending induced by cruciform formation. Nature *313*, 154–156. 5.2

Gough, G., Sullivan, K. and Lilley, D. (1986). The structure of cruciforms in supercoiled DNA: probing the single-stranded character of nucleotide bases with bisulphite. EMBO J. *5*, 191–196. 5.2

Grassian, V., Brady, G. W. and Benham, C. J. (1983). X-ray conformational study of the DNA duplex in solution. Biopolymers *22*, 1523–1543. 2.2.4

Grindley, N. D., Lauth, M. R., Wells, R. G., Wityk, R. J., Salvo, J. J. and Reed, R. R. (1982). Transposon-mediated site-specific recombination: identification of three binding sites for resolvase at the *res* sites of $\gamma\delta$ and Tn3. Cell *30*, 19–27. 4.4, 4.4, 6.1

Grindley, N. D. F. (1994). Resolvase-mediated site-specific recombination. In Nucleic Acids and Molecular Biology, Eckstein, F. and Lilley, D. M. J., eds. (Berlin - Heidelberg: Springer-Verlag), pp. 236–267. 1.4, 1.8, 1.8, 4.1, 6.1

Grindley, N. D. F. (2002). The movement of Tn3 like elements: transposition and cointegrate resolution. In Mobile DNA II, Craig, N., Craigie, R., Gellert, M. and Lambowitz, A., eds. (Washington, DC: American Society of Microbiology Press), pp. 272–302. 1.3, 1.8, 4.1, 6.1

Guinier, A. (1963). X-ray Diffraction in crystals, imperfect crystals and amorphous bodies. Freeman and Co, San Francisco and London. A, A.0.1, A.0.1

Guinier, A. and Fournet, G. (1955). Small angle scattering of x-rays. Wiley, New York. 2.2.1, 3.2.5, 6.3

Guo, F., Gopaul, D. N. and Van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature *389*, 40–46. 1.1, 2

Haber, J. E. (2000). Partners and pathways - repairing a double-strand break. Trends Genet. *16*, 259–264. 5.2

Hammermann, M., Brun, N., Klenin, K., May, R., Toth, K. and Langowski, J. (2000). The diameter of the DNA superhelix decreases with salt concentration: SANS measurements and Monte Carlo simulations. J. Appl. Cryst. *33*, 526–529. 2.2.4

Hansen, S. (1990). Calculation of small-angle scattering profiles using Monte Carlo simulation. J. Appl. Cryst. *23*, 344–346. 2.2.3.1, 2.2.3.1

Hatfull, G. F., Sanderson, M. R., Freemont, P. S., Raccuia, P. R., Grindley, N. D. and Steitz, T. A. (1989). Preparation of heavy-atom derivatives using site-directed mutagenesis. Introduction of cysteine residues into $\gamma\delta$ resolvase. J. Mol. Biol. *208*, 661–667. 1.6

He, J., McIlwraith, M. J., Burke, M. E., Boocock, M. R. and Stark, W. M. (2002). Synapsis of Tn3 recombination sites: unpaired sites destabilize synapses by a partner exchange mechanism. J. Mol. Biol. *319*, 385–393. 6.2.1

Heichman, K. A. and Johnson, R. C. (1990). The Hin invertasome: protein-mediated joining of distant recombination sites at the enhancer. Science *249*, 511–517. 6.5.1

Heller, W. T., Krueger, J. K. and Trewhella, J. (2003). Further insights into calmodulin-myosin light chain kinase interaction from solution scattering and shape restoration. Biochemistry *42*, 10579–10588. 2.2.3.2, 7.2, 7.5

Henderson, S. J. (1996). Monte Carlo modeling of small-angle scattering data from non-interacting homogeneous and heterogeneous particles in solution. Biophys. J. *70*, 1618–1627. 2.2.3.1

Heyduk, T., Ma, Y., Tang, H. and Ebright, R. (1996). Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interactions. In Methods in Enzymology vol. 274,. (New York: Academic Press, Inc.), chapter 37, pp. 493–503. 2.4

Holliday, R. (1964). A mechanism for gene conversion in fungi. Genet. Res. *5*, 382–304. 5.2

Huffman, K. E. and Levene, S. D. (1999). DNA-sequence asymmetry directs the alignment of recombination sites in the Flp synaptic complex. J. Mol. Biol. *286*, 1–13. 6.1

Hughes, R. E., Rice, P. A., Steitz, T. A. and Grindley, N. D. (1993). Protein-protein interactions directing resolvase site-specific recombination: a structure-function analysis. EMBO J. *12*, 1447–1458. 1.6, 1.6

Humphrey, W., Dalke, A. and Schulten, K. (1996). VMD – Visual Molecular Dynamics. J. Mol. Graph. *14*, 33–38. 3.3.5

Jameson, D. and Sawyer, W. (1995). Fluorescence anisotropy applied to biomolecular interactions. In Methods in Enzymology, Sauer, K., ed. vol. 246,. (New York: Academic Press, Inc.), chapter 12, pp. 283–300. 2.4

Johnson, R. (2002). Bacterial site-specific DNA inversion. In Mobile DNA II, Craig, N., Craigie, R., Gellert, M. and Lambowitz, A., eds. (Washington, DC: American Society of Microbiology Press), pp. 230–271. 1.1.1

Kallenbach, N., Ma, R., Wand, A., Veeneman, G., van Boom, J. and Seeman, N. (1983). Fourth rank immobile nucleic acid junctions. J. Biomol. Struct. Dyn. *1*, 159–168. 5.2

King, S., Washington, C., Attwood, D., Booth, C., Mai, S., Yang, Y.-W. and Cosgrove, T. (2000). Polymer bristles: a SANS study. J. Appl. Cryst. *33*, 664–668. 3.2.5

Kirkpatrick, S., Gelatt, C. D. and Vecci, M. P. (1983). Optimisation by simulated annealing. Science *220*, 671–680. 7.3.6

Kitts, P. and Nash, H. (1987). Homology-dependent interactions in phage lambda site-specific recombination. Nature *329*, 346–348. 5.2

Koch, M. H., Vachette, P. and Svergun, D. I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q. Rev. Biophys. *36*, 147–227. 2.2.3.2, 5.5, 7.2

Koch, M. H. J. and Bordas, J. (1983). X-ray diffraction and scattering on disordered systems using synchroton radiation. Nucl. Instrum. Meth. in Phys. Res. *208*, 461–469. 3.2.6

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. and Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Cryst. *36*, 1277–1282. 3.2.5, 3.2.6

Kornberg, A. and Baker, T. (1991). DNA replication. W.H. Freeman and Co., New York. 1.1

Kosak, H. and Kemper, B. (1990). Large-scale preparation of T4 endonuclease VII from over-expressing bacteria. Eur. J. Biochem. *194*, 779–784. 5.2

Kozin, M. and Svergun, D. (2001). Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. *34*, 33–41. 7.2, 7.5

Kratky, O. (1982). In Small Angle X-ray Scattering, O., G. and Kratky, O., eds. (London: Academic Press), chapter 1. A, 4, 4

Kratky, O. and Pils, I. (1972). Recent advances and applications of diffuse x-ray small-angle scattering on biopolymers in dilute solutions. Q. Rev. Biophys. *5*, 481–537. 2.2.3.1

Landau, D. P. and Binder, K. (2000). A guide to Monte Carlo simulations in statistical physics. Cambridge University Press, Cambridge, UK. 7.3.6

Lattman, E. E. (1989). Rapid calculation of the solution scattering profile from a macromolecule of known structure. Proteins *5*, 149–155. 2.2.3.1, 2.2.3.1, 2.2.3.2

Laue, T. M., Shah, B. D., Ridgeway, T. M. and Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science, Harding, S. E.,

Rowe, A. J. and Horton, J. C., eds. (Cambridge, United Kingdom: Royal Society for Chemistry), pp. 90- 125. 3.2.1

Lebowitz, J., Lewis, M. S. and Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci. *11*, 2067–2079. 2.3.1, 2.3.2, 7.2

Lee, B. and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. *55*, 379–400. 2.2.3.1, 5.4.2.2, B.4

Leschziner, A. E. and Grindley, N. D. F. (2003). The architecture of the $\gamma\delta$ resolvase crossover site synaptic complex revealed by using constrained DNA substrates. Mol. Cell *12*, 775–781. 6.1, 6.3.3.2, 6.5

Lilley, D. (1980). The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. Proc. Natl. Acad. Sci. USA *77*, 6468–6472. 5.2

Lilley, D. and Clegg, R. (1993). The structure of branched DNA species. Q. Rev. Biophys. *26*, 131–175. 5.2, 5.4.2.1, 5.4.2.1, 5.5, 7.2

Liu, T., DeRose, E. F. and Mullen, G. P. (1994). Determination of the structure of the DNA binding domain of $\gamma\delta$ resolvase in solution. Protein. Sci. *3*, 1286–1295. 1.5, 1.6

Lohman, T. and Mascotti, D. (1992). Nonspecific ligand-DNA equilibrium binding parameters determined by fluorescence methods. In Methods in Enzymology vol. 212,. (New York: Academic Press, Inc.), chapter 25, pp. 425–459. 2.4

Macke, T. and Case, D. A. (1998). Modeling unusual nucleic acid structures. In Molecular Modeling of Nucleic Acids, Leontes, N. B. and Santa Lucia, J., eds. (Washington, DC: American Chemical Society), pp. 379–393. 3.3.4, 6.2.5

Maluf, N. K. and Lohman, T. M. (2003). Self-association equilibria of *escherichia coli* UvrD helicase studied by analytical ultracentrifugation. J. Mol. Biol. *325*, 889–912. 6.3.2

Mariani, P., Carsughi, F., Spinozzi, F., Romanzetti, S., Meier, G., Casadio, R. and Bergamini, C. M. (2000). Ligand-induced conformational changes in tissue transglutaminase: Monte Carlo analysis of small-angle scattering data. Biophys. J. *78*, 3240–3251. 2.2.3.1

Matsunaga, N. and Nagashime, A. (1983). Transport-properties of liquid and gascous D_2O over a wide range of temperature and pressure. J. Phys. Chem. Ref. Data *12*, 933–966. 3.2.1

Mattinen, M.-L., Paakkonen, K., Ikonen, T., Craven, J., Drakenberg, T., Serimaa, R., Waltho, J. and Annila, A. (2002). Quaternary structure built from subunits combining NMR and small-angle x-ray scattering data. Biophys. J. *83*, 1177–1183. 7.2

McCulloch, R., Coggins, L., Colloms, S. and Sherratt, D. (1994). Xer-mediated site-specific recombination at *cer* generates Holliday junctions *in vivo*. EMBO J. *13*, 1344–1355. 5.2

McIlwraith, M. J., Boocock, M. R. and Stark, W. M. (1996). Site-specific recombination by Tn3 resolvase, photocrosslinked to its supercoiled DNA substrate. J. Mol. Biol. *260*, 299–303. 1.9

McIlwraith, M. J., Boocock, M. R. and Stark, W. M. (1997). Tn3 resolvase catalyses multiple recombination events without intermediate rejoining of DNA ends. J. Mol. Biol. *266*, 108–121. 1.9, 6.5.1

Müller, J. (1983). Calculation of scattering curves for macromolecules in solution and comparison with results of methods using effective atomic scattering factors. J. Appl. Cryst. *16*, 74–82. 2.2.4

Müller, J., Hansen, S. and Pürschel, H. (1996). The use of small-angle scattering and the maximum-entropy method for shape-model determination from distance-distribution functions. J. Appl. Cryst. *29*, 547–554. 2.2.3.1

Murley, L. L. and Grindley, N. D. (1998). Architecture of the $\gamma\delta$ resolvase synaptosome: oriented heterodimers identity interactions essential for synapsis and recombination. Cell *95*, 553–562. 1.8

Ninio, J., Luzzatti, V. and Yaniv, M. (1972). J. Mol. Biol. *71*, 217–229. 2.2.3.1

Nöllmann, M., He, J., Byron, O. and Stark, W. M. (2004a). Solution structure of the Tn3 resolvase-crossover site synaptic complex. Molecular Cell *16*, 127–137. 6, 7.5

Nöllmann, M., Stark, W. M. and Byron, O. (2004b). Low-resolution reconstruction of a synthetic DNA Holliday junction. Biophys. J. *86*, 3060–3069. 2.2.3.2, 5.1, 7.2

Normark, B. H. and Normark, S. (2002). Evolution and spread of antibiotic resistance. J. Intern. Med. *252*, 91–106. 1.2

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. Protein Sci. *4*, 2411–2423. 3.2.1

Pan, B., Maciejewski, M. W., Marintchev, A. and Mullen, G. P. (2001). Solution structure of the catalytic domain of $\gamma\delta$ resolvase. implications for the mechanism of catalysis. J. Mol. Biol. *310*, 1089–1107. 1.5, 1.6, 4.3.1, 4.4, 6.1

Panayotatos, N. and Wells, R. (1981). Cruciform structures in supercoiled DNA. Nature *289*, 466- 470. 5.2

Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., Baker, S., Basham, D., Bentley, S. D., Brooks, K., Cerdeno-Tarraga, A. M., Chillingworth, T., Cronin, A., Davies, R. M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A. V., Leather, S., Moule, S., Oyston, P. C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B. G. (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature *413*, 523–527. 1.2

Pavlov, M. Y. and Fedorov, B. A. (1983). Improved technique for calculating x-ray scattering intensity of bio-polymers in solution - evaluation of the form, volume and surface of a particle. Biopolymers *22*, 1507–1522. 2.2.3.1, 5.4.2.2

Perkins, S. (1988). X-ray and neutron solution scattering. In New comprehensive biochemistry, Neuberger, A. and van Deenen, L., eds. (Amsterdam: Elsevier), chapter 6, pp. 143–263. 1

Perrin, F. (1926). Polarization de la lumière de fluorescence. Vie moyenne de molecules dans l'état excité. J. Phys. Radium. *7*, 390. 2.4

Petoukhov, M. V., Eady, N. A. J., Brown, K. A. and Svergun, D. I. (2002). Addition of missing loops and domains to protein models by x-ray solution scattering. Biophys. J. *83*, 3113–3125. Evaluation Studies. 5.4.2.2, 7.2

Petoukhov, M. V., Svergun, D. I., Konarev, P. V., Ravasio, S., van den Heuvel, R. H. H., Curti, B. and Vanoni, M. A. (2003). Quaternary structure of *Azospirillum brasilense* NADPH-dependent glutamate synthase in solution as revealed by synchrotron radiation x-ray scattering. J. Biol. Chem. *278*, 29933–29939. 7.2

Petry, I., Wisniewski, J. R. and Szewczuk, Z. (2001). Conformational stability of six truncated cHMG proteins studied in their mixture by H/D exchange and eletrospray ionization mass spectrometry. Acta Biochim. Pol. *48*, 1131–1136. 6.3.2

Porod, G. (1948). Acta Phys. Austriaca *2*, 255–292. 2.2.2

Porod, G. (1951). Die röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. Kolloïd Zeitschrift *124*, 83–114. 3

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1995). Numerical recipes in C. The art of scientific computing. Cambridge University Press. B.3

Rafferty, J., Sedelnikova, S., Hargreaves, D., Artymiuk, P.J., Baker, P., Sharples, G., Mahdi, A., Lloyd, R. and Rice, D. (1996). Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. Science *274*, 415–421. 5.2

Reed, R. (1981). Resolution of cointegrates between transposons $\gamma\delta$ and Tn3 defines the recombination site. Proc. Natl. Acad. Sci. USA *78*, 3428–3432. 1.5, 1.7

Reed, R. R. and Grindley, N. D. F. (1981). Transposon-mediated site-specific recombination *in vivo*: DNA cleavage and protein-DNA linkage at the recombination site. Cell *25*, 721–728. 1.4, 4.1

Rice, P. (2002). Theme and variation in tyrosine recombinases: structure of a Flp-DNA complex. In Mobile DNA II, Craig, N., Craigie, R., Gellert, M. and Lambowitz, A., eds. (Washington, DC: American Society of Microbiology Press), pp. 219–229. 1.1.1, 2

Rice, P. and Steitz, T. A. (1994a). Model for a DNA-mediated synaptic complex suggested by a crystal packing of $\gamma\delta$ resolvase subunits. EMBO J. *13*, 1514–1524. 1.6, 1.8, 1.8, 6.1, 6.1

Rice, P. and Steitz, T. A. (1994b). Refinement of $\gamma\delta$ resolvase reveals a strikingly flexible molecule. Structure *2*, 371–384. 1.6, 1.6, 4.3.1, 4.3.1, 4.4, 6.1

Richards, E. (1980). An introduction to physical properties of large molecules in solution. Cambridge University Press. 2.3.2

Rosenzweig, A. C., Frederick, C. A., Lippard, S. J. and Nordlund, P. (1993). Crystal structure of a bacterial non-haem iron hydroxylase that catalyses the biological oxidation of methane. Nature *366*, 537–543. 7.2, 7.5

Salvo, J. and Grindley, N. D. F. (1988). Resolvase bends the *res* site into a recombinogenic complex. EMBO J. *7*, 3609–3616. 1.8

Sambrook, J. and Russell, D. (2001). Molecular cloning. Cold Spring Harbor Laboratory Press, New York. 3.1.2, 3.1.4

Sanderson, M. R., Freemont, P. S., Rice, P. A., Goldman, A., Hatfull, G. F., Grindley, N. D. F. and Steitz, T. A. (1990). The crystal structure of the catalytic domain of the site-specific recombination enzyme $\gamma\delta$ resolvase at 2.7 Å resolution. Cell *63*, 1323–1329. 1.1, 1.6, 1.6, 1.8, 4.4, 6.1

Sano, Y., Inoue, H., Kajiwara, K., Urakawa, H. and Hiragi, Y. (1994). Self-assembling process of cylindrical virus coat proteins as observed by synchrotron small-angle x-ray scattering. J. Biochem. (Tokyo) *115*, 1058–1063. 2.2.3.1

Sarkis, G. J., Murley, L. L., Leschziner, A. E., Boocock, M. R., Stark, W. M. and Grindley, N. D. F. (2001). A model for the $\gamma\delta$ resolvase synaptic complex. Mol. Cell *8*, 623–631. 1.8, 2, 3.1.7, 4.4, 6.1, 6.2.5, 6.3.1, 6.3.3.1, 4, 6.5, 6.5.1

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys. J. *78*, 1606–1619. 2.3.1.1, 2.3.1.1, 2.3.2, 3.2.2, 4.3.1, 4.2, 5.3.2, 5.3, 6.2.3, 6.3.2, 6.2

Schuck, P. (2003a). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal. Biochem. *320*, 104–124. 2.3.2, 6.3.2

Schuck, P. (2003b). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal. Biochem. *320*, 104–124. 2.3.2

Schwacha, A. and Kleckner, N. (1995). Identification of double Holliday junctions as intermediates in meiotic recombination. Cell *83*, 783–791. 5.2

Scott, D. J., Grossmann, J. G., Tame, J. R. H., Byron, O., Wilson, K. S. and Otto, B. R. (2002). Low resolution solution structure of the Apo form of *Escherichia coli* haemoglobin protease Hbp. J. Mol. Biol. *315*, 1179–1187. 7.2

Seeman, N. and Kallenbach, N. (1994). DNA branched junctions. Ann. Rev. Biophys. Biomol. Struct. *23*, 53–86. 5.4.2.1

Sigal, N. and Alberts, B. (1972). Genetic recombination: the nature of a crossed strand-exchange between two homologous DNA molecules. J. Mol. Biol. *71*, 789–793. 5.2

Sobell, H. (1972). Molecular mechanism for genetic recombination. Proc. Natl. Acad. Sci. USA *59*, 2483–2487. 5.2

Spotorno, B., Piccinini, L., Tassara, G., Ruggiero, C., Nardini, M., Molina, F. and Rocco, M. (1997). BEAMS (BEAds Modelling System): a set of computer programs for the generation, the visualisation and the computation of the hydrodynamic and conformational properties of bead models of proteins. Eur. Biophys. J. Biophys. Lett. *25*, 373–384. 3.3.1

Stark, W. and Boocock, M. (1994). The linkage change of a knotting reaction catalysed by Tn3 resolvase. J. Mol. Biol. *239*, 25–36. 1.9

Stark, W., Grindley, N., Hatfull, G. and Boocock, M. (1991). Resolvase-catalysed reactions between *res* sites differing in the central dinucleotide of subsite I. EMBO J. *10*, 3541–3548. 1.9

Stark, W., Sherratt, D. and Boocock, M. R. (1989). Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. Cell *58*, 779–790. 1.9, 2, 6.1

Stassinopoulos, A., Ji, J., Gao, X. and Goldberg, I. H. (1996). Solution structure of a two-base DNA bulge complexed with an enediyne cleaving analog. Science *272*, 1943–1946. 7.2

Strater, N., Sherratt, D. J. and Colloms, S. D. (1999). X-ray structure of aminopeptidase A from *Escherichia coli* and a model for the nucleoprotein complex in Xer site-specific recombination. EMBO J. *18*, 4513–4522. 3.3.4, 6.2.5

Stühmeier, F., Clegg, R., Hillisch, A. and Diekmann, S. (2000). Practical aspects of fluorescence resonance energy transfer (FRET) and its applications in nucleic acid biochemistry. In DNA-protein interactions, Travers, A. and Buckle, M., eds. (New York: Oxford University Press), chapter 6, pp. 77–94. 7.2, 7.3.3

Stuhmeier, F., Hillisch, A., Clegg, R. M. and Diekman, S. (2000). Fluorescence energy transfer analysis of DNA structures containing several bulges and their interaction with CAP. J. Mol. Biol. *302*, 1081–1100. 7.2, 7.4.0.1, 7.4.0.1, 7.5

Stuhrmann, H. B. (1970). Ein neues Verfahren zur Bestimmung der Oberflaechenform und der inneren Struktur von geloesten globularen Proteinen aus Röntgenkleinwinkelmessungen. Zeitschr. Physik. Chem. Neue Folge. *72*, 177–198. 2.2.3.1, 7.2

Svedverg, T. and Pedersen, K. (1940). The Ultracentrifuge. Oxford University Press. 2.3.1

Svergun, D., Richard, S., Koch, M., Sayers, Z., Kuprin, S. and Zaccai, G. (1998). Protein hydration in solution: experimental observation by x-ray and neutron scattering. Proc. Nat. Acad. Sci. USA *95*, 2267–2272. 3.3.3, 4.3.1, 7.3.2

Svergun, D., Volkov, V.V., Kozin, M. and Stuhrmann, H. (1996). New developments in direct shape determination from small-angle scattering. Acta Crystallogr. *A52*, 419–426. 2.2.3.2

Svergun, D. I. (1992). Determination of the regularisation parameter in indirect transform using perceptual criteria. J. Appl. Cryst. *25*, 495–503. 3.2.6, 4.3.1, 5.4

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. *76*, 2879–2886. 2.2.3.2, 5.4.2.2, 7.2, 7.5

Svergun, D. I., Barberato, C. and Koch, M. H. J. (1995). CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic structures. J. Appl. Cryst. *28*, 768-773. 2.2.3.1, 2.2.3.1, 2.2.3.1, 2.2.3.1, 2.2.3.2, 3.3.3, 3.3.3, 5.4.2.1, 6.2.4, 6.3.3, 7.3.2, 7.3.2, 7.4.0.1

Svergun, D. I., Malfois, M., Koch, M. H., Wigneshweraraj, S. R. and Buck, M. (2000). Low resolution structure of the $\sigma 54$ transcription factor revealed by x-ray solution scattering. J. Biol. Chem. *275*, 4210–4214. 2.2.3.2

Svergun, D. I. and Nierhaus, K. H. (2000). A map of protein-rRNA distribution in the 70 S *Escherichia coli* ribosome. J. Biol. Chem. *275*, 14432–14439. 7.2

Svergun, D. I., Petoukhov, M. V. and Koch, M. H. J. (2001). Determination of domain structure of proteins from x-ray solution scattering. Biophys. J. *80*, 2946–2953. 2.2.3.2, 5.3.3, 5.4.2.2, 5.8, 7.2

Svergun, D. I., Richard, S., Koch, M. H., Sayers, Z., Kuprin, S. and Zaccai, G. (1998). Protein hydration in solution: experimental observation by x-ray and neutron scattering. Proc. Natl. Acad. Sci. U S A *95*, 2267–2272. 7.2

Tanford, C. (1961). Physical Chemistry of Macromolecule. J. Wiley & Sons, New York. 3.3.1, 3.3.2

Tung, C. S. and Carter, E. S. (1994). Nucleic acid modeling tool (NAMOT): an interactive graphic tool for modeling nucleic acid structures. Comput. Appl. Biosci. *10*, 427–433. 3.3.4, 5.3.2, 5.4.2.1, 5.9, 5.10, 6.2.5, B.9

Vachette, P. and Svergun, D. (2000). Small angle x-ray scattering by solutions of biological macromolecules. In Structure and Dynamics of Biomolecules, Franchon, E., Geissler, E., Hodeau, J.-L., Regnard, J.-R. and Timmins, P., eds. (Oxford, UK: Oxford University Press), chapter 11, pp. 199–237. 2.2.1, 2.2.1, 2.2.3.2

Van Duyne, G. (2002). A structural view of tyrosine recombinase site-specific recombination. In Mobile DNA II, Craig, N., Craigie, R., Gellert, M. and Lambowitz, A., eds. (Washington, DC: American Society of Microbiology Press), pp. 93–117. 1.1.1

Van Duyne, G. D. (2001). A structural view of Cre-loxP site-specific recombination. Annu. Rev. Biophys. Biomol. Struct. *30*, 87–104. 6.1

van Holde, K. (1985). Physical Biochemistry. Prentice Hall. 2.3.2, 2.3.2, 3.3.1

Volkov, V. and Svergun, D. (2003). Uniqueness of *ab initio* shape determination in small-angle scattering. J. Appl. Crystallogr. *36*, 860–864. 2.2.3.2, 7.2, 7.5

Wasserman, S. A. and Cozzarelli, N. R. (1985). Determination of the stereostructure of the product of Tn3 resolvase by a general method. Proc. Natl. Acad. Sci. USA *82*, 1079–1083. 1.7

Watson, M. A., Boocock, M. R. and Stark, W. M. (1996). Rate and selectivity of synapsis of *res* recombination sites by Tn3 resolvase. J. Mol. Biol. *29*, 317 329. 4.4, 6.1

Weber, G. (1952). Polarization of the fluorescence of macromolecules. I. Theory and experimental method. Biochem. J. *51*, 145. 2.4

Weeks, K. M. and Crothers, D. M. (1991). RNA recognition by Tat-derived peptides: interaction in the major groove. Cell *66*, 577–588. 7.2

Wemmer, D., Wand, A., Seeman, N. and Kallenbach, N. (1985). NMR analysis of DNA junctions: imino proton NMR studies of individual arms and intact junction. Biochemistry *24*, 5745–5749. 5.2

Wenwieser, S. (2001). Subunit interactions in regulation and catalysis of site-specific recombination. PhD thesis, University of Glasgow , UK. 3.1.5, 3.1.6

White, M., Giraud-Panis, M., Pohler, J. and Lilley, D. (1997). Recognition and manipulation of branched DNA structure by junction-resolving enzymes. J. Mol. Biol. *269*, 647–664. 5.2

Winzor, D. and Sawyer, W. (1995). Quantitative characterisation of ligand binding. Wiley-Liss, New York. 2.4.0.1, 2.4.0.1, 4.5

Wriggers, W., Milligan, R. A., Schulten, K. and McCammon, J. A. (1998). Self-organizing neural networks bridge the biomolecular resolution gap. J. Mol. Biol. *284*, 1247–1254. 3.3.5, 5.4.2.2

Yang, W. and Steitz, T. A. (1995). Crystal structure of the site specific recombinase $\gamma\delta$ resolvase complexed with a 34 bp cleavage site. Cell *82*, 193–207. (document), 1.5, 1.6, 1.6, 1.7, 2, 4, 4.1, 4.3.1, 4.4, 4.3.1, 4.4, 6.1, 6.1, 6.2.5, 6.3.3.1, 6.3.3.1, 6.5

Zipper, P. and Durchschlag, H. (2000). Prediction of hydrodynamic and small-angle scattering parameters from crystal and electron microscopic structures. J. Appl. Cryst. *33*, 788–792. 2.2.3.1