UNIVERSITY OF SOUTHAMPTON
FACULTY OF SCIENCE
School of Chemistry

On the structure and dynamics of cytochrome c_3 and HU

by

Michael Anthony Durney

A thesis submitted for the degree of

Doctor of Philosophy

March 2004

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE

SCHOOL OF CHEMISTRY

Doctor of Philosophy

ON THE STRUCTURE AND DYNAMICS OF CYTOCHROME C3 AND HU

Michael Anthony Durney

This thesis presents the results of research on the structure and dynamics of proteins using NMR spectroscopy and molecular dynamics. The introductory Chapter 1 presents an overview of NMR and molecular dynamics methodology as applied to proteins and nucleic acids in solution.

Chapter 2 presents the solution structure of cytochrome c_3 from *Desulfomicrobium norvegicum* (c3*Dn*). c3*Dn* is a tetrahaem protein involved in redox reactions in the bacterium. The oxidized and therefore paramagnetic isoform of c3*Dn* was assigned using standard sequential assignment methods. An ensemble of structures was calculated from restraints derived from NOESY spectra using the molecular dynamics program PARADYANA. The structure ensemble accurately represents the structure of the protein in solution as determined by the available data. The parameters of the paramagnetic tensors of the haem groups were fitted using calculated dipolar shifts and the X-ray coordinates. The NMR ensemble is also compared to the X-ray structure confirming that the structures in solution and in the crystal are highly similar.

Chapter 3 presents the assignment of the HU protein from the thermophilic bacterium *Thermotoga maritima* (HU*Tmar*). This homodimeric protein is extremely stable and acts as an architectural factor during the initiation of replication in bacteria. Unexpectedly for a thermostable protein HU*Tmar* exists in two forms in solution as evidenced by the complete resonance assignment by triple resonance NMR techniques. The secondary and tertiary structure of form A resembles the crystal structure according to an analysis of chemical shift and NOESY data. Form B has very similar secondary structure according to chemical shift data with the notable exception of the C-terminus. In contrast, the NOESY data for form B are much less complete indicating a more labile structure.

Chapter 4 examines the backbone dynamics of the two forms of HU*Tmar* using NMR relaxation methods. Form A is found to have similar overall flexibility compared to the previously characterized HU protein from *Bacillus stearothermophilus*. Form B is shown to have a much higher degree of flexibility which is in agreement with the data from Chapter 3. The extent of this phenomenon is unusual for a thermostable protein and is discussed in reference to recent reports in the literature.

to Niall, Anita, Adrian, Mary, and Michael

Finis coronat opus

Contents

List of Figures

List of Tables

DECLARATION OF AUTHORSHIP

I, Michael Anthony Durney, declare that the thesis entitled *On the structure and dynamics of cytochrome* c_3 *and HU* and the work presented in it are my own. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given, and, with the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signature and date: [redacted]

DUBLIN 15th OCTOBER 2004

Acknowledgements

The fact that this thesis exists at all is due to the very generous help that I have received from colleagues and friends throughout Europe.
First, I would like to thank David Turner, my supervisor at Southampton, for taking me on as a graduate student.
Geoffrey Luckhurst, my advisor, gave me honest and encouraging advice which was highly appreciated.
Antonio Xavier and Lorraine Brennan gave me a warm welcome in Lisbon.
Jill Queen played her own important role which I would like to acknowledge. Thank you also to Nicola Crombie for her help with the submission process.
Life in Southampton and Utrecht was very much enlivened by my dear friends who made a bigger contribution than perhaps they realize: Giuseppe Aliberti, Stefano Arnone, Stefania Beccafichi, Marc Caraway, Paolo Fagone, Antonio Gatti, Pauline Gu (all the way from Berkeley), Carl Kearney, William Kehoe, Tiziana Leone, P. K. Madhu (see you in Mumbai), Ina O'Murchu, Pierpaolo Pilla, Scott Powell, Sharon Richards, Gavin Rowe, and Niall Tucker. Julien Dubuc was a great colleague both in Southampton and Utrecht. Colin Jennings, Harald Keddes, Barbara Liesveld, Lilian Liesveld, and Roland Vernooij were good company in Southampton, Utrecht, and Amsterdam.
Rolf Boelens gave me the opportunity to work in his group in Utrecht and this was an excellent experience. I learned so much there and am grateful to him for choosing the projects, discussions, encouraging collaborations, and teaching me how to write a paper. Robert Kaptein also kindly extended my Marie Curie stipend. In addition to being a superb colleague both inside and outside of the lab, Rainer Wechselberger taught me how to operate an NMR spectrometer. I would like to thank also: Eiso AB, Daniela Beccati, Bernd van Buuren, Nocky Derix, Tammo Diercks, Cyril Dominguez, Adrien Favier, Patrick Fuchs, Albert George, Michiel Hilbers, Klaartje Houben, Danny Hsu, Rob de Jong, Henry Jonkers, Aart Nederveen, Roberto Kopke Salinas (perfect flatmate), Eugene Tischenko (we'll play pool again), Simona Tomaselli, Vincent Truffault, Hans Wienk, and Johan van der Zwan.
Tanya, for so many things.
Finally, my family. David, for encouragement and great company. To my grandmother Mary and my late grandfather Michael. Niall, Anita, and Adrian: it wouldn't have been possible without you.

Abbreviations

2D,3D	two-, three-dimensional
BMRB	BioMagResBank
c3*Da*	cytochrome c_3 from *Desulfovibrio africanus*
c3*Dg*	cytochrome c_3 from *Desulfovibrio gigas*
c3*Dn*	cytochrome c_3 from *Desulfomicrobium norvegicum*
c3*Dv*	cytochrome c_3 from *Desulfovibrio vulgaris*
COSY	correlated spectroscopy
CSA	chemical shift anisotropy
CSI	chemical shift index
DYANA	dynamics algorithm for NMR applications
EPR	electron paramagnetic resonance
HMQC	heteronuclear multiple quantum coherence
HSQC	heteronuclear single quantum coherence
HU*Bst*	HU protein from *Bacillus stearothermophilus*
HU*Bsu*	HU protein from *Bacillus subtilis*
HU*Tmar*	HU protein from *Thermotoga maritima*
lov	lower volume limit
MD	molecular dynamics
NOE	nuclear Overhauser enhancement
NOESY	NOE spectroscopy
PARADYANA	modified version of DYANA for paramagnetic proteins
PDB	Protein Data Bank
ppm	parts per million
R_1, R_2	longitudinal and transverse relaxation rate constants
RF	radio frequency
RMSD	root mean square deviation
TAD	torsion angle dynamics
TOCSY	total correlation spectroscopy
TPPI	time proportional phase incrementation
TROSY	transverse relaxation optimized spectroscopy
upl	upper volume limit

Chapter 1

Methodology

1.1 Introduction

The major advances in molecular biology research over the past half century have demonstrated that a deeper understanding of biomolecular function requires detailed knowledge of the three-dimensional structure [1, 2, 3, 4, 5]. This is more true than ever today in the post-genomic era with the availability of an abundance of data on biological systems. Such massive amounts of data require new methods of evaluation if useful knowledge is to be extracted. Approaches such as proteomics (the simultaneous study of all proteins encoded by the genome of an organism) and structural genomics (the systematic study of protein structures in order to identify a complete set of folds) initially generated intense attention in the structural biology community [6, 7]. However hypothesis-driven research remains the focus of the majority of structure determination projects [8].

During the past two decades NMR has emerged as an alternative to crystallography for structure determination. This new area of research was made possible by advances in two-dimensional (2D) [1] NMR which made it possible to assign spectra recorded from samples of small proteins. Pioneering work was performed in the laboratories of Bax, Clore, and Gronenborn at the National Institutes of Health; Boelens and Kaptein in Groningen and later Utrecht; and Ernst and Wüthrich at the ETH Zürich. Ernst and Wüthrich have both subsequently received the Nobel Prize (in 1991 and 2002 respectively) for their roles in the development and application of NMR to biological systems [9, 10].

Once these initial successes were achieved attention was turned to refining and extending the methodology. Two-dimensional techniques are only effective for proteins with a maximum molecular weight of approximately 10kDa. Spectra recorded from samples of proteins with larger molecular weights usually exhibit many overlapped peaks and increased linewidth which makes the assignment process very difficult or impossible. Advances in isotope labelling (including deuteration) and pulse sequence development made the assignment and structure determination of larger proteins and nucleic acid fragments possible. A new method was developed

[1] Abbreviations page viii

which combined the relative simplicity of multidimensional spectra with unambiguous assignments based on scalar couplings. The development of labelling methods also opened up the study of biomolecular dynamics on a much more comprehensive scale than before. During this time spectrometer technology also advanced and today the major centres have instruments operating at proton frequencies of 900 MHz. The parallel development of cryoprobes will allow recording time to be dramatically reduced. Another important development is the TROSY based pulse sequences which have extended the size limit of systems which can be studied [11, 12]. The optimal frequency for TROSY is calculated to be 1GHz and so development of spectrometer hardware will continue to explore this frontier. The study of membrane proteins in functional contexts is one of the new applications now accessible with newly developed NMR methods [13, 14]. Ideas from liquid crystal science have also been extended to biomolecular systems and the measurement of residual dipolar couplings in alignment media has provided access to a new class of information previously not available in liquid state biomolecular NMR [15, 16]. In addition to these advances in experimental techniques there have also been advances in the methodology of structure calculation. In some favourable cases completely automated structure calculations may one day be performed. The structural genomics projects hope to combine all of these advances to make NMR an essential technique in biomolecular structure and dynamics research.

The aims of this thesis were to study and understand the behaviour in solution of two very different proteins. The first protein, cytochrome c_3 from the bacterium *Desulfomicrobium norvegicum* (c3*Dn*), is an important protein in the electron transfer pathway in bacteria. Since c3*Dn* exists in two states, oxidized and reduced, during different stages of electron transfer it would be informative to compare these as accurately as possible. The results presented here focus on the structure of oxidized c3*Dn* in solution. Although a crystal structure of the protein in this form is available it was still necessary to determine the structure in solution in order to eventually compare with the structure of the reduced form which can only be determined in solution. This will allow a comparison of two solution structures rather than comparing a crystal structure with a solution structure. The second protein studied was the HU protein from the thermophilic bacterium *Thermotoga maritima* (HU*Tmar*). This protein is extremely stable and a model system for studies of

thermostability. A crystal structure of HU*Tmar* is available but is lacking information on the very mobile DNA-binding region. The aim was therefore to gain information on the behaviour of the protein in solution as compared to the crystal and to use this information to understand the results of earlier biophysical studies on mutant proteins.

1.2 Biomolecular NMR

The general protocol for studying the solution structure and dynamics of proteins and nucleic acids by NMR techniques is now well established [17, 18, 19, 20]. Since this thesis presents work on protein samples only the discussion will be mainly concerned with protein samples but the principles are the same for nucleic acid fragments. There are several stages required and each one presents potential problems which determine if a successful outcome is possible.

The first stage is to produce a sample either by purifying a protein from natural or recombinant sources or chemical synthesis or *in vitro* transcription in the case of nucleic acid fragments. Sample preparation and evaluation is perhaps the most critical stage of the whole procedure. The sample must be purified to homogeneity, highly concentrated, and stable over a reasonable time period. In addition the conditions must emulate as closely as possible the actual physiological conditions *in vivo*. Unfortunately, these stringent requirements are often incompatible for concentrated protein solutions and the optimization of conditions can be time consuming.

Isotopically labelled samples are now routinely used in biomolecular NMR. The main advantages of labelled samples are the unambiguous sequential assignment; larger systems are amenable to analysis and the relatively straighforward study of dynamic processes. The upper size limit for successful studies of proteins depends on the level of labelling used ranging from appoximately: 10kDa for non-labelled samples; 15kDa for ^{15}N-labelling; 25-30kDa for ^{13}C-^{15}N-labelling and 40kDa and upwards for ^{2}H-^{13}C-^{15}N-labelling which can be used in combination with site-specific labelling schemes. There are a variety of methods to produce isotopically enriched protein samples for NMR analysis [21, 22, 23]. The most effective methods use an organism

that overexpresses the protein of interest. This can be achieved using biotechnological procedures utilizing bacterial, yeast, insect, or mammalian cells. The general aim for structural studies is to achieve a high level of uniform enrichment, i.e. a situation in which all possible nuclei in the protein are replaced by NMR active isotopes to a maximum level throughout the macroscopic sample. Once a suitable sample has been prepared NMR spectra can be recorded and the process of resonance assignment can start. The following sections outline how data are recorded, assigned, and converted into structural or dynamical information for further analysis.

1.3 NMR spectroscopy

This section contains an overview of the various experimental methods which can be used to record spectra. The basic principles of multi-dimensional NMR are outlined followed by sections which introduce the different methods used to record spectra from labelled and unlabelled samples. There are many excellent books and reviews available on this subject [17, 18, 24, 25, 26, 27, 28, 29, 30].

1.3.1 Basic principles of multidimensional NMR

The most simple multidimensional NMR experiments are two-dimensional. The basic scheme for a two-dimensional NMR experiment consists of four sequential time periods. The preparation period is required to establish thermal equilibrium. This is followed by one or more radio frequency (RF) pulses and the evolution period (t_1). The evolution time is the period during which magnetization evolves. The mixing period begins with one or more pulses followed by an optional delay. Magnetization is transferred between spins during this period. The mixing period ends with one or more pulses and is followed by the detection period (t_2) during which the signal is recorded. The signal recorded in real time (the FID or free induction decay) is therefore modulated independently by interactions occuring during the evolution and detection periods yielding a 2D data matrix. This matrix is Fourier transformed in both dimensions to give a 2D spectrum.

Multidimensional experiments can be designed by concatenating two or more 2D experiments. A 2D pulse sequence can easily be extended into further dimensions by introducing extra evolution periods. An extra evolution period can have a signal which is modulated by either a homonucleus or (as is usually the case) a heteronucleus. Another feature is that scalar and dipolar interactions can be combined in the same pulse sequence. For example, a 3D experiment can be derived from a combination of two 2D experiments. The resulting 3D experiment contains three independent time periods. The FID recorded during the acquisition time t_3 would then be modulated by two evolution periods, t_1 and t_2. Fourier transformation of this 3D time domain data matrix yields a 3D frequency domain spectrum. Such a spectrum can be represented as a cube but it is more convenient to analyze 2D planes or slices.

1.3.2 Homonuclear and heteronuclear correlation

The first 2D studies on protein samples mainly used homonuclear correlation pulse sequences since this was the easiest way to obtain resonance assignments and structural restraints. The correlation of resonance frequencies can be obtained by coherence transfer mediated by *through-bond* (scalar) or *through-space* (dipolar) interactions. Heteronuclear 2D spectra are also extremely useful in certain cases, for example, resonance assignments in paramagnetic systems or backbone dynamics studies. The basic method is outlined for the example of heteronuclear coherence transfer used in the HSQC experiment and can be generalized for different experiments.

The HSQC (heteronuclear single-quantum correlation) experiment is widely used in biomolecular NMR and is a good example of an experiment in which coherence is transferred by scalar interactions. The basic pulse sequence for the HSQC experiment performed on a heteronuclear ($I = {}^{15}\mathrm{N}$, $S = {}^{1}\mathrm{H}$) spin system can be written as:

$$\begin{aligned} S:\ & (\pi/2)-\tau-(\pi)-\tau-(\pi/2)_y-\tfrac{t_1}{2}-(\pi)-\tfrac{t_1}{2}-(\pi/2)-\tau-(\pi)-\tau-acq. \\ I:\ & \qquad\qquad (\pi)-\tau-(\pi/2)\ -\tfrac{t_1}{2}-(\pi)-\tfrac{t_1}{2}-(\pi/2)-\tau-(\pi)-\tau-dec. \end{aligned} \tag{1.1}$$

where the terms in brackets indicate $\pi/2$ and π RF pulses; $\tau = 1/(4J_{IS})$, and t_1 is the chemical shift evolution time for the I spins. The pulse sequence proceeds with the following events: transfer of coherence from protons to the heteronucleus; chemical shift evolution of heteronuclear magnetization; transfer back to protons for signal acquisition while decoupling the heteronucleus.

The HSQC pulse sequence is used in many 3D heteronuclear-edited experiments, for example, to separate overlapped 2D homonuclear correlation spectra according to the resonance frequencies of the heteronuclei. The most widely used experiments in this category are the complementary TOCSY-HSQC and NOESY-HSQC.

1.3.3 Triple resonance experiments

The use of triple resonance pulse sequences and doubly-labelled samples has now become the standard method for reliable resonance assignment of proteins which can be expressed in the required yields. This category of experiments result in 3D spectra with resonance frequencies for three different nuclei along the orthogonal axes of the data matrix. Detailed discussions of triple resonance NMR as applied to biomolecules are given in the book by Cavanagh *et al.* and the reviews by Wider and by Sattler *et al.* [18, 28, 29].

The triple resonance pulse sequences have become quite advanced when compared with earlier 2D and 3D experiments. An example of a complex sequence is shown in Figure 1.1 which presents the pulse sequence for the COCCH-TOCSY experiment. This pulse sequence has typical features of many modern triple resonance experiments.

The pulse sequence is preceded by presaturation and homospoil pulses to destroy any signal from the water resonance. The first pulse of the coherence transfer is in the middle of the carbonyl (^{13}CO) region. Chemical shift evolution of the ^{13}CO resonances then takes place. Another chemical shift evolution period immediately follows and the ^{13}Cα resonances are frequency-labelled. At the end of this second evolution period longitudinal ^{13}Cα magnetization is created and mixed throughout

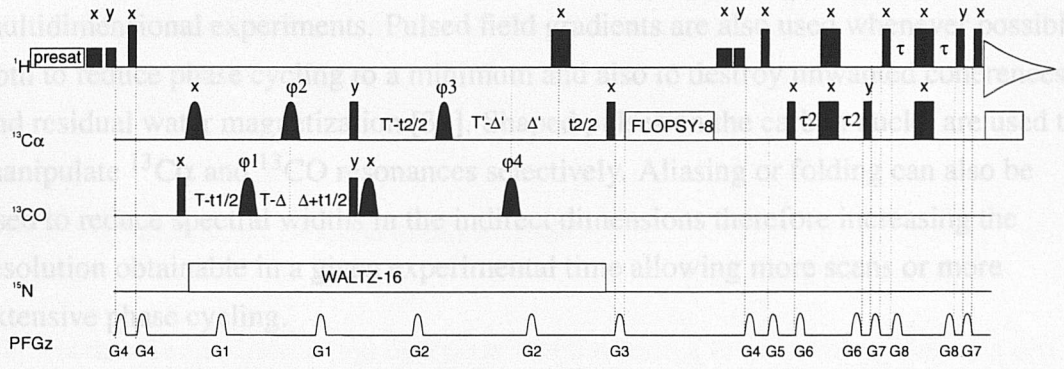

Figure 1.1: Pulse sequence of the COCCH-TOCSY experiment. Narrow and broad rectangles denote rectangular non-selective $\pi/2$ and π pulses, respectively; the remaining pulses were selective π pulses shaped like the central lobe of a $\sin(x)/x$ function and 100 μs (Cα) or 200 μs (CO) long. The following delays are used: T=8.7 ms; Δ=4.5 ms; T$'$=3.83 ms; Δ'=3.83 ms; τ=1.7 ms; τ2=1.0 ms. A FLOPSY-8 mixing sequence was used. The pulse lengths and maximum field strengths of the gradients were G1=0.5 ms, 20 G/cm; G2=0.5 ms, 15 G/cm; G3=0.5 ms, 31 G/cm; G4=4.0 ms, 30 G/cm; G5=3.0 ms, 30 G/cm; G6=0.5 ms, 3 G/cm; G7=2.0 ms, 25 G/cm; G8=0.5 ms, 7 G/cm. All gradients were of the same sign. Phase cycling was done as follows: φ1 : x,y ; ϕ2 : 8x, 8(-x) : ϕ3: 2x,2y : φ4: 4y, 4(-y) ; acquisition x,-x,-x,x. The first Cα $\pi/2$ pulse and the last Cα $\pi/2$ pulse before the mixing sequence were phase cycled according to the TPPI protocol. Figure courtesy of Eiso AB [31].

the spin systems using the FLOPSY broadband isotropic mixing scheme to cover the wide frequency range of the Cα spins. After the mixing period the magnetization is transferred back to protons and after a refocussed INEPT stage to generate in-phase signal acquisition is started while decoupling the ^{13}Cα spins. The resulting 3D spectrum therefore has two ^{13}C dimensions and one ^{1}H dimension and correlates all protons in a spin system with the ^{13}Cα and ^{13}CO resonances of the intra-residue spins.

1.3.4 Technical tips and tricks

The pulse sequence in Figure 1.1 also illustrates some of the main technical aspects which are used to obtain maximum S/N (signal-to-noise) ratio in triple resonance experiments. For this purpose the ^{1}H nucleus is always detected in all

multidimensional experiments. Pulsed field gradients are also used whenever possible both to reduce phase cycling to a minimum and also to destroy unwanted coherences and residual water magnetization [32]. Shaped pulses on the carbon nuclei are used to manipulate $^{13}C\alpha$ and ^{13}CO resonances selectively. Aliasing or folding can also be used to reduce spectral widths in the indirect dimensions therefore increasing the resolution obtainable in a given experimental time allowing more scans or more extensive phase cycling.

1.4 Assignment strategy

Once the required set of spectra has been recorded the resonance assignment process can begin. The strategy deployed for this stage depends on the isotope labelling of the protein sample. For unlabelled and ^{15}N-labelled samples the traditional method of sequential assignment based on NOE through-space interactions must be followed. This strategy can be successfully applied to relatively large proteins, especially in the case of ^{15}N-labelling. For larger proteins which can be recombinantly expressed to produce doubly-labelled samples the newer method of triple resonance through-bond assignment can be used. All assignment methods require a knowledge of typical chemical shifts observed for the various nuclei in proteins. The information contained in chemical shift data is extremely informative both in amino acid identification and also in correlation with various structural properties.

The chemical shift index (CSI) method can be used to qualitatively determine the secondary structure elements in an assignment sequence by comparison with random coil chemical shifts [33, 34]. Chemical shift referencing is therefore very important both for correlation with structural features and reproducibility. The usual procedure is to directly reference the proton chemical shifts to an internal standard and to indirectly reference the heteronuclei [35].

The first subsection outlines how chemical shifts in proteins can be assigned. The prediction of chemical shifts for protons in biomolecules can be a very powerful indication of structural features especially when effects due to aromatic ring currents

or paramagnetic centres are considered. The following subsections describe how protein sequences are assigned by NMR using both of the methods introduced above.

1.4.1 Prediction of chemical shifts

The chemical shift distribution of the various nuclei in a large group of proteins can be analyzed statistically to produce average values which can be used as a reference when analyzing spectra. These average values reflect the mean effect of protein secondary and tertiary structure on chemical shift values. Variations in local diamagnetic, paramagnetic and electrostatic fields can cause deviations from the average values which depend on the structure of the protein in question. Knowledge of the geometrical dependance of these shifts in model systems allows the design of reliable methods to predict the modified shifts in the protein of interest [36].

In proteins containing prosthetic groups such as haem the ring current shift must be accounted for when predicting chemical shifts. The ring current shift arises from the interaction of a nucleus and the current loop produced by a nearby aromatic group. Haem groups produce a very large ring current which causes a wide dispersion of the chemical shifts of nearby nuclei. The ring current shift can be calculated from the Pople equation [37]:

$$\Delta\delta_r = -\frac{10^6 n e^2 a^2 (1 - 3\cos^2\theta)}{6\pi m c^2 r^3} \tag{1.2}$$

where:

- n is the number of circulating electrons;
- e is the electronic charge;
- m is the electronic mass;
- a is the radius of the ring;

- r is the distance between the paramagnetic centre and the nucleus of interest;
- θ is the angle between the centre of the ring and the nucleus.

The paramagnetic shift arises from the interaction between an unpaired electron and the nucleus of interest. A bis-histidinyl ligated haem group can exist in both oxidized and reduced states. In the reduced form the iron atom is in the Fe(II) state with a low-spin d^6 orbital occupancy and the haem group is diamagnetic. In the oxidized form the iron atom is in Fe(III) low-spin d^5 state and is therefore paramagnetic.

An unpaired electron can cause paramagnetic shifts of the resonance frequencies of nearby nuclei via the contact (scalar) and pseudocontact (dipolar) interactions. The contact shift is due to the partial transfer of unpaired electron density from the iron(III) atom via covalent bonds to the nucleus in question. The contact shift can be calculated from the following equation:

$$\Delta\delta_c = -\frac{(A/h)g\beta S(S+1)}{3h\gamma_N kT} \tag{1.3}$$

where:

- A/h is the contact interaction constant (Hz) which can be positive or negative;
- g is the rotationally averaged electron g-factor;
- β is the Bohr magneton;
- S is the electronic spin;
- γ_N is the gyromagnetic ratio.

The spin 1/2 electron does not split the proton resonances however because the rate of electron spin relaxation is extremely fast. The magnitude of A depends on the percentage of the unpaired electron density residing on the nucleus of interest. For

some protons the contact shift can be much larger than the ring-current shift due to a haem group. It can also be equal but opposite in sign to the pseudocontact shift.

The pseudocontact shift is an anisotropic interaction between the magnetic moment of the unpaired electron and nearby nuclei. The pseudocontact shift is given by [38]:

$$\Delta\delta_{pc} = \frac{1}{12\pi r^3}[\Delta\chi_{ax}(3\cos^2\theta - 1) + 1.5\Delta\chi_{eq}\sin^2\theta\cos 2\phi] \qquad (1.4)$$

where:

- $\Delta\chi_{ax}$ and $\Delta\chi_{eq}$ are the effective values of the axial and equatorial anisotropies respectively;
- r is the distance from the iron atom to the resonating nucleus of interest;
- ϕ is measured from the z-axis, i.e. perpendicular to the haem plane;
- θ is measured in the haem plane.

If the equation for the pseudocontact shift is compared with the equation for the ring-current shift it can be seen that for an equivalent spatial arrangement of protons the ring-current and pseudocontact shift will be opposite in sign.

In order to predict the shifts it is necessary first to know both the magnitude and the orientation of the magnetic susceptibility tensor. However it is known that the principal axes of the tensor do not necessarily coincide with the axes of the haem. Usually the tilt of the z-axis of the tensor is small and the axis is therefore approximately perpendicular to the haem. In this case the rotation in the plane of the haem can be deduced from ^{1}H or ^{13}C chemical shifts of the haem substituents as described [39, 40]. Alternatively the tensors can also be fitted using data from dipolar couplings measured at different field strengths [41].

The use of information obtained from an analysis of dipolar shifts and the susceptibility tensor can then be used as an extra input when calculating the solution

structure of a paramagnetic protein. The program PARADYANA simultaneously calculates the structure and magnetic tensor(s) of paramagnetic proteins [42]. In addition to the normal NOE-based distance restraints the program uses dipolar shifts and the complete description of the magnetic axes. This approach improves the structural resolution obtained in the region of the haem group. The added accuracy in the haem region is very important because NOE signals from the imidazole protons of the axial ligands are never observed due to the efficient electron-induced relaxation. The information obtained in this way can be used to determine differences between a crystal structure and a solution structure and also between the two oxidation states of a solution structure.

1.4.2 Sequence-specific assignment of spin systems

This was the first method developed to assign protein NMR spectra and can also be applied to nucleic acids [17]. The essence of the method is a two-step strategy: (1) identification; and (2) sequential assignment of spin systems present in homonuclear correlation spectra (which may be heteronuclear-edited if a ^{15}N-labelled sample is available). The identification step is achieved by examining the COSY and TOCSY spectra and identifies the number and type of amino acid spin systems which give observable signals under the sample conditions. The results of this step are compared to the protein sequence and an assessment of the possibility of a complete assignment can be made. The sequential assignment stage is achieved by an exhaustive analysis of NOESY spectra to assign characteristic NOEs between adjacent residues in the sequence.

This method has an important limitation which is that care must be taken not to confuse NOEs between residues which are not adjacent in the sequence but are close in space due to the secondary and tertiary structures of the protein. NOEs due to secondary structure can be seen to fit into a consistent pattern if the local fold is in an α-helix or a β-pleated sheet. An implicit advantage of this method therefore is that the characteristic secondary structure NOEs are identified at an early stage which can be an added confirmation of the assignment in the case that a crystal structure is available and in general is very important for preliminary structure calculations. In

the course of the complete assignment of the sequence the NOEs due to tertiary structure should become obvious. If a ^{15}N-labelled sample (but not doubly-labelled) is available this method can be facilitated by recording heteronuclear-edited spectra with the advantage that extra information can be identified from the ^{15}N chemical shift data (see below). This method was mainly developed by the Wüthrich group and is described in the now classic textbook [17].

1.4.3 Triple resonance assignment strategy

The starting point for an assignment using this method is the ^{1}H-^{15}N HSQC spectrum. In principle this spectrum should have one correlation for each non-proline residue in the sequence. In addition there will be a number of correlations in which two amide protons have the same ^{15}N shift. These correlations are readily identified as NH_2 side-chain groups. An initial assessment of the distribution of the peaks in the HSQC spectrum can indicate typical positions for various types of residues as well as the probable secondary structure of the protein. Proteins with β-sheet secondary structures have larger chemical shift dispersion for all backbone nuclei and this is especially pronounced for amide protons. A recommended precaution is to record HSQC spectra at regular intervals throughout a lengthy period during which triple-resonance spectra are recorded. This is in order to continually assess the stability of the sample. Therefore in many ways the HSQC spectrum is a very informative fingerprint of the protein under investigation.

The number of expected peaks can of course differ from the number counted in the HSQC spectrum. The number can be smaller due to a few different factors: overlap; very fast amide proton exchange with the aqueous solvent; conformational broadening which is at a intermediate rate on the NMR time scale. The number can be larger due to two or more protein conformers being significantly populated under the experimental conditions. The conformers may or may not have equal populations. If the sample is a complex or a oligomer equilibria may be observed between protein and ligand or between monomer and multimer or between different conformers. This situation can complicate the analysis considerably and was in fact the case for HU*Tmar*.

In addition it is wise to record HSQC spectra over a relatively wide temperature range to check if some overlapped peaks move to different positions and if the overall chemical shift dispersion improves. In general working at a higher temperature will ensure narrower lines in 3D spectra but the choice of temperature must also be safe for the sample. Many non-native proteins degrade very quickly and the basic goal is to record the required data as quickly as possible. For especially unstable or dilute samples access to a spectrometer equipped with a cryoprobe is a major advantage.

The next stage is the assignment of the backbone resonances using the standard set of triple-resonance experiments recorded with a doubly-labelled sample:

- HNCA and HN(CO)CA;
- HNCO and HN(CA)CO;
- CBCA(CO)NH, HBHA(CO)NH, and HNCACB.

This set of experiments correlate backbone and/or sidechain resonances for nuclei of an amino acid and also the preceding amino acid in the sequence. The important factor that makes this approach possible is the large one-bond coupling constants observed for backbone resonances.

In addition ^{15}N TOCSY-HSQC and ^{15}N NOESY-HSQC can be recorded with either a singly or doubly-labelled sample to facilitate the analysis. These experiments can unambiguously assign fragments consisting of two or more amino acids which can usually be assigned to specific residues in the sequence. Different amino acids have characteristic chemical shift values for their backbone resonances and this knowledge combined usually makes an assignment completely unambiguous. Programs are available which can attempt to perform the assignment in an automated or semi-automated manner but care must be taken in peak picking routines and it is in fact wise to do this by hand. The set of triple-resonance spectra allow the compilation of a set of backbone chemical shifts for residues i and $i-1$. Sometimes correlations can be absent from the spectra, especially in less sensitive experiments but manual inspection can often solve the apparent ambiguity.

The sequence of the protein to be assigned is first checked for fragments which will assist or complicate the assignment. Since all of the experiments in the set involve detection of backbone amide protons there will be breaks in the assignment at proline residues. The sequence is then checked for unique fragments of two or more residues. Almost all amino acids have a unique pattern of backbone and β-carbon chemical shifts which will allow the amino acid type to be easily determined. In this way the non-specifically assigned fragments can gradually be specifically assigned to residues in the sequence as progressively more information is obtained from the different spectra. Any assignments of fragments of different backbone conformations should eventually lead back to a common starting point. As an example, a selection of the sequence-specific assignments obtained for the IIB^{Chb} protein is shown in Figure 1.2. The HNCA spectrum correlates the Cα resonances of residue i and $i-1$ with the ^{15}N and NH resonances of residue i. The HN(CO)CA provides the sequential information only by correlating the Cα resonance of residue $i-1$ with the ^{15}N and NH resonances of residue i. The information from these two spectra can be used with the CBCA(CO)NH spectrum to assign the majority of backbone resonances. In cases of overlap or missing correlations the HNCO and HN(CA)CO spectra can help complete the assignment.

The next stage is the assignment of the side-chain resonances. The backbone Cα and sidechain Cβ frequencies can be used as starting points in the search. The assignment is achieved using a combination of spectra to assign different types of signals. The majority of side-chain resonances involve carbon-carbon correlations and therefore can be assigned from COSY and TOCSY type experiments. The first step is to record a HCCH-COSY experiment which will mainly be used as an aid in assigning the HCCH-TOCSY spectra in the same way that a 2D COSY spectrum can help in the analysis of a 2D TOCSY spectrum by differentiating signals in certain residues.

Several different variants of the HCCH-TOCSY pulse sequence can be used to record 3D spectra with either two proton or two carbon dimensions to assign both nuclei in the side-chains. The hCCH-TOCSY has two carbon dimensions and is used to assign the sidechain ^{13}C resonances. Similarly the HcCH-TOCSY has two proton dimensions and is used to assign the sidechain proton resonances.

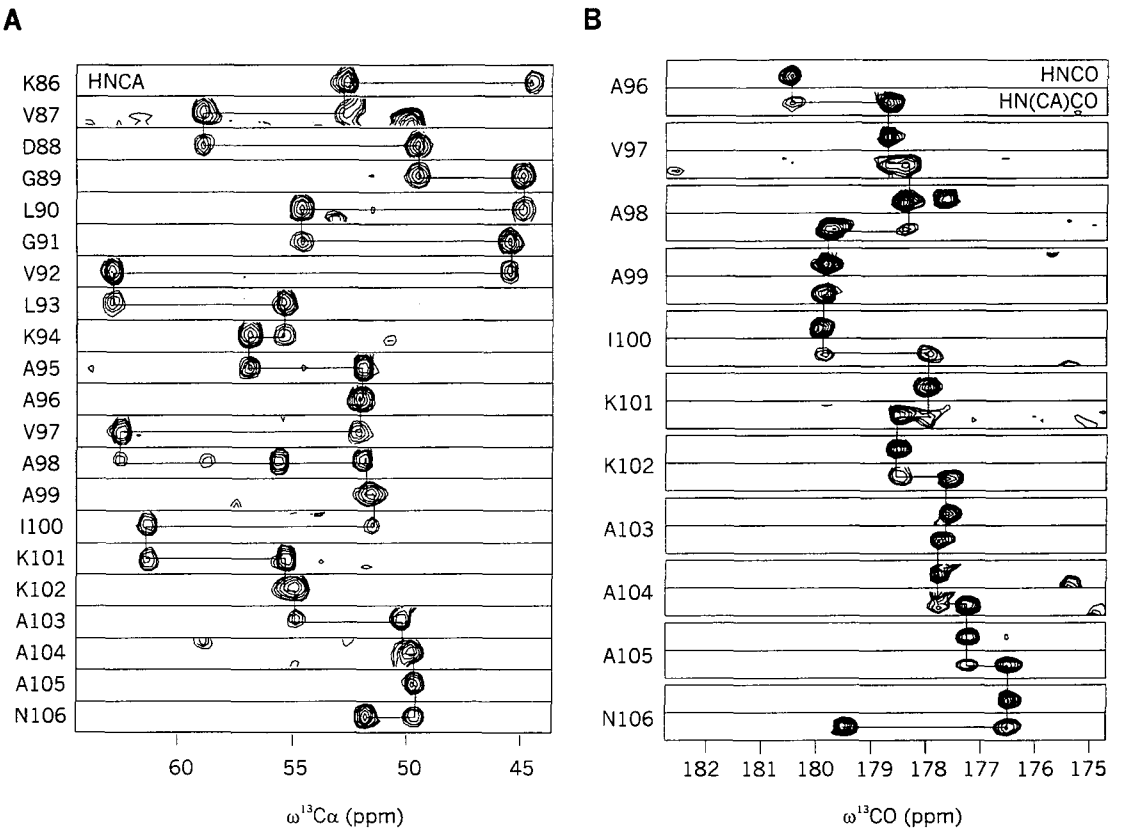

Figure 1.2: (A) Slices from the HNCA spectrum for the 21 C-terminal residues of IIB^{Chb} taken at their corresponding ^{1}H-^{15}N frequencies. (B) Alternating slices of the HN(CA)CO and HNCO for the 11 C-terminal residues of IIB^{Chb}. For each residue two slices are shown taken at the same ^{1}H-^{15}N frequencies; the upper slice is from the HNCO for each residue and the lower slice is from the HN(CA)CO. Figure courtesy of Eiso AB [31].

1.5 Relaxation and protein backbone dynamics

The structures of proteins and other biomolecules determined by crystallography and NMR do not usually provide direct information on the dynamic nature of the conformational changes which take place in solution. NMR is a very powerful technique for probing the dynamics of proteins over a wide range of timescales and has been comprehensively reviewed by several authors
[43, 44, 45, 46, 47, 48, 49, 50, 51].

Nuclear spin relaxation is the process by which the spins in the ensemble return to Boltzmann equilibrium with the surroundings. The relaxation rates reflect the physical properties of the molecule and therefore can be used to extract information about the molecular dynamics. Relaxation is mediated by local random magnetic fields created by rotational molecular motion. These random fields cause transitions between energy levels in the ensemble. The random time-dependent amplitude and direction of these local fields is exactly what drives the ensemble to equilibrium. Since molecular motion in solution occurs at frequencies similar in magnitude to those required for coupling of the spins to the lattice the relaxation rates are very sensitive to this. The lattice is assumed to have an infinite heat capacity and consequently to be at thermal equilibrium at all times. Spin-lattice relaxation is the mechanism by which the Boltzmann population distribution is re-established after a perturbation such as a pulse. Spin-spin relaxation occurs as coherences decay. These two relaxation processes have different rate constants which can be measured experimentally.

The two most widely used concepts in explanations of relaxation phenomona are the correlation function and the spectral density. The correlation function is basically a description of the magnitude and rate of the randomly fluctuating fields which cause relaxation. Since these random fields are modulated by the tumbling of the molecule in solution (and therefore to its molecular weight and approximate shape) the correlation function can be thought of as determining the relaxation parameters characteristic of the molecule under a given set of experimental conditions (such as temperature). Therefore fast fluctuations will result in a rapidly decaying correlation function while slower fluctuations lead to a slower decay. The correlation time for the molecular tumbling is inversely proportional to the rate of the fluctuations and is typically in the ps-ns range for molecules in solution.

The spectral density is defined as the Fourier transform of the correlation function in a similar way as a spectrum is the transformed counterpart of an FID. If the correlation time is short the fluctuations are fast and the spectral density function is broad. Conversely, a long correlation time gives rises to slower fluctuations resulting in a narrow spectral density function. The area under the spectral density curve remains the same for the molecular system under consideration and therefore any

changes in the correlation time give rise to differing contributions to the relaxation at different frequencies. Similarly, if the experimental conditions are preserved and the molecular system is changed the effects of differences in correlation times can be examined. The spectral density is perhaps the key concept in understanding relaxation phenomona (especially in the qualitative sense presented here) and is central to understanding more sophisticated models and for the interpretation of data measured for biomolecules.

The following subsections expand on the use of these concepts in the interpretation of relaxation data measured with protein samples. There are to date two methods well-established in the literature, namely: spectral density mapping and the model-free formalism. The model-free formalism was developed first by Lipari and Szabo with later input from Clore and co-workers. The spectral density mapping method was developed by Lefèvre, Peng, and Wagner and co-workers with additional independent input from Ishima and Nagayama and also from Kay and co-workers. The model-free method is more widely used to date yet it does have some important limitations. The major requirement for successful use of the model-free method is an accurate definition of the molecular diffusion tensor. An additional limitation of the model-free method is that the derived order parameters report only on processes which occur on a timescale of nanoseconds or faster. This means that motions on a timescale comparable to or slower than the molecular tumbling cannot be described accurately. Spectral density mapping overcomes both of these limitations because: firstly, the spectral densities at the various frequencies can be obtained directly from the measured relaxation rates without recourse to any models describing the molecular tumbling, and; secondly, the derived spectral densities are of course sensitive to motions occuring at all frequencies which can induce relaxation. This means that spectral density mapping can be applied to unfolded states or other systems with dynamics that occur on unusually wide timescale ranges.

The interactions which produce spectral density at the relevant frequencies can be produced by both internal and overall motions of the molecule [51]. An accurate analysis must therefore account for both types of motion and any possible correlation between them. The model free formalism has an explicit assumption that there is no correlation between the two types of motion provided that the internal motions are

small which is generally the case. For non-folded or partially folded proteins which share characteristics of polymers in solution this separability of internal and overall motions is no longer valid and the spectral density mapping approach may yield a more accurate interpretation of the dynamics. While spectral density mapping does not require separability to produce results care must be taken in the interpretation of the motions and timescales which are assumed to occur in the molecule.

1.5.1 Relaxation mechanisms

For spin 1/2 nuclei only the dipole-dipole and chemical shift anisotropy (CSA) mechanisms are significant in solution. Excellent introductions to the subject of relaxation are given in the books by Cavanagh *et al.* and Levitt [18, 30].

The Fourier transform of a simple exponentially decaying autocorrelation function is the normalized spectral density function which for the isotropic diffusion of a rigid molecule with correlation time τ_c is:

$$J(\omega) = \frac{\tau_c}{1 + \omega^2 \tau_c^2} \tag{1.5}$$

where ω is the frequency of the fluctuations. The general form of the spectral density function is relatively constant while $\omega^2\tau_c^2 \ll 1$ and changes at $\omega^2\tau_c^2 \approx 1$. For rapid molecular motion due to short correlation times the *extreme narrowing* condition of $\omega^2\tau_c^2 \ll 1$ is reached and the spectral density function is broad and the values are relatively constant over the frequency range. In contrast, for slow tumbling the *spin diffusion* limit of $\omega^2\tau_c^2 \gg 1$ is reached and the spectral density function is narrow and peaked at zero frequency.

This basic form of the spectral density function (Equation 1.5) describes the contributions to relaxation at different frequencies for a given set of conditions. The next stage is to connect this function with the transitions possible in a given spin system. The example is a dipolar-coupled two-spin (homonuclear or heteronuclear)

system for which the Solomon equations for relaxation are presented. The two-spin system has characteristic transitions which can be related to both the Solomon equations and the spectral density function to explain different relaxation effects.

The magnitude of the dipolar coupling interaction depends on the orientation of the internuclear vector with respect to the main magnetic field and also on the inverse cube of the internuclear distance. As the molecule tumbles in solution both the orientation and the size of the vector can change. This time-dependent variation produces local magnetic fields and sufficient spectral density is produced at frequencies corresponding *either* to the overall molecular tumbling *and/or* internal motions at the relevant timescales. The rate is proportional to the square of the coupling. In solution the actual dipolar coupling is averaged to zero and is not observed in the spectra of non-aligned systems.

For an isolated covalent two-spin homonuclear system in an idealized rigid molecule there are four energy levels and twelve different transitions of which eight are single-quantum, two are double-quantum and the remaining two are zero-quantum. The transition probabilities between the different energy levels can be used to derive the Solomon equations which describe the rates of change for the longitudinal magnetization of the two spins in terms of the relaxation rate constants. For a general system of N spins the Solomon equations can be written in terms of the auto- (ρ_k) and cross-relaxation (σ_{kj}) rate constants:

$$\frac{d\Delta M(t)}{dt} = -\rho_k \Delta M(t) - \sum_{j \neq k} \sigma_{kj} \Delta M(t), \tag{1.6}$$

where ρ_k is the auto-relaxtion rate constant for spin k due to all other spins and the cross-relaxation rate constant $\sigma_{kj} = W_2 - W_0$. The cross-relaxation rate is non-zero only if $W_2 - W_0 \neq 0$ which requires that W_2 and/or W_3 must be non-zero which is the case for biological macromolecules.

Theoretical calculations can relate the transition probabilities to the values of the spectral density function at different frequencies:

$$W_1 = (3/20)b^2J(\omega_0);\ W_2 = (3/5)b^2J(2\omega_0);\ W_0 = (1/10)b^2J(0); \qquad (1.7)$$

where W_1, W_2, and W_0 are, respectively, the single-, double-, and zero-quantum transition probabilities, J is the spectral density at zero and the Larmor (ω_0) frequencies, and

$$b = -\frac{\mu_0 h\gamma^2}{8\pi^2 r^3} \qquad (1.8)$$

is the homonuclear dipole-dipole coupling constant. For short correlation times the spectral density function is very flat and therefore the relative order of these probabilities depends mainly on the numerical factors and can be summarized as: $W_2 > W_1 > W_0$. For long correlation times the spectral density function is narrow and symmetrically peaked at zero frequency and the relative order of the probabilities is dominated simply by the shape of the function: $W_0 > W_1 > W_2$.

The NOE is a spin relaxation phenomenon which varies as a function of the product of the correlation time of the molecule and the Larmor frequency at which the NMR data is recorded. For small molecules τ_c is short relative to ω_0^{-1} and $\omega_0\tau_c \ll 1$. In this extreme motional narrowing regime the rotational motions can efficiently couple with the energy-dissipating nuclear spin transitions. In contrast for biomolecules the motional frequencies are too low and $\omega_0\tau_c \gg 1$. In this slow tumbling regime energy conserving cross-relaxation transitions are favoured. The most important NOE measured in biomolecular studies is the proton-proton NOE which has a maximum value of $+0.5$ for extreme narrowing and a maximum of -1.0 when the molecular tumbling is slow. The NOE is in any case dependent on the gyromagnetic ratio of the interacting nuclear species and on the field strength at which the measurements are made.

The basic spectral density function (Equation 1.5) can be used to calculate the cross-relaxation rate:

$$\sigma_{IS}^{NOE} = \frac{h^2\mu_0^2\gamma^4\tau_c}{20\pi^3 r_{IS}^6}\left[-1+\frac{6}{1+4\omega_0^2\tau_c^2}\right]; \tag{1.9}$$

and autorelaxation rate constants for a homonuclear spin system:

$$\rho_{IS} = \sum_{j\neq i}^{n} \frac{h^2\mu_0^2\gamma^4\tau_c}{20\pi^3 r_{IS}^6}\left[1+\frac{3\tau_c}{1+\omega_0^2\tau_c^2}+\frac{6\tau_c}{1+4\omega_0^2\tau_c^2}\right]. \tag{1.10}$$

The NOE competes with other mechanisms for spin relaxation, for example paramagnetic relaxation. Spin diffusion is also a very important phenomenon resulting from relayed NOE effects particularly in biomolecules where there are many spins close to the spin of interest. Even with these limitations accurate measurements are possible under the correct experimental conditions. According to theory it can be shown that in a homonuclear system the initial rate of NOE buildup on an object spin due to irradiation of a source spin develops as:

$$\frac{dM_2}{dt} = -\rho_2 M_2 - \sum_{j\neq 2} \sigma_{2j} M_j \tag{1.11}$$

where M_2 is the difference of the object spin magnetization from its initial value and M_j represents the same quantity for the other j spins in the molecule. The quantites ρ_2 and σ_{2j} are the spin lattice and cross relaxation rates respectively. At time $t = 0$, $M_2 = 0$ and all $M_j = 0$ except the magnetization on the source spin M_1. The equation then reduces to:

$$\left[\frac{dM_2}{dt}\right]_{t=0} = -\sigma_{21} M_1, \tag{1.12}$$

and the initial slope of the NOE buildup curve is equal to the cross relaxation rate σ_{21}. Using this model NOE buildup curves can be measured for protons in biomolecules. In most practical situations it is only possible to record a single

NOESY spectrum with a mixing time short enough that the direct cross relaxation should be dominant and spin diffusion can be neglected. The intensity of the NOE between two nuclei is given by:

$$V = < r^{-6} > f(\tau_c). \tag{1.13}$$

The averaging indicates that r may vary due to flexibility within the molecule. The function τ_c can be assumed to be constant which would imply that the protein is relatively rigid and the correlation time in that case would be the same for all regions of the molecule. This is of course a gross assumption but it is used anyway to make the process of structure calculation possible. The second major assumption in using the above equation is that it is valid only for an isolated pair of nuclear spins. Magnetization can also be transferred indirectly via other nearby nuclei, a phenomenon known as spin diffusion. Because of spin diffusion the isolated spin pair approximation is only valid for NOESY acquired with very short mixing times. However if the experimental mixing time is too short the NOE intensity becomes proportional to the mixing time. Usually experiments acquired with a mixing time of 80-100ms achieve a balance between these two opposing factors.

The Solomon equations also predict the steady-state nuclear Overhauser effect. This effect occurs when a RF field is continuously applied at the resonance frequency of one of the partners in a coupled two-spin (IS) system and has an enhancing effect on the magnetization of the second spin. This phenomenon is very important in studies of protein dynamics where the heteronuclear steady state NOE enhancement of the ^{15}N magnetization due to the irradiation of the amide protons reports directly on the local dynamics and provides input for spectral density mapping and model-free analyses. The transition probabilities are very similar to those for the homonuclear system except for different contributions at the resonance frequency of the heteronucleus and the dipolar coupling constant:

$$\begin{aligned} &W_1^I = (3/20)b_{IS}^2 J(\omega_I^0);\ W_1^S = (3/20)b_{IS}^2 J(\omega_S^0); \\ &W_2 = (3/5)b_{IS}^2 J(\omega_I^0 + \omega_S^0);\ W_0 = (1/10)b_{IS}^2 J(\omega_I^0 - \omega_S^0); \end{aligned} \tag{1.14}$$

where W_1, W_2, and W_0 are the single-, double-, and zero-quantum transtions respectively, J is the spectral density at the two Larmor frequencies ω_I and ω_S, and the heteronuclear dipolar coupling constant is:

$$b_{IS} = -\frac{\mu_0 h \gamma_I \gamma_S}{8\pi^2 r^3}. \tag{1.15}$$

A measurement of the heteronuclear NOE therefore samples the spectral density at four frequencies. The derivation of the heteronuclear NOE yields the following expression [52]:

$$NOE = 1 + \left[(\gamma_I/\gamma_S) d^2 \left[6J(\omega_I + \omega_S) - J(\omega_I - \omega_S)\right] T_1\right] \tag{1.16}$$

with;

$$d^2 = 0.1 \gamma_I^2 \gamma_S^2 h^2/(4\pi^2) < 1/r_{IS}^3 >^2 . \tag{1.17}$$

In these expressions $I = {}^{15}\mathrm{N}$, $S = {}^{1}\mathrm{H}$, r_{IS} is the internuclear distance, H_0 is the static field strength, $\sigma_{\|}$ and $\sigma_{\perp}$ are the parallel and perpendicular components of the axially symmetric chemical shift tensor and the other symbols have the same meanings as defined above. The r_{IS} bond length is usually set to 1.02 Å according to results from neutron diffraction [53] and the chemical shift anisotropy has been measured to be: $\sigma_{\|} - \sigma_{\perp} = -160$ ppm [54].

The longitudinal (R_1) and transversal (R_2) relaxation rates can be derived using the same theoretical ideas and the following expressions are obtained:

$$R_1 = d^2 \left[J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)\right] + c^2 J(\omega_S); \tag{1.18}$$

$$R_2 = \begin{aligned} & 0.5d^2[4J(0)+J(\omega_I-\omega_S)+3J(\omega_S)+6J(\omega_I)+6J(\omega_I+\omega_S)] \\ & +(1/6)c^2[3J(\omega_S)+4J(0)]; \end{aligned} \quad (1.19)$$

where:

$$c^2 = (2/15)\gamma_S^2 H_0^2(\sigma_{\|}-\sigma_{\perp})^2. \quad (1.20)$$

The dipolar mechanism can also operate between electrons and nuclei in the case of paramagnetic relaxation as mentioned below in the context of relaxation matrix calculations.

In solution state NMR nuclei experience an average chemical shift which is the result of rapid molecular tumbling. Chemical shift anisotropy (CSA) arises because local fields modulated by tumbling and varying in three dimensions can give rise to relaxation. This is distinct from the dipolar mechanism in that the CSA does not cause changes in populations of energy levels since it is caused by fields parallel to rather than perpendicular to the main static field.

Relaxtion induced through CSA is proportional to the shift anisotropy (chemical shift range) and the square of the field strength. Accordingly, CSA relaxation is expected to be significant for nuclei with large shift ranges observed at high field, for example ^{13}C or ^{31}P.

1.5.2 Spectral density mapping

The spectral density mapping approach uses the expressions for the heteronuclear NOE and the R_1 and R_2 relaxation rates to directly determine the values of the spectral density function at the relevant frequencies independently of any model. The measured rates are used as input to matrix calculations to determine the contributions

at the different frequencies. The results of this calculation can therefore give direct insight into the motional properties along the protein backbone.

The original procedure for spectral density mapping involved determining the values for the spectral density function at five frequencies: $0, \omega_S, \omega_{I-S}, \omega_I$, and ω_{I+S}. The authors proposed measuring six relaxation rates and solving a set of linear equations to find values for the spectral densities at the five frequencies and the dipolar relaxation rate for the amide proton [43, 44]. Practical problems in the form of negative values for the spectral densities at high frequencies for large proteins and/or high field strengths prompted the search for an improved approach.

The reduced spectral density mapping method was therefore introduced independently by different research groups. This method is based on the observation that, at high frequencies (400-600 MHz), the spectral density function is almost flat. This suggests that the values for $J(\omega)$ at ω_{I-S}, ω_I, and ω_{I+S} can be considered as equal which simplifies both the experimental measurements (since only three rates need to be measured) and the analysis. The following matrix equation relates the spectral densities to the measured relaxation rates [45]:

$$\begin{bmatrix} J(0) \\ J(\omega_S) \\ J_{avg}(\omega_I) \end{bmatrix} = \begin{bmatrix} -\frac{3}{4E} & \frac{3}{2E} & -\frac{9}{10E} \\ \frac{1}{E} & 0 & -\frac{7}{5E} \\ 0 & 0 & \frac{1}{5A} \end{bmatrix} \begin{bmatrix} R_N(N_Z) \\ R_N(N_X) \\ R_N(H_Z - N_Z) \end{bmatrix}, \qquad (1.21)$$

where: $E = 3A + B$; $A = (\mu_0/4\pi)^2(\gamma_I^2\gamma_S^2h^2)/8\pi^2r_{IS}^6$; $B = (\Delta^2\omega_S^2)/3$; and Δ is the chemical shift anisotropy.

1.5.3 Model free formalism

The relaxation rates therefore provide information on the overall tumbling and localized dynamics in proteins. This information is directly encoded in the form of the spectral density function. The next stage is to formulate models which connect the

motional parameters of proteins in solution with a spectral density function. The necessary link with the molecular dynamics is then available since the spectral density function is defined as the Fourier transformation of the autocorrelation function. Many models have been presented in the literature but the most powerful and widely-used is the model-free formalism proposed by Lipari and Szabo [55, 56]. Since the idea of a model-free model is initially confusing it is important to point out that the formalism is model free because it does not invoke any model for the internal motions. This is in contrast to other procedures with defined models for the internal motions. Such models can lead to overinterpretation of limited experimental data and the possibility of non-unique desriptions of internal motions in proteins. The model free formalism is very powerful in this respect since the relaxation data are completely described by a model-independent order parameter and an effective correlation time. An important assumption in the formalism is that the overall molecular motion and the internal motion are independent. The following expression is derived for the spectral density:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_c}{1+(\omega_0\tau_c)^2} + \frac{(1-S^2)\tau}{1+(\omega_0\tau)^2}\right] \tag{1.22}$$

where: $\tau^{-1} = \tau_c^{-1} + \tau_e^{-1}$; τ_e is an effective correlation time for the rates of internal motions in the extreme narrowing limit; τ_c is the correlation time for the overall molecular tumbling; and S is the order parameter which is inversely proportional to the degree of restriction of the motion. This form of the spectral density is the central equation in the model free formalism. This equation is derived without reference to any specific motional model and the parameters S and τ_e have model-independent significance therefore justifying the model-free description of the formalism. In practice the experimental and calculated relaxation parameters are least-squares fitted using this equation with S and τ_e as the only variable parameters.

In some cases internal motions in the protein can occur on two distinct timescales and the expression for the spectral density must be modified to allow successful fitting of the experimental data. A procedure for the analysis of such data was introduced by Clore and co-workers [57, 58]. In this approach it is assumed that there are two types of internal motions influencing the relaxation behaviour: very fast motions described

as diffusion in a cone; and slower motions involving jumps between two different orientations of the NH bond vectors. These two types of internal motions occur on timescales differing by one or two orders of magnitude and both are faster than the overall molecular correlation time. The modified expression for the spectral density is:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_c}{1+(\omega_0\tau_c)^2} + \frac{(1-S_f^2)\tau_f^*}{1+(\omega_0\tau_f^*)^2} + \frac{(S_f^2-S^2)\tau_s^*}{1+(\omega_0\tau_s^*)^2}\right] \tag{1.23}$$

where: $\tau_i^* = \tau_c\tau_i/(\tau_c+\tau_i), i = f,s$; τ_f and τ_s are the correlation times for the internal motions on the fast and slow timescales respectively; and the order parameter is a product of two corresponding fast and slow order parameters, $S^2 = S_f^2 S_s^2$. This expression for the spectral density requires six parameters for fitting the experimental data and therefore models must be selected to reduce the number of parameters.

Some protein molecules diffuse anisotropically in solution and the data analysis must take this into account. In this situation the relaxation of the backbone nuclei may be affected. However since the model free formalism allows separate descriptions of the overall tumbling and the internal dynamics the results from the data analysis using different models for the diffusion reflect this effect.

1.5.4 Relaxation measurements in proteins

The advances in two-dimensional NMR and isotope labelling completely changed the approach to studying biomolecular dynamics in solution. For proteins in particular, pulse sequences based on the HSQC experiment allow the measurement of relaxation parameters for individual backbone nuclei. The basic pulse sequences for measuring the heteronuclear NOE, the longitudinal, and the transversal relaxation rates are presented in Chapter 4. The pulse sequence for measuring the heteronuclear NOE is recorded in an interleaved mode with and without proton saturation. The heteronuclear NOE values are extracted simply by measuring the differences in the intensities of the cross peaks in the spectra recorded with and without proton

saturation. The longitudinal and transversal relaxation rates are extracted by fitting intensities of cross peaks in spectra recorded with different values of the relaxation delays. The intensities are fitted to the equations for longitudinal and transversal magnetization decay and the derived rates are evaluated using statistical methods which are now well established in the literature.

1.5.5 Hydrogen exchange

Amide proton exchange rates are another source of essential information about both the global stability and localized fluctuations in protein structures [59, 60]. The method chosen to measure the rates depends on the timescale of the exchange being monitored. In the slow exchange regime (from minutes to days) the rates can be calculated by following the decay in intensity of amide proton signals in proteins dissolved in D_2O. The exchange rates calculated from these measurements can provide information about the solvent accessibility of secondary and tertiary structure. The exchange rates vary predictably with both pH and temperature. At pH 3 the acid and base catalyzed exchange rates are equal which leads to a minimum exchange rate overall. The rates increase tenfold with each pH unit change. The rates increase approximately threefold with each 10°C increase in temperature allowing the calculation of the corresponding activation energy which is on the order of 17-20 kcal/mol. In proteins the exchange rates are modulated by local electronic and steric effects due to the structure. These effects can lead to large variations in the measured exchange rates in contrast to those obtained for model compounds. The pH dependence of the exchange rates is also attenuated compared with model compounds. The amide proton exchange rates in a protein can be probed at amino acid level in the same way as amide nitrogen nuclei relaxation rates by recording HSQC spectra at discrete time intervals. The general pattern is that protons in the interior of the structure and those involved in hydrogen bonding in secondary structural elements have extremely reduced exchange rates. The separation of contributions due to solvent inaccessibility and hydrogen bonding can be difficult to separate however. As a consequence there is no consensus on how hydrogen exchange occurs in proteins and which dynamic events are responsible. The exchange of interior protons is usually explained by invoking localized flexibility which allows

"open" conformations of the protein to be transiently populated.

1.6 Structure calculation

Once an almost complete set of assignments have been made the process of structure calculation can begin [19, 61]. Molecular dynamics simulations combine experimental information with simple physical models to calculate biomolecular structures [62]. This method simulates a system based on the interactions between the constituent particles. The interactions are governed by the forces between the particles which are combined with the equations of motion to allow simulation of the system. The motions of the particles can be calculated from the Newtonian equations of motion. Geometrical conformational restraints for use in the calculation can be derived from a number of different NMR parameters. The information derived from NMR experiments is supplemented with information about the covalent structure of the molecule. An advantage of such simulations is that the inherent flexibility of biomolecules can be tested by examining the results of a simulation of a flexible region in parallel with the experimental input data.

The usual strategy to calculate a structure is to minimize a target function which is a measurement of the agreement between a conformer and a set of restraints [63]. This is achieved by gradually fitting an initially random structure to the experimental restraints. The target function contains terms for both steric and experimental restraints. The target function is varied during the calculation to avoid local minima. At the beginning of a calculation only intraresidual restraints are used, followed by sequential constraints and, finally, medium and long-range restraints are introduced. In this way the local conformation around each amino acid is first built up followed by the global fold towards the end of the calculation.

Molecular dynamics in Cartesian space has the advantage of kinetic energy to escape local minima. This method is based on solving Newton's equations of motion to calculate trajectories for the atoms in the system. The atomic positions represent the degrees of freedom. In principle any MD program can be used for NMR calculations

provided that terms for NMR derived restraints are included in the energy calculation. The most widely used programs are CNS (XPLOR) and CYANA (DYANA) [64, 65, 66, 67].

Restraints can be derived from NOE data, scalar couplings, residual dipolar couplings, chemical shifts (diamagnetic and paramagnetic). The most widely used restraints are those which are derived from nuclear Overhauser effects between nuclei pairs in the molecule (see below).

Vicinal scalar coupling constants can also be used to calculate structural restraints according to the Karplus equation:

$$^3J(\theta) = A\cos^2\theta + B\cos\theta + C, \qquad (1.24)$$

where the parameters A, B, and C are determined by fitting measured coupling constants to known protein structures.

Pseudocontact shifts have been used in the refinement of protein structures [68, 69, 70, 71]. The pseudocontact shift for a proton due to the influence of a metal ion can be approximated as described above. This assumes that there are no redox state dependent structural changes which is a valid assumption for haem groups and the majority of amino acid residues. Ideally, experimental pseudocontact shifts could be derived from a comparison of shifts for the reduced and oxidized states of the protein. In the absence of assignments a crystal structure can be used to calculate the shifts.

At high field strengths molecules can be aligned due to their anisotropic magnetic susceptibility tensors [72]. Alternatively biomolecules can be dissolved in solutions containing alignment media such as liquid crystals which allow the measurement of residual dipolar couplings (RDC) which can be used as restraints in structure calculations or, in some cases, in studies of protein dynamics [72, 73, 74, 75].

1.6.1 The nuclear Overhauser effect

Quantitative information from crosspeaks in a NOESY spectrum is determined by integrating over the peak volume. For isolated crosspeaks the simplest method is to integrate within an ellipsoidal area surrounding the peak. For overlapping clusters of peaks deconvolution can be achieved by lineshape integration. An important factor to bear in mind is that the error in the derived restraint is only one sixth of the error of integration. Unknown proton distances (r_u) can be calculated from known covalent distances (r_k) by measuring cross peak intensities:

$$r_u = r_k \left[\frac{I_k}{I_u}\right]^{1/6}, \tag{1.25}$$

where I_k and I_u are the respective cross peak intensities. This relationship must be used with the implicit approximations of the isolated spin pair model in mind. Even though the derived restraints are not used as exact restraints on the distances during the calculation it is important to consider the range within which a given interatomic distance can vary. The distance restraints are often only used as upper bounds on the interatomic distances. This can lead to the situation where two nuclei can be in van der Waals contact even though no NOE has been observed between them. For this reason it is prudent to use both upper and lower bounds on the interatomic distances.

1.6.2 Variable target function method

As stated above the basic idea of the target function method is to fit a random starting structure to the experimental input starting with the intraresidual restraints and working up to include sequential, medium, and long range restraints. This method works in torsion angle space and therefore the covalent geometry is set at the beginning of the calculation. This method was first used in the program DISMAN [63] and later in the program DIANA [76].

1.6.3 Molecular dynamics in Cartesian space

Molcular dynamics in Cartesian space solves Newton's equations of motion to find trajectories for all of the atoms in a system. The number of degrees of freedom is equal to the number of atoms. The aim is to search the conformational space as efficiently as possible using a target function which represents the potential energy of the system. Violated restraints contribute to the target function and therefore in order to minimize the energy of the system as many as possible of the experimental data must be satisfied. The main advantage is the presence of kinetic energy which helps the system escape local minima in the energy landscape. Newton's equations of motion for a system of n particles with masses m_i and positions r_i are given by:

$$m_i \frac{d^2 r_i}{dt^2} = F_i \ (i = 1,..,n). \tag{1.26}$$

The forces F_i are given by the negative slope of the potential energy function. The potential energy functions normally used contain terms for bond lengths, bond angles, torsion angle potentials, chirality and planarity, and nonbonded repulsions in addition to NMR specific terms representing distance and torsion angle restraints. For example the potential energy function used in the program XPLOR is given by [64]:

$$\begin{aligned} E = \ & \Sigma_{bonds} k_b (r - r_0)^2 \\ & + \Sigma_{angles} k_\theta (\theta - \theta_0)^2 \\ & + \Sigma_{dihedrals} k_\phi (1 + \cos(n\phi + \delta)) \\ & + \Sigma_{impropers} k_\phi (\phi - \delta)^2 \\ & + \Sigma_{nonbonded} k_{repel} (max(0, (sR^2_{min} - R^2)))^2 \\ & + \Sigma_{distances} k_d \Delta_d^2 \\ & + \Sigma_{angles} k_a \Delta_a^2. \end{aligned} \tag{1.27}$$

The terms in this equation are explained as follows:

- $k_b, k_\theta, k_\phi, k_{repel}, k_d$, and k_a represent the various force constants;
- r and r_0 are the actual and correct bond lengths respectively;
- θ and θ_0 are the actual and correct bond angles;
- ϕ is the actual torsion angle or improper angle value;
- δ is an offset of the torsion angle and improper potentials;
- R_{min} is the distance where the van der Waals potential has its minimum;
- R is the actual distance between a non-bonded pair;
- s is a scaling factor;
- Δ_d and Δ_a are the size of the distance or torsion angle restraint violation.

A start structure can be generated by metric matrix distance geometry or the variable target function method. Alternatively an extended structure or even atoms randomly distributed in space can be used. Actual trajectories are calculated by integrating the equations of motion over very small finite time steps. The temperature can be controlled by coupling the system to a heat bath. To ensure sufficient sampling of the conformational space simulated annealing is initially performed at a high temperature with a low weighting for the steric repulsion terms to allow atoms to penetrate each other. The weighting of the steric repulsion terms is gradually increased and the temperature is decreased before a final energy minimization stage. This procedure is repeated for all of the start conformations.

1.6.4 Torsion angle dynamics

The essence of this approach is that the covalent geometry of the molecule is kept fixed while only the torsion angles are allowed to vary. The main advantage is that

strong potentials to retain the covalent structure are not required and therefore a simpler potential energy function can be used allowing more time for the calculation of the equations of motion. This means that the algorithm is much more efficient than the conventional molecular dynamics programs used for protein structure calculations. The success of the calculation is measured in terms of a target function which represents the potential energy of the system. The target function is equal to zero only if all the experimental input data is satisfied and there is no steric overlap between all non-bonded atom pairs. The exact form of the target function is given by the following equation:

$$V = \sum_{c=u,l,v} \omega_c \sum_{(\alpha,\beta)\in I_c} f_c(d_{\alpha\beta}, b_{\alpha\beta}) + \omega_d \sum_{k\in I_d} \left[1 - \frac{1}{2}\left(\frac{\Delta_k}{\Gamma_k}\right)^2\right] \Delta_k^2, \tag{1.28}$$

where:

- $d_{\alpha\beta}$ is the distance between two atoms α and β;
- $b_{\alpha\beta}$ are upper and lower bounds between α and β;
- I_u, I_l, and I_v are the sets of atom pairs (α,β) with upper, lower, or van der Waals distance bounds respectively and I_d is the set of restrained torsion angles;
- ω_c and ω_d are weighting factors for the different types of restraints;
- $\Gamma_k = \pi - (\theta_k^{max} - \theta_k^{min})/2$ denotes the half width of the forbidden range of torsion angle values;
- Δ_k is the size of the torsion angle restraint violation.

Violated distance restraints contribute to the target function according to the function $f_c(d,b)$ which can take various forms depending on the magnitude of the violation. The form used most often is:

$$f_c(d,b) = \left[\frac{d^2 - b^2}{2b}\right]^2, \tag{1.29}$$

where d and b denote the actual distance and the distance bound respectively.

Since the potential energy landscape of a protein is very complex with many local minima, successful annealing requires careful use of temperature during the schedule. The kinetic energy, *i.e.* the temperature, enables the structure to overcome the local minima. The calculation starts with a conformation in which all the torsion angles are random. The standard simulated annealing protocol used in DYANA has eight stages:

- a start conformation with random values for all of the torsion angles is generated;
- a steric check excluding all hydrogen atoms is performed;
- 200 conjugate gradient minimization steps to reduce high-energy interactions;
- $1N/5$ TAD steps at constant high temperature e.g. 10^4K;
- $4N/5$ TAD steps with slow cooling to almost zero temperature;
- 100 conjugate gradient minimization with all hydrogen atoms to check for steric overlap;
- 200 TAD steps at zero reference temperature;
- final minimization with 1000 conjugate gradient steps.

where N is the total number of TAD steps. The efficiency of a structure calculation can be assessed in terms of two factors: the time required to calculate one conformer and the success rate, *i.e.* the convergence to a well-defined ensemble which satisfies the input data. On both of these measures DYANA has a far superior performance to other programs.

1.6.5 Relaxation matrix calculations

Experimental NOE intensities can deviate from their theoretical values because of spin diffusion and internal mobility within the molecule. The calculated structures

can be used to calculate the expected NOE intensities using the complete relaxation matrix method [77, 78]. The calculated NOE intensities can then be compared to the experimental values to assess how well the structures satisfy the experimental data. The calculated values can be used to derive a more accurate set of distance restraints. In the relaxation matrix the diagonal elements represent the nuclear longitudinal relaxation rate for each spin and the off-diagonal elements represent the cross-relaxation rate:

$$\mathbf{R} = \begin{bmatrix} \rho_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{n1} & \cdots & \cdots & \rho_{nn} \end{bmatrix}, \tag{1.30}$$

where the cross-relaxation and the autorelaxation rate constants are given in Equations 1.9 and 1.10 respectively. Crosspeaks in a NOESY spectrum of a paramagnetic protein can be attenuated due to an increase in the relaxation rates of nuclei close to the iron atoms caused by dipolar coupling to the unpaired electrons. The contribution to longitudinal relaxation from the dipolar interaction between a nuclear spin and an unpaired electron can be calculated from the following equation [38]:

$$\begin{aligned} \rho_{iM}^{para} = & \frac{2}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r_i^6} \\ & \times \left[\frac{\tau_S}{1+(\omega_0-\omega_S)\tau_S^2} + \frac{3\tau_S}{1+\omega_0^2\tau_S^2} + \frac{6\tau_S}{1+(\omega_0+\omega_S)^2\tau_S^2}\right] \end{aligned} \tag{1.31}$$

where:

- γ_I is the nuclear magnetogyric ratio;
- g_e is the electronic g-factor;
- μ_B is the Bohr magneton;

- S is the electron spin angular momentum;
- r_i is the distance between the metal nucleus and the proton;
- τ_S is the correlation time for electron spin relaxation;
- ω_0 and ω_S are the Larmor frequencies of the proton and electron.

Paramagnetic contributions can be included by adding the value electron-nucleus dipolar relaxation to the diagonal elements of the relaxation matrix [79]. This method is approximate and in practice some error must be allowed for by relaxing the resulting distance constraints. This can be achieved by comparing calculated NOE volumes with those for which experimental data exist. These experimental data can replace the entries for theoretical entries in the relaxation matrix. This hybrid matrix can be retransformed to give a new set of predicted distances. The distances that are outside the range of the experimental data can then be used to calculate a ratio that can be used to relax all distance restraints.

1.6.6 Automated NOE assignment

The assignment of NOESY spectra to collect distance restraints for use in structure calculations is one of the most time consuming stages in a structure calculation project. Consequently, some groups have attempted to develop programs that perform this step automatically. The two main problems in completely assigning NOESY spectra are incomplete resonance assignments and chemical shift degeneracy. If the resonance assignments are incomplete there can be many peaks in the NOESY spectra for which there are simply no information and therefore this potential information is essentially discarded. The heteronuclear assignment strategy outlined above prevents this situation in the majority of cases. Chemical shift degeneracy can make the assignment of NOESY peaks very difficult although this is of course much simpler with heteronuclear-edited spectra. Another very important reason for the development of automated assignment programs is simply to save time. Some groups have shown that their automated assignment and structure calculation programs produce output that is very similar to those resulting from manual analysis. There are

a few different approaches to this problem that are currently implemented in programs such as ARIA and CANDID [67, 80, 81, 82, 83]. However it remains true that the most basic requirement for such programs is complete or nearly complete resonance assignments.

A recent and very important development in biomolecular structure calculation from NMR data is the use of ambiguous distance constraints [82]. The basic idea is that at a given cross peak position there may be NOE contributions from two or more peaks based on degenerate chemical shifts in one or both of the frequency dimensions. ARIA uses a summed distance which is an ambiguous distance restraint (ADR) for all of the assignment possibilities of the peak in question:

$$\bar{D} = \left[\sum_{i=1}^{N} d_i^{-6} \right]^{-1/6}, \tag{1.32}$$

where the index i runs through all of the N possible assignments for the crosspeak and d_i is the distance between the corresponding pairs of protons. This summed distance can be calculated from a structure and used as input to the next calculation in an iterative cycle. The ADRs can in effect be used as distance restraints contributing to a target function in a way analogous to standard distance restraints. This method can be very powerful as the NOE assignment can be performed in successive iterations of the structure calculation cycle. The reason for this is because the summed distance $\bar{D}$ is strongly weighted towards the shortest of the contributing distances.

The efficiency of the calculation can be increased by combining ADRs with unambiguous NOE assignments. Initially the unassigned NOEs are given possible assignments based on chemical shift values only. These assignments are combined with the unambiguous assignments and submitted to the first iteration. The best structures are selected for further intepretation of the input data. Distances for each assignment possibility are measured from these structures. The volumes of the unassigned NOE peaks are converted into distance restraints and compared to the summed distances for each assignment possibility obtained from the structures. At this stage some of the assignment possibilities can be discarded because the

corresponding distances in the structures are greater than a specified cutoff value. New structures can then be calculated using the updated restraint lists. In principle this procedure should result in an increased percentage of assigned peaks at each iteration and the structures should converge torwards a well-defined ensemble.

An alternative approach to automated NOESY assignment is implemeted in the program NOAH which is integrated into DYANA [80]. The essential difference from the ARIA method is that NOAH creates an unambiguous restraint for each possibility while ARIA creates one ADR for each ambiguous peak. Another way to view this difference is that NOAH incorporates incorrect input into a calculation with the aim of identifying such data by violation analysis. In contrast ARIA tries to avoid errors in the assignment of unknown peaks by using ADRs to account for all assignment possibilities. Recently, the program CANDID [67] has been developed to incorporate ambiguous NOE data into the program DYANA for structure calculation.

1.6.7 Structure analysis

This is the most crucial stage of the actual calculation process because only by a critical analysis of the results can any conclusions be drawn about the accuracy and precision of the ensemble of structures [84, 85]. The first distinction to be made is that between accuracy and precision. Accuracy is defined as the degree to which the calculated ensemble represents the actual situation in solution. Precision refers to the reproducibilty of the results and is thus an indication that the calculation is probably on the right track and hopefully not trapped in a local minimum. The goal is to derive an accurate and precise ensemble that accounts for the experimental data. Agreement with the experimental data is a good indication that the ensemble is not a very precise yet inaccurate group of conformers.

One of the most basic criteria for an accurate protein structure is that the values for the torsion angles ψ and ϕ are within the favoured ranges allowed sterically [2]. This distribution is normally analyzed by means of a Ramachandran plot [86]. Software tools such as PROCHECK [87, 88] and WHATIF [89] have been developed to aid this type of validation. Another important criterion is the packing quality which can be

measures using the DACA (directional atomic contact analysis) method proposed by Vriend and Sander [90].

Questions about accuracy and precision have given rise to studies which try to determine the limits of these two parameters. A study by Zhao and Jardetzky [91] has shown that the precision mainly depends on the number of input restraints and is relatively insensitive to the quality of that data. The authors also concluded that accuracy does depend on the quality of the input data. Recently, an important study by the Utrecht group examined the correlation between the quality of various NMR structures in relation to the amount of experimental input data [92]. The experimental data were examined using the program AQUA [88]. The authors concluded that the precision does increase with the amount of experimental data, as expected.

Since the structures calculated by the methods outlined above are *in vacuo* water refinement can be used in an attempt to mimic the natural environment of biomolecules more closely. Comparisons can be made between structures simulated *in vacuo* and those refined in implicit or explicit water [93, 94].

Databases have become important research resources in this field and deposition of experimental and structural data in the BioMagResBank (BMRB) and the Protein Data Bank (PDB) is encouraged [95, 96, 97] [2].

[2]BMRB, `http://www.bmrb.wisc.edu`; PDB, `http://www.rcsb.org`

Chapter 2

Solution structure of cytochrome c_3 from *Desulfomicrobium norvegicum*

2.1 Introduction

Cytochrome c_3 is a small protein found in all species of the *Desulfovibrionaceae* family of sulphate-reducing bacteria [98, 99, 100, 101]. Tetrahaem cytochromes c_3 have an average molecular weight of about 15kDa and the four haem groups have bis-histidinyl ligation of the iron atoms. The four haem groups have a conserved spatial arrangement as confirmed by all of the structural studies so far. The tertiary structure is conserved between the different cytochromes c_3 despite the very low sequence homology. The postulated biological role of c_3 is as a coupling factor to the enzyme hydrogenase in the periplasm (the structureless region between the plasma membrane and the cell wall) of these bacteria [102, 103, 104].

According to the proposed model cytoplasmic hydrogenase produces molecular hydrogen which diffuses across the plasma membrane where it is taken up and oxidized by periplasmic hydrogenase. The reaction catalyzed by hydrogenase produces two electrons and two protons. The transduction performed by c_3 results in the transfer of energy from the electrons to the protons. The electrons are donated to a transmembrane electron transfer complex and then passed to sulphate reducing enzymes in the cytoplasm. The protons are utilized by ATP synthase in the production of ATP.

The structural basis of energy transduction involves cooperation between acid-base and redox centres in c_3. Each of the four haem groups of a given cytochrome c_3 has a different pH-dependent redox potential. The ranges of redox potentials for cytochromes c_3 from the different organisms are relatively constant with the notable exception of c_3 from *Desulfomicrobium norvegicum* (c3*Dn*). c3*Dn* has the broadest range of redox potentials as confirmed by EPR measurements [105]. Another anomalous feature of c3*Dn* is that it is the only c_3 for which the intermolecular electron transfer is fast on the NMR scale [106] thus precluding the possibility of studying the thermodynamics using the method of following the haem methyl resonances through the five different oxidation levels as has been achieved for cytochromes c_3 from *D. gigas* (c3*Dg*) and *D. vulgaris* (c3*Dv*) [107, 108]. This "redox-Bohr" effect is made possible by conformational changes and/or electronic changes in haem group and amino acid sidechains. Therefore in order to understand

this behaviour the structures of the protein in both the reduced and oxidized states are required to identify which conformational changes are due to redox events. NMR allows the variation of experimental conditions such as pH and oxidation state making it a very useful technique to study redox proteins and enzymes in general. In addition it would also make sense to compare structures obtained by the same technique rather than, for example, comparing an X-ray structure of the oxidized state with an NMR structure of the reduced state.

The structures of several cytochromes c_3 have been determined by early NMR studies (Table 2.1) [109, 110, 111, 112]. The three-dimensional structures of several cytochromes c_3 have been determined by X-ray crystallography [113, 114, 115, 116, 117] and NMR [118, 119, 120]. The first crystal structure of c3*Dn* [121] was corrected by NMR studies [118] and subsequently a higher resolution structure incorporating the corrections was published [114]. The electronic structure has also been elucidated and was found to be similar to those from the other organisms [39]. Recently the structure of a novel tetrahaem cytochrome with a different haem core architecture has been determined [122].

protein	method	oxidation state	pH value	citation
c3*Dv* (Miyazaki)	X-ray	oxidized		[113]
c3*Dn*	X-ray	oxidized		[114]
c3*Dg*	X-ray	oxidized	6.5	[115]
c3*Dv*	X-ray	oxidized	5.5	[116]
c3 ATCC 27774	X-ray			[116]
c3*Da*	X-ray	oxidized,reduced		[117]
c3*Dv*	NMR	oxidized,reduced	7.1,8.5	[119, 123]
c3*Dg*	NMR	oxidized,reduced	5.1,7.3	[120]

Table 2.1: Structural studies of cytochromes c_3.

A very interesting development has been the successful expression of tetraheme cytochromes c_3 in *E. coli* [124, 125] which will allow the efficient production of labelled samples which will in turn speed up the assignment process and open up possibilities for dynamics studies. This expression technique has recently been successfully applied to produce and assign a ^{15}N-labelled sample of c_3 from *D. vulgaris* [126].

The aim of this project was to determine the solution structure of c3*Dn* by NMR methods. This project was part of an EU contract to study the biophysics of haem proteins and to develop further the NMR methodology for determining the solution structures of paramagnetic haem proteins.

2.2 Materials and methods

All of the experimental work for this project was carried out at the ITQB in Lisbon in the laboratory of Prof. Antonio Xavier as a partner in the TMR network funded by the EU.

2.2.1 Sample preparation

Cytochrome c_3 from *Desulfomicrobium norvegicum* was purified as described [127]. After purification the solution was dialyzed at 4 °C firstly with 0.1 M NaCl to remove residual buffer and then with deionized water (Millipore) to remove excess salt. The sample was filtered through a Millex-GS unit with 0.22 μm pore size and then lyophilized. For NMR experiments in H_2O the protein was dissolved in 92% H_2O - 8% D_2O. For NMR experiments in D_2O the protein was lyophilized four times from D_2O, incubated at 321 K for two hours and then lyophilized again. This procedure was performed in order to ensure efficient deuterium exchange of the protected exchangeable protons. The sample was then dissolved in D_2O (99.98 atom %). The volume of the samples in the NMR tube was approximately 550 μL. The final protein concentration was approximately 2.7 mM and the salt concentration was in the μM range. An antibiotic cocktail (127 μM ampicillin, 78 μM kanamicin, and 141 μM chloramphenicol) was added to prevent bacterial growth. The pH was adjusted to 4.75 (uncorrected pH reading) and monitored with a glass electrode (Ingold) inserted directly into the NMR tube.

2.2.2 NMR spectroscopy

NMR spectra were recorded on a Bruker AMX-500 spectrometer equipped with a 5mm inverse detection probe head with internal B_0 gradient coil. The sample temperature was controlled using a Eurotherm 818 temperature control unit with a BCU 05 cooling unit. Assignments were performed using spectra of samples at pH 4.75 acquired at 303K but additional spectra were acquired at 300K to help resolve peak overlap. All 2D spectra were acquired in the phase-sensitive mode by the States-TPPI method [128] collecting 4096 (t_2) $\times$ 1024 (t_1) data points with at least 40 scans per increment. NOESY spectra [129, 130] were acquired with mixing times of 25 and 100ms and with spectral widths of 32 and 22 kHz. NOESY spectra of the D_2O sample were recorded with standard pulse sequences with continuous low-power saturation of the water resonance during the relaxation delay and the mixing time. NOESY spectra of the H_2O sample were recorded in the same manner for the larger spectral width. The spectra with the smaller spectral width were recorded using the WATERGATE sequence [131]. Total correlation spectra were recorded using the clean TOCSY pulse sequence [132] with spin-lock times of 20, 25, 50 and 75ms and 22kHz spectral width. COSY [133, 134, 135] spectra were recorded with a 13 kHz spectral width. Raw data were multiplied in the F_2 dimension by a Gaussian function with line-broadening of -5 Hz and by a pure cosine-squared function in the F_1 dimension except for the COSY spectra which were multiplied by a pure sine-squared function in both dimensions. The spectra were typically processed to a final size of 4096 $\times$ 1024 data points except the COSY spectra which were processed with 1024 $\times$ 1024 data points. Polynomial baseline corrections were applied on both dimensions of all spectra. The data were processed using XWIN-NMR software (Bruker, Rheinstetten).

2.2.3 Assignment

The assignment was performed using XEASY [136] to display and annotate spectra and for volume integration of NOESY crosspeaks. The NOEs were measured in the 100ms NOESY spectra in H_2O and D_2O at 303 K and pH 4.75. TOCSY and NOESY

spectra acquired at 300 K were used to perform sequential assignments as described in Chapter 1.

2.2.4 Paramagnetic tensor analysis

The analysis of the dipolar shifts was performed as described in Chapter 1. Once a significant number of shifts had been assigned the dipolar shifts were calculated by subtraction of calculated shifts for the reduced form and used to calculate the magnetic tensors with respect to the crystal structure. The predicted dipolar shifts were used to indicate possible assignments.

The orientation of the axial histidine ligands determined from the NMR analysis were used to derive dihedral angle restraints. The analysis determines the relative orientations of the histidine planes to each other and also their resultant position as a pair but does not distinguish between the two ligands. Preliminary structures were therefore used as far as possible to determine which angles correspond to which ligands.

Note however that neither the dipolar shifts or the histidine angle restraints were used to calculate the structures analyzed in this chapter.

2.2.5 Determination of volume restraints

Integration was performed either by manual integration for isolated peaks or with line-shape integration for overlapped peaks. The baseline around each individual peak (or cluster of peaks) was measured and used to correct the volume. Peak volumes were measured in NOESY spectra recorded with 25 and 100 ms mixing times and for protein samples prepared in H_2O and D_2O. Peaks assigned to protons separated by fixed distances and all intrahaem peaks (except those assigned to the propionate groups) were excluded. The two sets of NOE volumes from the two mixing times were then processed separately to convert them into upper and lower

volume limits. PARADYANA converts the input NOE volumes into proton distances using the function:

$$r = \frac{k}{\sqrt[n]{V}} \tag{2.1}$$

where r is the inter-proton distance, k is a scaling factor and V is the NOE volume. Separate scaling factors are used for NOEs involving two methyl groups, one methyl group and a single proton, two single protons in sidechains or two backbone protons. A power of $n = 4$ was used in the calculations.

2.2.6 Paramagnetic leakage correction

The correction factors for paramagnetic leakage were calculated following the procedure described in Chapter 1. All of the constants in Equation 1.31 may be expressed by a single parameter k such that:

$$\rho_i = k r_i^{-6}. \tag{2.2}$$

This model assumes that the unpaired electron densities are localized on the iron atoms. It is also assumed that τ_S is equal for the four haem groups. The values of k and τ_c were the same as those used for c3*Dg* which were estimated from the intensities of crosspeaks in the 25 and 100 ms spectra [120]. The values used were $k = 10^6 \text{Å}^6\text{s}^{-1}$ and $\tau_c = 6$ ns. Crosspeak intensities were then calculated with and without paramagnetic effects using averaged distances from an ensemble of preliminary structures. For each calculated intensity a correction factor:

$$c_{ij} = \frac{I_{ij}^{para}}{I_{ij}^{dia}} \tag{2.3}$$

was obtained and each upper and lower volume limit was divided by its correction factor. As mentioned above this approach is approximate and the distance constraints were relaxed appropriately.

2.2.7 Structure calculation

Six non-standard residues were used in the structure calculations: fast-flipping aromatic residues; flexible haem groups and proline residues; a linking cysteine residue to the first haem binding site; and two types of axial histidine residues [42, 119, 120]. Both histidine residues contain a set of pseudoatoms that are used to maintain the hexacoordinate geometry of the iron through several fixed *upls*. The fifth axial histidine ligand also contains several pseudoatoms that represent the principal axes of the magnetic susceptibility tensor. In c3*Dn* there are four haem groups and eight proline residues and their design with allowance for flexibility requires 150 fixed *upls*.

The contribution of the restraints to the target function was defined as DYANA type 1 except for the fixed *upls* used in the non-standard residues which were defined as type 2 [66]. The torsion angles, susceptibility tensors, scaling factors, dipolar shifts (if used), and angle restraints are all optimized simultaneously. Each calculated structure is therefore calibrated independently and has its own set of scaling factors. After conjugate gradient minimization to remove strong overlaps the structures were annealed with 60,000 steps. The weighting of the van der Waals repulsion was then increased to 2.0 followed by conjugate gradient minimization and 200 steps of molecular dynamics at constant temperature and a final 2000 steps of conjugate gradient minimization.

2.2.8 Structure analysis

The program MOLMOL version 2.6 [137] was used for superimposition, visual inspection, drawing of the ensemble and calculation of solvent exposures. Average

solvent exposures were determined using for each H_2O molecule a radius of 1.4 Å. The NMR structures were analyzed with respect to the experimental restraints with the program AQUA [88]. The stereochemical analysis of the structures was performed with the program WHATIF [89] and with both PROCHECK [87, 138] and PROCHECK-NMR [88]. WHATIF was also used for the identification of possible hydrogen bonds in the ensemble of conformers and in the X-ray structure of the ferricytochrome [114] and for the identification of crystal contacts in this structure. Identification and classification of the consensus secondary structure elements in the NMR ensemble, defined as those present in at least 50% of the structures, was performed using PROMOTIF [139].

2.3 Results and discussion

2.3.1 Sequence specific assignment

The assignment of the proton resonances in c3*Dn* was expected to be quite straightforward due to the presence of the four paramagnetic centres which increase chemical shift dispersion. This was indeed the case for c3*Dv* and c3*Dg*, the assignments being achieved by two Ph.D. students in reasonable time [123]. However for c3*Dn* the assignment was a very difficult and frustrating task lasting about two years. A major problem has been incomplete spin systems in the TOCSY spectra. Eventually patterns were recognised that allowed the assignment of spin systems that are far from complete in the TOCSY but which have all their intra-residual NOEs in the NOESY spectra. All of these spin systems have been confirmed by sequential assignments and by long-range NOEs. Also by comparing the shifts of certain residues from c3*Dv* and c3*Dg* a qualitative pattern is seen which supports the assignment of the known residues and gives an indication of what types of shifts the unassigned ones should have.

The assignment of the NMR spectra of c3*Dn* presented the main bottleneck stage in proceeding with the project to determine the solution structure. Progress was

eventually made due to the observation of particular spectral features. An attempt to explain follows. In principle, the TOCSY spectrum of a protein sample should include complete correlation grids for the constituent amino acids residues of the protein. Normally the grid is complete for the majority of the residues although overlap can of course make some of the grid points very difficult to find. If the grid is not fully complete it is usually sufficiently complete as to leave little doubt that it represents a spin system. However in the spectra of c3*Dn* many of the assignments are for residues that have substantially incomplete grids in the TOCSY spectra. This was noticed at first for certain key residues with many NOE connectivities. This apparent fact was then used to make many new assignments and at this stage it seems to just be a feature to be included in the reckoning when making assignments. Also some of the residues which have been assigned are incomplete but their identity has been confirmed by the observation of sequential NOEs. This is usually the case for longer side chains but also for an unknown reason for threonine residues.

Louro *et al.* [39] had already assigned the haem resonances as part of a study of the electronic structure of porphyrins in cytochromes c_3. In that work a ^{13}C-^{1}H HMQC spectrum was used in conjunction with a NOESY spectrum run with a mixing time of 25 ms to specifically assign the haem protons. Table A.1 presents the chemical shifts of the haem assignments.

The second approach was to try to identify as many of the individual spin systems as possible. The first step was to identify the backbone NH-αH crosspeaks. To ensure that these correlations are due to NH protons the spectra recorded in D_2O were checked to see if the signals were attenuated due to exchange with the solvent. The exchange rates for NH protons in regions of the protein with different exposure to solvent will vary widely. It is possible that some NH protons in the protein interior will be exchanging so slowly that the signal may not have disappeared during the acquisition time. Sometimes the two αH protons of a glycine residue may be degenerate which can reduce the number of identified correlations.

The NH-αH crosspeaks were taken as starting points to search for complete spin systems corresponding to amino acids in the protein sequence. The spin systems can be identified by examination of the COSY [140] and TOCSY [141] spectra. Two

COSY spectra were used, one of which was processed with a line-narrowing window function applied to help resolve peaks in crowded areas such as the fingerprint region.

A series of TOCSY spectra were recorded: at 303K in H_2O with mixing times of 50 and 75ms; at 303K in D_2O with mixing times of 20 and 50 ms; at 306 K in H_2O with mixing times of 25 and 50 ms; at 300 K in H_2O with mixing time of 50 ms. The spectra run with longer mixing times proved useful for delineating long spin systems such as lysine because the magnetization transfer will have time to propagate through the whole system. The spectra run at the lower and higher temperatures were used to help resolve confusion due to overlapping NH resonances.

A series of complementary NOESY spectra were recorded to facilitate the analysis. The actual sequence-specific assignments were obtained after an exhaustive analysis of the NOESY spectra. The main high quality NOESY spectrum used for this analysis is shown in Figure 2.1. As outlined above the COSY, TOCSY, and NOESY spectra were used in combination to achieve the assignment. The problem of the incomplete TOCSY grids was compounded by the severe overlap for all but the most shifted signals in the spectra. Therefore for the signals with substantial dipolar shifts the assignments were easily confirmed by reference to the earlier work. This allowed the identification of NOE peaks for intra-haem, haem-haem, and haem-protein interactions.

The assignment of the fingerprint region in both the correlation and the NOESY spectra was relatively straightforward. A useful feature was the observation of rotating nuclear Overhauser effect (ROE) peaks in the TOCSY spectra. ROE interactions in TOCSY spectra are very useful in identifying interactions which are unambiguously due to dipolar coupling and not scalar coupling. These peaks are easily recognized due to their opposite sign to the main TOCSY peaks. Normally the ROE will have a slower build-up rate than the TOCSY effect for scalar coupled nuclei and so will not interfere to a large extent with the assignment. For certain interactions the ROE assignments were confirmed in spectra recorded at different temperatures.

The regions with especially severe overlap mainly concerned the assignment of the amino acid side chains. This was particularly evident in the region connecting the NH

resonances with the upfield side chain resonances. This region was, however, still essential for the assignment. The variable temperature spectra were also helpful here since ladders of NOE interactions associated with a particular NH resonance could be tracked at different temperatures.

In all of the spectra the most severe overlap was encountered in the aliphatic region. This region presented two almost intractable problems. The correlations of the Hα protons with the upfield methylene and methyl groups were very difficult to assign. This was even observed in the spectra recorded in D_2O due to inefficient suppression of the residual water signal during data acquisition. This made the assignment of many lysine and proline side chains impossible. Overlap in the methyl region close to the diagonal prevented the assignment of certain side chain methyl groups. The overlap problem in the aliphatic region made a major contribution to the lack of side chain assignments in the loop from residue M16 to P25. This, in turn, led to a paucity of medium and long range restraints for this segment of the protein.

In total, some assignments were achieved for the majority of residues (Table A.2). The residues for which no assignments were found are: P4, P20, K21, P25, P33, H36, H39, C44, H48, H49, R73, K75, A76, E85, N86, D94, H96, P105, G107, P108. These 20 residues from a total of 118 represent a sizeable fraction and should be kept in mind regarding the analysis of the structures presented below.

2.3.2 Paramagnetic tensor analysis

The NMR signals in the region of paramagnetic haem groups are severely attenuated due to the effects of paramagnetic relaxation. Structural information in these regions is however essential for determining small differences between structures which may be relevant. Accordingly, new methods have been developed to assign crucial information on the haem groups and the attached ligands. The magnetic and structural parameters in paramagnetic haem systems are highly correlated and a careful analysis of experimental data can yield information on the relationship between the geometry of the ligands and the form of the frontier molecular orbitals [39].

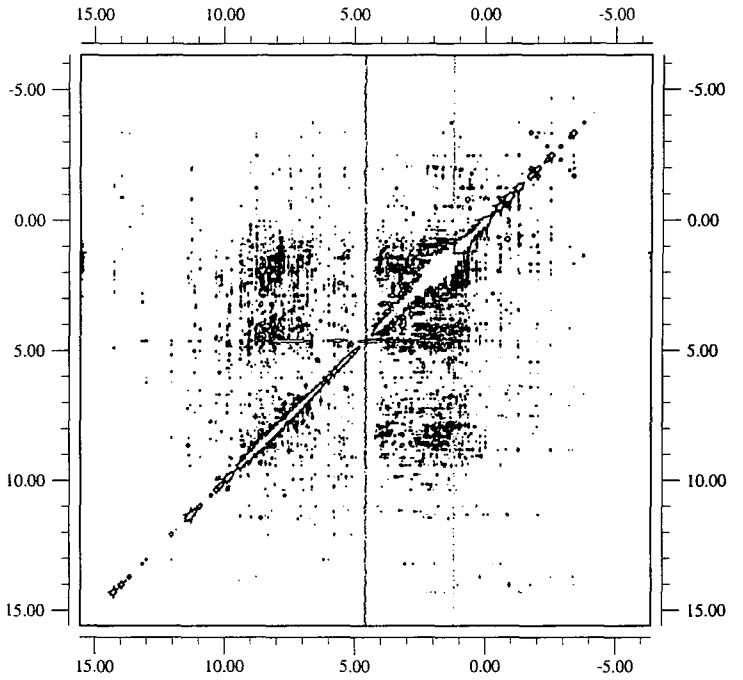

Figure 2.1: 2D NOESY spectrum of c3*Dn* recorded at a proton frequency of 500 MHz and 303 K.

The geometry of the axial ligands in low-spin bis-histidinyl haems *c* determines the form of the two frontier molecular orbitals of the tetrapyrrole ring. These molecular orbitals undergo a rhombic perturbation (θ) with a concomitant energy splitting (ΔE). A linear correlation exists between both of these parameters and geometric properties of the axial ligands. Specifically, the resultants of the normals to the histidine planes (ϕ) are correlated with θ and the dihedral angle between the planes (ψ) is correlated with ΔE. The angle ϕ is defined as the angle between the bisector of the acute angle between the normals to the histidine planes and the vector defined by the pyrrole nitrogens A and C of the haem . The definition of the angle ψ follows from the definition of ϕ (see Figure 2.2).

The ^{13}C resonance frequencies of the haem α-subsituents are dominated by the contact shift interactions. A detailed molecular orbital theory analysis yields relationships between the electronic and geometrical parameters. The energy splitting

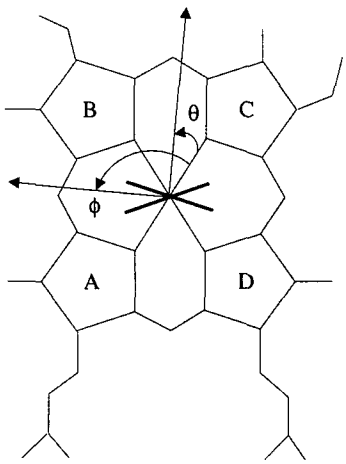

Figure 2.2: Geometric parameters used to describe the haem ligands (adapted from [142]).

is a maximum if the ligands have a zero dihedral angle and a minimum when the dihedral angle is 90°. The values for the energy splitting can be combined with those for the rhombic perturbation in an empirical expression to estimate the dihedral angle between the histidine planes (ψ) [39]:

$$\Delta E = (5 + \cos 4\theta)\cos\psi. \tag{2.4}$$

Further analysis investigated the relationship between the ligand geometry and the paramagnetic susceptibility tensor [40]. In the case where the magnetic z-axis is close to the normal of the haem plane it was found that the x and y axes are inversely correlated with the resultant ligand orientation ϕ. Furthermore, since ϕ is also directly correlated with θ, the expected inverse correlation of the x and y axes and θ was observed. The data were analyzed using a model in which dipolar shifts were expressed in terms of the spherical polar coordinates of the nucleus with repect to the principal axis system of the paramagnetic susceptibilty tensor with equatorial and axial anisotropies according to Equation 1.4. The Euler angle β represents the angle between the magnetic z-axis and the normal to the haem plane. In the case where this angle is small the sum of the angles α and γ is approximately equal to the in-plane rotation of the y-axis of the tensor.

The relationship between ΔE and ψ is paralleled by a correlation between the axial and equatorial anisotropies and ψ. This correlation is at best qualitative and is only reliable in the case of nearly perpendicular ligands.

The two models are however in agreement and both can be used to assign structural information near haem groups in paramagnetic proteins. For example, the orientation of the ligands can reliably be obtained from ^{13}C data and used as restraints in structure calculations which use the proton dipolar shifts to fit the paramagnetic tensor parameters. The consistency of this approach can then be validated by comparing the tensors from the structure calculations with those obtained from fitting the shifts to a crystal structure.

The procedure to fit the paramagnetic tensor requires therefore both dipolar shifts and atomic coordinates. The dipolar shifts were obtained from the difference of the experimental shifts for the oxidized state and the calculated diamagnetic shifts of the reduced state. The shifts for the reduced state were calculated using the program `total` (Prof. M. Williamson, Sheffield) and the coordinates of the crystal structure of c3*Dn*. Since these coordinates are for the oxidized form of the protein the procedure assumes that there are no changes in conformation associated with the change in redox state. The parameters for the molecular orbital analysis and the fitting of the paramagnetic susceptibility tensors are presented in Table 2.2.

A total of 225 dipolar shifts were used for the fitting of the tensor parameters to the X-ray coordinates of c3*Dn* (Figure 2.3). The uncertainties were set to ± 0.2 ppm and ± 0.5 Å for the dipolar shifts and the atomic coordinates respectively. Five shifts were rejected after statistical testing. Least squares fitting of the shifts produced a good fit to a line of slope 0.95 and an intercept of 0.014 resulting in 91.5% of variance explained. The quality of the fit is reasonable and allowed a preliminary evaluation of the tensor parameters. The Euler angle β representing the tilt of the z-axis of the tensor away from the normal to the haem plane is small for haems 1, 2, and 4 and slightly larger for haem 3. The tilt of the z-axis is expected to increase as the dihedral angle ψ decreases, *i.e.* for haems with nearly parallel axial ligands. The value of ψ for haem 3 in the X-ray structure is larger than that for haem 1 yet a larger tilt of the z-axis is calculated for haem 3 than for haem 1. The values for haem 2 and 4 are as

	haem 1	haem 2	haem 3	haem 4
θ (deg)[a]	-35.8	-16.9	72.9	-8.2
ΔE (kJ.mol^{-1})[a]	3.83	3.42	5.34	1.47
ϕ (X-ray) (deg)[a]	-32.7	-19.8	58.8	-1.1
ψ (X-ray) (deg)[a]	5.0	50.0	25.6	77.2
ψ (NMR) (deg)[b]	24.2	50.5	6.1	75.4
α (deg)[c]	-0.8 (56.7)	37.3 (17.5)	89.9 (6.8)	93.9 (17.1)
β (deg)[c]	-3.1 (2.7)	-5.8 (2.4)	17.4 (5.2)	-6.3 (1.5)
γ (deg)[c]	36.9 (53.3)	-15.7 (20.2)	28.8 (8.1)	-86.0 (18.2)
$\alpha+\gamma$ (deg)[c]	36.1 (7.9)	21.6 (6.5)	118.7 (4.3)	7.9 (5.5)
$\Delta\chi_{ax}.10^{32}$ (m^3)[c]	2.82 (0.24)	3.20 (0.20)	3.07 (0.28)	4.63 (0.24)
$\Delta\chi_{eq}.10^{32}$ (m^3)[c]	-1.34 (0.25)	-1.50 (0.32)	-2.41 (0.28)	-1.31 (0.28)

Table 2.2: Molecular orbital and empirical magnetic susceptibility tensor parameters for c3*Dn*. [a] From [39]; [b] calculated from Equation 2.4 [c] calculated from the c3*Dn* X-ray coordinates, 2cy3.pdb [114]. Standard deviations are given in parentheses.

expected however. Similarly, the sum $\alpha+\gamma$ which approximates the in-plane rotation of the y-axis of the tensor can be compared with both the negative value of θ and the ligand orientation ϕ [40]. As explained in the paper the correlation of $\alpha+\gamma$ has better agreement with the negative value of θ than with the value for ϕ as seen for the data for c3*Dn*. For $\alpha+\gamma$ the values for haems 1, 2 and 4 agree well while the values for haem 3 are not in agreement. This suggests that either the input data require improvement or possibly that the X-ray structure is different than the solution structure for haem 3. Such differences can of course be expected when incomplete data measured in solution are fitted to X-ray coordinates.

This preliminary analysis confirms that the dipolar shifts can be reliably extracted from the experimental data and the derived tensor parameters are consistent with earlier published results. However tests of the structure calculation incorporating dipolar shifts resulted in structures which converged but had extremely high target functions. Therefore these data are not included in the calculated structures analyzed in the following sections.

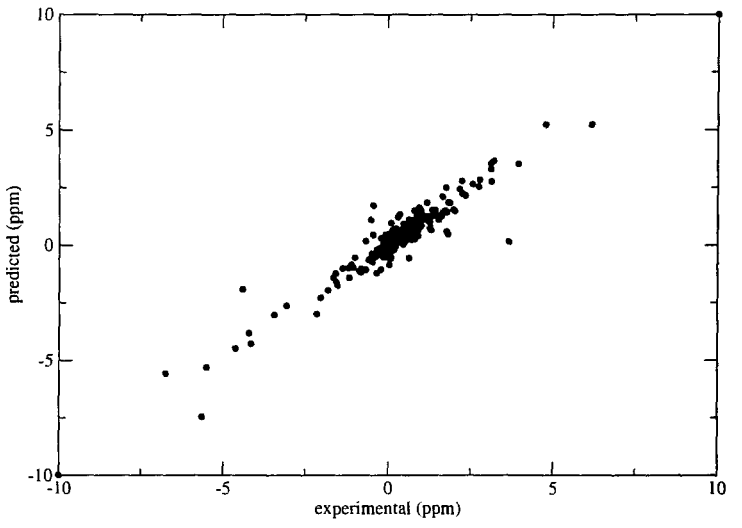

Figure 2.3: Correlation of predicted and experimental dipolar shifts for c3*Dn*.

2.3.3 Solution structure calculation

The input data for the structure calculations were obtained from the assigned and integrated peaks from the 25 and 100 ms spectra in H_2O and D_2O (Figures 2.4, 2.5). The peak volumes were converted into volume restraints using PARADYANA. The 100ms spectra generated 980 lower volume limits and 818 upper volume limits. The 25 ms spectra generated an additional 103 lower volume limits and 103 upper volume limits. An extra set of 56 fixed upper distance limits were used to maintain the haem and proline geometries. Since the volume calibration is a fully automatic procedure identification of redundant restraints must wait until after the calculation has been performed. Medium range restraints are between amino acids which are separated by between two and four positions in the protein sequence. Long range restraints are between amino acids which are separated by five or more positions in the sequence. Redundant distance restraints are removed by DYANA before the calculation begins and lower distance restraints below the van der Waals radii are also removed. Upper and lower limits from the same NOE were counted as one. Haem-haem and intra-haem restraints were excluded from the count. One hundred random conformers

were submitted to the annealing schedule and the structures with the 20 lowest values of the target function were accepted as being representative of the solution structure of the protein.

Type of volume restraint	lower volume restraint	upper volume restraint
A. *100ms mixing time*		
intra-residue	319	310
sequential	226	193
medium range	192	131
long range	243	184
B. *25ms mixing time*		
intra-residue	0	0
sequential	4	4
medium range	30	30
long range	69	69
C. *Total*	1139	921
D. *Per residue*	9.7	7.8

Table 2.3: Restraints used for the structure calculation of c3*Dn*.

2.3.4 Quality analysis of the calculated structures

The NMR structure ensemble could be described as medium resolution (Figure 2.6). A summary of the structure factors for the NMR ensemble is given in Table 2.4. The target function values of the ensemble members range from 12.93 to 14.31 Å^2.

The structures were superimposed on both the set of backbone heavy atoms and the set of complete heavy atoms. The average RMSD relative to the coordinates of the mean structure is 1.74 Å for the backbone heavy atoms of all residues. These data are plotted in Figure 2.7 where it can be seen that some regions of the structure have a much higher RMSD relative to the remainder. There are few assignments for these regions leading in turn to high variability in the calculated structures. For example, the RMSD of the backbone heavy atoms is reduced to 1.07 Å when residues 14-27 in

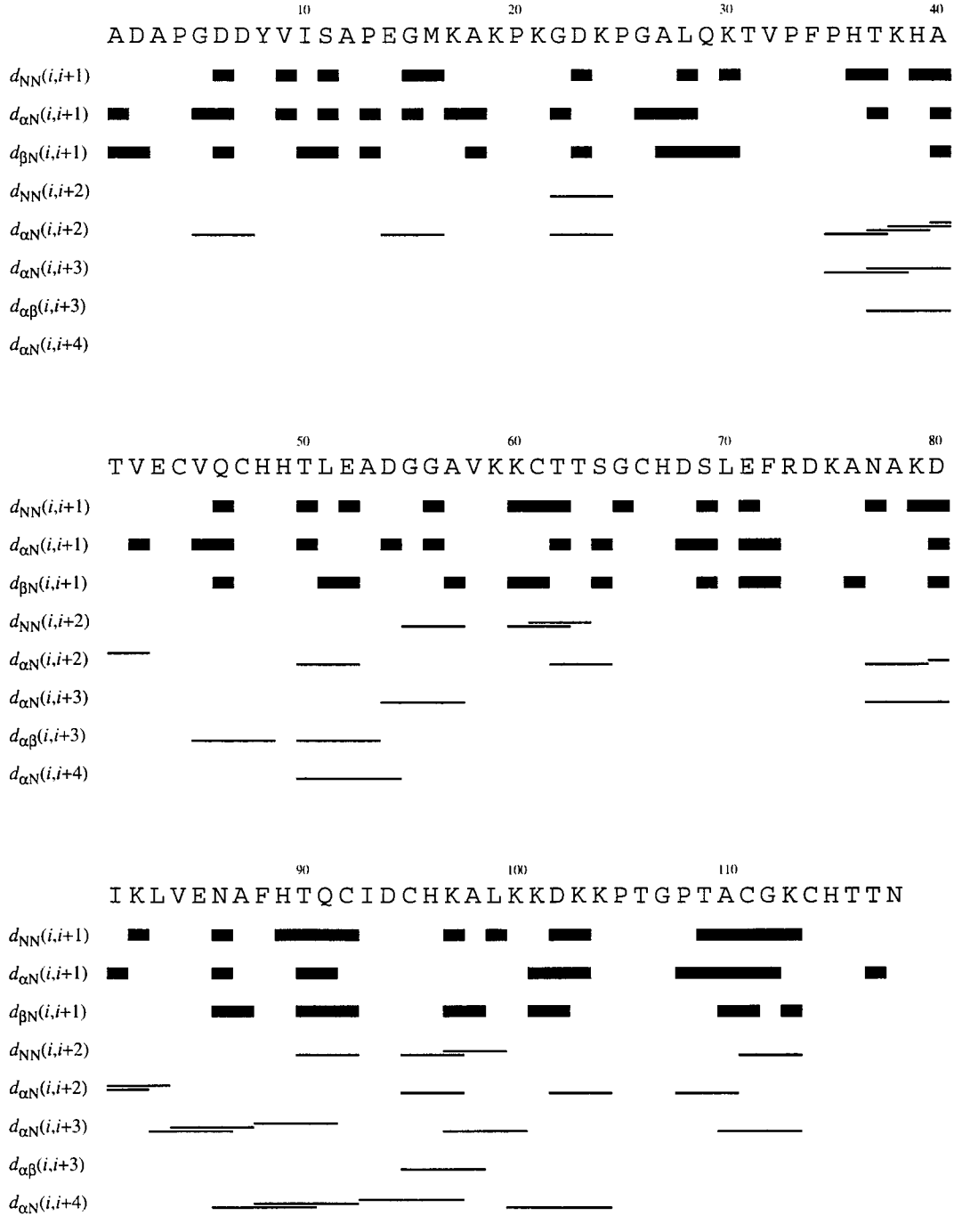

Figure 2.4: Summary of the sequential and medium range NOEs for c3*Dn*. The thickness of the lines correlates with the intensity of the NOE.

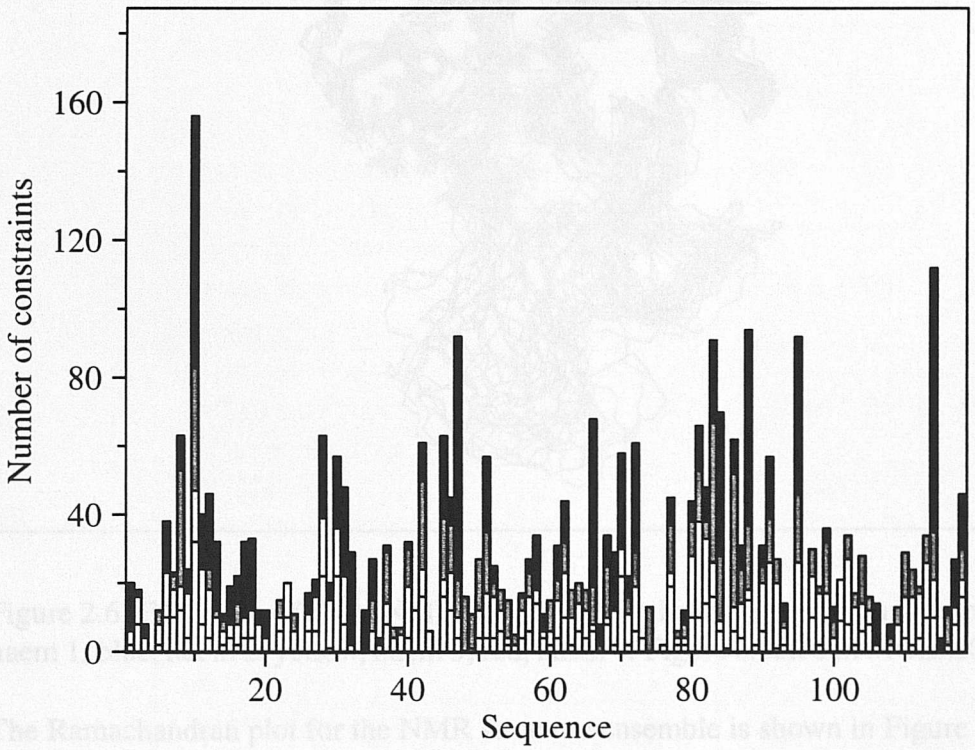

Figure 2.5: Summary of the number of NOE restraints per residue. The bars are shaded to represent the different classes of restraints as follows: white, intra-residue; light grey, sequential; dark grey, medium range; black, long range.

the large variable loop are excluded. The overall RMSD as well as that in the unassigned regions could be reduced with further NOE assignments.

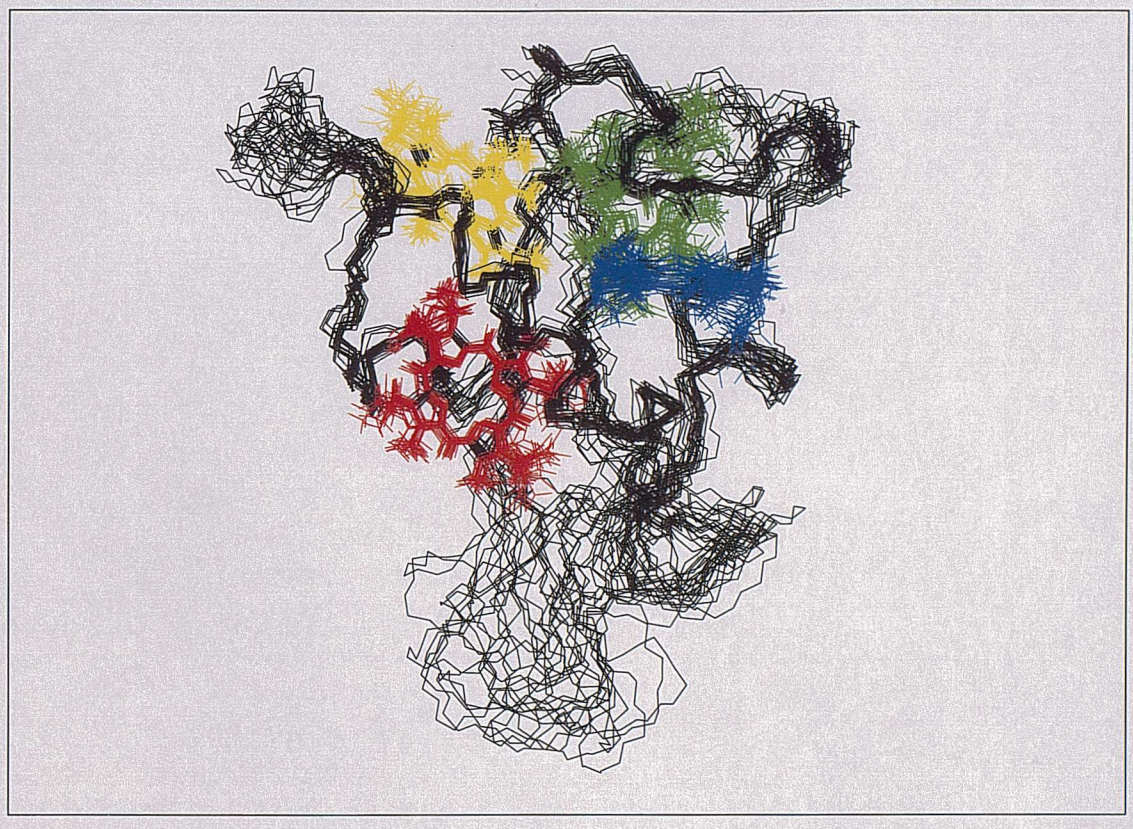

Figure 2.6: Ensemble of c3*Dn* NMR structures with haem group colour coding: green, haem 1; blue, haem 2; yellow, haem 3; red, haem 4. Figure created in MOLMOL [137].

The Ramachandran plot for the NMR structure ensemble is shown in Figure 2.8. The percentages of residues in the different regions of the Ramachandran plot are also in Table 2.4. These compare with the values for the crystal structure of 82.8% and 17.2% in the most favoured and additionally allowed regions respectively. The crystal structure had no non-glycine residues in the generously allowed and disallowed regions. Since the structure is at an intermediate stage of refinement it is expected that these values will converge to the values expected for NMR structures as the refinement is continued.

Bad contacts were calculated using PROCHECK [87]. Bad contacts were defined if the atoms are separated by more than four bonds and the interatomic distance ≤ 2.6Å. For the lowest energy NMR structure in the ensemble 37 bad contacts were found.

Quantity	
A. *DYANA target function (type 1)*	
average total (Å^2)	13.70 ± 0.45
B. *Upper distance limit violations*	
average maximum (Å)	1.14 ± 0.03
number of consistent violations (> 0.2Å)	6
C. *Lower distance limit violations*	
average maximum (Å)	0.73 ± 0.05
number of consistent violations (> 0.2Å)	7
D. *Van der Waals violations*	
average maximum (Å)	0.33 ± 0.05
E. *Ramachandran plot (%)*	
most favoured	34.3
additionally allowed	44.4
generously allowed	16.2
disallowed	5.1
F. *Packing*	
bad contacts	37
G. *Precision*	
RMSD backbone (Å)	1.74

Table 2.4: Summary of restraint violations and quality analysis for the final structure ensemble of oxidized c3*Dn*.

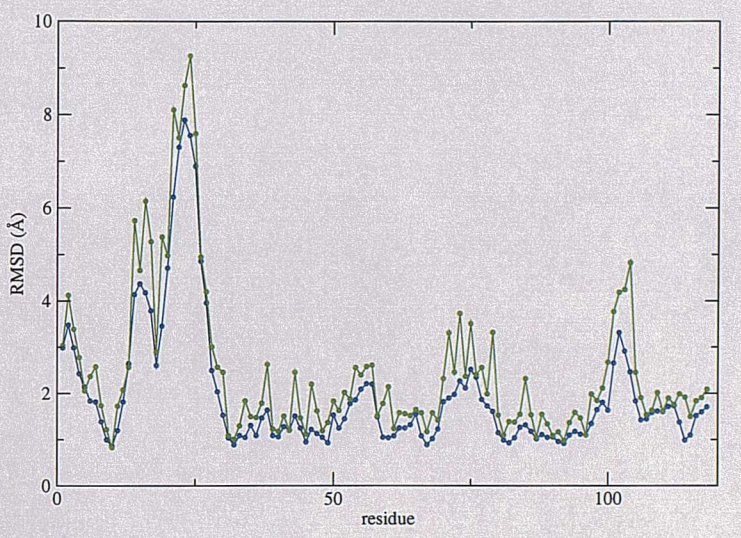

Figure 2.7: Average backbone (blue) and heavy atom (green) RMSD values per residue with respect to the mean structure of the NMR ensemble of c3*Dn* structures.

This number is reasonable when compared with 16 bad contacts found for the crystal structure. Again, it is expected that this value will be reduced with further refinement.

2.3.5 Comparison with the crystal structure

Comparison of the X-ray and NMR structures of c3*Dn* confirms that the general fold and the haem positions are reproduced. The average iron-iron distances were calculated using COFIMA [76] and are similar to those in the X-ray structure as reported in Table 2.5. The RMSD for the backbone heavy atoms in the X-ray structure and the lowest energy NMR structure is 3.29 Å over all residues (excluding the second cysteines coordinating each haem due to coordinate file incompatability for these residues). In general cytochrome c_3 structures display very little secondary structure. This is because the four haem groups are quite bulky and the protein backbone can only barely wrap itself around the the haems to ligate them. There is one notable secondary structure element in the NMR ensemble which is an α-helix

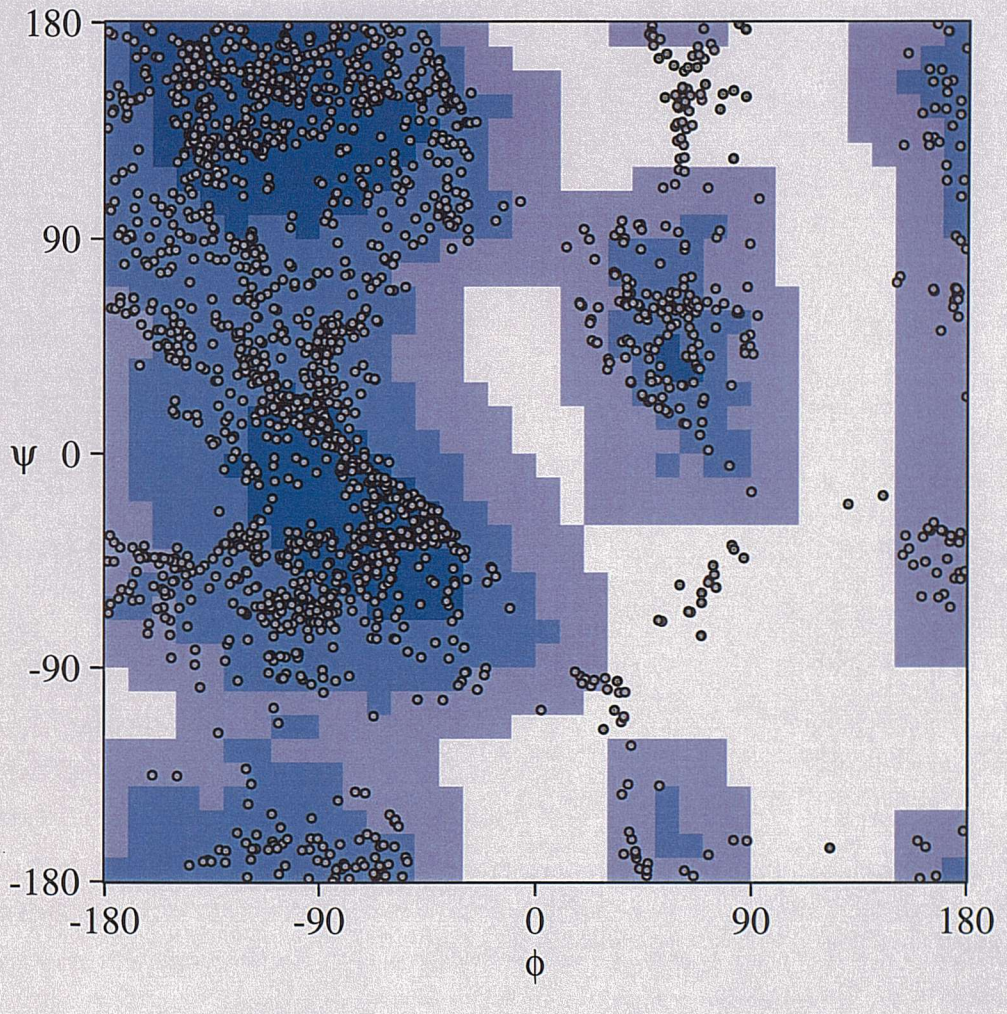

Figure 2.8: Ramachandran plot for the ensemble of the 20 lowest energy c3*Dn* structures calculated by PARADYANA. Backbone torsion angles for all non-glycine, non-proline, and non-terminal residues are included.

which extends from L83 to A98. This helix is shorter than in the X-ray structure in which it extends to K101. The probable reason for this is that the loop from L99 to T106 is variable in the NMR ensemble. It is possible that with some more restraints for residues L99 to K101 the helix would be extended.

haem		II	III	IV
I	NMR	11.67 (0.38)	12.51 (0.60)	17.96 (0.47)
	X-ray	12.45	11.06	17.10
II	NMR		16.45 (0.25)	16.49 (0.26)
	X-ray		16.65	15.86
III	NMR			12.97 (0.19)
	X-ray			12.66

Table 2.5: Average iron-iron distances (Å) in the NMR and X-ray structures of c3*Dn*. Standard deviations are given in parentheses.

2.4 Concluding remarks

This chapter presents the work done on the NMR structure of c3*Dn* as of January 2002. The structure is at an intermediate level of refinement compared to the previous c_3 structures. However the main features of the structure of the protein in solution have been confirmed and compared to the crystal structure.

The parameters of the paramagnetic tensors of the haem groups in c3*Dn* have been investigated based on the results of earlier work on cytochromes c_3. Dihedral angle restraints from the combined haem ^{13}C NMR and molecular orbital analysis have been determined and can be included in future structure calculations.

The results illustrate that the most important factor in the success of a structure determination project by NMR is the completeness of the resonance assignment. As noted this stage took up the majority of the time spent on this project. There are still some assignment ambiguities, notably in the solvent exposed loops. In this respect progress could be made by the recording data at higher field strengths and the preparation of a ^{15}N-labelled sample. Additional work remaining includes increasing

the general quantity of data used as input for structure calculations especially from the short mixing time spectra. This data will have to be evaluated carefully since at short mixing times paramagnetic effects on crosspeak volumes are attenuated and internuclear distances can be derived with greater accuracy. Further data which can also be included are pseudocontact shifts and stereospecific assignments.

2.5 Acknowledgements

The research staff at the laboratory of Prof. Antonio Xavier at the ITQB Lisbon produced the sample and recorded the spectra. In particular, Dr. Lorraine Brennan recorded a very high quality NOESY spectrum which was indispensable for the progress described. I would also like to personally thank Prof. Xavier for a very kind welcome during a working visit in August 1999. Financial support was provided by the EU from a TMR network award.

Chapter 3

Assignment of the HU protein from *Thermotoga maritima*

3.1 Introduction

The protein HU is found in prokaryotic organisms and serves a role as a multipurpose DNA binding factor. HU is closely related to IHF (integration host factor) and the phage protein TF1 (transcription factor 1) [143, 144, 145]. There are both significant similarities and differences between these proteins in terms of both structure and function. HU binds DNA non-specifically while both IHF and TF1 bind to recognized consensus sequences. The common theme in the mode of action of these proteins is the bending of the DNA for the purpose of general chromosomal compaction or more specific roles in the cell [146]. While IHF is always found as a heterodimer both HU and TF1 can exist as either homo- or heterodimers. The eukaryotic equivalents of these bacterial type II DNA-binding proteins are the high mobility group (HMG) domain proteins which have completely different structures but similar roles as architectural factors. Indeed, the striking structural similarities of HU, IHF, and TF1 when considered with their functional differences allows a detailed comparison of specific and non-specific protein-DNA complexes.

The structures of members of the IHF-HU family are very similar and can be described as a compact body from which two β-ribbon arms extend. The body is formed by several intertwined α-helices capped by a pair of three-stranded β-sheets. The hydrophobic core between this cap and the helices contains a cluster of phenylalanine residues which are highly conserved. The β-ribbon arms bind and bend DNA by wrapping around the substrate in both specific and non-specific complexes [147, 148]. In the absence of DNA the β-ribbon arms of HU are highly flexible yet internally structured and become less mobile in complex with DNA [149]. Several structures of HU have been determined including: X-ray [150, 151] and NMR [152] structures of HU from *Bacillus stearothermophilus* (HU*Bst*); an X-ray structure of HU from *Thermotoga maritima* (HU*Tmar*) [153]; and very recently the X-ray structure of a HU-DNA complex from *Anabaena* [148].

Ultrastable proteins have been the subject of many recent biophysical studies aimed at understanding the relationship between flexibility and stability in these systems [154, 155]. *Thermotoga maritima* is a hyperthermophilic bacterium originally found in geothermally heated marine sources in Italy. Consequently, the constituent proteins

of this organism are highly thermostable. HU*Tmar* (T_m = 80.5°C) has been the focus of studies aimed at determining the contributions to thermostability at an amino acid level [156]. The X-ray structure of HU*Tmar* [153] confirms that the structure is conserved and also that no electron density was found for the flexible β-ribbon arms. The origin of the protein is reflected in its unusually high affinity temperature-independent DNA binding properties which have allowed the development of biotechnological applications as an effective gene delivery agent [157, 158]. The aim of this project was to determine the structure and dynamics of HU*Tmar* in solution by NMR methods and to compare this with HU*Bst*.

3.2 Materials and methods

3.2.1 Sample preparation

Cloning of the HU*Tmar* gene and overproduction of the wild type protein was performed as described in the literature [159, 160]. Labelled protein was produced using ^{15}N-NH_4Cl and/or ^{13}C-glucose in minimal medium. The samples were purified by affinity chromatography on heparin sepharose and mono-S FPLC. The purity was over 95% according to SDS-PAGE. The recombinant HU*Tmar* retains the DNA binding properties observed for samples purified from the bacterium. The protein was suspended in a 50mM phosphate buffer with 200mM potassium chloride at pH 5.8. Addition of the protein to the buffer solution changes the pH to approximately 6.2 since the protein is concentrated and highly positively charged. NMR experiments were performed at a temperature of 311K with one ^{15}N-labelled sample (1 mM) and one ^{13}C-^{15}N-labelled sample (1 mM).

3.2.2 NMR spectroscopy

NMR spectra were recorded on Bruker Avance spectrometers operating at proton frequencies of 500, 600, 700, and 750 MHz at a temperature of 311K. All of the

spectrometers are equipped with a triple-resonance probe with a shielded gradient coil. The 2D (^{15}N and ^{13}C) spectra were recorded to give an overview of the chemical shift ranges of the heteronuclei [161].

The constant-time versions of the HNCO, HNCA, and HN(CO)CA experiments were recorded essentially as described by Grzesiek and Bax [162]. The HN(CA)CO experiment was recorded as desribed by Clubb and co-workers [163]; a refocused INEPT step was used for the magnetization transfer from ^{1}H to ^{15}N for the HNCA and the HN(CA)CO experiments. The CBCA(CO)NH and HBHA(CBCACO)NH experiments were recorded as described [164]. The CC(CO)NH-TOCSY and the HCCH-TOCSY experiments experiments were implemented as described [165, 29]. Efficient water suppression was achieved by the use of pulsed field gradients in all experiments. In most of the experiments the carrier frequencies were positioned to obtain unfolded spectra.

All NMR data were processed using NMRPipe [166]. The program NMRView was used to analyze the spectra [167].

3.3 Results and discussion

3.3.1 Sequential assignment

A sequence alignment of the HU proteins from several organisms is presented in Figure 3.1. Almost identical secondary structure elements have been observed in all of the HU structures determined to date, including the crystal structure of HU*Tmar* [153]. The conserved topology is composed of a helix-turn-helix motif (α1 and α2); β-strands β1, β2, and β3; the β-arms and the C-terminal α-helix α3. The overall structure of the HU dimer is presented in Figure 3.2. Comparisons of this structure with the sequence shows how the α1 and α2 helices form the compact core of the dimer which is capped by the β-strand formed from β1, β2, and β3. The C-terminal α3 is also involved in intermonomer contacts with α2. The β-arms do not contribute

to the dimer interface but extend from the core and are very mobile yet internally structured in the free protein (Figure 3.2). The HU structure has therefore evolved to maximize the usage of a relatively short primary sequence to achieve considerable structural efficiency.

The starting point for the analysis was an examination of the ^{15}N HSQC spectrum (Figure 3.3). Excluding side chain NH_2 groups, 84 correlations were expected in this spectrum since each monomer has 90 residues and no peaks will be observed for the N-terminal methionine and the five prolines.

Surprisingly, approximately 140 peaks were counted. At first it was thought that the extra resonances could be due to slow conformational exchange of the flexible β-ribbon arms of HU and the sequential assignment process was started. As the assignment progressed it became clear that the extra resonances could be assigned to residues throughout the protein and that two approximately equally populated conformations of the protein exist under the NMR sample conditions used.

The conformational equilibrium persists over many months as evidenced by the HSQC spectra recorded at regular intervals. This idea was tested by varying the experimental conditions as much as possible. The pH, salt concentration, and protein concentration were all varied with no significant change in the spectrum. One sample of the protein was thermally denatured and was observed to refold into the same conformational equilibrium. A temperature series was also recorded in the range 290-330 K and while certain key residues do undergo chemical shift changes the overall ratio of the populations of the two conformations was not significantly perturbed.

Eventually the backbone assignment of the two conformers was completed. In some segments of the sequence the two conformers have very different chemical shifts for the backbone nuclei while in other segments the protein may only be in one conformation or two conformations could have degenerate same chemical shifts.

The sequential assignment followed the standard procedure of correlating shifts from a given residue with shifts from the preceding residue depending upon which

```
Tmar -----MTKKELIDRVAKKAGAKKKDVKLILDTILETITEALAKGEKVQIVGFGSF 50
Bst  -----MNKTELINAVAETSGLSKKDATKAVDAVFDSITEALRKGDKVQLIGFGNF 50
Bsu  -----MNKTELINAVAEASELSKKDATKAVDSVFDTILDALKNGDKIQLIGFGNF 50
Tth  AAKKTVTKADLVDQVAQATGLKLLDVKAMVDALLAKVEEALANGSKVQLTGFGTF 55
Tvo  ----MVGISELSKDVAKKANTTQKVARTVIKSFLDEIVAQANGGQKINLAGFGIF 51
          : : :*:  **   :  :::     :     ::::: :* *::  *** *

Tmar EVRKAAARKGVNPQTRKPITIPERKVPKFKPGKALKEKVK-- 90
Bst  EVRERAARKGRNPQTGEEMEIPASKVPAFKPGKALKDAVK-- 90
Bsu  EVRERSARKGRNPQTGEEIEIPASKVPAFKPGKALKDAVAGK 92
Tth  EVRKRKARTGVKPGTKEKIKIPATQYPAFKPGKALKDKVK-- 95
Tvo  ERRTQGPRKARNPQTKKVIEVPSKKKFVFRASSKIKYQQ--- 90
     *:*  ::*:  :*:*  :*  ::: *:::::::*  ::
```

Figure 3.1: ClustalW [168] alignment of HU sequences from *Thermotoga maritima* (Tmar), *Bacillus stearothermophilus* (Bst), *Bacillus subtilis* (Bsu), *Thermus thermophillus* (Tth) and *Thermoplasma volcanium* (Tvo). Asterisks indicate residues conserved in all five sequences and colons indicate residues additionally conserved between HU*Tmar* and HU*Bst*.

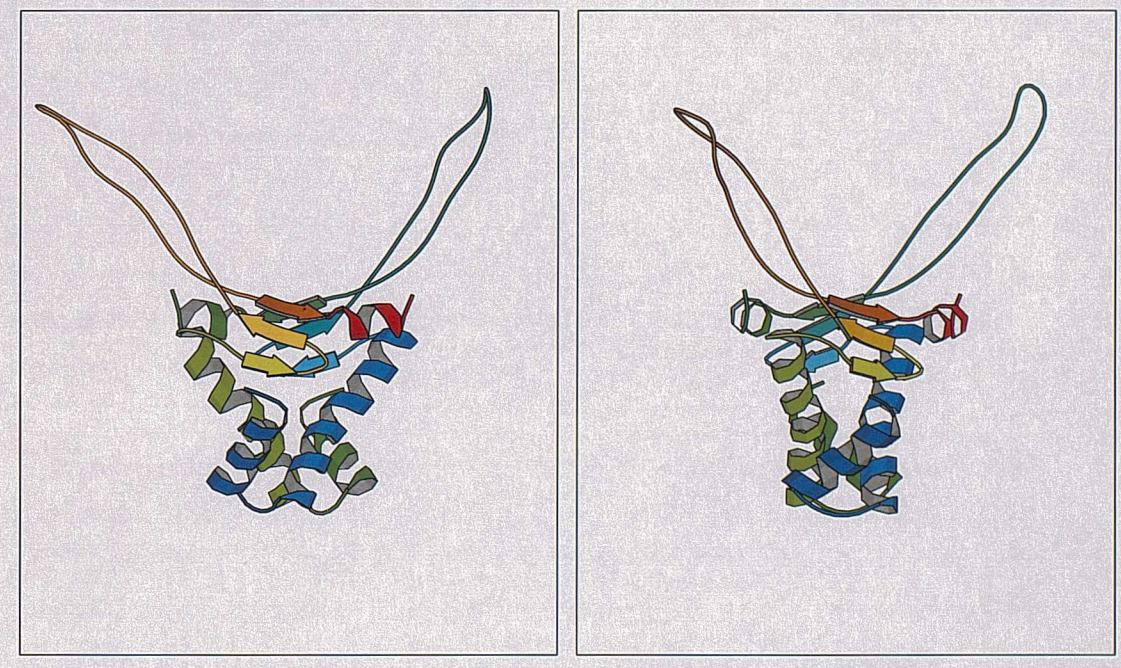

Figure 3.2: Alternate views (rotated around z-axis by 45°) of one structure from the NMR ensemble (1hue.pdb [152]) of HU*Bst*. Figure created using MolScript [169].

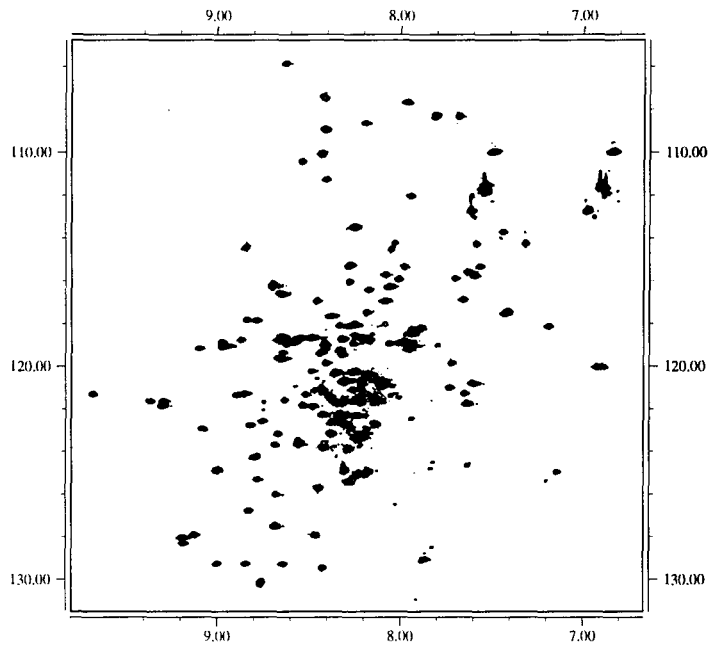

Figure 3.3: 2D ^{15}N HSQC spectrum of uniformly ^{15}N-labelled HU*Tmar* recorded at 750 MHz proton frequency and 311 K.

spectrum is examined [18, 25, 29, 170]. As segments of sequential connectivities are built up the amino acid types can be determined by the chemical shifts of the Cα and Cβ nuclei and the segments can be placed in the protein sequence. The CC(CO)NH-TOCSY spectrum gives additional information on the amino acid types and, in principle, by following this procedure an assignment can be achieved and confirmed very quickly. The backbone assignment of HU*Tmar* was however complicated by the equilibrium between the two conformations.

The first spectra examined in the assignment procedure were the complementary HNCA and HN(CO)CA which were used to establish the intraresidue and sequential $^{13}C^{\alpha}$ resonances. The HNCA spectrum contains both the intraresidue $^{13}C_i^{\alpha}$ and the sequential $^{13}C_{i-1}^{\alpha}$ resonances for each ^{15}N-NH crosspeak while the HN(CO)CA spectrum contains only the sequential $^{13}C_{i-1}^{\alpha}$ resonances. A comparison of the peaks in these two spectra yields the resonance frequencies of these nuclei for all the peaks observed in the ^{15}N HSQC spectrum.

In principle the HNCA and HN(CO)CA pair of spectra can provide the backbone assignment for the complete protein sequence and therefore correlate the peaks in the ^{15}N HSQC with each other and also with the intraresidue $^{13}C_i^{\alpha}$ resonances which will be used to assign the amino acid sidechain resonances. In the event of overlap or missing correlations in these spectra the missing assignments can usually be obtained from an analysis of the HNCO and HN(CA)CO spectra which are complementary to the HNCA and HN(CO)CA pair. The HNCO spectrum correlates each ^{15}N-NH crosspeak with the sequential $^{13}CO_{i-1}$ carbonyl resonance. This spectrum is analysed in combination with the HN(CA)CO which correlates each ^{15}N-NH crosspeak with the intraresidue $^{13}CO_i$ and sequential $^{13}CO_{i-1}$ carbonyl resonances. In combination these four spectra should allow the assignment of the complete 2D ^{15}N HSQC spectrum. Proline residues in the sequence can also be assigned in this way but any assigned segment will end at the proline since there is no corresponding crosspeak in the ^{15}N HSQC spectrum from which to continue the assignment.

The next stage of the assignment procedure was the identification of the amino acid types in the sequentially assigned segments obtained from the backbone correlation spectra. This identification is usually done in parallel with the sequential assignment depending on the complexity of the task which in turn depends on the size of the protein or the types of sequence segments present. The position of a crosspeak in the ^{15}N HSQC spectrum is an indication of the type of amino acid present for a limited subset of residue types. For example, glycine residues usually have an upfield shifted ^{15}N resonance. An analysis of the $^{13}C^{\alpha}$ shifts though will provide a much clearer indication of the amino acid types. If the $^{13}C^{\alpha}$ shifts can be combined with the $^{13}C^{\beta}$ shifts the identification of amino acid type is usually unambiguous and combined with the positions of the residues in the segments the assignment can be confirmed beyond doubt. This amino acid identification step is based on the statistical distribution of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shifts. The $^{13}C^{\beta}$ shifts were obtained from the CBCA(CO)NH spectrum which correlates the $^{13}C_{i-1}^{\alpha}$ and $^{13}C_{i-1}^{\beta}$ resonances with the ^{15}N-NH crosspeaks. Additionally the complementary HBHA(CBCACO)NH spectrum allowed the assignment of the corresponding $^{1}H_{i-1}^{\alpha}$ and $^{1}H_{i-1}^{\beta}$ resonances.

The assignment of the dual conformations of HU*Tmar* was achieved using the spectra recorded as described above (Section 3.2.2). The actual assignment procedure was

relatively straightforward once the source of the extra peaks was established. The availability of the complementary spectra was essential for the assignment. For example, for certain segments of the sequence there was overlap of the resonance frequencies for all three nuclei. This is especially evident in certain regions of the HNCA spectrum. Although there was sufficient resolution in the proton and nitrogen dimensions for the majority of residues this was not always the case for the carbon resonances. This was noted mainly for lysine and arginine residues in the region centered around carbon frequencies of 52 ppm. There are 20 lysine and 5 arginine residues in each conformation of HU*Tmar*. In these cases the HNCO spectrum allowed the unambiguous assignment of the resonances to the correct conformation. This is due mainly to the greater chemical shift dispersion of the ^{13}CO resonances compared to the $^{13}C^{\alpha}$ resonances. Specific examples of this are discussed below (Section 3.3.3)."

Apart from the minor overlap problem the sequential assignment was achieved in a timely manner. Complete or partial assignments were made for all residues in form A and 75 residues in form B (Table B.1). No assignments were found for 15 residues in form B, namely: K17, K18, V21, K22, I44, V45, G46, F47, F50, E51, P77, K78, F79, K80, K83."

3.3.2 Side chain assignment

Once the backbone and $^{13}C^{\beta}$ and $^{1}H^{\beta}$ sidechain resonances were assigned the next stage was to continue to assign as many as possible of the sidechain carbon and proton resonances. This stage was achieved through the use of TOCSY type experiments. The 2D ^{13}C HSQC spectrum (see Figure 3.4) gives an overview of the range of ^{13}C chemical shifts present. The backbone resonances were used as starting points in the analysis.

The HCCH-TOCSY experiment can be recorded in two different modes depending on whether two proton (HcCH) or two carbon (hCCH) dimensions are present in the resulting spectrum. The spectra recorded in both modes are essential for the assignment of the remaining sidechain carbon and proton resonances. These

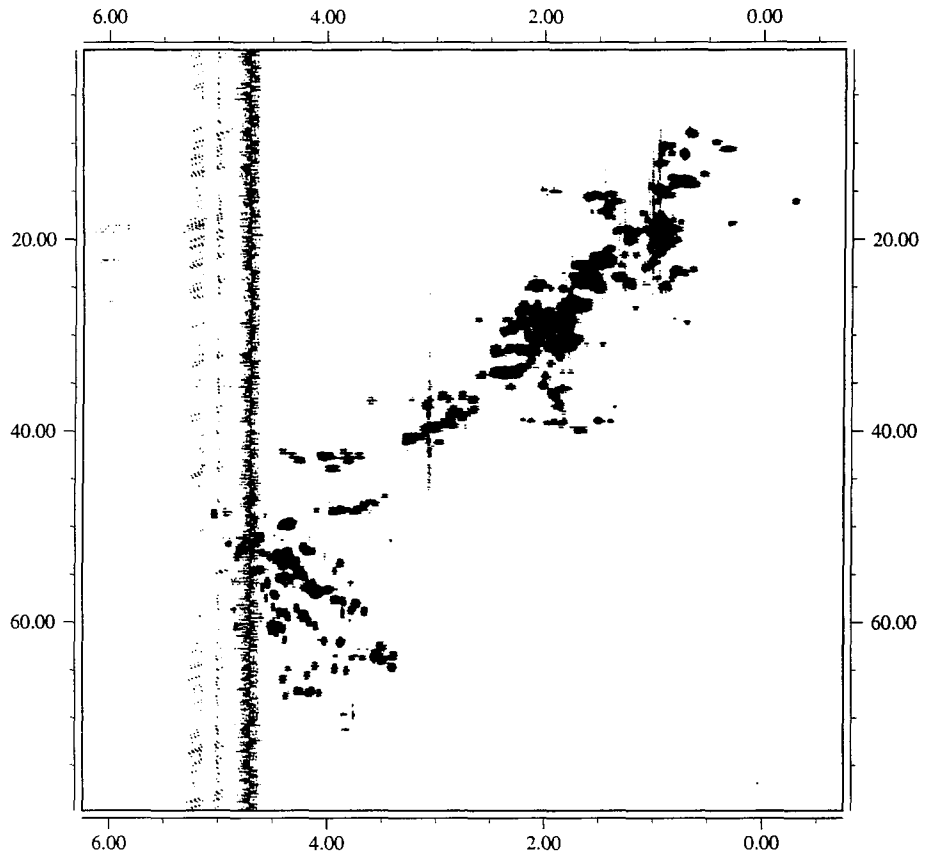

Figure 3.4: 2D ^{13}C HSQC spectrum of uniformly ^{13}C-^{15}N-labelled HU*Tmar* recorded at 600 MHz proton frequency and 311 K.

complementary spectra are used to correlate the sidechain proton resonances with the directly attached carbon resonances to obtain the complete assignment for each amino acid. A knowledge of the chemical shift ranges for the different sidechain resonances is essential for this stage. The statistically averaged shifts of the different nuclei in the various amino acid types has been compiled in the BMRB and can give very reliable indications.

The carbon resonances for the complete sidechain of the $i-1$ residue can be obtained from the CC(CO)NH-TOCSY spectrum which provided an ideal starting point for the analysis of the HCCH-TOCSY spectra. The complete proton resonances of the sidechains could be assigned by an experiment similar to the CC(CO)NH-TOCSY by

changing the ^{13}C evolution period to a ^{1}H period with an appropiate magnetization transfer. However the assignment of specific protons to their attached carbon nuclei requires an analysis of HCCH-TOCSY spectra. This stage is important if a complete analysis of NOESY spectra is required for a structure determination. These spectra were analyzed by picking peaks in the common ^{13}C planes and correlating the chemical shifts with those known from the backbone assignment.

The sidechain amide nitrogen and proton resonances were assigned through a combination of the 2D ^{15}N HSQC and the CBCA(CO)NH spectra. In the ^{15}N HSQC spectra pairs of signals with the same ^{15}N shift correspond to the $N^{\delta}H_2$ and $N^{\varepsilon}H_2$ resonances of asparagine and glutamine residues respectively. These resonances were conveniently assigned in the CBCA(CO)NH spectrum. The CBCA(CO)NH spectrum was searched at the nitrogen frequencies of the NH_2 groups. For the simpler case of the asparagine residues four symmetrically displaced peaks were observed corresponding to the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances at the two NH_2 frequencies. For the glutamine residues four peaks were again assigned corresponding to the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ resonances at the two NH_2 frequencies. In this case the CBCA(CO)NH spectrum could provide an extra confirmation of the $^{13}C^{\gamma}$ assignments.

3.3.3 Analysis of the dual assignment

As HU*Tmar* is conformationally heterogeneous under NMR conditions the assignment procedure was considerably complicated and eventually a dual assignment was found for a majority of the sequence. This section presents the details of the dual assignment and describes how the two assignments are internally consistent yet mutually distinct. There are many examples described in the literature of proteins which display localized slow conformational exchange under NMR conditions. However it is very unusual to observe slow conformational exchange for such a large majority of residues in the sequence. In addition the exchange is between two approximately equally populated native-like conformational substates rather than between major-minor, folded-unfolded or complexed-uncomplexed conformations which is frequently observed [171, 172, 173, 174]. Isomerization of *cis-trans* prolyl peptide bonds or changes in the conformations of disulfide bridges are also well

known causes of conformational exchange in proteins in solution [175, 176]. This slow exchange raises the questions of why are the conformations almost equally populated and what are the structural determinants of this exchange. A description of the kinetic and thermodynamic barriers separating the two conformations would be required to explain the exchange process. NMR is the most suitable method for investigating such exchange processes because of the availability of information at residue level.

The segments of the sequence for which dual assignments including resolved resonances in the ^{15}N HSQC spectrum were found are: 2-9, 11-13, 15, 16, 20, 24-43, 53-62, 64-67, 69-71, 73, 75, 76, 82, 85-87, 89, 90. In addition dual connectivities corresponding to residues 10, 23, 63, 68, 72, 74, 81, 84 were assigned in the HNCA spectrum but were either overlapped or had exchanged amide protons in the ^{15}N HSQC spectrum. A few residues could only be unambiguously assigned in the HNCO spectrum. The internal consistency of each assignment was checked as far as possible in the *J*-correlated and NOESY-HSQC spectra.

The segment M1-K13 was unambiguously assigned from the HNCA spectrum. The signals for residues A14 and G15 in both forms are almost completely overlapped in the ^{15}N HSQC although separate peaks can be discerned for each in the HNCA and HNCO spectra. The degeneracy is broken for residue A16 allowing the correct assignment of each resonance for this residue. The degeneracy continues for the segment K17-K22 for which only one correlation could be assigned in the ^{15}N HSQC. The segment L23-K41 was unambiguously assigned solely from the HNCA spectrum. This segment was extended by the assignment of a sequential NOE to V42 in form A. In this segment the two assignments do show some important differences in that certain key residues are present only in one conformation or are very weak or broadened in the other conformer. Examples of these are the residues Q43, I44, V45, G46, F47, G48, S49, and F50. A very weak peak was assigned for residue Q43 in form B while residues I44-V52 could only be assigned in form A. Residue G46 could be assigned in the ^{15}N NOESY-HSQC spectrum only where unambiguous NOEs to V45 and F47 were observed. This segment of the sequence contains the highly conserved GFGXF consensus sequence (the dimerization signal) which creates a hydrophobic core between the two HU monomers allowing the formation of a very

stable homodimer. Interestingly, this same subset of residues has long range NOEs to different segments of the sequence. The assignment of both conformations was confirmed unambiguously for the segment R53-V76 from a combination of the HNCA and NOESY spectra. This segment contains proline residues P63, P68, and P72 which break the sequential connectivity in the HNCA and HNCO spectra. Therefore the consistency of this segment was confirmed by sequential and long range NOE assignments. The degeneracy of the two conformations resumes for residues P77-K80. For the purposes of our analysis it was crucial to obtain the correct assignment of the two conformations for the final segment P81-K90. This was achieved through a combination of information from the correlation and NOESY spectra. For these residues form A displays convincing chemical shift homology with the equivalent residues in HU*Bst*. In summary the unambiguous dual assignments of the two conformations could be achieved only by a combination of all of the available chemical shift data.

Major chemical shift changes are observed for almost all amino acids that were assigned in two conformations (Figure 3.5). Note that significant chemical shift differences are observed for residues in the β-arms and the C-terminus. These chemical shift data will be discussed in the next chapter in the context of the backbone mobility of HU*Tmar*. Note also that the chemical shifts of amide pairs in proteins are more reflective of three-dimensional structure rather than secondary structure (with the proviso that hydrogen-bonding in α-helices and β-sheets can affect the shifts of the amide proton considerably). In addition, conformational exchange or unfolding events can affect these shifts and these possibilities are discussed below.

The assignments of HU from *Bacillus stearothermophilus* [177] were very useful as starting points since the conserved residues in the form A of HU*Tmar* have very similar chemical shifts to those of HU*Bst* (Figure 3.1). Form A of HU*Tmar* has higher chemical shift homology with the assignments for HU*Bst*. Furthermore, certain key residues could only be assigned in the major form as discussed below. The chemical shift homology combined with long-range NOESY assignments and relaxation data (see the following chapter) indicate that form A is much more similar to both the X-ray structure of HU*Tmar* and the NMR and X-ray structures of HU*Bst* than form B. We propose therefore that form A is the native structure in solution and

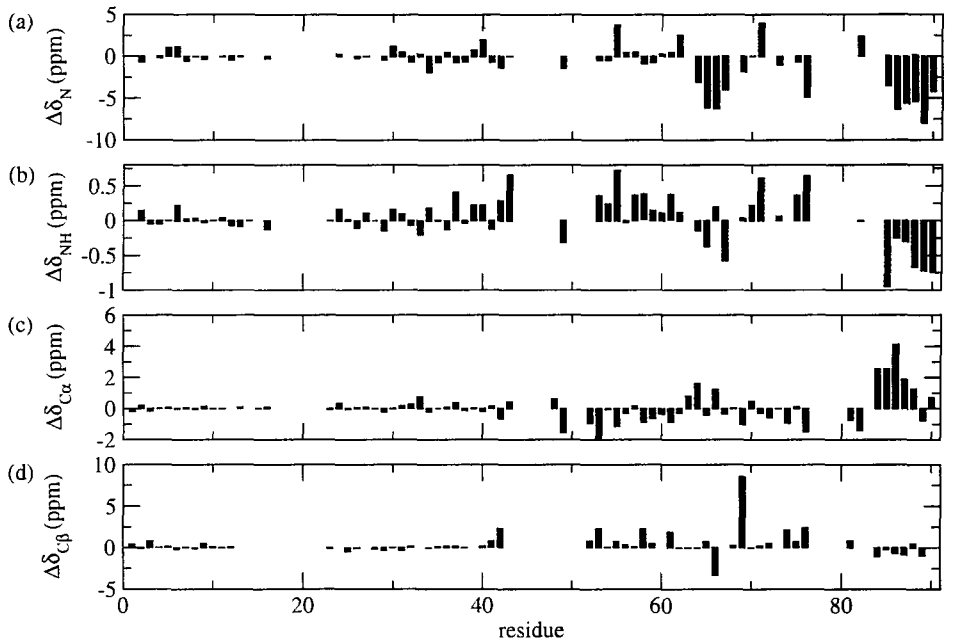

Figure 3.5: Chemical shift mapping, $\Delta\delta$ = form A − form B (ppm), to identify conformational differences between the two forms of HU*Tmar*. Residues for which there are no entries correspond to those which are not assigned in form B.

form B is partially unfolded or unstructured (as discussed in the following chapter).

Long range backbone-to-backbone NOE patterns assigned in the β-arms (presented in Figure 3.6) confirm the expected structure of this region as was achieved for HU*Bst* [177]. Since no electron density was observed for the β-arms in the crystal structure of HU*Tmar* these NOE assignments provide crucial evidence for the structural integrity of this region of the protein. Although the NOE data asssigned for the β-arms are less complete than for HU*Bst* due to spectral overlap the subset of assignments confirm that the antiparallel β-sheet topology is maintained.

As noted above the assignment of these crucial long range NOEs also provide an additional confirmation of the internal consistency of the main assignment. The peaks assigned in the NOESY spectra for form A are in general much less intense and the

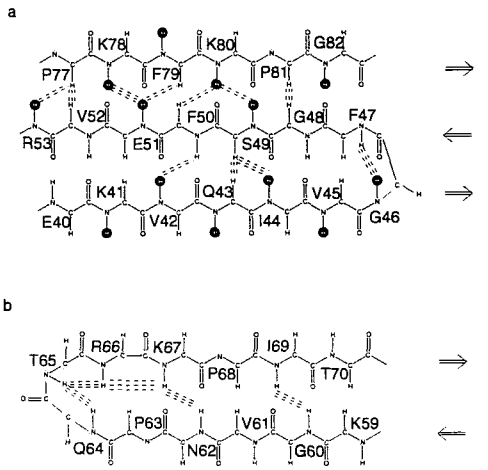

Figure 3.6: Schematic arrangement of the three-stranded antiparallel β-sheet (a) and of the β-arm (b) of HU*Tmar* based on long range backbone-to-backbone NOEs. The number of dashed lines indicates the relative intensities of the NOEs.

equivalent set of long range NOEs were not present. Indeed, the NOEs assigned for form B are mostly sequential connectivities. In some cases, mainly for residues which have small chemical shift differences in the ^{15}N HSQC spectrum (and therefore can be assumed to be in an almost equivalent structural environment), an equivalent set of NOEs could be assigned in both conformations. The NOE data when combined with the chemical shift data suggest that form A is closer to the expected native structure and that form B is less structured.

The chemical shifts of the backbone ^{15}N, ^{13}C, and ^{1}H nuclei have also been used to compare the predicted secondary structures of the dual assignments using the chemical shift index (CSI) [33, 34]. The combined chemical shifts of all assigned backbone nuclei can be compared with random coil values to give a reliable indication of secondary structure elements in proteins. The most reliable indicators are the $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO. A mapping of the observed secondary structure shifts (for both assignments) onto the primary sequence of HU*Tmar* (Figures 3.7, 3.8, 3.9, 3.10) can be compared with the secondary structure elements in the X-ray structure. Upfield shifts of $^{13}C^{\beta}$ combined with downfield shifts of $^{13}C^{\alpha}$ and ^{13}CO are strong indications of an α-helix. The opposite pattern is usually observed for a β-strand

where downfield shifts of $^{13}C^{\beta}$ combined with upfield shifts of the $^{13}C^{\alpha}$ and ^{13}CO resonances are expected.

Three important conclusions are immediately evident from an examination of the secondary shifts for HU*Tmar*. The first conclusion is that there are hardly any significant differences between the shifts for those nuclei that could be assigned in both forms and in general between the two forms considered over the whole sequence. The second conclusion is that, while the $^{13}C^{\alpha}$ and the $^{1}H^{\alpha}$ indices are reliable indicators of the secondary structure for HU*Tmar*, the $^{13}C^{\beta}$ and ^{13}CO indices are not reliable. The third conclusion is that, except for the C-terminus, the secondary structure elements of the X-ray structure are confirmed by the $^{13}C^{\alpha}$ and the $^{1}H^{\alpha}$ chemical shift data. Furthermore, the secondary structure predicted by all indices for the C-terminus in both form A and form B is unambiguously different from that in the X-ray structure. This difference can be attributed to either inaccurate predictions or differences between the solution and X-ray structures. Relaxation data (discussed in the next chapter) implicate this region of the protein as a possible source of the conformational heterogeneity. The relaxation data also present conflicting suggestions for the helical elements which could in principle help explain the observed values for the indices in these regions. These data allow us to conclude provisionally that the CSI predictions present a realistic view of the structure of HU*Tmar* in solution.

Statistical analysis of chemical shift data can give a reliable indication of the isomerization of *cis-trans* prolyl peptide bonds [178, 179]. Accordingly, if the chemical shift difference between the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ resonances is in the range 4.51 $\pm$ 1.37 ppm the *trans* isomer is favoured and if the difference is in the range 9.64 $\pm$ 1.62 ppm the *cis* isomer is favoured. Table 3.1 presents the differences in these resonances assigned for the five proline residues in HU*Tmar* in the dual assignments. According to these values all of the prolyl peptide bonds in HU*Tmar* are likely to be in the *trans* conformation. Note that P77 was assigned in one conformation only since the segment P77-K80 does not have a dual assignment.

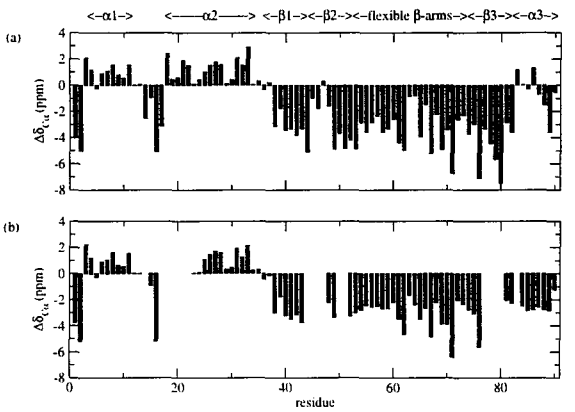

Figure 3.7: Backbone $^{13}C^{\alpha}$ secondary chemical shifts for (a) form A and (b) form B of HU*Tmar*. Positive and negative shifts are indicative of α-helical and β-strand conformations respectively.

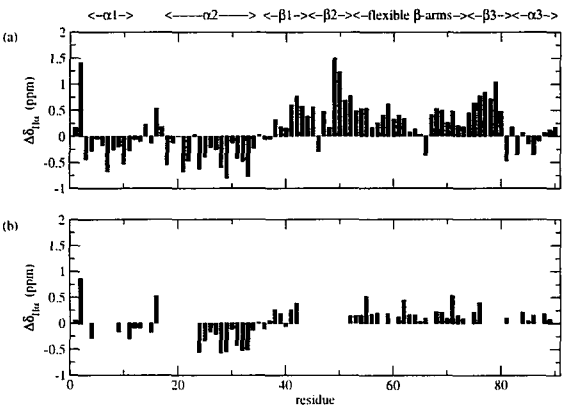

Figure 3.8: Backbone $^{1}H^{\alpha}$ secondary chemical shifts for (a) form A and (b) form B of HU*Tmar*. In contrast to the $^{13}C^{\alpha}$ and ^{13}CO indices positive and negative shifts are indicative of β-strand and α-helical conformations respectively.

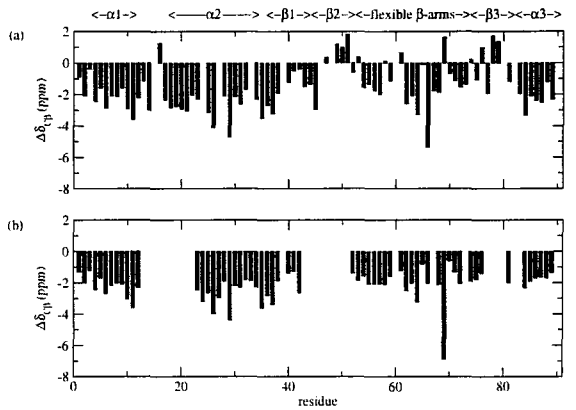

Figure 3.9: Backbone $^{13}C^{\beta}$ secondary chemical shifts for (a) form A and (b) form B of HU*Tmar*. In contrast to the $^{13}C^{\alpha}$ and ^{13}CO indices positive and negative shifts are indicative of β-strand and α-helical conformations respectively.

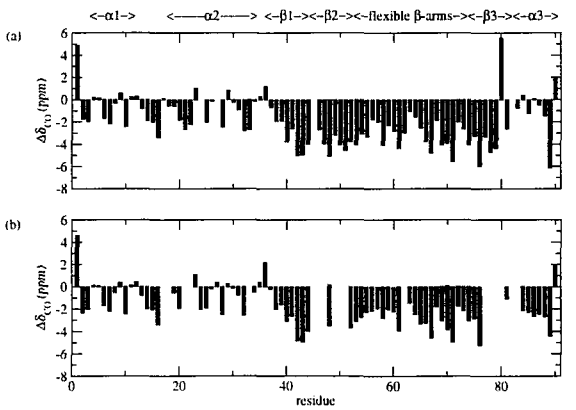

Figure 3.10: Backbone ^{13}CO secondary chemical shifts for (a) form A and (b) form B of HU*Tmar*. Positive and negative shifts are indicative of α-helical and β-strand conformations respectively.

	form A		form B	
residue	$\Delta_{\beta\gamma}$(ppm)	isomer	$\Delta_{\beta\gamma}$(ppm)	isomer
P63	4.85	*trans*	4.89	*trans*
P68	4.78	*trans*	4.70	*trans*
P72	5.43	*trans*	4.70	*trans*
P77	4.98	*trans*	no data	no data
P81	5.84	*trans*	4.92	*trans*

Table 3.1: Chemical shift differences, $\Delta_{\beta\gamma} = \Delta[^{13}C^{\beta} - {}^{13}C^{\gamma}]$, calculated for the five prolines in the dual assignments of HU*Tmar*.

3.4 Concluding remarks

The HU protein from *Thermotoga maritima* exists in two conformations in solution under NMR conditions. The backbone and side chains of both of these conformations have been assigned as far as possible by heteronuclear multidimensional NMR spectroscopy. These assignments provide a starting point for further NMR studies which can address the question of why HU*Tmar* exists in two conformations and what are the structural and dynamical reasons for the heterogeneity.

3.5 Acknowledgements

This project was performed during a Marie Curie training site fellowship under the supervision of Prof. dr. Rolf Boelens in Utrecht. The project was initiated by Dr. C. G. Kalodimos in collaboration with Dr. C. E. Vorgias (Athens) who expressed and purified the protein samples. Eugene Tischenko programmed the CC(CO)NH-TOCSY experiment. Dr. Rainer Wechselberger provided training in data recording and pulse sequence implementation. The NMR spectra were recorded at the SON NMR Large Scale Facility in Utrecht, which is funded by the *Access to Research Infrastructures* program of the European Union (HPRI-CT-2001-00172).

Chapter 4

Backbone dynamics of HU from *Thermotoga maritima*

4.1 Introduction

Proteins from thermophilic organisms display properties that reflect the environment in which they have evolved. The biophysical properties of such proteins has recently been the focus of several projects designed to investigate the relationship between molecular dynamics and thermal stability. The general consensus reached from interpretation of results from many different systems is that thermal stability results from conformational rigidity due in turn to the increased packing efficiency in the structures of thermophilic proteins [154]. Recent results from selected thermostable systems with unexpected flexibility challenge this paradigm however [155].

The HU-IHF group of dimeric histone-like proteins have various roles in DNA bending and compaction [180]. Members of this group include HU, IHF, H-NS, TF1, and Fis, and can bind DNA specifically or non-specifically. The crystal structures of the DNA complexes of both IHF an HU present a dramatic bending of the DNA mediated by proline intercalation and asymmetric charge neutralization [143, 147, 148, 146]. The structures of non-complexed HU from different prokaryotic species species have been determined by both crystallography and NMR. In all of the crystal structures no electron density is observed for the highly mobile yet internally structured β-arms which wrap around the DNA in the structures of the complexes [181]. The NMR structure of HU*Bst* provided the first details on the structure and dynamics of the β-arms [152]. The crystal structure of HU from the hyperthermophilic bacterium *Thermotoga maritima* has been refined to a resolution of 1.53Å although no electron density was observed for the β-arms [153]. Our interest was to characterize the structure and dynamics of HU*Tmar* in solution and compare the results with those for HU*Bst* [152, 182].

The behaviour of HU*Tmar* in solution probed by NMR is complementary to the results from a recent study characterizing destabilizing mutations [156]. In that work the authors designed and characterized mutant HU proteins from different sources based on sequence alignments. In particular for HU*Tmar* the single G15E, E34D, and V42I mutants displayed the expected decreases in thermostability as evidenced by results from circular dichroism. The E34D and V42I mutants have moderate decreases in thermostability while the G15E mutant converts the extremely

thermostable HU*Tmar* into a protein with similar stability to HU proteins from mesophilic sources. In this chapter the solution properties of these three mutants with the wild-type protein are compared using NMR methods. The expected similarity of the wild-type and E34D and V42I mutants was observed in the NMR spectra while the G15E mutant displayed markedly different behaviour in solution.

The results show that the wild-type HU*Tmar* is unexpectedly heterogeneous in solution on the NMR timescale. We show that our data suggest a two-state model in which form A and form B are approximately equally populated. The use of the A/B terminology therefore refers to structural integrity rather than populations of the two conformations. Therefore we propose that form A is more native-like while form B seems to result from localized destabilizing fluctuations in both homotypic and heterotypic intermonomer interfaces. Specifically, these fluctuations appear to arise in the α2 and α3 helices and not in the β-arms. In addition the two destablizing mutants (E34D and V42I) do not change this behaviour in contrast to the G15E mutant. The observed flexibility of HU*Tmar* in solution is unexpected when compared with other thermostable proteins.

4.2 Materials and methods

4.2.1 NMR spectroscopy

The ^{15}N-labelled samples of wild-type and mutant HU*Tmar* proteins were prepared as described in the literature [156, 159]. HSQC spectra were recorded for all samples. In addition, for the wild-type protein, ^{15}N heteronuclear NOE, longitudinal, and transverse relaxation data sets were recorded using modified versions of the standard Bruker pulse sequences on a 500 MHz spectrometer at a temperature of 310K. The modified pulse sequences were implemented as described in the literature [52, 171].

The pulse sequences for the measurement of the heteronuclear NOE, the longitudinal, and transverse relaxation rates are presented in Figure 4.1. These pulse sequences all

use the sensitivity enhancement scheme [183]. In addition, pulsed field gradients are used for coherence selection and solvent suppression [32]. These additions to the basic pulse sequences ensure excellent water suppression, absence of artifacts, and increased sensitivity.

The heteronuclear NOE experiments were programmed to record data in an interleaved mode where experiments with (NOE) and without (NONOE) the initial proton saturation are recorded in alternate scans. The NOE and NONOE datasets were separated using a Bruker macro and processed individually.

The R_1 and R_2 experiments were programmed so that the signal intensity decayed to zero as a function of the relaxation delay. A 90° ^{15}N pulse followed by a gradient ensures that magnetization originates on the protons. The R_1 experiment was repeated for eight values of the relaxation delay: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, and 1.0 s. The experiment with delays of 0.1 and 0.4 s were duplicated for the error analysis. The R_2 experiment was recorded with values for the delay of: 0, 2, 16, 32, 48, 64, 80, 96, 128, and 160 ms where the datasets with delays of 0 and 64 ms were duplicated.

A modified experiment for the simultaneous measurement of R_1 and slow conformational exchange rates was implemented in the Bruker pulse programming language [172]. However the conformational exchange in HU*Tmar* is very slow and therefore the expected cross peaks were not observed. In this experiment the t_1 frequency labelling of the ^{15}N magnetization takes place before the relaxation delay so that it is possible to discriminate between nuclei which exchange during the relaxation period. If exchange does occur this is a very powerful method for measuring the exchange rates of conformationally heterogeneous systems [173].

4.2.2 Data analysis

Two-dimensional data sets were recorded with 360 t_1 and 1280 t_2 complex points and zero-filled once before processing. The spectral widths were 1.5 kHz and 6 kHz in the indirect and direct dimensions respectively. All data were processed using NMRPipe [166] and analyzed using NMRView [167]. The analysis was performed as described

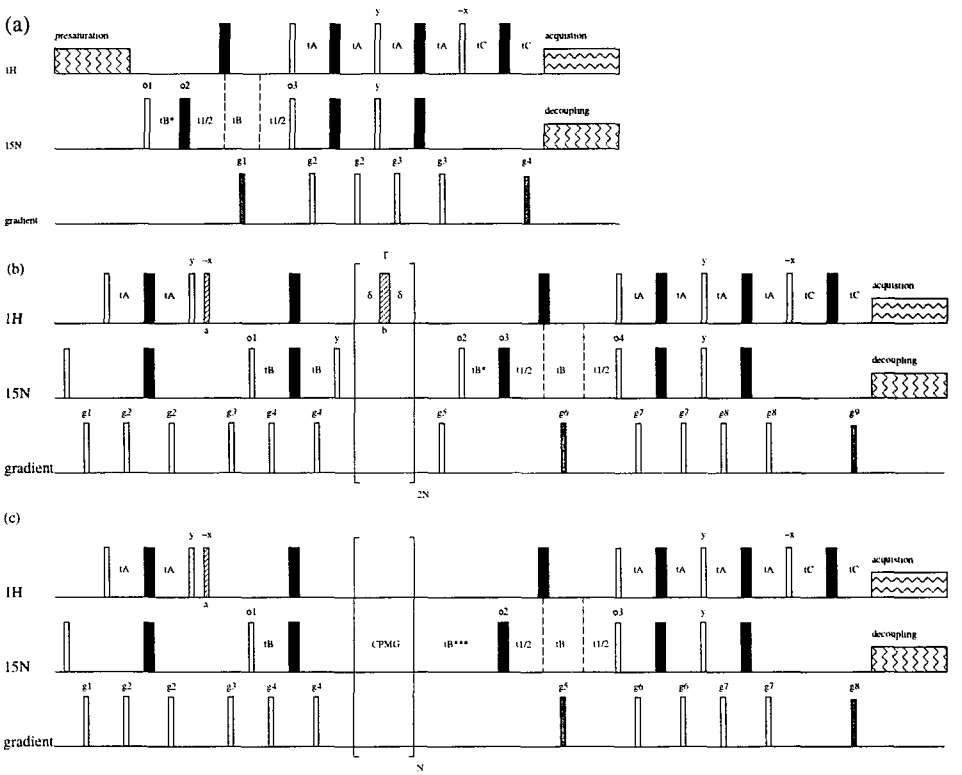

Figure 4.1: Pulse sequences for the measurement (a) NOE, (b) R_1, and (c) R_2. Narrow white and wide black bars indicate 90° and 180° pulses respectively and applied along the x-axis unless indicated otherwise. The narrow and wide dashed pulses on the proton channels are shaped 90° and 180° water selective pulses. These shaped pulses are used for water flip-back and suppression of cross-correlated relaxation (with a frequency null on water to cause minimal perturbation of the water signal) respectively. Dark bars on the gradient channel indicate pulses used for coherence selection and the light bars pulses used for artifact suppression. The delays tA, tB, and tC were set to 2.25ms ($< 1/4J_{NH}$), 2.75ms ($1/4J_{NH}$), and 0.5ms respectively.

in the literature [171, 184, 185].

Steady-state heteronuclear Overhauser enhancements are:

$$\eta = \frac{(I_{sat} - I_0)}{I_0} - 1 \tag{4.1}$$

where I_{sat} and I_0 are cross peak intensities with and without proton saturation respectively. The errors in the heteronuclear NOE for each residue were determined by dividing the standard deviation of the noise in each spectrum by the peak intensities, calculating the overall relative error from the square root of the summed squares of these errors and finally multiplying the overall relative error by the NOE value.

The intensities of the peaks in the R_1 and R_2 experiments as a function of the relaxation delay t are given by:

$$I(t) = I_0 exp(-R_{1,2}t), \tag{4.2}$$

where $I(t)$ is the intensity after a delay of time t and I_0 is the ^{15}N intensity at the beginning of the relaxation period $t = 0$. The rate constants were determined from peak volumes using a script with NMRView to interface with the program CurveFit (Prof. A. G. Palmer, Columbia University). The program uses the Levenburg-Marquardt algorithm [186] to minimize the value of χ^2 given by:

$$\chi^2 = \sum_{i=1}^{n} \frac{[I(t_i) - I_p(t_i)]^2}{\sigma^2}, \tag{4.3}$$

where $I(t_i)$ is the actual peak intensity at time point i; $I_p(t_i)$ is the value predicted during the fitting for the same time point; σ is the uncertainty in the experimental data and n is the number of data points recorded [184]. The uncertainties in the peak

heights were estimated from the duplicate experiments. Differences between peak heights in the duplicate spectra were evaluated for all cross peaks and the standard deviation of theose differences was divided by $\sqrt{2}$ to give the uncertainties in the actual peak heights. The χ^2 goodness-of-fit test was used to evaluate the monoexponential decay functions and the accuracy of the value of σ. If the residual χ^2 is greater than the 0.95 critical value there is a 95% confidence level that the model fit does not accurately describe the data.

Peak heights rather than peak volumes were used for deriving the NOE and rate constants from the HSQC spectra [184, 185]. In all cases the actual maximum peak height was used rather than the same data point in each spectrum. Peaks with moderate to severe overlap were excluded from the analysis. The results of the fitting procedure are presented in Appendix C. The data for both conformations of HU*Tmar* were fitted simultaneously. Where the signals of the two forms overlap the derived relaxation rates are an average for both conformations.

4.3 Results and discussion

4.3.1 Relaxation rate measurements

In order to give an overview of the structure and of the flexibility of the β-arms the NMR ensemble of HU*Bst* [152] is presented in Figure 4.2. The relaxation data for HU*Tmar* are mapped onto the primary sequence for the two conformations in Figures 4.5 and 4.6. The dual assignments described in the previous chapter will now be discussed in terms of the information available from the relaxation data. The overall pattern is that the β-arms are observed to be more flexible than the core of the protein in both conformations. In addition the form B is more flexible overall than form A as discussed in the following paragraphs.

The results of the non-linear regression fitting of the R_1 and R_2 rates for residue F50 are presented in Figures 4.3 and 4.4. These results are representative of those

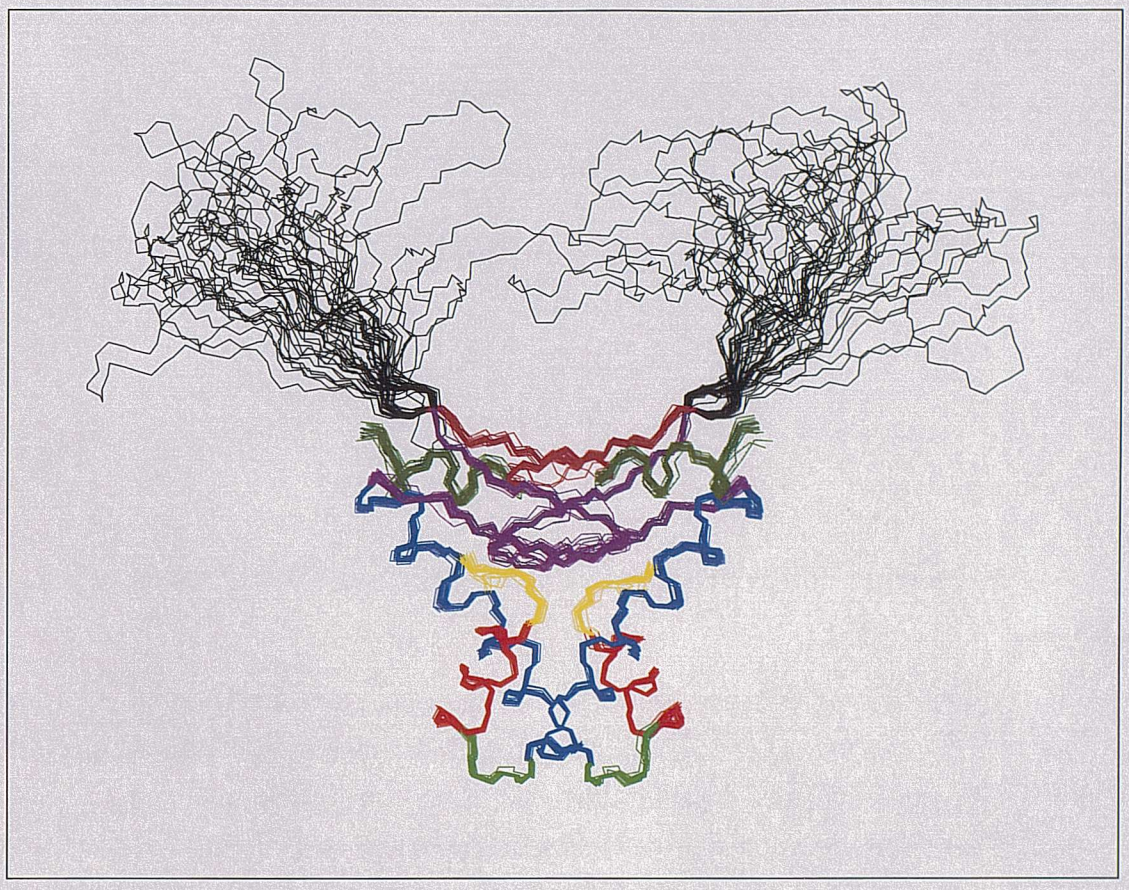

Figure 4.2: NMR ensemble of HU*Bst* (1hue.pdb) [152]. Residue colour coding: yellow, 1-3 (β1); red, 4-13 (α1); lime, 14-17 (turn); blue, 18-38 (α2); violet, 39-52 (β2,β3); black ,53-75 (β-arms); brown, 76-81 (β4); green, 82-90 (α3). Figure created in MOLMOL [137].

obtained for the majority of both the R_1 and the R_2 data (see Appendix C).

In form A more negative values for the ^{15}N-NH heteronuclear NOE and lower values for transverse relaxation rates compared to the core of the protein indicate an increase in flexibility. Furthermore, the values of the NOE and R_2 for residues N62-I69 indicate a maximum degree of flexibility for the tips of the β-arms. For the core of the protein the NOE values are fairly uniform for the majority of residues. The R_1 rates show variability over the protein sequence with higher rates in the β-arms and also in the the helix-turn-helix motif of α1 and α2.

The two forms of the protein clearly have different relaxation rates for residues for

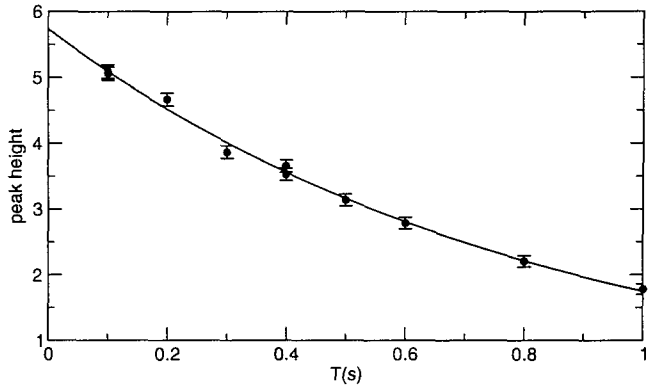

Figure 4.3: Optimized nonlinear fit of R_1 data for residue F50 in HU*Tmar*.

which data is available. Relaxation data for 41 (42 for the heteronuclear NOE) residues were obtained for form B (Figure 4.6). The remaining residues were either degenerate or overlapped in the HSQC spectrum. In particular, the substantially more negative values of the heteronuclear NOE and lower R_2 values for the β-arms indicate even higher mobility in form B compared to form A.

The values for the residues of the C-terminus are also interesting since according to the data this region is highly flexible in form B. The X-ray structure of HU*Tmar* has a short α-helix (1.5 helical turns) for the segment of residues numbered A84-V89. This C-terminal α-helix is characteristic of HU proteins and has been observed in all of the structures determined to date. The data for form A are compatible with an α-helix. The relaxation data for the C-terminus in form B, in contrast, are incompatible with an α-helix. The apparent lack of secondary structure for the C-terminal segment of form B may be significant for the overall structural integrity of the protein. The pattern of increased flexibilty is also observed in the long helix α2. For the second half of this helix, namely residues E30-E40, the R_1 rates are considerably higher and the R_2 rates are lower. The difference in the heteronuclear NOE is small but discernable for these residues. This information when combined with the secondary structure differences indicated by backbone chemical shifts and comparison of the assignments from the NOESY spectra indicate a high degree of flexibility throughout the structure. This effect may help explain the observed conformational heterogeneity.

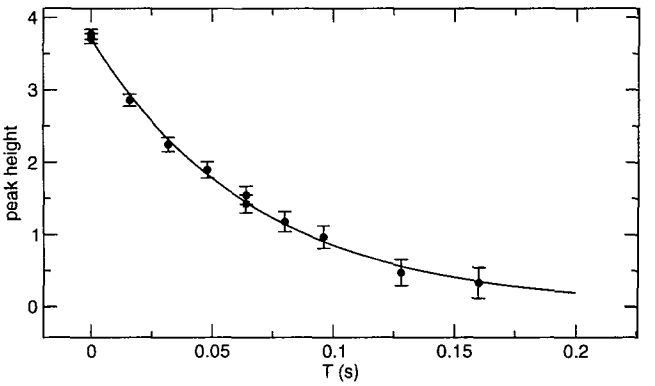

Figure 4.4: Optimized nonlinear fit of R_2 data for residue F50 in HU*Tmar*.

The combined information from the relaxation measurements, resonance assignments, and NOESY spectra are correlated. The major structural differences between the two conformations are mainly localized in the second half of helix α2, the β2 strand, the β-arms and the C-terminal helix α3. The majority of residues participate in the interface between the two monomers. Therefore any disruption of this interface might explain why there are two conformations. Perhaps such a disruption could propagate throughout the structure with a lessening of any effects further away from the source. Such a mechanism could account for the observation of less pronounced differences in the experimental results in more distal regions of the structure.

The conformational heterogeneity in this region is not confined to the backbone since for certain residues differences in the side chain conformations could be assigned from the ^{13}C-edited TOCSY spectra. These residues can be divided into two groups. The residues in the first group participate in the packing between the α2 helices in the intermonomer interface. The side chains of residues L23, I24, and L25 were completely assigned in the form A but not in form B presumably because the signals are conformationally broadened. The residues in the second group are again involved in the dimeric interface but between different secondary structure elements of the monomers, namely α1, β1, and α3. In this group residues in the second half of the α2 helix, I28-V45, are involved in intermonomer contacts with both the α1 and α3 helices. For these residues the side chains of I28, L29, I32, L36, V42, Q43, I44, V45,

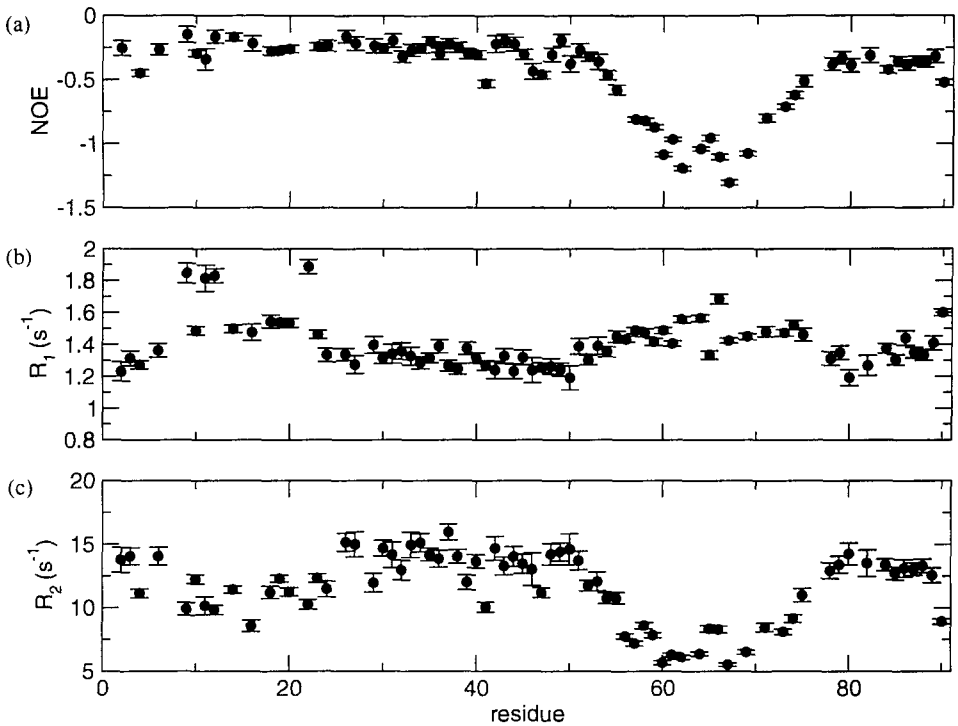

Figure 4.5: ^{15}N-NH (a) heteronuclear NOE, (b) R_1, and (c) R_2 relaxation rates for form A of HU*Tmar*.

G46, and F50 could only be completely assigned in form A. The general conclusion of a more structured form A is supported by the lack of side chain assignments for these residues in form B.

The mean R_2/R_1 ratio is a rough estimate of the overall correlation time of a protein and the value of 8.47 ns (form A) is in reasonable agreement with the value of 8.9 $\pm$ 0.6 ns obtained for HU*Bst* [182]. The R_2/R_1 ratios within one standard deviation of the mean can be fitted to obtain accurate estimates of the overall correlation times of proteins in solution [52, 58]. However since the following discussion of the backbone dynamics of HU*Tmar* is based on the basic relaxation rates and not a model free analysis no further analysis of the ratios to characterize the diffusion tensor was performed.

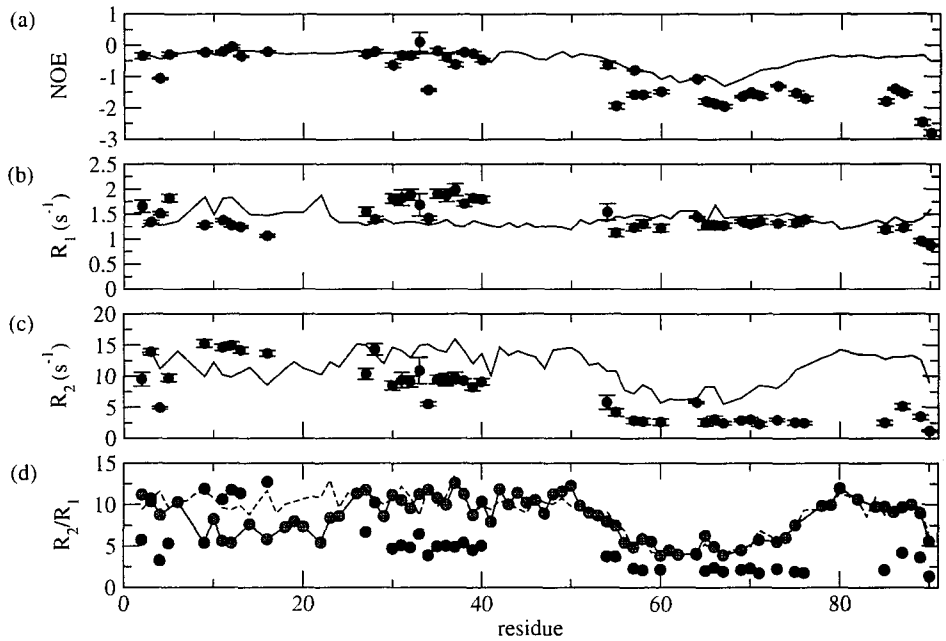

Figure 4.6: Comparison of (a) NOE, (b) R_1, and, (c) R_2 values for form A (line) and form B (circles) of HU*Tmar*; (d) comparison of R_2/R_1 rates for form A (shaded circles) and form B (filled circles) of HU*Tmar* with the values for HU*Bst* (dashed line) for reference.

The R_2/R_1 ratio reflects the orientation of secondary structure elements (or complete molecules) with respect to the overall rotational symmetry axis. The data clearly indicate differences between the two forms (Figure 4.6). Form A of HU*Tmar* and HU*Bst* have similar values of the R_2/R_1 ratios throughout the sequence. This indicates that form A has a diffusional anisotropy which is comparable to that of HU*Bst*. Furthermore, the R_2/R_1 values for α1 are lower on average than those for α2. In form B the opposite effect is seen since the R_2/R_1 values for α1 are higher than those for α2. This clearly shows that the relative positions of the two helices with respect to the diffusional symmetry axis are different for form A and form B.

The time scale of the conformational fluctuations can also be deduced from the NMR data. Longitudinal and transverse relaxation rates report on dynamical processes

occuring on different time scales [52]. R_1 values are sensitive to motions with frequencies in the picosecond to nanosecond range. R_2 values are also dependent on motions in this range but in addition can report on dynamical processes occuring on the microsecond to millisecond range. Conformational fluctuations on the μs-ms timescale broaden NMR signals leading to an increase in the measured R_2 rates.

4.3.2 Backbone dynamics of HU*Tmar* in solution

Proteins from thermophilic organisms have been the subject of many recent studies investigating the correlations between thermostability and conformational rigidity [154, 187]. The general consensus is that the thermostability required for successful biological functioning of such proteins is a result of increased packing efficiency due in turn to many small incremental improvements in hydrogen bonding, hydrophobic, coulombic, and van der Waals interactions. These packing improvements imply a concomitant increase in conformational rigidity which in turn can allow corresponding dynamical states at elevated temperatures. Recently however data to challenge this paradigm have appeared from studies on proteins from thermophilic sources [155]. Higher structural flexibility was observed for a thermophilic α-amylase as compared to the mesophilic homologue [188]. Interestingly, a recent study of amide hydrogen exchange in rubredoxin from *Pyrococcus furiosus* demonstrated unexpected flexibility in the most thermostable protein characterized to date [189, 190].

The crystal structure of HU*Tmar* confirms that the overall fold is very similar to the previous NMR and crystal structures [153]. Sequence alignment with HU*Bst* and HU*B.su* was used to design and characterize several mutant proteins with altered thermostabilities. Wild-type HU*Tmar* has a melting temperature of 80.5°C which can be compared with the point mutations G15E (T_m = 55.8°C), E34D (T_m = 72.7°C), and V42I (T_m = 70.9°C); and the combined triple mutant G15E-E34D-V42I (T_m = 35.9°C) [156]. NMR spectra recorded from samples of the three point mutations confirm this trend. The spectra of the mutants E34D (data not shown) and V42I (Figure 4.7) are almost identical to the spectrum for the wild type protein. In contrast the spectrum of the G15E sample (Figure 4.7) is quite different as might be predicted

from the dramatically reduced melting temperature of this mutant. This residue is known to be critical to the stability of both HU*Bst* and HU*Tmar* [153, 191, 192].

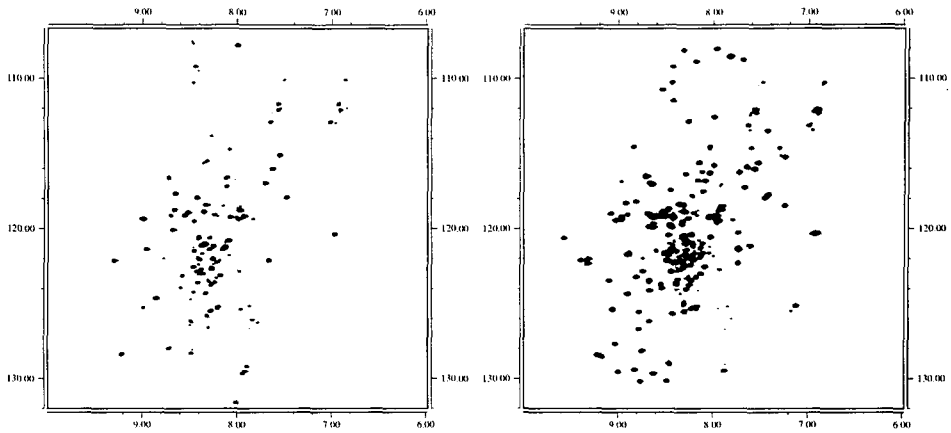

Figure 4.7: 2D ^{15}N HSQC spectra of uniformly ^{15}N-labelled G15E (left) and V42I (right) mutant HU*Tmar* proteins recorded at 500 MHz proton frequency and 311K.

The spectra of the three mutant proteins allow us to provisionally conclude that the structure of two of the mutants (E34D and V42I) are very similar to the wild-type protein while the G15E mutant may adopt a different conformation in solution. Comparison of the spectra for the wild-type and G15E proteins allowed the unambiguous assignment of 53 residues in the mutant in addition to confirming the the absence of the peak assigned to G15. The 53 residues assigned in the spectrum of the G15E mutant were based on their identical resonance frequencies as compared to form A of the wild type protein. In addition the G15E mutant exists in one form only which perhaps highlights the importance of the glycine residue in the turn between the two helices. Recall that the possible rearrangement of these two helices may give rise to the dual conformations in the wild type protein. In addition, since there is only one form a comparison of the shifts with form B of the wild type protein would not have any meaning. Further details on the dynamics of the G15E will require a complete assignment of the backbone and comparison with a three-dimensional structure.

The *B*-factors of the backbone nitrogen atoms of the two monomers of HU*Tmar* are also indicative of structural fluctuations. Although the *B*-factor data give no information on the time scale of the motions a general correlation with the NMR data can be discerned. Residues at the N- and C-termini and in turns are usually expected

to display higher than average *B*-factor values due to heterogeneity in the crystal. However for the HU*Tmar* crystal structure the values for the C-terminus are distinctly higher than the mean and the values in the second half of helix α2 and β-sheet elements β2 and β3 also have slightly higher values. A similar pattern is present for the *B*-factor data for the crystal structure of HU*Bst* but the values for the C-terminus are relatively lower than for the HU*Tmar* structure.

Several sources of data illustrate the different conformational properties of the two forms of HU*Tmar* in solution. The two most likely explanations for this conformational heterogeneity are: (1) monomer-dimer equilibrium; and (2) different dimeric structures due to dynamic effects possibly leading to a structural rearrangement. The available data suggest that the latter is the case for HU*Tmar*. Furthermore, based on the chemical shift, NOESY, and relaxation data it may be concluded that form A bears close resemblance to both the X-ray and NMR structures of HU*Tmar* and HU*Bst* respectively.

The monomeric state of HU*Tmar* is likely to exist only very transiently in solution due to the high stability of the dimer [193] and therefore cannot account for the observed differences in the relaxation data. This is supported by the observation that changes in the protein and salt concentrations do not significantly change the spectra. In addition the average R_2/R_1 values for the core of the protein are very similar for the two forms which would not be expected for a monomeric form B. A comparison of the R_2/R_1 values for the α1 and α2 helices however indicates a difference in the relative positions of the two helices with respect to the molecular diffusion axis for the two forms of HU*Tmar*. Whether this is due to a molecular rearrangement of the helices or due to a reorientation of the major diffusion axis cannot be concluded at this stage. This would require a high-resolution structural analysis of the form B.

The possibility of association of dimers to form tetramers also seems unlikely based on the differences in linewidths and relaxation rates of the NMR signals that would be expected in that case. An additional possibility is the interaction of phosphate in the buffer with the protein. While this cannot be ruled out it is unlikely based on the observation of only one species on SDS gels.

The high stability of the HU*Tmar* dimer results from intermonomer contacts in the compact core of the protein. The flexible β-arms extend from the core and do not participate in formation of the dimer but are instead available to wrap around a DNA substrate [152]. A recent Raman spectroscopy deuterium exchange study of the solution conformation of HU*Bst* found unexpectedly high exchange rates for helical elements which the authors attributed to flexibility at the dimeric interface [194]. The backbone relaxation rates measured for both forms of HU*Tmar* also indicate unusual flexibility at both the homotypic and heterotypic dimeric interfaces formed by the two pairs of α1 and α2 helices. Flexibility at the dimeric interface could account for the critical role of residue G15 in the turn between the α1 and α2 helices since mutation of this residue is known to dramatically reduce the thermostability of HU proteins [156, 192].

The biophysical properties of HU proteins from various sources have been documented in detail [151]. In marked contrast to HU*Tmar*, protein concentration and ionic strength have been shown to affect both the structure and stability of HU*Bsu* in solution as measured by CD [195]. These observations were extended with ESI-MS measurements of the influence of the ionic strength on the monomer-dimer equlibrium in HU*Bsu* [196]. The NMR data presented here expand on the earlier biophysical results for HU*Tmar* [193] since information is now available for individual amino acids and the existence of dual conformations is surprising.

4.4 Concluding remarks

This chapter demonstrates that HU*Tmar* is unexpectedly heterogeneous and dynamic in solution on a broad timescale range. The two conformations are approximately equally populated according to peak intensities. Form A in solution has features consistent with those determined in the X-ray structure as evidenced by chemical shift, NOE, and ^{15}N relaxation data. The presence of the two forms in both wild-type and several mutants of HU*Tmar* suggests that the equilibrium is an intrinsic property of the native structure. The dynamical behaviour of HU*Tmar* suggests that this protein may be a new example of the small class of extremely thermostable proteins

which display unexpected flexibility in solution. In addition HU*Tmar* may be the first example of a homodimeric thermostable protein displaying these properties.

4.5 Acknowledgements

The NMR spectra were recorded at the SON NMR Large Scale Facility in Utrecht, which is funded by the *Access to Research Infrastructures* program of the European Union (HPRI-CT-2001-00172). Klaartje Houben and Rainer Wechselberger modified the Bruker pulse sequences and KH set up the scripts for the analysis in NMRView. RW and MAD implemented the longitudinal exchange experiment as described in the original paper. Hans Wienk read and improved the condensed version of this and the previous chapter submitted for publication (Appendix D).

Bibliography

[1] H. F. Judson. *The eighth day of creation: the makers of the revolution in biology.* Simon & Schuster, New York, 1979.

[2] C. R. Cantor and P. R. Schimmel. *Biophysical chemistry part I: the conformation of biological macromolecules.* Freeman, San Francisco, 1980.

[3] W. Saenger. *Principles of nucleic acid structure.* Springer-Verlag, New York, 1984.

[4] C. Branden and J. Tooze. *Introduction to protein structure.* Freeman, San Francisco, 1998.

[5] R. F. Gesteland, T. R. Cech, and J. F. Atkins, editors. *The RNA world.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999.

[6] A. Šali. 100,000 protein structures for the biologist. *Nat. Struct. Biol.*, 5:1029–1032, 1998.

[7] L. Stein. Creating a bioinformatics nation. *Nature*, 417:119–120, 2002.

[8] S. C. Harrison. Whither structural biology? *Nat. Struct. Mol. Biol.*, 11:12–15, 2004.

[9] R. R. Ernst. Nuclear magnetic resonance Fourier transform spectroscopy (Nobel lecture). *Angew. Chem. Int. Ed. Engl.*, 31:805–830, 1992.

[10] K. Wüthrich. NMR studies of structure and function of biological macromolecules (Nobel lecture). *Angew. Chem. Int. Ed. Engl.*, 42:3340–3363, 2003.

[11] G. Wider and K. Wüthrich. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr. Opin. Struct. Biol.*, 9:594–601, 1999.

[12] J. Fiaux, E. B. Bertelsen, A. L. Horwich, and K. Wüthrich. NMR analysis of a 900K GroEL-GroES complex. *Nature*, 418:207–211, 2002.

[13] C. Fernández, C. Hilty, S. Bonjour, K. Adeishvili, K. Pervushin, and K. Wüthrich. Solution NMR studies of the integral membrane proteins OmpX and OmpA from *Escherichia coli*. *FEBS Lett.*, 504:173–178, 2001.

[14] S. J. Opella and F. M. Marassi. Structure determination of membrane proteins by NMR spectroscopy. *Chem. Rev.*, 104:3587–3606, 2004.

[15] A. Bax. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.*, 12:1–16, 2003.

[16] J. H. Prestegard, C. M. Bougault, and A. I. Kishore. Residual dipolar couplings in structure determination of biomolecules. *Chem. Rev.*, 104:3519–3540, 2004.

[17] K. Wüthrich. *NMR of proteins and nucleic acids.* Wiley, New York, 1986.

[18] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, and N. J. Skelton. *Protein NMR spectroscopy: principles and practice.* Academic Press, New York, 1996.

[19] P. Güntert. Structure calculation of biological macromolecules from NMR data. *Q. Rev. Biophys.*, 31:145–237, 1998.

[20] S. S. Wijmenga and B. N. M. van Buuren. The use of NMR methods for conformational studies of nucleic acids. *Prog. NMR Spectrosc.*, 32:287–387, 1998.

[21] B. J. Stockman. Preparation of ^{2}H, ^{13}C, and ^{15}N isotopically-enriched proteins for NMR spectroscopic investigations. In S. K. Sarkar, editor, *NMR spectroscopy and its application to biomedical research*, pages 159–185. Elsevier, Amsterdam, 1996.

[22] L. E. Kay and K. H. Gardner. Solution NMR spectroscopy beyond 25kDa. *Curr. Opin. Struct. Biol.*, 7:722–731, 1997.

[23] K. H. Gardner and L. E. Kay. The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.*, 27:357–406, 1998.

[24] A. E. Derome. *Modern NMR techniques for chemistry research.* Pergamon, Oxford, 1987.

[25] A. Bax and S. Grzesiek. Methodological advances in protein NMR. *Acc. Chem. Res.*, 26:131–138, 1993.

[26] F. J. M. van de Ven. *Multidimensional NMR in liquids.* VCH, New York, 1996.

[27] R. Freeman. *Spin choreography: basic steps in high resolution NMR.* Spektrum, Oxford, 1997.

[28] G. Wider. Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. *Prog. NMR Spectrosc.*, 32:193–275, 1998.

[29] M. Sattler, J. Schleucher, and C. Griesinger. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.*, 34:93–158, 1999.

[30] M. H. Levitt. *Spin dynamics: basics of nuclear magnetic resonance*. Wiley, New York, 2001.

[31] E. AB. *Structural studies on IIB^{Chb}*. PhD thesis, Rijksuniversiteit Groningen, 2002.

[32] L. E. Kay. Field gradient techniques in NMR spectroscopy. *Curr. Opin. Struct. Biol.*, 5:674–681, 1995.

[33] D. S. Wishart and B. D. Sykes. The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J. Biomol. NMR*, 4:171–180, 1994.

[34] D. S. Wishart and B. D. Sykes. Chemical shifts as a tool for structure determination. *Methods Enzymol.*, 239:363–392, 1994.

[35] D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley, and B. D. Sykes. ^{1}H, ^{13}C, and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, 6:135–140, 1995.

[36] G. J. Pielak, D. S. Auld, S. F. Betz, S. E. Hilgen-Willis, and L. L. Garcia. Nuclear magnetic resonance studies of class I cytochromes *c*. In R. A. Scott and A. G. Mauk, editors, *Cytochrome c: a multidisciplinary approach*, pages 203–284. University Science Books, Sausalito, California, 1996.

[37] J. A. Pople. Proton magnetic resonance of hydrocarbons. *J. Chem. Phys.*, 24:1111, 1956.

[38] I. Bertini and C. Luchinat. *NMR of paramagnetic molecules in biological systems*. Benjamin-Cummings, Menlo Park, California, 1986.

[39] R. O. Louro, I. J. Correia, L. Brennan, I. B. Coutinho, A. V. Xavier, and D. L. Turner. Electronic structure of low-spin ferric porphyrins: ^{13}C studies of the influence of axial ligand orientation. *J. Am. Chem. Soc.*, 120:13240–13247, 1998.

[40] D. L. Turner, L. Brennan, A. C. Messias, M. L. Teodoro, and A. V. Xavier. Correlation of empirical magnetic susceptibility tensors and structure in low-spin heme proteins. *Eur. Biophys. J.*, 29:104–112, 2000.

[41] B. F. Volkman, S. J. Wilkins, A. L. Lee, B. Xia, W. M. Westler, R. Beger, and J. L. Markley. Redox-dependent magnetic alignment of *Clostridium pasteurianum* rubredoxin: measurement of magnetic susceptibility anisotropy and prediction of pseudocontact shift contributions. *J. Am. Chem. Soc.*, 121:4677–4683, 1999.

[42] D. L. Turner, L. Brennan, S. G. Chamberlin, R. O. Louro, and A. V. Xavier. Determination of solution structures of paramagnetic proteins by NMR. *Eur. Biophys. J.*, 27:367–375, 1998.

[43] G. Wagner. NMR relaxation and protein mobility. *Curr. Opin. Struct. Biol.*, 3:748–754, 1993.

[44] J. W. Peng and G. Wagner. Investigation of protein motions via relaxation measurements. *Methods Enzymol.*, 239:563–596, 1994.

[45] K. T. Dayie, G. Wagner, and J. F. Lefèvre. Theory and practice of nuclear spin relaxation in proteins. *Annu. Rev. Phys. Chem.*, 47:243–282, 1996.

[46] A. G. Palmer. Probing molecular motion by NMR. *Curr. Opin. Struct. Biol.*, 7:732–737, 1997.

[47] L. E. Kay. Protein dynamics from NMR. *Nat. Struct. Biol.*, 5:513–517, 1998.

[48] A. G. Palmer, C. D. Kroenke, and J. P. Loria. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.*, 339:204–238, 2001.

[49] R. Ishima and D. A. Torchia. Protein dynamics from NMR. *Nat. Struct. Biol.*, 7:740–743, 2000.

[50] M. Akke. NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Curr. Opin. Struct. Biol.*, 12:642–647, 2002.

[51] R. Brüschweiler. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr. Opin. Struct. Biol.*, 13:175–183, 2003.

[52] L. E. Kay, D. A. Torchia, and A. Bax. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28:8972–8979, 1989.

[53] E. A. Keiter. *Deuterium and nitrogen-14 nuclear quadrupole double resonance studies of hydrogen bonding*. PhD thesis, University of Illinois, 1986.

[54] Y. Hiyama, C. H. Niu, J. V. Silverton, and A. Bavoso. Determination of ^{15}N chemical shift tensor via ^{15}N-^{2}H dipolar coupling in boc-glyclyglycyl[^{15}N]glycine benzyl ester. *J. Am. Chem. Soc.*, 110:2378–2383, 1988.

[55] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.*, 104:4546–4559, 1982.

[56] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.*, 104:4559–4570, 1982.

[57] G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.*, 112:4989–4991, 1990.

[58] G. M. Clore, P. C. Driscoll, P. T. Wingfield, and A. M. Gronenborn. Analysis of the backbone dynamics of interleukin-1β using two-dimensional inverse detected heteronuclear ^{15}N-^{1}H NMR spectroscopy. *Biochemistry*, 29:7387–7401, 1990.

[59] S. W. Englander and M. M. G. Krishna. Hydrogen exchange. *Nat. Struct. Biol.*, 8:741–742, 2001.

[60] C. B. Arrington and A. D. Robertson. Kinetics and thermodynamics of conformational equilibria in native proteins by hydrogen exchange. *Methods Enzymol.*, 323:104–124, 2000.

[61] P. L. Weber. Protein structure determination from NMR data. In S. K. Sarkar, editor, *NMR spectroscopy and its application to biomedical research*, pages 187–239. Elsevier, Amsterdam, 1996.

[62] M. Karplus and G. A. Petsko. Molecular dynamics simulations in biology. *Nature*, 347:631–639, 1990.

[63] W. Braun and N. Go. Calculation of protein conformations by proton-proton distance constraints: a new efficient algorithm. *J. Mol. Biol.*, 186:611–626, 1985.

[64] A. T. Brünger. *X-PLOR, version 3.1: a system for X-ray crystallography and NMR*. Yale University Press, New Haven, 1992.

[65] A. T. Brünger, P. D. Adams, G. M. Clore, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice,

T. Simonson, and G. L. Warren. Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta. Cryst. D.*, 54:905–921, 1998.

[66] P. Güntert, C. Mumenthaler, and K. Wüthrich. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, 273:283–298, 1997.

[67] T. Herrmann, P. Güntert, and K. Wüthrich. Protein NMR structure determination with automated assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, 319:209–227, 2002.

[68] L. Banci, I. Bertini, K. L. Bren, M. A. Cremoni, H. B. Gray, C. Luchinat, and P. Turano. The use of pseudocontact shifts to refine the solution structures of paramagnetic metalloproteins: Met80Ala cyano-cyt *c* as an example. *J. Biol. Inorg. Chem.*, 1:117–126, 1996.

[69] L. Banci, I. Bertini, M. A. Cremoni, G. Gori-Savellini, P. Güntert, C. Luchinat, and K. Wüthrich. PSEUDYANA for NMR structure calculation of paramagnetic metalloproteins using torsion angle molecular dynamics. *J. Biomol. NMR*, 12:553–557, 1998.

[70] M. Gochin and H. Roder. Protein structure refinement based on paramagnetic NMR shifts: applications to wild-type and mutant forms of cytochrome *c*. *Protein Sci.*, 4:296–305, 1995.

[71] M. Gochin and H. Roder. Use of pseudocontact shifts as structural constraints for macromolecules in solution. *Bull. Magn. Reson.*, 17:1–4, 1995.

[72] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci. USA*, 92:9279–9283, 1995.

[73] N. Tjandra and A. Bax. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, 278:1111–1114, 1997.

[74] A. Bax and N. Tjandra. Are proteins even floppier than we thought? *Nat. Struct. Biol.*, 4:254–256, 1997.

[75] J. R. Tolman. Dipolar couplings as a probe of molecular dynamics and structure in solution. *Curr. Opin. Struct. Biol.*, 11:532–539, 2001.

[76] P. Güntert, W. Braun, and K. Wüthrich. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.*, 217:517–530, 1991.

[77] R. Boelens, T.M.G. Koning, and R. Kaptein. Determination of biomolecular structures from proton-proton NOEs using a relaxation matrix approach. *J. Mol. Struct.*, 173:299–311, 1988.

[78] R. Boelens, T.M.G. Koning, G.A. van der Marel, J.H. van Boom, and R. Kaptein. Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach: application to a DNA octamer. *J. Magn. Reson.*, 82:290–308, 1989.

[79] I. Bertini, A. Donaire, I. C. Felli, and A. Rosato. From NOESY cross peaks to structural constraints in a paramagnetic metalloprotein. *Magn. Reson. Chem.*, 34:948–950, 1996.

[80] C. Mumenthaler, P. Güntert, W. Braun, and K. Wüthrich. Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *J. Biomol. NMR*, 10:351–362, 1997.

[81] M. Nilges, M. J. Macias, S. I. O'Donoghue, and H. Oschkinat. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.*, 269:408–422, 1997.

[82] M. Nilges and S. I. O'Donoghue. Ambiguous NOEs and automated NOE assignment. *Prog. NMR Spectrosc.*, 32:107–139, 1998.

[83] J. P. Linge, S. I. O'Donoghue, and M. Nilges. Assigning ambiguous NOEs with ARIA. *Methods Enzymol.*, 339:71–90, 2001.

[84] R. W. W. Hooft, G. Vriend, C. Sander, and E. E. Abola. Errors in protein structures. *Nature*, 381:272, 1996.

[85] R. W. W. Hooft, C. Sander, and G. Vriend. Objectively judging the quality of a protein structure from a Ramachandran plot. *Comput. Appl. Biosci.*, 13:425–430, 1997.

[86] G. N. Ramachandran, C. Ramakrishna, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 7:95–99, 1963.

[87] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26:283–291, 1993.

[88] R. A. Laskowski, J. A. C. Rullmann, M. W. MacArthur, R. Kaptein, and J. M. Thornton. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, 8:477–486, 1996.

[89] G. Vriend. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, 8:52–56, 1990.

[90] G. Vriend and C. Sander. Quality control of protein models: directional atomic contact analysis. *J. Appl. Cryst.*, 26:47–60, 1993.

[91] D. Zhao and O. Jardetzky. An assessment of the precision and accuracy of protein structures determined by NMR: dependence on distance errors. *J. Mol. Biol.*, 239:601–607, 1994.

[92] J. F. Doreleijers, J. A. C. Rullmann, and R. Kaptein. Quality assessment of NMR structures: a statistical survey. *J. Mol. Biol.*, 281:149–164, 1998.

[93] J. P. Linge, M. A. Williams, C. A. Spronk, A. M. Bonvin, and M. Nilges. Refinement of protein structures in explicit solvent. *Proteins*, 50:496–506, 2003.

[94] B. Xia, V. Tsui, D. A. Case, H. J. Dyson, and P. E. Wright. Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J. Biomol. NMR*, 22:317–331, 2002.

[95] B. R. Seavey, E. A. Barr, W. M. Westler, and J. L. Markley. A relational database for sequence-specific protein NMR data. *J. Biomol. NMR*, 1:217–236, 1991.

[96] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–542, 1977.

[97] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucl. Acids Res.*, 28:235–242, 2000.

[98] G. W. Pettigrew and G. R. Moore. *Cytochromes c: biological aspects*. Springer-Verlag, Berlin, 1987.

[99] G. R. Moore and G. W. Pettigrew. *Cytochromes c: evolutionary, structural, and physicochemical aspects*. Springer-Verlag, Berlin, 1987.

[100] I. B. Coutinho and A. V. Xavier. Tetraheme cytochromes. *Methods Enzymol.*, 243:119–140, 1994.

[101] A. V. Xavier. A mechano-chemical model for energy transduction in cytochrome *c* oxidase: the work of a Maxwell's god. *FEBS Lett.*, 532:261–266, 2002.

[102] R. O. Louro, T. Catarino, C. A. Salgueiro, J. LeGall, and A. V. Xavier. Redox-Bohr effect in the tetrahaem cytochrome c_3 from *Desulfovibrio vulgaris*: a model for energy transduction mechanisms. *J. Biol. Inorg. Chem.*, 1:34–38, 1996.

[103] R. O. Louro, T. Catarino, C. A. Salgueiro, J. LeGall, and A. V. Xavier. Redox-Bohr effect in electron-proton energy transduction: cytochrome c_3 coupled to hydrogenase works as a proton thruster. *J. Biol. Inorg. Chem.*, 2:488–491, 1998.

[104] R. O. Louro, T. Catarino, D. L. Turner, M. A. Picarra-Pereira, I. Pacheco, J. LeGall, and A. V. Xavier. Functional and mechanistic studies of cytochrome c_3 from *Desulfovibrio gigas*: thermodynamics of a proton-thruster. *Biochemistry*, 37:15808–15815, 1998.

[105] B. Guigiarelli, P. Bertrand, L. More, R. Haser, and J. P. Gayda. Single-crystal electron paramagnetic resonance study of cytochrome c_3 from *Desulfovibrio desulfuricans* Norway strain: assignment of the heme midpoint redox potentials. *J. Mol. Biol.*, 216:161–166, 1990.

[106] F. Guerlesquin, M. Bruschi, and K. Wüthrich. 1H-NMR studies of *Desulfovibrio desulfuricans* Norway strain cytochrome c_3. *Biochim. Biophys. Acta*, 830:296–303, 1985.

[107] C. A. Salgueiro, D. L. Turner, H. Santos, J. LeGall, and A. V. Xavier. Assignment of the redox potentials to the four haems in *Desulfovibrio vulgaris* cytochrome c_3 by 2D-NMR. *FEBS Lett.*, 314:155–158, 1992.

[108] M. A. Picarra-Pereira, D. L. Turner, J. LeGall, and A. V. Xavier. Structural studies on *Desulfovibrio gigas* cytochrome c_3 by two-dimensional ^{1}H nuclear magnetic resonance spectroscopy. *Biochem. J.*, 294:909–915, 1993.

[109] D. J. Detlefsen, V. Thanabal, V. L. Pecoraro, and G. Wagner. Solution structure of Fe(II) cytochrome *c*551 from *Pseudomonas aeruginosa* as determined by two-dimensional ^{1}H NMR. *Biochemistry*, 30:9040–9046, 1991.

[110] M. Cai, E. G. Bradford, and R. Timkovich. Investigation of the solution conformation of cytochrome *c*551 from *Pseudomonas stutzeri*. *Biochemistry*, 31:8603–8612, 1992.

[111] P. X. Qi, D. DiStefano, and A. J. Wand. Solution structure horse heart ferrocytochrome *c* determined by high resolution NMR and restrained simulated annealing. *Biochemistry*, 33:6408–6417, 1992.

[112] M. J. Blackledge, S. Medvedeva, M. Poncin, F. Guerlesquin, M. Bruschi, and D. Marion. Structure and dynamics of ferrocytochrome *c*553 from *Desulfovirio vulgaris* studied by NMR spectroscopy and restrained molecular dynamics. *J. Mol. Biol.*, 245:661–681, 1995.

[113] Y. Higuchi, M. Kusunoki, Y. Matsuura, N. Yasuoka, and M. Kakudo. Refined structure of cytochrome c_3 at 1.8Å resolution. *J. Mol. Biol.*, 172:109–139, 1984.

[114] M. Czjzek, F. Payan, F. Guerlesquin, M. Bruschi, and R. Haser. Crystal structure of cytochrome c_3 from *Desulfovibrio desulfuricans* at 1.7Å resolution. *J. Mol. Biol.*, 243:653–667, 1994.

[115] P. M. Matias, L. M. Saraiva, C. M. Soares, A. V. Coelho, J. LeGall, and M. A. Carrondo. Nine-haem cytochrome *c* from *Desulfovibrio desulfuricans* ATCC 27774: primary sequence determination, crystallographic refinement at 1.8Å and modelling studies of its interaction with the tetrahaem cytochrome c_3. *J. Biol. Inorg. Chem.*, 4:478–494, 1999.

[116] P. Simoes, P. M. Matias, J. Morais, K. Wilson, Z. Dauter, and M. A. Carrondo. Refinement of the three-dimensional structures of cytochromes c_3 from *Desulfovibrio vulgaris* Hildenborough at 1.67Å resolution and from *Desulfovibrio desulfuricans* ATCC 27774 at 1.6Å resolution. *Inorg. Chim. Acta.*, 273:213–224, 1998.

[117] S. Norager, P. Legrand, L. Pieulle, C. Hatchikian, and M. Roth. Crystal structure of the oxidised and reduced acidic cytochrome c_3 from *Desulfovibrio africanus*. *J. Mol. Biol.*, 290:881–902, 1999.

[118] I. B. Coutinho, D. L. Turner, J. LeGall, and A. V. Xavier. Revision of the haem-core architecture in the tetrahaem cytochrome c_3 from *Desulfovibrio bacalatus* by two-dimensional ^{1}H NMR. *Eur. J. Biochem.*, 209:329–333, 1992.

[119] A. C. Messias, D. H. W. Kastrau, H. S. Costa, J. LeGall, D. L. Turner, H. Santos, and A. V. Xavier. Solution structure of *Desulfovibrio vulgaris* (Hildenborough) ferrocytochrome c_3: structural basis for functional cooperativity. *J. Mol. Biol.*, 281:719–739, 1998.

[120] L. Brennan, D. L. Turner, A. C. Messias, M. L. Teodoro, J. LeGall, H. Santos, and A. V. Xavier. Structural basis for the network of functional cooperativities in cytochrome c_3 from *Desulfovibrio gigas*: solution structures of the oxidised and reduced states. *J. Mol. Biol.*, 298:61–82, 2000.

[121] M. Pierrot, R. Haser, M. Frey, F. Payan, and J. P. Astier. Crystal structure and electron transfer properties of cytochrome c_3. *J. Biol. Chem.*, 257:14341–14348, 1982.

[122] M. Pessanha, L. Brennan, A. V. Xavier, P. M. Cuthbertson, G. A. Reid, S. K. Chapman, D. L. Turner, and C. A. Salgueiro. NMR structure of the haem core of a novel tetrahaem cytochrome isolated from *Shewanella frigidimarina*: identification of the haem-specific axial ligands and order of oxidation. *FEBS Lett.*, 489:8–13, 2001.

[123] A. C. Messias. *NMR solution structures of cytochromes c_3: insight into the structural origins of functional cooperativities.* PhD thesis, Universidade Nova de Lisboa, 2000.

[124] P. Da Costa, C. Conte, and L.M. Saraiva. Expression of a *Desulfovibrio* tetraheme cytochrome *c* in *Escherichia coli*. *Biochem. Biophys. Res. Commun.*, 268:688–691, 2000.

[125] M. L. Herbaud, C. Aubert, M. C. Durand, F. Guerlesquin, L. Thöny-Meyer, and A. M. Dolla. *Escherichia coli* is able to produce heterologous tetraheme cytochrome c_3 when the *ccm* genes are co-expressed. *Biochim. Biophys. Acta*, 1481:18–24, 2000.

[126] L. ElAntak, O. Bornet, X. Morelli, A. Dolla, and F. Guerlesquin. Sequential NMR assignment of the ferri-cytochrome c_3 from *Desulfovibrio vulgaris* Hildenborough. *J. Biomol. NMR*, 23:69–70, 2002.

[127] M. Bruschi, C. E. Hatchikian, L. A. Golovleva, and J. LeGall. Purification and characterization of cytochrome c_3, ferredoxin, and rubredoxin isolated from *Desulfovibrio desulfuricans* Norway. *J. Bacteriol.*, 129:30–38, 1977.

[128] D. Marion, M. Ikura, R. Tschudin, and A. Bax. Rapid recording of 2D NMR spectra without phase cycling: application to the study of hydrogen exchange in proteins. *J. Magn. Reson.*, 85:393–399, 1989.

[129] J. Jeener, B. H. Meier, P. Bachmann, and R. R. Ernst. Investigation of exchange processes by two-dimensional n.m.r. spectroscopy. *J. Chem. Phys.*, 71:4546–4553, 1979.

[130] A. Kumar, R. R. Ernst, and K. Wüthrich. A two-dimensional nuclear overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.*, 95:1–6, 1980.

[131] M. Piotto, V. Saudek, and V. Sklenár. Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, 2:661–665, 1992.

[132] J. Briand and R. R. Ernst. Computer-optimized homonuclear TOCSY with suppression of cross relaxation. *Chem. Phys. Lett.*, 185:276–285, 1991.

[133] D. Marion and K. Wüthrich. Application of phase-sensitive two-dimensional correlated spectroscopy (COSY) for measurement of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.*, 113:967–974, 1983.

[134] M. Rance, O. W. Sorenson, G. Bodenhausen, G. Wagner, R. R. Ernst, and K. Wüthrich. Improved spectral resolution in COSY ^{1}H spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.*, 117:479–485, 1983.

[135] A. E. Derome and M. P. Williamson. Rapid-pulsing artifacts in double quantum filtered COSY. *J. Magn. Reson.*, 88:177–185, 1990.

[136] C. Bartels, T. Xia, M. Billeter, P. Güntert, and K. Wüthrich. The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. *J. Biomol. NMR*, 6:1–10, 1995.

[137] R. Koradi, M. Billeter, and K. Wüthrich. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, 14:51–55, 1996.

[138] A. L. Morris, M. W. MacArthur, E. G. Hutchinson, and J. M. Thornton. Stereochemical quality of protein structure coordinates. *Proteins*, 12:345–364, 1992.

[139] E. G. Hutchinson and J. M. Thornton. PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Sci.*, 5:212–220, 1996.

[140] W. P. Aue, E. Bartholdi, and R. R. Ernst. Two-dimensional spectroscopy: application to nuclear magnetic resonance. *J. Chem. Phys.*, 64:2229–2246, 1976.

[141] L. Braunschweiler and R. R. Ernst. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.*, 53:521–528, 1983.

[142] R. O. Louro, E. C. de Waal, M. Ubbink, and D. L. Turner. Replacement of the methionine axial ligand in cytochrome c_{550} by a lysine: effects on the haem electronic structure. *FEBS Lett.*, 510:185–188, 2002.

[143] P. A. Rice. Making DNA do a U-turn: IHF and related proteins. *Curr. Opin. Struct. Biol.*, 7:86–93, 1997.

[144] X. Jia, A. Grove, M. Ivancic, V. L. Hsu, E. P. Geiduscheck, and D. R. Kearns. Structure of the *Bacillus subtilis* phage SPO1-encoded type II DNA-binding protein TF1 in solution. *J. Mol. Biol.*, 263:259–268, 1996.

[145] K. K. Swinger and P. A. Rice. IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.*, 14:28–35, 2004.

[146] A. Travers. DNA-protein interactions: IHF - the master bender. *Curr. Biol.*, 7:R252–R254, 1997.

[147] P. A. Rice, S. W. Yang, K. Mizuuchi, and H. A. Nash. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, 87:1295–1306, 1996.

[148] K. K. Swinger, K. M. Lemberg, Y. Zhang, and P. A. Rice. Flexible DNA bending in HU-DNA cocrystal structures. *EMBO J.*, 22:3749–3760, 2003.

[149] H. Vis. *Structure and dynamics of the DNA binding protein HU of Bacillus stearothermophilus by NMR*. PhD thesis, Universiteit Utrecht, 1996.

[150] I. Tanaka, K. Appelt, K. Dijk, and S. W. White. 3Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature*, 310:376–381, 1984.

[151] S. W. White, K. Appelt, K. S. Wilson, and I. Tanaka. A protein structural motif that bends DNA. *Proteins*, 5:281–288, 1989.

[152] H. Vis, M. Mariani, C. E. Vorgias, K. S. Wilson, R. Kaptein, and R. Boelens. Solution structure of the HU protein from *Bacillus stearothermophilus*. *J. Mol. Biol.*, 254:692–703, 1995.

[153] E. Christodoulou, W. R. Rypniewski, and C. E. Vorgias. High-resolution X-ray structure of the DNA-binding protein HU from the hyper-thermophilic *Thermotoga maritima* and the determinants of its thermostability. *Extremophiles*, 7:111–122, 2003.

[154] R. Jaenicke and G. Böhm. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.*, 8:738–748, 1998.

[155] R. Jaenicke. Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc. Natl. Acad. Sci. USA*, 97:2962–2964, 2000.

[156] E. Christodoulou and C. E. Vorgias. The thermostability of DNA-binding protein HU from the mesophilic, thermophilic, and extreme thermophilic bacteria. *Extremophiles*, 6:21–31, 2002.

[157] A. Grove and L. Lim. High-affinity DNA binding of HU protein from the hyperthermophile *Thermotoga maritima*. *J. Mol. Biol.*, 311:491–502, 2001.

[158] D. Esser, H. Amanuma, A. Yoshiki, M. Kusakabe, R. Rudolph, and G. Böhm. A hyperthermostable bacterial histone-like protein as an efficient mediator for transfection of eukaryotic cells. *Nat. Biotechnol.*, 18:1211–1213, 2000.

[159] P. M. Padas, K. S. Wilson, and C. E. Vorgias. DNA binding protein from mesophilic and thermophilic *Bacilli*: cloning, overexpression and purification. *Gene*, 117:39–44, 1992.

[160] E. Christodoulou and C. E. Vorgias. Cloning, overproduction, purification and crystallization of the DNA binding protein HU from the hyperthermophilic eubacterium *Thermotoga maritima*. *Acta Cryst. D*, 54:1043–1045, 1998.

[161] A. Bax, M. Ikura, L. E. Kay, D. A. Torchia, and R. Tschudin. Comparison of different modes of two-dimensional reverse correlation NMR for the study of proteins. *J. Magn. Reson.*, 86:304–318, 1990.

[162] S. Grzesiek and A. Bax. Improved 3D triple resonance NMR techniques applied to a 31kDa protein. *J. Magn. Reson.*, 96:432–440, 1992.

[163] R. T. Clubb, V. Thanabal, and G. Wagner. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^{1}H^{N}$, ^{15}N, and $^{13}C'$ chemical shifts in ^{15}N-^{13}C-labeled proteins. *J. Magn. Reson.*, 97:213–217, 1992.

[164] S. Grzesiek and A. Bax. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N enriched proteins. *J. Biomol. NMR*, 3:185–204, 1993.

[165] A. Bax, G. M. Clore, and A. M. Gronenborn. ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization: a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.*, 88:425–431, 1990.

[166] F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, 6:277–293, 1995.

[167] B. A. Johnson and R. A. Blevins. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, 4:603–614, 1994.

[168] D. G. Higgins, J. D. Thompson, and T. J. Gibson. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, 266:383–402, 1996.

[169] P. J. Kraulis. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, 24:946–950, 1991.

[170] M. Ikura, L. E. Kay, and A. Bax. A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, 29:4659–4667, 1990.

[171] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay. Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, 33:5984–6003, 1994.

[172] N. A. Farrow, O. Zhang, J. D. Forman-Kay, and L. E. Kay. A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay

and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR*, 4:727–734, 1994.

[173] N. A. Farrow, O. Zhang, J. D. Forman-Kay, and L. E. Kay. Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry*, 34:868–878, 1995.

[174] M. Tollinger, N. R. Skrynnikov, F. A. A. Mulder, J. D. Forman-Kay, and L. E. Kay. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.*, 123:11341–11352, 2001.

[175] G. Otting, E. Liepinsh, and K. Wüthrich. Disulfide bond isomerization in BPTI and BPTI(G36S): an NMR study of correlated mobility in proteins. *Biochemistry*, 32:3571–3582, 1993.

[176] U. Reimer and G. Fischer. Local structural changes caused by peptidyl-prolyl *cis/trans* isomerization in the native state of proteins. *Biophys. Chem.*, 96:203–212, 2002.

[177] H. Vis, R. Boelens, M. Mariani, R. Stroop, C. E. Vorgias, K. S. Wilson, and R. Kaptein. ^{1}H, ^{13}C, and ^{15}N resonance assignments and secondary structure analysis of the HU protein from *Bacillus stearothermophilus* using two- and three-dimensional double- and triple-resonance heteronuclear magnetic resonance spectroscopy. *Biochemistry*, 33:14858–14870, 1994.

[178] M. Schubert, L. J. Ball, H. Oschkinat, and P. Schmieder. Bridging the gap: a set of selective ^{1}H-^{15}N-correlations to link sequential neighbors of prolines. *J. Biomol. NMR*, 17:331–335, 2000.

[179] M. Schubert, D. Labudde, H. Oschkinat, and P. Schmieder. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. *J. Biomol. NMR*, 24:149–154, 2002.

[180] R. T. Dame and N. Goosen. HU: promoting or counteracting DNA compaction? *FEBS Lett.*, 529:151–156, 2002.

[181] S. W. White, K. S. Wilson, K. Appelt, and I. Tanaka. The high-resolution structure of the DNA-binding protein HU from *Bacillus stearothermophilus*. *Acta Cryst. D*, 55:801–809, 1999.

[182] H. Vis, C. E. Vorgias, K. S. Wilson, R. Kaptein, and R. Boelens. Mobility of NH bonds in DNA-binding protein HU of *Bacillus stearothermophilus* from reduced spectral density mapping at multiple NMR fields. *J. Biomol. NMR*, 11:265–277, 1998.

[183] A. G. Palmer, J. Cavanagh, P. E. Wright, and M. Rance. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson.*, 93:151–170, 1991.

[184] A. G. Palmer, M. Rance, and P. E. Wright. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance ^{13}C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.*, 113:4371–4380, 1991.

[185] M. J. Stone, W. J. Fairbrother, A. G. Palmer, J. Reizer, M. H. Saier Jr., and P. E. Wright. Backbone dynamics of the *Bacillus subtilis* glucose permease IIA domain determined from ^{15}N NMR relaxation measurments. *Biochemistry*, 31:4394–4406, 1992.

[186] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical recipes in* FORTRAN*: the art of scientific computing*. Cambridge University Press, Cambridge, 1986.

[187] P. Závodszky, J. Kardos, A. Svingor, and G. A. Petsko. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. USA*, 95:7406–7411, 1998.

[188] J. Fitter and J. Heberle. Structural equilibrium fluctuations in mesophilic and thermophilic α-amylase. *Biophys. J.*, 79:1629–1636, 2000.

[189] G. Hernández, F. E. Jenney Jr., M. W. W. Adams, and D. M. LeMaster. Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *Proc. Natl. Acad. Sci. USA*, 97:3166–3170, 2000.

[190] G. Hernández and D. M. LeMaster. Reduced temperature dependence of collective conformational opening in a hyperthermophile rubredoxin. *Biochemistry*, 40:14384–14391, 2001.

[191] S. Kawamura, Y. Kakuta, I. Tanaka, K. Hikichi, S. Kuhara, N. Yamasaki, and M. Kimura. Glycine-15 in the bend between two α-helices can explain the thermostability of DNA binding protein HU from *Bacillus stearothermophilus*. *Biochemistry*, 35:1195–1200, 1996.

[192] S. Kawamura, Y. Abe, T. Ueda, K. Masumoto, T. Imoto, N. Yamasaki, and M. Kimura. Investigation of the structural basis for thermostability of DNA-binding protein HU from *Bacillus stearothermophilus*. *J. Biol. Chem.*, 273:19982–19987, 1998.

[193] D. Esser, R. Rudolph, R. Jaenicke, and G. Böhm. The HU protein from *Thermotoga maritima*: recombinant expression, purification and physicochemical characterization of an extremely hyperthermophilic DNA-binding protein. *J. Mol. Biol.*, 291:1135–1146, 1999.

[194] D. Serban, S. F. Arcineigas, C. E. Vorgias, and G. J. Thomas Jr. Structure and dynamics of the DNA-binding protein HU of *Bacillus stearothermophilus* investigated by Raman and ultraviolet-resonance Raman spectroscopy. *Protein Sci.*, 12:861–870, 2003.

[195] H. Welfle, R. Misselwitz, K. Welfle, N. Groch, and U. Heinemann. Salt-dependent and protein-concentration-dependent changes in the solution structure of the DNA-binding histone-like protein, HBsu, from *Bacillus subtilis*. *Eur. J. Biochem.*, 204:1049–1055, 1992.

[196] H. Vis, U. Heinemann, C. M. Dobson, and C. V. Robinson. Detection of a monomeric intermediate associated with dimerization of protein Hu by mass spectrometry. *J. Am. Chem. Soc.*, 120:6427–6428, 1998.

Appendix A

Chemical shift data for c3*Dn*

	haem 1		haem 2		haem 3		haem 4	
proton	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H	^{13}C
QM1	20.34	-48.8	7.51	-29.6	13.69	-31.3	9.07	-28.6
HT2A	2.41	nd	nd	nd	-2.20	-28.6	-1.26	nd
QT2	0.67	nd	nd	nd	-3.33	78.2	0.20	nd
HAM	11.88	nd	6.07	29.3	-3.89	nd	6.38	nd
QM3	8.73	-14.9	19.81	-34.0	14.00	-37.1	14.30	-32.2
HT4A	nd	-43.5	1.28	nd	-0.45	nd	nd	nd
QT4	-4.00	80.2	2.28	nd	-0.86	nd	nd	nd
HBM	-5.99	52.6	-1.98	nd	nd	nd	nd	nd
QM5	17.21	-47.9	11.33	-35.3	19.07	-33.1	8.83	-31.4
HA62	0.66	-5.6	-0.80	-1.3	24.45	-67.2	-1.22	-13.7
HA63	-3.71	-5.6	0.74	-1.3	19.71	-67.2	0.55	-13.7
HB62	2.01	70.8	-0.07	51.0	-0.77	169.2	-0.44	77.3
HB63	nd	70.8	0.09	51.0	-0.49	169.2	0.66	77.3
HGM	10.83	nd	6.54	nd	nd	nd	nd	nd
HA72	7.53	-5.5	11.42	-22.7	3.32	-13.9	13.20	-32.1
HA73	2.49	-5.5	8.68	-22.7	3.56	-13.9	3.11	-32.1
HB72	2.10	nd	2.20	94.7	-2.82	75.5	-0.23	111.3
HB73	nd	nd	2.14	94.7	-2.29	75.5	2.82	111.3
QM8	27.28	-58.1	27.76	-57.7	-0.94	-2.4	19.69	-43.8
HDM	-3.02	49.9	-0.57	nd	nd	nd	0.38	nd

Table A.1: ^{1}H and ^{13}C chemical shifts (ppm) of haem resonances of c3*Dn*.

Table A.2: Proton chemical shifts (ppm) for assigned residues in c3*Dn*.

residue	NH	Hα	Hβ	others
A1		3.13	0.01	
D2	7.19	3.39	1.81,1.51	
A3	7.06	-0.90	0.87	
P4				
G5	9.12	5.17,4.75		
D6	9.85	6.06	3.52,3.27	
D7	9.43	4.92	3.10,2.95	
Y8		5.01	4.04,3.17	Hδ 7.59,7.59 Hε 6.69,6.69
V9	8.02	4.44	1.85	1.07,1.03
I10	8.81	3.47	0.54	(0.75,-0.73) -1.21 -2.46
S11	7.49	4.91	3.88,3.79	

continued on next page

continued from previous page

residue	NH	Hα	Hβ	others
A12	9.00	3.57	1.05	
P13		4.26		
E14	8.38	3.98	2.30,1.94	
G15	8.60	4.19,3.59		
M16	7.47	4.25		
K17	8.33	4.43		
A18	7.95	4.21	0.68	
K19	7.87			
P20				
K21				
G22	8.29	3.85,3.51		
D23	8.51	4.45	2.74,2.63	
K24	7.61	4.67		
P25				
G26	8.30	3.74,3.21		
A27	7.83	4.51	1.26	
L28	8.51	4.18	1.87	1.64,0.92,0.40
Q29	11.00	6.09	5.03	Hε 6.50,5.36
K30	11.21	5.86	4.04	2.76,2.66,2.54,3.66
T31	10.25	6.70	5.15	2.15
V32	9.09	6.70		
P33				
F34	8.57	4.77	2.08,1.63	
P35	6.25			
H36				
T37	12.09	5.84	5.19	4.80
K38	8.37	5.48		
H39				
A40	9.02	5.63	2.42	
T41	8.54	4.58	1.61	
V42	8.32	3.99	2.24	0.25,0.04
E43	8.55			
C44				
V45	8.62	5.47	2.93	2.03,1.47
Q46	8.08	4.21	1.86,1.66	1.34, Hε 6.94,6.37
C47	6.55	3.09	1.33,-1.98	
H48				
H49				

continued on next page

continued from previous page

residue	NH	Hα	Hβ	others
T50	11.38	7.10	6.06	
L51	9.44	4.90	3.26,2.55	2.25,2.03,0.95
E52	9.28	4.40	2.48	2.71,2.58
A53	8.63	4.40	1.77	
D54	8.63	4.90	3.23	
G55	8.18	4.57,3.98		
G56	8.67	4.07,3.25		
A57	6.92	3.98	1.17	
V58	7.95	3.98	0.96	0.46,-1.94
K59	9.00			
K60	8.81	4.05		
C61	7.96	2.91	1.20,0.49	
T62	6.81	5.30	4.89	1.47
T63	7.05	4.04		1.40
S64	8.22	4.38	3.85,3.78	
G65	9.14	3.92,3.19		
C66	8.23	1.98	3.35,3.22	
H67	11.17	8.41	10.59,7.75	
D68	8.72	5.04	3.30,2.82	
S69	8.15	4.72	4.30,4.24	
L70	9.37	5.00	1.91,1.58	1.72,0.93,0.72
E71	8.84	4.54	2.43	
F72	10.38	5.05	2.93	7.38,7.38,Hε 6.88,6.88
R73				
D74	7.24	4.75	3.12,2.62	
K75				
A76				
N77	7.81	5.13	3.61,2.90	Hε 7.74,7.07
A78	8.57	4.36	1.52	
K79	8.66	4.94		
D80	8.09	4.86	3.64,2.91	
I81	9.16	5.56	3.40	2.43,1.45,0.80
K82	9.78	6.58	2.76,2.69	1.92,2.87,2.52
L83	8.41	5.42	2.68,2.57	1.41,1.18
V84	9.44	3.78	1.62	1.36,-0.54
E85				
N86				
A87	9.96	4.43	2.89	

continued on next page

continued from previous page				
residue	NH	Hα	Hβ	others
F88	7.13	1.04	-1.89,-3.17	4.17,4.17,Hε 4.81,5.28
H89	10.99	7.33	18.11,16.42	
T90	9.10	4.58	4.08	1.24
Q91	5.45	1.69		
C92	6.42	1.31	-0.07,-1.04	
I93	8.11	4.99	3.79	1.95
D94				
C95	6.06	3.98	1.23,-0.52	
H96				
K97	10.16	4.53	2.48,2.33	2.25,2.14,1.95,3.36
A98	8.13	4.23	1.50	
L99	8.20	4.10	2.31	1.84,1.11,0.81
K100	7.93	4.17	2.31	
K101	8.03	4.10	2.08,2.03	1.62,1.54,3.06
D102	7.29	4.74	2.84,2.54	
K103	7.94	3.82	2.02,1.84	1.32
K104	7.77	4.55	1.81	1.45,1.32
P105				
T106	7.23	3.99	3.87	1.03
G107				
P108				
T109	9.10	4.90		
A110	7.14	3.88	0.84	
C111	6.38	0.59	-1.35,-1.79	
G112	8.38	4.75,3.67		
K113	6.78	4.36	1.45,1.34	1.16,1.01
C114	7.23	5.82	2.82,2.28	
H115	11.48	10.36	13.87,9.49	
T116		4.81	2.31	
T117	9.96	5.69	4.87	1.47
N118	9.88	5.39	3.52,3.42	7.80,7.37

Appendix B

Chemical shift data for HU*Tmar*

Table B.1: Chemical shifts (ppm) of ^{1}H, ^{13}C, and ^{15}N nuclei in HU*Tmar*.

residue	^{15}N	C=O	Cα	Cβ	others
M1		180.6	52.6(4.48)	31.8	
b		180.3	52.7(4.38)	31.4(2.22)	Cγ 28.4(2.65,2.41)
N2	127.1(8.84)	173.2	48.5(6.04)	36.8(3.64,3.10)	Nδ n.d.(7.60,7.34)
b	127.7(8.70)	172.7	48.3(5.49)	36.9(3.63,3.09)	
K3	119.1(8.63)	174.4	58.7(3.79)	31.8(2.09,n.d.)	Cγ 23.6(1.84), Cε n.d.(2.89)
b	119.1(8.68)	174.4	58.8	31.0	
K4	118.4(7.90)	176.6	57.8(3.95)	29.8(2.03,1.95)	Cγ 22.4(1.66,1.53), Cδ n.d.(1.80)
b	118.6(7.95)	176.5	57.8(3.94)	29.8(1.98)	
E5	119.8(8.33)	176.1	56.4(4.19)	28.0(2.35,n.d.)	Cγ 34.0(2.61,2.40)
b	118.8(8.34)	176.1	56.3	27.9	
L6	122.1(8.55)	175.3	56.5(4.18)	39.0(2.21,1.55)	Cγ 24.7(0.91), Cδ 21.5(0.99)
b	120.9(8.33)	175.3	56.5	39.1	
I7	118.4(8.28)	174.7	63.6(3.42)	35.3(2.05)	Cγ 28.6(1.93,0.74), Cγ 14.7(0.97), Cδ 12.0(0.97)
b	118.9(8.26)	174.6	63.5	35.3	
D8	119.3(8.01)	176.8	55.6(4.38)	38.6(2.90,2.79)	
b	119.4(7.97)	176.6	55.6	38.7	
R9	118.8(8.57)	177.0	57.0(4.13)	28.6(2.05,1.94)	Cγ 25.1, Cδ 41.1
b	119.1(8.59)	176.8	56.8(4.16)	28.1(2.01,n.d.)	
V10	120.0(8.66)	174.6	63.5(3.59)	28.7(2.26)	Cγ 22.5(1.05), Cγ 19.4(0.95)
b	120.0(8.66)	174.6	63.4	28.6	
A11	121.7(8.90)	177.2	53.9(3.92)	15.3(1.57)	
b	121.7(8.86)	177.2	53.9(3.89)	15.3(1.57)	
K12	115.8(7.64)	176.7	56.6(4.17)	30.0(2.04,n.d.)	
b	116.3(7.71)	176.8	56.6(4.14)	29.2(2.03,n.d.)	
K13	119.1(7.99)	175.6	56.7(4.13)	31.1(1.88,n.d.)	
b	119.1(8.07)	175.6	56.6(4.14)	n.d.(1.89)	
A14	117.0(8.66)	175.1	49.9(4.14)	15.9(1.39)	
b	117.0(8.66)	175.0			
G15	108.5(7.82)	171.5	44.0(3.99,n.d.)		
b		171.5	44.0(3.94,n.d.)		
A16	120.9(7.62)	173.6	47.3(4.72)	20.2(1.26)	
b	121.2(7.74)	173.6	47.2(4.72)	n.d.(1.27)	
K17	119.3(8.99)	176.5	53.5(4.41)	29.9(2.07,1.92)	Cγ 22.5(1.74,1.64), Cδ 26.6, Cε 39.7
K18	125.2(9.02)	175.9	59.0(3.68)	29.4(1.91,1.85)	Cγ 24.0(1.69,1.36), Cδ 26.9, Cε 39.6(3.10)
K19	116.5(8.71)	175.8	57.0(4.11)	29.5(1.98,1.82)	Cγ 22.0, Cδ 26.8, Cε 39.6(3.10)
b		175.9			
D20	120.2(6.92)	175.3	54.6(4.63)	37.8(2.87,2.69)	
b	120.2(6.92)	175.2			
V21	119.3(7.97)	174.4	64.8(3.43)	28.6(2.25)	Cγ 20.6(0.93), Cγ 19.4(1.02)
K22	118.8(8.66)	174.2	58.2(3.76)	30.2(1.99,1.89)	Cγ 22.7, Cδ 26.8, Cε 39.9
L23	117.9(7.44)	178.1	55.6	39.4	
b	117.9(7.43)	178.0	55.7(4.38)	39.5(2.02,1.87)	Cγ 25.2(1.54), Cδ 22.8(1.11), Cδ 20.8(1.06)
I24	121.2(8.48)		62.9(3.46)	n.d.(1.99)	
b	120.9(8.31)	174.8	62.6(3.53)	34.2(2.02)	Cγ 27.2(1.65,1.20), Cγ 15.4(0.92), Cδ 8.9(0.70)
L25	122.1(9.32)	175.0	56.6(3.96)	38.7	Cγ 24.9, Cδ 20.2
b	122.1(9.31)	175.1	56.7(4.01)	39.2(2.15)	Cγ 24.9, Cδ 20.9(0.95)
D26	119.1(8.89)	177.0	55.4(4.42)	36.6(2.89,2.63)	
b	119.3(8.99)	177.0	55.5(4.46)	36.7(2.92,2.68)	
T27	116.6(8.18)		64.8(4.11)	n.d.(4.36)	Cγ n.d.(1.41)
b	116.5(8.07)	175.5	64.7(4.16)	65.0(4.38)	Cγ 21.1(1.44)
I28	128.2(9.20)	177.8	64.1(3.49)	35.3	Cδ 13.7
b	128.2(9.20)	174.4	64.1(3.52)	35.5(2.34)	Cγ 28.3(2.13), Cγ 14.4(0.83), Cδ 13.9
L29	118.4(8.35)	177.8	55.7(3.55)	37.1(1.22,n.d.)	Cγ n.d.(1.87), Cδ 23.6(0.74) Cδ 16.2(-0.26)
b	118.8(8.49)	177.3	56.0(3.81)	37.4(1.91,1.38)	Cγ 23.2(0.82), Cδ 18.4(0.33)

continued on next page

continued from previous page

residue	^{15}N	C=O	Cα	Cβ	others
E30	121.7(8.53)	175.8	57.1(4.11)	27.4(2.19,n.d.)	Cγ 33.7
b	120.5(8.37)	175.9	57.1(4.13)	27.4(2.27,2.17)	Cγ 33.9(2.49,2.29)
T31	117.7(8.19)	174.2	65.1(3.95)	65.4(4.19)	Cγ n.d.(1.23)
b	117.2(8.09)	174.3	65.0(3.95)	65.7(4.20)	Cγ 19.5(1.25)
I32	121.9(8.36)	174.0	64.0(3.61)	35.7	
b	122.6(8.43)	174.2	63.8(3.56)	35.6(1.89)	Cγ 26.7, Cγ 14.4(0.68), Cδ 10.6(0.36)
T33	115.8(8.09)	172.4	65.9(3.60)	n.d.(4.42)	Cγ n.d.(1.27)
b	115.6(8.29)		65.2(3.86)	66.1(4.42)	Cγ 19.1(1.38)
E34	120.0(8.42)	175.9	56.7(4.01)	27.3(2.25,2.13)	Cγ 33.4
b	121.9(8.23)	175.5	56.9(4.10)	27.3(2.25,2.13)	Cγ 33.9
A35	121.4(7.66)	177.3	52.7(4.21)	15.4(1.43)	
b	122.1(7.65)	177.4	52.7(4.21)	15.3(1.42)	
L36	118.4(8.26)	178.2	55.3(4.29)	39.1	Cδ 19.4
b	117.9(8.39)	179.2	55.2(4.23)	39.0(1.94)	Cγ 31.5, Cδ 23.5, Cδ 20.2
A37	121.9(8.65)	176.3	52.6(4.15)	15.7(1.62)	
b	122.6(8.24)	176.8	52.3(4.24)	15.5(1.63)	
K38	115.4(7.57)	174.5	53.5(4.55)	30.3(2.14,n.d.)	Cγ 22.5(1.70), Cε 39.7
b	116.0(7.61)	174.4	53.6(4.49)	30.3(2.12)	Cγ 22.6(1.67), Cδ 26.8, Cε 39.6
G39	108.8(8.20)	171.6	43.1(4.29,3.67)		
b	108.1(7.97)	171.9	43.1(4.30,3.85)		
E40	123.1(8.32)	172.2	53.2(4.39)	28.4	
b	121.2(8.10)	172.9	53.4(4.18)	28.2(2.26)	Cγ 34.1
K41	120.3(8.19)	173.8	53.3(4.84)	31.7(1.92,1.86)	Cγ n.d.(1.55) Cδ n.d.(1.77) Cε 39.5
b	121.0(8.32)	173.8	53.1(4.48)	30.9(1.80)	Cγ 22.3(1.50) Cδ 26.6(1.77) Cε 39.7(3.08)
V42	123.1(9.09)	172.0	59.1(4.88)	31.2(2.39)	Cγ 18.5(1.10), Cγ n.d.(1.04)
b	124.5(8.81)	172.1	59.8(4.49)	29.0(2.23)	Cγ 19.1(1.18), Cγ 18.0(1.09)
Q43	128.2(9.14)	171.2	52.8(4.85)	28.5(2.35,2.08)	
b	128.2(8.48)		52.3		
I44	124.0(8.69)	172.8	57.4(4.47)	36.1(2.54)	Cγ 16.1(1.02)
V45	129.6(8.44)		61.9(4.67)	28.7(2.35)	Cγ 19.1(1.20), Cγ 17.3(1.14)
G46	116.8(8.96)	170.8	43.1(3.82,n.d.)		
F47	121.7(8.41)	171.2	58.1(4.90)	39.6(3.18,2.95)	
G48	106.0(8.63)	168.4	43.3(4.28,3.84)		
b		170.0	42.7		
S49	112.3(7.95)	170.4	53.3(5.88)	63.8(3.79,3.69)	
b	113.7(8.26)		54.9		
F50	121.5(9.69)	171.7	54.1(5.66)	40.2(3.12,3.12)	
E51	119.3(9.11)	171.5	51.8(4.92)	31.4(2.16,2.00)	Cγ 33.0
V52	118.1(8.85)	173.2	58.7(4.88)	31.0(2.00)	Cγ 19.7(1.05), Cγ 19.5(0.91)
b		173.3	59.6	30.2(2.08)	
R53	125.4(8.79)	172.3	51.3(4.80)	30.6(1.98,1.72)	Cγ 22.2
b	125.9(8.47)	173.3	53.2(4.46)	28.3(1.86,1.77)	
K54	123.5(8.68)	173.3	53.8(4.75)	30.7(1.79,nd)	
b	124.0(8.44)	173.6	53.8(4.38)	30.6(1.90,1.83)	
A55	129.4(9.02)	173.7	48.8(4.73)	17.6(1.45)	
b	125.6(8.30)	174.7	49.9(4.70)	16.8(1.46)	
A56	123.8(8.23)	175.2	49.6(4.36)	17.1(1.52)	
b	123.3(8.26)	174.8	49.8(4.36)	16.8(1.46)	
A57	123.8(8.58)	175.0	50.1(4.45)	16.9(1.48)	
b	123.3(8.21)	175.0	49.9(4.38)	16.8(1.46)	
R58	119.6(8.66)	172.4	52.6(4.72)	30.3(1.95,1.82)	
b	120.5(8.27)	173.6	53.5	28.1	
K59	122.1(8.50)	174.1	53.3(4.85)	31.1(1.96,1.86)	Cγ 22.3(1.54), Cδ 26.2, Cε 39.6
b	122.9(8.35)	174.4	53.9(4.42)	30.6(1.94,1.86)	Cγ 22.2, Cδ 26.5, Cε39.6
G60	110.7(8.55)	170.7	42.3(4.44,3.88)		
b	110.4(8.45)	171.3	42.7		
V61	119.3(8.36)	172.7	58.5(4.51)	32.3(1.89)	Cγ 18.1(0.87), Cγ n.d.(0.79)

continued on next page

continued from previous page					
residue	^{15}N	C=O	Cα	Cβ	others
b	118.8(7.96)	173.1	59.4(4.23)	30.4(2.12)	
N62	126.3(8.70)	172.1	48.6(4.96)	36.3(3.25,2.74)	Nδ n.d.(7.63,7.00)
b	123.8(8.58)		48.8(5.07)	36.4(2.98,2.80)	
P63	n.d.	174.9	62.0(4.41)	29.5(2.43,2.08)	Cγ 24.8, Cδ 48.3(4.12,3.91)
b		174.5	61.1(4.49)	29.6(2.39,2.05)	Cγ 24.7, Cδ 48.3
Q64	116.3(8.29)	174.6	55.3(4.14)	26.7(2.25,n.d.)	Cγ 31.6(2.50,n.d.), Nε n.d. (7.55,6.89)
b	119.4(8.44)	173.8	53.7(4.43)	26.8(2.22,2.11)	Cγ 31.6(2.49)
T65	108.5(7.69)	172.6	59.1(4.40)	67.9(4.29)	Cγ 18.8(1.26)
b	114.6(8.06)	171.8	59.5(4.39)	67.2(4.94)	
R66	117.0(8.47)	172.6	54.7(3.96)	24.8(2.10,n.d.)	Cγ n.d.(2.38), Cδ 40.5(3.25)
b	123.3(8.28)	173.1	53.5(4.42)	28.2(1.93,1.87)	Cγ 24.6(1.71), Cδ 40.9(3.28)
K67	120.1(7.73)	171.6	51.4(4.64)	30.4(1.98,n.d.)	Cγ 21.9(1.74), Cδ 26.8(1.53), Cε 39.6
b	124.1(8.31)	171.8	51.7	n.d.(1.84)	
P68	n.d.	174.0	60.6(4.85)	29.7(2.39,1.98)	Cγ 25.0, Cδ 48.2
b		174.1	60.5(4.55)	29.5(2.35,1.97)	Cγ 24.8, Cδ 48.2
I69	119.3(8.27)	172.8	57.6(4.58)	39.0(1.86)	Cγ 24.2, Cγ 15.5(0.96), Cδ 11.1(0.87)
b	121.2(8.23)	173.8	58.6(4.30)	30.5(1.94)	Cγ 24.8(1.58), Cγ 10.5(1.29), Cδ 15.0(0.99)
T70	119.1(8.44)	171.2	59.6(4.63)	67.3(4.10)	Cγ n.d.(1.17)
b	119.1(8.22)	171.3	59.1(4.46)	67.4	
I71	129.4(8.87)	171.3	55.8(4.57)	36.3(1.95)	Cγ 24.5(1.58,1.28), Cγ 14.8(1.03), Cδ 10.4(0.96)
b	125.4(8.26)	171.9	56.1(4.62)	36.1(1.98)	Cγ 24.3(1.61), Cγ n.d.(1.04)
P72	n.d.	173.9	60.2(4.54)	30.1(2.40,2.06)	Cγ 24.6, Cδ 48.5(3.96,3.78)
b		174.2	60.8(4.48)	29.6(2.38,1.97)	Cγ 24.8, Cδ 48.5
E73	120.5(8.50)	173.3	54.3(4.43)	28.3(2.08,2.04)	Cγ 33.9(2.39,2.32)
b	121.5(8.44)	173.8	54.3(4.33)	n.d.(2.37,2.09)	
R74	121.9(8.26)	172.4	52.5(4.77)	30.4(1.90,n.d.)	Cγ 23.6, Cδ 41.0
b			53.4	28.3	
K75	122.8(8.77)	173.2	53.6(4.87)	31.1(1.88,1.81)	Cγ 22.4(1.48), Cδ 26.7, Cε 39.6(3.08)
b	123.5(8.40)	173.6	53.5(4.44)	30.4(1.84)	
V76	118.1(8.80)	171.0	55.8(4.89)	32.6(2.15)	Cγ 18.8(0.98), Cγ 17.5
b	123.1(8.15)	171.8	57.3(4.51)	30.2(2.16)	Cγ 18.4(1.03), Cγ 17.7
P77	n.d.	172.6	59.5(5.18)	29.7	Cγ 22.7, Cδ 48.3(4.13,3.86)
K78	121.9(9.38)	171.7	52.2(4.95)	33.9(1.82,1.80)	Cγ 23.2(1.43), Cδ 26.6, Cε 39.5(3.04)
F79	123.1(8.83)	171.4	52.1(5.46)	40.6(2.96,2.52)	
K80	129.6(8.65)	182.0	49.1(4.70)	n.d.(1.63,1.49)	
P81	n.d.	173.3	60.0(3.86)	30.4	Cγ 24.7, Cδ 48.5
b		174.8	60.8(4.43)	29.6(2.41,2.04)	Cγ 24.7, Cδ 48.1
G82	111.6(8.42)	n.d.	41.3(4.29,3.53)		
b	109.2(8.43)		42.7		
K83	n.d.	175.7	57.8(3.88)	30.3	
A84	119.6(8.45)	177.4	52.4(4.25)	15.6(1.51)	
b		174.9	49.9(4.40)	16.6(1.45)	
L85	118.4(7.20)	175.8	55.3(4.21)	39.7(1.95,1.54)	Cγ 20.9(1.00), Cδ 23.5(0.84)
b	121.7(8.15)	174.7	52.8(4.40)	39.9(1.70,1.68)	Cγ 24.5, Cδ 22.4(1.02), Cδ 21.2(0.97)
K86	115.6(7.99)	176.5	57.9(3.88)	29.8(1.96,n.d.)	Cγ 22.2
b	121.9(8.24)	173.8	53.8(4.38)	30.5(1.92,1.82)	Cγ 22.2, Cδ 26.6 Cε 39.6
E87	116.3(8.02)	175.5	55.9(4.15)	27.2(2.22,2.17)	Cγ 33.8(2.47,2.44)
b	121.9(8.33)	173.5	54.0(n.d.)	28.0(n.d.)	Cγ 33.8
K88	117.2(7.67)	174.9	55.1(4.29)	31.0(2.03,1.65)	
b	122.5(8.34)	173.7	53.8(4.41)	30.5(1.91,1.86)	Cγ 22.2, Cδ 26.7 Cε 39.6
V89	113.9(7.45)	170.9	59.3(4.24)	29.3(2.28)	Cγ 18.6(0.92), Cγ 17.7
b	121.9(8.17)	172.6	60.1(4.19)	30.3(2.18)	Cγ 18.6(1.03), Cγ 18.0
K90	125.2(7.15)	178.5	56.1(4.41)	n.d.(1.97,1.55)	
b	129.4(7.90)	178.4	55.4(4.24)	n.d.(1.89,n.d.)	Cγ n.d.(1.48), Cδ n.d. Cε n.d.(3.09)

Solution conditions: approximately 1mM HU*Tmar*, 200mM potassium chloride and 50mM potassium phosphate in a mixture of 95% H_2O and 5% D_2O at pH 5.6 (neat buffer, the protein changes the pH to 6.3) and 311K. ^{1}H chemical shifts of protons attached to the assigned heteronucleus are indicated in parentheses. Proton chemical shifts are referenced to internal TSP and the heteronuclei were referenced indirectly using published data [35]. The accuracy (ppm) of the chemical shifts are 0.04, 0.2 and 0.2 respectively. Assignments marked *b* denote the minor form.

Appendix C

Relaxation data for HU*Tmar*

Table C.1: Relaxation data for ^{15}N nuclei in HU*Tmar*. *b* denotes data for the minor form.

residue	NOE	error	R_1 (s^{-1})	± s.d.	R_2 (s^{-1})	± s.d.
M1						
N2	-0.2539	0.0585	1.2246	0.0298	13.7909	0.4807
b	-0.3384	0.0935	1.6519	0.0629	9.6044	0.5529
K3			1.3207	0.0228	14.0071	0.3218
b			1.3405	0.0155	13.9324	0.2344
K4	-0.4485	0.0257	1.2698	0.0125	11.2634	0.1710
b	-1.0609	0.0266	1.5079	0.0148	5.1292	0.1026
E5						
b	-0.2869	0.0600	1.8205	0.0372	9.6670	0.3049
L6	-0.2640	0.0431	1.3637	0.0211	14.1599	0.3366
I7						
D8						
R9	-0.1473	0.0615	1.8453	0.0314	9.9360	0.2516
b	-0.2272	0.0352	1.2808	0.0163	15.3539	0.2991
V10	-0.2962	0.0251	1.4844	0.0132	12.3186	0.1708
A11	-0.3409	0.0834	1.8170	0.0410	10.1807	0.3519
b	-0.1983	0.0346	1.3767	0.0151	14.7031	0.2513
K12	-0.1640	0.0510	1.8288	0.0219	9.8436	0.1825
b	-0.0413	0.0479	1.2739	0.0147	14.9674	0.2575
K13						
b	-0.3465	0.0300	1.2418	0.0140	14.1240	0.2412
A14	-0.1676	0.0290	1.4981	0.0115	11.5259	0.1353
G15						
A16	-0.2169	0.0596	1.4743	0.0255	8.5409	0.2220
b	-0.2016	0.0350	1.0732	0.0132	13.6505	0.2221
K17						
K18	-0.2752	0.0310	1.5426	0.0204	11.2658	0.2365
K19	-0.2708	0.0200	1.5352	0.0105	12.3849	0.1284
D20	-0.2590	0.0287	1.5322	0.0139	11.3059	0.1560
V21	0.7664 0	.0401				
K22			1.8884	0.0226	10.2436	0.1887
L23	-0.2408	0.0294	1.4641	0.0132	12.4251	0.1673
I24	-0.2312	0.0424	1.3342	0.0210	11.8602	0.3120
L25						
D26	-0.1603	0.0489	1.3361	0.0188	15.1917	0.3371

continued on next page

continued from previous page

residue	NOE	error	R_1 (s^{-1})	± s.d.	R_2 (s^{-1})	± s.d.
T27	-0.2146	0.0541	1.2757	0.0277	14.9908	0.4754
b	-0.2744	0.0706	1.5542	0.0439	10.2915	0.4282
I28	-0.1936	0.0607	1.3992	0.0281	14.4234	0.4424
L29	-0.2331	0.0537	1.3931	0.0263	11.9570	0.3626
E30	-0.2543	0.0375	1.3160	0.0181	14.7358	0.3082
b	-0.6310	0.0668	1.8068	0.0469	8.5306	0.3594
T31	-0.1926	0.0514	1.3463	0.0299	14.0413	0.4805
b	-0.3089	0.1142	1.8221	0.0798	9.5736	0.6174
I32	-0.3184	0.0440	1.3590	0.0264	13.2171	0.3826
b	-0.3135	0.0853	1.8951	0.0554	9.2944	0.4404
T33	-0.2655	0.0476	1.3326	0.0279	14.9489	0.4778
b	0.1193	0.3035	1.6864	0.1112	11.1865	1.0527
E34	-0.2534	0.0367	1.2839	0.0189	15.0886	0.3435
b	-1.4196	0.0368	1.4163	0.0196	5.6164	0.1510
A35	-0.1967	0.0332	1.3173	0.0150	14.1679	0.2419
b	-0.1688	0.0865	1.9039	0.0465	9.4703	0.3489
L36	-0.2945	0.0461	1.3906	0.0209	13.9301	0.3278
b	-0.3727	0.0985	1.8745	0.0639	9.4576	0.4893
A37	-0.2153	0.0403	1.2653	0.0167	16.0237	0.2998
b	-0.6101	0.0734	1.9805	0.0616	9.6440	0.4574
K38	-0.2445	0.0391	1.2486	0.0165	14.1103	0.2650
b	-0.2239	0.0569	1.7120	0.0249	9.2461	0.2067
G39	-0.2934	0.0409	1.3735	0.0197	11.9712	0.2817
b	-0.2551	0.0643	1.8245	0.0367	8.2210	0.2642
E40	-0.3017	0.0337	1.3199	0.0164	13.7381	0.2527
b	-0.4618	0.0572	1.7956	0.0322	9.1483	0.2520
K41	-0.5286	0.0304	1.2715	0.0149	10.0617	0.1910
V42	-0.2181	0.0659	1.2387	0.0268	14.7790	0.4489
Q43	-0.1930	0.0515	1.3292	0.0229	13.3063	0.3395
I44	-0.2203	0.0534	1.2325	0.0230	14.1148	0.3807
V45	-0.3008	0.0401	1.3130	0.0231	13.4837	0.3608
G46	-0.4341	0.0620	1.2351	0.0391	12.9623	0.6251
F47	-0.4603	0.0280	1.2578	0.0143	11.2602	0.1914
G48	-0.3031	0.0530	1.2564	0.0244	14.2304	0.4071
S49	-0.1904	0.0469	1.2420	0.0205	14.4937	0.3444
F50	-0.3744	0.0644	1.1888	0.0353	14.7151	0.5934
E51	-0.2690	0.0533	1.3900	0.0237	13.6557	0.3492
V52	-0.3180	0.0347	1.3034	0.0152	11.7151	0.2079

continued on next page

continued from previous page

residue	NOE	error	R_1 (s^{-1})	± s.d.	R_2 (s^{-1})	± s.d.
R53	-0.3591	0.0583	1.3852	0.0271	12.0327	0.3621
K54	-0.4633	0.0350	1.3509	0.0162	10.8120	0.2127
b	-0.6245	0.1129	1.5328	0.0822	5.8340	0.5650
A55	-0.5778	0.0378	1.4491	0.0186	10.8535	0.2212
b	-1.9371	0.0893	1.1242	0.0372	4.2749	0.3048
A56			1.4320	0.0095	7.8207	0.0969
b	-0.7906	0.0167				
A57	-0.8075	0.0180	1.4884	0.0101	7.2560	0.0956
b	-1.5677	0.0674	1.2351	0.0355	2.8904	0.2311
R58	-0.8193	0.0216	1.4738	0.0129	8.6537	0.1276
b	-1.5706	0.0708	1.3108	0.0385	2.6432	0.2488
K59	-0.8706	0.0187	1.4157	0.0108	7.9499	0.1055
G60	-1.0843	0.0174	1.4863	0.0100	5.7177	0.0836
b	-1.4779	0.0579	1.2091	0.0378	2.6242	0.2531
V61	-0.9656	0.0145	1.4049	0.0087	6.3481	0.0749
N62	-1.1914	0.0179	1.5570	0.0101	6.1530	0.0853
P63						
Q64	-1.0388	0.0157	1.5642	0.0098	6.4189	0.0791
b	-1.0803	0.0153	1.4395	0.0094	5.7848	0.0750
T65	-0.9538	0.0227	1.3333	0.0134	8.4567	0.1308
b	-1.7823	0.0884	1.2855	0.0443	2.5582	0.2928
R66	-1.1026	0.0208	1.6824	0.0152	8.3502	0.1341
b	-1.8658	0.0808	1.2745	0.0386	3.1489	0.2847
K67	-1.3028	0.0194	1.4243	0.0080	5.5881	0.0658
b	-1.949	0.0590	1.2733	0.0234	2.5306	0.1394
P68						
I69	-1.0767	0.0182	1.4512	0.0097	6.5596	0.0843
b	-1.6381	0.0387	1.3545	0.0161	2.9315	0.1036
T70						
b	-1.5059	0.0528	1.2915	0.0273	3.0918	0.1795
I71	-0.7991	0.0324	1.4815	0.0159	8.3783	0.1665
b	-1.6237	0.0731	1.3418	0.0284	2.2695	0.1768
P72						
E73	-0.7107	0.0198	1.4736	0.0105	8.1117	0.1059
b	-1.3023	0.0348	1.3141	0.0168	3.0395	0.1135
R74	-0.6207	0.0254	1.5191	0.0143	9.1999	0.1524
K75	-0.5124	0.0435	1.4602	0.0206	11.0840	0.2607
b	-1.5265	0.0724	1.3290	0.0337	2.6194	0.2015

continued on next page

continued from previous page						
residue	NOE	error	R_1 (s^{-1})	± s.d.	R_2 (s^{-1})	± s.d.
V76						
b	-1.7027	0.0719	1.3986	0.0262	2.5357	0.1469
P77						
K78	-0.3785	0.0481	1.3102	0.0211	12.9605	0.3028
F79	-0.3293	0.0477	1.3498	0.0216	13.4625	0.3270
K80	-0.3883	0.0522	1.1917	0.0247	14.1808	0.4162
P81						
G82	-0.3122	0.0589	1.2656	0.0312	13.5471	0.5037
K83						
A84	-0.4231	0.0268	1.3734	0.0151	13.4331	0.2270
L85	-0.3589	0.0376	1.3007	0.0165	12.6972	0.2388
b	-1.796	0.0708	1.1897	0.0290	2.5485	0.2020
K86	-0.3863	0.0408	1.4411	0.0225	13.1794	0.3368
E87	-0.3541	0.0359	1.3447	0.0167	13.0655	0.2409
b	-1.4021	0.0377	1.2219	0.0262	5.2561	0.2226
K88	-0.3619	0.0415	1.3338	0.0176	13.3240	0.2631
V89	-0.3180	0.0509	1.4034	0.0217	12.7046	0.2845
b	-2.4659	0.0970	0.9596	0.0231	3.5934	0.1916
K90	-0.5233	0.0219	1.5993	0.0105	8.9800	0.1014
b	-2.8124	0.0935	0.8765	0.0112	1.1210	0.0824

Appendix D

Submitted manuscript

An alternate conformation of the hyperthermostable HU protein from *Thermotoga maritima* has unexpectedly high flexibility

Michael A. Durney[a,b], Rainer W. Wechselberger[a], Charalampos G. Kalodimos[a,1], Robert Kaptein[a], Constantinos E. Vorgias[c], and Rolf Boelens[a,*]

[a] Bijvoet Center for Biomolecular Research, Universiteit Utrecht, Padualaan 8, 3584 CH Utrecht, the Netherlands.
[b] Department of Chemistry, University of Southampton, Southampton SO17 1BJ, UK.
[c] Department of Biochemistry and Molecular Biology, National and Kapodistrian University of Athens, 15784 Athens, Greece.
* Corresponding author. Fax: 31-30-2537623. e-mail: boelens@nmr.chem.uu.nl
[1] Present address: Department of Chemistry, Rutgers University, Newark, NJ 07102-1811, USA.

Keywords: conformational flexibility; protein dynamics; thermostability; nuclear magnetic resonance spin relaxation

Abbreviations: HU*Tmar*, *Thermotoga maritima* HU protein; HU*Bst*, *Bacillus stearothermophilus* HU protein; HU*Bsu*, *Bacillus subtilis* HU protein; IHF, integration host factor; NOE, nuclear Overhauser enhancement; $R_{1,2}$, longitudinal and transverse relaxation rates; HSQC, heteronuclear single quantum coherence; NOESY, nuclear Overhauser enhancement spectroscopy; ESI-MS, electrospray ionization mass spectrometry

Abstract
The homodimeric HU protein from the hyperthermophile *Thermotoga maritima* is a model system which can yield insights into the molecular determinants of thermostability in proteins. Unusually for a thermostable protein, HU*Tmar* exists in a structurally heterogeneous state as evidenced by the assignment of two distinct and approximately equally populated forms in solution. Relaxation measurements combined with chemical shift, hydrogen exchange, and NOE data confirm the main structural features of both forms. In addition, these data support a two-state model for HU*Tmar* in which the major form closely resembles the X-ray structure while the very flexible minor form is less structured. HU*Tmar* may therefore be a new example of the small class of hyperthermostable proteins with unexpected flexibility.

D.1 Introduction

Proteins from thermophilic organisms display properties that reflect the environment in which they have evolved. The general consensus reached from biophysical studies of many different systems is that thermal stability results from conformational rigidity due to the increased packing efficiency in the structures of thermophilic proteins [154]. Recent data from studies of selected thermostable systems with unexpected flexibility challenge this paradigm however [155].

The HU-IHF family of dimeric histone-like proteins have various roles in DNA remodelling, bending, compaction, and negative supercoiling of the bacterial nucleoid [143, 146, 180]. Members of this family include HU, IHF, H-NS, TF1, and Fis, and can bind DNA specifically or non-specifically. The intrinsic dynamics of HU proteins is highlighted by X-ray structures of HU*Bst* and HU*Tmar* (refined to resolutions of 2.1Å and 1.53Å respectively) in which no electron density was observed for the DNA-binding β-arms [151, 153]. The NMR structure of HU*Bst* provided the first details on the structure and dynamics of the highly mobile yet internally structured β-arms in the free protein [152, 182]. The β-arms stabilize and wrap around the DNA upon binding as clearly shown by the crystal structures of the DNA complexes of both IHF and HU [147, 148, 181] These studies present a dramatic bending of the DNA mediated by proline intercalation by residues in the β-arms which in turn become much less mobile. The dynamics of the β-arms therefore play an essential role in mediating DNA binding.

The HU*Tmar* protein (T_m = 80.5°C) is a model thermostable system with unusually high affinity and temperature independent DNA binding properties [157]. The extreme stability and versatility of HU*Tmar* ensures protection of tightly-bound DNA substrates which has allowed the development of biotechnological applications, for instance, as a highly effective gene delivery agent [158].

The results from an NMR analysis of the behaviour of HU*Tmar* in solution are complementary to those from a recent site-directed mutagenesis study characterizing destabilizing mutations [156]. Our results characterize an anomalous dynamic property of HU proteins. We noted that wild-type and mutants of HU*Tmar* are unexpectedly heterogeneous in solution. The data also support a two-state model in which the major form resembles the X-ray structure while the flexible minor form seems to result from a structural rearrangement of the dimer architecture. The use of the major/minor terminology therefore refers to structural integrity rather than populations of the two conformations. The observed dynamics of HU*Tmar* in solution are unforeseen when compared with other thermostable proteins.

D.2 Materials and methods

D.2.1 Sample preparation

Cloning of the HU*Tmar* gene, overexpression, and purification of wild-type and mutant proteins were performed as described [159]. Fully labelled protein was produced using ^{15}N-NH_4Cl and ^{13}C-glucose in minimal medium. Protein samples (1 mM) were prepared in a 50 mM phosphate buffer with 200 mM KCl at pH 5.8. Addition of the protein to the buffer solution changes the pH to approximately 6.2 since the protein is positively charged. 7.5% D_2O was added for NMR measurements.

D.2.2 NMR spectroscopy

NMR spectra were recorded on Bruker Avance spectrometers operating at proton frequencies of 500, 600, 700, and 750 MHz at a temperature of 311 K. The backbone and sidechain resonances were assigned using standard triple resonance techniques [29]. The ^{15}N heteronuclear NOE, R_1 and R_2 data sets were recorded using modified versions of the standard Bruker pulse sequences. The NOE experiment was repeated three times, R_1 datasets were recorded with relaxation delays of: 0.1 (×2), 0.2, 0.3, 0.4 (×2), 0.5, 0.6, 0.8, 1.0 s; and R_2 datasets with delays of 0 (×2), 2, 16, 32, 48, 64 (×2), 80, 96, 128, and 160 ms. All NMR data were processed using NMRPipe [166] and analyzed using NMRView [167]. Relaxation data were fitted as described [184]. Hydrogen exchange protection factors were calculated as described [60].

D.3 Results

A sequence alignment of HU proteins and the secondary structure elements of the crystal structure of HU*Tmar* are presented in Figure D.1. Surprisingly, 140 peaks (excluding sidechains) were counted in the HSQC spectrum where 84 were expected (Figure D.2). The extra peaks could be assigned to residues throughout the protein and two slowly exchanging and approximately equally populated forms of the protein exist under the experimental conditions. This equilibrium seems to persist indefinitely and no time-dependent changes in the intensities of peaks from either form were observed. All residues in the primary sequence were assigned in the major form while 75 residues were assigned in the minor form. The remaining residues in the minor form are degenerate with those in the major form. The temperature, pH, salt concentration, and protein concentration were all varied with no significant change in the spectra and no degradation was observed on SDS gels. Urea-triton gels, which are very sensitive and can identify point mutations, show a single homogeneous band which indicates that HU*Tmar* is chemically homogeneous (data not shown). One fraction of the protein sample was thermally denatured and refolded into the same conformational equilibrium. Major chemical shift differences indicating conformational differences between

the two forms are found in the second half of the sequence and notably in the short C-terminal α-helix (Figure D.3). The HSQC spectra of two destabilizing point mutants (E34D and V42I) are almost identical to that of the wild-type protein, showing the same doubled set of NMR signals. The heterogeneity of HU*Tmar* in solution is therefore an intrinsic property of the structure. Analysis of the ^{13}C chemical shift data [179] rule out *cis-trans* proline isomerization as an explanation for the two forms.
Considering first the major form, the secondary structure determined by the $^{13}C^{\alpha}$ secondary chemical shifts is consistent with that observed in the NMR and X-ray structures of HU*Bst* and HU*Tmar* (Figure D.4). The general pattern observed in the relaxation data (Figure D.5) is that the residues of the β-arms are more flexible than the core of the protein. The more negative values for the ^{15}N heteronuclear NOE and the lower values for transversal relaxation rates compared to the core of the protein indicate enhanced mobility. In addition the values of the NOE and R_2 for residues N62-I69 indicate a maximum degree of flexibility for the turns at the tips of the β-arms. The dynamics of the major form are therefore highly similar to those observed for HU*Bst* [182].
In contrast, the minor form shows a number of critical differences compared to the major form. The $^{13}C^{\alpha}$ shifts (Figure D.4) indicate a change in conformation for the C-terminal residues G82-K90. In addition, the minor form clearly has different relaxation rates for residues for which data is available (Figure D.6). In particular, the substantially more negative values of the heteronuclear NOE and the lower values of the R_2 rates for the β-arms and the C-terminus indicate even higher mobility compared to the major form. In support of this the NOESY spectra show no crosspeaks for the majority of residues of the β-arms and the C-terminus in the minor form. This is in contrast with the major form in which the characteristic NOEs defining the structure of the β-arms and the C-terminal α-helix were assigned in the same way as for HU*Bst* [177]. Protection factors calculated from amide proton exchange also indicate a less stable α2 in the minor form (Figure D.7).
The R_2/R_1 ratio reflects the orientation of secondary structure elements (or complete molecules) with respect to the overall rotational symmetry axis. The data clearly indicate differences between the two forms (Figure D.6). The major form of HU*Tmar* and HU*Bst* have similar values of the R_2/R_1 ratios throughout the sequence. This indicates that the major form has a diffusional anisotropy which is comparable to that of HU*Bst*. Furthermore, the R_2/R_1 values for α1 are lower on average than those for α2. In the minor form the opposite effect is seen since the R_2/R_1 values for α1 are higher than those for α2. While local increased R_2 rates (and consequently R_2/R_1 ratios) can be due to conformational exchange, decreased R_2 rates can only be due to increased mobility. Since the average magnitudes of the deviations are similar for both the increased and the decreased rates and uniform in the α1 and α2 helices conformational exchange is not likely. The R_2/R_1 values show rather that the relative positions of the two helices with respect to the diffusional symmetry axis are different for the major and minor forms.

D.4 Discussion

There are several sources of data which illustrate the different conformational properties of the two forms of HU*Tmar* in solution. The two most likely explanations for this conformational heterogeneity are: (1) monomer-dimer equilibrium; and (2) different dimeric structures due to dynamic effects possibly leading to a structural rearrangement. We propose that the latter is the case for HU*Tmar* based on our analysis of the available data. Furthermore, based on the chemical shift, NOESY, and relaxation data we conclude that the major form bears close resemblance to both the X-ray and NMR structures of HU*Tmar* and HU*Bst* respectively.

The monomeric state of HU*Tmar* is likely to exist only very transiently in solution due to the high stability of the dimer [193] and therefore cannot account for the observed differences in the relaxation data. This is supported by the observation that changes in the protein and salt concentrations do not significantly change the spectra. In addition the average R_2/R_1 values for the core of the protein are very similar for the major and minor forms which would not be expected for a monomeric minor form. A comparison of the R_2/R_1 values for the α1 and α2 helices however indicates a difference in the relative positions of the two helices with respect to the molecular diffusion axis for the two forms of HU*Tmar*. Whether this is due to a molecular rearrangement of the helices or due to a reorientation of the major diffusion axis cannot be concluded at this stage. This would require a high-resolution structural analysis of the minor form.

The high stability of the HU*Tmar* dimer results from intermonomer contacts in the compact core of the protein. The flexible β-arms extend from the core and do not participate in formation of the dimer but are instead available to wrap around a DNA substrate [152]. A recent Raman spectroscopy deuterium exchange study of the solution conformation of HU*Bst* found unexpectedly high exchange rates for helical elements which the authors attributed to flexibility at the dimeric interface [194]. The backbone relaxation rates measured for both forms of HU*Tmar* also indicate unusual flexibility at both the homotypic and heterotypic dimeric interfaces formed by the two pairs of α1 and α2 helices. Flexibility at the dimeric interface could account for the critical role of residue G15 in the turn between the α1 and α2 helices since mutation of this residue is known to dramatically reduce the thermostability of HU proteins [156, 192].

The biophysical properties of HU proteins from various sources have been documented in detail [151]. In marked contrast to HU*Tmar*, protein concentration and ionic strength have been shown to affect both the structure and stability of HU*Bsu* in solution as measured by CD [195]. These observations were extended with ESI-MS measurements of the influence of the ionic strength on the monomer-dimer equilibrium in HU*Bsu* [196]. The NMR data presented here expand on the earlier biophysical results for HU*Tmar* [193] since information is now available for individual amino acids and the existence of dual conformations is surprising.

Recent research has raised interesting questions about correlating thermostability with rigidity or flexibility in thermostable proteins [154, 187, 188]. For instance, the most thermostable protein known, rubredoxin from *Pyrococcus furiosus*, with a melting

temperature close to 200°C has been shown to have similar flexibility to the mesophilic rubredoxin from *Clostridium pasteurianum* [189].

D.5 Conclusion

We have demonstrated that HU*Tmar* is unexpectedly heterogeneous and dynamic in solution on a broad timescale range. The major form in solution has features consistent with those determined in the X-ray structure as evidenced by chemical shift, NOE, and ^{15}N relaxation data. The presence of the two forms in both wild-type and several mutants of HU*Tmar* suggests that the equilibrium is an intrinsic property of the native structure. The dynamical behaviour of HU*Tmar* suggests that this protein may be a new example of the small class of extremely thermostable proteins which display unexpected structural heterogeneity in solution. To the best of our knowledge HU*Tmar* is the first example of a homodimeric thermostable protein displaying these properties.

Acknowledgements

The NMR spectra were recorded at the SON NMR Large Scale Facility in Utrecht, which is funded by the *Access to Research Infrastructures* program of the European Union (HPRI-CT-2001-00172). We thank Klaartje Houben for setting up the relaxation experiments and analysis scripts and Hans Wienk for reading the manuscript. This work was supported by a Marie Curie training site fellowship (HPMT-2000-00045) awarded to MAD.

```
Tmar -----MTKKELIDRVAKKAGAKKKDVKLILDTILETITEALAKGEKVQIVGFGSF 50
Bst  -----MNKTELINAVAETSGLSKKDATKAVDAVFDSITEALRKGDKVQLIGFGNF 50
Bsu  -----MNKTELINAVAEASELSKKDATKAVDSVFDTILDALKNGDKIQLIGFGNF 50
Tth  AAKKTVTKADLVDQVAQATGLKLLDVKAMVDALLAKVEEALANGSKVQLTGFGTF 55
Tvo  ----MVGISELSKDVAKKANTTQKVARTVIKSFLDEIVAQANGGQKINLAGFGIF 51
          : : :*:  **   :  :::     :     ::::: :* *::  *** *

Tmar EVRKAAARKGVNPQTRKPITIPERKVPKFKPGKALKEKVK-- 90
Bst  EVRERAARKGRNPQTGEEMEIPASKVPAFKPGKALKDAVK-- 90
Bsu  EVRERSARKGRNPQTGEEIEIPASKVPAFKPGKALKDAVAGK 92
Tth  EVRKRKARTGVKPGTKEKIKIPATQYPAFKPGKALKDKVK-- 95
Tvo  ERRTQGPRKARNPQTKKVIEVPSKKKFVFRASSKIKYQQ--- 90
     *:*  ::*:  :*:*  :*  ::: *:::::*  ::
```

Figure D.1: ClustalW alignment of HU sequences from *Thermotoga maritima* (Tmar), *Bacillus stearothermophilus* (Bst), *Bacillus subtilis* (Bsu), *Thermus thermophillus* (Tth) and *Thermoplasma volcanium* (Tvo). Asterisks indicate residues conserved in all five sequences and colons indicate residues additionally conserved between HU*Tmar* and HU*Bst*.

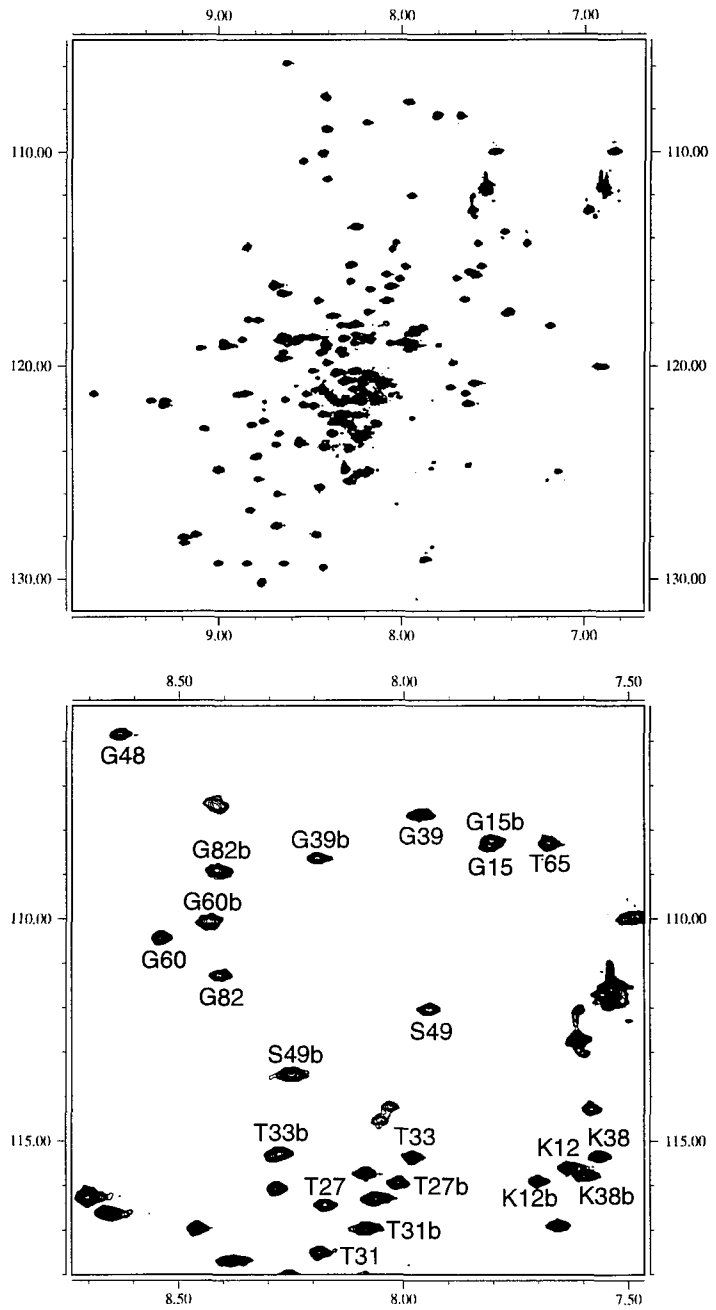

Figure D.2: 2D ^{15}N HSQC spectrum (above) and excerpt (below) of uniformly ^{15}N-labelled HU*Tmar* recorded at 750 MHz proton frequency and 311 K. Selected resonances are marked with their assignments where b indicates the minor form.

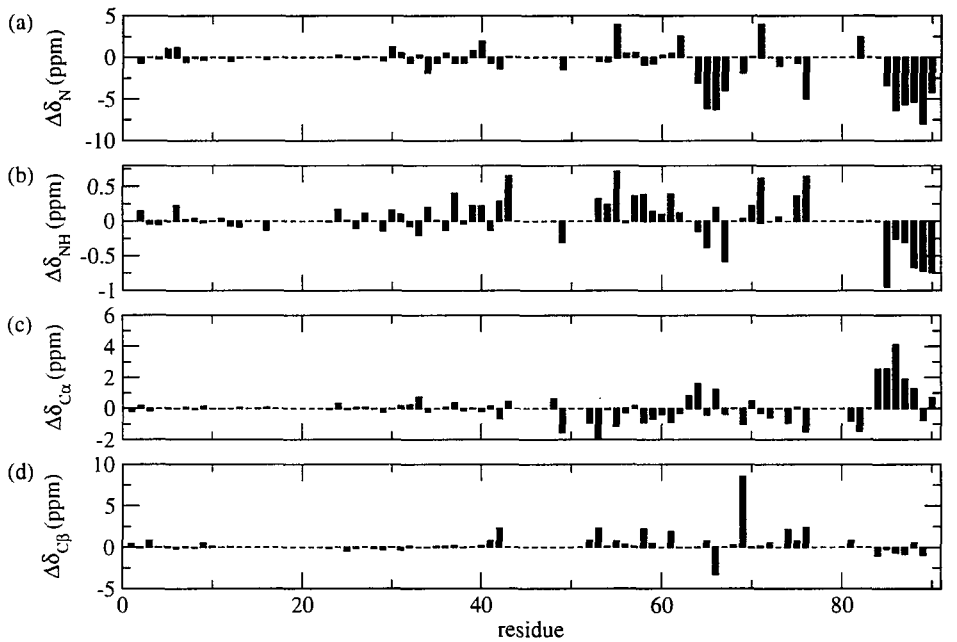

Figure D.3: Chemical shift mapping, $\Delta\delta$ = major − minor (ppm), to identify conformational differences between the major and minor forms of HU*Tmar*.

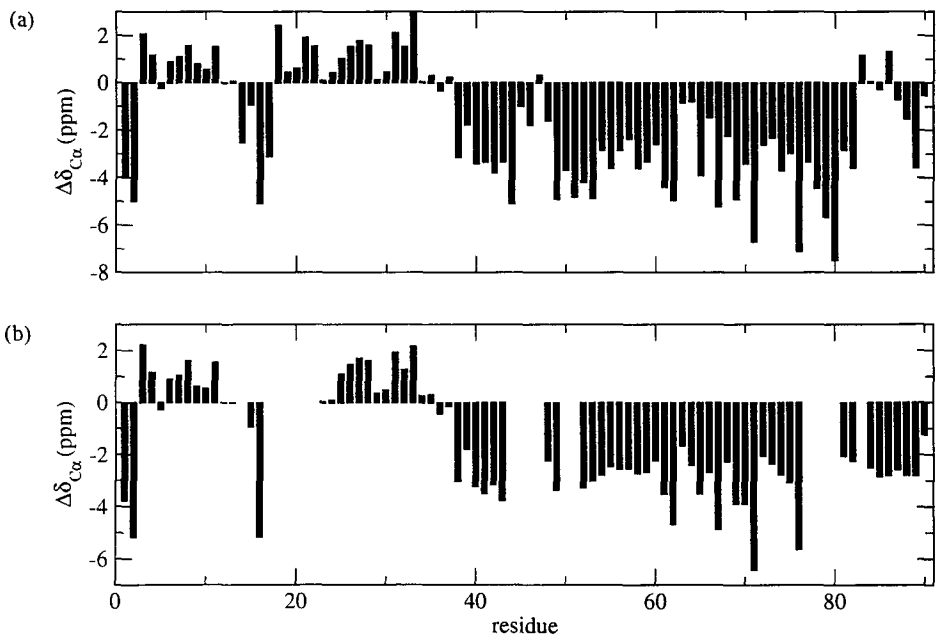

Figure D.4: Backbone $^{13}C^{\alpha}$ secondary chemical shifts for the (a) major and (b) minor forms of HU*Tmar*. Positive and negative shifts are indicative of α-helical and β-strand conformations respectively.

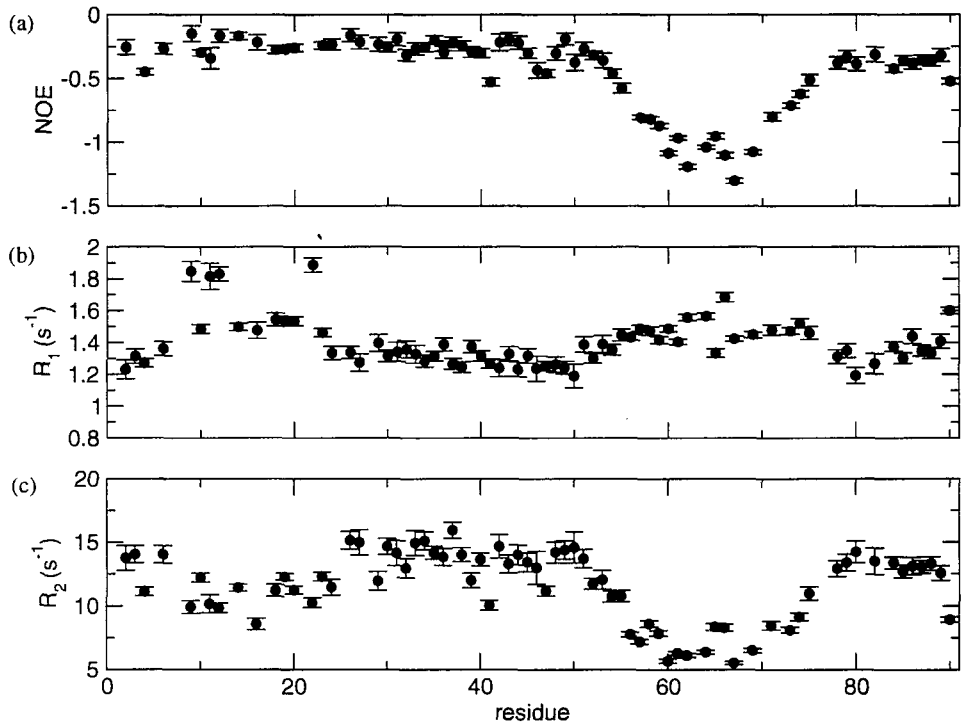

Figure D.5: Backbone ^{15}N (a) NOE, (b) R_1, and (c) R_2 data for the major form of *HUTmar*. Lower values of NOE and R_2 indicate increased flexibility on a ps-ns timescale.

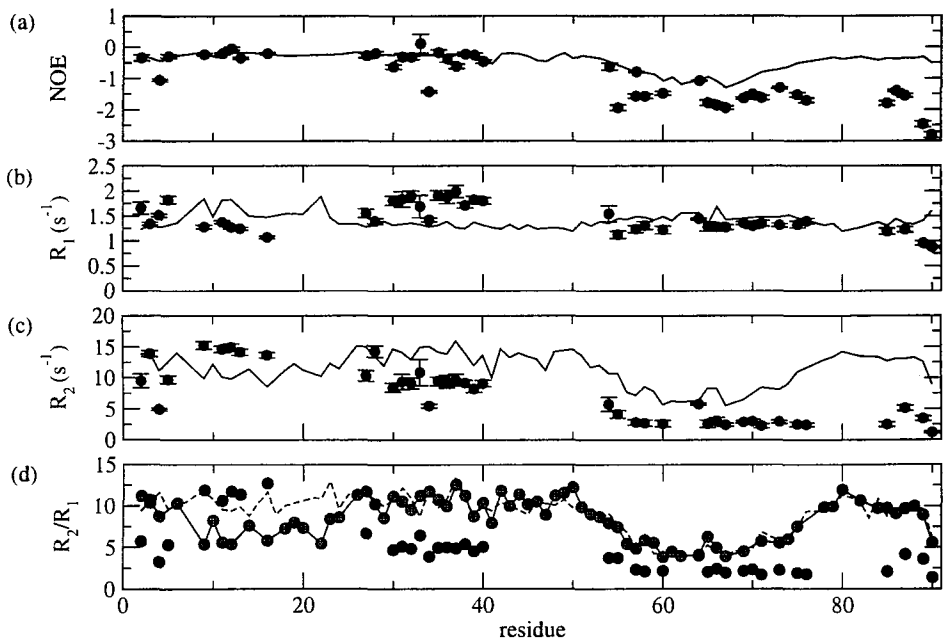

Figure D.6: Comparison of (a) NOE, (b) R_1, and, (c) R_2 values for the major (line) and minor (circles) forms of HU*Tmar*; (d) comparison of R_2/R_1 ratios for the major (shaded circles) and minor (filled circles) forms of HU*Tmar* with the values for HU*Bst* (dashed line) for reference.

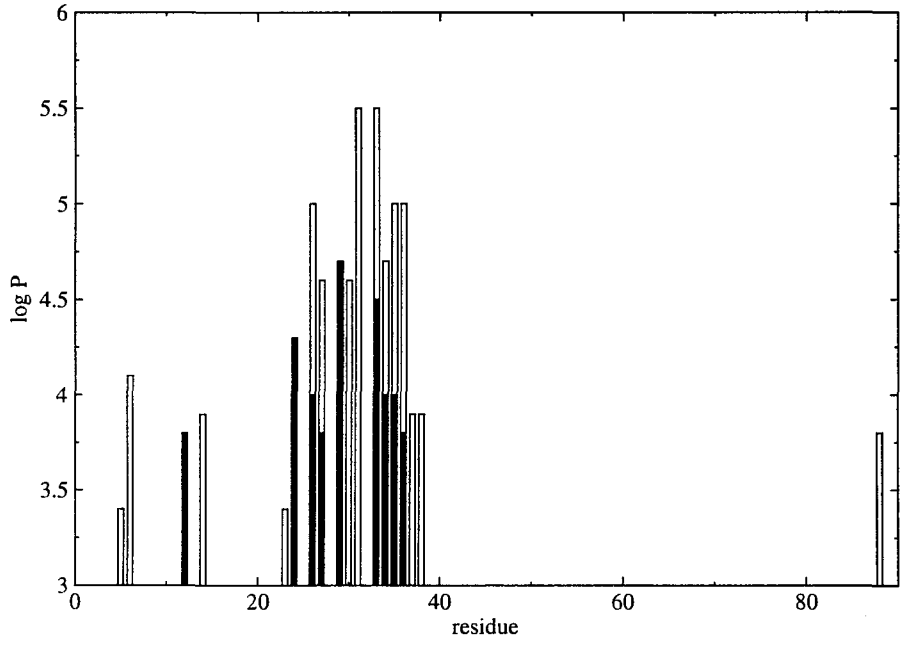

Figure D.7: Protection factors (P) calculated for the major (white bars) and minor (black bars) forms of HU*Tmar*. Higher values indicate backbone amide protons which are less solvent accessible due to either structural location or conformational fluctuations.