Journal of Biomolecular NMR (2005) 33: 175–186
DOI 10.1007/s10858-005-3207-9

Article

Describing partially unfolded states of proteins from sparse NMR data

Gloria Fuentes, Aart J. Nederveen, Robert Kaptein, Rolf Boelens & Alexandre M. J. J. Bonvin*

*NMR Research Group, Bijvoet Center for Biomolecular Research, Utrecht University,
Padualaan 8, 3584 CH, Utrecht, The Netherlands*

Received 9 June 2005; Accepted 13 September 2005

Key words: chemical shifts, native contacts, partially unfolded states, structure calculations

Abstract

Proteins involved in signal transduction can usually be present in two states: an inactive and an active (signaling) state. In the case of photoreceptors such as PYP, it has been shown that the signaling state has a large degree of structural and dynamic disorder. Conventional structural NMR approaches present difficulties in describing such partially unfolded states. Owing to the disordered dynamical and transient nature of such states classical NOE-based information, when present, is sparse. Chemical shift changes upon partial unfolding can, however, be easily monitored from HSQC spectra. We show here that such states can be modeled by defining native-like inter-residue contacts for those residues that do not shift significantly upon partial unfolding. The feasibility of this approach is demonstrated using lysozyme as a test case and applied to model the partially unfolded signaling state (pB) of a truncated form of the photoactive yellow protein for which a “classical” NOE-based structure is available for validation. This approach should be generally applicable to systems in which part of the structure remains in a well-defined native-like conformation.

Introduction

There is an increasing interest in characterizing the structural features and evaluating the roles of unfolded and partially folded protein conformations (Wright and Dyson, 1999). In the cell, partially unfolded forms of proteins are involved in a broad range of biological processes, such as translocation across membrane (Schwartz et al., 1999) and protein degradation within the cell (Dill and Shortle, 1991; Ptitsyn, 1995). In the case of proteins involved in signal transduction, it has been shown that they present a large degree of structural and dynamic disorder in the active (signaling) state. This is the case for the photoactive yellow protein (PYP) (Rubinstenn et al., 1998), a photoreceptor thought to be involved in a phototactic response of the bacterium *Ectothiorhodhospira halophila* to intense blue light. Apart from their significance in normal cellular processes, some of these states appear to be associated with protein aggregation, which is of significance in regard to our understanding of amyloid-associated diseases such as Alzheimer’s and spongiform encephalopathies (Thomas et al., 1995; Bucciantini et al., 2002). Structural information on such states will further enhance our understanding of one of the most intriguing problems in biology: the mechanism of protein folding.

NMR spectroscopy is a particularly important tool for investigating protein folding allowing the study of conformation and dynamics of unfolded, partially folded and native states at atomic resolution (Dyson and Wright, 1996; Shortle and

*To whom correspondence should be addressed. E-mail: a.m.j.j.bonvin@chem.uu.nl

Ackerman, 2001). There are, however, NMR spectral features that hamper the structure determination of a protein in a (partially) unfolded state. The NMR data can no longer be interpreted in the context of a single conformation; instead they reflect a dynamical average over an ensemble of conformations. Although some methods to deal with such averaging have been developed for native proteins (Bonvin et al., 1994; Mierke et al., 1994; Bonvin and Brunger, 1995), in the case of partially unfolded states they need to be adapted to describe larger conformational ensembles. The low dispersion of ^{1}H and ^{13}C chemical shifts, the dynamic, mobile character of any residual structure, and the frequent presence of severe line broadening, lead to a lower density of structural restraints that can be collected from NMR experiments (Shortle, 1996). The sparse NOE information will thus have to be complemented by other types of experimental restraints (such as chemical shift-derived native structure information).

Conventional molecular dynamics simulations have been used to complement the analysis of the denatured states by NMR in order to obtain more detailed structural information on components of the denatured ensemble (Daggett, 2002). However, they are in themselves not always adequate for dealing with large ensembles of partially unfolded protein conformations, due to simulation time limitations (Daggett, 2000). Therefore, new computational tools need to be developed to characterize partially unfolded states of proteins (Fersht and Daggett, 2002). Vendruscolo and collaborators have been particularly active in this area. Using ϕ values obtained from kinetic data on engineered mutants (Fersht et al., 1986; Matouschek et al., 1989; Fersht et al., 1992) they have developed a Monte-Carlo approach based on native-like contacts to describe the structure of transition states (Vendruscolo et al., 2001; Paci et al., 2002, 2003). Recently, they have applied a similar approach based on ensemble-average MD simulation using different sources of experimental NMR data. Data obtained from relaxation dispersion (Korzhnev et al., 2004) experiments, paramagnetic relaxation enhancement (PRE) (Lindorff-Larsen et al., 2004; Dedmon et al., 2005) experiments and experimental order parameters (Best and Vendruscolo, 2004) (S^2) have been used to model ensembles of conformations of non-native states for several proteins.

In this work, we build on the idea of using native-like contacts to characterize, by NMR, partially unfolded states. By partially unfolded states we mean here systems in which a fraction of the structure remains in a well-defined native-like conformation (excluding for example molten globule states). Provided that the native structure is known, chemical shift information from HSQC spectra can be used to define native-like inter-residue contacts for those residues that do not shift upon partial unfolding. The exact definition of those native inter-residue contacts (or "ground state" contacts in the context of signaling proteins) and how chemical shift information can be best translated into them is first investigated using synthetic data for lysozyme. We then apply this method to describe the partially unfolded state (pB) of a truncated form of the PYP (van der Horst et al., 2001), Δ25PYP, and compare our results with a structure recently solved using classical NOE information (Bernard et al., 2005).

Materials and methods

Generation of a reference partially unfolded state of lysozyme using MD simulations

All simulations were performed with the GROMACS 3.1.3 molecular dynamics package (Lindahl et al., 2001), using the GROMOS 43a3 force field (Daura et al., 1998). The MD run in water was performed at 300 K in a truncated octahedron box, filled with 17225 SPC (Berendsen et al., 1981) water molecules and eight additional Cl^- ions to electro-neutralize the system.

In the case of the MD simulation, solute, solvent and counterions were independently weakly coupled to reference temperature baths at 300 K ($\tau = 0.1$ ps) (Berendsen et al., 1984), and the pressure was maintained by weakly coupling the system to an external pressure bath at one atmosphere. For the stochastic dynamic (SD) simulations, the temperature of the system (600 K) was regulated by stochastic forces. In both cases, The LINCS algorithm (Hess et al., 1997) was used to constrain bond lengths, allowing an integration time step of 0.002 ps (2 fs) to be used. The nonbonded interactions were calculated with a twin-range cutoff (van Gunsteren and Berendsen, 1990) of 0.8 and 1.4 nm. The long-range electrostatic interactions beyond the 1.4 nm cutoff were treated with the generalized reaction field model (Tironi et al., 1995) using a dielectric constant of 54. The non-bonded interaction pair list was updated every five steps. (for further details refer to Hsu and Bonvin (2004)).

Simulation of the HSQC spectra

SHIFTX (Neal et al., 2003) was used to simulate the ^{15}N and ^{1}H chemical shifts for the native ensemble (10 conformers taken every 50 ps for the last 500 ps of the MD simulation) and for the partially unfolded ensemble (20 conformers, taken every 100 ps from the last 2 ns of the SD simulation). The ^{15}N and ^{1}H chemical shifts were calculated for all the conformers belonging to the ensemble and, subsequently averaged. Since SHIFTX does not provide chemical shifts for side-chains carbon atoms, the generation of the ^{13}C chemical shifts was performed with the server PROSHIFT (Meiler, 2003).

Definition of native-like restraints

In order to define the native contacts in the unfolded state, we use chemical shift differences of ^{15}N–^{1}H HSQC peaks between the native and the partially unfolded state. The differences are calculated as $\mathrm{sqrt}[(\Delta\sigma_{HN})^2 + [(\Delta\sigma_N/6.515)^2]$ (Farmer, 1996; Mulder et al., 1999). In the definition of the amino acids that do not shift during the unfolding event, a cutoff was calculated using an iterative process. First, the average chemical shift differences and standard deviations are calculated for all the amino acids in the protein. Then, those amino acids, presenting a chemical shift difference larger than the calculated average plus one standard deviation, are excluded and a new averaging round in performed. The convergence of the average chemical shift difference and standard deviation is monitored. The cutoff is chosen as the value at which the latter starts to reach a plateau. All the residues with values below the cutoff are considered to be in their native environment and their C^{α}, C^{β} contacts, from these residues to all other residues within a 7.5 Å cutoff and at least two residues further in the sequence, are included in the restraint set. While native-like contacts are defined only for those residues that do not show significant chemical shift changes, the contacting residues to which the restraints are defined might well have significant shifts since part of their 3D environment might be changing.

Native-like restraints for methyl groups of VAL, ILE and LEU that do not shift during partial unfolding are defined by calculating all distances within a 5 Å cutoff between the C^{δ} and C^{γ} of these amino acids to any C atom of the protein.

Native contacts restraints are defined from the native state structure as followed: the upper bound is set to the averaged distance calculated over the ensemble plus one standard deviation, while the lower bound is taken as the sum of the van der Waals radii.

Secondary chemical shifts of available nuclei (H^{α}, HN, C′, C^{α} and C^{β}) in the partially unfolded state were used in Talos (Cornilescu et al., 1999) to define secondary structure restraints (ϕ/ψ dihedral angles). This was however only done for those residues that are in a helical or stranded conformation in the native state. The Talos 'good' predictions were transformed into dihedral angle restraints as the average ϕ and ψ angles ± two times the standard deviation with a minimum error of 30°.

Structure calculations

Structure calculations were performed with CNS (Brunger et al., 1998) using a simulated annealing protocol derived from ARIA (Linge et al., 2001) followed by refinement in explicit solvent (Nederveen et al., 2005). The PARALLHDG5.3 force field with the PROLSQ parameters was used (Linge et al., 2003) during the simulated annealing protocol, while the OPLS non-bonded parameters (Jorgensen and Tirado-Rives, 1988) were applied during the water refinement. In the case of the Δ25PYP, the topology file was manually adjusted to describe the chromophore. Atomic charges for the chromophore were taken from Groenhof et al. (2002).

During the simulated annealing protocol, 300 structures were calculated and only the 50 lowest energy structures were submitted to water refinement. A final ensemble of the 20 lowest energy structures with no violations was chosen from the water-refined structures and subjected to validation, in order to obtain an indication of its quality and structural statistics. We have used the following programs: PROCHECK (Laskowski et al., 1993), ROCHECK_NMR (Laskowski et al., 1996) and WHAT IF (Vriend, 1990). Violations of distance and dihedral restraints for the models in the final ensemble were calculated using CNS protocols.

Methyl group chemical shift analysis

It is possible to use the information from methyl groups as 3D reporters, only if there is no correlation between the different rotameric states and the chemical shifts. In order to analyze this correlation, we searched the BMRB (Seavey et al., 1991) for references to PDB (Bernstein et al., 1977) entries. We collected 14 BMRB-PDB matches of high resolution structures, for which side chain chemical shifts were also deposited. A total of 337 side chain torsion angles (χ_1) and 497 side-chain chemical shifts (for Cγ for VAL, LEU and ILE were analyzed.

Depending on the angle, these chemical shifts were grouped in the three possible staggered conformations, defined as trans, gauche$^+$ and gauche$^-$. For each conformation and residue, histograms were plotted, with a bin size of 0.50 ppm.

Results and discussion

Test case: lysozyme

Lysozyme was chosen as a model system to develop and validate our approach. It is composed of two domains (α and β) with the active site cleft situated between them. The α-domain consists of four α-helices and a C-terminal 3_{10} helix. The primary component of the β-domain is a three-stranded β-sheet, followed by a long loop and a 3_{10} helix (Figure 1a). These structural features make this protein a suitable model for the study of partially unfolded states.

Generation of the reference native and partially unfolded state ensembles

For generating the reference, folded state of lysozyme, allowing for some degree of flexibility to be expected in solution, we performed a 2 ns MD simulation in water at 300 K starting from the high-resolution X-ray structure (Artymiuk et al., 1982) (PDB entry 1AKI). The native state was modeled by an ensemble of 10 energy minimized snapshots taken every 50 ps from the last 500 ps (Figure 1b). As model for the partially unfolded state, we decided to unfold the β-domain of lysozyme. This was achieved by running a SD simulation at 600 K with position restraints on the heavy atoms of the α-domain (corresponding to the 4 α-helices in the structure: αA, αB, αC and αD). This simulation was started from the last structure of the MD simulation in water and run for 4 ns (for details see Material and methods). The partially unfolded ensemble was described by 20 snapshots taken every 100 ps from the last 2 ns of the SD simulation (Figure 1c). The α-domain in this ensemble deviates from the reference native structure (the closest to the mean from the native ensemble) by 1.01 ± 0.03 Å and 1.75 ± 0.02 Å for backbone and heavy atoms, respectively, while the β-domain shows a large conformational dispersion, with no apparent well-defined motif (backbone RMSD of 15.5 ± 0.2 Å from the reference native structure).

Definition of native-like inter-residue contacts

We have investigated various schemes of defining native-like restraints, evaluating their performance by calculating native-like structures. The native fold could be properly reproduced using distance restraints defined between C$^\alpha$ and C$^\beta$ atoms within a 7.5 Å cutoff for amino acids at least two residues apart (i, $i+n$; $n \geq 2$) in the sequence.

Using chemical shift information to describe partially unfolded states

Structural information to be used in modeling partially unfolded states from chemical shifts can be in principle derived from three sources depending on their availability:

i. chemical shift perturbation (CSP) data from ^{1}H–^{15}N HSQC spectra
ii. secondary chemical shifts from backbone nuclei (H$_\alpha$, C$'$, C$_\alpha$, C$_\beta$)
iii. chemical shift perturbation data from ^{1}H–^{13}C HSQC spectra.

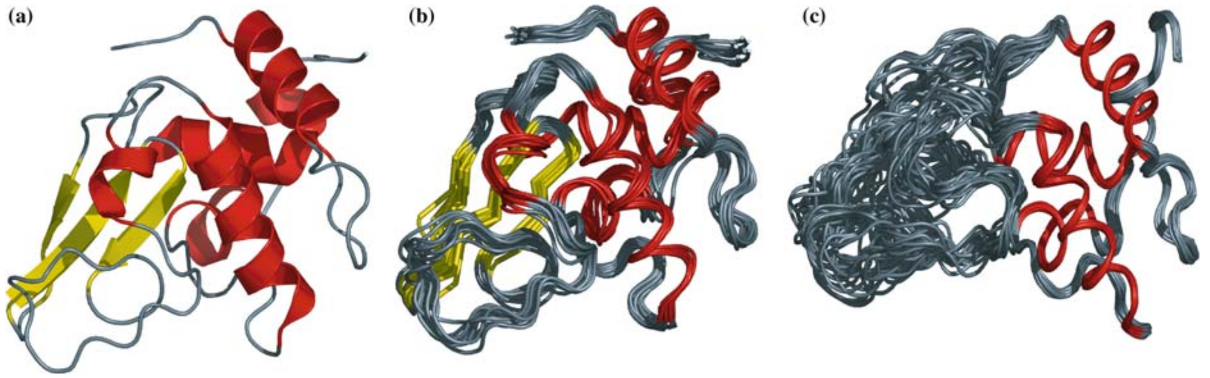

Figure 1. Structure representation of the different states for lysozyme. (a) Lysozyme native state reference structure; (b) ensemble of 10 structures for the partially folded state; (c) ensemble of 19 structures for the partially unfolded state. The α-domain is shown in red, and those residues belonging to the β-strands have been colored in yellow.

The main idea is to define native-like contacts from CSP data for those residues that do not significantly shift upon partial unfolding, while secondary chemical shifts can be used to define backbone ϕ,ψ dihedral angle restraints using approaches such as CSI (Wishart and Sykes, 1994) or Talos (Cornilescu et al., 1999). CSP data derived by comparing ^{1}H–^{15}N HSQC spectra of native and partially unfolded species are typically the most readily available, while the other two sources rely on the availability of (backbone) ^{13}C chemical shift assignments.

The feasibility of this approach was tested with the native and partially unfolded states of our model system, lysozyme. For this, we calculated average chemical shifts using SHIFTX (Neal et al., 2003) for the native and partially unfolded reference ensembles. These chemical shifts were then used to simulate HSQC spectra and calculate chemical shift perturbations (differences) resulting from the partial unfolding. In contrast to NMR titration experiments used in interface mapping, which is based on identifying large shifts, here one should select residues showing only small shifts. The selection cutoff was determined by calculating the average CSP, removing all outlier above the calculated average value and recalculating the average for the new set until convergence was found. All residues with shifts below this cutoff were then assumed to be in a native-like structural environment. Note that it can happen that a residue in the native structure has chemical shifts close to random coil values. For those cases, unfolding will not lead to significant chemical shift changes and those residues should be excluded when defining native-like constraints.

In the case of the ^{1}H–^{15}N HSQC CSP data for lyzozyme, this procedure led to a cutoff value of 0.27 ppm. With this cutoff, 35 amino acids out of 129 were considered to remain in their native-like environment corresponding to 27% secondary structure compared to the native state (Figure 2). Native contacts for those amino acids resulted in 603 distance restraints, which were subsequently used in structure calculations (see Material and methods) to model the partially unfolded state. The resulting ensemble of structures is shown in Figure 3a. It corresponds to a moderate quality ensemble, with very low RMS distance restraint violation and no consistent violation larger than 0.30 Å. The positional RMSD values for the α-domain (Table 1) indicate that the structure of the still folded domain is well reproduced while the β-domain is unfolded.

When complete backbone assignments are available, the native-like restraints derived from ^{1}H–^{15}N CSP data can be complemented by secondary structure restraints derived from secondary chemical shifts. This should however only be done for residues remaining in a native-like environment to avoid possible conformational averaging problems. For lysozyme, ϕ/ψ dihedral angle restraints were defined for 24 residues. Since TALOS could not be used in this case because of the absence of experimental chemical shifts, we took the dihedral angle values from the native state. These restraints were combined with the previously defined native-like restraints for a

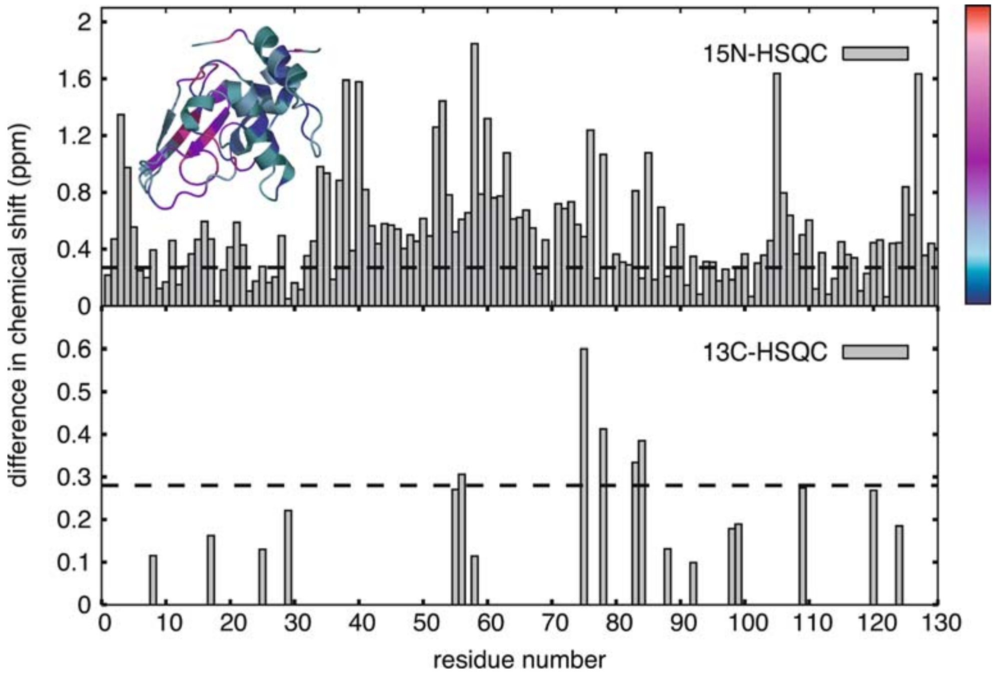

Figure 2. (a) Chemical shift differences between the folded and partially unfolded states for lysozyme: top panel shows the chemical shift perturbation found in the H, N^{15}-HSQC, and the bottom one that found in the H, C^{13}-HSQC for the methyl groups of VAL, LEU and ILE. (b) Color-coding according to the extent of the ^{15}N chemical shift perturbation of the partially unfolded state, with respect to the native one. The color spectrum ranges linearly from blue (0 ppm) to red (2 ppm).

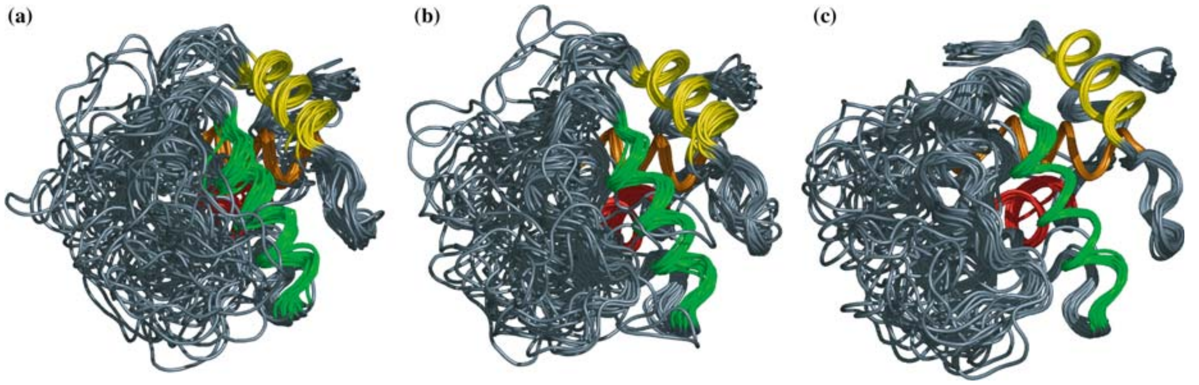

Figure 3. Ensembles of structures (20) calculated for the partially unfolded state of lysozyme using: (a) only distance restraints between Cα and Cβ atoms; (b) Cα Cβ distances and secondary structure dihedral restraints; (c) Cα Cβ distances, secondary structure dihedral restraints and methyl–methyl distances. Structures were superimposed on the backbone atoms of the secondary elements including residues 5–15, 25–35, 88–101,110–115. The α-helices are colored as followed: α1 in yellow, α2 in orange, α3 in green and α4 in red.

new round of structure calculations. The resulting ensemble (Figure 3b) shows smaller RMSDs for the native-like domain (both with respect to the mean and from the reference α-domain structure indicating a more precise and accurate ensemble) compared to the case where only native-like distance restraints were considered (Table 1). The inclusion of the dihedral angle restraints also led to an improvement in the RMS distance restraint violation (from 0.042 ± 0.015 to 0.026 ± 0.010), indicating a better convergence of the calculations.

Finally, in cases where chemical shift assignments of the usually well dispersed methyl group

Table 1. Structural statistics of the 20 best structures of the ensembles calculated for lysozyme

	CαCβ	+ Dihedrals	+ Methyl groups
Restraints statistics[a]			
Number of native-contact distances	603	603	853
Number of dihedral angles	–	48	48
RMS distance violations (Å)	0.04 ± 0.02	0.02 ± 0.01	0.02 ± 0.00
RMS dihedral angle violations (°)	–	0.60 ± 0.25	0.56 ± 0.22
Rmsd (Å) from the mean			
All backbone atoms	4.2 ± 0.9	3.7 ± 1.0	2.7 ± 0.7
All heavy atoms	5.3 ± 0.9	4.7 ± 1.0	3.6 ± 0.7
α-domain secondary structure backbone atoms[a]	1.5 ± 0.5	0.8 ± 0.2	0.5 ± 0.1
Rmsd (Å) from the reference structure[b]			
α-domain secondary structure backbone atoms[c]	1.6 ± 0.4	1.2 ± 0.1	1.0 ± 0.1
α-domain secondary structure heavy atoms[c]	2.7 ± 0.4	2.4 ± 0.2	2.0 ± 0.1
β-domain backbone atoms[d]	16.3 ± 2.0	15.4 ± 1.4	14.9 ± 1.0
Second generation packing quality (Z-score)	−3.4 ± 0.5	−3.1 ± 0.6	−1.9 ± 0.5

[a] No distance and dihedral angle restraints were consistently violated by more than 0.3 Å and 5°, respectively, in more than 50% of the structures.

[b] The reference structure is referred to the closest to the mean from the MD partially unfolded ensemble.

[c] Secondary structure elements comprise residues 5–15, 25–35, 88–101, 110–115.

[d] β-domain comprises the stretch 36–87.

^{1}H,^{13}C resonances are available, we could, in principle, derive native contact restraints for the methyl groups of VAL, ILE and LEU by comparing ^{1}H–^{13}C-HSQC spectra of native and partially unfolded states. This would be particularly valuable since those residues are the most common amino acids in protein cores. Such an approach is, however, only valid provided the corresponding methyl group chemical shifts are sensitive to the 3D structure and do not only report on side-chain rotameric states. To check for this, we therefore performed an analysis of methyl group chemical shifts in the BMRB (Seavey et al., 1991) for which a 3D structure is available. We did not find any correlation between rotameric states and methyl group chemical shift values (Figure 4), which makes the latter attractive reporters of 3D native structures. A correlation between leucine δ-carbon chemical shifts and C$^{\delta}$–C$^{\alpha}$ J couplings has been reported before (MacKenzie et al., 1996); it was however concluded to be too weak to reliable predict rotamer states. This is in line with our findings.

For lysozyme, 12 methyl groups remain in a native-like environment, as indicated by the small chemical shift variations upon partial unfolding (a cutoff of 0.26 ppm was used; bottom panel in Figure 2). From these, 250 native-like distance restraints could be defined from the C$^{\delta}$ and C$^{\gamma}$ methyls to any C atoms within a 5 Å cutoff. These restraints were added, to the previous defined restraints (native contacts from ^{1}H^{15}N-HSQC and dihedral angle restraints) and the structure calculations were rerun. The ensemble obtained combining these three types of restraints (Figure 3c) is

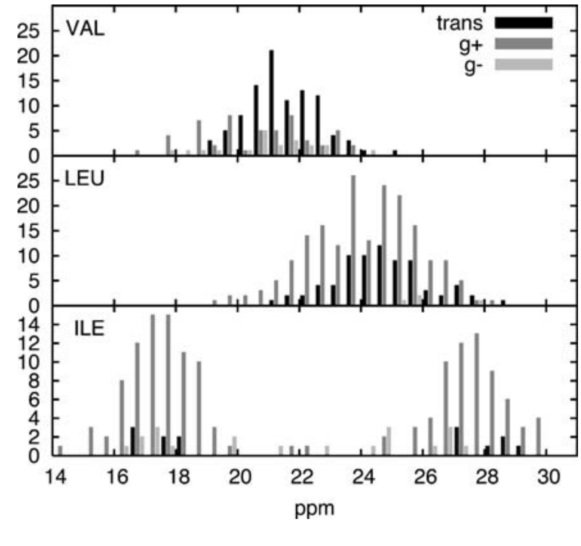

Figure 4. Distribution of the χ_1 side-chain dihedral of selected residues (Val, Ile and Leu) versus the chemical shifts (ppm).

even better defined than the previous cases. The largest impact of the inclusion the methyl group native-like restraints is, however, on the packing of the structure as indicated by an averaged 43% reduction in the second generation packing quality Z-score (see Table 1).

In conclusion, inclusion of restraints from various sources helps in defining the partially unfolded state ensemble.

Application case: Δ25PYP

Having demonstrated the feasibility of our approach on synthetic data for lysozyme, we applied it to model the partially unfolded pB state of Δ25PYP using NMR experimental data. This deletion mutant of PYP, lacking the 25 N-terminal residues, undergoes a similar photocycle albeit with strong decelerated kinetics. The removal of these residues does not affect significantly the fold of the PAS domain (Vreede et al., 2003). These two features make this truncated form a good candidate for structural studies of the long-lived intermediate (pB). The NOE-based structure of the pB state solved in our laboratory (Bernard et al., 2005) (PDB entry: 1XFQ) serves as a reference to validate our approach using real data. It has been shown that upon illumination, the pB state of the system exhibits a β-sheet folding pattern, similar to that of the native state (pG). On the contrary, a much more pronounced and generalized destabilization is observed for the α-helices.

The input data in this case were amide proton and nitrogen chemical shift differences extracted from the ^{1}H–^{15}N-HSQC for the native state (pG) and the lit, partially unfolded state (pB) (see Figure 5). As starting structure, the NMR ensemble of the native state (pG) (Bernard et al., 2005) (PDB entry 1XFN) was considered. The native contacts were calculated between C$^{\alpha}$ and C$^{\beta}$ atoms within 7.5 Å from all 20 structures of the NMR ensemble. Only those present in at least 50% of the structures were considered to define native-like distance restraints, using the calculated average distance plus one standard deviation as upper bound.

Using the procedure described above, we chose a 0.28 ppm cutoff to select from chemical shift differences the residues still considered in their native state environment. Fifty residues were selected corresponding to 50% of the structure in its native-like environment (Figure 5), from which 585 distance restraints were generated. These restraints resulted in an ensemble of structures for the partially unfolded state with low positional RMSDs for the remaining secondary structure elements, both from the mean and from the reference NOE-based structure (Table 2). No consistent distance violation larger than 0.3 Å was found.

It is also possible to validate the generated ensemble with the backbone NOE restraints available for the pB state for those residues we have considered in the native-like environment. These restraints were extracted from the available NOE data selecting only those involving backbone amide protons and corresponding to residues with a chemical shift deviation below the considered cutoff. The RMS distance restraint violation is slightly higher than that of the NOE-based ensembles, with five consistent violations larger than 1 Å (max. 2.2 Å).

A second set of calculations was run, including 94 additional dihedral angle restraints for 47 residues derived from the available chemical shifts for the partially unfolded pB state using Talos

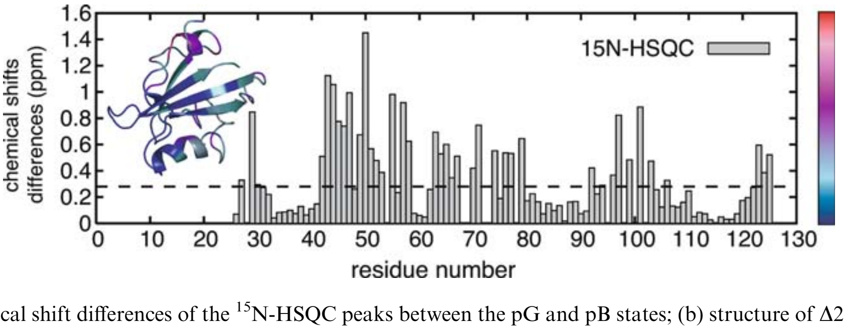

Figure 5. (a) Chemical shift differences of the ^{15}N-HSQC peaks between the pG and pB states; (b) structure of Δ25PYP with chemical shift differences color-coded as in Figure 2b.

Table 2. Structural statistics of the NMR ensembles of pB Δ25PYP

Restraints statistics	NOE based-	Chemical shift based-(CαCβ)	Chemical shift based-(+ dihedrals)
Number of native-contact distances	–	585	585
Number of dihedral angles	–	–	94
Number of NOE restraints[a] (not used here)	157	157	157
RMS distance violations (Å)	–	0.02 ± 0.004	0.01 ± 0.005
RMS dihedral angle violations (°)	–	–	0.78 ± 0.28
RMS NOE distance violations (Å)	0.05 ± 0.01	0.43 ± 0.12	0.32 ± 0.07
Consistent native-contact distances (> 0.30 Å)	–	0	0
Consistent NOE distances (> 1 Å)	0	5	3
Rmsd (Å) from the mean			
All backbone atoms	4.1 ± 0.9	3.4 ± 0.7	2.9 ± 0.6
All heavy atoms	4.9 ± 0.9	4.4 ± 0.7	3.7 ± 0.7
Secondary structure backbone atoms[b]	0.8 ± 0.2	1.5 ± 0.5	1.0 ± 0.3
Secondary structure heavy atoms[b]	1.5 ± 0.2	3.1 ± 0.7	2.4 ± 0.5
Rmsd (Å) from the reference structure[b, c]			
Secondary structure backbone atoms[b]	0.7 ± 0.2	2.2 ± 0.4	1.9 ± 0.3
Secondary structure heavy atoms[b]	1.3 ± 0.4	3.3 ± 0.6	2.8 ± 0.4

[a] From the NOE restraints only those involving backbone protons for the amino acids that do not shift were considered.
[b] Secondary structure elements comprise residues 30–34, 39–43, 77–85, 90–95, 104–111, 117–123.
[c] The reference structure is the structure closest to the mean from the NMR *ensemble of the pB state*.

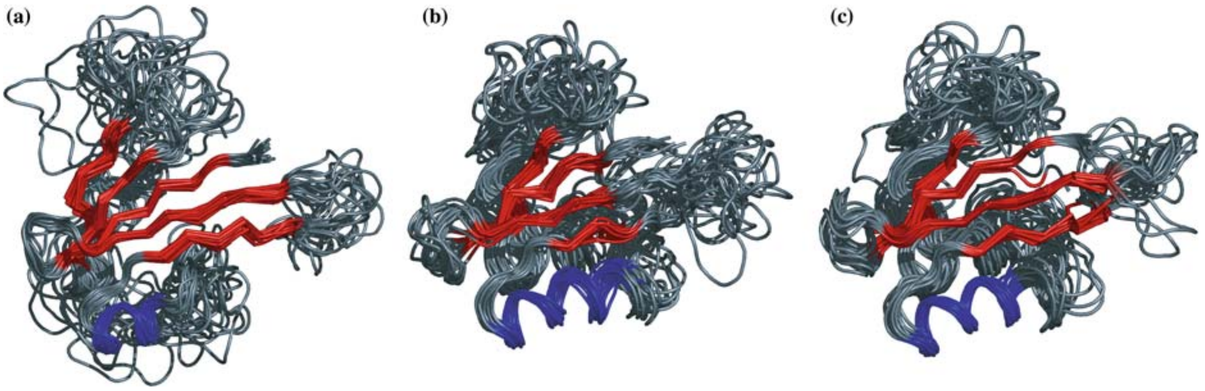

Figure 6. Ensembles of the partially unfolded state of Δ25PYP calculated using different approaches. (a) NOE-based ensemble of 20 NMR structures of pB Δ25PYP; (b) chemical shift-based (^{1}H–^{15}N-HSQC) ensemble of 20 NMR structures of pB Δ25PYP including only CαCβ distances; (c) chemical shift-based (^{1}H–^{15}N-HSQC) ensemble of 20 NMR structures of pB Δ25PYP including only CαCβ distances and secondary structure dihedral restraints. Structures were superimposed on the backbone atoms of the secondary elements determined for the pB NOE based-ensemble defined in Table 3. They are represented and colored in blue for the α-helix (α5) and red for the β-sheets.

(Cornilescu et al., 1999). The resulting ensemble (Figure 6c) presents better positional RMSD values, and a noticeable decrease of the RMS distance violations for the NOE set used for independent validation. The number of consistent NOE violations has been reduced to 3, with a maximum value of 1.70 Å. These few NOE violations should not be considered alarming, since due to the nature of the restraints we are using in our protocol (C–C restraints), only the force field is imposing restraints on the protons, in contrast to the NOE-based procedure, where the NOE restraints are directly acting on protons during the structure calculations.

Figure 6b, c shows the superposition of the 20 lowest energy structures of the light-induced state of Δ25PYP obtained using this protocol, only considering native-contact restraints and with the inclusion of dihedral restraints respectively. The overall fold is similar to the NOE-based ensemble for pB shown in Figure 6a, and it contains four β-strands, βI, βIV, βV and βVI and one α-helix (α5). These results are summarized in Table 3, together with the secondary structure elements identified in the NOE-based ensembles for the pG and pB states of Δ25PYP.

There are some small differences in the length of the β-strands between the NOE-based and chemical shift-based structures. However, these are minor and within the variation observed within each ensemble separately, especially when comparing the NOE based-ensemble with the chemical shift based-ensemble obtained including dihedral angle restraints from secondary chemical shifts. Although, as indicated by DSSP (Kabsch, 1983), the second strand, βII, is missing in both chemical shift based-ensembles (see Table 3), the backbone dihedral angle values of the corresponding residues indicate a high tendency towards this conformation. A more pronounced difference is found in the definition of the α5 helix: this helix is perfectly defined (as in the pG state) in the chemical shift-based structures while its N-terminus is lost in the NOE-based pB state structure. However, if we consider the deviations of $^{13}C^{\alpha}$ chemical shifts from their random coil values, for this particular stretch of amino acids (residue number: 75–86), there is a weak tendency for those to occur in an α-helical conformation($\Delta\sigma_{H\alpha}0$) (see Figure 7). A similar observation has been previously reported for the case of Apocytochrome b_{562} (D Amelio et al., 2002), where the chemical shifts suggested a higher helical content than what was observed in the NOE-based structure. The lack of NOEs can be explained by a lower density in the spectra caused by fast exchange between helical and non-helical conformations.

For comparison purpose, we have used the structure closest to the mean for the pG NMR ensemble, as the reference structure, in order to calculate the RMSD for those secondary structure elements found in the ground state. The backbone RMSDs from the reference pG state for still existing secondary elements are within the same range in the pB NOE based- and chemical shift based-ensembles. They are, however, all larger than with the pG NMR ensemble, as expected for more flexible/disordered states. The α3 helix is absent in all calculated pB state ensembles, independently of the protocol and data used to calculate them.

Conclusion

We have described a protocol that allows the description of partially unfolded states of proteins from chemical shift information only provided

Table 3. Comparison among all the available structures for Δ25PYP

Parameters		NMR pG	NOE-based pB	Chemical shift-based pB (CACB)	Chemical shift-based pB (+dihedrals)
Secondary structure elements[a]	βI	29–33	30–34	30–33	30–33
	βII	39–41	39–43		
	α3	44–49			
	α5	76–85	79–85	76–85	76–85
	βIV	89–96	90–95	90–92	90–96
	βV	103–112	104–111	107–112	103–111
	βVI	117–124	117–123	117–122	117–123
Rmsd (Å) from reference structurc[b,c]	*Backbone*	0.7 ± 0.2	2.2 ± 0.2	1.8 ± 0.6	0.9 ± 0.4
	Heavy	1.5 ± 0.4	3.2 ± 0.3	3.2 ± 0.7	2.3 ± 0.4
	α5	1.4 ± 0.4	3.5 ± 0.3	3.0 ± 0.7	2.1 ± 0.4

[a] The program DSSP was used in determination of the secondary elements. Only those conserved in at least 50% of the structures in the ensemble, and predicted strictly either like α-helix ("H") or extended strand ("E") were included.

[b] The structure closest to the mean from the NMR ensemble for pG was considered as the reference structure.

[c] Only secondary elements of the pG state (excluding α3).

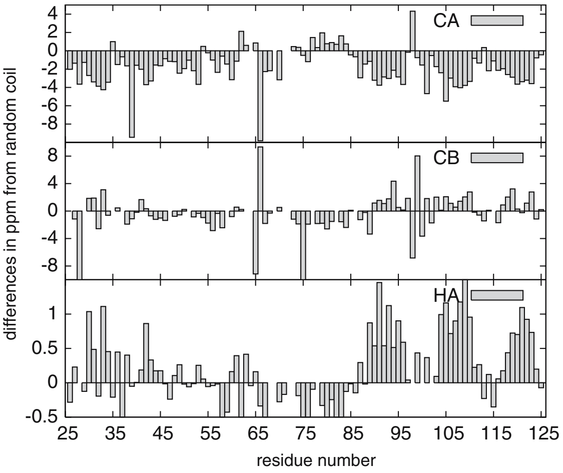

Figure 7. Chemical shift deviations from the random coil for the pB state of Δ25PYP.

the native state 3D structure is available. The main idea is based on the use of native-contact restraints derived from chemical shift information obtained by comparing native and partially unfolded state HSQC spectra. This approach allows the modeling of partially unfolded states, when classical NOE-based approaches fail. Its feasibility was demonstrated first with synthetic data for lysozyme. This approach was then validated with real NMR data by modeling the partially unfolded signaling state (pB) of a truncated form of the PYP. The resulting chemical shift based-ensembles are in good agreement with the reference NOE-based structure demonstrating that our protocol to describe non-native states is robust. We have also shown that, if available, methyl group chemical shift information can also be used to define native-contact restraints and results in a better definition of the core and improved packing of the protein under study.

Finally, while in our approach chemical shifts provide mainly information on the residual, folded part of the system, additional information for the unfolded regions could be derived from residual dipolar coupling and/or relaxation data if available to provide a more complete picture of the partially unfolded state. It has been already shown that they can be a powerful tool as sensitive probes of conformational changes and residual structure in the unfolded states (Shortle and Ackerman, 2001; Bertoncini et al., 2005). Their application in the case of short-lived partially unfolded states of photoactive systems such as PYP might however be limited.

Acknowledgements

Dr. Fuentes acknowledges the Fundacion Areces (Spain) and the European Community (6th Framework Marie Curie grant Meif-ct-2003-501495) for financial support.

References

Artymiuk, P.J., Blake, C.C.F., Rice, D.W. and Wilson, K.S. (1982) *Acta Crystallogr. B*, **38**, 778 .

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J. (1981) In *Intermolecular Forces* Pullman, B. (Ed.), Reidel Publishing Company, Dordrecht, pp. 331–342.

Bernard, C., Houben K., Derix, N., Marks, D., van der Horst, M., Hellingwerf, K., Boelens, R., Kaptein, R. and van Nuland, N. (2005) *Structure*, **13**, 953–962.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rogers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Bertoncini, C.W., Jung, Y.S., Fernandez, C.O., Hoyer, W., Griesinger, C., Jovin, T.M. and Zweckstetter, M. (2005) *Proc. Natl. Acad. Sci. USA*, **102**, 1430–1435.

Best, R.B. and Vendruscolo, M. (2004) *J. Am. Chem. Soc.*, **126**, 8090–8091.

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1994) *J. Biomol. NMR*, **4**, 143–149.

Bonvin, A.M.J.J. and Brunger, A.T. (1995) *J. Mol. Biol.*, **250**, 80–93.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr. D*, **54**, 905–921.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J.S., Taddei, N., Ramponi, G., Dobson, C.M. and Stefani, M. (2002) *Nature*, **416**, 507–511.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

D Amelio, N., Bonvin, A.M.J.J., Czisch, M., Barker, P. and Kaptein, R. (2002) *Biochemistry*, **41**, 5505–5514.

Daggett, V. (2000) *Curr. Opin. Struct. Biol.*, **10**, 160–164.

Daggett, V. (2002) *Accounts Chem. Res.*, **35**, 422–429.

Daura, X.M., Mark, A.E. and van Gunsteren, W.F. (1998) *J. Comput. Chem.*, **19**, 535–547.

Dedmon, M.M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. and Dobson, C.M. (2005) *J. Am. Chem. Soc.*, **127**, 476–477.

Dill, K.A. and Shortle, D. (1991) *Annu. Rev. Biochem.*, **60**, 795–825.

Dyson, H.J. and Wright, P.E. (1996) *Annu. Rev. Phys. Chem.*, **47**, 369–396.

Farmer, B.T. (1996) *Nat. Struct. Biol.*, **3**, 995–997.

Fersht, A.R. and Daggett, V. (2002) *Cell*, **108**, 573–582.

Fersht, A.R., Leatherbarrow, R.J. and Wells, T.N.C. (1986) *Nature*, **322**, 284–286.

Fersht, A.R., Matouschek, A. and Serrano, L. (1992) *J. Mol. Biol.*, **224**, 771–782.

Groenhof, G., Lensink, M.F., Berendsen, H.J.C., Snijders, J.G. and Mark, A.E. (2002) *Proteins*, **48**, 202–211.

Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) *J. Comput. Chem.*, **18**, 1463–1472.

Hsu, S.-T.D. and Bonvin, A.M.J.J. (2004) *Proteins*, **55**, 582–593.

Jorgensen, W.L. and Tirado-Rives, J. (1988) *J. Am. Chem. Soc.*, **110**, 1657–1666.

Kabsch, W.A.S.C. (1983) *Biopolymers*, **22**, 2577–2637.

Korzhnev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M. and Kay, L.E. (2004) *Nature*, **430**, 586–590.

Laskowski, R.A., Macarthur, M.W., Moss, D.S. 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.

Lindahl, E., Hess, B. and van der Spoel, D. (2001) *J. Mol. Model*, **7**, 306–317.

Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C.M., Poulsen, F.M. and Vendruscolo, M. (2004) *J. Am. Chem. Soc.*, **126**, 3291–3299.

Linge, J.P., O'Donoghue, S.I. and Nilges, M. (2001) *Methods Enzymol.*, **339**, 71–90.

Linge, J.P., Williams, M.A., Spronk, C.A.E.M., Bonvin, A.M.J.J. and Nilges, M. (2003) *Proteins*, **50**, 496–506.

MacKenzie, K.R., Prestegard, J.H. and Engelman, D.M. (1996) *J. Biomol. NMR*, **7**, 256–260.

Matouschek, A., Kellis, J.T., Serrano, L. and Fersht, A.R. (1989) *Nature*, **340**, 122–126.

Meiler, J. (2003) *J. Biomol. NMR*, **26**, 25–37.

Mierke, D.F., Kurz, M. and Kessler, H. (1994) *J. Am. Chem. Soc*, **116**, 1042–1049.

Mulder, F.A.A., Schipper, D., Bott, R. and Boelens, R. (1999) *J. Mol. Biol.*, **292**, 111–123.

Neal, S., Nip, A.M., Zhang, H.Y. and Wishart, D.S. (2003) *J. Biomol. NMR*, **26**, 215–240.

Nederveen, A.J., Doreleijers, J.F., Vranken, W.F., Miller, Z., Spronk, C.A.E.M., Nabuurs, S.B., Güntert, P., Livny, M., Markley, J.L., Nilges, M., Ulrich, E.L., Kaptein, R. and Bonvin, A.M.J.J. (2005) *Proteins*, **59**, 662–672.

Paci, E., Clarke, J., Steward, A., Vendruscolo, M. and Karplus, M. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 394–399.

Paci, E., Vendruscolo, M., Dobson, C.M. and Karplus, M. (2002) *J. Mol. Biol.*, **324**, 151–163.

Ptitsyn, O.B. (1995) *Adv. Protein Chem.*, **47**, 83–229.

Rubinstenn, G., Vuister, G.W., Mulder, F.A.A., Duex, P.E., Boelens, R., Hellingwerf, K.J. and Kaptein, R. (1998) *Nat. Struct. Biol.*, **5**, 568–570.

Schwartz, M.P., Huang, S.H. and Matouschek, A. (1999) *J. Biol. Chem.*, **274**, 12759–12764.

Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J.L. (1991) *J. Biomol. NMR*, **1**, 217–236.

Shortle, D. and Ackerman, M.S. (2001) *Science*, **293**, 487 489.

Shortle, D.R. (1996) *Curr. Opin. Struct. Biol.*, **6**, 24–30.

Thomas, P.J., Qu, B.H. and Pedersen, P.L. (1995) *Trends Biochem. Sci.*, **20**, 456–459.

Tironi, I.G., Sperb, R., Smith, P.E. and Vangunsteren, W.F. (1995) *J. Chem. Phys.*, **102**, 5451–5459.

van der Horst, M.A., van Stokkum, I.H., Crielaard, W. and Hellingwerf, K.J. (2001) *Febs. Lett.*, **497**, 26–30.

van Gunsteren, W.F. and Berendsen, H.J.C. (1990) *Angew Chem. Int. Edit.*, **29**, 992–1023.

Vendruscolo, M., Paci, E., Dobson, C.M. and Karplus, M. (2001) *Nature*, **409**, 641–645.

Vreede, J., van der Horst, M.A., Hellingwerf, K.J., Crielaard, W. and van Aalten, D.M.F. (2003) *J. Biol. Chem.*, **278**, 18434–18439.

Vriend, G. (1990) *J. Mol. Graphics*, **8**, 52–56.

Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.*, **239**, 363–392.

Wright, P.E. and Dyson, H.J. (1999) *J. Mol. Biol.*, **293**, 321–331.