

NMR in the SPINE Structural Proteomics project

E. AB,^a A. R. Atkinson,^b L. Banci,^c
I. Bertini,^c S. Ciofi-Baffoni,^c
K. Brunner,^d T. Diercks,^a
V. Dötsch,^e F. Engelke,^f
G. E. Folkers,^a C. Griesinger,^g
W. Gronwald,^d U. Günther,^e
M. Habeck,^h R. N. de Jong,^a
H. R. Kalbitzer,^d B. Kieffer,^h
B. R. Leeflang,^a S. Loss,ⁱ
C. Luchinat,^c T. Marquardsen,^f
D. Moskau,ⁱ K.-P. Neidig,^f
M. Nilges,^h M. Piccioli,^c
R. Pierattelli,^c W. Rieping,^h
T. Schippmann,^f H. Schwalbe,^e
G. Travé,^d J. Trenner,^d
J. Wöhnert,^d M. Zweckstetter^g
and R. Kaptein^{a*}

^aBijvoet Center for Biomolecular Research, NMR Spectroscopy, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, ^bInstitut de Génétique et de Biologie Moléculaire et Cellulaire, 1 Rue Laurent Fries, BP 163, 67404 Illkirch CEDEX, France, ^cCIRMMP, CERM, University of Florence, Via Sacconi 6, 50019 Sesto Fiorentino, Italy, ^dDepartment of Biophysics and Physical Biochemistry, University of Regensburg, Germany, ^eCenter for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe-University Frankfurt, Germany, ^fBruker-Biospin GmbH, Rheinstetten, Germany, ^gMax-Planck-Institut für Biophysikalische Chemie, Göttingen, Germany, ^hBioinformatique Structurale, Institut Pasteur, 25–28 Rue du Dr Roux, 75724 Paris, France, and ⁱBruker-Biospin AG, Fällanden, Switzerland

Correspondence e-mail: r.kaptein@chem.uu.nl

This paper describes the developments, role and contributions of the NMR spectroscopy groups in the Structural Proteomics In Europe (SPINE) consortium. Focusing on the development of high-throughput (HTP) pipelines for NMR structure determinations of proteins, all aspects from sample preparation, data acquisition, data processing, data analysis to structure determination have been improved with respect to sensitivity, automation, speed, robustness and validation. Specific highlights are *protonless* ¹³C-direct detection methods and inferential structure determinations (ISD). In addition to technological improvements, these methods have been applied to deliver over 60 NMR structures of proteins, among which are five that failed to crystallize. The inclusion of NMR spectroscopy in structural proteomics pipelines improves the success rate for protein structure determinations.

1. Introduction

The objectives of the SPINE project are twofold: firstly to improve the methodology for high-throughput (HTP) structure determination of proteins and secondly to solve structures of proteins relevant for human health and disease. These structures can be exploited for functional characterization and might serve as a starting point for drug discovery.

Recent developments in both hardware and software have enabled NMR to be a serious player in structural genomics. However, structure determination by NMR is still slower than by X-ray crystallography (where high-quality crystals are available) and is not yet automated to the same extent. The NMR groups involved in the SPINE consortium have identified a number of bottlenecks in this process and have contributed with a number of technological and methodological improvements in both the efficiency and the quality of protein structure determination.

One of the most challenging tasks for both crystallography and NMR groups is the preparation of samples suitable for structure determination. For NMR, size limitations (<20 kDa for HTP structure determination) compel NMR laboratories to work either with small proteins or with protein domains. Owing to the poor success rate for human proteins in prokaryotic expression systems and the demanding sample requirements for NMR, many targets have to be screened. This requires an automated and efficient cloning and expression system that allows the screening of multiple expression conditions. These small-scale expression-screening protocols serve as a good predictor for subsequent large-scale production (Folkers *et al.*, 2004). The proteins are subsequently used for screening of solubility, monodispersity and foldedness by ¹⁵N-HSQC spectroscopy. To facilitate HTP screening of

Received 13 December 2005

Accepted 14 August 2006

Table 1

Flow-cell parameters and basic NMR performance data.

	60 μl cell at 600 MHz	120 μl cell at 600 MHz	175 μl cell at 500 MHz
Outer diameter (mm)	3	4	5
NMR active volume (μl)	60	120	175
Variable temperature range (K)	263–353	263–353	293–303 [†]
Maximum pressure (MPa)	3	3	0.5 [†]
¹ H NMR line-shape hump [‡] (0.55/1.1% amplitude) (Hz)	7/14	7/14	10/20
¹ H FWHM [‡] (Hz)	0.7	0.7	0.7
¹ H signal-to-noise ratio [§]	250:1	400:1	350:1

[†] The reduced temperature range and restricted pressure limit for 175 μl cells as built here originates from the fact that Kel-F inserts were glued into the glass cylinders using a special epoxy adhesive. Manufacturing such a cell as the 60 or 120 μl cells, namely completely of glass, would also improve these specifications. [‡] Sample to determine ¹H line shape: CHCl_3 in acetone-*d*₆. [§] Sample to determine ¹H sensitivity: anomeric proton NMR line in 0.2 mM sucrose in D₂O, single-scan acquisition.

sample conditions, a novel flow cell to be used in cryoprobes has been developed in association with the industrial partner Bruker-Biospin.

NMR data-collection times have been further reduced by enhanced sensitivity obtained from unused proton polarization (Diercks & Orekhov, 2005; Diercks *et al.*, 2006). The use of residual dipolar couplings (RDCs) combined with more conventional structural parameters provides an interesting possibility for semi-automated structure determination (Bax, 2003), as well as for obtaining structural information on systems sampling different conformational substates (Bertini, Del Bianco *et al.*, 2004). The use of *protonless* NMR using direct detection of heteronuclei such as ¹⁵N and ¹³C has been pioneered by the Florence group within SPINE. This circumvents the problem of fast proton relaxation as occurs in paramagnetic proteins as well as in large proteins.

The most prominent obstacle for HTP structure determination by NMR is the labour-intensive spectral analysis including backbone and side-chain assignment. Novel data-analysis protocols have been implemented in various software tools. A final highlight is the inferential structure-determination (ISD) method developed by the Pasteur group (Rieping *et al.*, 2005). This probabilistic method is much more objective than conventional structure-determination protocols and provides greater structural quality and more realistic figures of merit.

While the primary focus was on methods development, the SPINE NMR groups have collectively solved over 60 protein structures to date. The purpose was to reconcile the HTP approach to structure determination with research aiming at understanding protein functions within the cell. Various human targets were selected from proteins or domains involved in nuclear events, including transcription, ubiquitination and DNA repair. Furthermore, different classes of targets are implemented with the goal of providing new insights into the role of metal ions in prokaryotes and eukaryotes by elucidating the structures of metal-binding proteins as well as their network of interactions with other proteins and, possibly, with nucleic acids (Banci & Rosato, 2003). The criteria for target selection are thus largely determined by the absence of a homologous structure, the putative function of the protein and by the possible involvement in human disease.

2. Technological developments

2.1. NMR data collection

2.1.1. Flow cells for NMR cryogenic probes. The hardware developments within the SPINE project have been focused on the creation of a liquid-handling system for high-throughput recording of NMR spectra (Utrecht Group with Bruker Biospin GmbH). The combination of flowthrough technology with NMR probes based on commercially available cryogenic probes pursues two goals: (i) to provide the high sensitivity of a cryoprobe in a flowthrough cell and simultaneously (ii) to achieve the maximum flexibility and handling in an NMR system by providing a flow cell that can be inserted into and removed from a cryoprobe almost as easily as a conventional NMR tube.

The innovative aspects of these developments deal with overcoming a recognized problem with dedicated cryo-flow probes (cryoprobes with a rigidly built-in flow cell): blocking of the inlet or outlet capillaries would require a complete warm up of the cryoprobe before the problem can be accessed and fixed, a time-consuming and costly operation.

Two types of flow cells have been developed. Initially, a 5 mm flow cell with 175 μl active volume was designed where the inlet and outlet capillaries both come from the top of the magnet. The construction of the cells is shown in Fig. 1. The cell is inserted, with tubing, from the top into a holder, which is basically the spinner body familiar from conventional sample tubes. A second design was made for flow cells with smaller active volumes of 120 and 60 μl . These cells have a straight design and are made entirely from glass. Once inserted into the cryoprobe, the inlet capillary comes from the bottom of the probe and the outlet capillary goes to the top of the magnet. Basic flow-cell data including some NMR performance results are summarized in Table 1.

2.1.2. Faster NMR data acquisition. In many NMR experiments, only the polarization of a limited subset of all protons is converted into observable coherence. Yet, recovery of the unused polarization can speed up NMR measurements considerably. Faster repetition of the same experiment can be achieved by accelerating the re-equilibration of the selected proton subset (Pervushin *et al.*, 2002). Alternatively, two experiments can be compacted into the same measurement time (Diercks & Orekhov, 2005).

Table 2

Selection of the *protonless* NMR experiments available.

The correlations observed in each experiment and the transfer pathways of the sequence are also reported. Alternative ways to obtain the removal of the *J*-splitting in the acquisition dimension have been implemented in the sequences, either based on the inclusion of spin-state selection or with band-selective homo-decoupling.

Experiment	Correlations observed	Transfer pathway	Reference
Backbone assignment			
CACO	$C_i^\alpha - C_i'$	$F1(C_i^\alpha, t_1) \rightarrow F1(CO, t_2)$ or $F1(CO) \rightarrow F1(C_i^\alpha, \text{HMOC}, t_1) \rightarrow F1(CO, t_2)$	<i>a, b</i>
COCA	$C_i^\alpha - C_i'$	$F1(C_i^\alpha, t_1) \rightarrow F1(CO, t_2)$	<i>a, c</i>
CON	$N_i - C_{i-1}'$	$F1(CO) \rightarrow F3(N, t_1) \rightarrow F1(CO, t_2)$	<i>c, d</i>
CAN	$C_i^\alpha - N_i - C_{i-1}'; C_{i-1}^\alpha - N_i - C_{i-1}'$	$F1(C_i^\alpha) \rightarrow F3(N, \text{HMOC}, t_1) \rightarrow F1(C_i^\alpha, t_2)$ or $F1(C_i^\alpha) \rightarrow F3(N, t_1) \rightarrow F1(C_i^\alpha, t_2)$	<i>e</i>
CANCO sel	$C_{i-1}^\alpha - N_i - C_{i-1}'$	$F1(C_i^\alpha, t_1) \rightarrow F3(N, t_2) \rightarrow F1(CO, t_3)$	<i>f</i>
CANCO	$C_i^\alpha - N_i - C_{i-1}'; C_{i-1}^\alpha - N_i - C_{i-1}'$	$F1(C_i^\alpha, t_1) \rightarrow F3(N, t_2) \rightarrow F1(CO, t_3)$	<i>g</i>
Side-chain assignment			
CBCACO	$C_i^\beta - C_i^\alpha - C_i', C_i^\alpha - C_i^\alpha - C_i'$	$F1(C_i^{\alpha/\beta}, t_1) \rightarrow F1(C_i^\alpha, t_2) \rightarrow F1(CO, t_3)$	<i>a</i>
CCCO	$C_i^{\text{ali}} - C_i^\alpha - C_i', C_i^\alpha - C_i^\alpha - C_i'$	$F1(C_i^{\text{ali}}, t_1) \rightarrow F1(C_i^\alpha, t_2) \rightarrow F1(CO, t_3)$	<i>d</i>
CCCON	$C_i^{\text{ali}} - C_i^\alpha - N_{i+1}, C_i^\alpha - C_i^\alpha - N_{i+1}$	$F1(C_i^{\text{ali}}, t_1) \rightarrow F1(C_i^\alpha) \rightarrow F1(CO) \rightarrow F3(N, t_2) \rightarrow F1(CO, t_3)$	<i>d</i>
CBCACON	$C_i^\beta - C_i^\alpha - N_{i+1}, C_i^\alpha - C_i^\alpha - N_{i+1}$	$F1(C_i^{\alpha/\beta}, t_1) \rightarrow F1(C_i^\alpha) \rightarrow F1(CO) \rightarrow F3(N, t_2) \rightarrow F1(CO, t_3)$	<i>d</i>
CGCB	$C_i^\gamma - C_i^\beta$	$F1(C_i^\gamma, t_1) \rightarrow F1(C_i^\beta, t_2)$	<i>h</i>
Multipurpose sequences			
CC-COSY	$C_i^\alpha - C_i', C_i^\alpha - C_i^\beta, C_i - C_i$	$F1(C, t_1) \rightarrow F1(C, t_2)$	<i>i, j, k</i>
C_i^α -TOCSY	$C_i^{\text{ali}} - C_i^\alpha$	$F1(C_i^{\text{ali}}, t_1) \rightarrow F1(C_i^\alpha, t_3)$	<i>a, l</i>
CACO-NOESY	$C_i^\alpha - C_i'$	$F1(C_i^\alpha, t_1) \rightarrow F1(CO, t_2)$	<i>m, n</i>

References: (a) Bermel, Bertini, Duma *et al.* (2005); (b) Bertini, Jiménez & Piccioli (2005); (c) Kostic *et al.* (2002); (d) Bermel, Bertini, Felli *et al.* (2005); (e) Bermel, Bertini, Felli, Kümmerle *et al.* (2006); (f) Bermel, Bertini, Felli, Piccioli *et al.* (2006); (g) Bertini, Duma *et al.* (2004); (h) Babini *et al.* (2005); (i) Bertini *et al.* (2001); (j) Bertini, Jiménez, Piccioli *et al.* (2005); (k) Machonkin *et al.* (2002); (l) Eletsky *et al.* (2003); (m) Bertini, Felli, Kümmerle, Luchinat *et al.* (2004); (n) Bertini, Felli, Kümmerle, Moskau *et al.* (2004).

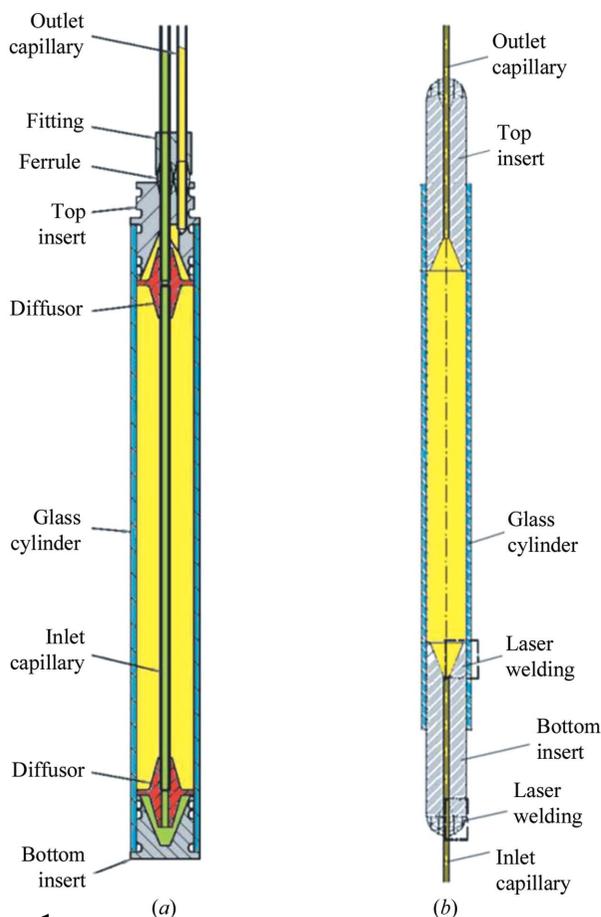


Figure 1
(a) 5 mm (175 μ l) flow cell with inlet and outlet capillaries both leading to the top of the NMR magnet when the cell is inserted into a cryo-probe.
(b) 4 mm (120 μ l) flow cell with inlet capillary arriving from the bottom and outlet capillary leading to the top of the magnet.

Novel schemes for polarization recovery based on HSQC-type or TROSY-type (Pervushin *et al.*, 1997) transfers have been developed by the Utrecht group: the latter is mostly employed for deuterated proteins. The use of recovered H^N anti-TROSY polarization in queued TROSY (Diercks & Orekhov, 2005) has been introduced, affording up to 50% time saving for recording experiment pairs. In contrast, HSQC-type transfer is the method of choice for standard non-deuterated proteins. Our novel extended flip-back (efb) schemes (Diercks *et al.*, 2006) can recover more than 50% of the previously wasted non-amide proton polarization. This fast pulsing regime afforded a sensitivity enhancement by more than 40% from accelerated recovery of H^N polarization. The next goal is to combine these concepts for physically accelerating NMR experiments by recovery and constructive use of unused proton polarization with schemes for faster sampling of the frequency space, such as time-shared continuous sampling with subsequent G-matrix Fourier transform (GFT; Szyperki *et al.*, 2002).

2.1.3. ^{13}C direct-detection *protonless* NMR. The interest in ^{13}C direct detection arises from the fact that the contributions to relaxation owing to a paramagnetic centre on ^{13}C with respect to 1H is smaller by a factor of 16 owing to the different gyromagnetic ratios of the two nuclei. Therefore, NMR experiments that avoid 1H transfers can improve signal detection in paramagnetic systems. Furthermore, in many other cases 1H may be difficult to detect and ^{13}C NMR can be helpful, *e.g.* for protein regions characterized by high conformational/chemical exchange. Such regions, for which the backbone NHs cannot be detected and the standard sequence-specific procedures for assignment fail, are often important for protein function. Another example is provided by unfolded

systems, where chemical shift dispersion of ^1H resonances and amide proton exchange may be unfavourable (Bermel, Bertini, Felli, Lee *et al.*, 2006).

Starting from the available building blocks for triple-resonance experiments (Sattler *et al.*, 1999; Kay *et al.*, 1990), it is possible to implement a set of pulse sequences enabling the complete assignment of a protein without involving ^1H transfers. In Table 2, a selection of the sequences currently available is reported, highlighting the correlations observed and the transfer pathway providing the correlations.

A peculiarity of ^{13}C direct-detection experiments is the presence of signal splitting in the acquisition dimension owing to the large J_{CC} coupling. For a CO signal, the main coupling is with the C^α , giving a relatively constant splitting of some 55 Hz. The uniformity of the splitting value allows the use of a 'trick' for its removal based on spin-state-selective methods (Ottiger *et al.*, 1998; Andersson *et al.*, 1998; Duma, Hediger, Brutscher *et al.*, 2003; Bermel, Bertini, Felli *et al.*, 2005). Two FIDs for each increment of a 2D/3D experiment are recorded, one for the anti-phase and one for the in-phase components; each pair of FIDs is then combined to separate the two multiplet components. These are then shifted to the centre of the original multiplet (by $J_{\text{C}^\alpha}/2$ Hz) and summed to obtain a singlet (Bermel, Bertini, Felli *et al.*, 2005; Nielsen *et al.*, 1995). In the case of spectra based on acquisition of C^α , a nucleus having two large couplings [with the CO (55 Hz) and with the C^β (35 Hz) nuclei], a double in-phase/anti-phase scheme can be implemented to remove the double splitting (Bermel, Bertini, Duma *et al.*, 2005; Duma, Hediger, Brutscher *et al.*, 2003; Duma, Hediger, Lesage *et al.*, 2003). The removal of the splitting can be also obtained, with some loss in the signal-to-noise ratio, by band-selective homodecoupling (Bermel *et al.*, 2003; Vögeli *et al.*, 2005).

These novel experiments have been tested in the Florence group and have been found to be useful for several systems, either paramagnetic (Arnesano, Banci, Bertini, Felli *et al.*, 2003; Bermel *et al.*, 2003; Babini, Bertini, Capozzi, Felli *et al.*, 2004) or diamagnetic (Bertini, Duma *et al.*, 2004; Bertini, Felli, Kümmerle, Luchinat *et al.*, 2004; Bertini, Felli, Kümmerle, Moskau *et al.*, 2004; Arnesano, Balatri *et al.*, 2005; Arnesano,

Banci, Bertini, Fantoni *et al.*, 2005; Bermel, Bertini, Duma *et al.*, 2005; Bermel, Bertini, Felli *et al.*, 2005; Bermel, Bertini, Felli, Piccioli *et al.*, 2006).

2.1.4. Specific labelling applied to the assignment of a membrane protein. Membrane proteins represent one of the greatest challenges in the area of structural biology based on the fact that they are hard to express and to purify and difficult to analyze by high-resolution structural methods. The large amount of α -helical secondary structures in membrane proteins that have a narrower chemical shift dispersion compared with β -sheets induces signal overlap. The overlap problem is further aggravated by the often large size of the proteins and by the fact that the proteins are solubilized in detergent micelles, adding to the molecular weight of the protein/micelle particles and resulting in broader line width. Combined, these disadvantages pose a challenge to the backbone-assignment process. In order to overcome this problem, an assignment strategy that is based on the use of an *in vitro* transcription/translation system has been developed by the University of Frankfurt (SPINE subcontractor; Klammt *et al.*, 2005; Trbovic *et al.*, 2005). The method is based on using non-selective standard triple-resonance experiments in order to obtain as many assignments as possible during the first stage of the assignment process. In our experience, this procedure provides approximately 50% of the backbone assignment. The rest of the assignment remains ambiguous owing to overlap and missing peaks. In order to obtain more assignments, a combinatorial approach is used to label different protein samples with ^{15}N - and ^{13}C -labelled amino acids. From the combination of HSQC and 2D-HNCO spectra, the sequence specific assignment for certain amino acids can be obtained which act as anchor points for further sequence-specific assignments. With this method, 85% of the backbone assignment of the 24 kDa bacterial membrane protein TehA was obtained. The labelling scheme used for this combinatorial approach was optimized by a computer program that takes into account the amino-acid sequence as well as the amino-acid stretches that have so far been assigned. This protocol maximizes the information that is obtained in each sample, thus minimizing the number of samples needed for the assignment.

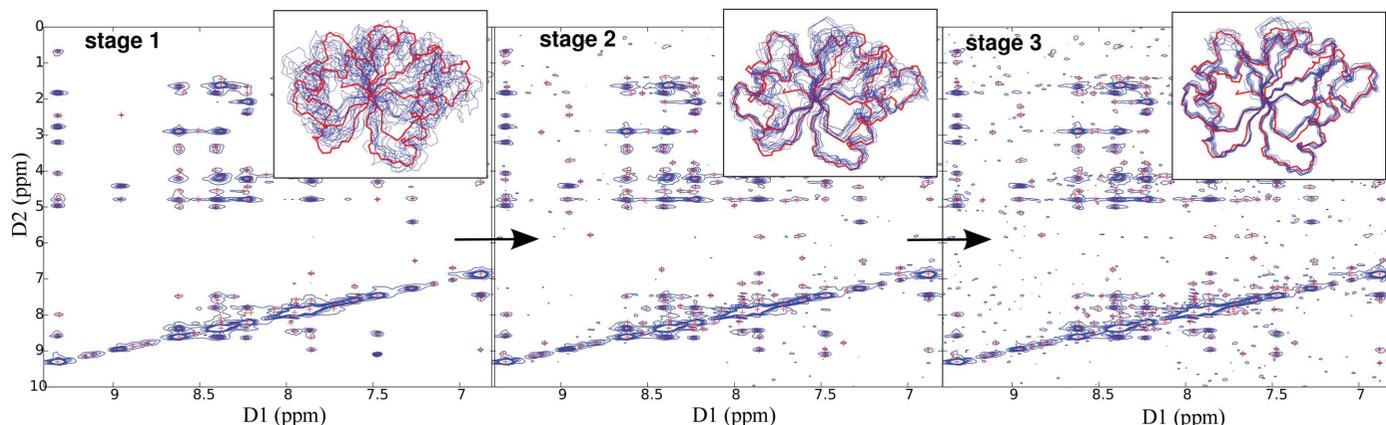


Figure 2
Multi-stage process of automated peak identification and structure determination using incremental wavelet denoising.

Table 3

Benchmarking of the homology mapping.

Protein	Protein model was based on (PDB code)	Sequence identity (%)	R.m.s.d. (Å)	Success rate for stretch of two residues (%)		
				Struct.	Model	Ext. chain
<i>Thermotoga maritima</i> L11 CTD	<i>Bacillus stearothermophilus</i> L11 CTD (1hc8)	68	1.58	79	79	25.8
Lysozyme HEW	Lysozyme human (1jsf)	60	2.66	81.1	73.6	17.3
Integrin	LFA (1dgg)	26	4.50	58.8	47.8	19.8
Trigger factor	FKBP (1fkf)	22	3.11	58.2	44.2	31.0

2.2. NMR data analysis

2.2.1. Application of wavelet-denoised NOESY spectra.

The generation of reliable NOESY peak lists is a major time-consuming step of protein NMR structure determination. A new software package has been developed by the SPINE subcontractor University of Frankfurt (Dancea & Günther, 2005) which couples the automated identification of peaks to automated structure determination by denoising NMR spectra using wavelet methods. The core concept of this procedure is the generation of incremental peak lists by applying different wavelet-denoising procedures, which lead to spectra with different noise content (Fig. 2). The first structure is calculated using a peak list which has been strongly denoised and lacks some 20–30% of peaks but has a high reliability that the peaks are real signals. A second peak list is generated with more moderate denoising, yielding a peak list which is 90% complete but has some residual noise signals included. In a third round all signals are included. The wavelet filter is further enhanced by spin-system network anchoring and symmetry considerations within and between various NOESY peak lists. To optimize the overall process, a highly reliable peak-picking algorithm has been designed to provide reliable peak integrals even for overlapping signals. These algorithms were embedded in the context of the *ARIA* software (Linge *et al.*, 2003) for automated NOE assignment and structure determination. Fig. 3 shows as an example NOESY spectra and structures of polysulfide-sulfur transferase (Lin *et al.*, 2004) for subsequent stages of structure determination. The final structure was similar to a manually identified structure of the same protein and was of similar quality (r.m.s.d. = 0.85).

2.2.2. Assignment strategy using structures of homologous proteins. The NMR resonance-assignment process can be accelerated if chemical shift prediction from homologous proteins is performed (contribution by University of Frankfurt). The difference between experimental and predicted shifts can be minimized using a target function

$$\text{r.m.s.d.}(A) = \left\{ \left[\sum_{k=A}^{A+l-1} \sum_{i=1}^S (\hat{I}_{k,i}^{\text{NMR}} - \hat{I}_{k,i}^{\text{calc}})^2 / l \right]^{1/2} \right\},$$

which computes the deviation between experimental and predicted shifts for amino acid *l* at any possible location *A* within the sequence. An analysis of the success rates for models that were obtained using *SWISSMODEL* (Kopp & Schwede, 2004) shows significantly better results than those obtained for an extended chain, even in the case of low

sequence identities and correspondingly high r.m.s.d. values between the modelled and experimental structures (Table 3).

2.2.3. AUREMOL. The software package *AUREMOL*, developed at the University of Regensburg in cooperation with Bruker-Biospin, aims to computerize the steps from assignment to structural models (Gronwald & Kalbitzer, 2004; <http://www.auremol.de>). It employs a top-down ‘molecule-centred approach’ (MCA) in which a trial three-dimensional starting structure is iteratively refined until it fits the experimental data as well as possible using a minimal set of data. The goal is the best possible structure but not necessarily the most complete resonance assignment. Structure validation is an important step which is performed at intermediate stages of evaluation and is mandatory at the end of the procedure.

AUREMOL has been enhanced in the SPINE project. In order to improve the quality of the input data, a new and promising ICA (independent component analysis) based method for automated baseline correction and artefact suppression in *n*-dimensional spectra is under development (Stadlhanner *et al.*, 2003; Böhm *et al.*, 2005). Furthermore, the program *RELAX-JT2* (Ried *et al.*, 2004) was developed to allow the proper simulation of line shapes by the integrated calculation of *T*₂ times and multiplet structures caused by *J*-couplings. Moreover, the effects of relaxation mediated by chemical shift anisotropy are taken into account. This allows an improved simulation of multidimensional NOESY spectra that are used for assignment and structure-validation purposes within *AUREMOL*.

Based on heteronuclear triple-resonance spectra, the backbone resonance line assignment can be automatically obtained by sequentially linking pseudo-residues and mapping them on the known primary sequence based on the relationship between amino-acid type and ¹³C^α and ¹³C^β shifts. *AUREMOL* can now automatically perform most of the stages required for protein structure determination from NMR data.

2.2.4. Structure determination with missing resonance assignments: PROXY residue. SPINE (Utrecht Group) has developed and implemented the new concept of PROXY atoms (AB *et al.*, 2006), which allows NOE-based structure calculation protocols such as *ARIA* (Linge *et al.*, 2003) and *CANDID* (Herrmann *et al.*, 2002) to use the information present in the unassigned portion of the NOE spectra which makes them much more robust with regard to missing resonance assignments. The key idea is that unassigned resonances are not ignored during structure calculations, but are instead

represented by dummy constructs, so-called PROXY atoms or residues. NOE-based distance restraints acting on these PROXY atoms are expected to place them close to the atoms corresponding to their correct assignment. Additional information from the chemical shift values or J -coupling-based spectra can be efficiently incorporated in the form of identity restraints, short-range ambiguous distance restraints that carry information about the possible assignment. This approach has been tested with the widely used structure-calculation protocols *ARIA* and *CANDID*. Test cases have shown that the use of PROXIES in *CANDID* allows up to 30% missing resonance assignments before the results become unreliable, whereas for the original *CANDID* implementation this is approximately 10%. In another case, correctly folded structures were obtained using only backbone and C^β/H^β resonance assignments.

2.2.5. Residual dipolar couplings (RDCs) and other non-NOE constraints. RDCs can be obtained using a paramagnetic metal ion as an internal orienting device. Paramagnetic RDC restraints as well as pseudocontact shifts (PCS) have been implemented by the Florence group in *DYANA*, *CYANA* and more recently in *XPLOR-NIH* (Banci *et al.*, 2004; <http://www.postgenomicnmr.net>) and used to refine the NMR structures of several metalloproteins either containing a native paramagnetic metal or specifically substituted with paramagnetic ions (Arnesano, Banci, Bertini, Felli *et al.*, 2003; Arnesano, Banci, Bertini, Mangani *et al.*, 2003; Baig *et al.*, 2004; Babini, Bertini, Capozzi, Del Bianco *et al.*, 2004; Bertini, Del Bianco *et al.*, 2004). A medium-resolution structure of a test protein could also be obtained without NOEs, largely based on chemical shift index and paramagnetic constraints (Barbieri *et al.*, 2004), and a structural model of a four-helix bundle cytochrome (Bertini, Faraone-Mennella *et al.*, 2004) could be effectively validated by the use of a restricted

number of paramagnetic constraints. As PCS and RDC can also be obtained on heteronuclei in *protonless* experiments and heteronuclei are less sensitive to line broadening, paramagnetic constraints are also promising for high-molecular-weight proteins.

In contrast to RDCs generated from external orienting devices, paramagnetic RDCs are intrinsically able to provide information on the relative conformational freedom of different domains in multi-domain proteins where the paramagnetic metal resides in one domain. An approach involving two sets of RDC and PCS constraints obtained from substitution of Tb^{3+} and Tm^{3+} in the N-terminal calcium-binding site II of human calmodulin allowed the description of the conformational space sampled by the C-terminal domain with respect to the N-terminal domain (Bertini, Del Bianco *et al.*, 2004).

In collaboration with B. Imperiali (MIT, USA), the University of Frankfurt has designed short peptides that bind lanthanide ions providing lanthanide-binding tags (LBTs) with significantly improved properties compared with conventional EF-hand motifs. These LBTs have been incorporated into fusion proteins, allowing facile overexpression of a protein containing a minimally invasive versatile protein tag.

The LBT chosen for this study has the sequence YIDTNNNDGWYEGDELLA, which was appended to the N-terminus of human ubiquitin by standard cloning techniques. The peptide's affinity for trivalent lanthanide ions follows a parabolic relationship dependent on the ionic radii across the series, with the tightest apparent dissociation constant occurring for Tb^{3+} ($K_d = 57$ nM). The affinity is reduced moving to the largest (La^{3+} , $K_D = 4$ μ M) and smallest (Lu^{3+} , $K_d = 130$ nM) lanthanide (unpublished results). The K_d of the fusion protein for Tb^{3+} was determined to be 130 nM by fluorescence titration, affirming that the LBT retains its lanthanide affinity in the context of a protein fusion (Wöhnert *et al.*, 2003).

2.3. NMR structure determination

2.3.1. Inferential structure determination (ISD). ISD is an entirely novel and independent program to sample the probability density of a structure given prior information and experimental data (Rieping *et al.*, 2005; Habeck, Nilges *et al.*, 2005a,b; Habeck, Rieping *et al.*, 2005). The SPINE subcontractor at the Pasteur Institute is using ideas from the probabilistic viewpoint implemented into standard structure-calculation approaches (determination of Karplus coefficients from the data; optimal choice of the weight on the experimental data term).

Macromolecular structures calculated from NMR data are not fully determined by experimental data, but

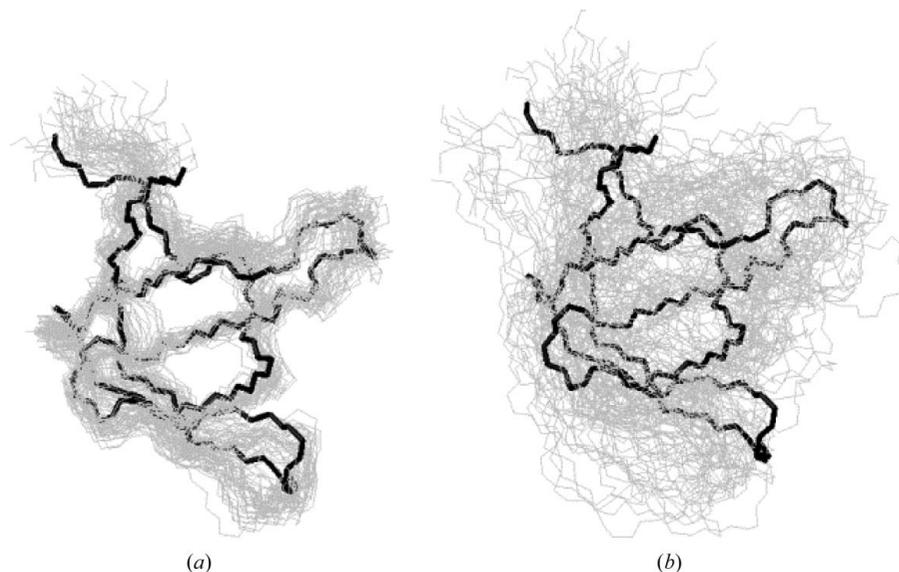


Figure 3

Application of the inferential structure-determination method is shown in (a), whereas (b) illustrates the less well defined protein structure using tradition methods.

depend on subjective choices in data treatment and parameter settings. This makes it difficult to objectively judge the precision of the structures. It is possible to use Bayesian inference to derive a probability distribution that represents the unknown structure and its precision. This probability distribution also determines additional unknowns, such as theory parameters, that previously had to be chosen empirically. Implementation of this approach using Markov chain Monte Carlo techniques provides an objective figure of merit and improves structural quality.

Structure determination has always been considered an optimization problem: a probabilistic approach solves many practical problems and also increases structural quality. The probability distribution that represents the unknown structure and its precision comprises prior assumptions about physical interactions and a likelihood function for the data. Both terms result in a complex posterior distribution, making its simulation particularly difficult. In order to deal with these difficulties, we have combined multiple Markov chain Monte Carlo techniques. Our algorithm is a multi-parameter generalization of Replica-exchange Monte Carlo. The strategy relies on gradual weighing of experimental data and on Tsallis generalized statistics and has demonstrated the effectiveness of the method on NMR data for several folded proteins. In particular, the probabilistic method has advantages for sparse data.

The ISD approach was extended to analyse three-bond scalar coupling constants in an objective and consistent way. The Karplus curve and a Gaussian error law are used to model scalar coupling measurements. By applying Bayes' theorem, a joint posterior density was obtained for all unknowns, *i.e.* the torsion angles, the Karplus parameters and the standard deviation of the Gaussian. Unlike traditional approaches, which require a predetermined reference structure to deter-

mine the Karplus curve, all these unknowns are inferred from scalar coupling data using Markov chain Monte Carlo sampling and analytically derive a probability density that only involves the torsion angles.

The determination of macromolecular structures requires weighting of the experimental evidence relative to the prior physical information. Although the weighting can critically affect structural quality, data are routinely weighted on an empirical basis. At present, cross-validation is the method of choice to determine the best weight. However, using Bayesian inference to determine the weight along with the structure, it is possible to eliminate the weight completely from the structure calculation. Bayesian weighting of data turns out to be optimal in terms of structural accuracy and, in contrast to cross-validation, demands almost no additional computational cost. This has been incorporated into the ISD software package and all standard NMR experimental terms have been implemented in the program (NOE distance restraints, torsion-angle restraints, scalar coupling constants, residual dipolar coupling constants). Furthermore, the program supports different error distributions (Gaussian, lognormal, von Mises, two-component mixture model). The slow convergence of the replica algorithm is a problem for the general application of the ISD software. To tackle this problem, minimization schemes will be integrated into the program with the aim of developing hybrid approaches between standard structure calculation and sampling around the final solution (Fig. 3).

3. Methods

3.1. Automated expression screening and HTP sample preparation

PCR products were amplified using an rtPCR reaction from cDNA prepared from various human cell lines. After purification using AmpPure PCR cleanup magnetic beads and T4 treatment, PCR products were cloned using ligation-independent cloning in a pET15b-derived N-terminal His-tag expression vector. After transformation in a suitable host, plasmid DNA was isolated and retransformed in BL21 Rosetta. Expression screening was performed in 100 μ l cultures where growth was monitored every 20 min and expression induced when the required OD was reached. After induction at 291 K, cells were lysed using detergents or sonication and proteins were purified under native conditions using His-tag affinity purification. Protein expression was evaluated using SDS-PAGE. All cloning, transformation, expression and purification steps were undertaken on a Hamilton STAR liquid-handling station equipped with multiple temperature-controllable shakers, temperature-controlled storage and a plate reader. The near-complete automation of all steps from gene to purified protein allowed expression screening under many different conditions including temperature, OD of induction, induction time, expression strain and solubility tag, parameters previously shown to critically influence total and soluble protein expression (Folkers *et al.*, 2004). We found that protein

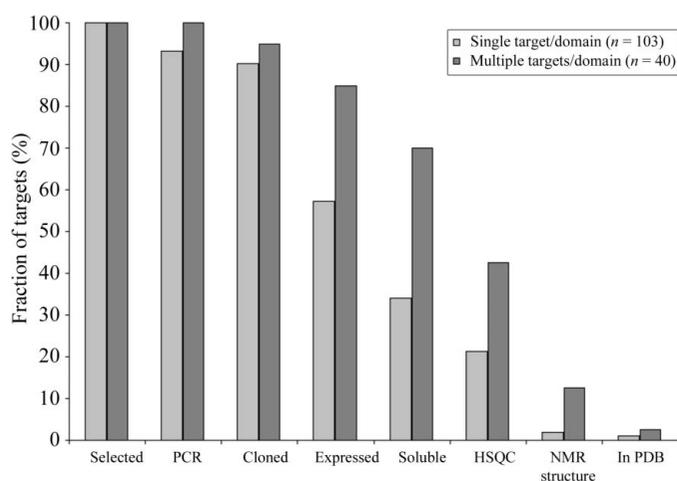


Figure 4 The overall efficiency, relative to the total number of targets selected, at various stages of the procedure from target selection to NMR structure. PCR, successful PCR amplification. Expressed, total expression >0.5 mg per millilitre of bacterial culture. Soluble, soluble expression >0.5 mg per millilitre of bacterial culture. HSQC, sufficient amount of protein was purified for ¹⁵N HSQC analysis. For the indicated number of selected targets either one (grey bar) or multiple (black) constructs were evaluated.

production in small-scale experiments correlates well with large-scale expression (Folkers *et al.*, 2004). Combined with efficient large-scale protein purification (Folkers *et al.*, 2004), this allowed the screening of a large number of samples. Proteins were classified in terms of suitability for structure determination by NMR on the basis of ^{15}N - ^1H -HSQC spectra, using peak dispersion, homogeneity and long-term sample stability as criteria.

3.2. Standardized NMR protocol for protein structure determination

NMR spectroscopy for protein structure elucidation at Utrecht has been adjusted in both organisational and methodological aspects for the special needs of the SPINE project.

In order to enable successful recording of all relevant NMR data by different users (including non-experts and visiting scientists from consortium partners) on the various spectrometers installed at the facility (with field strengths of 500, 600, 700, 750 and 900 MHz), a routine measurement protocol with easy set-up was established. NMR experiments were selected and implemented with the primary object of facilitating a reliable and straightforward automated data analysis, rather than absolutely minimizing spectrometer times. Initially, frequency ranges were determined from ^{15}N and ^{13}C HSQC spectra and the aggregate state was verified by a DOSY experiment. With NMR measurement times mostly limited by resolution requirements, optimally folded frequency windows were defined using in-house software (*FoldIt*, available on request). Inter-residual 3D HNCOC, CBCA[CO]NH and HBHA[CO]NH correlation spectra along with their intra-residual complements HN[CA]CO, HNCACB and HN[CA]HA then provided verified sequential connectivity information on four nuclei (CO, CA, CB and HA) for efficient automated backbone assignment by *AutoAssign* (Zimmermann *et al.*, 1997) and *PASTA* (Leutner *et al.*, 1998). Side chains were assigned from 3D H[C]CH-, [H]CCH- and C[CO]NH-TOCSY spectra. Finally, a combination of 3D HNH, HCH and CNH (optionally also CCH) edited NOESY spectra provided pseudo-four-dimensional resolved data (Diercks *et al.*, 1999) for distance-restraints structure calculation.

To find additional resonance assignments and to improve the quality of the peak lists, *CANDID* (Herrmann *et al.*, 2002) was used in an iterative cycle with the NOE spectra and preliminary *CANDID*-derived structures as input. Manual NOE peak assignments are generally not fixed in the structure-calculation runs. Instead, they were used to create accurate spectrum-specific chemical shift lists for consistency checks of consequent *CANDID* runs and to check the manual assignments. Hydrogen-bond restraints were applied when the data from the secondary shift data, expected NOE contacts and the structures were consistent. During the final *CANDID* run, Ramachandran and side-chain rotamer dihedral angle restraints were used for every cycle except the very last. In the final cycle fixed stereospecific assignments of prochiral groups were used if available. To complete the structure determina-

tion, water refinement was performed using *CNS* (Brünger *et al.*, 1998), where the final *CANDID*-based NOE restraints were used together with restraints for hydrogen bonds and dihedral restraints from *TALOS*. Structures were validated using *WHATIF* (Vriend, 1990) and *PROCHECK* (Laskowski *et al.*, 1996).

4. Structures

4.1. Human protein domains

Within the SPINE project, 257 targets were selected from 143 different human domains derived from the domain databases PFAM and SMART for which no structure was known and where *BLAST* searches against the PDB revealed no homologues. Domain boundaries were determined using publicly available programs for disorder prediction, secondary-structure prediction and multiple sequence alignment. On the basis of these predictions either one ($n = 103$) or multiple domain boundaries ($n = 40$) were chosen. As shown in Fig. 4, all steps in the pipeline from gene to structure benefit significantly from the multi-domain approach, with an overall increase in percentage of interpretable HSQCs from 21 to 42% of the domains (Fig. 4). Most importantly for multiple protein domains, addition or removal of terminal residues converted a protein with poor expression, solubility or biophysical behaviour to a protein characterized as suitable for structure determination by NMR. For 12.5% of all the domains where multiple targets were selected a structure was solved, while for one target per domain only 2% of the targets structure were successfully finished. Note that the number of good HSQCs per domain is larger than the number of structures solved, as work is still in progress or work was stopped owing to PDB submission of a homologous structure by other groups. Finally, a few samples failed to give a structure owing to poor long-term sample stability during NMR measurement.

These data clearly show that the number of samples as well as sample quality significantly increases by domain boundary optimization, underscoring that current bioinformatics tools are not sophisticated enough to accurately determine domain boundaries and that an experimental HTP expression-screening procedure is essential for HTP structure determination.

The expectations for the structural genomics approach were that this would increase the chance for identification of novel folds. Two new folds have been identified [PDB codes 1w6v (de Jong *et al.*, 2005) and 2bze]. Furthermore, for six different human proteins (PDB codes 1w6v, 2bze, n4bp2, 2aq0, 1rjy, 1ttx) and one protein complex (PDB code 1z00; Tripsianes *et al.*, 2005) functional insight was provided into the role of the domain by structural similarity or characteristic structural features. Information about the protein structures of the NMR node of SPINE (Utrecht, Florence and Strasbourg) is provided as supplementary material¹ available with this article.

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: GX5089). Services for accessing this material are described at the back of the journal.

4.2. Metal-binding proteins

Bioinformatic tools have been developed (Andreini *et al.*, 2004) by the Florence partner that allow the selection of protein targets with respect to their ability to bind metal ions. Genomic context analysis (conserved neighbourhood, gene fusions, phylogenetic occurrence) combined with homology-based methods (genome search, structure modelling, correlated mutation analysis) can predict both the pathway in which a protein operates and its molecular function (Arnesano, Banci, Bertini & Martinelli, 2005). The latter approach helps to identify new candidate metalloproteins (Banci, Bertini, Ciofi-Baffoni *et al.*, 2005). This approach, applied by the Florence group to proteins involved in copper homeostasis, identified 780 copper protein targets within selected prokaryotic and eukaryotic genomes and 25 structures of diamagnetic Cu^I and paramagnetic Cu^{II} proteins were solved (see supplementary material). Of all these structures solved, 17% had, at time of deposition, less than 30% sequence identity to a known structure.

When diamagnetic metal ions are present, X-ray absorption spectroscopy (XANES, edge and EXAFS) can be exploited to learn about the metal-coordination geometry (Banci, Bertini & Mangani, 2005). In the case of Cu^{II}-binding proteins, new NMR experiments based on ¹³C direct detection were developed. The first solution-structure determination of a type II copper(II) (paramagnetic) protein, CopC from *Pseudomonas syringae*, was obtained in this way (Arnesano, Banci, Bertini, Felli *et al.*, 2003). This protein binds copper(I) and copper(II) at two different sites, either sequentially or simultaneously (Arnesano, Banci, Bertini, Mangani *et al.*, 2003). The two sites are approximately 30 Å far apart. Oxidation of Cu^I-CopC or reduction of Cu^{II}-CopC causes migration of copper from one site to the other. CopC resides in the periplasm of Gram-negative bacteria where there is a multicopper oxidase, CopA, which may modulate the redox state of copper within the cell. Another example where the structure gave novel hints on protein function is represented by Sco1, a protein probably involved in the copper transfer to cytochrome *c* oxidase. The solution structure of Sco1 from *B. subtilis* is the first of this class and serves as a model for eukaryotic and bacterial homologues (Balatri *et al.*, 2003). Sco1 has a thioredoxin-like fold, which is generally devoted to a thiol:disulfide oxidoreductase activity. The structure therefore suggests that Sco1 may be involved in the maturation of cytochrome *c* oxidase, providing another step in the understanding of the COX complex-assembly process. Finally, the combination of NMR and docking methods allowed fast structure determination of protein complexes. As an example, the structure of a protein complex involved in copper(I) trafficking was solved (Arnesano *et al.*, 2004). Overall, the structural characterization of the proteins listed in the supplementary material, contributed at the molecular level to understanding the mechanisms for copper-ion uptake and transport in cells.

A combined X-ray/NMR approach addressed the details of the conformational degrees of freedom of the catalytic domain of MMPs, which are relevant for drug design. Besides solving

the X-ray structure of the catalytic domain of MMP-10 (Bertini, Calderone *et al.*, 2004), the catalytic domain of MMP-12 in three adducts with drug-like molecules was re-examined by NMR: the differences observed between the three structures do not arise from the nature of the ligand but from intrinsic conformational freedom, all conformations being sampled actually in solution and interconverting on a variety of time scales (Bertini, Del Bianco *et al.*, 2004).

4.3. Collaborations with other SPINE partners

The HTP protein-production facility (Folkers *et al.*, 2004) and the infrastructure for fast HSQC measurements at Utrecht has been offered to other members of the SPINE consortium to screen for feasibility of structure determination by NMR. Training facilities for SPINE members are available to solve structures of proteins by NMR under the supervision of an experienced NMR spectroscopist. A total of 44 target constructs underwent an NMR feasibility screen, from which 36 proteins were purified and NMR spectra were recorded from 31; eight had an HSQC that permitted structure determination (*e.g.* Fig. 5*a*). The value of screening samples for both NMR and X-ray is underscored by the large number of HSQCs classified as good (18%) for samples that failed to give well diffracting crystals. For 5 of these proteins the three-dimensional structure was determined (Fig. 5*b*). Importantly, all these samples were extensively screened in crystallization trials but failed to give a structure and all these structures were solved within four months by inexperienced NMR spectroscopists. Furthermore, using both methods in parallel, several protein structures have been determined by X-ray crystallography (see supplementary material and Fig. 5*c*). Various NMR experiments were conducted to either permit docking of protein-DNA (Rumpel *et al.*, 2004) or protein-protein complexes (Tripsianes *et al.*, 2005; Verdier *et al.*, 2005; Arnesano *et al.*, 2004).

The tight interplay between biochemical and structural approaches conducted at Strasbourg proves to be efficient in providing key functional data on protein targets involved in human cancer. The combined use of proteolysis experiments and bioinformatics allowed the identification of several structural domains in the human transcription factor TFIIF that could be studied in solution using NMR. Among the results of this study, the presence of a new type of C4C4 RING domain within the core subunit p44 is of particular interest since this domain may be involved in the regulation of the intracellular TFIIF level *via* the ubiquitination pathway (Kellenberger *et al.*, 2005). Another insight into TFIIF function revealed by this study arose from the identification of a PH/PTB domain within the p62 subunit (Gervais *et al.*, 2004). Biochemical experiments revealed that the p62 PH/PTB-like domain is required for nucleotide-excision repair and physically interacts with the 3' endonuclease XPG, providing a unique example of a PH/PTB-like domain potentially involved in the shuttling of TFIIF between DNA repair and transcription. Monodispersity-based quality-optimization strategies together with sequence engineering were used to

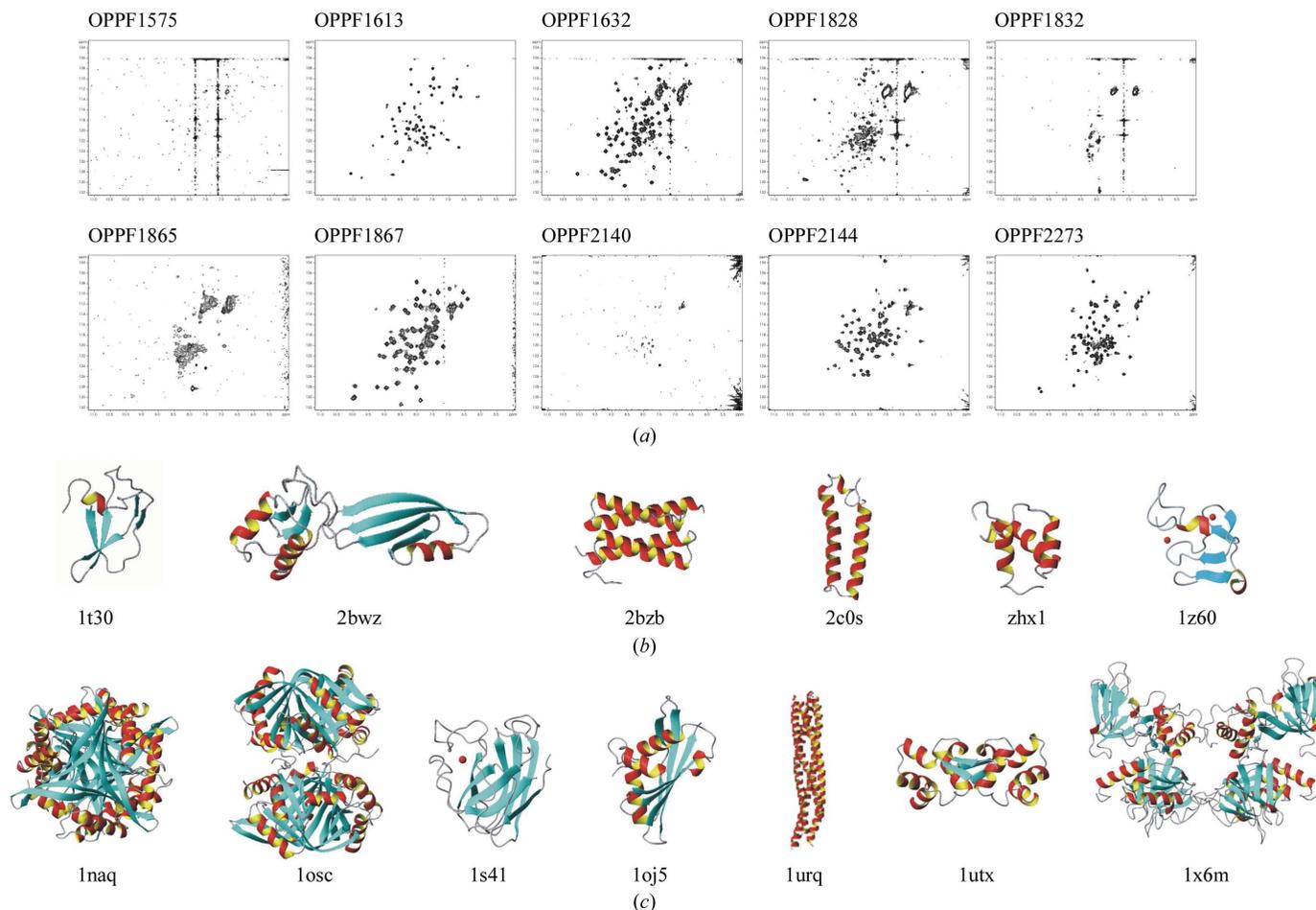


Figure 5
 (a) ¹⁵N HSQC screening of targets that failed to crystallize. (b) Structures determined in a collaborative effort between X-ray groups and the Utrecht NMR department. (c) X-ray structures determined in a collaborative effort between NMR and X-ray groups within SPINE

obtain samples of the C-terminal zinc finger of the oncoprotein E6 that were amenable for structural studies (Nominé *et al.*, 2003). The solution structure of this domain revealed a new type of zinc-binding fold and provided the first structural insights in the molecular pathways of human papillomavirus-mediated pathogenesis.

5. Conclusions and future perspectives

It is clear that NMR spectroscopy, with a modest share (15%) in the resources of SPINE, has significantly contributed to its overall objectives. Methodological improvements have been made in various aspects of HTP structure determination such as protein expression and screening, data collection and analysis and structure calculation. Important novel developments are the *protonless* NMR strategy for paramagnetic or large proteins (Arnesano, Banci, Bertini, Felli *et al.*, 2003; Bertini, Duma *et al.*, 2004) and the inferential structure-determination method (Rieping *et al.*, 2005) that provides a more unbiased and statistically sound basis for NMR structure calculations.

NMR is well integrated in the SPINE consortium. Several NMR structures have been derived for proteins that failed to

crystallize (Fig. 5). In the case of the York–Utrecht collaborations these were carried out by exchange students, thus providing a valuable educational experience.

With high-throughput methods and more efficient protocols in place, NMR is now ready for the next challenge: the structure determination of protein–protein complexes. Interactions between proteins in signalling networks are often weak and transient and NMR is exceptionally good in the characterization of these complexes. Protein–protein docking methods have recently been developed (*e.g.* HADDOCK; Dominguez *et al.*, 2003) that generate accurate structures for these complexes based on easily obtainable interaction data such as chemical shift perturbations. With these and other ongoing technical developments, NMR will continue to play an important role in structural genomics projects.

This work was funded by the European Commission as SPINE (Structural Proteomics In Europe) contract No. QLG2-CT-2002-00988 under the Integrated Programme ‘Quality of Life and Management of Living Resources’.

References

- AB, E., Pugh, D. J. R., Kaptein, R., Boelens, R. & Bonvin, A. M. J. J. (2006). *J. Am. Chem. Soc.* **128**, 7566–7571.

- Andersson, P., Weigelt, J. & Otting, G. (1998). *J. Biomol. NMR*, **12**, 435–441.
- Andreini, C., Bertini, I. & Rosato, A. (2004). *Bioinformatics*, **20**, 1373–1380.
- Arnesano, F., Balatri, E., Banci, L., Bertini, I. & Winge, D. R. (2005). *Structure*, **13**, 713–722.
- Arnesano, F., Banci, L., Bertini, I. & Bonvin, A. M. J. J. (2004). *Structure*, **12**, 669–676.
- Arnesano, F., Banci, L., Bertini, I., Fantoni, A., Tenori, L. & Viezzoli, M. S. (2005). *Angew. Chem.* **44**, 6341–6344.
- Arnesano, F., Banci, L., Bertini, I., Felli, I. C., Luchinat, C. & Thompsett, A. R. (2003). *J. Am. Chem. Soc.* **125**, 7200–7208.
- Arnesano, F., Banci, L., Bertini, I., Mangani, S. & Thompsett, A. R. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 3814–3819.
- Arnesano, F., Banci, L., Bertini, I. & Martinelli, M. (2005). *J. Proteome Res.* **4**, 63–70.
- Babini, E., Bertini, I., Capozzi, F., Del Bianco, C., Holleder, D., Kiss, T., Luchinat, C. & Quattrone, A. (2004). *Biochemistry*, **43**, 16076–16085.
- Babini, E., Bertini, I., Capozzi, F., Felli, I. C., Lelli, M. & Luchinat, C. (2004). *J. Am. Chem. Soc.* **126**, 10496–10497.
- Babini, E., Felli, I. C., Lelli, M., Luchinat, C. & Pierattelli, R. (2005). *J. Biomol. NMR*, **33**, 137.
- Baig, I., Bertini, I., Del Bianco, C., Gupta, Y. K., Lee, Y.-M., Luchinat, C. & Quattrone, A. (2004). *Biochemistry*, **43**, 5562–5573.
- Balatri, E., Banci, L., Bertini, I., Cantini, F. & Ciofi-Baffoni, S. (2003). *Structure*, **11**, 1431–1443.
- Banci, L., Bertini, I., Cavallaro, G., Giachetti, A., Luchinat, C. & Parigi, G. (2004). *J. Biomol. NMR*, **28**, 249–261.
- Banci, L., Bertini, I., Ciofi-Baffoni, S., Katsari, E., Katsaros, N., Kubicek, K. & Mangani, S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 3994–3999.
- Banci, L., Bertini, I. & Mangani, S. (2005). *J. Synchrotron Rad.* **12**, 94–97.
- Banci, L. & Rosato, A. (2003). *Acc. Chem. Res.* **36**, 215–221.
- Barbieri, R., Luchinat, C. & Parigi, G. (2004). *ChemPhysChem*, **21**, 797–806.
- Bax, A. (2003). *Protein Sci.* **12**, 1–16.
- Bermel, W., Bertini, I., Duma, L., Emsley, L., Felli, I. C., Pierattelli, R. & Vasos, P. (2005). *Angew. Chem. Int. Ed.* **44**, 3089–3092.
- Bermel, W., Bertini, I., Felli, I. C., Kümmerle, R. & Pierattelli, R. (2003). *J. Am. Chem. Soc.* **125**, 16423–16429.
- Bermel, W., Bertini, I., Felli, I. C., Kümmerle, R. & Pierattelli, R. (2006). *J. Magn. Reson.* **178**, 56–64.
- Bermel, W., Bertini, I., Felli, I. C., Lee, Y.-M., Luchinat, C. & Pierattelli, R. (2006). *J. Am. Chem. Soc.* **128**, 3918–3919.
- Bermel, W., Bertini, I., Felli, I. C., Piccioli, M. & Pierattelli, R. (2006). *Prog. NMR Spectrosc.* **48**, 25–45.
- Bermel, W., Bertini, I., Felli, I. C., Pierattelli, R. & Vasos, P. R. (2005). *J. Magn. Reson.* **172**, 324–328.
- Bertini, I., Calderone, V., Fragai, M., Luchinat, C., Mangani, S. & Terni, B. (2004). *J. Mol. Biol.* **336**, 707–716.
- Bertini, I., Del Bianco, C., Gelis, I., Katsaros, N., Luchinat, C., Parigi, G., Peana, M., Provenzani, A. & Zoroddu, M. A. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 6841–6846.
- Bertini, I., Duma, L., Felli, I. C., Fey, M., Luchinat, C., Pierattelli, R. & Vasos, P. (2004). *Angew. Chem. Int. Ed.* **43**, 2257–2259.
- Bertini, I., Faraone-Mennella, J., Gray, B. H., Luchinat, C., Parigi, G. & Winkler, J. R. (2004). *J. Biol. Inorg. Chem.* **9**, 224–230.
- Bertini, I., Felli, I. C., Kümmerle, R., Luchinat, C. & Pierattelli, R. (2004). *J. Biomol. NMR*, **30**, 245–251.
- Bertini, I., Felli, I. C., Kümmerle, R., Moskau, D. & Pierattelli, R. (2004). *J. Am. Chem. Soc.* **126**, 464–465.
- Bertini, I., Jiménez, B. & Piccioli, M. (2005). *J. Magn. Reson.* **174**, 125–132.
- Bertini, I., Jiménez, B., Piccioli, M. & Poggi, L. (2005). *J. Am. Chem. Soc.* **127**, 12216–12217.
- Bertini, I., Lee, Y.-M., Luchinat, C., Piccioli, M. & Poggi, L. (2001). *ChemBioChem*, **2**, 550–558.
- Böhm, M., Stadlhanner, K., Gruber, P., Theis, F. J., Lang, E. W., Tomé, A. M., Teixeira, A. R., Gronwald, W. & Kalbitzer, H. R. (2005). *IEEE Trans. Biomed. Eng.* **53**, 810–820.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Dancea, F. & Günther, U. (2005). *J. Biomol. NMR*, **33**, 139–152.
- De Jong, R. N., AB, E., Diercks, T., Truffault, V., Daniëls, M., Kaptein, R. & Folkers, G. E. (2005). *J. Biol. Chem.* **281**, 5026–5031.
- Diercks, T., Coles, M. & Kessler, H. (1999). *J. Biomol. NMR*, **15**, 177–180.
- Diercks, T., Daniels, M. & Kaptein, R. (2006). *J. Biomol. NMR*, **33**, 243–259.
- Diercks, T. & Orekhov, V. (2005). *J. Biomol. NMR*, **32**, 113–127.
- Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. (2003). *J. Am. Chem. Soc.* **125**, 1731–1737.
- Duma, L., Hediger, S., Brutscher, B., Bockmann, A. & Emsley, L. (2003). *J. Am. Chem. Soc.* **125**, 11816–11817.
- Duma, L., Hediger, S., Lesage, A. & Emsley, L. (2003). *J. Magn. Reson.* **164**, 187–195.
- Eletsky, A., Moreira, O., Kovacs, H. & Pervushin, K. (2003). *J. Biomol. NMR*, **26**, 167–179.
- Folkers, G. E., van Buuren, B. N. M. & Kaptein, R. (2004). *J. Struct. Funct. Genomics*, **5**, 119–131.
- Gervais, V., Lamour, V., Jawhari, A., Frindel, F., Wasielewski, E., Dubaele, S., Egly, J. M., Thierry, J. C., Kieffer, B. & Poterszman, A. (2004). *Nature Struct. Mol. Biol.* **11**, 616–622.
- Gronwald, W. & Kalbitzer, H. R. (2004). *Prog. NMR Spectrosc.* **44**, 33–96.
- Habeck, M., Nilges, M. & Rieping, W. (2005a). *Phys. Rev. E*, **72**, 031912.
- Habeck, M., Nilges, M. & Rieping, W. (2005b). *Phys. Rev. Lett.* **94**, 018105.
- Habeck, M., Rieping, W. & Nilges, M. (2005). *J. Magn. Reson.* **177**, 160–165.
- Herrmann, T., Güntert, P. & Wüthrich, K. (2002). *J. Mol. Biol.* **319**, 209–227.
- Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990). *J. Magn. Reson.* **89**, 496–514.
- Kellenberger, E., Dominguez, C., Fribourg, S., Wasielewski, E., Moras, D., Poterszman, A., Boelens, R. & Kieffer, B. (2005). *J. Biol. Chem.* **280**, 20785–20792.
- Klammt, C., Schwarz, D., Fendler, K., Haase, W., Dotsch, V. & Bernhard, F. (2005). *FEBS J.* **272**, 6024–6038.
- Kopp, J. & Schwede, T. (2004). *Nucleic Acids Res.* **32**, D230–D234.
- Kostic, M., Pochapsky, S. S. & Pochapsky, T. C. (2002). *J. Am. Chem. Soc.* **124**, 9054–9055.
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). *J. Biomol. NMR*, **8**, 477–448.
- Leutner, M., Gschwind, R. M., Liermann, J., Schwarz, C., Gemmecker, G. & Kessler, H. (1998). *J. Biomol. NMR*, **11**, 31–43.
- Lin, Y., Dancea, F., Löhr, F., Klimmek, O., Pfeiffer-Marek, S., Nilges, M., Wien, H., Kröger, A. & Rüterjans, H. (2004). *Biochemistry*, **43**, 1418–1424.
- Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. (2003). *Bioinformatics*, **19**, 315–316.
- Machonkin, T. E., Westler, W. M. & Markley, J. L. (2002). *J. Am. Chem. Soc.* **124**, 3204–3205.
- Nielsen, N. C., Thøgersen, H. & Sørensen, O. W. (1995). *J. Am. Chem. Soc.* **117**, 11365–11366.
- Nominé, Y., Charbonnier, S., Ristriani, T., Stier, G., Masson, M., Cavusoglu, N., Van Dorsselaer, A., Weiss, E., Kieffer, B. & Travé, G. (2003). *Biochemistry*, **42**, 4909–4917.
- Ottiger, M., Delaglio, F. & Bax, A. (1998). *J. Magn. Reson.* **131**, 373–378.

- Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 12366–12371.
- Pervushin, K., Vögeli, B. & Eletsy, A. (2002). *J. Am. Chem. Soc.* **124**, 12898–12902.
- Ried, A., Gronwald, W., Trenner, J. M., Brunner, K., Neidig, K.-P. & Kalbitzer, H. R. (2004). *J. Biomol. NMR*, **30**, 121–131.
- Rieping, W., Habeck, M. & Nilges, M. (2005). *Science*, **309**, 303–306.
- Rumpel, S., Razeto, A., Pillar, C. M., Vijayan, V., Taylor, A., Giller, K., Gilmore, M. S., Becker, S. & Zweckstetter, M. (2004). *EMBO J.* **23**, 3632–3642.
- Sattler, M., Schleucher, J. & Griesinger, C. (1999). *Prog. NMR Spectrosc.* **34**, 93–158.
- Stadlhanner, K., Theis, F. J., Lang, E. W., Gronwald, W. & Kalbitzer, H. R. (2003). *Neural Inform. Process.* **1**, 103–110.
- Szyperski, T., Yeh, D. C., Sukumaran, D. K., Moseley, H. N. B. & Montelione, G. T. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 8009–8014.
- Trbovic, N., Klammt, C., Koglin, A., Lohr, F., Bernhard, F. & Dotsch, V. (2005). *J. Am. Chem. Soc.* **127**, 13504–13505.
- Tripsianes, K., Folkers, G. E., AB, E., Das, D., Odijk, H., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R. & Boelens, R. (2005). *Structure*, **13**, 1849–1858.
- Verdier, L., Al-Sabi, A., Rivier, J. E., Olivera, B. M., Terlau, H. & Carlomagno, T. (2005). *J. Biol. Chem.* **280**, 21246–21255.
- Vögeli, B., Kovacs, H. & Pervushin, K. (2005). *J. Biomol. NMR*, **31**, 1–9.
- Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
- Wöhnert, J., Franz, K. J., Nitz, M., Imperiali, B. & Schwalbe, H. (2003). *J. Am. Chem. Soc.* **125**, 13338–13339.
- Zimmermann, D. E., Kulikowski, C. A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C.-Y., Powers, R. & Montelione, G. T. (1997). *J. Mol. Biol.* **269**, 592–610.