Journal of Biomolecular NMR, **21:** 235–248, 2001.
KLUWER/ESCOM

Joint refinement as a tool for thorough comparison between NMR and X-ray data and structures of HU protein

Mia L. Raves[a], Jurgen F. Doreleijers[a,†], Hans Vis[**], Constantin E. Vorgias[c], Keith S. Wilson[d] & Robert Kaptein[a,*]

[a]*Department of NMR Spectroscopy, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands;* [b]*Insituto de Tecnologia Quimíca e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal;* [c]*Department of Biochemistry & Molecular Biology, National and Kapodistrian University of Athens, Greece;* [d]*Structural Biology Laboratory, Department of Chemistry, University of York, U.K.;* [**]*Current address: BioMagResBank, University of Wisconsin, Madison WI, U.S.A.*

Received 7 August 2001; Accepted 21 September 2001

Key words: HU protein, NMR, structure validation, X-ray chrystallography

Abstract

Joint refinement, i.e., the simultaneous refinement of a structure against both nuclear magnetic resonance (NMR) spectroscopic and X-ray crystallographic data, was performed on the HU protein from *Bacillus stearothermophilus* (HUBst). The procedure was aimed at investigating the compatibility of the two data sets and at identifying conflicting information. Wherever important differences were found, such as peptide flips in the main-chain conformation, the data were further analyzed to find the cause. The NMR data showed some errors arising either from the manual interpretation of the spectra or from the incorrect account for spin diffusion. The most important artefact inherent to the X-ray data is the crystal packing of the molecules: the effects range from the limitation of the freedom of the flexible parts of the HUBst molecule to possibly one of the peptide flips.

Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy and X-ray crystallography are both techniques that are currently widely used nowadays for determination of the molecular structure of proteins and nucleic acids. Both methods utilize experimentally obtained structural data in combination with a force field containing information on the ideal geometry of macromolecules in order to derive the best model. Opposing these global similarities, there are some important differences between the two methods. First and foremost is the experimental environment of the macromolecule: in NMR the molecules move freely in solution, in X-ray crystallography they are packed with multiple copies of themselves onto a crystalline lattice. This packing has important consequences for flexible regions of the molecule, which will either be restricted in motion by contacts with symmetry-related molecules or be 'invisible' if they are still disordered in the crystal. The freedom of motion in an NMR experiment makes this technique much more suitable for studying the dynamic behavior of macromolecules. A second difference is the information content of the experimental data: in X-ray crystallography the observed reflection amplitudes predominantly reflect the positions of the 'heavy' atoms (carbon and higher atomic weights), whereas in NMR the bulk of the structural data is derived from the relative positions of hydrogen atoms in the molecule, observed with the Nuclear Overhauser Effect (NOE). In general, X-ray reflections in a data set for a given macromolecule are more numerous than NMR constraints, which makes the problem in the latter method underdetermined. Thirdly, an X-ray structure is refined in an iterative process of calculations, in which one attempts to min-

*To whom correspondence should be addressed. E-mail: kaptein@nmr.chem.uu.nl

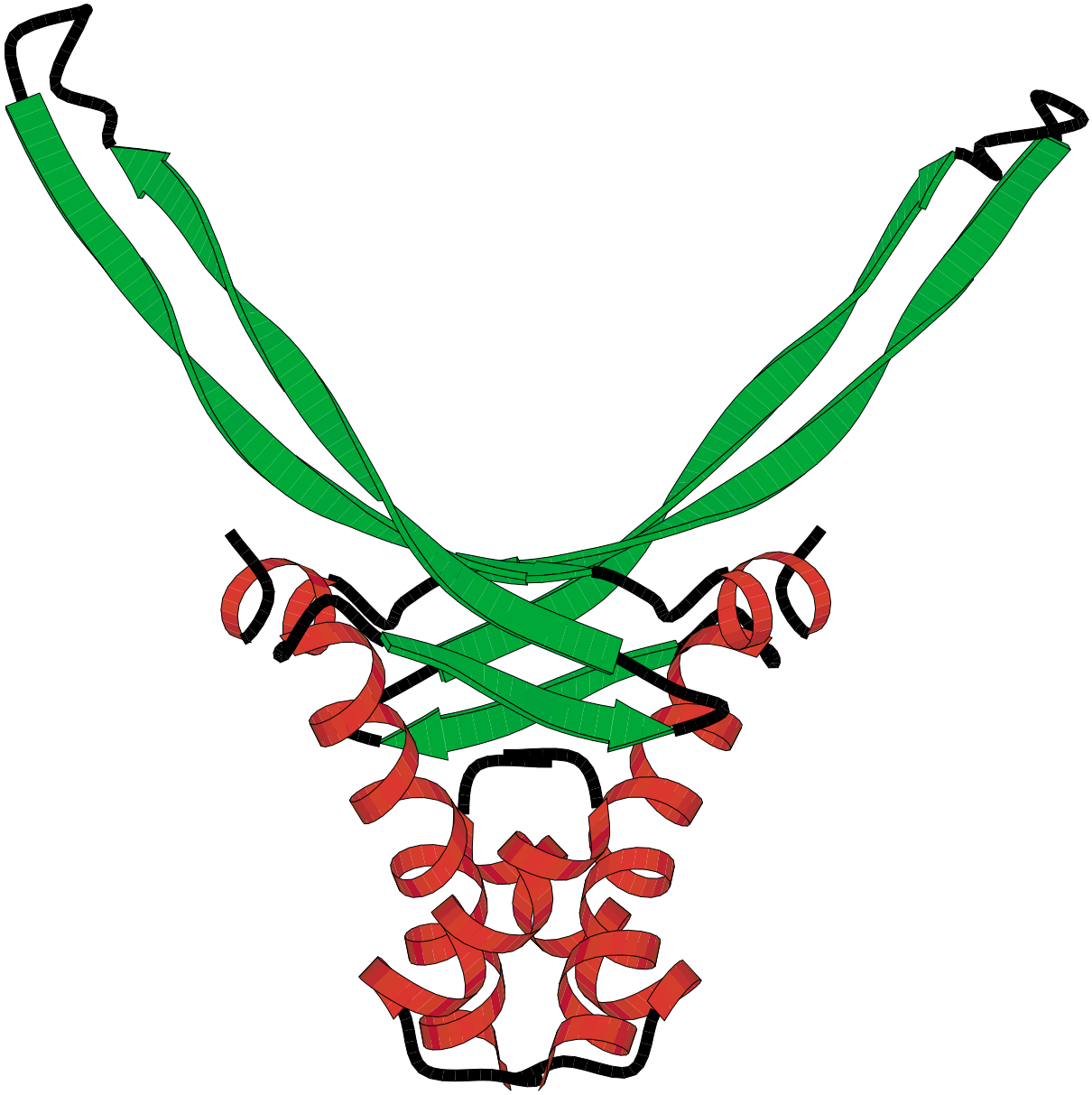

Figure 1. Secondary structure of a HUBst dimer from the NMR structure; α-helices are depicted as red ribbons, β-strands as green arrows.

imize the difference between the measured structure factors (F^{obs}) and those calculated from the model (F^{calc}), alternated with manual intervention in order to globally position residues into areas of electron density and to locate water molecules. In contrast, NMR structure refinement is accomplished on the basis of the resonance assignment of the spectra, followed by integration of the NOE cross peaks and their conversion into interproton distance restraints; problematic regions in the models are identified, whereupon individual interpretations of data are revised or more data are extracted from the spectra, and the structure is recalculated with the new set of experimental data. Water molecules are usually not included in NMR structures, as their determination requires very specific experimental procedures (for instance, see Karimi-Nejad et al., 1999).

A few compelling questions arise at this point: Do NMR and X-ray data provide complementary or contradictory information? What are the strengths and

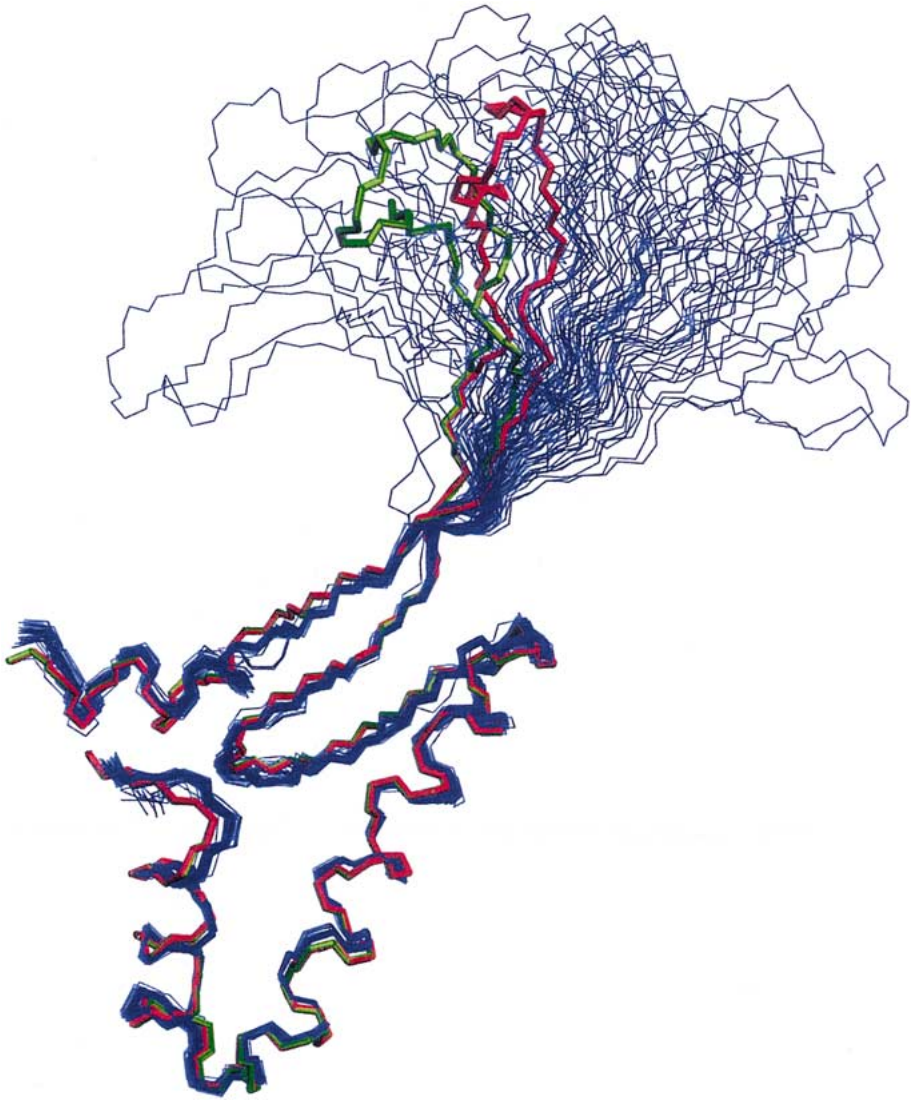

Figure 2. Superposition of 50 NMR monomers of PDB entry 1HUE (blue) and 2 independent X-ray monomers (green and red).

weaknesses of the two techniques in terms of the quality of the resulting structures? We have studied the HU protein in order to address these questions, using joint refinement of a model against both X-ray and NMR data simultaneously.

Joint refinement

The first test case for joint refinement was interleukin-1β (Shaanan et al., 1992). The authors of that study constructed a model with a crystallographic R-factor and geometric quality comparable to those derived from refinement against the X-ray data alone. The few residual NOE violations observed in the combined model could then provide a reliable measure of genuine differences between the solution and crystal structures. A few years later, a similar procedure was applied to BPTI (Schiffer et al., 1994), for which one model was found which could accommodate both data sets, and ribosomal protein L9 (Hoffman et al., 1996) in which only the NMR restraints for the 40 disordered N-terminal residues were used. For the oligomerization domain of p53 (Miller et al., 1996), the NMR data were found not to interfere with the X-ray data, but to aid in getting a model with improved geometry. In each of the four examples, the authors found only a few conflicting differences which could be explained from close examination of the experimental data. Most

of the discrepancies concerned side chains of residues on the surface of the proteins, as expected.

Protein HU

The eubacterial protein HU is a highly abundant DNA-binding protein with a function similar to that of eukaryotic histones, viz., the bending of long strands of DNA in order to facilitate the formation of higher-order nucleoprotein complexes (Kobryn et al., 1999). The native dimeric protein consists of a core formed by a three-stranded antiparallel β-sheet and three α-helices in each of the monomers, and a β-stranded arm (see Figure 1). Two such arms in the dimeric molecule are thought to be able to wrap around the DNA double-helix, similar to DNA binding by the homologous IHF (Rice et al., 1996) and TF1 (Silva et al., 1997).

The solution structure of HU from *Bacillus stearothermophilus* (HUBst) was solved in 1995 (Vis et al., 1995a,b). The 50 models for the monomer in the NMR structure (see Figure 2) clearly show the rigidity of the core region, as well as the high flexibility of the arms, which can adopt a wide variety of conformations, all of which are consistent with the NMR data. The flexibility of the arms, both in the presence and in the absence of DNA (Vis et al., 1996), has been thoroughly investigated using heteronuclear relaxation experiments. In addition, an analysis of the per-residue completeness of the NOE restraints, i.e., the ratio of the number of observed NOEs and those expected to be observable based on the structure (Doreleijers et al., 1999), shows that the flexibility is not just an artefact due to the underdetermination of this region of the structure; on the contrary, the completeness per residue is higher on average for the arm than for the core residues.

The X-ray structure of HUBst was solved in 1984 (Tanaka et al., 1984) and rerefined to higher resolution (2.0 Å) in 1999 (White et al., 1999). Of the three independent monomers in the asymmetric unit, only the core is well-defined; most of the residues in the arm regions are disordered, in agreement with the inherent structural flexibility of the arms observed in the NMR structure determination. The dimers are formed by crystallographic two-fold symmetry. A more recent X-ray structure determination of HUBst in a different space group (Dauter and Wilson, manuscript in preparation) shows the entire structure including the complete arms (see Figure 2): Only two residues on the tip of the arm are partially disordered in both independent monomers in the asymmetric unit. The flexibility of the arms has, in this crystal form, been restricted by crystal contacts.

Validation

Validation tools which are currently used as a standard way to check the quality of the generated models can only provide insight into the precision in terms of the correctness of the geometry of the model and of its fit to the experimental data. They may reveal less of the relevance of the model, i.e., whether it represents the conformation of the macromolecules *in vivo*. Even the comparison of models determined with different techniques such as NMR and X-ray crystallography is complicated by differences in experimental conditions (pH, additives, temperature, etc.) and the software packages and force fields used for structure determination. A careful comparison of the information in the different datasets, using the same software and force field, may point out genuine differences under the different experimental conditions, as well as artefacts caused for example by crystal contacts in X-ray crystallography or spin diffusion for NMR. The former is caused by inter-molecular contacts in the crystal which would not occur if the molecules were free in solution. The latter lies in the fact that transfer of magnetization between two protons, as observed in an NOE experiment, can be influenced by a third proton close to the first two, thus altering the intensity of the corresponding peak in the spectrum, which is directly translated into an incorrect distance restraint between the two protons.

Using the experimental NMR data (consisting mostly of NOE restraints) and X-ray reflections obtained for HuBst, we initiated a joint-refinement project. In order to start with an unbiased model, the procedure followed consists of:

(1) starting with two randomly structured, elongated polypeptides;
(2) folding them into the approximate dimeric shape using only the NOE restraints;
(3) placing the dimer into the crystallographic unit cell using molecular replacement;
(4) refinement against both data sets simultaneously.

The goal of the joint-refinement procedure was to determine to what extent the NMR and X-ray data and structures are in agreement, where the major discrepancies are located, and whether these can be explained. Indeed, there are interesting differences in the main-chain conformation, which have direct consequences

Table 1. Experimental NMR data* on HUBst

Total number of NOE restraints	1245
Intra-monomer	1162
Main-chain - main-chain	393
Main-chain - side-chain	558
Side-chain - side-chain	211
Intra-residual	173
Sequential	420
Medium range[b]	316
Long range[c]	336
Inter-monomer	83
Number of hydrogen bond restraints	25
Number of dihedral angle restraints	80
φ angles	41
χ^1 angles	39

*Number of restraints per monomer.
[b] $1 < |i - j| \leqslant 4$.
[c] $|i - j| > 4$.

Table 2. Quality of X-ray data on HUBst

Space group	C2
Unit cell dimensions	$a = 100.5$, $b = 37.3$, $c = 73.5$ Å
	$\beta = 134.5°$
Number of unique reflections	18 200
Resolution	25–1.8 Å
Completeness	99.8%
R_{sym}	5.9%

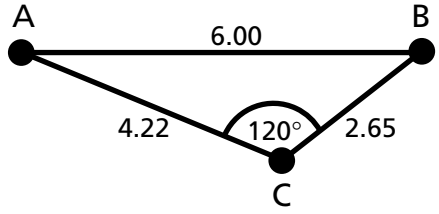

Figure 3. Schematic illustration of spin diffusion, showing exemplifying proton-proton distances. Transfer of magnetization between two protons which are 6 Å apart, normally too far to give an NOE signal, can occur *via* a third proton along shorter distances. The angle of 120° was used as cut-off.

for the definition of the secondary-structure elements, as well as in the dimer interface.

Methods

The NMR structure and experimental data were obtained from the Protein Data Bank (PDB) (Sussman et al., 1998), entry code 1HUE. The ensemble consists of 25 models for the dimer, or 50 independent models for the monomer since no internal symmetry was imposed during structure determination (Vis et al., 1995a). The experimental data consists of 2490 NOE distance restraints, 1245 per monomer (see Table 1). The constraints are identical for the two monomers and were converted to X-plor format using the AQUA software package (Laskowski et al., 1996).

The 1.8 Å X-ray structure of the complete HUBst was kindly provided to us by Dr. Z. Dauter prior to publication, along with the crystallographic reflections. Details on the quality of the data are given in Table 2. The asymmetric unit contains two independent monomers, each of which forms a dimer across the crystallographic two-fold axis.

Tests using different sets of NMR restraints were conducted to see if the quality of the models, generated using only NMR data, could be improved. A procedure was used which is a combination of the simulated annealing and refinement scripts (X-plor tutorial files `sa.inp` and `refine.inp`). Without the use of an acceptance criterion as is usually applied in NMR structure determination (Nilges, 1996), 50 models of the dimer were generated. To remove low-quality structures from the set, the models were ranked according to their rms deviation of the NOE violations and the percentage of residues in the most favoured region of the Ramachandran plot. The best 20 structures, in terms of NOE violations and Ramachandran score, were selected for further statistical analysis.

The tests included:

– omission of individual NOEs, selected on the basis of severe violations when compared to the X-ray models: omitting the restraints between 80 H^{α} and 83 H^{N} or the one between 2 H^{N} and 143 H^{α} (inter-monomer restraint) yielded a significant improvement in the Ramachandran score for residues 2, 3 and 80;
– correction of the hydrogen-bond restraint between 82 O and 84 H^{N} to 82 O and 86 H^{N}: this greatly improved the φ/ψ angles for residue 83, going from a majority of conformations (71%) in the disallowed region to a majority (92%) in the favourable areas of the Ramachandran plot;
– addition of H-bond restraints for atoms for which there is slow exchange data (Vis et al., 1995a): the list of 25 H-bond restraints (part of PDB entry 1HUE) was found to be incomplete when compared to the secondary-structure elements. For example, in the first α-helix restraints are listed

between 4 O and 8 H^N and between 6 O and 10 H^N, but not between 5 O and 9 H^N. Inclusion of the 12 intermediate H-bond restraints yielded only a slight overall improvement in the Ramachandran score, but the added restraints are consistent with the X-ray models.

– variations in the non-bonded repulsion in the X-plor protocol: Varying the scale factor with which the van der Waals radii of the atoms are multiplied during simulated annealing (Nilges et al., 1988) greatly affected both the quality of the Ramachandran plot (from 73% of residues in the most-favoured region for a (default) scale factor of 0.75 to 83% at 0.95) and the geometry (the length of the second α-helix as measured by the distance between the C^α atoms of residues 19 and 37 ranges from 25.3 Å for a scale factor of 0.75 to 26.9 at 0.95; cf. 26.6 Å in X-ray models).

The 50 models in the NMR ensemble (PDB entry 1HUE) were examined on a per-residue basis using PROCHECK-NMR (Laskowski et al., 1996). Residues with φ/ψ combinations in the disallowed or other less favoured regions of the Ramachandran plot were flagged, as well as residues with an unfavourable side-chain conformation. A new dimer model was then constructed from pieces of the models, pasted together by least-squares fitting of the fragments to a hybrid dimer model consisting of the two independent X-ray monomers. One cycle of simulated annealing in X-plor at $T = 400$ K against the NMR restraints ensured both a correct geometry for the 'consensus' model and an optimal fit to the NMR data.

Joint refinement

The joint-refinement calculations were all performed using X-plor (Brünger et al., 1987) version 3.854. The starting structure contained an elongated polypeptide chain with ideal geometry for each of the monomers. Using only the NMR restraints, the chains were folded to an initial dimeric model. In the NMR data set, the restraints between 80 H^α and 83 H^N and between 2 H^N and 143 H^α were omitted, the H-bond restraint between 82 O and 84 H^N was corrected and 12 intermediate H-bond restraints were added, as described above.

Molecular replacement using the core residues of this model (residues 4–52 and 77–88) was carried out with the program Amore (Navaza, 1994) using all data between 25 and 4.5 Å resolution. A molecular replacement search for the arm fragments was not successful, and their position was fixed with a least-squares fit to the X-ray models. The same was done for the N- and C-terminal residues.

In parallel, the 'consensus' NMR model, described above, was used as a starting model for joint refinement. Since the fragments from which this model was built up were least-squares fitted to the X-ray models, molecular replacement could be performed on the complete dimeric molecule.

The crystallographic data and the two molecular replacement solutions were transformed to space group P1 in order to be able to use the inter-monomer NOE restraints. The unit-cell dimensions were reduced to half the size of the original cell (new cell dimensions: $a = 37.3$, $b = 73.5$, $c = 53.6$ Å, $\alpha = 134.5$, $\beta = 69.6°$) so that the asymmetric unit in P1 contains two dimers, and strong non-crystallographic symmetry (NCS) restraints were used between identical monomers. Six percent of the reflections, in thin resolution shells, were excluded from the refinement and used for calculation of R_{free}.

During each step in joint refinement, the scale factors for the NMR data were kept high (with a scale factor of 150 both for NOEs and dihedrals) so as to avoid domination by the more numerous X-ray data. The scale factor for the X-ray reflections was taken as roughly one third of the weight suggested by standard X-plor protocols ($W_A = 200,000$). No σ cut-off was applied to the X-ray data and the full resolution range was used.

Manual intervention was performed using the graphics program O (Jones et al., 1991) version 6.2.1. The startup files for O were adjusted for displaying and manipulating hydrogen atoms. Several small PERL programs were written for conversion of hydrogen labels, and for displaying NOE restraints and violations in O. Geometric analyses were performed using PROCHECK (Laskowski et al., 1993), PROCHECK-NMR (Laskowski et al., 1996) and OOPS (Kleywegt et al., 1996).

The effect of spin diffusion on the NOE data was analyzed by calculating all interproton distances in a reference structure. All protons A and B which are closer than 6 Å and for which a third proton C could be found with an angle ACB greater than 120° were listed (see Figure 3). This list of proton pairs was then compared to the NOE restraint list, to identify NOEs which could possibly have been affected by spin diffusion. The upper-bound distances for these NOEs were lengthened by 2 Å. Both the dimer of the first X-ray model (generated by applying crystallographic

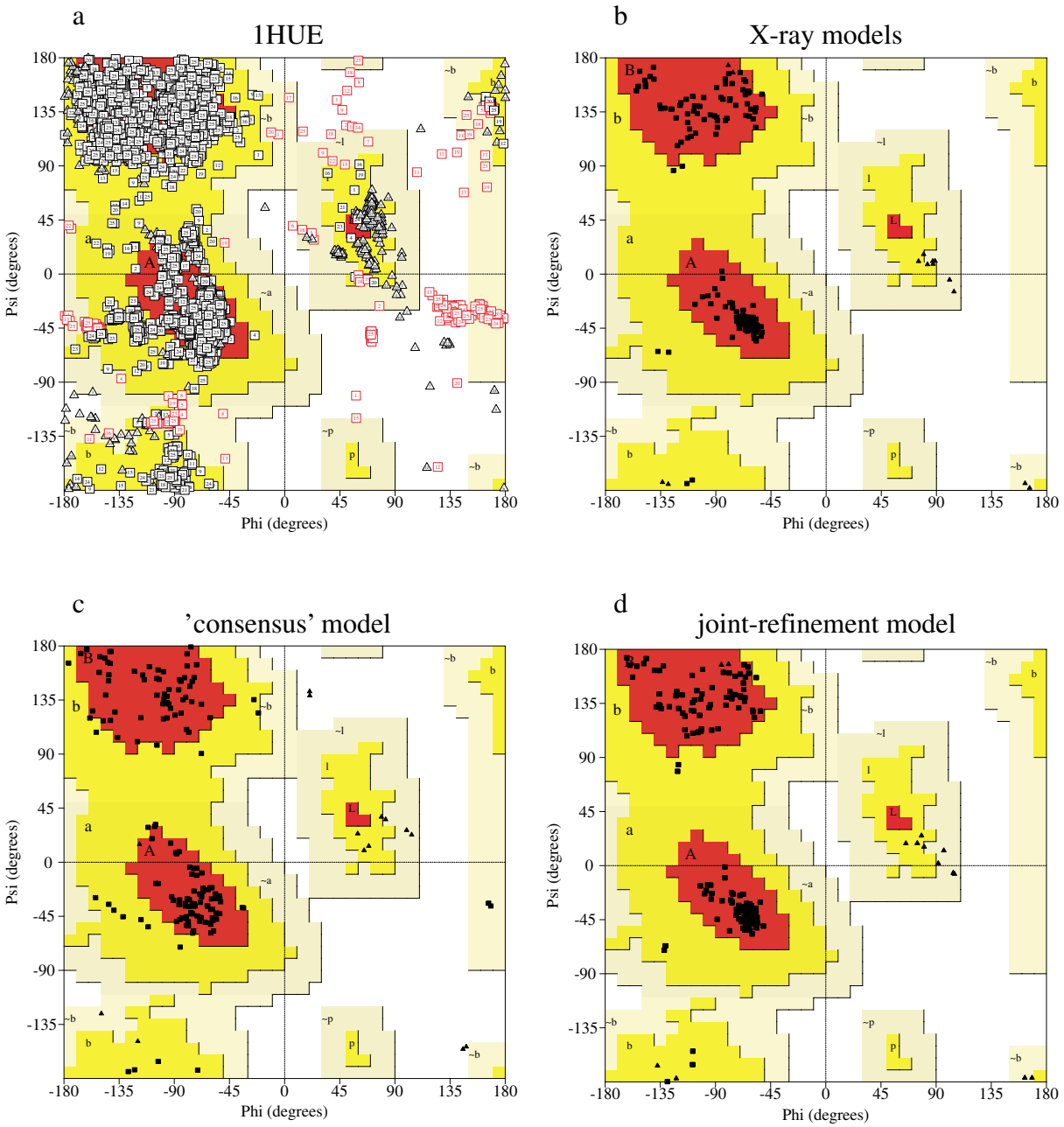

Figure 4. Ramachandran plot for HUBst: (a) 50 NMR models of PDB entry 1HUE; (b) 2 X-ray models; (c) 'consensus' model; (d) joint-refinement model.

twofold symmetry on the first monomer in the X-ray structure) and the joint-refinement model were used as reference models.

Results and discussion

A superposition on the C^{α} positions of the core residues (4–52 and 77–88) for the 50 NMR models for the monomeric HUBst obtained from PDB entry 1HUE clearly shows the agreement in the structure of the core region, as well as the high flexibility of

the arm of the molecule (Figure 2). The rms deviation from the average structure is 0.34 Å for the C^{α} atoms in the core region, 0.39 Å for the backbone atoms (N, C^{α}, C and O) and 0.88 Å for all non-hydrogen atoms in this region. Figure 2 also shows that the two independent monomers in the X-ray structure have conformations that could easily fit into this ensemble. The rms deviations for C^{α} atoms in the core residues between each of the 50 NMR models and the two X-ray models range from 0.79 to 1.12 Å. The arms in the crystal structure make contacts with symmetry-related molecules, thus limiting their conformational freedom to two distinct states. This limitation will inevitably be present in the joint-refinement model as well.The quality of the φ/ψ angle combinations in the Ramachandran plot is markedly lower for the ensemble of NMR models (on average 77.5% in most-favoured, 15.2% in additionally allowed, 4.4% in generously allowed and 2.9% in disallowed regions) than for the X-ray models (96.1% in most-favoured and 3.9% in additionally allowed regions). The Ramachandran plots are shown in Figures 4a and b, respectively. Similarly, an analysis of the side chains shows a preference for more favoured conformations in the X-ray models; in the NMR models, even eclipsed dihedral angles can be found. These facts are not surprising if one considers that the NMR data (and thus the models) represent a dynamic molecule in motion. While a protein, when immobilized into a certain conformation by the crystallization process, is likely to select an energetically favourable conformation, the molecule in solution (often at elevated temperatures during NMR data collection) can adopt many conformations at different energy levels. A model in which a reasonable geometry is combined with the NMR restraints was constructed by combining fragments of the 50 models in the ensemble. The resulting 'consensus' model has a greatly improved Ramachandran score (84.4% in most-favoured, 14.3% in additionally allowed and 1.3% in generously allowed regions, see Figure 4c) as well as a similar number and rms deviation of NOE violations as the individual NMR models from the ensemble. As an extra advantage, the 'consensus' model has a more correct geometry for protons in the protein backbone, since the BIOSYM software package which was used for the original structure determination allows much more freedom in the positions of these protons than X-plor, often resulting in distorted proton geometry.

Spin diffusion

A factor that influences the NMR data is the effect of spin diffusion in the NOE experiment. This phenomenon causes the transfer of magnetization observed between two protons to occur via a third proton that is in close proximity to the first two (see Figure 3). The observed intensity of the NOE cross peak in the spectrum, which is inversely proportional to the distance between the two protons, is thus increased, causing the observed distance to be shorter than it is in reality. This effect is stronger with longer mixing times during the recording of the NOE spectrum (Kalk and Berendsen, 1976) and in the case of HUBst [mixing time of 150 ms (Vis et al., 1995a)] it is certain to have played a role. Spin diffusion is not usually accounted for during structure calculations. Experimentally, if one collects NOE spectra at short mixing times in order to avoid spin diffusion, the signal-to-noise ratio is not favourable. Methods such as relaxation matrix refinement, in which the NOE spectrum is recalculated from the atomic coordinates, followed by refinement to minimize the difference between the theoretical and experimental spectra, have been shown to be very successful in a few cases [see for example (Nilges et al., 1991; Bonvin et al., 1994)], but the method is time-consuming and not applied often.

In another approach, one can determine where the effect may have occurred by calculating distances between proton pairs in a reference structure and looking for other protons in the vicinity which may aid in the transfer of magnetization. Using (one of) the NMR models for this kind of calculations is tricky, since the geometry of those models may already have been influenced locally by NOE restraints affected by spin diffusion. We have performed this analysis using one of the X-ray models or the joint-refinement model as a reference structure. Since the crystallographic structure does not necessarily represent the conformation of the molecule in solution, one cannot place too much importance on the exact results, but it can give an indication of problematic regions. The joint-refinement model, having been refined against the NMR data and 'guided' by the X-ray data, seems to be a more realistic model for this analysis. According to the joint-refinement model, one third of the NOE restraints have probably been affected by spin diffusion. Even more pronounced is the effect on the NOEs in the core of the molecule: at the interface between the two monomers, 62% of the restraints between side chain protons show a susceptibility to spin diffusion.

This is likely to be realistic, as the interface consists entirely of side chains of hydrophobic residues, rich in protons.

Molecular replacement

The apparent overall similarity between the NMR and the X-ray models seems to be in contrast with the difficulties we encountered when attempting molecular replacement in preparation for the joint-refinement process. An initial unbiased search model was obtained by folding two polypeptide chains in a random conformation into the familiar dimeric structure of HUBst using only the NMR restraints. However closely this search model resembles the X-ray structures, the molecular replacement search is impeded by the absence of information on (relative) temperature factors. As an illustration to this problem, the final X-ray structure with fully refined individual B-factors (without alternative conformations for the side chains) has an R-factor of 18.3%. Giving all atoms a B-factor of 15 $Å^2$ and setting all occupancies to one increases the R-factor to 30.9%. Furthermore, the arms in the search model are probably not in the relevant conformation, and are thus more likely to hamper the molecular replacement search than to aid it. The X-ray structure without the arm residues and without the water molecules, which are absent in the NMR-generated search model, has an R-factor of 40.2%. At this point, the combined differences between the truncated search model and the corresponding part of the X-ray structure, small as they may be, only serve to worsen the fit of the model to the X-ray data.

An important feature of the molecular replacement search is that the standard procedure in X-plor for generating NMR structures uses a scale factor for the van der Waals radii of all atoms of less than one (Nilges et al., 1988). This scale factor ensures that the atoms are not hindered too much by other atoms, thus accomplishing a more thorough sampling of the conformational space available to the molecule. The resulting structures obtained with the standard scale factor of 0.75 are, however, too compact. In our case, increasing the value to 0.85 finally gave a search model with which molecular replacement was successful. The quality of the Ramachandran plot in a test set of structures generated with X-plor using only the NMR restraints for HUBst increased from 72.9% in the most favoured region for a scale factor of 0.75 to 82.9% for a scale factor of 0.95. It has been noted in a study of geometric properties of a large set of NMR structures (Doreleijers et al., 1998) that structures generated using X-plor often show this exaggerated compactness and, concomitantly, poor geometric quality.

The rotation search in molecular replacement gave two clear solutions for most search models, but the translation search only gave good results for one search model generated with scale = 0.85. The R-factor at this point was 50.3% (resolution range 25 to 4.5 Å) with a correlation coefficient of 41.3; values which are far from ideal. Molecular replacement searches on various arm fragments, with the core regions taken as fixed, were unsuccessful. The fragments from models generated with a scale factor of 0.85 which had the lowest rms deviation to the conformation of the arms in the X-ray models had to be pasted onto the core regions manually, yielding an R-factor of 53.4%. Molecular replacement on the 'consensus' model, which consists of fragments of some of the 50 NMR models from PDB entry 1HUE with favourable geometry least-squares fitted onto the X-ray models, was less problematic. Both the search for the core and for the entire HUBst dimer gave good results, the latter giving an R-factor of 51.5% with a correlation coefficient of 35.9. With these two starting models, joint refinement was initiated.

Joint refinement

The scale factor for the NMR data was kept high throughout the entire process. Since the crystallographic reflections greatly outnumber the NMR restraints in quantity (18 200 versus 2700) their information tends to dominate the direction of the refinement. In order to balance the two datasets properly, the weighting factor used for the X-ray reflections was kept lower than the value that would have been used in refinement against the X-ray data alone.

The joint-refinement process halted at an R-factor above 30%, with a free R-factor about 5% above the working R and about twenty NOE violations greater than 0.5 Å with an rmsd of 0.12 Å (75 waters in the model and individual B-factors). At this point, either the X-ray reflections could be given more weight, thus decreasing the R-factors, or the scale factor on the NMR restraints could be increased to lower the NOE violations. Either option would have disturbed the carefully maintained balance between the two datasets. The results obtained with the 'consensus' model as a starting model for joint-refinement were very similar.

Performing the spin-diffusion analysis by calculating proton-proton distances in the model obtained at this stage and lengthening those NOEs that have probably been affected by spin diffusion allowed us to continue the joint refinement, with a higher weight on the X-ray data than was used previously. The final model showed good geometry and a high Ramachandran score (95.5% in most favoured, 4.5% in additionally allowed regions, see Figure 4d). The final R-factor was 27.4%, R_{free} 30.8% and 4 NOE violations greater than 0.5 Å, with an rmsd of 0.06 Å.

Main chain differences

Right from the beginning of the joint-refinement procedure, three problematic areas in the protein, where differences in the main-chain conformation were observed, were identified. They concerned peptide planes which were 'flipped' in the NMR structure with respect to the X-ray structure:

- between residues 2 and 3: the N-terminal α-helix in the X-ray structure starts at residue 2, in the NMR models at residue 3;
- between residues 42 and 43: these two residues are part of the first β-strand;
- between residues 82 and 83: the C-terminal α-helix in the X-ray structure starts at residue 82, in the NMR models at residue 83.

The causes for these so-called 'peptide flips' (Jones et al., 1991) were closely examined. In the last case, illustrated in Figure 5, the culprit was identified as a hydrogen-bond restraint which was erroneously assigned between the oxygen atoms of residue 82 to the amide hydrogen of residue 84. Secondary-structure analysis on the NMR data (Vis et al., 1994) had already indicated that the C-terminal α-helix spanned from residue 83, possibly 82, to residue 90. The hydrogen-bonding partners for slowly exchanging amide hydrogens, which are assumed to be involved in secondary-structure forming hydrogen bonds, are always difficult to identify unequivocally. In the case of HUBst, the assignment was done based on initial models in the structure determination process. A small hydrogen-bond assignment error in this case had the effect of pushing the intermediate residue, 83, into the disallowed region of the Ramachandran plot. Since hydrogen-bond restraints are usually short distances, they have a major effect on the structure calculations. Correction of the aforementioned restraint to 82 O–86 H^N resolved the conflict for this particular peptide plane: residue 82 became a part of the C-terminal

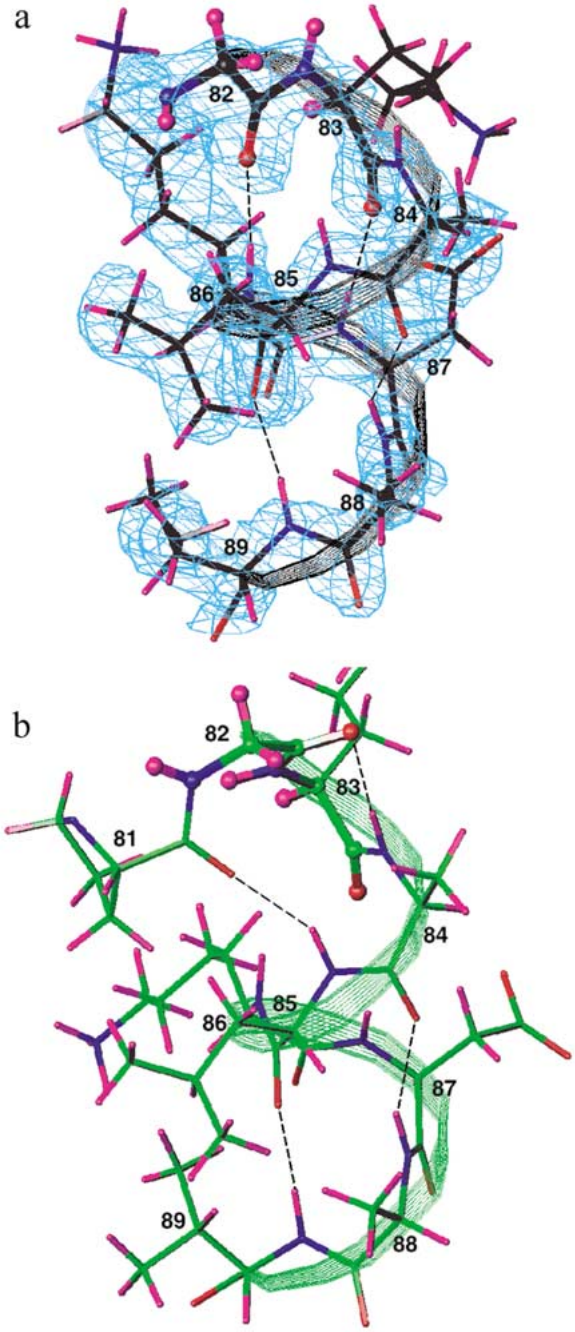

Figure 5. Orientation of peptide bond between residues 82 and 83 (highlighted): (a) In the X-ray structure; the $3F^{obs}$-$2F^{calc}$ electron density map is shown as a light blue mesh, hydrogen bonds are shown with dashed lines. Part of the α-helix is indicated by a ribbon; (b) in an NMR structure; the dashed lines represent hydrogen-bond restraints.

α-helix and residue 83 moved to the more favoured regions of the Ramachandran plot.

The second case, of the peptide plane between residues 42 and 43, is more puzzling. In the X-ray structure, the carbonyl oxygen of residue 42 forms a hydrogen bond with the amide nitrogen of residue 50 in the second β-strand. There is no hydrogen-bond restraint for these particular atoms in the NMR data, even though both flanking hydrogen bonds (between 42 H^N and 50 O and between 44 H^N and 48 O, see Figure 6) are defined as restraints. About half the NMR models in 1HUE (26 out of 50) have the hydrogen bond between 42 O and 50 H^N present, in the other models the peptide plane is 'flipped'. Even though it seems unlikely that the β-sheet would be interrupted midway while continuing on either side, it should be noted that studies on the mobility of NH bonds in HUBst have shown that this β-strand is more mobile than might be expected (Vis et al., 1998). Furthermore, the 'aberrant' models still fully satisfy the NMR restraints. To avoid the ambiguity, the hydrogen-bond restraint was incorporated into the dataset in calculations in this study.

In the case of the first peptide plane, the X-ray structure shows that the amide nitrogen of residue 3 is involved in a hydrogen bond with a carboxylic oxygen of a symmetry-related molecule (see Figure 7) in both monomers. In the earlier, incomplete HUBst structure (PDB entry 1HUU), all three independent monomers in the asymmetric unit equally show hydrogen bonds between different oxygens from symmetry-related molecules to this amide nitrogen. This indicates that the orientation of the peptide plane between residues 2 and 3 could be influenced by packing contacts in the two crystal structures. In this particular region, the NMR structure may well give a better indication of the molecule's conformation free in solution.

Side-chain differences

The dimer interface of HUBst consists entirely of hydrophobic residues: all five leucines, all four phenylalanines and several other hydrophobic residues are situated in this region. There were many differences in side-chain conformations; the most important ones are located in the interface between the two monomers. Leu16, for example, has an unusual conformation in 47 of the 50 NMR models ($\chi^1 = +49 \pm 1°$, $\chi^2 = +44 \pm 3°$). A study of the distribution of side-chain torsion angles in high-resolution X-ray structures (Kleywegt et al., 1998) has shown that 94% of all leucines have $\chi^1 = -60$ or 180°. Interestingly, one of the NOE restraints, between H^α and MD1 of Leu16, is severely violated for these 47 models; in the remaining 3 models, Leu16 has a much more acceptable conformation (average $\chi^1 = -68 \pm 23°$, $\chi^2 = +174 \pm 3°$) but in these models, three other NOEs are violated. This situation looks like a tug-of-war between different NOEs, with the randomized starting point of the structure calculation determining which conformation will prevail. The spin-diffusion analysis described earlier indicates that all NOEs involved have probably been affected by spin diffusion and their upper-bounds should therefore be lengthened for Leu16 to reach the correct conformation. This is probably a general phenomenon for all the residues in the dimer interface, since all the hydrophobic residues in this region contribute to a high local density in hydrogen atoms, which increases the chance of indirect transfer of magnetization via other protons.

Conclusions

When a protein structure has been independently determined by NMR spectroscopy and X-ray diffraction experiments, it is usually not satisfactory to simply compare the resulting structures. Differences in experimental conditions during data collection, in information content of the data, and in protocols and force fields used for the determination of the two structures can render a straightforward comparison inappropriate. In our experience, the combination of NMR and X-ray data in the determination of a protein structure may have presented some 'logistic' difficulties, but the insight gained in the process was extremely valuable. In addition, errors in the experimental data or structure determination process could be identified and corrected.

Globally, the NMR and X-ray data for HUBst are in agreement. Both methods give structures which are very similar in secondary-structure elements and their relative orientation. This is confirmed by the rms deviation of C^α positions of the core residues which is approximately 1 Å between separate X-ray and NMR models. One obvious difference lies in the flexibility of the arms, extensively studied by NMR techniques, which is limited to two distinct conformations in the X-ray structure, and therefore also in the joint-refinement model.

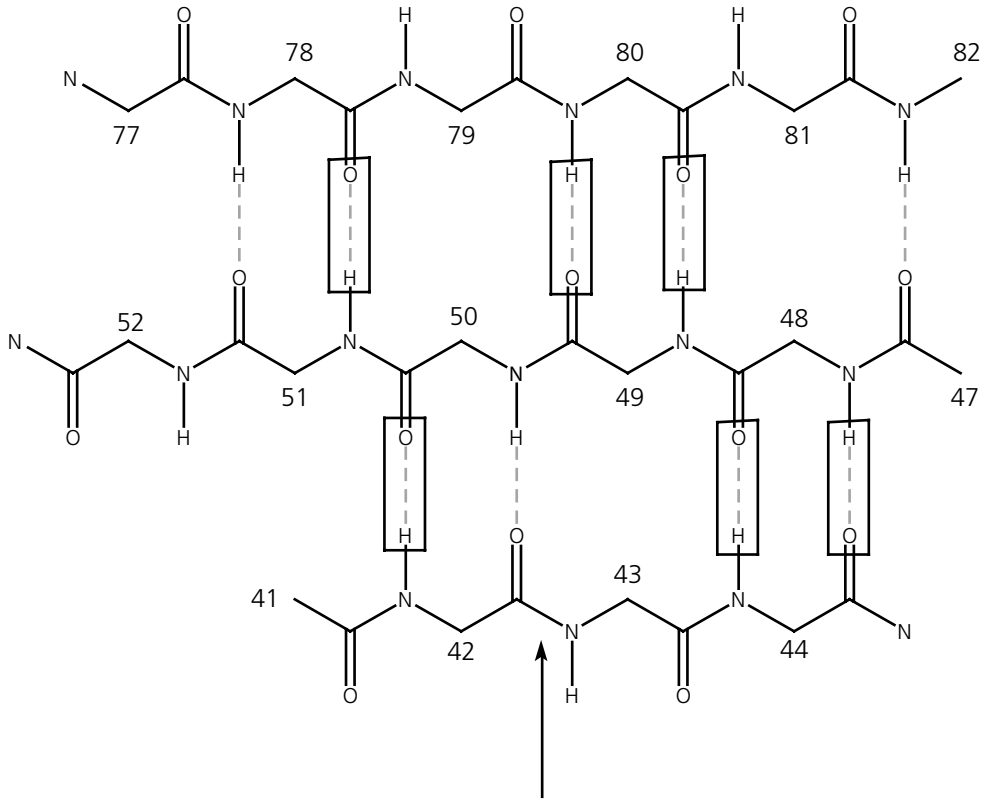

Figure 6. Schematic representation of the core β-sheet of HUBst. Dashed lines represent hydrogen bonds, boxed dashed lines are hydrogen-bond restraints. The arrow indicates the peptide bond between residues 42 and 43, discussed in the text.

The most difficult step in the joint-refinement procedure proved to be molecular replacement using NMR-generated models. This was due to several factors, including the fact that the arms are likely to be in a different position compared to that seen in the crystal structure, and they can therefore not be included in the search model. Furthermore, NMR data consists for the greater part of proton-proton distances; information on the geometry of heavy atoms depends on the force field used in the calculations. Especially for the non-bonded contacts, we have seen that there can be problems with the standard X-plor protocols.

Starting with two polypeptide chains of random conformation, we used only the NMR restraints to fold them into a dimeric molecule and performed molecular replacement with this initial model. When the joint-refinement process halted, we analyzed the model for possible spin-diffusion influences and adjusted the affected NOEs, which gave a significant improvement both in terms of R-factors and NOE violations.

The joint-refinement procedure greatly aided in identifying problematic regions, such as peptide flips in the main-chain conformation, where there were discrepancies between the NMR and X-ray data. In some cases the differences could be attributed to artefacts in either method, such as errors in the assignment of hydrogen-bonding partners or effects of spin diffusion in the NMR data, or crystal packing in the X-ray structure. The effects of spin diffusion could be accounted for by the use of relaxation matrix based structure calculation protocols (Boelens et al., 1988; Borgias and James, 1990; Bonvin et al., 1994). However, routine use of these methods would require renewed efforts in developing more efficient and user-friendly programs than are hitherto available.

Simultaneous refinement of a protein structure against NMR and X-ray data is a useful protein structure validation tool, as it allows the detailed examination of discrepancies at the level of the experimental data.

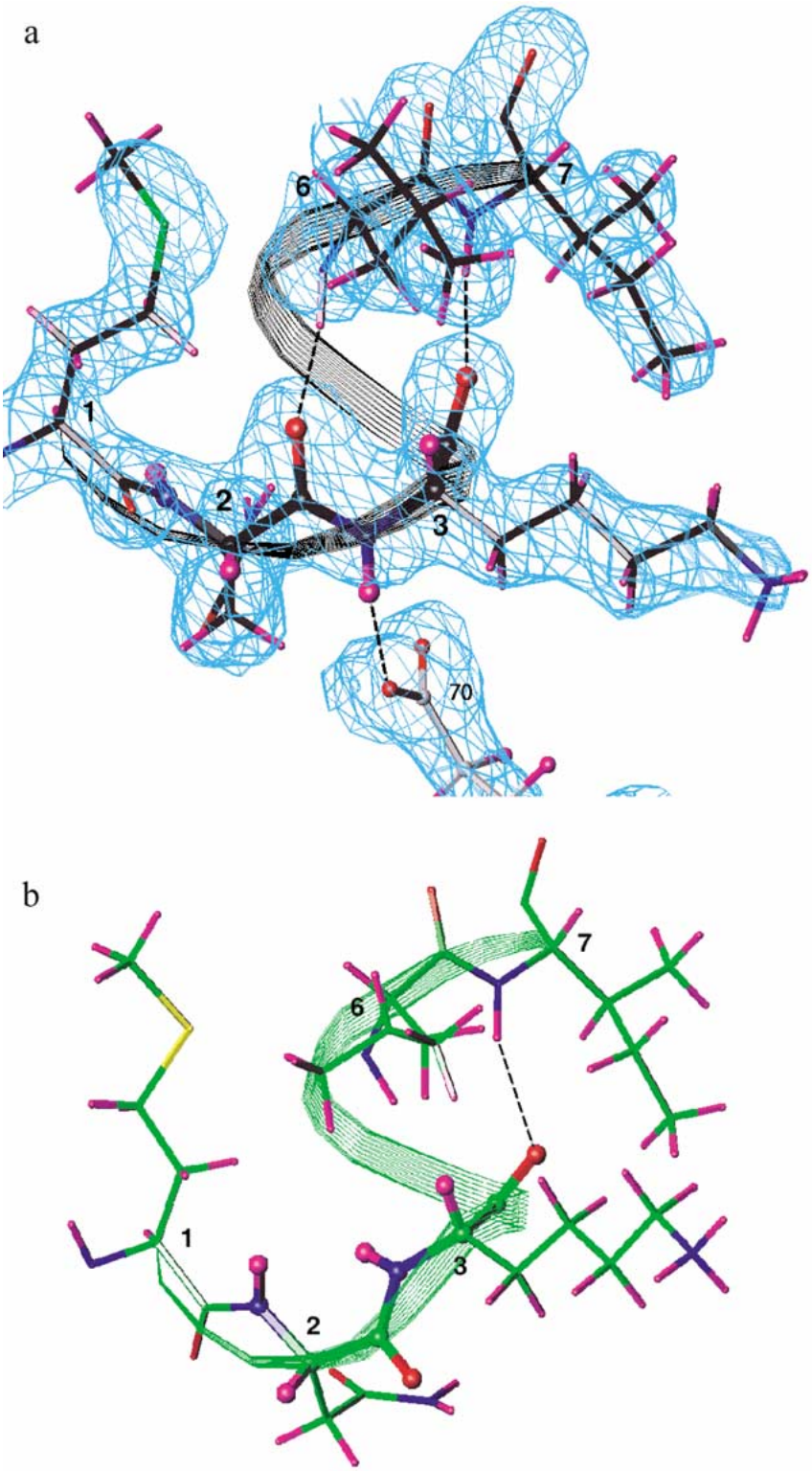

Figure 7. Orientation of peptide bond between residues 2 and 3 (highlighted): (a) In the X-ray structure; the $3F^{obs}$-$2F^{calc}$ electron density map is shown as a light blue mesh, hydrogen bonds are shown with dashed lines. Part of the α-helix is indicated by a ribbon. Residue 70 is a glutamic Acid from a symmetry-related molecule; (b) in an NMR structure.

Acknowledgements

We would like to acknowledge Dr Zbigniew Dauter and Prof Rolf Boelens for invaluable discussions on methods for and results of this work. MLR is funded by the BIOTECH program of DGXII of the Commission of the European Union (BIO4-CT96-0189), for which both RK and KSW are grant-holders.

References

Boelens, R., Koning, T.M.G. and Kaptein, R. (1988) *J. Mol. Biol.*, **173**, 299–311.

Bonvin, A.M.J.J., Vis, H., Breg, J.N., Burgering, M.J.M., Boelens, R. and Kaptein, R. (1994) *J. Mol. Biol.*, **236**, 328–341.

Borgias, B.A. and James, T.L. (1990) *J. Magn. Reson.*, **87**, 475–487.

Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) *Science*, **235**, 458–460.

Doreleijers, J.F., Raves, M.L., Rullmann, J.A.C. and Kaptein, R. (1999) *J. Biomol. NMR*, **14**, 123–132.

Doreleijers, J.F., Rullmann, J.A.C. and Kaptein, R. (1998) *J. Mol. Biol.*, **281**, 149–164.

Hoffman, D.W., Cameron, C.S., Davies, C., White, S.W. and Ramakrishnan, V. (1996) *J. Mol. Biol.*, **264**, 1058–1071.

Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Crystallogr.*, **A47**, 110–119.

Kalk, A. and Berendsen, H.J.C. (1976) *J. Magn. Reson.*, **24**, 343–366.

Karimi-Nejad, Y., Löhr, F., Schipper, D., Rüterjans, H. and Boelens, R. (1999) *Chem. Phys. Lett.*, **300**, 706–712.

Kleywegt, G.J. and Jones, T.A. (1996) *Acta Crystallogr.*, **D52**, 829–832.

Kleywegt, G.J. and Jones, T.A. (1998) *Acta Crystallogr.*, **D54**, 1119–1131.

Kobryn, K., Lavoie, B.D., Chaconas, G. (1999) *J. Mol. Biol.*, **289**, 777–784.

Laskowski, R.A., MacArthur, M.W., Moss, D. and Thornton, J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Miller, M., Lubkowski, J., Rao, J.K.M., Danishefsky, A.T., Omichinski, J.G., Sakaguchi, K., Sakamoto, H., Appella, E., Gronenborn, A.M. and Clore, G.M. (1996) *FEBS Lett.*, **399**, 166–170.

Navaza, J. (1994) *Acta Crystallogr.*, **A50**, 157–163.

Nilges, M. (1996) *Curr. Opin. Struc. Biol.*, **6**, 617–623.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.

Nilges, M., Habazettl, J., Brünger, A.T. and Holak, T.A. (1991) *J. Mol. Biol.*, **219**, 499–510.

Rice, P.A., Yang, S., Mizuuchi, K. and Nash, H.A. (1996) *Cell*, **87**, 1295–1306.

Schiffer, C.A., Huber, R., Wütrich, K. and van Gunsteren, W.F. (1994) *J. Mol. Biol.*, **241**, 588–599.

Shaanan, B., Gronenborn, A.M., Cohen, G., Gilliland, G.L., Veerapandian, B., Davies, D.R. and Clore, G.M. (1992) *Science*, **257**, 961–964.

Silva, M.V., Pasternack, L.B. and Kearns, D.R. (1997) *Arch. Biochem. Biophys.*, **348**, 255–261.

Sussman, J.L., Lin, D., Jiang, J., Manning, N.O., Prilusky, J., Ritter, O. and Abola, E. (1998), *Acta Crystallogr.*, **D54**, 1078–1084.

Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) *Nature*, **310**, 376–381.

Vis, H., Boelens, R., Mariani, M., Stroop, R., Vorgias, C.E., Wilson, K.S. and Kaptein, R. (1994) *Biochemistry*, **33**, 14858–14870.

Vis, H., Mariani, M., Vorgias, C.E., Wilson, K.S., Kaptein, R. and Boelens, R. (1995a) *J. Mol. Biol.*, **254**, 692–703.

Vis, H., Mariani, M., Vorgias, C.E., Wilson, K.S., Kaptein, R. and Boelens, R. (1995b) *Quart. Magn. Reson. Biol. Med.*, **2**, 107–117.

Vis, H., Vageli, O., Nagel, J., Vorgias, C.E., Wilson, K.S., Kaptein, R. and Boelens, R. (1996) *Magn. Reson. Chem.*, **34**, S81–S86.

Vis, H., Vorgias, C.E., Wilson, K.S., Kaptein, R. and Boelens, R. (1998) *J. Biomol. NMR*, **11**, 265–277.

White, S.W., Wilson, K.S., Appelt, K. and Tanaka, I. (1999) *Acta Crystallogr.*, **D55**, 801–809.