

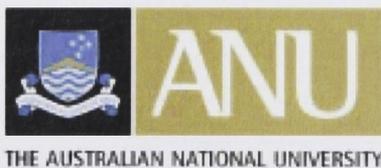


# **New Methods for Biomolecular NMR**

**PETER SHANG-CHE WU**

February 2008

A thesis submitted for the degree of Doctor of Philosophy  
of the Australian National University





New Methods  
for  
Biological NMR

THE AUSTRALIAN NATIONAL UNIVERSITY

A book which is essential for the biologist and a  
valuable reference for the chemist.

THE AUSTRALIAN NATIONAL UNIVERSITY

Abstract

# Declaration

The work in this thesis is my own except where otherwise stated



Peter Shang-Che Wu

30/06/2008

# Abstract

A collective theme of the papers that make up this thesis is improving the efficiency of nuclear magnetic resonance (NMR) research. In the post-genomic era where many NMR spectroscopists are inundated with genomic and cellular data to work with, speeding up NMR sample production and experimental time is ever more important. Three methods are described in this thesis that reduce the time taken to obtain NMR data, and in doing so, also facilitate new applications. In total, five papers and one review are presented.

Paper I is aimed at examining the binding mechanism of pyrin domains (PyD). This is achieved through studying the critical residues for self-association of the PyD from the inflammatory and cell-death signaling protein, ASC. An *in vivo* method was devised to track PyD filament formation, and the critical residues for filament formation were identified through site-directed mutagenesis. NMR spectroscopy was used to confirm the results *in vitro* and to ensure the structural fidelity of the mutants.

Paper II presents a new method for calibrating NMR pulse lengths based on the concept of nutation spectroscopy. The method is performed in a single scan, making it many times faster than traditional procedures. The method was found to give the same accuracy as traditional procedures (~1%) and is robust and easy to automate. It is envisaged that it will become the new routine method for pulse calibration.

Paper III describes SWET, a pulse sequence for solvent suppression on high-end spectrometers. SWET is resistant to radiation damping effects that become more prominent as spectrometer field strength and probe sensitivity increases. Therefore, in comparison, SWET was found to give similar solvent suppression but better selectivity than sequences susceptible to radiation damping, such as presaturation and WET. Also, solvent irradiation during evolution periods was found to significantly improve solvent suppression on high-field spectrometers, while having negligible effects on spectral appearance.

Paper IV reviews current practical knowledge on using *Escherichia coli* cell-free protein expression systems to produce labeled proteins for NMR analysis. Cell-free protein expression confers many advantages for NMR spectroscopists over *in vivo* expression. The review describes the advantages and characteristics of cell-free protein expression relevant to isotopic labeling, as well as some applications.

Paper V outlines a new labeling scheme to be used with cell-free protein expression. All 19 amino acid types (proline excluded as it is not observable in  $^{15}\text{N}$ -HSQC spectra) can be assigned in the  $^{15}\text{N}$ -HSQC spectra of 5 samples. The whole procedure from sample production to assignment can be accomplished in two days. The experiment also reduces spectral overlaps. The technique was applied for the assignment of two proteins.

Paper VI presents an advance that allows *Escherichia coli* cell-free protein expression to achieve the same yield using linear templates, such as those amplified from PCR, as for plasmid templates. The key is to construct linear templates with single-stranded overhangs which ligate upon addition to cell extracts. Besides saving time and cost in protein sample production, the advance has many practical applications. One example shown in the paper is that site-directed mutants could now be produced and ready for NMR analysis in less than a day.

## List of papers

This thesis is based on the following journal articles:

- I. M. Moriya, S. Taniguchi, **P. Wu**, E. Liepinsh, G. Otting, J. Sagara  
“Role of charged and hydrophobic residues in the oligomerization of the PYRIN domain of ASC”  
*Biochemistry* **44**, 575-583 (2005)
- II. **P. S. C. Wu**, G. Otting  
“Rapid pulse-length determination in high-resolution NMR”  
*J. Magn. Reson.* **176**, 115-119 (2005)
- III. **P. S. C. Wu**, G. Otting  
“SWET for secure water suppression on probes with high quality factor”  
*J. Biomol. NMR* **32**, 243-250 (2005)
- IV. K. Ozawa, **P. S. C. Wu**, N. E. Dixon, G. Otting  
“<sup>15</sup>N-labelled proteins by cell free protein synthesis: strategies for high-throughput NMR studies of proteins and protein-ligand complexes”  
*FEBS J.* **273**, 4154-4159 (2006)
- V. **P. S. C. Wu**, K. Ozawa, S. Jergic, X. C. Su, N. E. Dixon, G. Otting  
“Amino acid type identification in <sup>15</sup>N-HSQC spectra by combinatorial selective <sup>15</sup>N-labelling”  
*J. Biomol. NMR* **34**, 13-21 (2006)
- VI. **P. S. C. Wu**, K. Ozawa, S. P. Lim, S. Vasudevan, N. E. Dixon, G. Otting  
“Cell-free transcription/translation from PCR amplified DNA for high-throughput NMR studies”  
*Angew. Chemie Int. Ed.* **46**, 3356-3358 (2007)

# Table of Contents

<b>1. General introduction.....</b>	<b>1</b>
1.1 Background.....	1
1.2 Modern NMR instruments.....	3
1.3 <i>Escherichia coli</i> cell-free protein synthesis.....	4
1.4 Selective isotope labeling for NMR.....	6
<b>2. Studying interaction interfaces of death domain proteins.....</b>	<b>8</b>
2.1 The death domain superfamily of proteins.....	8
2.2 The oligomerisation interface of the ASC pyrin domain [Paper I].....	10
<b>3. Development of new methods in NMR.....</b>	<b>15</b>
3.1 A fast, automated procedure for pulse length determination [Paper II].....	15
3.2 A water suppression technique for modern NMR spectrometers [Paper III].....	16
3.2.1 Established water suppression techniques in homonuclear experiments.....	17
3.2.2 SWET.....	20
<b>4. Cell-free protein expression.....</b>	<b>22</b>
4.1 Review of cell-free labeling for high-throughput NMR applications [Paper IV].	22
4.2 Combinatorial labeling for fast identification of amino-acid types in HSQC spectra [Paper V].....	22
4.2.1 Combinatorial labeling.....	22
4.2.2 The $\tau$ subunit of <i>Escherichia coli</i> DNA Polymerase III.....	23
4.2.3 The flexible interaction regions of $\tau$ .....	25
4.3 High-yield <i>E. coli</i> cell-free protein expression from PCR products [Paper VI] ...	26
4.3.1 DNA and RNA stability in <i>Escherichia coli</i> cell-free extracts.....	26
4.3.2 Stabilizing linear double-stranded DNA in <i>E. coli</i> extracts.....	27
<b>5. Conclusions and future perspectives.....</b>	<b>29</b>
<b>6. Acknowledgments.....</b>	<b>32</b>
<b>7. References.....</b>	<b>34</b>
<b>8. Contribution to papers.....</b>	<b>42</b>
<b>9. Appendix I-VI.....</b>	<b>43</b>

# 1. General introduction

The discovery of the phenomenon of nuclear magnetic resonance (NMR) dates back over 60 years (Bloch et al., 1946; Purcell et al., 1946). Much of the theoretical background of NMR is well understood and is described in textbooks (Abragam, 1961; Ernst et al., 1987; Goldman, 1988; Cavanagh et al., 1996). Applications of NMR, however, continue to expand particularly in biomolecular areas. The development of better instrumentation and more sophisticated molecular biology methods provides strong impetus for further progress. This thesis outlines new NMR techniques for modern spectrometers as well as applications of cell-free protein expression to sample production for NMR experiments.

## 1.1 Background

Nuclear magnetic resonance was first observed in independent experiments by Bloch and Purcell (Bloch et al., 1946; Purcell et al., 1946). It was quickly taken up by physicists as a model system for quantum mechanical theory. The theoretical treatment of NMR by statistical quantum mechanics is known for its elegance and comparative simplicity. Nevertheless, the full statistical quantum mechanics treatment is cumbersome for calculating the results of the complex pulse experiments used today. The concept of the product operator formalism provided a major simplification for calculations (Sørensen et al., 1983). Spin mechanics can now be calculated by hand rather than requiring large computer matrix calculations. Relaxation effects are easily added as exponential factors (Cavanagh et al., 1996). This enables researchers to calculate the effects of pulse sequences rapidly and led to big improvements in pulse sequence design.

Chemists first took up NMR for the study of solids and organic molecules. Early experiments employed continuous wave instruments. The development of Fourier Transform instruments was a tremendous improvement to NMR instrumentation (Ernst

and Anderson, 1966). Recording times could be greatly reduced because the signal-to-noise was much better and it opened up the possibility of manipulating nuclear spins using sequences of pulses, and subsequently, two-dimensional NMR (Ernst, 1975). The demonstration of these ideas earned Richard Ernst the 1991 Nobel Prize in chemistry.

These developments were also instrumental for the first protein structure determination by NMR (Williamson et al., 1985; Kaptein et al., 1985). NMR spectra of amino acids and peptides have been well studied since the 1960s. Protein spectra, however, were too complex for interpretation at the time. Research by Kurt Wüthrich's group showed that by using two-dimensional correlation spectroscopy, sufficient resolution could be obtained to assign all NMR resonances of a protein (Wüthrich et al., 1982). This led to a strategy for *de novo* structure determination based on these sequential assignments, Nuclear Overhauser effect (NOE) measurements for distance information and scalar coupling measurements to obtain bond angle information (Wüthrich, 1986). NOEs are the basis of almost all NMR structure determinations today. Prior to this accomplishment, X-ray crystallography was the only proven method for determining protein structures at the atomic level.

The major limitation of NMR protein structure determination is a limit on molecular weight. Slow molecular tumbling and the spectral complexity of large proteins present major obstacles for NMR. Overcoming these issues has thus been a topical area for NMR research. The introduction of better molecular biology techniques and robust procedures for uniform  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeling of proteins has been very useful in this regard (Ikura et al., 1990). The associated introduction of three and higher dimensional experiments provided the same jump in resolution as the transition from one- to two-dimensional NMR (Ikura et al., 1990).

The concept of TROSY, where magnetization pathways are selected in which dipole-dipole (DD) and chemical shift anisotropy (CSA) relaxation destructively interfere, has also been vital for extending the molecular weight limit of NMR (Pervushin et al., 1997). It combines particularly well with protein deuteration which removes background dipolar

relaxation (Kay and Gardner, 1997). The current largest monomeric protein structure solved by NMR is a low resolution structure of the 82 kDa malate synthase G (Tugarinov et al., 2005).

The amount of information contained in NMR experiments means that NMR has also found applications in biomolecular research beyond three-dimensional structure determination. NMR has been particularly useful for studies of protein folding, protein dynamics and protein ligand interactions, areas where X-ray crystallography cannot provide the same level of information (Lipari et al., 1982; Miranker et al., 1991; Balbach et al., 1996; Eberstadt et al., 1998; Liu et al., 2000; Klein-Seetharaman et al., 2002; Eisenmesser et al., 2005; Lindorff-Larsen et al., 2005; Oltersdorf et al., 2005; Sadqi et al., 2006; Tang et al., 2006; Henzler-Wildman et al., 2007).

## 1.2 Modern NMR instruments

Modern NMR spectrometers are equipped with superconducting magnets producing highly homogeneous fields of up to 950 MHz (the first commercial 1 GHz NMR spectrometer will be installed this year). High magnetic fields confer useful advantages for NMR spectroscopists. Firstly, the magnitude of nuclear polarization and hence, signal-to-noise (S/N) in NMR experiments increase with the magnetic field strength to the power of 2.5  $\left( \frac{\gamma_I}{\gamma_S} \right)^{2.5}$  (Cavanagh et al., 1996). Therefore, a 20% increase in field strength provides an increase in S/N of 58%. Higher magnetic fields also afford better spectral resolution. To a first approximation, spectral dispersion increases linearly with magnetic field (Cavanagh et al., 1996).

In contrast to the incremental S/N gains from magnetic field strengths, the introduction of cryogenic probes has provided a big leap in S/N for NMR experiments. The concept of cryogenic probes was first noted in 1976 by Hoult (Hoult and Richards, 1976). Due to

the difficulty of cooling the radiofrequency (RF) circuitries to liquid helium temperatures while conforming to other stringent NMR probe requirements, the first experimental demonstration was not performed until 1984 (Styles et al., 1984). Commercial cryogenic NMR probes were launched in the late 1990s. In current cryogenic probes, both the preamplifier and the RF coil circuitries are cooled to ~15-30 K. Thermal noise in the electrical circuits is thereby minimized.

### **1.3 *Escherichia coli* cell-free protein synthesis**

Cell-free protein expression using *Escherichia coli* cell extracts was demonstrated in the 1960s (Nirenberg and Matthaei, 1961). The early experiments were established for the study of protein translation. Therefore, the low yields of those early experiments were not of major concern. Spirin's introduction of a continuous flow system where the reaction buffer is continuously exchanged, providing fresh amino acids and dNTPs, gave a major improvement in yield (Spirin et al., 1988). Each mRNA in the cell-free mixture can be translated hundreds of times, making cell-free protein expression an interesting method for protein production. Further optimizations allow current *E. coli* cell-free expression systems to produce yields comparable with *in vivo* expression, and the use of cell-free expression has increased correspondingly (Kigawa et al., 1995; Kigawa et al., 1999; Ozawa et al., 2004). The protocols for the preparation of *E. coli* cell-free extracts as well as reaction procedures are documented in numerous papers and reviews (see for instance, Kigawa et al., (2004); Apponyi et al., (2008)). Figure 1 shows my setup for a cell-free reaction.

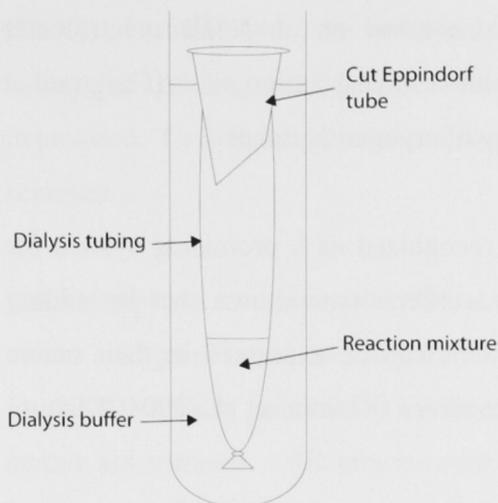


Figure 1: My setup for a continuous flow cell-free expression system.

Once cell extracts have been prepared and stored, cell-free protein expression is more convenient than *in vivo* expression. Toxic proteins which would otherwise affect the growth of *E. coli* cells can be expressed successfully using cell extracts. The user also has ready control of the reaction environment. The convenience and versatility makes cell-free expression increasingly of interest for high-throughput protein work such as structural genomics (Yokoyama, 2003). Another particularly advantageous application of cell-free protein expression is for selective isotope labeling of proteins. In cell-free protein expression, amino acids are directly supplied to the reaction. Therefore there are no dependencies on bacterial biosynthetic processes. A useful consequence is that amino acid inter-conversions are greatly diminished when labeling is performed *in vitro* compared to experiments *in vivo* (Kigawa et al., 1995). Moreover, the reaction volumes for cell-free expressions are in the order of milliliters in contrast to the hundreds of milliliters of culture media used for *in vivo* expression. This means that substantial cost savings can be achieved with expensive labeled amino-acids when making isotope labeled sample.

The characteristics of cell-free protein expression make it a good sample production method for NMR spectroscopy. The ability to selectively label one amino acid type with spin  $\frac{1}{2}$  nuclei greatly reduces spectral complexity. Guignard and co-workers have also demonstrated that due to the low levels of background expression in modern cell-free

systems, heteronuclear labeled samples can be measured on an NMR spectrometer without any purification other than a dialysis to remove excess amino acids (Guignard et al., 2002). This is greatly assisted by the sensitivity of cryogenic probes.

Cell-free protein expression is also increasingly recognized as a promising system for membrane protein expression. Klammt and co-workers have shown that by adding surfactants to cell-free reactions, membrane proteins can be expressed in their native folded states, saving time-consuming refolding processes (Klammt et al., 2004; Klammt et al., 2005; Klammt et al., 2006).

## 1.4 Selective isotope labeling for NMR

The demand to expand the range of NMR to larger proteins and proteins of poor spectral quality motivates the search for methods to obtain additional information and better resolution to facilitate assignments. Selective isotope labeling has become a major tool towards this end. One of the early applications of selective isotope labeling for NMR is the use of fractional  $^{13}\text{C}$  labeling (Neri et al., 1989). Amino acid biosynthesis in *E. coli* has the property that pro-R and pro-S methyl groups in valine and leucine originate from different pyruvate molecules. Therefore, by using a 9:1 mixture of  $^{12}\text{C}$  and  $^{13}\text{C}$  glucose as the sole carbon source for *E. coli*, the diastereomeric forms of valine and leucine methyl groups can be distinguished by their  $^{13}\text{C}$ - $^{13}\text{C}$  couplings to the  $\beta$  or  $\gamma$  carbons, respectively. More recently, it has been shown that isotope labeled  $\alpha$ -keto acids can be used as biosynthetic precursors to achieve a variety of aliphatic amino-acid labeling schemes (Rosen et al., 1996; Goto et al., 1999). This has been applied to good effect for the study of large proteins (Tugarinov et al., 2006).

While the biosynthetic pathways of *E. coli* can be advantageous, they can also be a major hinderance. Isotope labeling of specific amino acids reduces spectral overlap and identifies amino acid types. This strategy is now commonly used to assist with backbone assignment of larger proteins (see for example, Langer et al., (2004); Rodriguez et al.,

(2006)). Unfortunately, the various transamination reactions that occur in *E. coli* lead to isotope scrambling and limit the usefulness of selective amino-acid labeling in *in vivo* expression. Cell-free protein expression enables the possibility of more complex labeling schemes.

For example, a combinatorial  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling scheme can distinguish between amino acid pairs based on six pairs of  $^{15}\text{N}$ -HSQC and 2D-HNCO experiments (Parker et al., 2004). This is possible by having a unique  $^{13}\text{C}$  labeling pattern for every amino acid across six samples. All amino acids are  $^{15}\text{N}$  labeled in all samples but the isotopic labeling is either 100% or only 50%. This is again unique for each amino acid. Therefore, inspecting peak intensities in the six  $^{15}\text{N}$ -HSQC spectra will reveal the amino acid residue type and inspecting the existence of peaks in the six 2D-HNCOs will reveal the amino acid type of the preceding residue.

Recently, Kainosho and co-workers showed that using a specifically synthesized set of stereospecifically labeled amino acids, the complete stereospecific assignment and three-dimensional structure of proteins up to 40 kDa can readily be obtained (Kainosho et al., 2006).

## 2. Studying interaction interfaces of death domain proteins

### 2.1 The death domain superfamily of proteins

The death domain superfamily is one of the largest protein domain superfamilies (McEntyre and Gibson, 2004). It comprises four principal family members, the death domain (DD), the death effector domain (DED), the caspase recruitment domain (CARD) and the pyrin domain (PyD). Every member is well represented in the human genome. They appear in proteins involved in inflammatory response and apoptosis (Park et al., 2007). In general, their function seems to be to serve as adaptor domains to bring proteins into close proximity to facilitate the action of signaling enzymes such as kinases and caspases. Death domains take part in many major inflammatory and apoptotic signaling complexes, acting as anchors to hold them together.

Given the function of death domains as molecular joints, their structure and their mechanism of interaction is of primary interest for understanding their function. Three-dimensional structures are available for several proteins in every death domain subfamily. They all adopt homologous six-helix bundle structures (Figure 2). Differences lie in the length and orientation of the helices. Structural insight into the interaction mechanisms of death domains is harder to obtain as their role as molecular adhesives entails that they are prone to aggregation. Many of the currently available structures were determined at non-physiological pH values or using mutants that prevent the formation of intermolecular interactions.

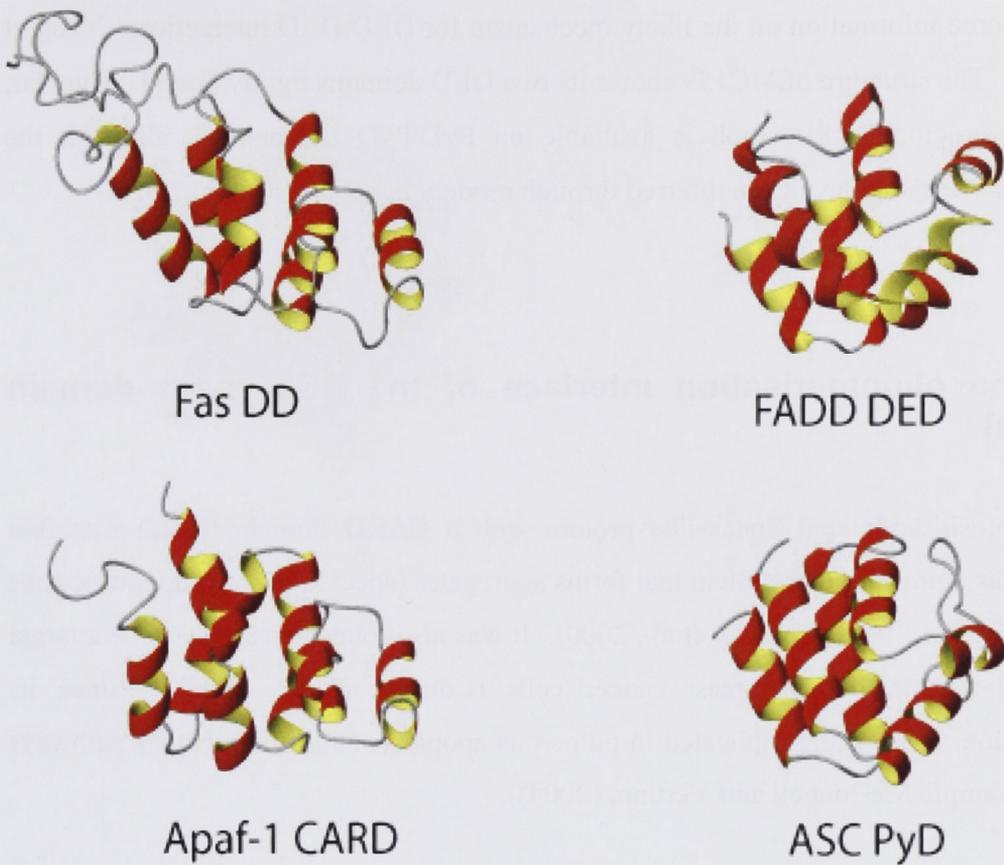


Figure 2: Three-dimensional structure of a member from each DD subfamily. DD from Fas (Huang et al., 1996). DED from FADD, F25G mutant (Eberstadt et al., 1998). CARD from Apaf-1 (Zhou et al., 1999). PyD from ASC (Liepinsh et al., 2003).

An interesting paradigm about death domain interactions is that members of each DD subfamily interact almost exclusively in homotypic interactions with other members of the same subfamily. Given the structural resemblance between subfamilies, this selectiveness is surprising. It also poses the question of whether the interaction mechanism might be different for each DD subfamily. Unfortunately due to the difficulty of working with DD complexes, only two DD complex structures have been solved thus far. They nevertheless provide some insights into these intriguing interactions. The two complex structures available are those of Tube and Pelle (DD:DD) and Apaf-1 and caspase-9 (CARD:CARD) (Qin et al., 1999; Xiao et al., 1999). They show clear differences in their interaction interfaces. The three-dimensional structure of MC159

provides some information on the likely mechanism for DED:DED interactions (Yang et al., 2005). The structure of MC159 shows its two DED domains tightly bound. Thus far, no direct structural information is available on PyD:PyD interactions, although the interaction interfaces have been inferred through modeling (Liepinsh et al., 2003).

## **2.2 The oligomerisation interface of the ASC pyrin domain [Paper I]**

The Apoptosis-associated Speck-like protein with a CARD domain (ASC) was first identified as a proapoptotic protein that forms aggregates (specks) in some apoptotic cells (Masumoto et al., 1999; Conway et al., 2000). It was also found separately to be a target of methylation in human breast cancer cells (Conway et al., 2000). Since its identification, it has been implicated in numerous apoptosis and inflammatory pathways (see for example McConnell and Vertino, (2004)).

ASC contains two domains, a CARD domain and a PyD domain. A solution state structure of the ASC PyD has been solved and the NMR assignments are available (Liepinsh et al., 2003), making it a suitable candidate for the investigation of PyD:PyD interactions. ASC PyD self-associates to form oligomers at neutral pHs. To perform the NMR assignment and structure determination, therefore, a pH of 3.7 was used to abolish oligomerisation. The structure of ASC PyD shows two charged patches on the surface of the protein, a positive and a negative patch (Figure 3). This, along with the fact that binding is disrupted at acidic pHs, led one to speculate that these charged patches are the interaction surfaces of PyD. The negative-positive association would also lead to concatenation, explaining the oligomerisation. The finding in paper I that mutations of the charged residues on these charged patches of ASC PyD abolish filament formation both *in vivo* and *in vitro* further support their role in PyD interactions. Based on this belief, it is anticipated that combining two ASC PyD mutants, each with a mutation on a differently charged surface, might yield a dimer which could be studied by NMR (see Figure 4).

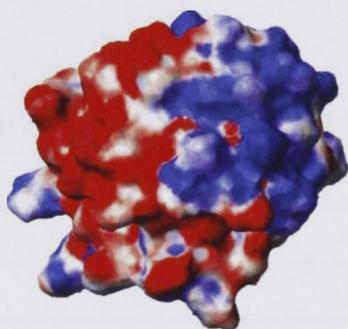


Figure 3: Potential surface map of ASC PyD, showing distinct positive and negative surfaces.

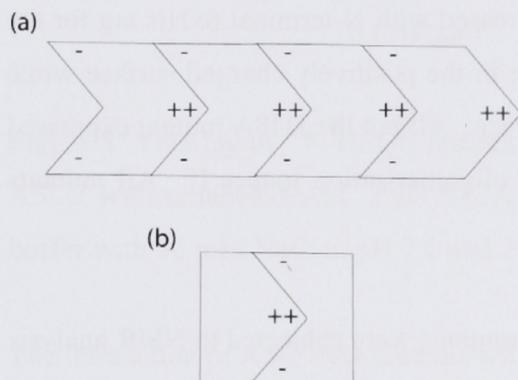


Figure 4: (a) Hypothetical scheme of the mechanism for ASC PyD oligomerisation. ASC PyD concatenates by interactions between positive to negative surfaces. (b) A heterodimer obtained by disrupting different charged patches on the surface of ASC PyD through site-directed mutagenesis could be suitable for NMR studies.

## Production and NMR spectroscopy of ASC PyD mutants

Table 1: Summary of ASC PyD constructs studied

Mutant	Surface	Expression	Solubility (pH )
Wild-type	N/A	+	-
K21A	+	+	+
L25A	+	+	+
R41A	+	+	+
D48A	-	-	?
D51A	-	+	+

Five mutants plus wild-type ASC PyD were expressed with N-terminal 6xHis tag for the study of their interaction. Three mutations were in the positively charged surface while the other two were in the negatively charged surface. All but the D48A mutant expressed well. Their solubility indicates a disruption of oligomerisation [paper I]. All mutants were at least partially soluble at pH 7.2.

Following purification by nickel-NTA columns, mutants were subjected to NMR analysis. All mutants exhibited NMR spectra closely resembling that of the wild-type, indicating that the structures are conserved. The mutants were subsequently combined in all possible pairings in the hope of making a complex. Unfortunately, no peak shifts, which would be indicative of complex formation, were observed for any pairings. It is possible that ASC PyD interactions require the collective association of multiple monomers. This would support the *in vivo* observations that ASC oligomers appear to be the functional form of ASC (Mariathasan et al., 2004).

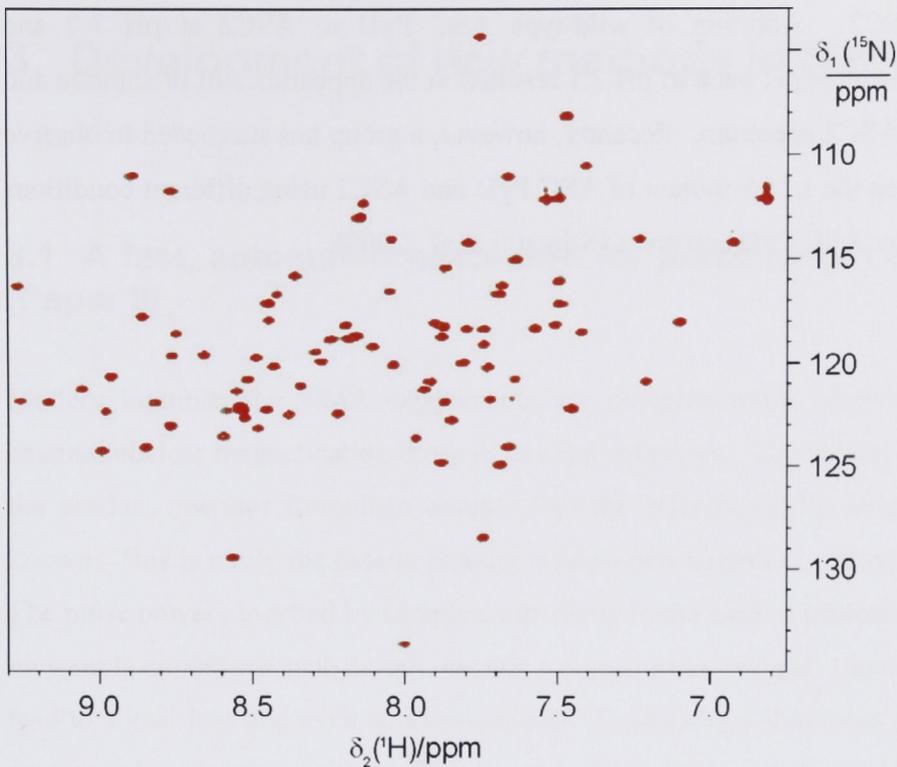


Figure 5: Overlay of  $^{15}\text{N}$ -HSQC spectra from  $^{15}\text{N}$ -labeled ASC2 (green) and  $^{15}\text{N}$ -labeled ASC2 with unlabeled ASC PyD R41A mutant. The spectra were run in 10 mM tris buffer with 50 mM NaCl at pH 7.1 and 25°C. No peak shifts can be observed.

The interaction of ASC PyD mutants with  $^{15}\text{N}$ -labeled ASC2 (supplied by Dr Justine Hill from the University of Queensland) was also tested. ASC2, also known as POP1 and ASC1, is a pyrin-domain-only protein that functions as a regulator of inflammatory pathways (Natarajan et al., 2006). It appears to function as an inhibitor to NF- $\kappa$ B and caspase-1 activation (Stehlik et al., 2003). Its interaction with ASC plays a key role in this inhibitory function (Stehlik et al. 2003). By binding to the ASC pyrin domain, ASC2 prevents the formation of complexes necessary for further downstream signaling. The structure of ASC2 is known through NMR (Natarajan et al., 2006).

Despite much *in vivo* evidence of the ASC/ASC2 interaction, including GST-pulldown assays (J. Hill personal communications), my NMR experiments gave negative results. No peak shifts could be observed when any of the soluble ASC PyD mutants were added

to  $^{15}\text{N}$ -labeled ASC2. Addition of wild-type ASC PyD to ASC2 at pH 3.7 and subsequent adjustment of pH back to pH 7.1 resulted in the appearance of precipitate and no changes to the ASC2 spectrum. Recently, however, a group has succeeded to observe interactions between the L25A mutant of ASC PyD and ASC2 using different conditions and a slightly longer ASC PyD construct (Srimathi et al., 2008).

## 3. Development of new methods in NMR

### 3.1 A fast, automated procedure for pulse length determination [Paper II]

Modern biomolecular NMR experiments use complex trains of pulses and delays to channel nuclear magnetization through desired pathways. Theoretical calculations using the product operator formalism assume that the effective pulse lengths are exact and known. This is rarely the case in practice when a new sample is placed in a spectrometer. The pulse power absorbed by samples, especially in the case of protons, varies depending on sample conditions such as salt, sample concentration, and pH. Incorrect pulse lengths lead to signal loss and even spurious signals. Calculations show that a 10% error in the proton pulse length results in S/N loss of about 16% in a basic HSQC experiment. Losses increase for every proton pulse added to the experiment. Therefore, pulse length calibration is a routine part of the setup procedure for NMR experiments.

The most frequently used method of pulse length calibration for high-resolution NMR used to be to search for minimum signal following a  $180^\circ$  or  $360^\circ$  pulse which leaves no magnetization in the transverse plane and then back-calculate the  $90^\circ$  pulse length (Barnaal and Lowe, 1963; Keifer, 1999). This method is robust and accurate to about 1%. This is within the errors caused by off-resonance effects, which causes spin off-resonance from a radiofrequency (RF) pulse to experience a slightly weaker RF pulse compared to spins directly on-resonance.

The drawback of the traditional calibration procedure is that it is rather time consuming. Automation would involve a systematic scan of pulse lengths at small intervals to find the null point. This takes a few minutes, which is not of great concern for long three-dimensional experiments, but represents a significant fraction of time for short HSQC experiments. The issue is particularly pertinent with the recent interest in fast multi-

dimensional experiments (Frydman et. al., 2002, Freeman and Kupce, 2003). There are also other described methods for pulse calibration but these have not become popular due to their slow speed or lack of robustness (Sørensen et al., 1986; Nielsen et al., 1987).

Paper II describes the use of nutation spectroscopy (Freeman and Kupce, 2003) for pulse length determination in high resolution NMR. The method is very effective. Proton pulse lengths can be evaluated in a single scan, the errors are no larger than those of the conventional procedure and the method lends itself to automation. It is also demonstrated in the paper that the method would work for samples with spectra containing many peaks, provided that there is one dominating resonance. An additional bonus is that the entire RF homogeneity profile can be checked from a simple nutation spectrum because Fourier transformation of a nutation spectrum delineates all of the frequencies observed in the sample. The procedure reported in Paper II has been implemented by the main spectrometer manufacturers Bruker and Varian as a standard pulse determination routine.

### **3.2 A water suppression technique for modern NMR spectrometers [Paper III]**

Water suppression is an important element of biomolecular NMR pulse sequences. The solvent water signal in protein samples needs to be reduced by more than 10000 fold to become comparable in size to the protein signals. For heteronuclear experiments, this could usually be achieved by the separate manipulation of protein and water resonances. Water suppression tends to be more problematic for homonuclear pulse sequences, particularly the simple COSY experiment (Aue et al., 1976). In paper III, a new water suppression sequence for homonuclear experiments is proposed and examined. The aim was to devise a water presaturation sequence which is unaffected by radiation damping.

Radiation damping describes the feedback mechanism between a strong solvent signal and the RF coil. In brief, a strong precessing solvent magnetization generates a current in the RF coil. This current in turn influences the trajectory of the solvent magnetization.

The resulting effect is that any transverse solvent magnetization is driven back to the Z axis with a rate constant of  $R_{rd}$  ((Abragam, 1961; Mao and Ye, 1997). This effect is independent of the solvent transverse relaxation rate,  $R_2$ , and the longitudinal relaxation rate,  $R_1$ .  $R_{rd}$  is directly related to the quality factor (Q) of the probe by  $R_{rd} = knQ$  where  $n$  is the filling factor and  $k$  is proportional to the equilibrium magnetization and the nuclear gyromagnetic ratio (Bloom, 1957). Radiation damping has an adverse effect on water suppression and with modern NMR instruments having higher magnetic fields and probes of higher quality factor, radiation damping is increasingly an issue. This provided the motivation to devise a water suppression scheme unaffected by radiation damping.

### 3.2.1 Established water suppression techniques in homonuclear experiments

#### Presaturation

Presaturation (Figure 6) is the earliest solvent suppression method invented and is still often used for homonuclear experiments (Campbell et al., 1974). The concept follows logically from continuous wave experiments. A weak RF irradiation is applied at the frequency of the solvent signal. The continuous irradiation during the recycle delay results in partial saturation of the solvent signal. The RF inhomogeneity of the probe is another major contributor to saturation (Figure 7).

The quality of solvent suppression in presaturation is limited by the extent of saturation and recovery of solvent magnetisation during the pulse sequence. The recovery of solvent magnetisation is particularly problematic in two-dimensional experiments where a long  $t_1$  evolution period would allow significant solvent signal recovery.



Figure 6: Presaturation sequence. The solid bar represents a long continuous pulse.

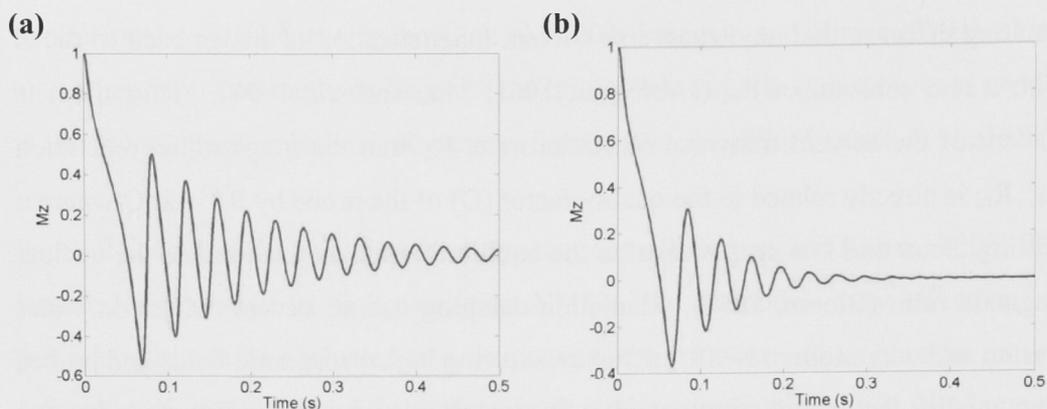


Figure 7: Simulation of solvent magnetization during presaturation. The presaturation is set on resonance at 30 Hz nutation frequency. **(a)** Solvent presaturation assuming perfect RF homogeneity across the sample. **(b)** Solvent presaturation assuming a Gaussian distribution (1 Hz standard deviation) in RF homogeneity across the sample. Simulation was performed using Matlab.

## WATERGATE

WATERGATE (Figure 8) is a popular solvent suppression technique for both homonuclear and heteronuclear experiments because it is very robust (Piotto et al., 1992; Sklenâr et al., 1993). The idea is to first dephase all magnetization using a gradient pulse and subsequently apply a 0 or 360 degree pulse to the solvent magnetization while all other signals receive a 180 degree pulse. A second gradient rephases the signals of interest, while further dephasing the solvent signal. The selective  $^1\text{H}$  pulses in WATERGATE can be implemented in two ways, either using shaped selective pulses or using the 3-9-19 sequence which relies on differences in spin precession frequencies in a manner similar to jump-return sequences (Plateau and Guéron, 1982).

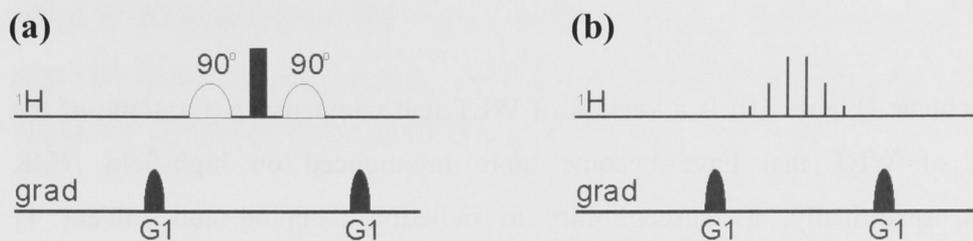


Figure 8: WATERGATE sequences. Bars represent rectangular pulses. Open arches represent shaped pulses. grad = gradient. Gradient pulses are shown with filled arches. (a) WATERGATE with shaped selective pulses. (b) WATERGATE with the 3-9-19 pulse. The rectangular pulses have relative power levels of 3, 9, 19, 19, 9, 3.

## WET

WET is a solvent signal suppression technique similar to presaturation in that the solvent signal is suppressed during the recycle delays of the experiment (Ogg et al., 1994; Smallcombe et al., 1995). This is achieved by a train of four solvent selective pulses and gradient pulses. The strength of each gradient pulse is halved at every step from G1 to G4 (Figure 9) to prevent accidental refocusing. The strengths of the solvent selective pulses have been optimized to account for RF inhomogeneity and water  $T_1$  relaxation (Ogg et al., 1994).

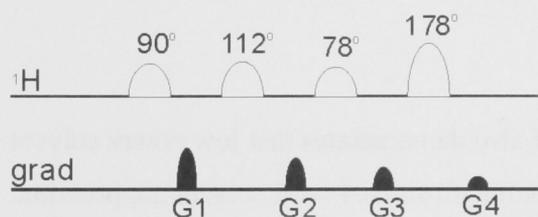


Figure 9: WET sequence. Open arches represent shaped pulses. The pulse flip angles are noted on top of the selective pulses. grad = gradient. Gradient pulses are shown with filled arches.

### 3.2.2 SWET

The SWET scheme (Figure 10) is a version of WET that compensates for some of the shortcomings of WET that have become more pronounced on high-field NMR spectrometers, specifically, its susceptibility to radiation damping and solvent  $T_1$  relaxation during pulse sequences. In SWET, the shaped selective pulses in WET are replaced by DANTE type pulses interspersed by bipolar gradients (Figure 10) (Böckmann and Guittet, 1996). The pulse sequence is thus much less affected by radiation damping while still maintaining selectivity. This allows lower pulse powers to be used, resulting in less saturation of protein signals.

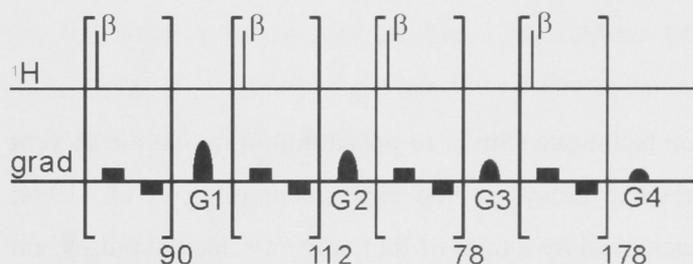


Figure 10: SWET sequence. Brackets represent repeated segments with the number of repetitions shown at the bottom right hand corner. Bars indicate square pulses with flip angle  $\beta$ , with  $\beta$  being approximately  $1^\circ$  (to be adjusted for optimal water suppression). grad = gradient. Filled bars and arches represent square and shaped gradients, respectively.

Besides presenting the SWET scheme, paper III also demonstrates that low power solvent irradiation during evolution periods can be beneficial to assist with solvent suppression. This idea was previously thought to be compromised by concomitant Bloch-Siegert shifts (Wider et al., 1983). These are peak shifts that occur as a result of limited selectivity of the RF pulses. As a consequence, if an RF field is applied during evolution periods, spins experience both the  $B_0$  field and the RF field. Their precession frequencies thus become influenced by both according to the formula  $\omega = \sqrt{\Omega^2 + \omega_1^2}$ , where  $\omega$  is the precession frequency,  $\Omega$  is the offset of the Larmor frequency of the spin from the solvent frequency,

and  $\omega_1$  is the amplitude of the solvent irradiation field. Simulations and experiments in paper III show, however, that the Bloch-Siegert shifts induced by weak irradiations are negligible for modern high-field spectrometers.

## **4. Cell-free protein expression**

### **4.1 Review of cell-free labeling for high-throughput NMR applications [Paper IV]**

Controlled conditions for isotope labeling present one of the biggest advantages of cell-free protein expression (Kigawa et al., 1995; Ozawa et al., 2005). Paper IV summarizes our current knowledge and experience with cell-free isotope labeling and optimal labeling schemes.

### **4.2 Combinatorial labeling for fast identification of amino-acid types in HSQC spectra [Paper V]**

#### **4.2.1 Combinatorial labeling**

Amino acid type identification is a very useful piece of information for protein NMR peak assignments. Some amino acids can be readily identified from their chemical shifts and cross-peak patterns in commonly used experiments (Wüthrich, 1986). Specialized experiments have been designed to identify other amino acids from NMR spectra based on J couplings between side-chain spins (Gehring and Guittet, 1995; Tashiro et al., 1995; Dötsch et al., 1996, Schubert et al., 1999). Not all amino acids, however, can be distinguished in this manner and many of the above-mentioned experiments suffer from poor signal-to-noise and ambiguities. Selective labeling of amino acids provides a much more robust way of identifying the individual amino acids (Yamazaki et al., 1991; Ozawa et al., 2004). Ultimately, however, it is cumbersome to produce all the samples required to identify all 20 amino acids (20 samples for 20 amino acids).

In combinatorial labeling, the mathematical concept of combinatorics is applied to isotope labeling. Each sample contains more than one labeled amino acid and the labeling pattern is varied from one sample to another. The number of distinct possible labeling patterns, therefore, increases combinatorially with the number of samples produced. Shortle applied this idea to assign 14 amino-acid residue types from five samples produced with partial  $^{15}\text{N}$ -labeling (Shortle, 1994). However, the isotope scrambling that occurs in *in vivo* expression limits the applicability of the procedure to 14 amino acids. It is also very cumbersome to express and purify 5 *in vivo* samples for NMR analysis.

In the implementation in Paper V, combinatorial labeling was applied using a cell-free protein expression system to assign all 20 amino acid types and also to track chemical shift changes between two constructs of a gene (domain V of the  $\tau$  subunit of *E. coli* DNA polymerase III (Gao and McHenry, 2001)). The labeling scheme was optimized for spectral resolution based on amino acid abundances compiled in the NCBI database.

#### **4.2.2 The $\tau$ subunit of *Escherichia coli* DNA Polymerase III**

The *E. coli* DNA polymerase III holoenzyme is a complex composed of at least ten proteins (Figure 11) which are responsible for replication of the *E. coli* genome (Clover and McHenry, 2001). The  $\tau$  subunit forms part of the  $\beta$  clamp loading complex. Through its dimerisation, it holds the  $\alpha$  core on both the leading and lagging strand in the appropriate orientation with respect to DNA and the  $\beta$  clamp.  $\tau$  contains five domains (Gao and McHenry, 2001). Domains I to III are shared with the  $\gamma$  subunit and their structures are known through crystal structures (Jeruzalmi et al., 2001; Kazmirski et al., 2004). Domains IV to V have resisted crystallization but are amenable to NMR studies. Domain IV plays a central role in  $\tau$ 's interaction with DNA and the  $\beta$  subunit (Gao and McHenry, 2001), while domain V is the unit responsible for the interaction with the  $\alpha$  subunit (Gao and McHenry, 2001). The  $\alpha$  interaction site has recently been mapped to the C-terminal 18 residues of domain V (Jergic et al., 2007). The solution structure of the

N-terminal 128 residues ( $\tau_{C14}$ ) of domain V has been determined (Su et al., 2007) (Figure 12).

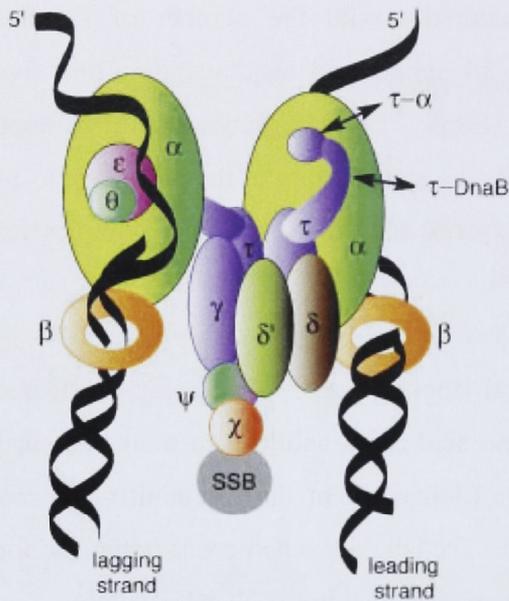


Figure 11: Schematic diagram of the *E. coli* DNA polymerase III holoenzyme. The  $\alpha$  and DnaB interaction regions of  $\tau$  are indicated. Diagram reproduced from Schaeffer et al., 2005.



Figure 12: Solution structure of  $\tau_{C14}$  (PDB – 2AYA)

### 4.2.3 The flexible interaction regions of $\tau$

In paper V it is shown that the  $\alpha$ -binding eighteen-residue polypeptide segment on the C-terminus of  $\tau$  has a random-coil structure and is highly flexible. The  $\beta$  and DNA binding domain IVa was also examined for structural features using NMR chemical shifts and linewidths. A 22 kDa construct of  $\tau$  comprising domain IVa and  $\tau_{C14}$  was named  $\tau_{C22}$ .

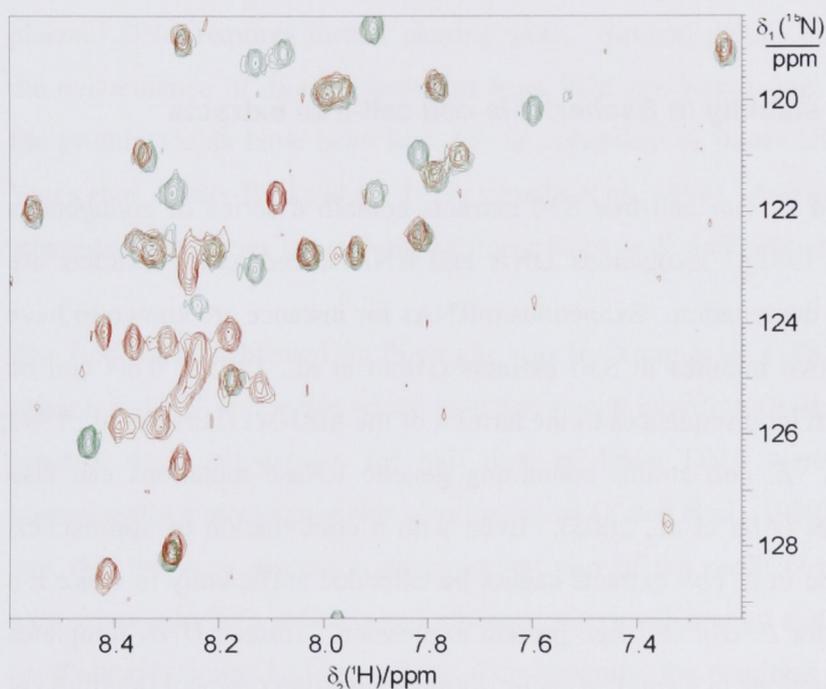


Figure 13: Overlay of the  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -Ala, -Lys, -Arg, -Phe, -Gln, -Met, -Cys and -Trp labeled  $\tau_{C14}$  (in green) and  $^{15}\text{N}$ -Ala labeled  $\tau_{C22}$  (in red). Both were expressed by cell-free protein synthesis.  $\tau_{C22}$  was purified on a DEAE sepharose column after expression to remove bound DNA. The overlaid spectra show that the  $\tau_{C14}$  resonances appear at the same position in the spectrum of the  $\tau_{C22}$  construct and that the additional resonances from  $\tau$  domain IVa all appear in random coil regions.

Remarkably, all the peptide segments of  $\tau$  domains IV and V that have been shown to be involved in intermolecular binding interactions appear to be intrinsically unstructured (Figure 13). This includes domain IVa ( $\tau$  residues 430-496), which binds to DNA and DnaB helicase, and  $\tau$  residues 625-643 which is involved in binding with the  $\alpha$  subunit of DNA polymerase III (Gao and McHenry, 2001; Jergic et al., 2007).

### **4.3 High-yield *E. coli* cell-free protein expression from PCR products [Paper VI]**

#### **4.3.1 DNA and RNA stability in *Escherichia coli* cell-free extracts**

Conventionally prepared *E. coli* cell-free S30 extracts contain a series of endogenous nucleases (Pratt et al., 1981). Exogenous DNA and RNA added to the extracts are therefore prone to rapid degradation. Exogenous mRNAs for instance are known to have half-lives of less than two minutes in S30 extracts (Hirao et al., 1993). This can be increased by adding stem loop sequences to the termini of the mRNAs (Hirao et al., 1993; Lee and Cohen, 2001). *E. coli* strains containing genetic RNase mutations can also prolong mRNA lifetimes (Ahn et al., 2005). Even with a combination of approaches, however, mRNA lifetime in *E. coli* extracts cannot be extended sufficiently to make it a practical starting point for *E. coli* cell-free protein expression. Instead DNA templates are used as inputs for reactions in a coupled transcription/translation protocol (Kudlicki et al., 1992; Kigawa et al., 2004).

DNA is not perpetually stable either in the nuclease-rich environment. Circular double-stranded plasmid DNA is, however, sufficiently stable over the time of the cell-free reaction. Therefore, plasmid DNA is currently the most frequently used template for cell-free reactions. Linear DNA is much less stable (Yang et al., 1980). Numerous exonucleases exist in *E. coli*, the most prominent of which is exonuclease V which is encoded by the *recBCD* gene in *E. coli* (Wright et al., 1971; Amundsen et al., 1986). In

addition, helicases present in *E. coli* unwind linear DNA, forming single-stranded overhangs which are highly susceptible to degradation (Korangy and Julin, 1993; Yu et al., 1998).

### 4.3.2 Stabilizing linear double-stranded DNA in *E. coli* extracts

The possibility of using linear DNA as template for cell-free reactions is very attractive because linear DNA is a direct product of PCR reactions whereas the production of plasmid DNA requires further cloning work. Several groups have attempted to exploit the convenience of direct expression from PCR products using cell-free expression but the protein yields have been low due to instability of linear DNA (Kung et al., 1978; Yates et al., 1980; Burks et al., 1997; Ohuchi et al., 1998). Accordingly, researchers have attempted to improve the stability of linear DNA in *E. coli* extracts.

The first paper published on this topic was by Yang et al. (1980). They found that by using a K-12 *E. coli* strain which contains a *recB* mutation (Barbour and Clark, 1970) to produce their cell extract, the half lives of linear DNA were greatly improved and consequently expression yields also improved (Yates et al., 1980). A problem, however, was that the *recB* protein is an important part of the *recBCD* recombination complex, participating in both helicase and nuclease activity (Korangy and Julin, 1993). Mutating *recB* greatly diminished its activity. Consequently, the resulting strain grows poorly. A *RecD* mutant strain which suppresses endonuclease V activity but does not inhibit recombination is a good alternative for cell extracts (Amundsen et al., 1986; Biek and Cohen, 1986; Lesley et al., 1991). The resulting stabilization effect for linear DNA is, however, limited (Michel-Reydellet et al., 2005). Michel-Reydellet and co-workers instead removed the whole *recBCD* operon from an A19 *E. coli* strain and introduced a  $\lambda$  phage recombination system to replace its function (Michel-Reydellet et al., 2005). They found that this led to increased stability of linear DNA and healthy cells. A different approach attempts to suppress the *recBCD* activity in the cell extracts using inhibitors. In particular, the addition of moderate concentrations of  $\lambda$  phage Gam protein, which is a

known inhibitor of the *recBCD* complex, was found to significantly improve DNA stability (Sitaraman et al., 2004). A recent investigation showed that reducing the incubation temperature during the production of *E. coli* cell extracts also reduces subsequent nuclease activity (Seki et al., 2008). Finally, there are ways to prolong the lifetime of the mRNA. For example, expression from PCR templates could be significantly improved using cell extracts from the *E. coli* strain BL21\*(DE3) which contains a genetic mutation that truncates and inactivates RNase E, a major RNase in *E. coli* (Ahn et al., 2005; Hahn and Kim, 2006).

Although important improvements in protein yields could be achieved by these approaches, cell-free expression from linear DNA has yet to achieve the yields reported for the optimized continuous flow cell-free system using plasmid DNA as templates (Kigawa et al., 1999). In Paper VI, we approached the problem from a different angle. Rather than modifying cell extracts, we modified our PCR protocol to produce linear DNA templates capable of cyclization to achieve the lifetimes of plasmid DNA. This approach led, for the first time, to identical yields in a continuous flow cell-free system irrespective of whether PCR templates or plasmid DNA were used.

The process to make sticky-end DNA capable of cyclization follows that used by other research groups to avoid restriction enzymes (Ailenberg and Silverman, 1996). The process is shown in Figure 14. A different idea using abasic sites to stall polymerases in order to generate overhangs (Gal et al., 1999) was also tested. However, the ligation efficiency over these abasic sites was not sufficient to achieve good yields in a cell-free system.

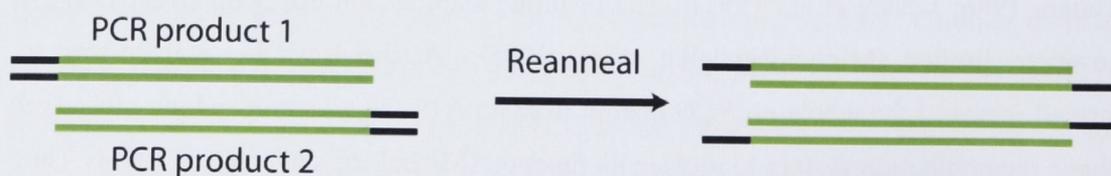


Figure 14: Process to make sticky-end DNA templates capable of cyclization. PCR product 1 and PCR product 2 are made in two separate PCR reactions. The template

DNA is shown in green. Reannealing can be achieved by heating at 96°C for 5 minutes with no formamide needed.

## 5. Conclusions and future perspectives

In this thesis four new NMR methods were developed with a focus on biomolecular applications. A review of cell-free isotope labeling and a study of the ASC pyrin domain were also undertaken. The new methods developed are:

**A new pulse length calibration procedure** - The nutation pulse calibration procedure requires only a single scan and is easy and robust to automate. It is comparable in accuracy to the traditional pulse calibration method. Its speed and simplicity has seen it adopted as the standard pulse calibration method for both Bruker and Varian instruments. It will be particularly useful in the future as ultra-fast NMR experiments become more prevalent and pulse calibration will represent an increasingly significant portion of spectrometer time. The speed and ease of automation also makes the method ideal for exchange experiments where dead time is a concern and in LC-NMR where many samples are processed sequentially. Varian has already developed a heteronuclear version to extend the method to non-proton nuclei (Murali, 2006).

**A new water suppression sequence** – SWET is designed to suppress radiation damping effects during presaturation. This means that for high-end spectrometers where radiation damping is prevalent, SWET is a good alternative to conventional presaturation in order to achieve solvent suppression with higher selectivity. On higher-end spectrometers, this results in significantly reduced saturation of protein signals. As NMR spectroscopists continue to improve their equipment, SWET should become more prominent as a water suppression method for homonuclear experiments. Furthermore solvent irradiation during  $t_1$  evolution periods should become a standard feature in all homonuclear 2D NMR experiments as it substantially improves water suppression while having negligible effects on spectral appearance.

**Amino acid type assignment in  $^{15}\text{N}$ -HSQC spectra using cell-free protein expression and combinatorial labeling** – The combinatorial labeling scheme devised can assign all 20 amino acid types using only 5 samples. The scheme is also designed to reduce the number of cross-peaks in the  $^{15}\text{N}$ -HSQC experiments by about 2/3, which considerably aids assignment. Using cell-free protein expression, the whole process from sample production to amino acid type assignment can be performed in less than 2 days. The information obtained is highly complementary to standard heteronuclear sequential assignment experiments. The speed and information garnered from the method would make it a good routine tool for future NMR backbone assignment processes, saving valuable spectrometer time from having otherwise to run complex 3D experiments. I have shown that the 14 kDa  $\tau_{\text{C14}}$  can be assigned using the amino acid type information and an HNCA experiment alone.

As demonstrated in the paper, the method is also useful for observing chemical shift changes such as those presented by mutants or ligand binding. The amino acid type information makes tracking large peak shifts much easier, while the reduced spectral overlaps assist in identifying peaks.

**Using PCR produced DNA templates for cell-free protein expression** – The new method enables the same yields to be achieved in an *E. coli* cell-free expression system independent of whether PCR products or whole plasmids are used as templates. The method frees users from time consuming cloning steps. NMR protein samples can therefore be prepared from genes in a day, needing only a PCR reaction, cell-free expression, and dialysis.

The method is ideal for high-throughput protein studies, particularly as it is easy to run both PCR and cell-free reactions in parallel. For example, the method could be employed in the future to perform high-throughput NMR screens from cDNA libraries to identify targets suitable for structure determination or ligand-binding experiments. It could also be combined with a front-end directed evolution technique to screen for protein mutants of desired properties, such as solubility, ligand binding ability, or structure.

The method also makes it very quick to produce and study protein mutants. In paper VI, this was employed to assign a protein exhibiting very poor NMR spectra. The assignment technique can assign proteins that would be impossible to work with using normal sequential assignment procedures. We are working on automating the whole procedure using pipetting robots so that tens, even hundreds of site-directed mutants can be produced automatically and quickly. The rapid production and NMR studies of site-directed mutants would also be useful for many other purposes, particularly for functional analysis of proteins. For example, the method could be used to search for critical ligand binding residues or residues critical for protein folding or to manipulate dynamic movements of a protein. The possible insights gained from these should greatly support future protein engineering endeavors.

## 6. Acknowledgments

A large number of people have assisted me during my PhD studies in ways big and small. Unfortunately, I cannot give a full acknowledgement to everyone who has helped me in my studies. That list would stretch very far and in the end be meaningless. Rather, I will say a big blanketing thank you to everyone who has helped, worked with, taught, supported, encouraged, entertained or generally been nice to me during my PhD studies. Having said that, I want to specifically acknowledge the people who have been particularly important.

I want to say a giant thank you to **Professor Gottfried Otting** who has been such a fantastic supervisor. Gottfried happily spent so much of his time teaching me, providing me with guidance, supporting my projects, giving me new inspiration and ideas, being my careers counselor etc. etc.

I would also like to express major gratitude to **Professor Nicholas Dixon** for his enthusiasm, support, streams of ideas, and being a living biological encyclopedia, and **Dr Kiyoshi Ozawa**, for mentoring me in the lab and being a major source of support for all my cell-free expression projects.

I would also like to thank:

- Dr Patrick Schaeffer who taught me many of the molecular biology techniques I now use routinely in the lab
- Dr Guido Pintacuda for being a major source of inspiration and for his hospitality in my trip to the Karolinska Institute in Sweden
- Dr Andrei Kaikkonen for his kind hospitality in Sweden
- Dr Justine Hill for collaborating with me in a lot of the experimental projects we tried relating to death domain proteins.
- Dr Michael John and Dr Xun-Cheng Su for their help with many NMR related work
- Penny Lilley and Dr Ruhu Qi for providing lab support

- Marilyn Holloway for being the best Academic Secretary
- Everyone else in Gottfried and Nick's group, particularly Dr Slobodan Jergic, Dr Madeleine Headlam and Dr Ah Young Park
- Members in Professor David Ollis' group, particularly Dr Jian-Wei Liu

More personally, I would like to thank my parents, Karen and Randy, for their constant encouragement; my sister, Alice, who packs lunch for me everyday and is always supportive and amusing; and my apartment mate, Zach Chew, for entertaining me with his antics, sharing dinner cooking and lending me a car whenever I needed one.

I would also like to thank the Australian National University, the Australian Research Council and the Research School of Chemistry for providing me with support and numerous scholarships.

## 7. References

- Abragam, A. (1961). Principles of nuclear magnetism. Oxford, Oxford University Press.
- Ahn, J. H., Chu, H. S., Kim, T. W., Oh, I. S., Choi, C. Y., Hahn, G. H., Park, C. G. and Kim, D. M. (2005) Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochem. Bioph. Res. Co.*, **338**: 1346-1352.
- Ailenberg, M. and Silverman, M. (1996) Description of a one step staggered reannealing method for directional cloning of PCR-generated DNA using sticky-end ligation without employing restriction enzymes. *Biochem. Mol. Biol. Int.*, **39**: 771-779.
- Amundsen, S. K., Taylor, A. F., Chaudhury, A. M. and Smith, G. R. (1986) RecD - the gene for an essential 3rd subunit of exonuclease-V. *Proc. Natl. Acad. Sci. USA*, **83**: 5558-5562.
- Apponyi, M., Ozawa, K., Dixon, N. E. and Otting, G. (2008). Cell-free protein synthesis for analysis by NMR spectroscopy. *Methods in Molecular Biology, Structural proteomics: high-throughput methods*. B. Kobe, M. Guss and T. Huber. Eds., Totowa, Humana Press.
- Aue, W. P., Bartholdi, E. and Ernst, R. R. (1976) 2-Dimensional spectroscopy - Application to nuclear magnetic resonance. *J. Chem. Phys.*, **64**: 2229-2246.
- Balbach, J., Forge, V., Lau, W. S., vanNuland, N. A. J., Brew, K. and Dobson, C. M. (1996) Protein folding monitored at individual residues during a two-dimensional NMR experiment. *Science*, **274**: 1161-1163.
- Barbour, S. D. and Clark, A. J. (1970) Biochemical and genetic studies of recombination proficiency in *Escherichia coli* .1. Enzymatic activity associated with RecB<sup>+</sup> and RecC<sup>+</sup> Genes. *Proc. Natl. Acad. Sci. USA*, **65**: 955-&.
- Barnaal, D. and Lowe, I. J. (1963) Effects of rotating magnetic fields on free-induction decay shapes. *Phys. Rev. Lett.*, **11**: 258-&.
- Biek, D. P. and Cohen, S. N. (1986) Identification and characterization of RecD, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. *J. Bacteriol.*, **167**: 594-603.
- Bloch, F., Hansen, W. W. and Packard, M. (1946) Nuclear induction. *Phys. Rev.*, **69**: 127.
- Bloom, S. (1957) Effects of radiation damping on spin dynamics. *J. Appl. Phys.*, **28**: 800-805.
- Böckmann, A. and Guittet, E. (1996) Suppression of radiation damping during selective excitation of the water signal: The WANTED sequence. *J. Biomol. NMR*, **8**: 87-92.
- Burks, E. A., Chen, G., Georgiou, G. and Iverson, B. L. (1997) In vitro scanning saturation mutagenesis of an antibody binding pocket. *Proc. Natl. Acad. Sci. USA*, **94**: 412-417.
- Campbell, I. D., Dobson, C. M., Jeminet, G. and Williams, R. J. (1974) Pulsed NMR methods for observation and assignment of exchangeable hydrogens - application to Bacitracin. *FEBS Lett.*, **49**: 115-119.
- Cavanagh, J., Fairbrother, W. J., Palmer, A. G. and Skelton, N. J. (1996). Protein NMR spectroscopy: principles and practice. San Diego, Academic Press.

- Clover, B. P. and McHenry, C. S. (2001) The DNA polymerase III holoenzyme: An asymmetric dimeric replicative complex with leading and lagging strand polymerases. *Cell*, **105**: 925-934.
- Conway, K. E., McConnell, B. B., Bowring, C. E., Donald, C. D., Warren, S. T. and Vertino, P. M. (2000) TMS1, a novel proapoptotic caspase recruitment domain protein, is a target of methylation-induced gene silencing in human breast cancers. *Cancer Res.*, **60**: 6236-6242.
- Dötsch, V., Oswald, R. E. and Wagner, G. (1996) Amino-acid-type-selective triple-resonance experiments. *J. Magn. Reson. B*, **110**: 107-111.
- Eberstadt, M., Huang, B. H., Chen, Z. H., Meadows, R. P., Ng, S. C., Zheng, L. X., Lenardo, M. J. and Fesik, S. W. (1998) NMR structure and mutagenesis of the FADD (Mort1) death-effector domain. *Nature*, **392**: 941-945.
- Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E. and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, **438**: 117-121.
- Ernst, R. R. (1975) 2-Dimensional spectroscopy. *Chimia*, **29**: 179-183.
- Ernst, R. R. and Anderson, W. A. (1966) Application of Fourier transform spectroscopy to magnetic resonance. *Rev. Sci. Instru.*, **37**: 93-102.
- Ernst, R. R., Bodenhausen, G. and Wokaun, A. (1987) Principles of nuclear magnetic resonance in one and two dimensions. Oxford, Clarendon.
- Freeman, R. and Kupce, E. (2003) New methods for fast multidimensional NMR. *J. Biomol. NMR*, **27**: 101-113.
- Frydman, L., Scherf, T., Lupulescu, A. (2002) The acquisition of multidimensional NMR spectra within a single scan. *Proc. Natl. Acad. Sci. USA*, **99**: 15858-15862.
- Gal, J., Schnell, R., Szekeres, S., Kalman, M. (1999) Directional cloning of native PCR products with preformed sticky ends (autosticky PCR). *Mol. Gen. Genet.*, **260**: 569-573.
- Gao, D. X. and McHenry, C. S. (2001) tau binds and organizes Escherichia coli replication proteins through distinct domains - Domain IV, located within the unique C terminus of tau, binds the replication fork helicase, DnaB. *J. Biol. Chem.*, **276**: 4441-4446.
- Gao, D. X. and McHenry, C. S. (2001) tau binds and organizes Escherichia coli replication proteins through distinct domains - Partial proteolysis of terminally tagged tau to determine candidate domains and to assign domain V as the alpha binding domain. *J. Biol. Chem.*, **276**: 4433-4440.
- Gehring, K. and Guittet, E. (1995) 2-Dimensional nuclear-magnetic-resonance method for identifying the <sup>N</sup>H /N signals of amino-acid-residues following glycine. *J. Magn. Reson. B*, **109**: 206-208.
- Goldman, M. (1988). Quantum description of high-resolution NMR in liquids. Oxford, Oxford University Press.
- Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. and Kay, L. E. (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated <sup>15</sup>N-, <sup>13</sup>C-, <sup>2</sup>H-labeled proteins. *J. Biomol. NMR*, **13**: 369-374.
- Guignard, L., Ozawa, K., Pursglove, S. E., Otting, G. and Dixon, N. E. (2002) NMR analysis of in vitro-synthesized proteins without purification: a high-throughput approach. *FEBS Lett.*, **524**: 159-162.

- Hahn, G. H. and Kim, D. M. (2006) Production of milligram quantities of recombinant proteins from PCR-amplified DNAs in a continuous-exchange cell-free protein synthesis system. *Anal. Biochem.*, **355**: 151-153.
- Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. and Kern, D. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, **450**: 913-U27.
- Hirao, I., Yoshizawa, S. and Miura, K. (1993) Stabilization of messenger-RNA in an Escherichia coli cell-free translation system. *FEBS Lett.*, **321**: 169-172.
- Hoult, D. I. and Richards, R. E. (1976) Signal-to-noise ratio of nuclear magnetic resonance experiment. *J. Magn. Reson.*, **24**: 71-85.
- Huang, B. H., Eberstadt, M., Olejniczak, E. T., Meadows, R. P. and Fesik, S. W. (1996) NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain. *Nature*, **384**: 638-641.
- Ikura, M., Kay, L. E. and Bax, A. (1990) A novel approach for sequential assignment of <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N spectra of larger proteins - Heteronuclear triple resonance 3-dimensional NMR spectroscopy - application to Calmodulin. *Biochemistry*, **29**: 4659-4667.
- Ikura, M., Marion, D., Kay, L. E., Shih, H., Krinks, M., Klee, C. B. and Bax, A. (1990) Heteronuclear 3-D NMR and isotopic labeling of calmodulin - Towards the complete assignment of the <sup>1</sup>H NMR spectrum. *Biochem. Pharmacol.*, **40**: 153-160.
- Jergic, S., Ozawa, K., Williams, N. K., Su, X. C., Scott, D. D., Hamdan, S. M., Crowther, J. A., Otting, G. and Dixon, N. E. (2007) The unstructured C-terminus of the tau subunit of Escherichia coli DNA polymerase III holoenzyme is the site of interaction with the alpha subunit. *Nucleic Acids Res.*, **35**: 2813-2824.
- Jeruzalmi, D., O'Donnell, M. and Kuriyan, J. (2001) Crystal structure of the processivity clamp loader gamma (gamma) complex of E-coli DNA polymerase III. *Cell*, **106**: 429-441.
- Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, A. M. and Guntert, P. (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature*, **440**: 52-57.
- Kaptein, R., Zuiderweg, E. R., Scheek, R. M., Boelens, R. and van Gunsteren, W. F. (1985) A protein structure from nuclear magnetic resonance data. lac repressor headpiece. *J. Mol. Biol.*, **182**: 179-182.
- Kay, L. E. and Gardner, K. H. (1997) Solution NMR spectroscopy beyond 25 kDa. *Curr. Opin. Struc. Biol.*, **7**: 722-731.
- Kazmirski, S. L., Podobnik, M., Weitze, T. F., O'Donnell, M. and Kuriyan, J. (2004) Structural analysis of the inactive state of the Escherichia coli DNA polymerase clamp-loader complex. *Proc. Natl. Acad. Sci. USA*, **101**: 16750-16755.
- Keifer, P. A. (1999) 90 degrees pulse width calibrations: How to read a pulse width array. *Concept Magn. Res.*, **11**: 165-180.
- Kigawa, T., Muto, Y. and Yokoyama, S. (1995) Cell-free synthesis and amino-acid selective stable isotope labeling of proteins for NMR analysis. *J. Biomol. NMR*, **6**: 129-134.

- Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A. and Yokoyama, S. (2004) Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics*, **5**: 63-68.
- Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. and Yokoyama, S. (1999) Cell free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.*, **442**: 15-19.
- Klammt, C., Lohr, F., Schafer, B., Haase, W., Dotsch, V., Ruterjans, H., Glaubitz, C. and Bernhard, F. (2004) High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.*, **271**: 568-580.
- Klammt, C., Schwarz, D., Fendler, K., Haase, W., Dotsch, V. and Bernhard, F. (2005) Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J.*, **272**: 6024-6038.
- Klammt, C., Schwarz, D., Lohr, F., Schneider, B., Dotsch, V. and Bernhard, F. (2006) Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *FEBS J.*, **273**: 4141-4153.
- Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. and Schwalbe, H. (2002) Long-range interactions within a nonnative protein. *Science*, **295**: 1719-1722.
- Korangy, F. and Julin, D. A. (1993) Kinetics and processivity of ATP hydrolysis and DNA unwinding by the RecBC enzyme from Escherichia coli. *Biochemistry*, **32**: 4873-4880.
- Kudlicki, W., Kramer, G. and Hardesty, B. (1992) High-efficiency cell-free synthesis of proteins - refinement of the coupled transcription translation system. *Anal. Biochem.*, **206**: 389-393.
- Kung, H. F., Tainsky, M. and Weissbach, H. (1978) Regulation of in vitro synthesis of alpha-peptide of beta-galactosidase directed by a restriction fragment of lactose operon. *Biochem. Biophys. Res. Co.*, **81**: 1000-1010.
- Langer, T., Vogtherr, M., Elshorst, B., Betz, M., Schieborr, U., Saxena, K. and Schwalbe, H. (2004) NMR backbone assignment of a protein kinase catalytic domain by a combination of several approaches: Application to the catalytic subunit of cAMP-dependent protein kinase. *Chembiochem*, **5**: 1508-1516.
- Lee, K. and Cohen, S. N. (2001) Effects of 3' terminus modifications on mRNA functional decay during in vitro protein synthesis. *J. Biol. Chem.*, **276**: 23268-23274.
- Lesley, S. A., Brow, M. A. D. and Burgess, R. R. (1991) Use of in vitro protein synthesis from polymerase chain reaction-generated templates to study interaction of Escherichia coli transcription factors with core RNA-polymerase and for epitope mapping of monoclonal-antibodies. *J. Biol. Chem.*, **266**: 2632-2638.
- Liepinsh, E., Barbals, R., Dahl, E., Sharipo, A., Staub, E. and Otting, G. (2003) The death-domain fold of the ASC PYRIN domain, presenting a basis for PYRIN/PYRIN recognition. *J. Mol. Biol.*, **332**: 1155-1163.
- Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. and Vendruscolo, M. (2005) Simultaneous determination of protein structure and dynamics. *Nature*, **433**: 128-132.

- Lipari, G., Szabo, A. and Levy, R. M. (1982) Protein dynamics and NMR relaxation - comparison of simulations with experiment. *Nature*, **300**: 197-198.
- Liu, Z. H., Sun, C. H., Olejniczak, E. T., Meadows, R. P., Betz, S. F., Oost, T., Herrmann, J., Wu, J. C. and Fesik, S. W. (2000) Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature*, **408**: 1004-1008.
- Mao, X. A. and Ye, C. H. (1997) Understanding radiation damping in a simple way. *Concept Magnetic Res.*, **9**: 173-187.
- Masumoto, J., Taniguchi, S., Ayukawa, K., Sarvotham, H., Kishino, T., Niikawa, N., Hidaka, E., Katsuyama, T., Higuchi, T. and Sagara, J. (1999) ASC, a novel 22-kDa protein, aggregates during apoptosis of human promyelocytic leukemia HL-60 cells. *J. Biol. Chem.*, **274**: 33835-33838.
- McConnell, B. B. and Vertino, P. M. (2004) TMS1/ASC: The cancer connection. *Apoptosis*, **9**: 5-18.
- McEntyre, J. R. and Gibson, T. J. (2004) Patterns and clusters within the PSM column in TiBS, 1992-2004. *Trends Biochem. Sci.*, **29**: 627-633.
- Michel-Reydellet, N., Woodrow, K. and Swartz, J. (2005) Increasing PCR fragment stability and protein yields in a cell-free system with genetically modified *Escherichia coli* extracts. *J. Mol. Microb. Biotech.*, **9**: 26-34.
- Miranker, A., Radford, S. E., Karplus, M. and Dobson, C. M. (1991) Demonstration by NMR of folding domains in lysozyme. *Nature*, **349**: 633-636.
- Murali, N. (2006) IDEAL - A fast single scan method for X pulse width calibration. *J. Magn. Reson.*, **183**: 142-144.
- Narajan, A., Ghose, R., Jill, J. M. (2006) Structure and dynamics of ASC2, a pyrin domain-only protein that regulates inflammatory signaling. *J. Biol. Chem.*, **281**:31863-31867.
- Neri, D., Szyperski, T., Otting, G., Senn, H. and Wüthrich, K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434-repressor by biosynthetically directed fractional <sup>13</sup>C labeling. *Biochemistry*, **28**: 7510-7516.
- Nielsen, N. C., Bildsoe, H., Jakobsen, H. J. and Sørensen, O. W. (1987) Two-dimensional pulse techniques for determination of radiofrequency field strengths and proton multiplicities in NMR spectroscopy. *J. Am. Chem. Soc.*, **109**: 901-902.
- Nirenberg, M. and Matthaei, J. H. (1961) Dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA*, **47**: 1588-&.
- Ogg, R. J., Kingsley, P. B. and Taylor, J. S. (1994) WET, a T-1 insensitive and B-1 insensitive water suppression method for in-vivo localized <sup>1</sup>H NMR spectroscopy. *J. Magn. Reson. B*, **104**: 1-10.
- Ohuchi, S., Nakano, H. and Yamane, T. (1998) In vitro method for the generation of protein libraries using PCR amplification of a single DNA molecule and coupled transcription/translation. *Nucleic Acids Res.*, **26**: 4339-4346.
- Oltersdorf, T., Elmore, S. W., Shoemaker, A. R., Armstrong, R. C., Augeri, D. J., Belli, B. A., Bruncko, M., Deckwerth, T. L., Dinges, J., Hajduk, P. J., Joseph, M. K., Kitada, S., Korsmeyer, S. J., Kunzer, A. R., Letai, A., Li, C., Mitten, M. J., Nettesheim, D. G., Ng, S., Nimmer, P. M., O'Connor, J. M., Oleksijew, A., Petros, A. M., Reed, J. C., Shen, W., Tahir, S. K., Thompson, C. B., Tomaselli, K. J.,

- Wang, B. L., Wendt, M. D., Zhang, H. C., Fesik, S. W. and Rosenberg, S. H. (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*, **435**: 677-681.
- Ozawa, K., Dixon, N. E. and Otting, G. (2005) Cell-free synthesis of N-15 labeled proteins for NMR studies. *IUBMB Life*, **57**: 615-622.
- Ozawa, K., Headlam, M. J., Schaeffer, P. M., Henderson, B. R., Dixon, N. E. and Otting, G. (2004) Optimization of an Escherichia coli system for cell-free synthesis of selectively <sup>15</sup>N-labelled proteins for rapid analysis by NMR spectroscopy. *Eur. J. Biochem.*, **271**: 4084-4093.
- Park, H. H., Lo, Y. C., Lin, S. C., Wang, L., Yang, J. K. and Wu, H. (2007) The death domain superfamily in intracellular signaling of apoptosis and inflammation. *Annu. Rev. Immunol.*, **25**: 561-586.
- Parker, M. J., Aulton-Jones, M., Hounslow, A. M. and Craven, C. J. (2004) A combinatorial selective labeling method for the assignment of backbone amide NMR resonances. *J. Am. Chem. Soc.*, **126**: 5020-5021.
- Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA*, **94**: 12366-12371.
- Piotto, M., Saudek, V. and Sklenâr, V. (1992) Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions. *J. Biomol. NMR*, **2**: 661-665.
- Plateau, P. and Guéron, M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.*, **104**: 7310-7311.
- Pratt, J. M., Boulnois, G. J., Darby, V., Orr, E., Wahle, E. and Holland, I. B. (1981) Identification of gene-products programmed by restriction endonuclease DNA fragments using an Escherichia coli in-vitro system. *Nucleic Acids Res.*, **9**: 4459-4474.
- Purcell, E. M., Torrey, H. C. and Pound, R. V. (1946) Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, **69**: 37-38.
- Qin, H. X., Srinivasula, S. M., Wu, G., Fernandes-Alnemri, T., Alnemri, E. S. and Shi, Y. G. (1999) Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature*, **399**: 549-557.
- Rodriguez, J. C., Wilks, A. and Rivera, M. (2006) Backbone NMR assignments and H/D exchange studies on the ferric azide- and cyanide-inhibited forms of Pseudomonas aeruginosa heme oxygenase. *Biochemistry*, **45**: 4578-4592.
- Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T. and Kay, L. E. (1996) Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.*, **263**: 627-636.
- Sadqi, M., Fushman, D. and Munoz, V. (2006) Atom-by-atom analysis of global downhill protein folding. *Nature*, **442**: 317-321.
- Schaeffer, P. M., Headlam and M. J., Dixon, N. E. (2005) Protein-protein interactions in the eubacteria replisome. *IUBMB Life*, **57**: 5-12.
- Schubert, M., Smalla, M., Schmieder, P. and Oschkinat, H. (1999) MUSIC in triple-resonance experiments: Amino acid type selective <sup>1</sup>H-<sup>15</sup>N correlations. *J. Magn. Reson.*, **141**: 34-43.

- Seki, E., Matsuda, N., Yokoyama, S. and Kigawa, T. (2008) Cell-free protein synthesis system from *Escherichia coli* cells cultured at decreased temperatures improves productivity by decreasing DNA template degradation. *Anal. Biochem.*, **377**: 156-161.
- Shortle, D. (1994) Assignment of amino-acid type in  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra by labeling with  $^{14}\text{N}$  amino acids. *J. Magn. Reson. B*, **105**: 88-90.
- Sitaraman, K., Esposito, D., Klarmann, G., Le Grice, S. F., Hartley, J. L. and Chatterjee, D. K. (2004) A novel cell-free protein synthesis system. *J. Biotechnol.*, **110**: 257-263.
- Sklenâr, V., Piotto, M., Leppik, R. and Saudek, V. (1993) Gradient-tailored water suppression for  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. A*, **102**: 241-245.
- Smallcombe, S. H., Patt, S. L. and Keifer, P. A. (1995) WET solvent suppression and its applications to LC NMR and high-resolution NMR spectroscopy. *J. Magn. Reson. A*, **117**: 295-303.
- Sørensen, O. W., Eich, G. W., Levitt, M. H., Bodenhausen, G. and Ernst, R. R. (1983) Product operator formalism for the description of NMR pulse experiments. *Prog. Nucl. Mag. Res. Sp.*, **16**: 163-192.
- Sørensen, O. W., Nielsen, N. C., Bildsoe, H., Jakobsen, H. J. (1986) Determination of radiofrequency field strengths and proton multiplicities by two-dimensional pulse techniques. *J. Magn. Reson.*, **70**: 54-70.
- Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y. and Alakhov, Y. B. (1988) A continuous cell-free translation system capable of producing polypeptides in high-yield. *Science*, **242**: 1162-1164.
- Srimathi, T., Robbins, S. L., Dubas, R. L., Chang, H., Cheng, H., Roder, H. and Park, Y.C. (2008) Mapping of POP1-binding site on pyrin domain of ASC. *J. Biol. Chem.*, **283**: 15290-15298.
- Stehlik, C., Krajewska, M., Welsh, K., Krajewski, S., Godzik, A., Reed, J. C. (2003) The PAAD/PYRIN-only protein POP1/ASC2 is a modulator of ASC-mediated nuclear-factor- $\kappa\text{B}$  and pro-caspase-1 regulation, **373**: 101-113.
- Styles, P., Soffe, N. F., Scott, C. A., Cragg, D. A., Row, F., White, D. J. and White, P. C. J. (1984) A high-Resolution NMR probe in which the coil and preamplifier are cooled with liquid helium. *J. Magn. Reson.*, **60**: 397-404.
- Su, X. C., Jergic, S., Keniry, M. A., Dixon, N. E. and Otting, G. (2007) Solution structure of domains IVa and V of the tau subunit of *Escherichia coli* DNA polymerase III and interaction with the alpha subunit. *Nucleic Acids Res.*, **35**: 2825-2832.
- Tang, C., Iwahara, J. and Clore, G. M. (2006) Visualization of transient encounter complexes in protein-protein association. *Nature*, **444**: 383-386.
- Tashiro, M., Rios, C. B. and Montelione, G. T. (1995) Classification of amino-acid spin systems using PFG HCC(CO)NH-TOCSY with constant-time aliphatic C-13 frequency labeling. *J. Biomol. NMR*, **6**: 211-216.
- Tugarinov, V., Choy, W. Y., Orekhov, V. Y. and Kay, L. E. (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. USA*, **102**: 622-627.

- Tugarinov, V., Kanelis, V. and Kay, L. E. (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.*, **1**: 749-754.
- Wider, G., Hosur, R. V. and Wüthrich, K. (1983) Suppression of the solvent resonance in 2D NMR spectra of proteins in H<sub>2</sub>O solution. *J. Magn. Reson.*, **52**: 130-135.
- Williamson, M. P., Havel, T. F. and Wüthrich, K. (1985) Solution conformation of proteinase inhibitor-Lia from bull seminal plasma by <sup>1</sup>H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.*, **182**: 295-315.
- Wright, M., Buttin, G. and Hurwitz, J. (1971) Isolation and characterization from *Escherichia coli* of an adenosine triphosphate-dependent deoxyribonuclease directed by Rec B, C genes. *J. Biol. Chem.*, **246**: 6543-&.
- Wüthrich, K. (1986). NMR of proteins and nucleic acids. New York, John Wiley & Sons.
- Wüthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) Sequential resonance assignments as a basis for determination of spatial by high resolution nuclear magnetic resonance. *J. Mol. Biol.*, **155**: 311-319.
- Xiao, T., Towb, P., Wasserman, S. A. and Sprang, S. R. (1999) Three-dimensional structure of a complex between the death domains of Pelle and Tube. *Cell*, **99**: 545-555.
- Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H. and Nagayama, K. (1991) Assignments of backbone <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N resonances and secondary structure of ribonuclease-H from *Escherichia coli* by heteronuclear 3-dimensional NMR-spectroscopy. *Biochemistry*, **30**: 6036-6047.
- Yang, H. L., Ivashkiv, L., Chen, H. Z., Zubay, G. and Cashel, M. (1980) Cell-free coupled transcription-translation system for investigation of linear DNA segments. *Proc. Natl. Acad. Sci. USA*, **77**: 7029-7033.
- Yang, J. K., Wang, L. W., Zheng, L. X., Wan, F. Y., Ahmed, M., Lenardo, M. J. and Wu, H. (2005) Crystal structure of MC159 reveals molecular mechanism of DISC assembly and FLIP inhibition. *Mol. Cell.*, **20**: 939-949.
- Yates, J. L., Arfsten, A. E. and Nomura, M. (1980) In-vitro expression of *Escherichia coli* ribosomal-protein genes - Autogenous inhibition of translation. *Proc. Natl. Acad. Sci. USA*, **77**: 1837-1841.
- Yokoyama, S. (2003) Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.*, **7**: 39-43.
- Yu, M., Souaya, J., Julin, D. (1998) The 30-kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **95**: 981-986.
- Zhou, P., Chou, J., Olea, R. S., Yuan, J. Y. and Wagner, G. (1999) Solution structure of Apaf-1 CARD and its interaction with caspase-9 CARD: A structural basis for specific adaptor/caspase interaction. *Proc. Natl. Acad. Sci. USA*, **96**: 11265-11270.

## 8. Contribution to papers

I made the following contributions to the papers.

**Paper I:** I expressed and purified wild-type ASC PyD plus the 5 mutants K21A, L25A, R41A, D48A and D51A, and performed the NMR experiments with these samples. I contributed to writing the Experimental section of the paper.

**Paper II:** I participated in the experimental design and performed the relevant experiments leading to the paper. I assisted with the interpretation of the results, wrote the first draft of the paper and prepared the figures.

**Paper III:** I performed the experiments leading to the paper. I carried out the mathematical modelling and simulations, and assisted with the interpretation of the results. I wrote the first draft of the paper and prepared the figures.

**Paper IV:** I provided intellectual input and prepared figures I and II.

**Paper V:** I was involved in the experimental design and performed all the experiments and data analysis in the paper. I also contributed to the interpretation of the results, wrote the first draft of the paper and prepared the figures.

**Paper VI:** I lead the experimental design and performed the experiments leading to the paper. I also contributed to the interpretation of the results, wrote the first draft of the paper and prepared the figures.

## 9. Appendix: I (VI)

# Paper I

M. Moriya, S. Taniguchi, **P. Wu**, E. Liepinsh, G. Otting, J. Sagara  
"Role of charged and hydrophobic residues in the oligomerization of the  
PYRIN domain of ASC"

Reproduced with permission from  
*Biochemistry* **44**, 575-583 (2005)  
Copyright 2005 The American Chemical Society

## Role of Charged and Hydrophobic Residues in the Oligomerization of the PYRIN Domain of ASC<sup>†</sup>

Mie Moriya,<sup>§</sup> Shunichiro Taniguchi,<sup>§</sup> Peter Wu,<sup>||</sup> Edwards Liepinsh,<sup>⊥</sup> Gottfried Otting,<sup>||</sup> and Junji Sagara<sup>\*§</sup>

*Molecular Oncology, Institute on Aging and Adaptation, Shinshu University Graduate School of Medicine, 3-1-1 Asahi, 390-8621, Japan, Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia, and Department of Medical Biochemistry and Biophysics, Karolinska Institute, S-17177 Stockholm, Sweden*

*Received July 30, 2004; Revised Manuscript Received October 19, 2004*

**ABSTRACT:** Apoptosis-associated speck-like protein containing a caspase recruitment domain (ASC) is an adaptor protein composed of two homophilic protein–protein interaction domains, a PYRIN domain (PYD) and a caspase recruitment domain. PYD-dependent oligomerization of ASC is thought to play a crucial role in formation of a molecular platform, the inflammasome, which activates caspase-1. When expressed in cells, the PYD of ASC was shown to form cytoplasmic filaments through self-association. Over 70 single point mutants were analyzed for filament formation in cells expressing the mutant proteins. The set of mutations comprised every single amino acid residue with a charged side chain (Arg, Lys, Asp, and Glu) and a large hydrophobic side chain (Ile, Leu, Met, Phe, Pro, and Val). Filament formation of the ASC PYD was prevented by mutation of Lys21, Leu25, Lys26, Pro40, Arg41, Asp48, and Asp51 of helices 2, 3, and 4. These data identify a coherent interaction surface, establishing a molecular model of PYD–PYD complexes with an important role for charge–charge interactions.

We previously identified ASC<sup>1</sup> as a proapoptotic protein containing a caspase recruitment domain (CARD) that formed a large aggregate (called a “speck”) in apoptotic cells (1). ASC (also termed TMS1 or Pycard) is an adaptor protein composed of two protein–protein interaction domains, an N-terminal PYRIN domain (PYD), and a C-terminal CARD, and has been implicated in apoptosis, inflammation, and cancer (2–7). The PYD, also called PAAD or DAPIN, is a new member of the six-helix bundle death domain-fold superfamily that includes the death domain (DD), death effector domain (DED), and CARD (8). In humans, more than 30 PYD-containing proteins have been identified, including the causative gene products of autoinflammatory diseases, pyrin and cryopyrin (9–11).

ASC interacts with the proform of caspase-1 via CARD–CARD interactions and stimulates its conversion to its active form (12–15). The induced proximity model for apoptosis signaling pathways is based on evidence that adaptor protein-induced oligomerization of initiator caspases is essential for their proteolytic activation (16, 17). It is thought that PYD-

mediated oligomerization of ASC is a critical prerequisite for assembly of the signal complex activating caspase-1. Therefore, it is important to determine the PYD structure and the critical residues for PYD oligomerization. Recently, the PYD structures of Nalp1 and ASC have been determined by NMR spectroscopy (18, 19). These structures allowed an improved sequence alignment between different PYDs, identifying a number of conserved surface residues that may play a role for a conserved mode of oligomerization. The conservation of PYD surface residues is, however, incomplete, and experimental verification has been missing.

Overexpression of the ASC PYD in cells has been shown to result in filamentous structures in the cytoplasm through self-oligomerization (4, 5). Such cytoplasmic filament-like structures have been observed for many domains of the death domain superfamily (20–23). For example, overexpression of the DEDs of FADD and caspase-8 results in cytoplasmic filaments, and the CARDS of caspase-2 and mE10 have also been shown to form intracellular filaments through homophilic interactions.

In the present study, we used the generation of filaments to determine the role of each amino acid of the ASC PYD in PYD–PYD interactions. No other assay is currently available that would allow efficient assessment of the activity of a large number of ASC PYD mutants. Seventy-two single point mutants of the charged (Asp, Glu, Arg, and Lys) and hydrophobic (Leu, Ile, Met, Val, Phe, and Pro) residues of the ASC PYD were prepared and analyzed with respect to production of PYD filaments. Combined with the solvent exposure of the respective amino acid side chains in the three-dimensional structure, a continuous surface area with positively and negatively charged residues was identified, which seems to be crucial for oligomerization of the ASC PYD. The data further support the notion that the R42W

<sup>†</sup> This work was supported in part by the Japan Society for the Promotion of Science (16021218) and the Australian Research Council for a Federation Fellowship (G.O.).

\* To whom correspondence should be addressed. Telephone: +81-263-37-2723; Fax: +81-263-37-2724. E-mail: sagara@sch.md.shinshu-u.ac.jp.

<sup>§</sup> Shinshu University Graduate School of Medicine.

<sup>||</sup> Australian National University.

<sup>⊥</sup> Karolinska Institute.

<sup>1</sup> Abbreviations: ASC, apoptosis-associated speck-like protein containing a CARD; CARD, caspase recruitment domain; DD, death domain; DED, death effector domain; PYRIN domain, pyrin N-terminal homology domain; PYD, PYRIN domain; PAAD, pyrin, AIM2, ASC and death domain-like; DAPIN, domain in apoptosis and interferon response; GFP, green fluorescence protein; NMR, nuclear magnetic resonance.

mutation in the pyrin PYD associated with familial Mediterranean fever (FMF) interferes with intermolecular protein–protein interactions.

## MATERIALS AND METHODS

**Expression Plasmids and Mutations.** ASC cDNA encoding the full-length polypeptide (195 residues) was inserted into pcDNA3 (Invitrogen) and pEGFP-c2 (Clontech) as described (1, 4). Constructs encoding GFP-tagged truncated ASC mutants (amino acids 1–100, 1–95, 1–90, and 1–85) were produced by PCR using pEGFP-c2 ASC (amino acids 1–195) as the template (24). The primers were designed according to the human ASC mRNA sequence (GenBank Accession No. AB023416). In this study, ASC (amino acids 1–90) is referred to as ASC PYD. A construct encoding C-terminally HA-tagged ASC PYD was produced by PCR using the template plasmid pcDNA3 ASC (amino acids 1–195) and two primers encoding the HA tag sequence (YPYDVPDYA). An N-terminally Flag-tagged ASC PYD construct was made by insertion of the PCR product (amino acids 1–100) into the EcoRI and Sall sites of pFLAG-CMV-4 (Sigma) as described (4). Point mutations were introduced into pEGFP-c2 ASC PYD by PCR using primer sets that included the mutations. Mutations in the ASC PYD were confirmed by sequencing. The Expand High-Fidelity system (Roche Molecular Biochemicals) was used for the PCR. The restriction enzymes, T4 polynucleotide kinase, T4 DNA ligase, Klenow Fragment, and M-MLV reverse transcriptase were purchased from Takara Biotechnology (Tokyo, Japan).

**Expression and Observation of GFP-Tagged ASC PYD Filaments.** Cos7 cells were cultured in Dulbecco's modified minimal essential medium supplemented with 10% fetal bovine serum. One day before transfection, Cos7 cells were seeded on a Lab-Tek 8-chambered coverglass (Nalge Nunc International). A total of 0.1–0.2  $\mu\text{g}$  of pEGFP-c2 ASC PYD and 0.3  $\mu\text{L}$  of the transfection reagent Fugene 6 (Roche Molecular Biochemicals) were mixed and added to cell cultures (0.25 mL of medium) according to the manufacturer's instructions. Fluorescent cells were observed from 12 to 72 h after transfection in the living state using a fluorescent microscope, Axiovert S100 (Carl Zeiss Corporation). Expression levels were measured by Western Blotting as follows. One day before the transfection, Cos7 cells were seeded on 24-well culture plates (Becton Dickinson Labware). A total of 0.3  $\mu\text{g}$  of plasmid DNA and 0.6  $\mu\text{L}$  of Fugene 6 were mixed and added to the cell cultures (0.5 mL of medium) according to the manufacturer's instructions. Twenty-four hours after transfection, the cells were lysed in 0.1 mL of a 1% SDS solution and boiled for 5 min. The cell lysates (20  $\mu\text{g}$  of protein) were electrophoresed in a 10% SDS polyacrylamide gel and electroblotted on a PVDF membrane (Millipore). After the blots were blocked in Tris-buffered saline containing 5% nonfat skim milk, they were incubated with rabbit anti-GFP antiserum (Santa Cruz Biotechnology) and finally detected with the ECL Western blotting detection system (Amersham). The percentage of PYD filament-producing cells was determined by the number of PYD filament-producing cells divided by the total number of fluorescent cells. For each experiment, more than 200 fluorescent cells from randomly chosen fields were counted. The PYD filament-positive mutants usually produced fila-

mentous structures in over 10% of the fluorescent cells, while the negative mutants produced filaments in none or much less than 1% of the fluorescent cells.

**Immunostaining.** One day before transfection, Cos7 cells were seeded on a coverglass in 30-mm culture dishes. A total of 1  $\mu\text{g}$  of plasmid DNA and 2  $\mu\text{L}$  of Fugene 6 (Roche Molecular Biochemicals) were mixed and added to cell cultures (2 mL of medium). Twenty-four or forty-eight hours after transfection, cells were fixed with 3.7% formaldehyde for 15 min at room temperature. The fixed cells were permeabilized with 0.2% Triton X-100 and then soaked for 15 min with Tris-buffered saline containing 1% bovine serum albumin and 2% fetal bovine serum at room temperature. The filaments of HA- and Flag-tagged ASC PYDs were stained with rabbit anti-HA and anti-Flag antibodies, respectively (Santa Cruz Biotechnology). A rhodamine-conjugated goat anti-rabbit immunoglobulin antibody (Amersham Biosciences) was used as the secondary antibody. Antibodies against tubulin, actin, vimentin, and cytokeratins were purchased from Sigma.

**Expression of ASC PYD Mutants in *E. coli*.** The wild-type and a set of mutant ASC PYD genes were inserted into the phage T7 promoter vector pETMCSIII (25) between the NdeI and EcoRI restriction sites for subsequent expression with a N-terminal (His)<sub>6</sub> tag. The nucleotide sequences were verified using an ABI 3730 sequencer. The proteins were produced in the *Escherichia coli* strain BL21(DE3)/pLysS (26). The cells were grown at 37 °C in Luria-Bertani broth containing ampicillin (50  $\mu\text{g}/\text{mL}$ ) and chloramphenicol (33  $\mu\text{g}/\text{mL}$ ) and gene expression induced by addition of isopropyl- $\beta$ -D-thiogalactoside (0.5 mM) when the OD<sub>595</sub> had reached between 0.5 and 0.8. After further growth for 3 h, the cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Na phosphate at pH 7.7, 300 mM NaCl, 2 mM  $\beta$ -mercaptoethanol) containing spermidine (10 mM), and lysed in a French press. Solubility was determined by comparing the amount of protein in the cell lysate with that in the supernatant fraction on an SDS–PAGE gel. For purification, His<sub>6</sub>-tagged soluble recombinant proteins from the supernatants were batch loaded onto Ni-NTA (Qiagen) affinity resins and washed with 20 column volumes of lysis buffer containing 25 mM imidazole, followed by elution in lysis buffer containing 250 mM imidazole.

**NMR Measurements.** Purified ASC PYD mutants were exchanged into pH 7.2 phosphate buffer (10 mM Na phosphate, 50 mM NaCl, 0.2% NaN<sub>3</sub>). D<sub>2</sub>O was added to 10% v/v prior to NMR analysis. One-dimensional <sup>1</sup>H NMR spectra were recorded on a Bruker Avance 800 MHz NMR spectrometer at 25 °C.

**Solvent Accessibility.** Solvent accessibilities of the side chains were calculated with a probe of 1.4 Å radius and compared with the hypothetical accessibility of a fully extended side chain in a helix, where both neighboring residues are Gly (27).

## RESULTS

**PYD Filaments.** To determine the minimal segment of ASC required for PYD filament formation, different C-terminal deletion mutants (amino acids 1–100, 1–95, 1–90, and 1–85) of ASC were expressed as fusion proteins with GFP in Cos7 cells. Except for the deletion mutant including

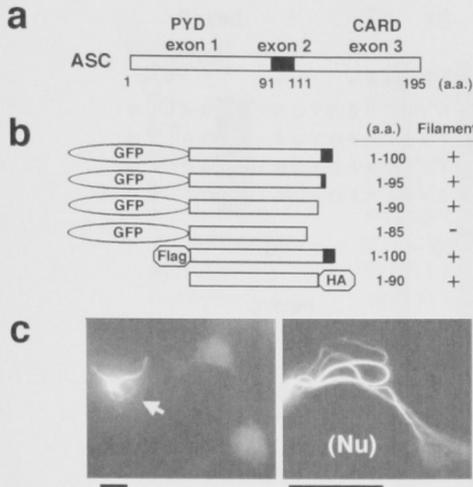


FIGURE 1: PYD filaments. (a) The domain structure of ASC. ASC is encoded by three exons and the PYD resides in exon 1. (b) Filament formation of GFP-tagged and HA-tagged PYD of ASC. ASC deletion mutants lacking the CARD were fused to GFP, Flag, or HA as indicated and expressed in Cos7 cells. Filament formation of GFP-tagged PYD was observed in vivo by fluorescence microscopy. Flag-tagged and HA-tagged PYD filaments were observed by immunostaining after fixation. All filament-positive (+) constructs produced PYD filaments in more than 20% of cells expressing each construct. Filament-negative (-) PYD constructs showed no filamentous structure. (c) Morphology of PYD filaments. GFP-tagged ASC PYD (residues 1-90) was expressed in Cos7 cells and photographed 24 h after transfection of the plasmid. Left panel: example of a living cell producing GFP-tagged PYD filaments (arrow) as seen by fluorescence microscopy. Right panel: photograph at a higher magnification. (Nu): nucleus. Bars: 10  $\mu$ m.

amino acids 1-85 of ASC, all the mutants produced PYD filaments (Figure 1). PYD filaments were observed in more than 10% of cells expressing the GFP-tagged ASC PYD 24 h after transfection. This result demonstrates that the polypeptide comprising amino acids 1-90 is sufficient for the filament formation.

Since GFP (27 kDa) is larger than the ASC PYD (~10 kDa), the large GFP tag could conceivably affect oligomerization of the ASC PYD. Therefore, we also explored the use of two small tags, Flag and HA, instead of GFP. When fused to a N-terminal Flag or a C-terminal HA (Figure 1), both Flag-tagged and HA-tagged ASC PYDs produced filamentous structures 24 h after transfection in more than 20% of the cells expressing them. The morphology of both Flag-tagged and HA-tagged PYD filaments was indistinguishable from that of GFP-tagged PYD filaments. These observations suggested that GFP did not interfere with PYD filament formation. The GFP-tagged ASC PYD (amino acids 1-90) was subsequently used for the point mutation study.

**PYD Filaments and Cytoskeletons.** The pattern of ASC PYD filaments resembles that of cytoskeletal structures. To examine the relationship of ASC PYD filaments with the cytoskeleton, we immunostained cells expressing ASC PYD with anti-tubulin, anti-actin, or anti-vimentin antibodies. We also used anti-cytokeratin antibodies recognizing cytokeratins 1, 4, 5, 6, 8, 10, 13, 18, and 19. None of these cytoskeletal markers were co-localized with ASC PYD filaments (data not shown). Although we cannot rule out association of ASC PYD filaments with some unknown cytoskeletal component, PYD filament formation seems to be driven by self-

Table 1: Filament Formation by ASC PYDs Containing Mutations of Charged Residues<sup>a</sup>

mutant	filament	solvent accessibility (%)	mutant	filament	solvent accessibility (%)
R3A	+	69	R41A	-	46
R5A	+	30	R41Q	-	
D6A	+	50	R41K	+	
D10A	+	64	R41W <sup>c</sup>	-	
E13A	+	60	D48A	-	57
E18A	+	68	D48N	-	
E19A	-	39	D48E	-	
E19Q	+		D48R <sup>b</sup>	-	
K21A	-	48	D51A	-	53
K21Q	-		D51N	-	
K21R	+		D51E	+	
K21E <sup>b</sup>	-		D51K <sup>b</sup>	-	
K22A	+	49	K54A	+	49
K24A	+	8	K55A	+	40
K26A	-	28	E62A	-	47
K26Q	-		E62Q	+	
K26R	+		E67A	-	61
R33A	+	82	E67Q	+	
E34A	+	98	R74A	+	69
R38A	+	41	D75A	+	32
			E80A	+	71

<sup>a</sup> GFP-tagged ASC PYD mutants were expressed in Cos7 cells and their filament formations were examined. Positive mutants (+) showed filamentous structures in more than 10% of the fluorescent cells, while negative mutants (-) showed no filamentous structures. The solvent accessibility reports the accessibility of the side chain of the unmutated residue X in the three-dimensional structure of the ASC PYD compared with the solvent accessibility in a helical Gly-X-Gly peptide with a fully extended side chain. <sup>b</sup> This mutation introduces a residue present in the protein pyrin (Figure 3). <sup>c</sup> The corresponding R42W mutation of pyrin is a cause of FMF disease (31).

association of ASC PYDs rather than association with the cytoskeleton.

**Mutations of Charged Residues.** On the basis of the pronounced electrostatic dipole moment present in the structure of the ASC PYD and the important role of charged residues played in complexes between DDs and CARDS (28, 29), charged residues are likely to be of great importance also for PYD-PYD interactions (5, 18, 19, 30). To verify this prediction, we replaced all the residues with charged side chains in the PYD of human ASC with alanine and examined whether these alanine mutants produced filamentous structures in transfected cells. Filament formation of each mutant was tested at least five times by using various amounts of plasmid DNAs for transfection and the expression levels measured with fluorescence intensity and immunoblotting. Immunoblotting analyses of total cell extracts of the transfected cells showed that all the ASC PYD mutants were expressed at almost the same levels as the wild-type ASC PYD. The mutants E19A, K21A, K26A, R41A, D48A, D51A, E62A, and E67A produced no filamentous structures (Table 1). To assess the importance of a negative charge at these positions, we subsequently generated eight point mutations, where these amino acid residues were substituted with glutamine (Q) or asparagine (N) (Table 1). The mutants E19Q, E62Q, and E67Q were found to produce PYD filaments. Clearly, the charge of these residues is not required for filament formation, suggesting that their mutation to alanine compromised filament formation by disturbing the structure of the domain. In contrast, the mutants K21Q, K26Q, R41Q, D48, and D51N did not show filaments,

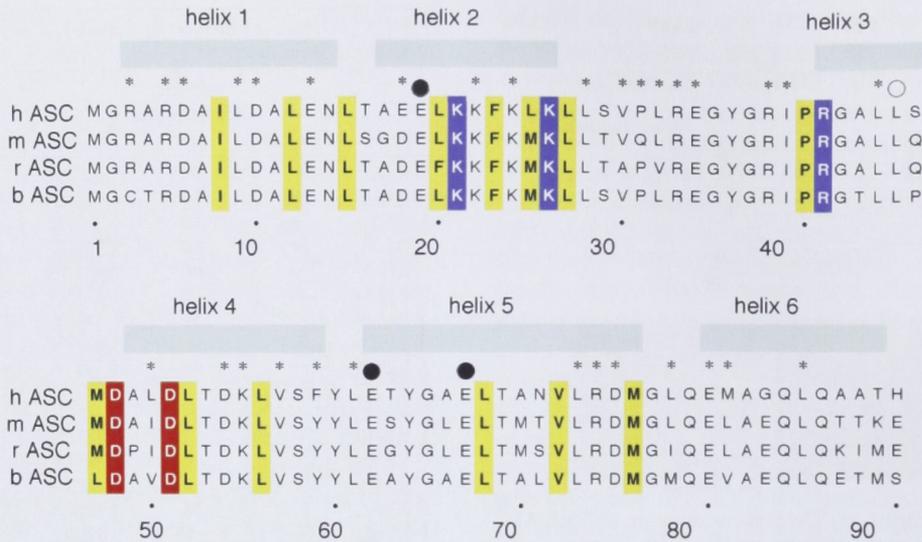


FIGURE 2: Amino acid sequence of ASC PYDs and overview of the present mutation analysis of the human ASC PYD. The PYD sequences of human (h), mouse (m), rat (r), and bovine (b) ASC are aligned. Asterisks identify charged and hydrophobic residues retaining the ability of filament formation upon mutation to alanine. The following colors were used for the amino acid residues where filament formation was prevented by mutation to alanine: blue, positively charged residues; red, negatively charged residues; yellow, hydrophobic residues. Filled circles above the alignment identify glutamic acid residues where the size of the side chain seems to be more important than its charge, as mutation to alanine and glutamine abolishes and retains the capability of filament formation, respectively. An open circle marks L45. Although this residue is solvent-exposed and mutation to alanine suppressed filament formation, it does not seem to be directly involved in intermolecular interactions, since filament formation was retained after mutations to larger residues, including hydrophilic and charged residues (Table 1). Bars at the top mark the helices determined by NMR spectroscopy (18).

indicating that the charges provided by the side chains of K21, K26, R41, D48, and D51 play critical roles in PYD filament formation. To confirm this hypothesis, we substituted each charged residue with another amino acid of the same charge. The K21R, K26R, R41K, and D51E mutants produced PYD filaments, highlighting the importance of charge at these positions, whereas the D48E mutant did not. Table 1 and Figure 2 provide a summary of these results. In the case of D48, it appears that both the size and negative charge of its side chain are critical for PYD filament formation in human ASC despite high solvent exposure (Table 1). Interestingly, sequence alignment of human and viral PYD proteins shows that, unlike D51, D48 is never substituted with E in any of the PYDs (Figure 3).

**Comparison of the Charged Residues of ASC with Those of Pyrin.** Of the five charged residues identified as critical for filament formation in the PYD of ASC, only K26 and R41 are conserved in the PYD of pyrin (Figure 3). To check whether correct charges are sufficient for filament formation, K21, D48, and D51 of ASC were substituted with the charged residues E, R, and K, respectively, present in the PYD of pyrin (Figure 3). As shown in Table 1, the K21E, D48R, and D51K mutants produced no filamentous structures. FMF is caused by point mutations in pyrin. One of these FMF-associated mutations, R42W, is located in the PYD of pyrin (31). This residue corresponds to R41 in ASC. R41W mutant of ASC PYD produced no filamentous structures in cells.

**Mutations of Hydrophobic Residues.** To probe the possible role of hydrophobic interactions in PYD–PYD interactions, all residues with large hydrophobic side chains were replaced by alanine in single point mutations except for the N-terminal methionine. Fifteen mutants (I8A, L12A, L15A, L20A, F23A, L25A, L27, P40, L45A, M47, L52, L56, L68, V72, and M76) of a total of 30 mutants lost the ability to form

filamentous structures (Table 1 and Figure 2). Immunoblotting analyses of total cell extracts of the transfected cells confirmed that all the ASC PYD mutants were expressed at almost the same levels as the wild-type ASC PYD (data not shown). Most of the filament-deficient mutants displayed a diffuse cytoplasmic and nuclear distribution.

The NMR structure of the PYD of ASC shows that many of the hydrophobic residues contribute to the hydrophobic core of the three-dimensional structure, displaying very little solvent exposure. Mutations of these residues to alanine are thus likely to prevent filament formation by affecting the three-dimensional structure of the domain. For example, F23 in helix 2 is completely buried in the PYD of ASC (Table 1) and its conservation among PYD proteins (Figure 3) suggests that it is structurally important. Disruption of filament formation observed for the F23A mutant may thus be explained by a disturbed protein structure, whereas the F23L mutant seems to be structurally sufficiently conserved to maintain filament formation (Table 2).

Disruption of filament formation upon mutation to alanine was observed for three hydrophobic residues, which display at least 30% solvent accessibility in the NMR structure of the ASC PYD (L25, P40, L45; Table 2), indicating their potential involvement in PYD–PYD interactions. P40 is spatially close to R41, which is also critical for filament formation (Table 1). P40 is thus likely to be involved in PYD–PYD interactions. L25 and, especially, L45 are only poorly conserved in PYDs (Figure 3). Therefore, additional mutants were made for these two residues to assess their importance in PYD–PYD interactions (Table 2). Hydrophobic mutants (M, I, V, F) of L25 produced filamentous structures. In contrast, filament formation was suppressed after mutation to glycine (G) or hydrophilic residues (Q, N, K, E), indicating that filament formation depends on a large hydrophobic side chain at this site. This pattern would be

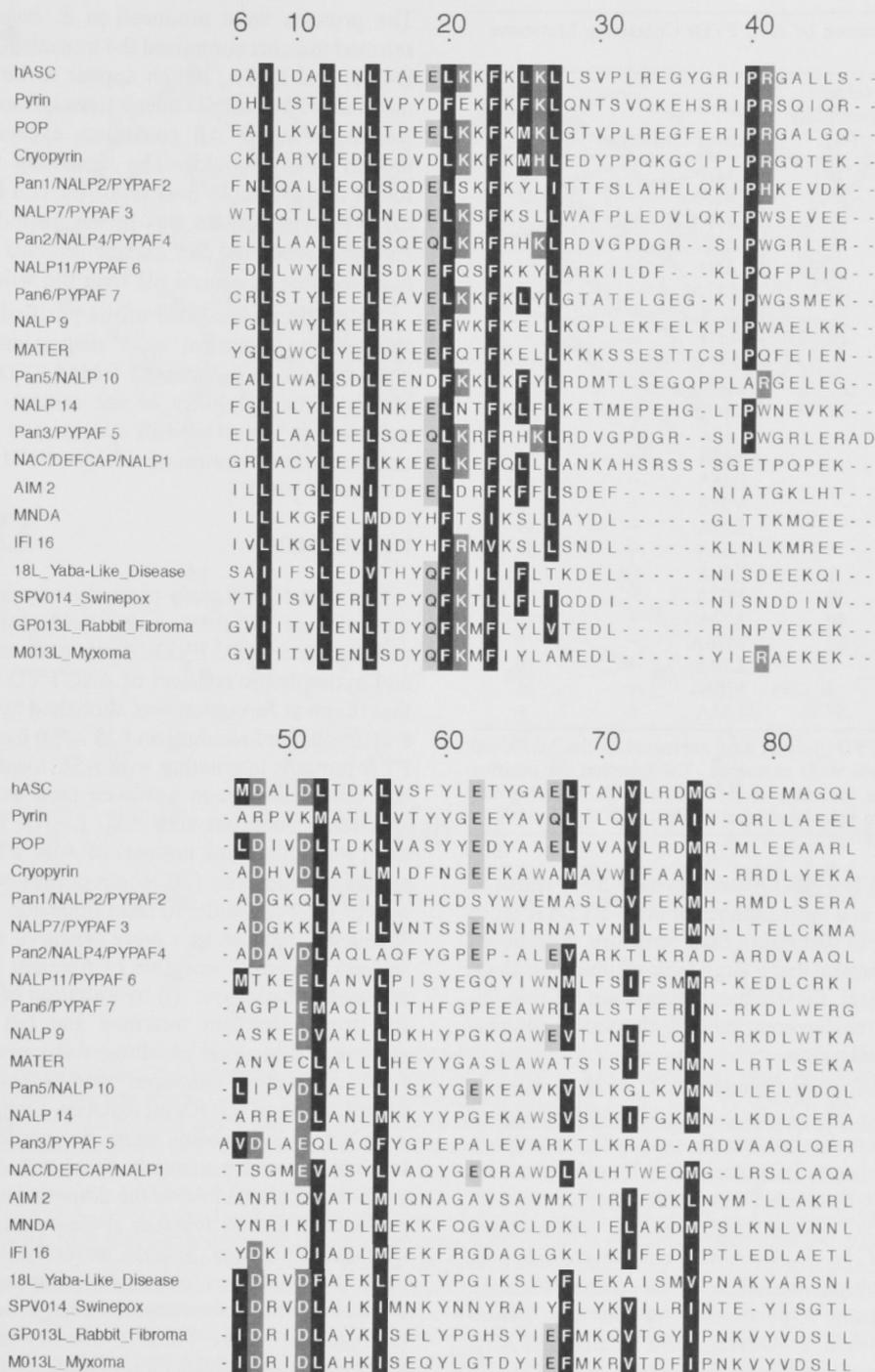


FIGURE 3: Amino acid sequence alignment of PYDs from human and viral proteins. The amino acid sequence data of the PYDs were obtained from the NCBI gene bank (<http://www.ncbi.nlm.nih.gov>) and previous reports (10, 11, 30). The alignment highlights the charged and hydrophobic residues found to be critical for PYD filament formation in the present study. White letters with black backgrounds identify nonpolar residues that are identical or similar to hydrophobic residues critical for PYD filament formation. White letters with gray backgrounds identify charged residues that have the same charge as charged residues critical for filament formation in human ASC PYD (K21, K26, R41, D48, and D51). Black letters with gray backgrounds identify glutamic acid (E) and glutamine (Q) residues corresponding to E19, E62, and E67 in human ASC PYD for which the charge is not required for PYD filament formation. The residue numbers of the ASC PYD are shown above the alignment.

expected, if L25 were involved in a hydrophobic intermolecular contact between PYDs. In contrast, hydrophilic and charged mutants (E, Q) of L45 retained the capability of filament formation. Since charged residues are difficult to accommodate in protein-protein interaction surfaces, L45 is probably located outside the PYD-PYD interface. Sup-

pression of filament formation by the L45A mutant may be explained by a structural change induced by the small alanine side chain.

Although M47 is not very solvent accessible in the NMR structure of the ASC PYD, the abolishment of filament formation by the M47A mutation may indicate involvement

Table 2: Filament Formation by ASC PYDs Containing Mutations of Nonpolar Residues<sup>a</sup>

mutant	filament	solvent accessibility (%)	mutant	filament	solvent accessibility (%)
I8A	—	0	L44A	+	0
L9A	+	24	L45A	—	32
L12A	—	0	L45I	+	
L15A	—	2	L45E	+	
L20A	—	16	L45Q	+	
F23A	—	0	M47A	—	7
F23L	+		M47L	+	
L25A	—	45	M47I	+	
L25M	+		M47V	+	
L25I	+		M47Q	—	
L25V	+		M47N	—	
L25F	+		L50A	+	65
L25G	—		L52A	—	1
L25Q	—		L56A	—	0
L25N	—		V57A	+	3
L25K	—		F59A	+	33
L25E	—		L61A	+	57
L27A	—	1	L68A	—	1
L28A	+	38	V72A	—	2
V30A	+	40	L73A	+	0
P31A	+	88	M76A	—	2
L32A	+	9	L78A	+	10
I39A	+	1	M81A	+	16
P40A	—	67	L85A	+	0

<sup>a</sup> GFP-tagged ASC PYD mutants were expressed in Cos7 cells and their filament formations were examined. The criterion for positive filament formation was the same as in Table 1. Side chain solvent accessibilities were also calculated as in Table 1.

of this residue in PYD–PYD interactions, as this residue is replaced by alanine in a large number of other PYDs (Figure 3), apparently without disturbing their structure. Additional mutations of M47 showed that a large hydrophobic side chain at this site is critical for filament formation (Table 2), suggesting that this residue contributes hydrophobic contacts at the PYD–PYD interface.

**Localization of Filament-Deficient ASC PYD Mutants in Specks Produced by Full-Length ASC.** Overexpression of full-length ASC in cells by transfection with expression plasmids produces large, irregularly distributed structures called “specks” (1, 4, 5). Since PYD proteins interacting with ASC localize in ASC specks when they are cotransfected with full-length ASC (1, 5, 14), ASC specks might also play a role under physiological conditions. In contrast to ASC, intracellular expression of pyrin alone results in a diffuse cytoplasmic distribution. However, pyrin localizes in ASC specks when cotransfected with full-length ASC (5). To examine whether filament-deficient alanine mutants localize in ASC specks, we coexpressed GFP-tagged ASC PYD mutants with Flag-tagged full-length ASC in Cos7 cells. All filament-deficient ASC PYD mutants of charged and hydrophobic residues were localized in ASC specks (data not shown), indicating the conservation of an intact surface for the interaction with ASC specks. This result suggests that the overall structure of the ASC PYD is still intact after the introduction of point mutations that disturb the structure sufficiently to abolish filament formation.

**In vitro Studies of ASC PYD Mutants.** To check the degree of structural perturbation introduced by the point mutations in in vitro experiments, His<sub>6</sub>-tag constructs were made of the wild-type ASC PYD and a number of selected mutants.

The proteins were produced in *E. coli* and purified. The selected mutants comprised the mutants K21A, L25A, R41A, D48A, and D51A, which appear to be involved in intermolecular PYD–PYD interactions according to the filament formation assay. All constructs except D48A expressed protein in high yields. The remaining four proteins were found in the soluble fraction of the cell lysates. In contrast, the wild-type protein was in the insoluble fraction. These data indicated that the mutant proteins were significantly more soluble at neutral pH than the wild-type.

NMR spectra recorded of the purified mutants at pH 7.2 showed the chemical shift dispersion and line widths characteristic of monomeric folded PYD domains (data not shown). The solubility of the mutant D51A was barely sufficient to record a NMR spectrum at neutral pH, but the characteristic spectrum of a folded PYD domain was easily verified at pH 3.7.

## DISCUSSION

The data of this study present a comprehensive survey of the roles of different amino acid residues for the self-association of ASC PYDs. Systematic mutation of charged and hydrophobic residues of ASC PYD to alanine revealed that filament formation was abolished by point mutations in 8 of 24 charged residues and 15 of 30 hydrophobic residues. PYD proteins interacting with ASC localize in ASC specks, and this phenomenon has been used as evidence for their specific interactions with ASC (5–7). We observed, however, that all alanine mutants of ASC PYD concentrated in specks produced by full-length coexpressed ASC, independent of their capability to form filaments. Therefore, we used filament formation as a more sensitive assay of ASC PYD self-association. Disruption of filament formation could be affected in two ways: (i) by mutation of residues located at the protein–protein interface and (ii) indirectly by the structural disruption resulting from mutation of a buried residue. Several measures were taken to ascertain the identification of interfacial residues. First, data from residues with buried side chains were disregarded, unless mutation to alanine could reasonably be expected not to disrupt the three-dimensional fold of the domain, based on the frequent presence of alanine residues at these positions in other PYD-containing proteins. Second, several residues were probed by further mutations to different polar and nonpolar residues to establish the importance of charge or hydrophobicity at the respective sites. Third, a few key mutants of the ASC PYD were expressed with a His-tag, purified, and analyzed by NMR spectroscopy, confirming that their three-dimensional structure was still intact. The residues identified in this way were distributed over coherent surface areas of the three-dimensional ASC PYD structure (Figure 4), lending additional support for their involvement in direct PYD–PYD interactions.

Notably, the NMR structure of the wild-type ASC PYD had to be determined at pH 3.7 from refolded protein (18), because the protein precipitated at neutral pH. In contrast, all His-tagged mutants expressed in *E. coli* in high yield in the present study (K21A, L25A, R41A, and D51A) were found in the soluble fraction of the cell lysate at neutral pH. These in vitro results confirm that the mutants are less likely to aggregate compared to the wild-type ASC PYD, strongly

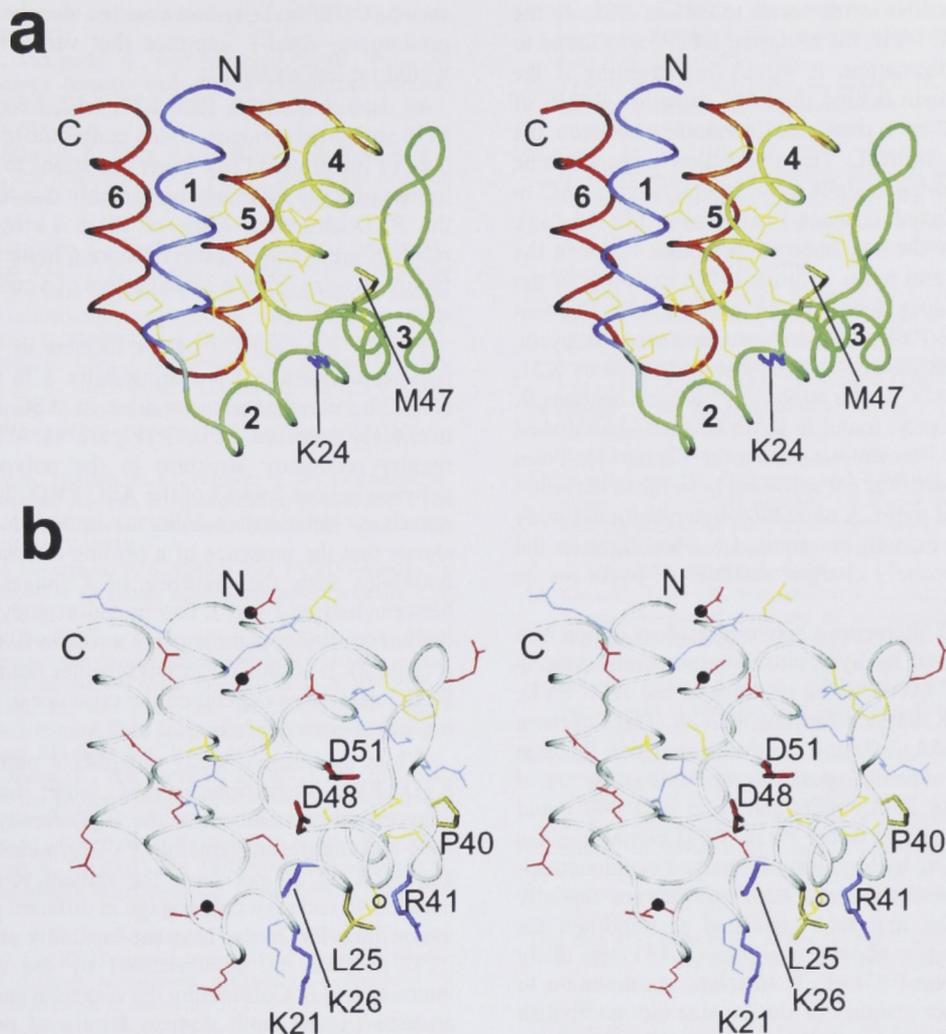


FIGURE 4: Stereoviews of the human ASC PYD, summarizing the mutation results. Side chains are shown only for those residues that were mutated to Ala in the present study. The following colors were used for the amino acid side chains: blue, positively charged side chains (K, R); red, negatively charged side chains (E, D); yellow, hydrophobic side chains (I, L, M, F, P, V). The N- and C-terminal ends of the domain are labeled. (a) View of mutated amino acid side chains with less than 20% average side chain solvent accessibility in the NMR structure. For these residues, mutations may have resulted in loss of function due to structural perturbation of the hydrophobic core. Among the mutated residues, K24 was the only charged residue with less than 20% side chain solvent accessibility. Mutation of M47 resulted in loss of filament formation either because of disruption of the protein structure or because of its location near the putative PYD–PYD interaction surface. The helices are identified with numbers. (b) View of mutated amino acid side chains with more than 20% average side chain solvent accessibility. Side chains that disrupted filament formation upon mutation are labeled with their one-letter amino acid symbol and their sequence number; for the charged residues among these, mutations to Ala as well as loss-of-charge mutations resulted in loss of filament formation. Mutations of unlabeled side chains did not disrupt filament formation. Black spheres identify the side chains of three Glu residues which led to loss of filament formation upon mutation to Ala but not upon mutation to Gln. The charge of these residues thus does not seem critical for PYD–PYD interactions in ASC. Similarly, a circle identifies the side chain of L45, which led to loss of filament formation upon mutation to Ala but not upon mutation to Ile, Gln, or Glu.

suggesting that the filament formation observed *in vivo* is based on direct PYD–PYD interactions and not mediated by other factors.

Interactions between residues of opposite charge are determinants of homo- and heterophilic interactions between DDs, DEDs, and CARDS (28, 29, 32, 33). Our present results provide experimental proof for earlier predictions that charged residues also play a crucial role for the self-association of PYDs (5, 18, 19, 30). In particular, the mutation data demonstrate the importance of K21, K26, R41, D48, and D51 for the production of ASC PYD filaments.

On the basis of the PYD structure of ASC, the positively charged side chains of K21, K26, and R41 form a positively charged cluster. Similarly, the negatively charged side chains

of D48 and D51 are close together (Figure 4b). These two clusters of opposite charge are separated by the backbone of helix 3. It would thus be plausible that intermolecular electrostatic interactions between the positively charged side chains of K21, K26, and R41 in one molecule and the negatively charged side chains of D48 and D51 in another molecule drive the polymerization and filament formation. As these interactions affect only a relatively small fraction of the surface area of the PYD at the far end of the N- and C-termini, it is not surprising that fusion proteins with GFP and other domains retain the propensity for filament formation.

FMF is caused by point mutations in pyrin. One of these FMF-associated mutations, R42W, is located in the PYD of

pyrin (31). This residue corresponds to R41 in ASC. In the context of the ASC PYD, the mutation R41W was found to abolish filament formation. It would be intriguing if the molecular mechanism behind the FMF mutation R42W of pyrin were based on a disrupted interaction between the PYDs from pyrin and ASC. This interaction is known to be of functional importance since pyrin interacts with ASC to inhibit ASC-mediated caspase-1 activation (5, 34, 35). Notably, however, the sequence conservation between the PYDs from ASC and pyrin is limited. For example, of the five charged residues found to be essential for filament formation of ASC PYD, only two are conserved in pyrin. Furthermore, mutation of the nonconserved residues K21, D48, and D51 in ASC to the oppositely charged residues E, R, and K, respectively, found in pyrin (Figure 3) abolished the formation of filamentous structures (Table 1). Point mutations of ASC are thus not sufficient to simulate the ASC-binding function of pyrin. A more thorough mutational study of the PYD of pyrin will be required to shed light on the role of these oppositely charged residues of pyrin for its inhibitory activity.

The PYD–PYD interactions between residues of opposite charge are supported by hydrophobic interactions. Among the 30 mutants of hydrophobic residues of the ASC PYD, 15 mutants lost the filament-forming activity. Three of these (L25A, P40A, L45A) affected residues for which the side chains are highly solvent exposed in the NMR structure of the domain (Table 2). In particular, L25 and P40 could plausibly be involved in direct PYD–PYD contacts, since mutation of L25 to hydrophilic or charged residues suppressed filament formation and both residues are spatially close to K21, K26, and R41 identified as important for filament formation (see above). In contrast, L45 is less likely to be located in the PYD–PYD interface, as mutation to glutamate and other residues of similar size did not abolish filament formation.

Overall, our results underline the role of helices 2 and 3 in PYD–PYD interactions and suggest that the binding between the ASC PYDs is supported by hydrophobic interactions. M47 and, possibly, K24 also form part of this interaction surface. Both residues are barely solvent accessible, yet only the M47A mutation interfered with filament formation (Tables 1 and 2). Although the side chain of M47 contributes to the hydrophobic core of the protein only superficially (Figure 4a), it cannot be ruled out that the M47A mutation affects the structure of the PYD. Notably, however, this residue is substituted by alanine in a large number of naturally occurring PYDs (Figure 3), suggesting that it is not a structurally important residue.

The extent of the present mutational analysis allows us to exclude a role of helix 1 for filament formation. Crystal structures of the complex between the CARD domains of the procaspase-9 prodomain and Apaf-1 and the complex between Pelle and Tube death domains showed that helix 1 was involved in both complexes of these members of the death domain superfamily (28, 29). Our results thus indicate that PYD–PYD interactions of human ASC use a different binding mode. As also helices 2, 3, and 4 are engaged in the interface of the complex between the procaspase-9 prodomain and Apaf-1, and residues from these three helices are also required for filament formation of ASC PYD, it may be tempting to model a PYD–PYD complex based on this

structure. Helix 1 is, however, so deeply involved in the prodomain–Apaf-1 interface that very little contact area would be left without it.

I8, L12, L15, L20, F23, L27, L52, L56, L68, V72, and M76 are buried residues, which contribute to the hydrophobic core of the protein (18). Their mutations to alanine resulted in loss of PYD filaments presumably due to perturbation of the PYD structure. Most of these 11 hydrophobic core residues are highly conserved among human and viral PYD family proteins (Figure 3), indicative of a conserved structural role.

P42 of the Nalp1 PYD is located in an unstructured polypeptide segment replacing helix 3 in the Nalp1 PYD (19). The corresponding residue in ASC PYD (P40) immediately precedes helix 3 (Figure 4). While there is no regular secondary structure in the polypeptide segment between helices 2 and 3 of the ASC PYD, this segment does not show enhanced mobility as in Nalp1 PYD. Figure 3 shows that the presence of a proline residue at position 40 correlates with the presence of a longer linker peptide between helices 2 and 3. Our mutation study shows that P40 but not the linker peptide plays a role in filament formation (Figure 2). It is thus uncertain whether homologous PYD–PYD interactions depend on the presence of a long connecting linker between helices 2 and 3.

While the present study suggests a particular mode of PYD–PYD interactions in ASC, other modes may occur between other members of the PYD family. For example, viral and interferon-inducible PYD proteins (IFI16, AIM2, and MND1) do not have the critical proline residue at position 40 and may thus engage in different protein–protein interactions. PYDs also have the capability of binding to non-PYD proteins, as demonstrated by the ASC PYD–Bax interaction (36). Considering the structural similarity between proteins of the death domain family of proteins and the documented variability of their intermolecular associations (37), the structure of a complex between PYDs presents an intriguing question. The present data provide the basis for the construction of a homodimeric complex between ASC PYDs, which will yield more detailed insight into PYD–PYD interactions.

## ACKNOWLEDGMENT

We thank Dr. J. C. Reed (Burnham Institute) for critical reading of this manuscript, Dr. J. Masumoto for the preparation of expression plasmids, and T. Shijo for preliminary mutation experiments. Access to the Biomolecular Resource and NMR facilities at the Australian National University is gratefully acknowledged.

## REFERENCES

- Masumoto, J., Taniguchi, S., Ayukawa, K., Sarvotham, H., Kishino, T., Niikawa, N., Hidaka, E., Katsuyama, T., Higuchi, T., and Sagara, J. (1999) ASC, a novel 22-kDa protein, aggregates during apoptosis of human promyelocytic leukemia HL-60 cells. *J. Biol. Chem.* 274, 33835–33838.
- Conway, K. E., McConnell, B. B., Bowring, C. E., Donald, C. D., Warren, S. T., and Vertino, P. M. (2000) TMS1, a novel proapoptotic caspase recruitment domain protein, is a target of methylation-induced gene silencing in human breast cancers. *Cancer Res.* 60, 6236–6242.
- McConnell, B. B., and Vertino, P. M. (2000) Activation of a caspase-9-mediated apoptotic pathway by subcellular redistribution

- of the novel caspase recruitment domain protein TMS1, *Cancer Res.* 60, 6243–6247.
4. Masumoto, J., Taniguchi, S., and Sagara, J. (2001) Pyrin N-terminal homology domain- and caspase recruitment domain-dependent oligomerization of ASC, *Biochem. Biophys. Res. Commun.* 280, 652–655.
  5. Richards, N., Schaner, P., Diaz, A., Stuckey, J., Shelden, E., and Wadhwat, A., and Gumucio, D. L. (2001) Interaction between pyrin and the apoptotic speck protein (ASC) modulates ASC-induced apoptosis, *J. Biol. Chem.* 276, 39320–39329.
  6. Stehlik, C., Fiorentino, L., Dorfleutner, A., Bruey, J., Ariza, M., Sagara, J., and Reed J. C. (2002) The PAAD/PYRIN-family protein ASC is a dual regulator of a conserved step in nuclear factor kappaB activation pathways, *J. Exp. Med.* 196, 1605–1615.
  7. Manji, G., Wang, L., Geddes, B. J., Brown, M., Merriam, S., Al-Garawi, A., Mak, S., Lora, J. M., Briskin, M., Jurman, M., Cao, J., DiStefano, P. S., and Bertin, J. (2002) PYPAF1, a PYRIN-containing Apaf1-like protein that assembles with ASC and regulates activation of NF-kappa B, *J. Biol. Chem.* 277, 115701–11575.
  8. Fairbrother, W. J., Gordon, N. C., Humke, E. W., O'Rourke, K. M., Starovasnik, M. A., Yin, J. P., and Dixit, V. M. (2001) The PYRIN domain: a member of the death domain-fold superfamily, *Protein Sci.* 10, 1911–1918.
  9. Bertin, J., and DiStefano, P. S. (2000) The PYRIN domain: a novel motif found in apoptosis and inflammation proteins, *Cell Death Differ.* 7, 1273–1274.
  10. Pawlowski, K., Pio, F., Chu, Z., Reed, J. C., and Godzik, A. (2001) PAAD – a new protein domain associated with apoptosis, cancer and autoimmune diseases, *Trends Biochem. Sci.* 26, 85–87.
  11. Staub, E., Dahl, E., and Rosenthal, A. (2001) The DAPIN family: a novel domain links apoptotic and interferon response proteins, *Trends Biochem. Sci.* 26, 83–85.
  12. Martinon, F., Burns, K., and Tschopp, J. (2002) The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta, *Mol. Cell* 10, 417–426.
  13. Srinivasula, S. M., Poyet, J. L., Razmara, M., Datta, P., Zhang, Z., and Alnemri, E. S. (2002) The PYRIN-CARD protein ASC is an activating adaptor for caspase-1, *J. Biol. Chem.* 277, 21119–21122.
  14. Stehlik, C., Lee, S. H., Dorfleutner, A., Stassinopoulos, A., Sagara, J., and Reed, J. C. (2003) Apoptosis-associated speck-like protein containing a caspase recruitment domain is a regulator of procaspase-1 activation, *J. Immunol.* 171, 6154–6163.
  15. Mariathasan, S., Newton, K., Monack, D. M., Vucic, D., French, D. M., Lee, W. P., Roose-Girma, M., Erickson, S., and Dixit, V. M. (2004) Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf, *Nature* 430, 213–218.
  16. Salvesen, G. S., and Dixit, V. M. (1999) Caspase activation: the induced-proximity model, *Proc. Natl. Acad. Sci., U.S.A.* 96, 10964–10967.
  17. Kumar, S., and Colussi, P. A. (1999) Prodomains—adaptors—oligomerization: the pursuit of caspase activation in apoptosis, *Trends Biochem. Sci.* 24, 1–4.
  18. Liepinish, E., Barbals, R., Dahl, E., Sharipo, A., Staub, E., and Otting, G. (2003) The death-domain fold of the ASC PYRIN domain, presenting a basis for PYRIN/PYRIN recognition, *J. Mol. Biol.* 332, 1155–1163.
  19. Hiller, S., Kohl, A., Fiorito, F., Herrmann, T., Wider, G., Tschopp, J., Grütter, M. G., and Wüthrich, K. (2003) NMR structure of the apoptosis- and inflammation-related NALP1 pyrin domain, *Structure* 11, 1199–1205.
  20. Siegel, R. M., Martin, D. A., Zheng, L., Ng, S., Bertin, J., Cohen, J., and Lenardo, M. J. (1998) Death-effector filaments: novel cytoplasmic structures that recruit caspases and trigger apoptosis, *J. Cell Biol.* 141, 1243–1253.
  21. Perez, D., and White, E. (1998) E1B 19K inhibits Fas-mediated apoptosis through FADD-dependent sequestration of FLICE, *J. Cell Biol.* 141, 1255–1266.
  22. Yan, M., Lee, J., Schilbach, S., Godard, A., and Dexit, V. (1999) mE10, a novel caspase recruitment domain-containing proapoptotic molecule, *J. Biol. Chem.* 274, 10287–10292.
  23. Guet, C., and Vito, P. (2000) Caspase recruitment domain (CARD)-dependent cytoplasmic filaments mediate bcl10-induced NF-kappaB activation, *J. Cell Biol.* 148, 1131–1139.
  24. Imai, Y., Matsushima, Y., Sugimura, T., and Terada, M. (1991) A simple and rapid method for generating a deletion by PCR, *Nucleic Acids Res.* 19, 2785.
  25. Neylon, C., Brown, S. E., Kralicek, A. V., Miles, C. S., Love, C. A., and Dixon, N. E. (2000) Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance, *Biochemistry* 39, 11989–11999.
  26. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, *Methods Enzymol.* 185, 60–89.
  27. Sevilla-Sierra, P., Otting, G., and Wüthrich, K. (1994) Determination of the nuclear magnetic resonance structure of the DNA-binding domain of the P22 c2 repressor (1–76) in solution and comparison with the DNA-binding domain of the 434 repressor, *J. Mol. Biol.* 235, 1003–1020.
  28. Xiao, T., Towb, P., Wasserman, S. A., and Sprang, S. R. (1999) Three-dimensional structure of a complex between the death domains of Pelle and Tube, *Cell* 99, 545–555.
  29. Qin, H., Srinivasula, S. M., Wu, G., Fernandes-Alnemri, T., Alnemri, E. S., and Shi, Y. (1999) Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1, *Nature* 399, 549–557.
  30. Liu, T., Rojas, A., Ye, Y., and Godzik, A. (2003) Homology modeling provides insights into the binding mode of the PAAD/DAPIN/pyrin domain, a fourth member of the CARD/DD/DED domain family, *Protein Sci.* 12, 1872–1881.
  31. Hull, K. M., Shoham, N., Chae, J. J., Aksentjevich, I., and Kastner, D. L. (2003) The expanding spectrum of systemic autoinflammatory disorders and their rheumatic manifestations, *Curr. Opin. Rheumatol.* 15, 61–69.
  32. Chou, J. J., Matsuo, H., Duan, H., and Wagner, G. (1998) Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment, *Cell* 94, 171–180.
  33. Eberstadt, M., Huang, B., Chen, Z., Meadows, R. P., Ng, S., Zheng, L., Lenardo, M. J., and Fesik, S. W. (1998) NMR structure and mutagenesis of the FADD (Mort1) death-effector domain. Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment, *Nature* 392, 941–945.
  34. Chae, J. J., Komarow, H. D., Cheng, J., Wood, G., Raben, N., Liu, P. P., and Kastner, D. L. (2003) Targeted disruption of pyrin, the FMF protein, causes heightened sensitivity to endotoxin and a defect in macrophage apoptosis, *Mol. Cell* 11, 591–604.
  35. Dowds, T. A., Masumoto, J., Chen, F. F., Ogura, Y., Inohara, N., and Nunez, G. (2003) Regulation of feveropyrin/Pypaf1 signaling by pyrin, the familial Mediterranean fever gene product, *Biochem. Biophys. Res. Commun.* 302, 575–580.
  36. Ohtsuka, T., Ryu, H., Minamishima, Y. A., Macip, S., Sagara, J., Nakayama, K. I., Aaronson, S. A., and Lee, S. W. (2004) ASC is a Bax adaptor and regulates the p53-Bax mitochondrial apoptosis, *Nat. Cell Biol.* 6, 121–128.
  37. McConnell, B. B., and Vertino, P. M. (2004) TMS1/ASC: the cancer connection, *Apoptosis* 9, 5–18.

BI048374I

**9. Appendix: II (VI)**

# Paper II

**P. S. C. Wu**, G. Otting

“Rapid pulse-length determination in high-resolution NMR”

Reproduced with permission from

*Journal of Magnetic Resonance* **176**, 115-119 (2005)

Copyright 2005 Elsevier

Communication

# Rapid pulse length determination in high-resolution NMR

Peter S.C. Wu, Gottfried Otting\*

*Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia*

Received 15 April 2005; revised 13 May 2005

Available online 21 June 2005

## Abstract

The  $90^\circ(^1\text{H})$  pulse length can be determined in a single scan using a simple homo-gated decoupling/nutation experiment. We show that the method is fast, accurate and readily amenable to automation.

© 2005 Elsevier Inc. All rights reserved.

*Keywords:* Nutation spectroscopy;  $90^\circ$  pulse; Homo-gated decoupling; Automation

## 1. Introduction

The sensitivity of  $^1\text{H}$  NMR experiments depends on the accurate knowledge of the  $90^\circ$  pulse length. On high-field NMR spectrometers, this pulse length critically depends on the chemical properties of the sample, in particular its ionic strength. Traditionally, the  $90^\circ$  pulse length is determined by a series of one-dimensional NMR experiments that aim to measure either the  $180^\circ$  or  $360^\circ$  pulse length [1,2]. Fourier transformation of a series of experiments recorded with systematically incremented pulse length can be used to derive the radiofrequency field strength from peak positions in a frequency spectrum [3]. Although pulse length determinations are easy to perform and take little time, they have to be performed so frequently that automation would be desirable. In the present project, we explored a nutation method that delivers the  $90^\circ(^1\text{H})$  pulse length in a single scan. Pulse length determination by nutation spectroscopy is an established method in MR imaging [4,5]. Our results show that the method produces accurate results with an accuracy of about 0.5% under the stringent conditions of high-resolution NMR spectroscopy for different situations of multi-line NMR spectra and is readily amenable to automation.

## 2. Results and discussion

Nutation spectroscopy is closely related to homo-gated decoupling [6], i.e., the sample is irradiated during the data acquisition by dividing each dwell time into a part for radiofrequency (RF) irradiation and a part for sampling a data point [7]. No excitation pulse precedes the scan. If the sample contains only a single resonance and the carrier frequency is positioned on-resonance, the magnetization nutates around the axis of the RF pulses. The resulting FID records the projection of the magnetization onto the transverse plane. Fourier transformation of the FID yields the nutation spectrum (Fig. 1). In the absence of off-resonance signals, the projection of the magnetization oscillates along a single axis, resulting in a nutation spectrum that is symmetric around zero frequency. The peak separation reflects the average nutation frequency during the scan and its inverse is proportional to the duration of a  $360^\circ$  pulse.

Several points must be observed to obtain the accurate  $90^\circ$  hard pulse, i.e., the length of the  $90^\circ$  pulse at high power. (i) It is recommended to perform the

\* Corresponding author. Fax: +61 2 61250750.

E-mail address: [gottfried.otting@anu.edu.au](mailto:gottfried.otting@anu.edu.au) (G. Otting).

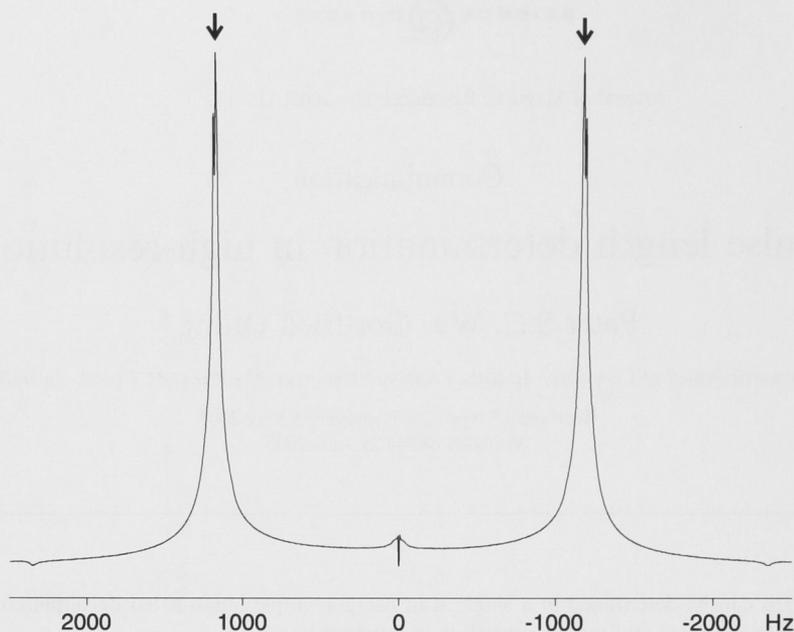


Fig. 1. Nutation spectrum of a 3.6 mM solution of hen egg white lysozyme in 90% H<sub>2</sub>O/10% D<sub>2</sub>O, pH 7.0 acquired at 25 °C on a Bruker Avance 800 MHz NMR spectrometer. The spectrum was recorded with a power level attenuated 16-fold compared to that of the hard 90° pulse, using a duty cycle of 80%. A magnitude spectrum was calculated after Fourier transformation. A hard 90° pulse of duration  $\tau_{90} = 10.65 \mu\text{s}$  was derived using  $\tau_{90} = d/(32 \times \Delta\nu)$ , where  $d$  is the duty cycle and  $\Delta\nu$  the frequency difference between the peaks identified by arrows. For comparison, a conventional pulse length determination searching for the null point of the water after a 360° pulse yielded  $\tau_{90} = 10.63 \mu\text{s}$ .

nutation experiment with reduced power to avoid sample heating. In addition, cryogenic probes may not tolerate the application of high power during the full duration of the FID in the case of large duty cycles. (ii) Calculation of the 90° hard pulse from an experiment recorded with attenuated power requires linear amplifiers or, alternatively, a calibration procedure. On our NMR spectrometer, the linearity of the amplifiers is achieved by software correction tables, resulting in predictable and accurate scaling of the RF power. (iii) The duty cycle must be taken into account. We use to record all NMR spectra with a digital filter and oversampling [8], i.e., the dwell times are 32-fold shorter than they would be without oversampling. In our hands, accurate pulse lengths were obtained with an 80% duty cycle (i.e., RF-irradiation during 80% of each dwell time), but not with 20% duty cycle. The duty cycle actually used by the spectrometer (displayed by the status parameter HDDUTY in the file `acqus` on our Bruker AV800 NMR spectrometer) was about 1% shorter than its nominal value in the parameter file `acqus`. Consequently, the value calculated for the hard 90° pulse length assuming an 80% duty cycle should have been too long by 1%. This effect was fortuitously compensated by the fact that the nutation peak shows a weak tail towards lower frequencies, corresponding to regions in the sample, where the RF field is both weak and inhomogeneous. The magnetization in these regions dephases rapidly during the nutation experiment. Therefore, the maximum of the

nutation peak is primarily determined by the most homogeneous region of the RF field distribution and the corresponding pulse length tends to be shorter than that determined from a 180° or 360° rotation, compensating the error made by assuming that the duty cycle was 1% longer than the actually used value. Keeping these deviations in mind, duty cycles other than 80% will work too. To maintain control over the actually used duty cycle, it is recommended to use the same acquisition parameters for all pulse length determinations. (iv) The nutation spectrum must be recorded with the carrier frequency at or close to an NMR signal. If the carrier frequency is off-resonance, the effective rotation frequency  $\omega_{\text{eff}}$  increases according to

$$\omega_{\text{eff}} = \sqrt{\omega_1^2 + \Omega^2}, \quad (1)$$

where  $\omega_1$  denotes the RF field strength and  $\Omega$  the difference between the RF and the Larmor frequency. Small off-resonance effects are well tolerated. For example, for a nutation frequency of 1000 Hz, a 100 Hz off-set creates less than 0.5% error in the pulse length determination. (v) Radiation damping during the nutation experiment reduces the average nutation frequency. However, the magnetization defocuses during the nutation experiment due to RF inhomogeneity [9–11], eliminating radiation damping towards the end of the FID. The positions of the peak maxima in the nutation spec-

trum are thus hardly affected by radiation damping, making pulse length determinations that use the water resonance particularly straightforward and accurate.

The nutation method also yields accurate  $^1\text{H}$  pulse length determinations for organic compounds in deuterated solvents which often display many  $^1\text{H}$  NMR peaks of similar intensity without a dominant solvent resonance. In this situation, the nutation spectrum displays several peaks and the  $90^\circ$  pulse must be derived from the symmetric pair of peaks with the lowest frequency (disregarding any peak at zero frequency). The carrier frequency must be placed at the frequency of one of the resonances, since any off-resonance peak will appear at an increased frequency, regardless of whether it is located high-field or low-field from the carrier (see Eq. (1)). An example is shown in Fig. 2. Placing the carrier frequency on the most intense resonance, several signals are obtained. The correct pulse length is derived from the peak shoulder near the peak maximum. Using the peak maxima instead, a 0.5% shorter pulse length would have been derived. This value still is within the error of pulse length determination by the conventional method of minimizing the signals after a  $360^\circ$  pulse, considering that off-reso-

nance effects result in different effective flip-angles at different offsets Eq. (1). For example, a  $360^\circ$  pulse based on the pulse length derived from the inner shoulders in Fig. 2B yielded the spectrum of Fig. 2C, where the resonances near the carrier frequency have nearly disappeared. Shortening the duration of the  $360^\circ$  pulse to minimize the integral of the off-resonance signals required an almost 1% shorter pulse (Fig. 2D). Picking the peak maxima instead of the peak shoulders in the nutation spectrum of Fig. 2B would thus have yielded the  $90^\circ$  pulse length with perfectly adequate accuracy.

The nutation method would be expected to become less accurate when the signal overlap is severe and there is no dominant line in the NMR spectrum. Yet, using a solution of hen egg white lysozyme in  $\text{D}_2\text{O}$  and placing the carrier frequency on the biggest methyl signal (Fig. 3A), the pulse length determined by the nutation method was only 0.3% shorter than that determined by a conventional  $360^\circ$ -pulse determination, because the signals from different methyl groups could not be resolved in the nutation spectrum (Fig. 3B).

It is an important advantage of the nutation method that it is readily amenable to automation. This works

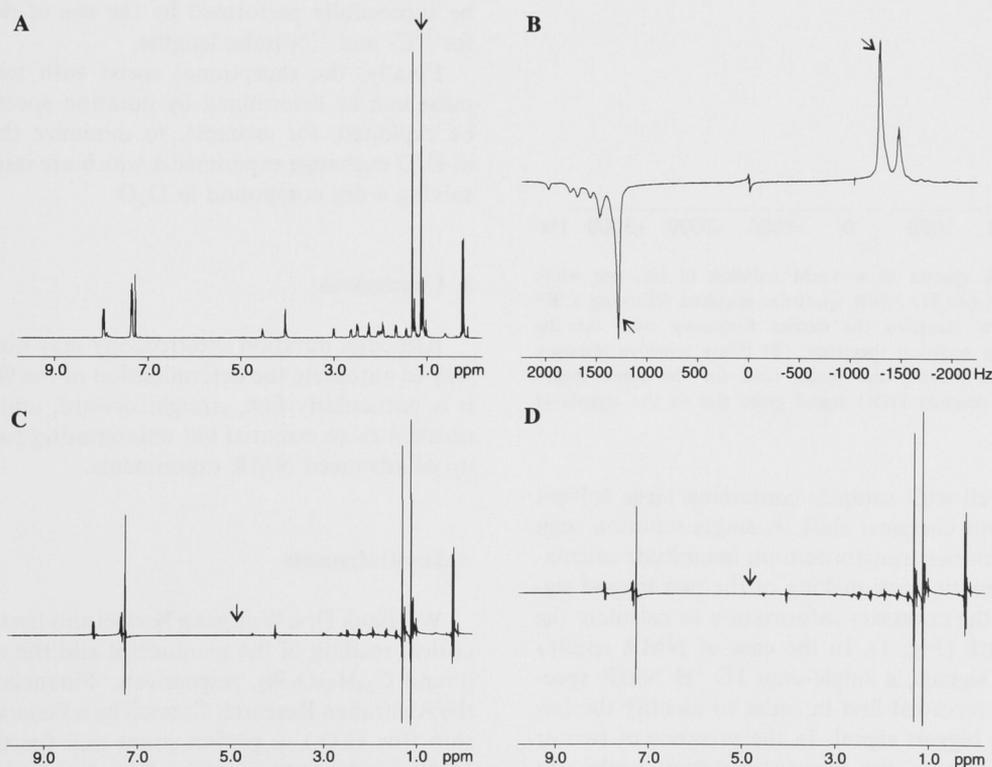


Fig. 2.  $^1\text{H}$  NMR spectra of an organic compound,  $\text{C}_{34}\text{H}_{53}\text{O}_5\text{Si}_2$ , in deuterobenzene. (A) 1D NMR spectrum acquired after a  $90^\circ$  pulse. The arrow identifies the RF carrier frequency used for the acquisition of the nutation spectrum. (B) Phase-sensitive nutation spectrum. The arrows identify the peak shoulders used for the pulse length calculation, yielding  $9.81\ \mu\text{s}$  for the width of the hard  $90^\circ$  pulse. Using the peak maxima would have yielded a hard  $90^\circ$  pulse of  $9.76\ \mu\text{s}$ . (C) 1D NMR spectrum recorded after application of a  $360^\circ$  pulse of  $4 \times 9.81\ \mu\text{s}$  duration. The carrier frequency is identified by an arrow. (D) 1D NMR spectrum acquired after a pulse of  $4 \times 9.73\ \mu\text{s}$  duration. The pulse minimized the residual intensities of the methyl resonances at 1.2 ppm, but was a bit too short for the resonances near the carrier frequency (arrow). Spectra (C and D) were vertically expanded about 25-fold compared to the spectrum in (A).

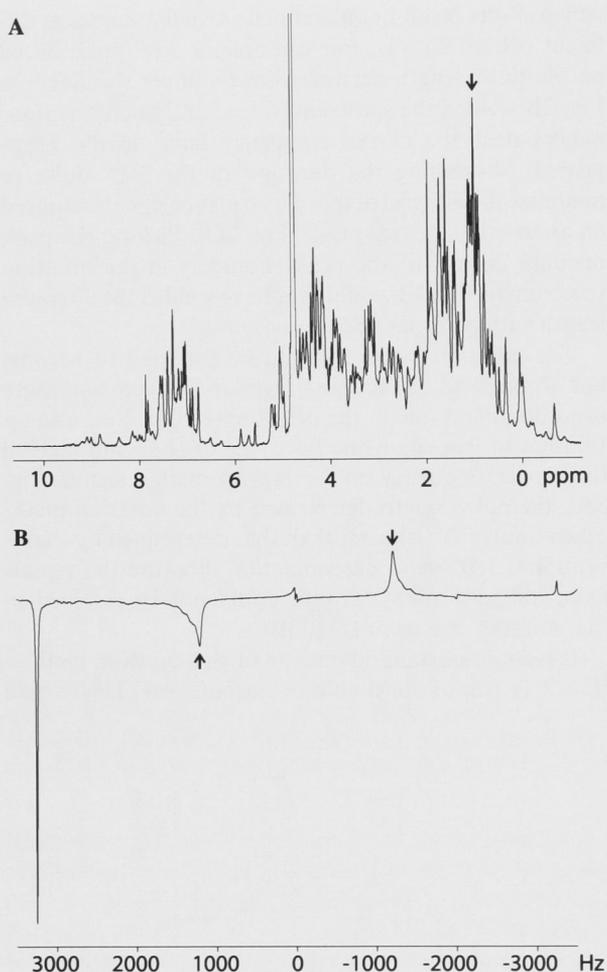


Fig. 3.  $^1\text{H}$  NMR spectra of a 4 mM solution of hen egg white lysozyme in  $\text{D}_2\text{O}$ . (A) 1D NMR spectrum acquired following a  $90^\circ$  pulse. The arrow identifies the carrier frequency used for the acquisition of the nutation spectrum. (B) Phase sensitive nutation spectrum. Arrows identify the peaks used for the pulse length calculation. The residual HDO signal gives rise to the signals at  $\pm 3246$  Hz.

particularly well with samples containing large solvent peaks of known chemical shift. A single nutation scan followed by Fourier transformation, magnitude calculation and automatic peak picking of the two biggest signals provides the necessary information to calculate the  $90^\circ$  pulse length (Fig. 1). In the case of NMR spectra with multiple signals, a single-scan 1D  $^1\text{H}$  NMR spectrum must be recorded first in order to identify the frequency of the biggest signal. In the presence of two or more intense signals, the relevant nutation peaks can be distorted by a magnitude calculation and may not be the most intense signal in the nutation spectrum (Figs. 2B and 3). In this situation we recommend a Fourier transform without magnitude calculation, picking of the peaks with the lowest absolute frequency that can be found in the nutation spectrum in the prospective re-

gions of interest and determination of their frequency with respect to the carrier frequency. If a small potential error in pulse length is acceptable, picking the largest peak in the relevant region of the nutation spectrum and its mirror image on the other side of the carrier frequency will be easier to automate.

Knowledge of the  $90^\circ(^1\text{H})$  pulse length is required for the optimal performance of a myriad of NMR experiments. Automation of its determination is important for experiments performed non-interactively with automatic sample changers and in LC-NMR, when different solvent and buffer compositions result in different ionic strength. In addition, automated determination of the  $90^\circ(^1\text{H})$  pulse length widens the scope of NMR spectroscopy for non-experts by automating the entire setup of NMR experiments, since parameter sets are usually available that only require an update of the  $90^\circ(^1\text{H})$  pulse length. Tuning, matching, and shimming of the probe head is already highly automated on current NMR systems. The nutation method may also be used for the determination of  $^{13}\text{C}$ - and  $^{15}\text{N}$ -pulses, if a sufficiently intense signal is available. These pulse lengths, however, are much less sensitive with regard to sample properties, and heteronuclear experiments usually can be successfully performed by the use of default values for  $^{13}\text{C}$ - and  $^{15}\text{N}$ -pulse lengths.

Finally, the exceptional speed with which the  $90^\circ$  pulse can be determined by nutation spectroscopy can be exploited, for example, to minimize the dead time in H/D exchange experiments which are initiated by dissolving a dry compound in  $\text{D}_2\text{O}$ .

### 3. Conclusions

Although nutation spectroscopy may not be the only way to automate the determination of the  $90^\circ(^1\text{H})$  pulse, it is particularly fast, straightforward, and accurate. It accelerates an essential but uninteresting part of the setup of advanced NMR experiments.

### Acknowledgments

We thank Drs. Wolfgang Bermel and Jens Renner for a critical reading of the manuscript and the organic compound  $\text{C}_{34}\text{H}_{53}\text{O}_5\text{Si}_2$ , respectively. Financial support by the Australian Research Council by a Federation Fellowship (for G.O.), a project grant and for the 800 MHz NMR spectrometer is gratefully acknowledged.

### References

- [1] P.A. Keifer, 90 degrees pulse width calibrations: How to read a pulse width array? *Concept Magn. Res.* 11 (1999) 165–180.

- [2] D. Barnaal, I.J. Lowe, Effects of rotating magnetic fields on free-induction decay shapes, *Phys. Rev. Lett.* 11 (1963) 258–260.
- [3] N.C. Nielsen, H. Bildsøe, H.J. Jakobson, O.W. Sørensen, Two-dimensional pulse techniques for determination of radio-frequency field strengths and proton multiplicities in NMR spectroscopy, *J. Am. Chem. Soc.* 109 (1987) 901–902.
- [4] I.V. Mastikhin, Rapid determination of the RF pulse flip angle and spin–lattice relaxation time for materials imaging, *J. Magn. Reson.* 172 (2005) 231–237.
- [5] K.R. Metz, J.P. Boehmer, J.L. Bowers, J.R. Moore, Rapid rotating-frame imaging using an RF pulse-train (RIPT), *J. Magn. Reson. B* 103 (1994) 152–161.
- [6] J.P. Jesson, P. Meakin, G. Kneissel, Homonuclear decoupling and peak elimination in Fourier-transform nuclear magnetic resonance, *J. Am. Chem. Soc.* 95 (1973) 618–620.
- [7] C.S. Yannoni, R.D. Kendrick, NMR nutation spectroscopy—a method for determining molecular-geometry in amorphous solids, *J. Chem. Phys.* 74 (1981) 747–749.
- [8] D. Moskau, Application of real time digital filters in NMR spectroscopy, *Concepts Magn. Res.* 15 (2002) 164–176.
- [9] D.I. Hoult, Solvent peak saturation with single-phase and quadrature Fourier transformation, *J. Magn. Reson.* 21 (1976) 337–347.
- [10] P.S.C. Wu, G. Otting, SWET for secure water suppression on probes with high quality factor, *J. Biomol. NMR* (2005), in press.
- [11] F.H. Larsen, P. Daugaard, H.J. Jakobsen, N.C. Nielsen, Improving RF field homogeneity in solid-state MAS NMR using a loop-gap resonator, *J. Magn. Reson. A* 115 (1995) 283–286.

**9. Appendix: III (VI)**

# Paper III

**P. S. C. Wu, G. Otting**

“SWET for secure water suppression on probes with high quality factor”

Reproduced with permission from

*Journal of Biomolecular NMR* **32**, 243-250 (2005)

Copyright 2005 Springer Netherlands

Article

## SWET for secure water suppression on probes with high quality factor

Peter S. C. Wu & Gottfried Otting\*

Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia

Received 4 April 2005; Accepted 26 May 2005

**Key words:** COSY, preirradiation, radiation damping, water suppression, WET

### Abstract

Water suppression by selective preirradiation is increasingly difficult to achieve on probeheads with high quality factor because of the opposing forces of radiation damping. Here we show that a simple modification to the WET scheme provides reliable water suppression in aqueous solutions of proteins and peptides with minimal saturation of the  $H^\alpha$  protons. The scheme is shown to work also with dilute peptide solutions. It is recommended to maintain the water suppression during the evolution time of COSY experiments by weak selective irradiation that causes only minimal Bloch-Siegert shifts. The new water-suppression scheme suppresses the water magnetization by spatial scrambling. Traditional water suppression by preirradiation is similarly based more on water scrambling due to the radiofrequency inhomogeneity than on relaxation effects.

**Abbreviations:** DANTE – delays alternating with nutations for tailored excitation; rf – radiofrequency; SWET – secure WET; WANTED – Water selective DANTE using gradient; WET – water suppression enhanced through  $T_1$  effects

### Introduction

Among many different water suppression schemes (Price, 1999), water suppression by selective irradiation (Campbell et al., 1974) stands out as one of the most universally applicable schemes. The high quality factor,  $Q$ , of modern probeheads on high-field NMR spectrometers, however, makes water suppression by selective irradiation an increasingly difficult proposition due to the counteracting force of radiation damping (Abragam, 1961; Guéron et al., 1991; Vlassenbroek et al., 1995; Augustine, 2002). The radiation damping field created by transverse water magnetization increases with the static magnetic field and the quality factor of the probehead. In order to rotate the water magnetization by on-resonance irradiation, the radiofre-

quency (rf) power needs to overcome the radiation damping field. Using a room-temperature triple-resonance, triple-axes gradient probehead on our 800 MHz NMR spectrometer, preirradiation field strengths of 75 Hz can be required to overcome the radiation damping field created by a dilute peptide sample in 90%  $H_2O$ /10%  $D_2O$  and obtain adequate water suppression. Clearly, such a strong preirradiation field causes considerable saturation of  $H^\alpha$  resonances and, hence, bleaching in homonuclear 2D NMR spectra (Wider et al., 1983). The problem becomes correspondingly more severe on cryogenic probes.

Due to the vagaries of water suppression by selective pre-irradiation, the much more robust Watergate scheme (Piotto et al., 1992; Sklenář et al., 1993) has become very popular. Unfortunately, it fails as a water suppression scheme in COSY (Aue et al., 1976) and DQF-COSY (Piantini et al., 1982) experiments, because the basic

\*To whom correspondence should be addressed. E-mail: gottfried.otting@anu.edu.au.

phase cycles of these experiments occasionally invert the water magnetization. Inverted water magnetization presents an unstable situation, where a minor disturbance creating some transverse magnetization triggers the return of the water magnetization to its equilibrium position along the positive  $z$ -axis by radiation damping, with fully transverse magnetization on the way (Otting and Liepinsh, 1995a; Augustine, 2002). In probeheads with high  $Q$ -factor this happens quickly and during the acquisition time of the COSY experiments. Although radiation damping during acquisition can, in principle, be suppressed by Q-switching (Maas et al., 1995), bipolar gradients applied between individual sampling points (Zhang and Gorenstein, 1996), addition of glycine in high concentrations (Rodriguez et al., 2002) or a feedback-loop to counteract radiation damping (Broekert and Jeneer, 1995; Louis-Joseph et al., 1995), all these schemes either reduce sensitivity, require special sample conditions or depend on specialized hardware that is not widely available. Furthermore, the non-uniform excitation profile of Watergate would affect any  $H^{\alpha}$ - $H^N$  COSY cross-peak for which the  $H^{\alpha}$  spin is not rotated by a multiple of  $180^{\circ}$  during the Watergate sequence. DQF-COSY spectra recorded with gradient selection during the double-quantum filter deliver outstanding water suppression (Hurd, 1990; Davis et al., 1991; John et al., 1992; van Zijl et al., 1995), but at the price of up to 4-fold reduced sensitivity compared with conventional COSY spectra. A similar reduction in sensitivity is observed in ZZCOSY (Zuiderweg, 1987).

Here we propose a variant of the WET scheme (Ogg et al., 1994; Smallcombe et al., 1995) as a generally applicable strategy to replace water saturation by selective preirradiation. The original WET scheme is based on a series of water-selective excitation pulses followed by pulsed field gradients to defocus the transverse water magnetization. On systems with high  $Q$ -factor, water-selective pulses are difficult to apply and they must be sufficiently intense to overcome radiation damping (Otting, 1997; Cutting et al., 2000). In our approach, each selective pulse is broken up in the DANTE fashion, inserting bipolar pulsed-field gradients in the delays (Böckmann and Guittet, 1996). The resulting scheme, dubbed secure WET (SWET), provides reliable water suppression with much less power than selective water preirradiation. In

addition, the Bloch-Siegert shifts (Ramsey, 1955) resulting from water irradiation during the evolution time of COSY experiments (Wider et al., 1983) were assessed and the water suppression mechanism behind conventional water preirradiation was investigated.

## Materials and methods

Experiments were performed on a Bruker Avance 800 MHz NMR spectrometer equipped with a triple-resonance ( $^1H/^{13}C/^{15}N$ ) probe operating at room-temperature. Measurements were performed at  $25^{\circ}C$  using a 3.5 mM sample of hen egg-white lysozyme in 90%  $H_2O/10\%$   $D_2O$  at pH 7.0 and a 100  $\mu M$  sample of C-peptide (31 residues, Ohtomo et al., 1998) in 90%  $H_2O/10\%$   $D_2O$  at pH 6.9, using conventional 5 mm NMR tubes. In all experiments, the carrier frequency was at the water frequency. The radiation damping field strength was determined by fitting the FID observed after a  $90^{\circ}$  pulse with the equation (Mao et al., 1994)

$$M_y(t) = M_0 \operatorname{sech}(t/T_r) \quad (1)$$

where  $M_y$  is the magnetization along the  $y$  axis,  $M_0$  the equilibrium magnetization,  $T_r$  the radiation damping time constant, and  $t$  the observation time of the FID. Measurements of the signal-to-noise ratio as a function of receiver gain, using the standard ethylbenzene sample in chloroform, showed that the sensitivity was invariant for receiver gain settings above 512, with about 15% loss in sensitivity for a receiver gain of 128. A receiver gain of 128 was subsequently deemed acceptable. Nutation experiments were performed by pulsing during the acquisition in homogated decoupling mode with a duty cycle of 20%. COSY spectra were baseline corrected in the spectral region  $\delta_1 = 1.0$ – $6.5$  ppm/ $\delta_2 = 5.5$ – $10.0$  ppm by subtracting 5th order polynomials in both dimensions in order to remove dispersive tails from the diagonal peaks.

## Results

### *Radiation damping field strength*

The rate of rotation  $\omega_{rd}$  of the water magnetization due to the radiation damping field depends on the

angle  $\theta$  between the water magnetization and the main magnetic field (the  $z$ -axis) (Abragam, 1961)

$$\omega_{rd} = \frac{-\sin \theta}{T_r} \quad (2)$$

On our spectrometer, fitting of the FID observed for a 90% H<sub>2</sub>O/10% D<sub>2</sub>O sample after a 90° pulse (Equation 1) yielded a  $T_r$  value of 4.3 ms, corresponding to a rotational frequency  $\omega_{rd}/2\pi$  of about 37 Hz for fully transverse water magnetization. Any selective water irradiation scheme must be applied with a higher field strength to overcome the radiation damping. For field strengths little above 37 Hz, radiation damping will significantly slow down the overall rate of rotation.

#### Conventional water preirradiation

Conventional water saturation by preirradiation typically uses about 1 s of selective irradiation before the first pulse of the experiment, combined with at least two dummy scans to achieve a steady state. A nutation experiment performed with data acquisition in homo-gated decoupling mode showed that the water signal was strongly attenuated after about 40 rotations (Figure 1). This attenuation was achieved in a time much shorter than the  $T_1$  and  $T_2$  relaxation times of the water which are about 2 s (Denisov and Halle, 2002). The same number of rotations produced a very similar attenuation also when much higher power levels were used, confirming that the decay is due to rf-inhomogeneity. Accordingly, a 2 ms trim-pulse applied at a rf-field strength of 20000 Hz provides good water suppression and is better than a 1 ms trim-pulse (Otting and Wüthrich, 1988; Otting, 1994). If relaxation effects can be neglected, Fourier transformation of the nutation data yields the rf-frequency distribution that reflects the rf-inhomogeneity. Figure 1b shows that the frequency distribution is narrow but of finite width. The width of the signals scaled with increasing nutation frequency, as expected for negligible effects from water relaxation (data not shown). The splitting of the peaks corresponds to a difference in nutation frequency of no more than 1.1%. The peaks are asymmetric with long tails of weak intensity towards zero frequency.

In principle, 40 nutations within a recovery delay of 1 s require a water-irradiation field strength of 40 Hz. In the presence of radiation

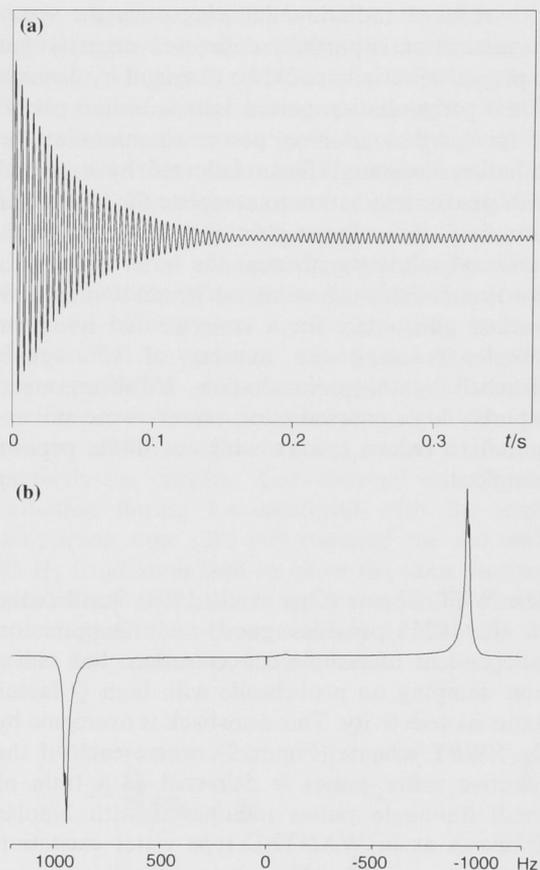


Figure 1. Nutation experiment for the measurement of rf-inhomogeneity. The data were acquired as a single FID while the <sup>1</sup>H NMR signal of a sample of 90% H<sub>2</sub>O/10% D<sub>2</sub>O was irradiated with homogated decoupling. (a) Real part of the FID. There was almost no signal in the imaginary part. (b) Fourier transform of the FID in (a), showing the distribution of nutation frequencies in the experiment.

damping, however, the field has to be higher, because the presence of a radiation damping field of up to 37 Hz slows down the overall nutation frequency according to

$$\frac{d\theta}{dt} = \omega_{rf} + \omega_{rd} = \omega_{rf} - \frac{\sin \theta}{T_r} \quad (3)$$

Integration between 0 and  $2\pi$  for  $\omega_{rf}/2\pi = 40$  Hz and  $T_r = 4.3$  ms yields an average nutation frequency of only 18 Hz.

#### Bilevel water irradiation

The necessity of at least 40 nutations for optimal water suppression by preirradiation combined

with reduced radiation damping once the water magnetization is partially defocused suggests that improved selectivity could be obtained by division of the preirradiation period into an initial period of increased irradiation power to minimize the radiation damping effects followed by a period with weaker irradiation to complete the number of nutations while allowing recovery of the magnetization of solute signals near the water resonance. We found that such a bilevel irradiation scheme worked adequately for a concentrated lysozyme sample, reducing the number of  $H^\alpha$  signals bleached by the preirradiation. However, unacceptably high preirradiation powers were still required to record spectra with our dilute peptide sample.

### Swet

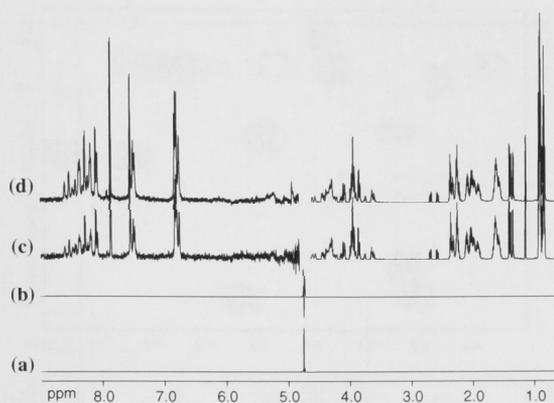
The WET scheme (Ogg et al., 1994; Smallcombe et al., 1995) provides good water suppression independent of sample concentration, but radiation damping on probeheads with high  $Q$  factor limits its selectivity. This drawback is overcome by the SWET scheme (Figure 2), where each of the selective water pulses is delivered as a train of small flip-angle pulses interleaved with bipolar gradients as in WANTED-type water excitation



**Figure 2.** COSY pulse sequence preceded by SWET. Typical parameters for the SWET scheme are: flip-angle  $\beta \approx 1^\circ$  with 1.4  $\mu\text{s}$  duration; the flip-angle must be interactively fine-tuned for maximum water suppression; amplitude of the bipolar gradients: 10 G/cm, with each pulsed field gradient of rectangular shape of 50  $\mu\text{s}$  duration and followed by a 50  $\mu\text{s}$  recovery delay; gradient pulses  $G_1$ – $G_4$ : sine shaped of 1 ms duration each and with amplitudes of 32, 16, 8 and 4 G/cm, respectively; each gradient pulse was followed by a 100  $\mu\text{s}$  recovery delay. Instead of bipolar gradients, a  $Q$ -switch could be used to suppress radiation damping (Otting and Liepinsh, 1995b). The pulse spacing must be less than twice the dwell time to avoid excitation sidebands within the spectral width. Continuous irradiation on the proton channel at a power level of 15 Hz during the  $t_1$  evolution period of the COSY pulse sequence serves to maintain the water suppression for long  $t_1$  values. Phase cycle:  $\phi = x, x, -x, -x$ ;  $\psi = x, -x$ ; receiver =  $x, x, -x, -x$ , with quadrature detection achieved by States-TPPI.

(Böckmann and Guittet, 1996). WANTED pulses suppress radiation damping and can be made highly selective. In our hands, rectangular WANTED pulses with an average field strength of 15 Hz provided adequate water suppression in the SWET scheme. This value, found by experimental optimization for best water suppression, was somewhat higher than predicted from the nominal flip angles of the SWET sequence which could be explained by finite rise and decay times of the pulses and by incomplete suppression of radiation damping by the bipolar gradients; doubling the duration of the SWET pulses by doubling the number of pulse-delay elements yielded adequate water suppression only if the pulse amplitudes were reduced by less than 50%. The flip-angles proposed for the WET sequence optimized for  $T_1$  and rf imperfections (Ogg et al., 1994) yielded the best water suppression with SWET. One-dimensional NMR spectra recorded of C-peptide with water suppression by SWET or conventional selective water irradiation with significantly increased power resulted in residual water signals of comparable size. While the intensities of the aliphatic signals of the peptide were indistinguishable, the signals of the exchangeable amide protons were more intense with SWET due to reduced saturation transfer, since SWET took up only the last 96 ms of the recovery delay between scans (Figure 3).

Since 2D NMR experiments are invariably recorded with recycling delays that allow only incomplete relaxation of the water magnetization, the remaining steady-state magnetization is less prone to radiation damping. In the case of the dilute solution of C-peptide on our 800 MHz NMR spectrometer, the steady-state water magnetization was at least 15% of the equilibrium water magnetization after recovery during 0.5 s of data acquisition and the water magnetization had grown to about 50% of its equilibrium value after 1.5 s. In this situation, a conventional WET scheme containing four Seduce-shaped pulses of 20 ms duration each (Smallcombe et al., 1995) did no longer yield adequate water suppression. In principle, the WET sequence could be implemented by replacing the first selective WET pulse by a radiation-damping compensated shaped pulse (Chen et al., 1999; Cutting et al., 2000), but this would require that the shape of the pulse is adjusted to the steady-state magnetization of the



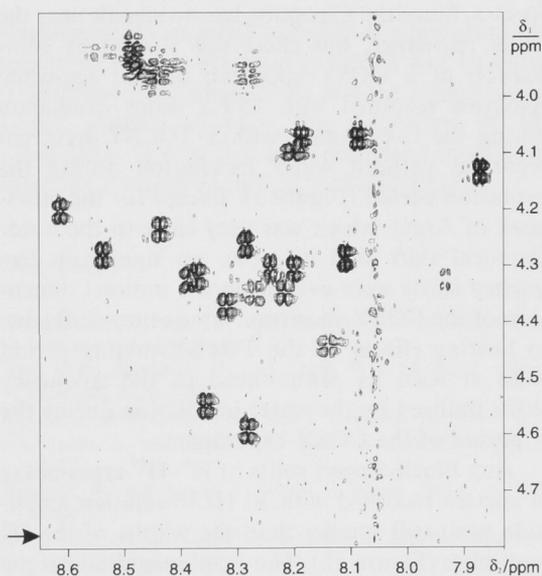
**Figure 3.** 1D NMR spectra recorded of a 100  $\mu\text{M}$  solution of C-peptide in 90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$  at pH 6.9 and 25  $^\circ\text{C}$ , using water suppression by 1 s of preirradiation with an amplitude of 75 Hz (a and c) or SWET with a total duration of 96 ms using an average pulse amplitude of 15 Hz (b and d), followed by a hard  $90^\circ$  pulse. Both spectra were recorded with an acquisition time of 0.51 s, using 128 scans, eight dummy scans and the same total recycling delay. No postacquisition processing was used to suppress the water resonance in (a) and (b), whereas the baseline was corrected manually near the water resonance in (c) and (d). (a) Residual water signal in the spectrum recorded with preirradiation. (b) Residual water signal in the spectrum recorded with SWET plotted on the same scale as (a). (c) Solute signals in the spectrum recorded with preirradiation. The vertical scale was expanded 1000-fold in the aliphatic region and 16700-fold in the amide region compared to the spectra in (a) and (b). (d) Solute signals in the spectrum recorded with SWET, plotted with the same magnification as in (c). SWET resulted in virtually identical signal intensities in the aliphatic region of the spectrum and less signal attenuation of the amide protons.

water present at the start of the WET sequence which in turn depends on the acquisition time and recycling delay used. Adequate water suppression was possible, when the shape of the first WET pulse was kept unchanged but its duration was shortened to 5 ms. This, however, would have resulted in substantial saturation of a wide band of  $\text{H}^\alpha$  resonances. SWET combines the advantages of a simple setup with selectivity of water suppression.

#### Water suppression in COSY

Although the SWET scheme provided adequate water suppression in 1D and 2D NMR experiments recorded with short evolution times, the water suppression deteriorated significantly during long  $t_1$  evolution times of COSY and DQF-COSY experiments due to the recovery of equilibrium

water magnetization. This problem was solved by application of weak selective water irradiation during  $t_1$  (Figure 2). A rf-irradiation strength of 15 Hz (i.e., of similar average power as the SWET pulses) was sufficient to maintain adequate water suppression throughout the 2D experiments. The scheme of Figure 2 was used to record a COSY spectrum of a 100  $\mu\text{M}$  solution of C-peptide (Figure 4). For consistent water suppression throughout the 2D experiment, the pulse power used for SWET was optimized using a one-dimensional experiment based on the pulse sequence SWET –  $90^\circ$ -pulse – acquisition. Baseline corrections in both dimensions removed the dispersive tails from the diagonal peaks, yielding a perfectly flat baseline. Conventional water preirradiation during 1 s combined with the same acquisition time (293 ms) required the use of a 75 Hz irradiation field to allow the same receiver gain setting and selective water irradiation during  $t_1$  would still have been necessary to maintain the water suppression for long  $t_1$  values.



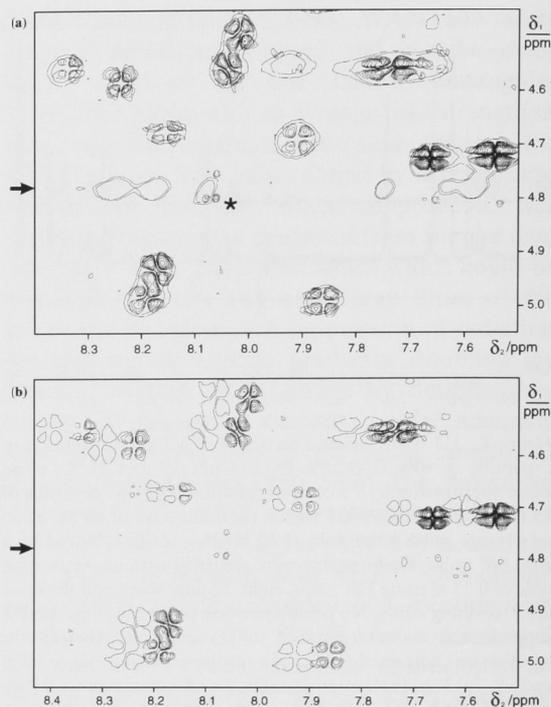
**Figure 4.** Fingerprint region of a COSY spectrum recorded with the pulse scheme of Figure 2, using a 100  $\mu\text{M}$  sample of C-peptide in 90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$  at pH 6.9 and 25  $^\circ\text{C}$ . The arrow identifies the frequency of the water resonance. The spectrum was recorded with  $t_{1\text{max}} = 100$  ms and  $t_{2\text{max}} = 293$  ms, using 28 scans per FID, spectral widths of 7000 Hz in both dimensions, a total recycling delay of 1 s (excluding the acquisition time but including the SWET duration of 96 ms) and a total experimental time of 16 h.

The shorter duration and the lower power needed for SWET compared to water suppression by selective preirradiation translates into reduced saturation of  $H^\alpha$  resonances near the chemical shift of the water. This effect was experimentally verified by COSY spectra recorded of hen egg-white lysozyme using, respectively, a 96 ms SWET sequence at 15 Hz average power or water preirradiation with 40 Hz field strength during 1 s, respectively (Supplementary material).

### Bloch-Siegert shifts

Water irradiation during evolution periods have been shown to result in frequency shifts of signals near the irradiation frequency (Wider et al., 1983). Since they are based on the same mechanism as frequency shifts due to the counter-rotating field contained in a linearly polarized radio-frequency field (Ramsey, 1955), we refer to these shifts in the following as Bloch-Siegert shifts. For the small irradiation power used in our experiments, however, these frequency shifts are small and would hardly interfere with the analysis of protein NMR spectra. Since the C-peptide has no signals near the water resonance, this effect was studied by comparison of a COSY spectrum of hen egg-white lysozyme recorded with 15 Hz water irradiation during the  $t_1$  evolution with a TOCSY spectrum recorded without water irradiation during the evolution period (Figure 5). Except for the cross-peak of Arg68 which was very close to the water chemical shift at 4.765 ppm, no significant frequency shifts were evident in the indirect dimension of the COSY spectrum. Frequency shifts due to heating effects by the TOCSY mixing period were at least as pronounced as the frequency shifts induced by the water irradiation during the  $t_1$  period of the COSY experiment.

Any Bloch-Siegert shifts of  $H^\alpha-H^N$  cross-peaks in spectra recorded with 30 Hz irradiation amplitude were still smaller than the widths of the  $H^\alpha$  multiplets (Figure 5b). The small magnitude of the frequency shifts to be expected from water irradiation during the evolution time is confirmed by the quantitative prediction presented in Figure 6. For example, a frequency shift of less than 3 Hz is predicted for any cross-peak further than 30 Hz from the water resonance, if the irradiation amplitude is 15 Hz.



**Figure 5.** Frequency shifts due to water irradiation during the evolution time. Selected spectral region from the fingerprint region of homonuclear COSY and TOCSY spectra recorded with a 3.5 mM solution of hen egg-white lysozyme in 90%  $H_2O$ /10%  $D_2O$  at pH 7.0 and 25 °C. The arrow identifies the frequency of the water resonance. (a) Superposition of a COSY spectrum recorded with continuous irradiation at 15 Hz during the  $t_1$  evolution period with a TOCSY spectrum recorded without water irradiation during  $t_1$ . The TOCSY spectrum contains additional cross-peaks at the water frequency due to chemical exchange. Only the two lowest contour lines are shown for the TOCSY spectrum. Water suppression in the TOCSY experiment was achieved by weak preirradiation during the recycle delay and a Watergate sequence following the mixing period. The  $H^\alpha-H^N$  cross-peak of Arg68 is identified by a star. (b) Superposition of the spectral regions from two COSY spectra recorded using irradiation during  $t_1$  evolution with amplitudes of 30 and 15 Hz, respectively. The spectrum with 15 Hz irradiation was plotted on a logarithmic scale with a factor of 1.4 between subsequent contour levels. Only the lowest contour level is shown for the COSY spectrum recorded with irradiation strength of 30 Hz. The latter spectrum was shifted horizontally to facilitate comparison with the COSY spectrum recorded with 15 Hz irradiation.

### Discussion and conclusions

$H^\alpha-H^N$  COSY cross-peaks contain valuable information for the resonance assignment of unlabelled proteins (Wüthrich, 1986). Since COSY spectra have a high intrinsic sensitivity, some

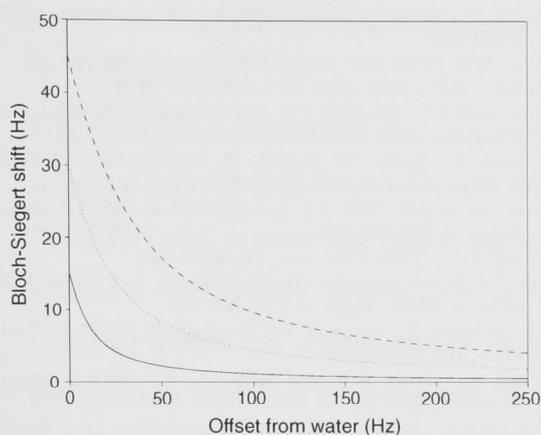


Figure 6. Prediction of the frequency shift of a protein spin as a function of its offset from the water irradiation frequency. The curves were calculated for water irradiation amplitudes of 45 (---), 30 (···) and 15 Hz (—), respectively, using the relation  $\omega_{BS} = \sqrt{\Omega^2 + \omega_1^2} - \Omega$ , where  $\omega_{BS}$  is the frequency shift,  $\Omega$  is the offset of the Larmor frequency of the spin from the water frequency, and  $\omega_1$  is the amplitude of the water irradiation field. The graph approximates the relation  $\omega_{BS} = \frac{\omega_1}{2\Omega}$  for  $\Omega \gg \omega_1$  (Ramsey, 1955).

attenuation of the protein signals by saturation transfer is often acceptable. For the spectral region containing the  $H^\alpha$ - $H^N$  cross-peaks, the dispersive tails of the diagonal peaks can readily be removed by baseline correction. The present study was prompted by the difficulty to record COSY spectra with conventional water suppression by selective preirradiation.

On NMR spectrometers 30 years ago, water suppression by selective irradiation relied on rf-inhomogeneity (Hoult, 1976) more than on the interplay of  $T_1$  and  $T_2$  relaxation (Torrey, 1949). Our present data show that this situation still holds today for the short irradiation times usually used in 2D NMR experiments. Therefore, the spatial scrambling of the water magnetization achieved by the WET or SWET schemes is conceptually not different from the spatial scrambling resulting from selective irradiation. In this situation, a WET scheme appears superior since the irradiation is shorter, giving rise to less saturation of solute resonances. For concentrated protein solutions and in the presence of compounds undergoing proton exchange with the water, the water resonance is broadened by the chemical exchange. In this case, conventional water irradiation can yield better water suppres-

sion than SWET because the effective  $T_2$  relaxation time of the water resonance is shortened by the exchange-broadening of the water, adding saturation of the water resonance (Torrey, 1949) as a significant mechanism of water suppression. With the advent of increasingly more sensitive NMR spectrometers, however, ever more dilute samples are being investigated, where the water resonance is little affected by the solute. Particularly under those circumstances, SWET provides superior water suppression compared with the conventional water preirradiation scheme by allowing the use of the same receiver gain setting with less power applied for a shorter time period resulting in reduced saturation of protein resonances.

#### Acknowledgements

Support by the Australian Research Council for the purchase of the 800 MHz NMR spectrometer and a Federation Fellowship for G.O. is gratefully acknowledged.

**Supplementary material** is available comparing the selectivity profiles achieved by SWET and 1 s water irradiation in (i) simulations and (ii) experimental COSY spectra showing the saturation of  $H^\alpha$  resonances. It is available in electronic format at: <http://dx.doi.org/10.1007/s10858-005-8531-6>.

#### References

- Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Oxford University Press.
- Aue, W.P., Bartholdi, E. and Ernst, R.R. (1976) *J. Chem. Phys.*, **64**, 2229–2246.
- Augustine, M.P. (2002) *Prog. NMR Spectrosc.*, **40**, 111–150.
- Böckmann, A. and Guittet, E. (1996) *J. Biomol. NMR*, **8**, 87–92.
- Broekert, P. and Jeneer, J. (1995) *J. Magn. Reson. A*, **113**, 60–64.
- Campbell, I.D., Dobson, C.M., Jeminet, G. and Williams, R.J. (1974) *FEBS Lett.*, **49**, 115–119.
- Chen, J.H., Jerschow, A. and Bodenhausen, G. (1999) *Chem. Phys. Lett.*, **308**, 397–402.
- Cutting, B., Chen, J.H., Moskau, D. and Bodenhausen, G. (2000) *J. Biomol. NMR*, **17**, 323–320.
- Davis, A.L., Laue, E.D., Keeler, J., Moskau, D. and Lohman, J. (1991) *J. Magn. Reson.*, **94**, 637–644.
- Denisov, V.P. and Halle, B. (2002) *J. Am. Chem. Soc.*, **124**, 10264–10265.
- Guéron, M., Plateau, P. and Decorps, M. (1991) *Prog. NMR Spectrosc.*, **23**, 135–209.

- Hoult, D.I. (1976) *J. Magn. Reson.*, **21**, 337–347.
- Hurd, R.E. (1990) *J. Magn. Reson.*, **87**, 422–428.
- John, B.K., Plant, D., Webb, P. and Hurd, R.H. (1992) *J. Magn. Reson.*, **98**, 200–206.
- Louis-Joseph, A., Abergel, D. and Lallemand, J.-Y. (1995) *J. Biomol. NMR*, **5**, 212–216.
- Maas, W.E., Laukien, F.H. and Cory, D.G. (1995) *J. Magn. Reson. A*, **113**, 274–277.
- Mao, X.A., Guo, J.X. and Ye, C.H. (1994) *Phys. Rev. B*, **49**, 15702–15711.
- Ogg, R.J., Kingsley, P.B. and Taylor, J.S. (1994) *J. Magn. Reson. Ser. B*, **104**, 1–10.
- Ohtomo, Y., Bergman, T., Johansson, B.L., Jörnvall, H. and Wahren, J. (1998) *Diabetologia*, **41**, 287–291.
- Otting, G. (1994) *J. Magn. Reson. B*, **103**, 288–291.
- Otting, G. (1997) *Prog. NMR Spectrosc.*, **31**, 259–285.
- Otting, G. and Liepinsh, E. (1995a) *J. Biomol. NMR*, **5**, 420–426.
- Otting, G. and Liepinsh, E. (1995b) *J. Magn. Reson. B*, **107**, 192–196.
- Otting, G. and Wüthrich, K. (1988) *J. Magn. Reson.*, **76**, 569–574.
- Piantini, U., Sørensen, O.W. and Ernst, R.R. (1982) *J. Am. Chem. Soc.*, **104**, 6800–6801.
- Piotto, M., Saudek, V. and Sklenář, V. (1992) *J. Biomol. NMR*, **2**, 661–665.
- Price, W.S. (1999) *Ann. Rep. NMR Spectrosc.*, **38**, 289–354.
- Ramsey, N.F. (1955) *Phys. Rev.*, **100**, 1191–1194.
- Rodriguez, J.C., Jennings, P.A. and Melacini, G. (2002) *J. Am. Chem. Soc.*, **124**, 6240–6241.
- Sklenář, V., Piotto, M., Leppik, R. and Saudek, V. (1993) *J. Magn. Reson. A*, **102**, 241–245.
- Smallcombe, S.H., Patt, S.L. and Keifer, P.A. (1995) *J. Magn. Reson. A*, **117**, 295–303.
- Torrey, H.C. (1949) *Phys. Rev.*, **76**, 1059–1068.
- Zijl, P.C.M.van, Johnson, M.O., Mori, S. and Hurd, R.E. (1995) *J. Magn. Reson. A*, **113**, 265–270.
- Vlassenbroek, A., Jeener, J. and Broekert, P. (1995) *J. Chem. Phys.*, **103**, 5886–5897.
- Wider, G., Hosur, R.V. and Wüthrich, K. (1983) *J. Magn. Reson.*, **52**, 130–135.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
- Zhang, S.M. and Gorenstein, D.G. (1996) *J. Magn. Reson. A*, **118**, 291–294.
- Zuiderweg, E.R.P. (1987) *J. Magn. Reson.*, **71**, 283–293.

## Supplementary Material

Peter S. C. Wu and Gottfried Otting

SWET for secure water suppression on probes with high quality factor

Figure S1. Comparison of the selectivity of water suppression by (a) SWET and (b) selective water preirradiation. The curves were simulated using the NMRSIM program in TOPSPIN and present the remaining z-magnetization after the two water suppression schemes. The SWET profile was approximated by a WET profile, simulating the selective pulses as continuous rectangular pulses rather than DANTE pulses and disregarding radiation damping effects. The profile was calculated by selecting the z-component of the magnetization after each of the four selective pulses, applied with flip-angles of 81.4, 101.4, 69.3 and 161.0 degrees and durations of 18.0, 22.4, 15.6 and 35.6 ms, respectively, followed by multiplication of the four profiles. The profile resulting from preirradiation was calculated by averaging ten profiles calculated for pulses of 1 s duration and amplitudes of 75.0, 75.1, 75.2, ..., 75.9 Hz, respectively, in order to simulate the radiofrequency inhomogeneity. 75 Hz was the experimentally determined field-strength required for adequate water suppression with a 100  $\mu\text{M}$  solution of C-peptide in 90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$ . A more accurate simulation of preirradiation would have to take into account the precise frequency distribution of the rf-field (Figure 1), the presence of radiation damping arising from steady-state magnetization after a number of dummy scans, and the  $T_1$  and  $T_2$  relaxation of the water magnetization during the irradiation. Several of these parameters are dependent on the sample and experimental parameters used. It is unlikely that such a simulation would change the overall conclusion that the selectivity of SWET is at least as good as that achieved by preirradiation. The selectivity of the SWET scheme could further be enhanced by the use of Seduce-shaped rather than rectangular pulse profiles as proposed for WET (Smallcombe et al., (1995) *J. Magn. Reson. A*, **117**, 295-303).

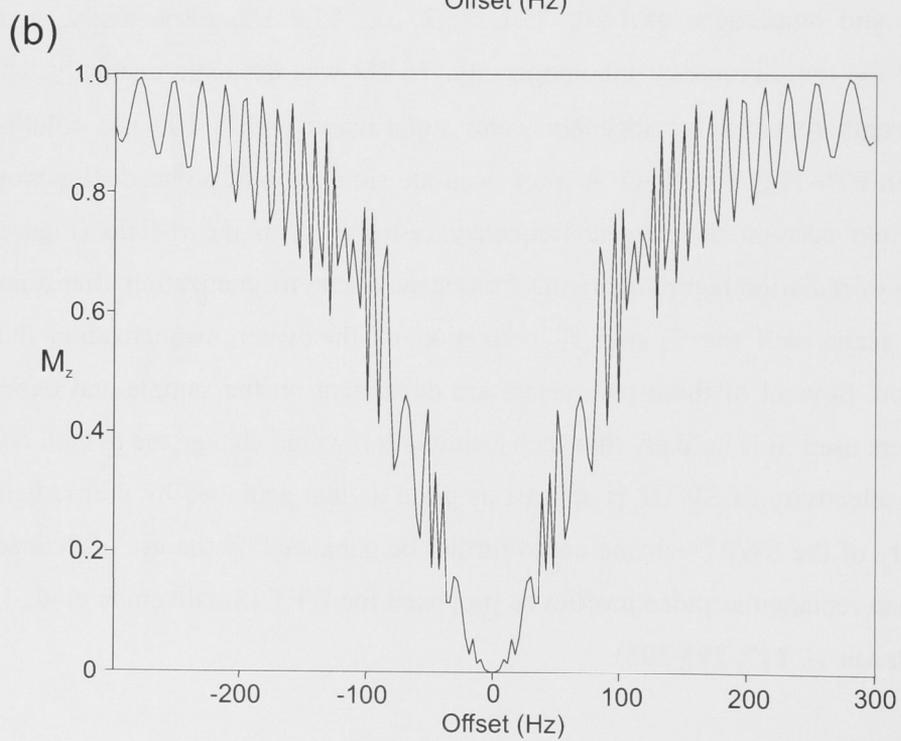
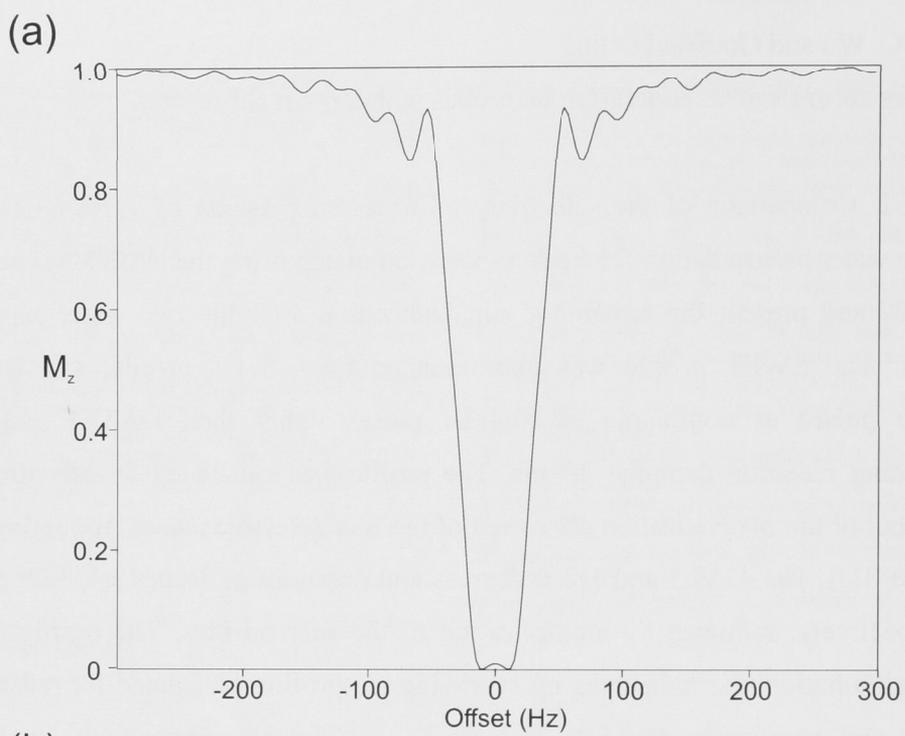


Fig. S1

Figure S2. COSY spectra recorded of a 3.6 mM solution of hen egg-white lysozyme in 90% H<sub>2</sub>O/10% D<sub>2</sub>O at pH 7.0 and 25 °C. The spectra were recorded under identical conditions on a Bruker 800 MHz NMR spectrometer, except that different water suppression schemes were used. The common parameters were:  $t_{1\max} = 51$  ms and  $t_{2\max} = 102$  ms, 16 scans per FID, sweep widths of 10000 Hz in both dimensions, recycle delay 1.1 s (excluding the acquisition time but including the water suppression period by selective water irradiation or SWET, respectively), and selective water irradiation during the evolution time with an amplitude of 15 Hz. The water signal was reduced by subtraction of a 5<sup>th</sup> order polynomial from each FID and the dispersive tails of the diagonal peaks were minimized by multiplication of the data with an unshifted sine-bell window function in both dimensions prior to Fourier transformation. No baseline correction was applied after Fourier transformation. In both spectra the contours were plotted at the same heights, using an exponential scale with a factor of 1.4 between subsequent contour levels. (a) Overview of the COSY spectrum recorded with water suppression by preirradiation during 1 s immediately prior to the first 90° pulse with an amplitude of 40 Hz. (b) Overview of the COSY spectrum recorded with water suppression by SWET during the 96 ms immediately prior to the first 90° pulse with an average irradiation amplitude of 15 Hz. The pulse sequence of Fig. 2 was used. (c) and (d) Expansions of the spectra shown in (a) and (b), respectively. The position of the water resonance in the indirect dimension is indicated by an arrow. Water suppression by SWET resulted in less attenuation of the H<sup>α</sup>-H<sup>N</sup> cross-peaks at chemical shifts near the water resonance and less saturation transfer to some of the H<sup>α</sup> resonances at other chemical shifts.

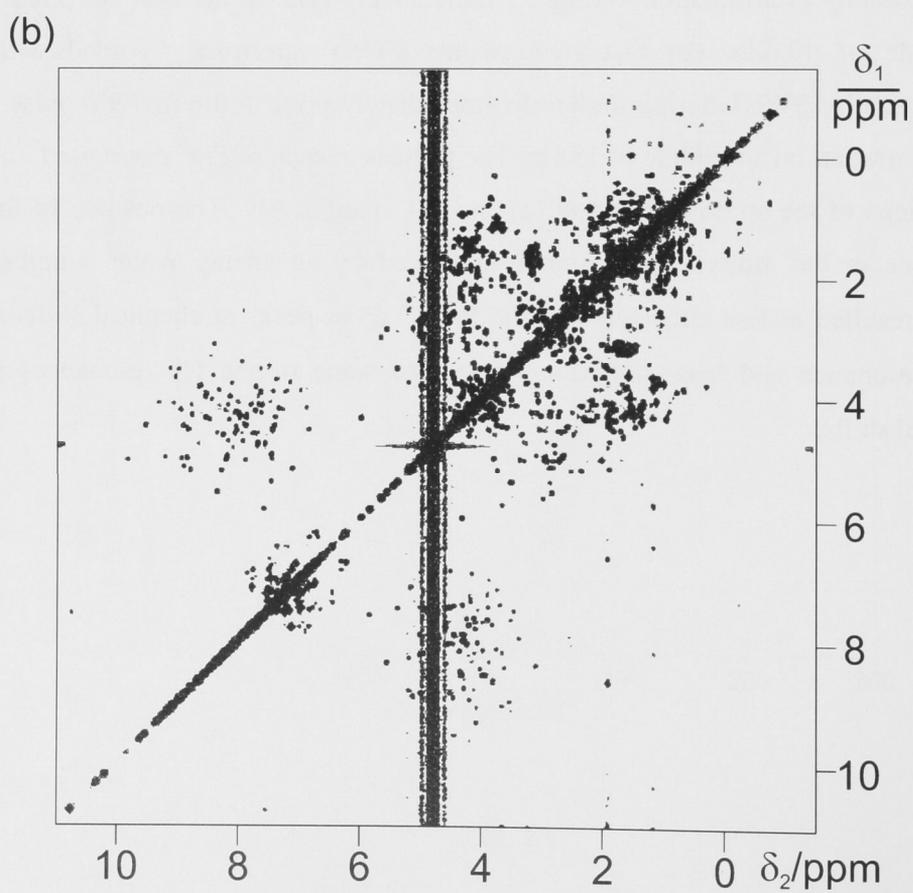
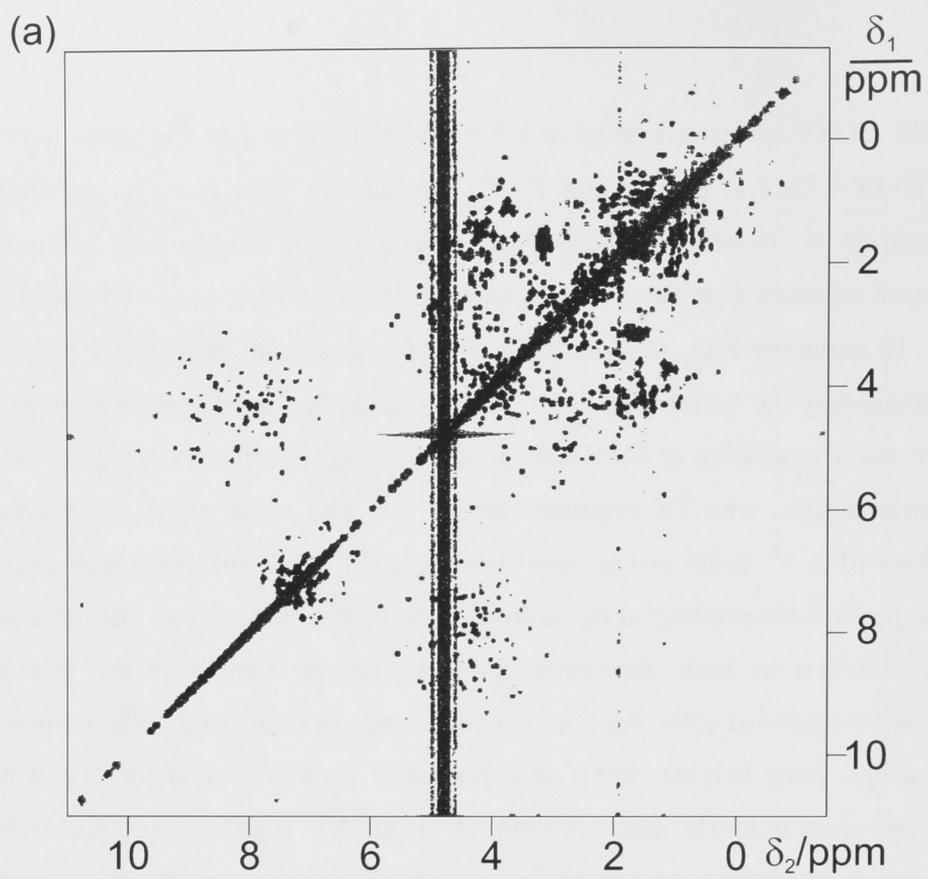


Fig. S2

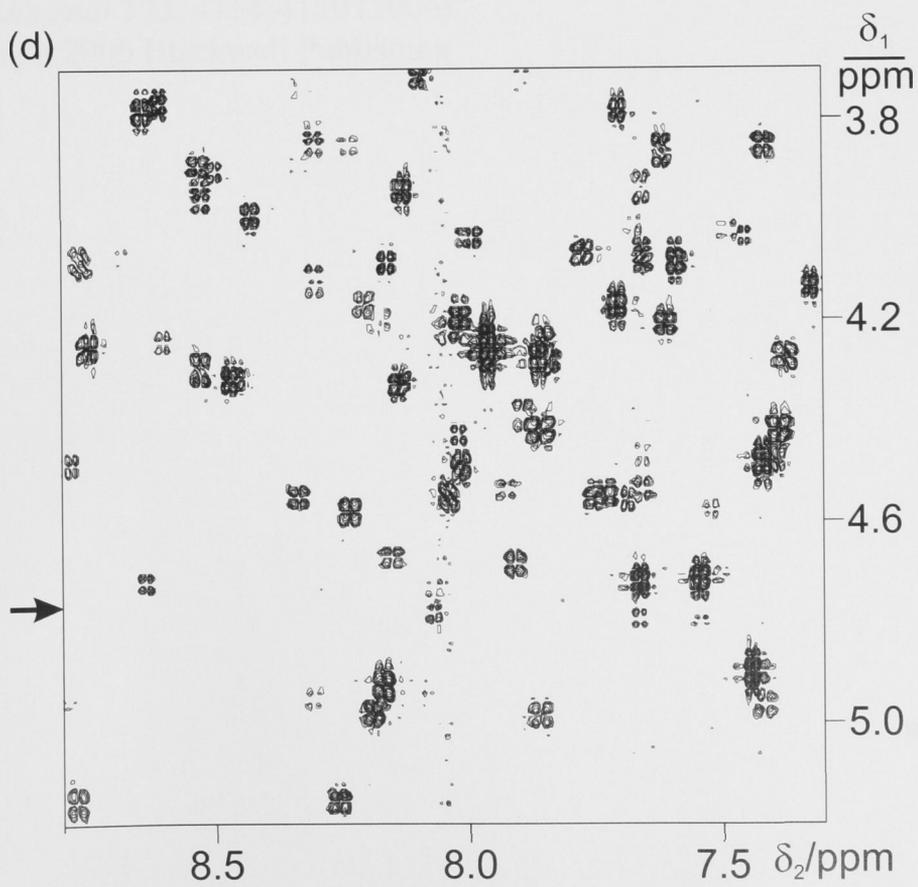
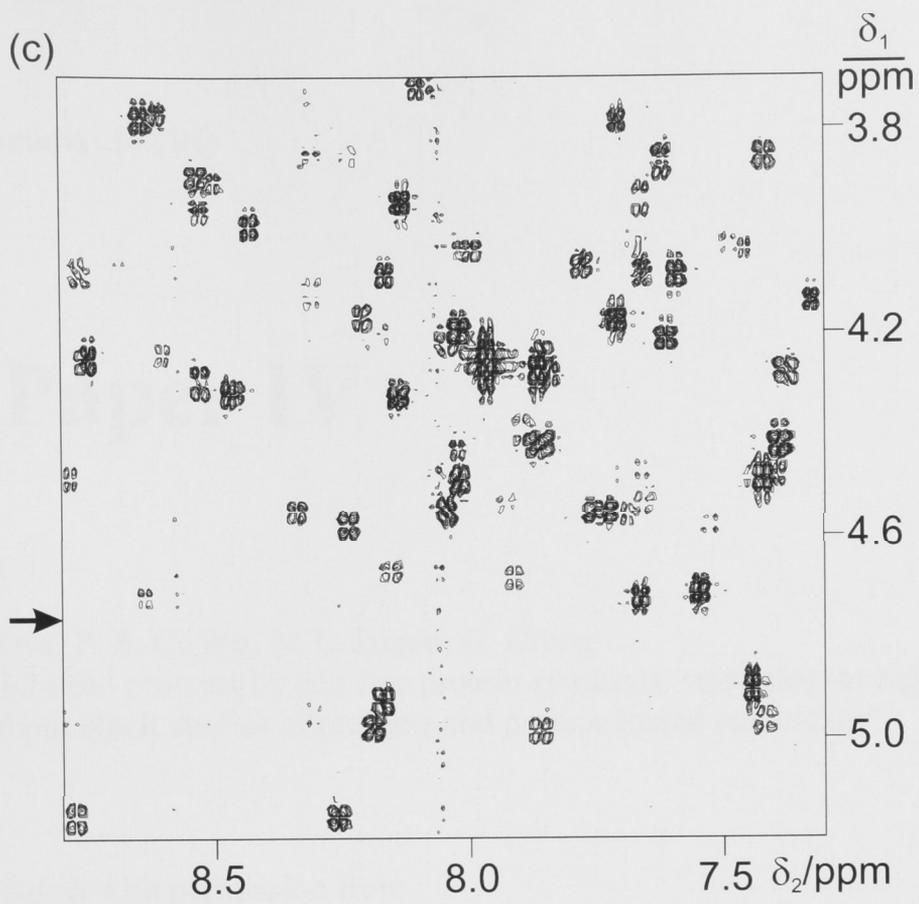


Fig. S2

## 9. Appendix: IV(VI)

# Paper IV

K. Ozawa, **P. S. C. Wu**, N. E. Dixon, G. Otting  
“<sup>15</sup>N-labelled proteins by cell free protein synthesis: strategies for high-throughput NMR studies of proteins and protein-ligand complexes”

Reproduced with permission from  
*FEBS Journal* **273**, 4154-4159 (2006)  
Copyright 2006 Blackwell Publishing

## MINIREVIEW

# <sup>15</sup>N-Labelled proteins by cell-free protein synthesis

## Strategies for high-throughput NMR studies of proteins and protein–ligand complexes

Kiyoshi Ozawa, Peter S. C. Wu, Nicholas E. Dixon and Gottfried Otting

Research School of Chemistry, Australian National University, Canberra, ACT, Australia

**Keywords**cell-free protein synthesis; combinatorial labelling; <sup>15</sup>N-HSQC; <sup>15</sup>N-labelled amino acids; protein–ligand interactions**Correspondence**G. Otting, Research School of Chemistry, Australian National University, Canberra, ACT, Australia  
Fax: +61 261250750  
Tel: +61 261256507  
E-mail: gottfried.otting@anu.edu.au  
Website: <http://rsc.anu.edu.au/~go/>

(Received 9 May 2006, accepted 23 June 2006)

doi:10.1111/j.1742-4658.2006.05433.x

[<sup>15</sup>N]-heteronuclear single quantum coherence (HSQC) spectra provide a readily accessible fingerprint of [<sup>15</sup>N]-labelled proteins, where the backbone amide group of each nonproline amino acid residue contributes a single cross-peak. Cell-free protein synthesis offers a fast and economical route to enhance the information content of [<sup>15</sup>N]-HSQC spectra by amino acid type selective [<sup>15</sup>N]-labelling. The samples can be measured without chromatographic protein purification, dilution of isotopes by transaminase activities are suppressed, and a combinatorial isotope labelling scheme can be adopted that combines reduced spectral overlap with a minimum number of samples for the identification of all [<sup>15</sup>N]-HSQC cross-peaks by amino acid residue type. These techniques are particularly powerful for tracking [<sup>15</sup>N]-HSQC cross-peaks after titration with unlabelled ligand molecules or macromolecular binding partners. In particular, combinatorial isotope labelling can provide complete cross-peak identification by amino acid type in 24 h, including protein production and NMR measurement.

**Introduction**

Cell-free protein synthesis in both the *Escherichia coli* coupled transcription-translation system and the wheat germ translation system has been remarkably improved so that milligram quantities of protein can routinely be prepared [1–6]. Compared to conventional recombinant protein production *in vivo*, cell-free protein synthesis offers a number of decisive advantages for the preparation of stable isotope labelled protein samples for analysis by NMR spectroscopy.

(a) The target protein is the only protein synthesized and labelled during the reaction. Consequently the isotope-labelled amino acids are used very efficiently, and because no new metabolic enzymes are expressed in the medium, isotope scrambling is kept to a minimum. Moreover, isotope-filtered NMR experiments allow the

selective observation of the isotope-labelled proteins without chromatographic purification.

(b) The reaction is fast. This is advantageous for the synthesis of proteins that are sensitive to proteolytic degradation and for high-throughput applications.

(c) The reaction can be carried out in small volumes. Therefore, isotope-labelled starting materials are used more efficiently and economically than for conventional *in vivo* labelling methods [7].

(d) The reaction is independent of cell growth. Therefore, toxic proteins and proteins containing non-natural amino acids can be made efficiently [8–10]. With the advent of cryogenic probe heads, heteronuclear single quantum coherence (HSQC) spectra of proteins made by cell-free expression can be recorded quickly at the concentration delivered by the reaction mixture.

**Abbreviations**

HSQC, heteronuclear single quantum coherence.

(e) The reaction mixture is accessible. This allows the synthesis of proteins in the presence of other proteins provided in excess at the start of or during the reaction, e.g., for the purpose of rescuing nascently produced insoluble proteins into soluble complexes with soluble binding partners [11].

This review summarizes our recent experience with cell-free protein synthesis, in particular with regard to the production of selectively [<sup>15</sup>N]-labelled proteins.

### Isotope scrambling

Selectively [<sup>15</sup>N]-labelled protein samples have long been made from a mixture of unlabelled and [<sup>15</sup>N]-labelled amino acids by *in vivo* protein synthesis in *E. coli* [12–15]. However, the amino acid metabolism of live *E. coli* cells can cause serious isotope scrambling for many of the amino acids, mostly due to transaminase activities [12,15–17]. In principle, this problem can be overcome by the use of auxotrophic *E. coli* strains [13], but this requires protein preparations from different strains.

Cell-free protein synthesis systems are far more inert with regard to isotope scrambling because the pool of metabolic enzymes present in the cell extract is not regenerated. Thus, cell extracts from nonauxotrophic *E. coli* strains such as A19 have been shown to yield selectively labelled proteins without significant interference from transaminases, except that conversion of [<sup>15</sup>N]aspartic acid to [<sup>15</sup>N]asparagine was still found to occur [18]. This conversion can, however, be suppressed by heat treatment of the *E. coli* S30 cell extract [7,19] or by replacing the originally recommended glutamate buffer [1] by acetate [7,18]. Different amino acids are susceptible to [<sup>15</sup>N]-scrambling in the wheat germ system than in *E. coli*. In particular, interconversion between Ala and Glu, Glu and Asp, and Glu and Gln is efficient in wheat germ extract but can effectively be suppressed by inhibitors of transaminases and glutamine synthase [20].

Among the multitude of metabolic enzymes present in the cell extract, only those leading to transfer of [<sup>15</sup>N]-amino groups to other amino acids can interfere with the subsequent NMR analysis. The NMR resonances of [<sup>15</sup>N]-amino groups, for example, are at a different chemical shift than the protein amide resonances and therefore do not interfere with the protein fingerprint represented by the amide cross-peaks in the [<sup>15</sup>N]-HSQC spectrum. Remaining free [<sup>15</sup>N]-amino acids are equally unproblematic because the amino protons of amino acids exchange too rapidly at neutral pH to yield a signal observable in [<sup>15</sup>N]-HSQC spectra. It is thus possible to obtain clean NMR spectra

directly of the reaction mixture without prior removal of low-molecular mass compounds [18,21,22].

### Selective [<sup>15</sup>N]-labelling

NMR resonance assignments and tracking of chemical shift changes is much easier if each amide cross-peak in the [<sup>15</sup>N]-HSQC spectrum of a protein can be attributed a priori to one of the 19 nonproline amino acid types. (Proline residues do not contain backbone amide protons.) Bacterial growth and *in vivo* overproduction of 19 different protein samples, each selectively [<sup>15</sup>N]-labelled with a different [<sup>15</sup>N]-amino acid, has been attempted [16] but is impractical because of transamination reactions, the expense associated with [<sup>15</sup>N]-labelled amino acids and the necessity to purify each individual sample. In contrast, cell-free systems allow the synthesis of [<sup>15</sup>N]-labelled proteins with very small quantities of [<sup>15</sup>N]-amino acids and they can be directly measured by NMR without chromatographic isolation or concentration [21]. The much improved selectivity of [<sup>15</sup>N]-labelling achieved by cell-free protein synthesis has been demonstrated for each of the 19 nonproline residues [18]. Time and expense can be drastically reduced by use of cell-free systems [11,18,21], opening many avenues for strategic applications of selectively isotope-labelled amino acids in protein production [23,24]. Because selective [<sup>15</sup>N]-amino acid labelling by cell-free protein synthesis can be carried out in parallel, it is possible in a single day to produce a complete set of 19 selectively isotope-labelled samples that are of sufficient concentration to record adequate NMR spectra in one hour per spectrum or less [10,22].

### Combinatorial selective [<sup>15</sup>N]-amino acids labelling

In general, proteins that can be produced in high yields *in vivo* are also suitable for efficient production by cell-free synthesis. In order to compensate for the increased effort and expense required for the production and selective isotope labelling of less efficiently produced proteins, a combinatorial labelling strategy can be adopted. Combinatorial labelling minimizes the number of samples that need to be prepared and analyzed in order to obtain the same information as that obtained from a much larger set of selectively labelled samples.

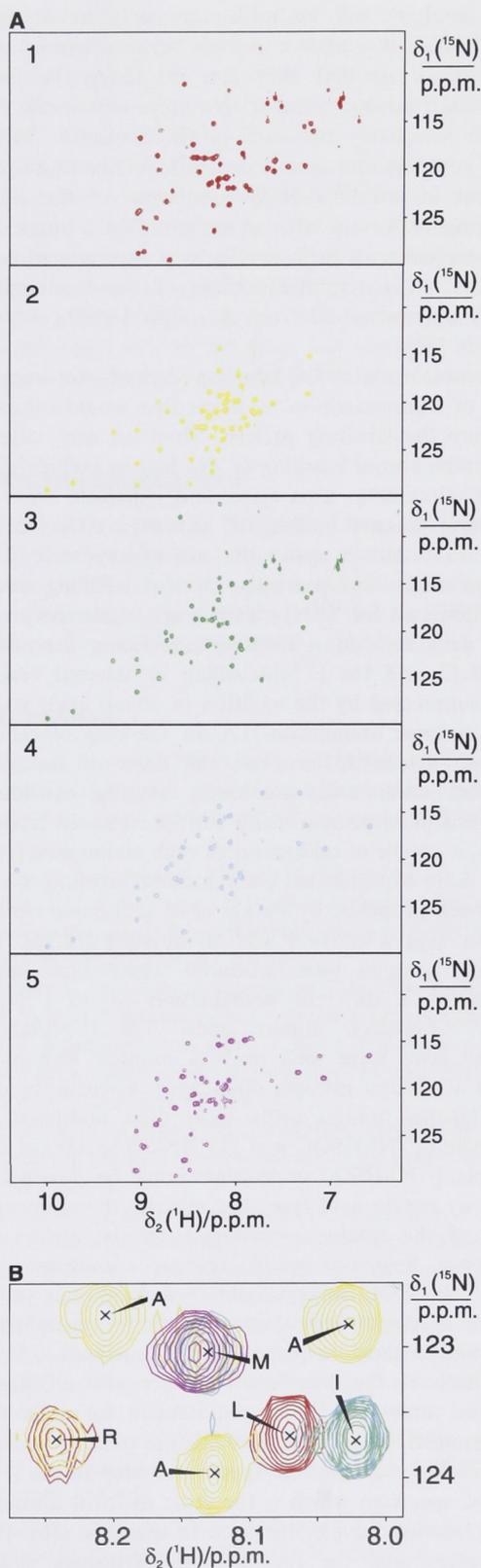
Different combinatorial strategies have been described. Figure 1 illustrates the most basic scheme, where the preparation of five samples leads to the assignment of every [<sup>15</sup>N]-HSQC cross-peak to one of

Res. Type	Sample No.					Freq. (%)
	1	2	3	4	5	
Leu	●					9.9
Ala		●				8.3
Gly			●			6.9
Ser				●		6.8
Glu					●	6.3
Val				●	●	6.7
Ile			●	●		6.0
Lys		●	●			5.6
Arg	●	●				5.5
Thr			●		●	5.4
Asp	●				●	5.3
Pro						4.7
Asn	●		●			4.2
Phe		●			●	4.1
Gln		●		●		3.9
Tyr	●			●		3.1
Met		●		●	●	2.4
His	●		●	●		2.2
Cys	●	●			●	1.3
Trp		●	●	●		1.2

**Fig. 1.** Combinatorial isotope labelling scheme. Oval symbols identify the <sup>15</sup>N-labelled amino acids used in the cell-free preparation of the five different samples. The last column displays the average amino acid abundance in proteins according to the NCBI database.

19 amino acid residue types [10]. The five samples are prepared with different combinations of [<sup>15</sup>N]-labelled amino acids. The most abundant amino acids are labelled in only one of the samples, while the least abundant amino acids are labelled in up to three of the samples. The pattern of occurrence and nonoccurrence of any particular cross-peak in the [<sup>15</sup>N]-HSQC spectra recorded of these five samples identifies the amino acid residue type associated with this cross-peak (Fig. 2).

**Fig. 2.** <sup>15</sup>N-HSQC spectra of five combinatorially <sup>15</sup>N-labelled samples of the C-terminal 16 kDa domain of the *E. coli* DNA polymerase III subunit τ. (A) Overview of the spectra. Numbers in the top left corner refer to the five different labelling patterns of Fig. 1. (B) Selected spectral region with all five spectra superimposed. The pattern of peak occurrence in the different spectra identifies the amino acid type.



This analysis will be misleading only in situations where there is complete overlap between two or more cross-peaks so that they can no longer be distinguished from one another. Notably, cross-peak overlap is less likely to occur in these spectra, because each contains only about one third of the cross-peaks present in the  $^{15}\text{N}$ -HSQC spectrum of the corresponding uniformly labelled sample. Not a single case of complete cross-peak overlap was encountered in the case of the C-terminal domain of the  $\tau$  subunit of DNA polymerase III from *E. coli*, a 16 kDa  $\alpha$ -helical protein [10].

Combinatorial  $^{15}\text{N}$ -labelling depends on suppression of transamination reactions that would otherwise obscure the labelling pattern. Thus, an early attempt of combinatorial labelling *in vivo* had to exclude glutamine, glutamate, asparagine and aspartate from the labelling scheme because of excessive cross-labelling [17]. In order to avoid the use of expensive  $^{15}\text{N}$ -amino acids, this particular *in vivo* labelling scheme was designed for ' $^{15}\text{N}$ -unlabelling', where the protein was produced on a medium containing inexpensive  $^{15}\text{NH}_4\text{Cl}$  and the  $^{15}\text{N}$ -labelling of selected residues was suppressed by the addition of amino acids at natural isotopic abundance [17]. In the case of cell-free protein synthesis, however, the costs of the  $^{15}\text{N}$ -labelled amino acids are hardly limiting, considering that adequate protein yields can be obtained from, at most, a couple of milligrams of each amino acid [18].

A more sophisticated combinatorial labelling scheme has been proposed by Parker *et al.* [25] based on dual amino acid selective  $^{13}\text{C}/^{15}\text{N}$ -labelling [12,26]. Five protein samples were produced where each sample contained a different combination of 16  $^{15}\text{N}$  or  $^{15}\text{N}/^{13}\text{C}$ -labelled amino acids. The  $^{15}\text{N}$ -labelled amino acids were used in 50% dilution with amino acids at natural isotopic abundance, whereas the doubly labelled amino acids were used undiluted. By recording  $^{15}\text{N}$ -HSQC and 2D HNCOC spectra of each sample,  $^{15}\text{N}$ -HSQC cross-peaks could be assigned not only by amino acid type, but also by the amino acid type of the residue preceding it in the amino acid sequence. Sequence specific resonance assignments of the  $^{15}\text{N}$ -HSQC peaks are obtained in this way so long as the corresponding amino acid pairs are unique in the amino acid sequence. The drawback of this approach is the significantly larger cost of doubly labelled amino acids, the requirement for more than five samples if all 20 amino acids are to be included in the labelling scheme, the spectral overlap in the  $^{15}\text{N}$ -HSQC spectrum which is the same as for a uniformly  $^{15}\text{N}$ -labelled sample, the need to quantify cross-peak intensities, and the fact that the sequence specific

assignments will almost always be incomplete because many amino acid pairs occur more than once in the amino acid sequence.

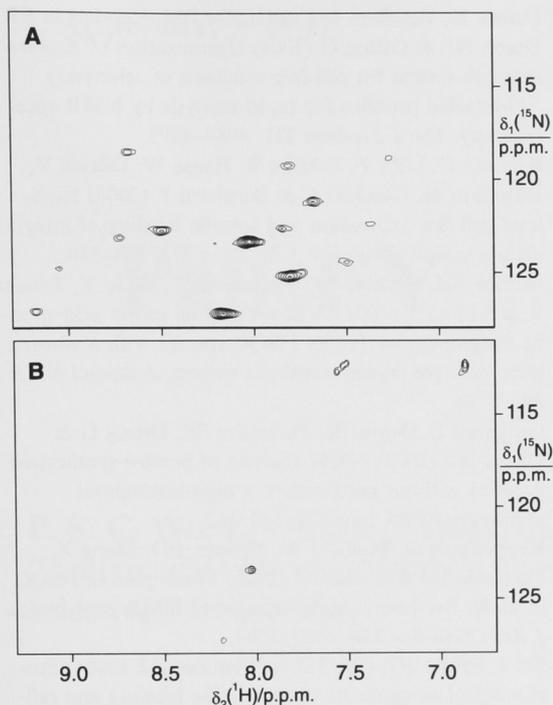
The basic combinatorial  $^{15}\text{N}$ -labelling scheme of Fig. 1 provides the benefit of improved spectral resolution, cost-efficiency and sensitivity (as no dilute labelling is employed and no experiments other than  $^{15}\text{N}$ -HSQC spectra are required). It has been shown that once the residue type assignment of the  $^{15}\text{N}$ -HSQC cross-peaks has been achieved by combinatorial  $^{15}\text{N}$ -labelling, a single 3D HNCA spectrum recorded of a uniformly  $^{15}\text{N}/^{13}\text{C}$ -labelled sample can be sufficient to complete the sequence specific resonance assignment of the backbone amides [10].

## Applications

The speed with which cell-free protein synthesis delivers  $^{15}\text{N}$ -HSQC spectra of selectively  $^{15}\text{N}$ -labelled proteins makes it an attractive tool for preliminary studies prior to the production of uniformly  $^{15}\text{N}/^{13}\text{C}$ -labelled samples for in-depth NMR analysis. Much information can be gleaned already from a single selectively labelled sample. For example, binding interactions with other (unlabelled) proteins can readily be assessed (Fig. 3), as the increase in effective molecular mass decreases the signal intensities in the  $^{15}\text{N}$ -HSQC spectrum [11].

Similarly, the presence of flexible polypeptide segments in the protein construct can be assessed by the observation of intense and narrow  $^{15}\text{N}$ -HSQC cross-peaks. Often, these unstructured segments can be localized in the amino acid sequence of the protein by their amino acid composition, which can be derived from all narrow  $^{15}\text{N}$ -HSQC cross-peaks observed in samples prepared with combinatorial  $^{15}\text{N}$ -labelling, without the need of sequence specific resonance assignments [10].

One of the most attractive applications of combinatorial  $^{15}\text{N}$ -labelling, however, may be for the identification of ligand binding sites on proteins with established sequence specific resonance assignments of the  $^{15}\text{N}$ -HSQC spectrum, where it is often difficult to assess the magnitude of chemical shift changes upon ligand binding in  $^{15}\text{N}$ -HSQC spectra of uniformly labelled proteins due to severe spectral overlap [27]. In this situation, combinatorial  $^{15}\text{N}$ -labelling allows the tracking of the cross-peaks at an effective spectral resolution equivalent to that of samples prepared with single  $^{15}\text{N}$ -labelled amino acids [10]. Although combinatorial labelling requires at least five samples to obtain complete residue type information, the protein–ligand interaction can be probed by  $^{15}\text{N}$ -HSQC



**Fig. 3.** Analysis of protein–protein interactions by NMR spectroscopy without sequence specific resonance assignment. This example shows  $^{15}\text{N}$ -HSQC spectra of selectively  $^{15}\text{N}$ -Ala labelled  $\psi$  in complex with  $\chi$  and  $\gamma$ , where  $\psi$ ,  $\chi$  and  $\gamma$  are subunits of the *E. coli* DNA polymerase III complex. (A)  $\psi$  was produced by cell-free synthesis in the presence of separately purified, unlabelled  $\chi$ ;  $\psi$  produced in the absence of  $\chi$  is insoluble. A cross-peak is observed for each of the 15 Ala residues of  $\psi$ . The most intense cross-peaks are from residues with increased mobility. The wide chemical shift distribution is indicative of a globular folded structure. The spectrum was recorded at pH 6.9 and 25 °C on a 600 MHz NMR spectrometer (Varian, Palo Alto, CA). The molecular mass of the  $\psi$ – $\chi$  complex is about 32 kDa. (B) Spectrum of the  $\psi$ – $\chi$  complex recorded in the presence of  $\gamma$ .  $\psi$  was selectively labelled with  $^{15}\text{N}$ -Ala, whereas  $\chi$  and  $\gamma$  were unlabelled. Due to the high molecular mass of the complex (about 150 kDa), most cross-peaks of  $\psi$  are broadened beyond detection, except for two cross-peaks from flexible residues. Signals near 112 p.p.m. in the  $^{15}\text{N}$  dimension arise from highly mobile  $\text{NH}_2$  groups of  $\gamma$  at natural isotopic abundance. The spectrum demonstrates that the  $\psi$ – $\chi$  complex binds to  $\gamma$ . It was recorded at pH 6.9 and 25 °C on a 800 MHz NMR spectrometer (Bruker, Karlsruhe, Germany) [11].

spectra of the reaction mixtures, which are quick to prepare [21].

## Conclusion

Over the past few years, cell-free protein synthesis has been developed into a fast and inexpensive tool for the production of stable isotope enriched proteins. Increased amino acid incorporation yields, reduced iso-

tope scrambling and easier sample handling compared to *in vivo* protein production render cell-free protein synthesis particularly attractive for high-throughput production of proteins and selective isotope labelling starting from relatively expensive isotope labelled amino acids. A straightforward combinatorial [ $^{15}\text{N}$ ]-labelling scheme carries particular promise for accelerated studies of proteins by NMR spectroscopy by assigning residue type information to every amide cross-peak observed in [ $^{15}\text{N}$ ]-HSQC spectra. We anticipate that high yield cell-free protein synthesis and combinatorial isotope labelling will become routine techniques in high-throughput NMR studies of proteins.

## Acknowledgements

GO and KO thank the Australian Research Council (ARC) for a Federation Fellowship, and an Australian Linkage (CSIRO) Postdoctoral Fellowship, respectively. Financial support by the ARC for the 800 MHz NMR facility at ANU is gratefully acknowledged.

## References

- 1 Kigawa T, Muto Y & Yokoyama S (1995) Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J Biomol NMR* **6**, 129–134.
- 2 Kigawa T, Yabuki T, Yoshida Y, Tsutsui M, Ito Y, Shibata T & Yokoyama S (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* **442**, 15–19.
- 3 Madin K, Sawasaki T, Ogasawara T & Endo Y (2000) A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: Plants apparently contain a suicide system directed at ribosomes. *Proc Natl Acad Sci USA* **97**, 559–564.
- 4 Yokoyama S (2003) Protein expression systems for structural genomics and proteomics. *Curr Opin Chem Biol* **7**, 39–43.
- 5 Vinarov DA, Lytle BL, Peterson FC, Tyler EM, Volkman BF & Markley JL (2004) Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nat Methods* **1**, 149–153.
- 6 Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM & Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52–57.
- 7 Ozawa K, Dixon NE & Otting G (2005) Cell-free synthesis of  $^{15}\text{N}$ -labeled proteins for NMR studies. *IUBMB Life* **57**, 615–622.
- 8 Kigawa T, Yamaguchi-Nunokawa E, Kodama K, Matsuda T, Yabuki T, Matsuda N, Ishitani R,

- Nureki O & Yokoyama S (2001) Selenomethionine incorporation into a protein by cell-free synthesis. *J Struct Funct Genomics* **2**, 29–35.
- 9 Ozawa K, Headlam MJ, Mouradov D, Watt SJ, Beck JL, Rodgers KJ, Dean RT, Huber T, Otting G & Dixon NE (2005) Translational incorporation of L-3,4-dihydroxyphenylalanine into proteins. *FEBS J* **272**, 3162–3171.
- 10 Wu PSC, Ozawa K, Jergic S, Su XC, Dixon NE & Otting G (2006) Amino-acid type identification in <sup>15</sup>N-HSQC spectra by combinatorial selective <sup>15</sup>N-labelling. *J Biomol NMR* **34**, 13–21.
- 11 Ozawa K, Jergic S, Crowther JA, Thompson PR, Wijffels G, Otting G & Dixon NE (2005) Cell-free *in vitro* protein synthesis in an autoinduction system for NMR studies of protein–protein interactions. *J Biomol NMR* **32**, 235–241.
- 12 Kainosho M & Tsuji T (1982) Assignment of the three methionyl carbon resonances in *Streptomyces subtilisin* inhibitor by a carbon-13 and nitrogen-15 double labeling technique. A new strategy for structural studies of proteins in solution. *Biochemistry* **21**, 6273–6279.
- 13 Le Master DM & Richards FM (1985) <sup>1</sup>H-<sup>15</sup>N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263–7268.
- 14 Griffey RH, Redfield AG, Loomis RE & Dahlquist FW (1985) Nuclear magnetic resonance observation and dynamics of specific amide protons in T4 lysozyme. *Biochemistry* **24**, 817–822.
- 15 McIntosh LP & Dahlquist FW (1990) Biosynthetic incorporation of <sup>15</sup>N and <sup>13</sup>C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q Rev Biophys* **23**, 1–38.
- 16 Yamazaki T, Yoshida M, Kanaya S, Nakamura H & Nagayama K (1991) Assignments of backbone <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N resonances and secondary structure of ribonuclease H from *Escherichia coli* by heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* **30**, 6036–6047.
- 17 Shortle D (1994) Assignment of amino acid type in <sup>1</sup>H-<sup>15</sup>N correlation spectra by labeling with <sup>14</sup>N-amino acids. *J Magn Reson B* **105**, 88–90.
- 18 Ozawa K, Headlam MJ, Schaeffer PM, Henderson BR, Dixon NE & Otting G (2004) Optimization of *Escherichia coli* system for cell-free synthesis of selectively <sup>15</sup>N-labelled proteins for rapid analysis by NMR spectroscopy. *Eur J Biochem* **271**, 4084–4093.
- 19 Klammt C, Löhr F, Schäfer B, Haase W, Dötsch V, Rüterjans H, Glaubitz C & Bernhard F (2004) High-level cell-free expression and specific labeling of integral membrane proteins. *Eur J Biochem* **271**, 568–580.
- 20 Morita EH, Shimizu M, Ogasawara T, Endo Y, Tanaka R & Kohno T (2004) A novel way of amino acid-specific assignment in <sup>1</sup>H-<sup>15</sup>N HSQC spectra with a wheat germ cell-free protein synthesis system. *J Biomol NMR* **30**, 37–45.
- 21 Guignard L, Ozawa K, Pursglove SE, Otting G & Dixon NE (2002) NMR analysis of *in vitro*-synthesized proteins without purification: a high-throughput approach. *FEBS Lett* **524**, 159–162.
- 22 Keppetipola S, Kudlicki W, Nguyen BD, Meng X, Donovan KJ & Shaka AJ (2006) From gene to HSQC in under five hours: high-throughput NMR proteomics. *J Am Chem Soc* **128**, 4508–4509.
- 23 Shi J, Pelton JG, Cho HS & Wemmer DE (2004) Protein signal assignments using specific labeling and cell-free synthesis. *J Biomol NMR* **28**, 235–247.
- 24 Trbovic N, Klammt C, Koglin A, Löhr F & Bernhard & Dötsch V (2005) Efficient strategy for the rapid backbone assignment of membrane proteins. *J Am Chem Soc* **127**, 13504–13505.
- 25 Parker MJ, Aulton-Jones M, Hounslow AM & Craven CJ (2004) A combinatorial selective labeling method for the assignment of backbone amide NMR resonances. *J Am Chem Soc* **126**, 5020–5021.
- 26 Yabuki T, Kigawa T, Dohmae N, Takio K, Terada T, Ito Y, Laue ED, Cooper JA, Kainosho M & Yokoyama S (1998) Dual amino acid-selective and site-directed stable-isotope labeling of the human *c*-Ha-Ras protein by cell-free synthesis. *J Biomol NMR* **11**, 295–306.
- 27 Emerson SD, Palermo R, Liu CM, Tilley JW, Chen L, Danho W, Madison VS, Greeley DN, Ju G & Fry DC (2003) NMR characterization of interleukin-2 in complexes with the IL-2R $\alpha$  receptor component, and with low molecular weight compounds that inhibit the IL-2/IL-R $\alpha$  interaction. *Protein Sci* **12**, 811–822.

## 9. Appendix: V (VI)

# Paper V

**P. S. C. Wu**, K. Ozawa, S. Jergic, X. C. Su, N. E. Dixon, G. Otting  
“Amino acid type identification in  $^{15}\text{N}$ -HSQC spectra by combinatorial selective  $^{15}\text{N}$ -labelling”

Reproduced with permission from  
*Journal of Biomolecular NMR* **34**, 13-21 (2006)  
Copyright 2006 Springer Netherlands

Article

## Amino-acid type identification in $^{15}\text{N}$ -HSQC spectra by combinatorial selective $^{15}\text{N}$ -labelling

Peter S. C. Wu, Kiyoshi Ozawa, Slobodan Jergic, Xun-Cheng Su, Nicholas E. Dixon & Gottfried Otting\*

Research School of Chemistry, Australian National University, Canberra, ACT, 0200, Australia

Received 9 September 2005; Accepted 2 November 2005

**Key words:**  $^{15}\text{N}$ -HSQC, cell-free protein synthesis, combinatorial  $^{15}\text{N}$ -labelling, DNA polymerase III, resonance assignment, subunit  $\tau$

### Abstract

The efficiency of cell-free protein synthesis combined with combinatorial selective  $^{15}\text{N}$ -labelling provides a method for the rapid assignment of  $^{15}\text{N}$ -HSQC cross-peaks to the 19 different non-proline amino-acid types from five  $^{15}\text{N}$ -HSQC spectra. This strategy was explored with two different constructs of the C-terminal domain V of the  $\tau$  subunit of the *Escherichia coli* DNA polymerase III holoenzyme,  $\tau_{\text{C}16}$  and  $\tau_{\text{C}14}$ . Since each of the five  $^{15}\text{N}$ -HSQC spectra contained only about one third of the cross-peaks present in uniformly labelled samples, spectral overlap was much reduced. All  $^{15}\text{N}$ -HSQC cross-peaks of the backbone amides could be assigned to the correct amino-acid type. Availability of the residue-type information greatly assisted the evaluation of the changes in chemical shifts observed for corresponding residues in  $\tau_{\text{C}16}$  vs. those in  $\tau_{\text{C}14}$ , and the analysis of the structure and mobility of the C-terminal residues present in  $\tau_{\text{C}16}$  but not in  $\tau_{\text{C}14}$ .

**Abbreviations:**  $\tau_{\text{C}14}$  – residues 499–625 of the  $\tau$  subunit of *E. coli* DNA polymerase III with an additional N-terminal methionine;  $\tau_{\text{C}16}$  – same as  $\tau_{\text{C}14}$ , but including the C-terminal 18 residues 626–643 of  $\tau$ .

### Introduction

$^{15}\text{N}$ -HSQC spectra provide well-resolved fingerprint information and are the cornerstone of backbone resonance assignments of  $^{15}\text{N}$ -labelled proteins. With the advent of high-yield cell-free protein synthesis systems, the preparation and NMR spectroscopic analysis of selectively  $^{15}\text{N}$ -labelled proteins has become both fast and inexpensive (Kigawa et al., 1995; Ozawa et al., 2005a). While maximal information could be obtained from 19 different samples, where each of the 19 non-proline residues is selectively  $^{15}\text{N}$ -labelled

(Yamazaki et al., 1991; Ozawa et al., 2004), the same information can be retrieved with much less effort by the use of combinatorial  $^{15}\text{N}$ -labelling, where several amino acids are simultaneously  $^{15}\text{N}$ -labelled in a limited number of samples (Shortle 1994; Parker et al., 2004). Combinatorial isotope labelling results in ambiguities only if chemical shift degeneracies lead to perfect superposition of the cross-peaks of two or more amino-acid residues.

Here we present a different combinatorial  $^{15}\text{N}$ -labelling scheme which delivers complete residue-type identifications for the  $^{15}\text{N}$ -HSQC cross-peaks from all 19 non-proline residues and where each  $^{15}\text{N}$ -HSQC spectrum produces only a third of the peaks generated by uniformly  $^{15}\text{N}$ -labelled protein. It was used for the structural

\*To whom correspondence should be addressed. E-mail: gottfried.otting@anu.edu.au

characterization of the C-terminal 16 kDa domain of the  $\tau$  subunit,  $\tau_C16$ , from the *Escherichia coli* replisome (Gao and McHenry, 2001). The protein is toxic to *E. coli* and could not be purified in stable form due to pronounced sensitivity with respect to proteolysis, whereas a shorter 14 kDa fragment,  $\tau_C14$ , that is missing the C-terminal 18 residues of  $\tau_C16$ , was sufficiently stable to allow the determination of its three-dimensional structure by NMR spectroscopy (X.-C. Su, S. Jergic and G. Otting, unpublished). We show that five combinatorially  $^{15}\text{N}$ -labelled samples, freshly prepared of each of the two proteins, readily identifies the  $^{15}\text{N}$ -HSQC cross-peaks of the C-terminal 18 residues, allowing their characterization from spectra containing only about one third of the cross-peaks that would be observed for a uniformly  $^{15}\text{N}$ -labelled protein.

## Materials and methods

### Sample preparation

Selectively  $^{15}\text{N}$ -labelled samples of  $\tau_C14$  and  $\tau_C16$  were synthesized in a cell-free *E. coli* coupled transcription-translation system, following a previously described protocol (Ozawa et al., 2005b), modified by heat treatment of the S30 extract prior to use as recommended by Klammt et al. (2004). Protein synthesis was programmed with plasmids pKO1296 and pSH1062 (K. Ozawa, S.M. Hamdan and N.E. Dixon, unpublished), which contain genes encoding  $\tau_C14$  and  $\tau_C16$ , respectively, in the phage T7-promoter vector pETMCSI (Neylon et al., 2000). T7 RNA polymerase was provided by non-competitive transcription/translation using plasmid pKO1166 (Ozawa et al., 2005b). Five  $^{15}\text{N}$ -labelled samples were prepared of each protein following the combinatorial labelling scheme of Table 1, using a dialysis system with 0.7 ml of reaction mixture in 7 ml of outside buffer.  $^{15}\text{N}$ -labelled amino acids (Cambridge Isotope Laboratories, Andover, MA, USA) were supplied in concentrations of 0.05, 0.15, 0.35 or 1 mM, depending on the  $K_m$  values of the respective aminoacyl-tRNA synthetases (Ozawa et al., 2004), except for Ala, which was supplied at 2 mM concentration to account for the large number of alanine residues present in the amino-acid sequence of T7 RNA polymerase and Glu, which

was supplied at 1.5 mM for improved protein yields. Unlabelled amino acids were supplied at a concentration of 1 mM. The reaction buffer contained 208 mM potassium glutamate except for the preparation of  $^{15}\text{N}$ -Glu labelled samples where only 1.5 mM of  $^{15}\text{N}$ -labelled glutamate was present. The reactions were carried out for 8 h at 37 °C. The product mixtures were subsequently centrifuged at 30000 $\times g$  for 60 min. Preparations 1–4 were dialyzed overnight at 4 °C against 2 l of NMR buffer (10 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, 0.1 mM  $\text{NaN}_3$ , pH 6.8), using the same beaker of buffer for all four samples, whereas the fifth sample was prepared later and dialyzed against freshly prepared NMR buffer. Except during NMR measurements the samples were kept frozen to minimize proteolytic digestion.

### NMR spectroscopy

Following addition of  $\text{D}_2\text{O}$  to 10% (v/v),  $^{15}\text{N}$ -HSQC spectra were recorded at 25 °C on a Bruker AV800 NMR spectrometer equipped with a  $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ -triple-resonance cryoprobe. All spectra were recorded using  $t_{1\text{max}} = 32$  ms,  $t_{2\text{max}} = 158$  ms and total recording times of 2.5 and 5 h for the  $\tau_C16$  and  $\tau_C14$  samples, respectively. The program Sparky (Goddard and Kneller, 2004) was used to overlay differently coloured contour plots of the NMR spectra recorded of the different samples.  $^3J(\text{H}^{\text{N}}, \text{H}^{\alpha})$  coupling constants of  $\tau_C16$  were measured with the CT-HMQC-HN experiment (Ponstingl and Otting, 1998), using  $T = 7.5$  ms,  $t_{1\text{max}} = 20$  ms,  $t_{2\text{max}} = 256$  ms, water suppression by SWET (Wu and Otting, 2005) and Watergate (Piotto et al., 1992) and a total recording time of 7.3 h per sample. Measured  $^3J(\text{H}^{\text{N}}, \text{H}^{\alpha})$  coupling constants were uniformly increased by 10% to account for different relaxation rates of antiphase and in-phase magnetization (Ponstingl and Otting, 1998).

## Results

### Cell-free protein synthesis

The protein yields obtained by cell-free protein synthesis were approximately 2 mg/ml of reaction mixture for samples 1–4 and 1 mg/ml for the  $^{15}\text{N}$ -glutamate labelled sample 5. This was

Table 1. Resolution-optimized combinatorial  $^{15}\text{N}$ -labelling scheme

Res. Type	Sample					Freq. (%) <sup>a</sup>
	1	2	3	4	5	
Leu	X					9.9
Ala		X				8.3
Gly			X			6.9
Ser				X		6.8
Glu					X	6.3
Val				X	X	6.7
Ile			X	X		6.0
Lys		X	X			5.6
Arg	X	X				5.5
Thr			X		X	5.4
Asp	X				X	5.3
Pro						4.7
Asn	X		X			4.2
Phe		X			X	4.1
Gln		X		X		3.9
Tyr	X			X		3.1
Met		X		X	X	2.4
His	X		X	X		2.2
Cys	X	X			X	1.3
Trp		X	X	X		1.2

<sup>a</sup> Data taken from the NCBI database. ([http://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa\\_explorer.cgi](http://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa_explorer.cgi))

sufficient to observe all  $^{15}\text{N}$ -HSQC cross-peaks with recording times of 2.5 h per spectrum without concentrating the reaction mixture prior to measurement. The peak positions of residues labelled in more than one sample were extremely well reproduced in the spectra recorded of samples 1–4, but less so in sample 5. This was attributed to the fact that sample 5 had been dialyzed against a similar but not the same sample of NMR buffer as the first four preparations. The preparation of the five samples was achieved in a day, followed by overnight dialysis into NMR buffer.

The peak intensities from amino acids labelled in more than one of the samples were closely reproduced. Differences in peak intensities, however, were observed between different amino-acid types as a consequence of undesired side reactions that diluted the pool of  $^{15}\text{N}$ -labelling for some but not all of the amino acids. The most pronounced effects were observed for glutamine residues for which the  $^{15}\text{N}$ -HSQC cross-peaks were consistently weaker (about 4-fold) than those of the other residue types. Furthermore, a transaminase

activity generated weak cross-peaks of glutamine in the samples prepared with  $^{15}\text{N}$ -glutamate, as reported previously for wheat germ cell extracts (Morita et al., 2004).

#### Combinatorial labelling scheme

A minimum of five different samples is required to assign the  $^{15}\text{N}$ -HSQC cross-peaks to the 19 different types of amino acids with backbone amide protons. Since five combinatorially labelled samples would allow the discrimination of  $2^5=32$  different amino acid types, different combinatorial labelling schemes are possible. Our labelling scheme (Table 1) was designed to minimize the number of cross-peaks in the  $^{15}\text{N}$ -HSQC spectra by labelling abundant amino acids only in a single one of the five samples. For general applicability, the scheme was based on the average amino-acid frequencies in proteins reported in the NCBI database. In addition, the scheme was designed to avoid any simultaneous  $^{15}\text{N}$ -labelling of Gln and Asn, because those residues increase the chance of signal overlap due to the cross-peaks from side-chain amides. Since backbone amides of threonine residues can sometimes overlap with the cross-peaks from side-chain amides, in at least one of the five samples Thr was labelled when Asn and Gln were not labelled and *vice versa*. Finally, the scheme labelled only one of the samples with  $^{15}\text{N}$ -Glu, since the protein yields were about 2-fold lower when the glutamate concentration, usually present at about 200 mM in the reaction buffer (Kigawa et al., 1999), was reduced to 1.5 mM. Given the amino-acid frequencies of the NCBI database, our scheme would label only about 31.5% of the residues in each of the five samples. For  $\tau_{\text{C}16}$ , the scheme resulted in five samples with, respectively, 32, 38, 21, 29 and 29% of the amino-acid residues labelled with  $^{15}\text{N}$ . The percentages of labelled residues were closely similar for  $\tau_{\text{C}14}$ . Accordingly, the overlap observed in the  $^{15}\text{N}$ -HSQC spectra of the five combinatorially labelled samples of  $\tau_{\text{C}16}$  and  $\tau_{\text{C}14}$  was greatly reduced compared to that expected for uniformly  $^{15}\text{N}$ -labelled samples (Figure 1).

#### Residue-type identification

Figure 2 illustrates how cross-peaks that are overlapped in one of the samples are well

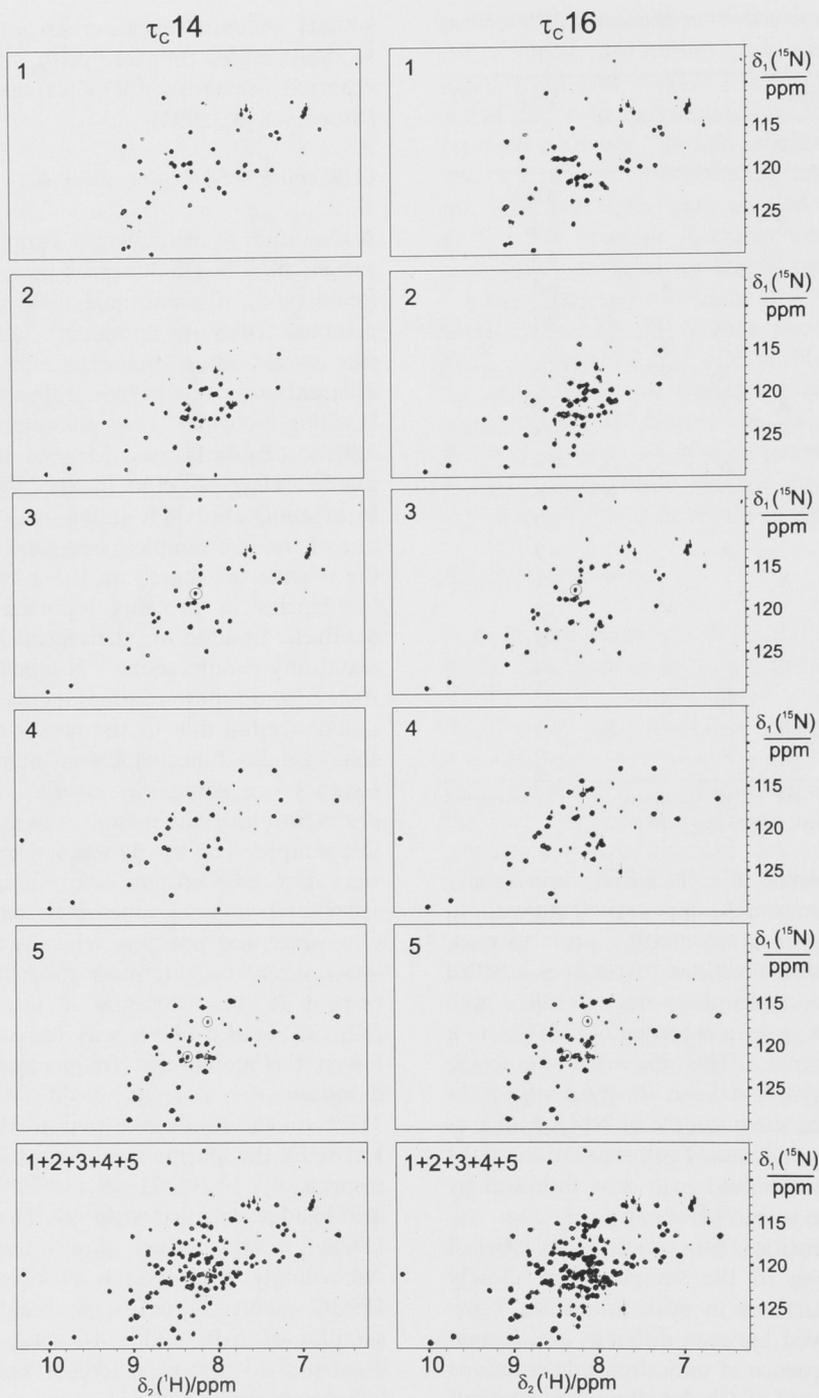


Figure 1.  $^{15}\text{N}$ -HSQC spectra recorded of samples of  $\tau_{\text{C}}14$  and  $\tau_{\text{C}}16$  produced with combinatorial  $^{15}\text{N}$ -labelling. The spectra were recorded at 25 °C and pH 6.9 at a  $^1\text{H}$ -NMR frequency of 800 MHz. Numbers in the top left corner of each spectrum correspond to the five different labelling patterns of Table 1. Circles identify cross-peaks in samples 3 and 5 that did not belong to the proteins of interest. The bottom panel illustrates the 3-fold increased density of cross-peaks expected in uniformly labelled samples.

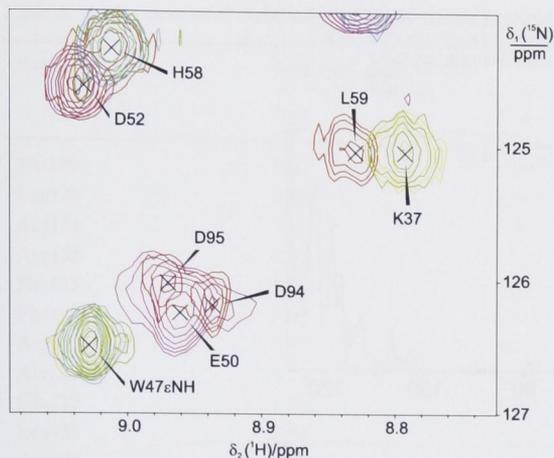


Figure 2. Spectral region selected from the  $^{15}\text{N}$ -HSQC spectra of  $\tau_{\text{C}16}$  shown in Figure 1. All five spectra of  $\tau_{\text{C}16}$  were overlaid, with spectra 1–5 plotted with red, yellow, green, cyan and purple contour lines, respectively. The selective labelling scheme resolves some of the spectral overlap between the cross-peaks of Asp94, Asp95 (purple and red contours) and Glu50 (purple contours only).

resolved in other samples made with a different combination of  $^{15}\text{N}$ -labelled amino acids. The cross-peaks of Asp94 and Asp95 (purple and red contours) can be distinguished from that of Glu50 (purple contours only). In the case of  $\tau_{\text{C}14}$  and  $\tau_{\text{C}16}$ , no  $^{15}\text{N}$ -HSQC cross-peaks were so badly overlapped that the assignments to amino-acid types would have been wrong or ambiguous. Only few signals were observed that did not seem to belong to the protein, including an apparent glycine peak in sample 3 and two glutamate peaks in sample 5 (circled in Figure 1). These additional peaks were identified as non-protein peaks by their narrower line shape and smaller intensities. The transaminase activity converting glutamate to glutamine did not lead to confusion between  $^{15}\text{N}$ -Gln and  $^{15}\text{N}$ -Met (Table 1), since all the  $^{15}\text{N}$ -Met peaks consistently had much higher intensities in samples 2, 4 and 5 than any of the  $^{15}\text{N}$ -Gln peaks.

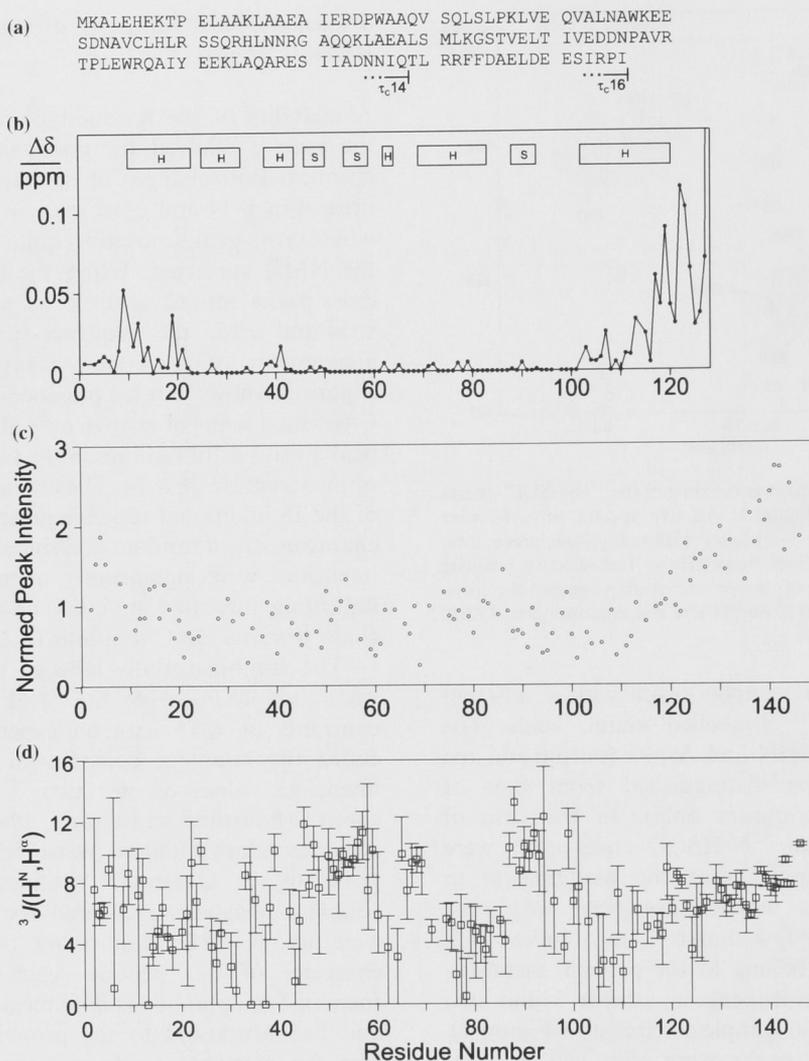
The present combinatorially labelled samples of  $\tau_{\text{C}14}$  allowed us to correct an error in the amide assignment made from a uniformly  $^{15}\text{N}/^{13}\text{C}$ -labelled sample (X.-C. Su and G. Otting, unpublished). The error involved a swap of the  $^{15}\text{N}$ -frequencies of His6 and Glu7 (Figure 3a), where chemical shift degeneracies had resulted in ambiguities in the HNCACB/CBCA(CO)NH pair of 3D-NMR experiments.

### Sequence-specific resonance assignments and structure of $\tau_{\text{C}16}$

Availability of the residue-type identifications of the spectra allowed the confident evaluation of chemical shift changes of corresponding residues present in  $\tau_{\text{C}14}$  and  $\tau_{\text{C}16}$ , even in those situations where cross-peaks moved to quite new positions in the NMR spectrum. Using the assumption that cross-peaks moved as little as possible between  $\tau_{\text{C}14}$  and  $\tau_{\text{C}16}$ , the sequence-specific resonance assignments of  $\tau_{\text{C}14}$  were transferred to  $\tau_{\text{C}16}$ . Figure 3b shows that the presence of the additional C-terminal segment in  $\tau_{\text{C}16}$  only affected the cross-peak positions of residues in the first and last helix of the structure of  $\tau_{\text{C}14}$ . The amide chemical shifts of the 18 additional residues present in  $\tau_{\text{C}16}$  were characteristic of random-coil shifts and their signal intensities were significantly increased (Table 2), indicating increased mobility of the polypeptide chain towards the C-terminus (Figure 3c).

The combinatorially labelled protein samples were used to measure the  $^3J(\text{H}^{\text{N}}, \text{H}^{\alpha})$  coupling constants of  $\tau_{\text{C}16}$  with an experiment that encodes the coupling constant in different peak intensities observed in two  $^{15}\text{N}$ -HMQC-type spectra (Ponstingl and Otting, 1998). Particularly accurate values could be measured for the intense peaks of the C-terminal residues. Most of the coupling constants measured for these residues were about 7 Hz, confirming the random-coil character of this peptide segment (Figure 3d). Increased coupling constants measured for Ile143 and Ile146 attested to the propensity of isoleucine for extended conformations. Notably, the  $^3J(\text{H}^{\text{N}}, \text{H}^{\alpha})$  couplings of the C-terminal segment were overestimated due to the multiplication factor of 1.1 used to compensate for differential relaxation between antiphase and in-phase magnetization (Ponstingl and Otting, 1998). This multiplication factor is appropriate for a rotational correlation time of more than 10 ns but not for highly mobile peptide segments.

In conclusion, all available data indicate that the C-terminal segment of  $\tau_{\text{C}16}$  is a random-coil peptide, and the small chemical shift changes observed for the N-terminal helix probably reflect non-specific transient interactions arising from the close proximity between the N- and C-terminal helices in the three-dimensional structure of  $\tau_{\text{C}14}$  (X.-C. Su and G. Otting, unpublished).



**Figure 3.** Amino-acid sequences of  $\tau_{C14}$  and  $\tau_{C16}$  and experimental structural information plotted against residue number. Lys2 of  $\tau_{C14}$  and  $\tau_{C16}$  corresponds to Lys499 in full-length  $\tau$ . (a) Amino-acid sequences of  $\tau_{C14}$  and  $\tau_{C16}$ . The two sequences differ only by the C-terminal 18 residues as indicated. In the present work, the residues were numbered Met1–Gln128 for  $\tau_{C14}$  and Met1–Ile146 for  $\tau_{C16}$ . (b) Changes in chemical shifts of the N-terminal 128 residues of  $\tau_{C16}$  compared with the same residues in  $\tau_{C14}$ . The vertical axis reports  $[(\Delta\delta_H)^2 + (0.1\Delta\delta_N)^2]^{0.5}$ . Bars indicate the location of helices (H) and  $\beta$ -strands (S) in the structure of  $\tau_{C14}$  (X.-C. Su and G. Otting, unpublished). (c) Intensities of  $^{15}\text{N}$ -HSQC cross-peaks of the backbone amides of  $\tau_{C16}$ . Different labelling efficiencies of different amino-acid types were accounted for by normalization of each of the peak intensities by the average for the corresponding amino-acid type. Data are shown only for residue types that occur at least five times in the amino-acid sequence and therefore allow adequate normalization. See Table 2 for the assignments of the C-terminal 18 residues used. (d)  $^3J(\text{H}^{\text{N}}, \text{H}^{\alpha})$  coupling constants measured from the CT-HMQC-HN experiment.

## Discussion

Complete residue-type identification of the backbone amide cross-peaks of  $^{15}\text{N}$ -HSQC spectra greatly facilitates sequence-specific resonance assignments. More important, it greatly enhances

the assessment of chemical shift changes observed in crowded spectral regions upon titration with binding partners and in mutant or chemically modified proteins, as illustrated by the comparison between  $\tau_{C16}$  and  $\tau_{C14}$ . Completeness of residue-type identification is of key importance for these applications.

Table 2. Assignment of C-terminal residues of  $\tau_C16^a$ 

Residue	Chemical shifts (ppm)	
	$^1\text{H}^N$	$^{15}\text{N}$
Thr129	8.03	115.0
Leu130	8.07	123.7
Arg131	8.10	120.9
Arg132	8.19	121.0
Phe133	8.11	120.1
Phe134	8.09	119.8
Asp135	8.19	121.7
Ala136	8.13	124.0
Glu137	8.35	119.0
Leu138	8.03	121.9
Asp139	8.25	121.3
Glu140	8.40	121.6
Glu141	8.43	120.7
Ser142	8.15	116.0
Ile143	7.93	121.6
Arg144	8.23	126.4
Ile146	7.71	125.2

<sup>a</sup> Sequence-specific assignments for residues occurring more than once in the sequence are tentative and based on the assumption of increasing peak heights towards the C-terminus due to increased mobility (Figure 3c).

*In vivo* production, purification and measurement of at least 15 different samples, each containing just one of the common amino acids in  $^{15}\text{N}$ -labelled form, has been achieved (Yamazaki et al., 1991), but it is both time-consuming and costly. Pulse sequences have been developed to achieve the selective observation of  $^{15}\text{N}$ -HSQC-type cross-peaks from different amino-acid types in  $^{15}\text{N}/^{13}\text{C}$ -labelled proteins (Yamazaki et al., 1995; Schmieder et al., 1998; Schubert et al., 1999, 2001a, b, c, 2005). Other experiments were designed for the classification by groups of amino-acid types or by residue type of the preceding residue (Gehring and Guittet, 1995; Tashiro et al., 1995; Dötsch and Wagner, 1996). Most of these experiments suffer from poor sensitivity and are applicable only to small proteins for which backbone resonance assignments are anyway easy to obtain. With the advent of efficient cell-free synthesis systems that require only small amounts of  $^{15}\text{N}$ -labelled amino acids and no chromatographic protein purification prior to NMR measurement (Guignard et al., 2002; Ozawa et al., 2004), the preparation of  $^{15}\text{N}$ -labelled protein samples has become much more attractive. Combinatorial

$^{15}\text{N}$ -labelling further shortens the sample preparation and analysis time. Although erroneous residue-type identifications could arise if two cross-peaks precisely overlap, not a single example of this occurred in the 146-residue protein  $\tau_C16$ .

Previous combinatorial labelling schemes (Shortle 1994; Parker et al., 2004; Trbovic et al., 2005) did not attempt residue-type identifications for all 19 non-proline residues. In the most ambitious scheme (Parker et al., 2004), all residues are labelled with  $^{15}\text{N}$  and pairs of sequential residues are identified using protein samples prepared with  $^{15}\text{N}$ - and  $^{13}\text{C}$ -labelled amino acids in different combinations and concentrations. As a drawback, spectral overlap is reduced only in the 2D-HNCO spectra required to identify the type of the previous residue and cross-peak intensities must be quantified in  $^{15}\text{N}$ -HSQC spectra that are not simplified compared to spectra of uniformly labelled samples. Since multiple occurrences of amino-acid pairs are common in the amino-acid sequences of proteins, the scheme by Parker et al. (2004) is not sufficient for complete sequence-specific backbone assignments.

Although our combinatorial labelling scheme yields no information about the type of the preceding residue, it appears more attractive: (i) It minimizes the number of cross-peaks observed in each spectrum, thereby improving spectral resolution. (ii) It only relies on sensitive  $^{15}\text{N}$ - $^1\text{H}$  correlation spectra, making it more applicable for large proteins than schemes requiring 2D versions of 3D triple-resonance experiments (Parker et al., 2004; Shi et al., 2004; Trbovic et al., 2005). (iii) It uses the full cross-peak sensitivity since no partial labelling is required. (iv) It is highly cost-efficient, since it requires  $^{15}\text{N}$ -labelled and unlabelled amino acids only. In the case of  $\tau_C14$ , we obtained complete and correct sequence-specific assignments by combining the residue-type identification derived from combinatorial labelling with a single 3D-HNCA spectrum of a uniformly  $^{15}\text{N}/^{13}\text{C}$ -labelled sample (data not shown). Combinatorial labelling thus presents an efficient strategy to reduce the number of 3D NMR experiments usually recorded for backbone resonance assignments.

In order to minimize the number of cross-peaks observed in each sample, it may be worthwhile to optimize our labelling scheme for the amino-acid abundances in the individual protein of interest. This can be achieved by replacing the amino-acids in the left column of Table 1 with those in the

protein sorted in descending order of abundance, while maintaining the constraints discussed above (e.g. minimal overlap with side-chain NH<sub>2</sub> groups).

Using exclusively inexpensive <sup>15</sup>N-labelled amino-acids in the combinatorial labelling scheme and a total of 10 <sup>15</sup>N-HSQC spectra of τ<sub>C</sub>14 and τ<sub>C</sub>16, we could show in the present example that the C-terminal 18 residues of τ<sub>C</sub>16 are highly mobile. This result is interesting, since these residues have been shown to be critical for the function of τ in binding to the α subunit of the *E. coli* DNA polymerase III (S. Jergic and N.E. Dixon, unpublished). The interaction with the α subunit may explain the observation that τ<sub>C</sub>16 is highly cytotoxic towards *E. coli* cells, whereas τ<sub>C</sub>14 exhibits no noticeable cytotoxicity (S. Jergic and N.E. Dixon, unpublished). In the cell-free protein production system, τ<sub>C</sub>16 was produced with similar yields to τ<sub>C</sub>14. The short measurement times of <sup>15</sup>N-HSQC spectra and the possibility to measure freshly synthesized samples were essential for the analysis of τ<sub>C</sub>16, for which samples prepared *in vivo* had proven to be very sensitive to proteolysis.

The main remaining problems were associated with a transaminase activity that converted glutamate to glutamine, the observation of a couple of non-protein cross-peaks and the 2-fold reduced yield for samples prepared in the absence of a high concentration of glutamate in the reaction mixture. The possible confusion of glutamate with methionine residues can be avoided by modification of the labelling scheme of Table 1, e.g. adding <sup>15</sup>N-Met to sample 1 instead of sample 5. We have not previously observed the appearance of non-protein cross-peaks that were not completely removed by dialysis (Ozawa et al., 2004).

Even in the face of possible uncertainties due to side reactions or spectral overlap, the combination of cell-free protein synthesis with combinatorial <sup>15</sup>N-labelling presents a remarkably efficient strategy: only one day was required for sample preparation of τ<sub>C</sub>14 and τ<sub>C</sub>16, and one day for recording and analysis of the <sup>15</sup>N-HSQC spectra. In addition, the samples produced can be used for further measurements, for example of <sup>3</sup>J(H<sup>N</sup>, H<sup>α</sup>) coupling constants.

In the absence of complete cross-peak overlap, classification of every <sup>15</sup>N-HSQC cross-peak by its residue type is equivalent to the spectral resolution achieved by 19 samples, where each is produced

with a different <sup>15</sup>N-labelled non-proline amino acid. We anticipate that our combinatorial labelling scheme will be most useful for the identification of chemical shift changes induced in <sup>15</sup>N-HSQC spectra upon ligand binding (Emerson et al., 2003; Zartler et al., 2003) and for the analysis of modified proteins as demonstrated here.

## Acknowledgements

We are grateful to Samir Hamdan for construction of the plasmid pSH1062. G.O. and K.O. thank the Australian Research Council (ARC) for a Federation Fellowship, and an Australian Linkage (CSIRO) Postdoctoral Fellowship, respectively. P.S.C.W. is the recipient of a Vice-Chancellor's Scholarship from the Australian National University, and S.J. was supported by an International Postgraduate Research Scholarship. Financial support by the ARC for this project and the 800 MHz NMR facility at ANU is gratefully acknowledged.

## References

- Dötsch, V. and Wagner, G. (1996) *J. Magn. Reson. B*, **111**, 310–313.
- Emerson, S.D., Palermo, R., Liu, C.M., Tilley, J.W., Chen, L., Danho, W., Madison, V.S., Greeley, D.N., Ju, G. and Fry, D.C. (2003) *Protein Sci.*, **12**, 811–822.
- Gao, D. and McHenry, C.S. (2001) *J. Biol. Chem.*, **276**, 4433–4440.
- Gehring, K. and Guittet, E. (1995) *J. Magn. Reson. B*, **109**, 206–208.
- Goddard, T.D. and Kneller, D.G. (2004) *SPARKY 3*, University of California, San Francisco.
- Guignard, L., Ozawa, K., Pursglove, S.E., Otting, G. and Dixon, N.E. (2002) *FEBS Lett.*, **524**, 159–162.
- Kigawa, T., Muto, Y. and Yokoyama, S. (1995) *J. Biomol. NMR*, **6**, 129–134.
- Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. and Yokoyama, S. (1999) *FEBS Lett.*, **442**, 15–19.
- Klammt, C., Löhr, F., Schäfer, B., Haase, W., Dötsch, V., Rüterjans, H., Glaubitz, C. and Bernhard, F. (2004) *Eur. J. Biochem.*, **271**, 568–580.
- Morita, E.H., Shimizu, M., Ogasawara, T., Endo, Y., Tanaka, R. and Kohno, T. (2004) *J. Biomol. NMR*, **30**, 37–45.
- Neylon, C., Brown, S.E., Kralicek, A.V., Miles, C.S., Love, C.A. and Dixon, N.E. (2000) *Biochemistry*, **39**, 11989–11999.
- Ozawa, K., Headlam, M.J., Schaeffer, P.M., Henderson, B.R., Dixon, N.E. and Otting, G. (2004) *Eur. J. Biochem.*, **271**, 4084–4093.
- Ozawa, K., Dixon, N.E. and Otting, G. (2005a) *IUBMB Life*, **57**, 615–622.

- Ozawa, K., Jergic, S., Crowther, J.A., Thompson, P.R., Wijffels, G., Otting, G. and Dixon, N.E. (2005b) *J. Biomol. NMR*, **32**, 235–241.
- Parker, M.J., Aulton-Jones, M., Hounslow, A.M. and Craven, C.J. (2004) *J. Am. Chem. Soc.*, **126**, 5020–5021.
- Piotto, M., Saudek, V. and Sklenář, V. (1992) *J. Biomol. NMR*, **2**, 661–665.
- Ponstingl, H. and Otting, G. (1998) *J. Biomol. NMR*, **12**, 319–324.
- Schmieder, P., Leidert, M., Kelly, M.J.S. and Oschkinat, H. (1998) *J. Magn. Reson.*, **131**, 199–201.
- Schubert, M., Smalla, M., Schmieder, P. and Oschkinat, H. (1999) *J. Magn. Reson.*, **141**, 34–43.
- Schubert, M., Oschkinat, H. and Schmieder, P. (2001a) *J. Magn. Reson.*, **148**, 61–72.
- Schubert, M., Oschkinat, H. and Schmieder, P. (2001b) *J. Magn. Reson.*, **153**, 186–192.
- Schubert, M., Oschkinat, H. and Schmieder, P. (2001c) *J. Biomol. NMR*, **20**, 379–384.
- Schubert, M., Labudde, D., Leitner, D., Oschkinat, H. and Schmieder, P. (2005) *J. Biomol. NMR*, **31**, 115–127.
- Shi, J.X., Pelton, J.G., Cho, H.S. and Wemmer, D.E. (2004) *J. Biomol. NMR*, **28**, 235–247.
- Shortle, D. (1994) *J. Magn. Reson. B*, **105**, 88–90.
- Tashiro, M., Rios, C.B. and Montelione, G.T. (1995) *J. Biomol. NMR*, **6**, 211–216.
- Trbovic, N., Klammt, C., Koglin, A., Löhr, F., Bernhard, F. and Dötsch, V. (2005) *J. Am. Chem. Soc.*, **127**, 13504–13505.
- Wu, P.S.C. and Otting, G. (2005) *J. Biomol. NMR*, **32**, 243–250.
- Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H. and Nagayama, K. (1991) *Biochemistry*, **30**, 6036–6047.
- Yamazaki, T., Pascal, S.M., Singer, A.U., Forman-Kay, J.D. and Kay, L.E. (1995) *J. Am. Chem. Soc.*, **117**, 3556–3564.
- Zartler, E.R., Hanson, J., Jones, B.E., Kiline, A.D., Martin, G., Mo, H., Shapiro, M.J., Wang, R., Wu, H. and Yan, J. (2003) *J. Am. Chem. Soc.*, **125**, 10941–10946.

## 9. Appendix: VI (VI)

# Paper VI

**P. S. C. Wu**, K. Ozawa, S. P. Lim, S. Vasudevan, N. E. Dixon, G. Otting  
“Cell-free transcription/translation from PCR amplified DNA for high-throughput NMR studies”

Reproduced with permission from  
*Angewandte Chemie International Edition* **46**, 3356-3358 (2007)  
Copyright 2007 Wiley-VCH Verlag GmbH & Co. KGaA

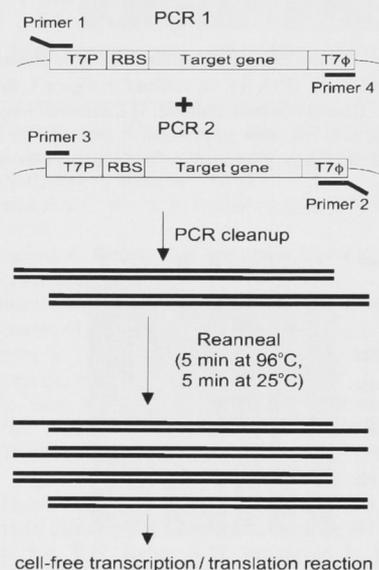
## Protein Production

## Cell-Free Transcription/Translation from PCR-Amplified DNA for High-Throughput NMR Studies\*\*

Peter S. C. Wu, Kiyoshi Ozawa, Siew P. Lim, Subhash G. Vasudevan, Nicholas E. Dixon, and Gottfried Otting\*

Cell-free coupled transcription/translation systems based on *Escherichia coli* extracts have become an increasingly attractive alternative to conventional expression systems *in vivo*.<sup>[1–3]</sup> In particular, they allow the fast synthesis of isotope-labeled proteins for direct analysis by NMR spectroscopy without chromatographic purification of the protein,<sup>[4]</sup> they suppress metabolic conversion between different amino acids,<sup>[5]</sup> and they use expensive isotope-labeled amino acids sparingly.<sup>[2]</sup> The best protein yields are obtained in a dialysis (continuous-exchange cell-free; CECF) system in which the outside buffer solution continuously replenishes the inside reaction mixture with amino acids and adenosine 5'-triphosphate (ATP).<sup>[6]</sup> In practice, the most time-consuming step is the large-scale preparation of plasmid DNA, as CECF reactions perform much better with plasmid DNA than with linear DNA templates. This is presumably due to degradation by exonucleases present in the *E. coli* extracts.<sup>[7]</sup> For example, reported yields of chloramphenicol acetyltransferase are up to 1 mg of protein per mL of reaction mixture when polymerase chain reaction (PCR)-amplified linear DNA is used,<sup>[8]</sup> whereas the use of plasmid DNA gives a yield of 6 mg mL<sup>-1</sup>.<sup>[6]</sup> Herein we present a technique that provides the same protein yields from PCR products as from plasmid DNA. This makes it possible to obtain protein NMR spectra within 24 h of gene amplification by PCR, including 7 h for cell-free protein synthesis and overnight dialysis against NMR buffer solution.

Our method is designed to generate complementary 8-base-pair, single-stranded overhangs at both ends of the PCR templates (Figure 1). These overhangs are suitable for



**Figure 1.** Generation of stable PCR templates from a construct with a T7  $\phi$ 10 promoter, a ribosome binding site, and a T7 terminator. The primer sequences were (single-stranded overhang regions underlined): primer 1: 5'-PO<sub>4</sub>-TTAGCTGGTTCGATCCCGCAAATTAATACG; primer 2: 5'-PO<sub>4</sub>-CCAGCTAACAAAAACCCCTCAAGACCCG; primer 3: 5'-PO<sub>4</sub>-TCGATCCCGCAAATTAATACG; primer 4: 5'-PO<sub>4</sub>-CAAAAAACCCCTCAAGACCCG. When starting from constructs without T7 elements, the T7 sequences can be introduced in an additional PCR step (see the Supporting Information).

ligation into exonuclease-insensitive cyclized templates by the endogenous ligase activity of *E. coli* S30 extracts. For all proteins tested in our laboratory, PCR products with 5'-phosphorylated overhangs produced the same protein yields in the CECF system as with plasmid DNA (Figure 2).

Evidence for cyclization was obtained indirectly. Figure 3 shows that the best yields were obtained with 1) longer overhangs and 2) 5'-phosphorylated overhangs; 3) single-stranded overhangs were essential; and 4) the precise nucleotide sequences of the overhangs were unimportant provided they were complementary.

The high protein yields obtained with the PCR protocol of Figure 1 open many important opportunities for high-throughput protein synthesis. For example, mutant proteins can readily be prepared and analyzed by NMR spectroscopy without cloning or chromatographic protein purification. Here we used this protocol to create site-directed mutants by PCR for site-specific resonance assignments of a dengue virus protease construct.

[\*] P. S. C. Wu, Dr. K. Ozawa, Prof. G. Otting

Research School of Chemistry  
The Australian National University  
Canberra ACT 0200 (Australia)

Fax: (+61) 2-6125-0750

E-mail: gottfried.otting@anu.edu.au

Homepage: <http://rsc.anu.edu.au/~go/>

Prof. N. E. Dixon

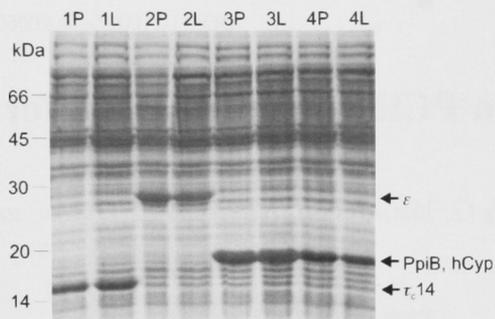
Department of Chemistry  
University of Wollongong, NSW 2522 (Australia)

Dr. S. P. Lim, Dr. S. G. Vasudevan

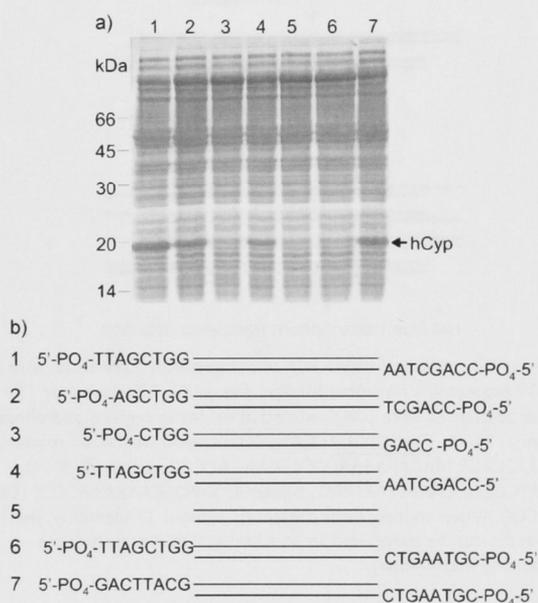
Novartis Institute for Tropical Diseases Pte Ltd  
10 Biopolis Road, Singapore 138670 (Singapore)

[\*\*] This work was supported by the Australian Research Council with a Federation Fellowship to G.O. and grant funding for the 800-MHz NMR spectrometer and this project. We thank Dr. Daying Wen and Dr. Xun-Cheng Su for a <sup>15</sup>N-HSQC spectrum of the uniformly labeled protease. PCR = polymerase chain reaction.

Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.



**Figure 2.** Comparison of protein yields obtained with plasmid DNA (P) or PCR-amplified linear DNA (L), as outlined in Figure 1, by using SDS-PAGE with Coomassie blue staining. 1) C-terminal fragment  $\tau_{c14}$  of the  $\tau$  subunit of *E. coli* DNA polymerase III (Pol III). 2)  $\epsilon$  Subunit of Pol III. 3) *E. coli* peptidylprolyl isomerase B (PpiB). 4) Human cyclophilin A (hCyp).

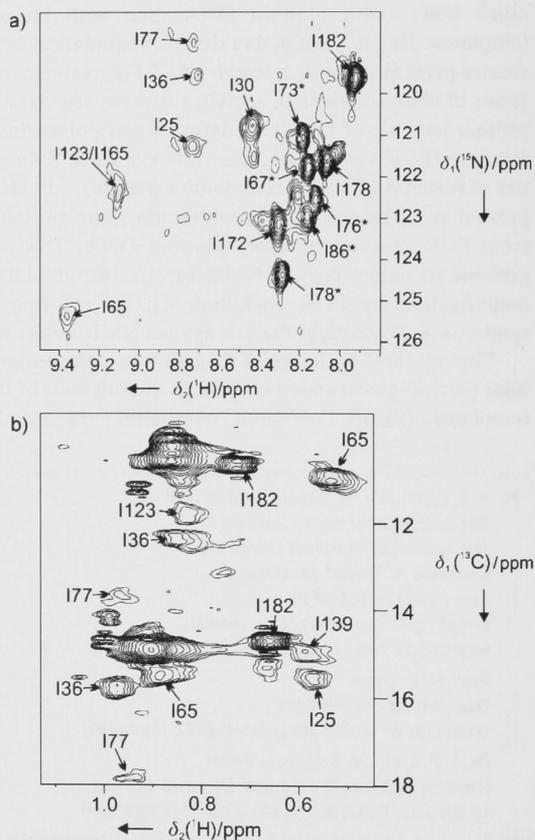


**Figure 3.** Expression yields achieved with different PCR templates. a) Coomassie blue-stained SDS-PAGE of hCyp synthesized by CEFV reactions by using PCR products ( $10 \mu\text{g mL}^{-1}$ ) with the overhangs shown in (b).

The dengue virus NS2B/NS3 protease complex is essential for dengue virus replication.<sup>[9]</sup>  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}/^{13}\text{C}$ -labeled protease yielded very broad signals (full-width at half maximum  $> 50$  Hz) owing to nonspecific aggregation or chemical exchange processes (see the Supporting Information).<sup>[10]</sup> In addition, many residues were highly mobile, resulting in poor spectral resolution, and the protein was prone to degradation. These properties prevent conventional resonance assignment strategies, yet, well-resolved cross-peaks were observed for samples that were selectively labeled with  $^{15}\text{N}/^{13}\text{C}$  isoleucine. We assigned these cross-peaks by systematic site-directed mutagenesis of each of the sixteen isoleucine residues in the amino acid sequence to valine. As

only isoleucine was isotope labeled, each peak missing from the HSQC spectrum of a mutant protein identified the assignment of the corresponding isoleucine residue in the wild-type protein.

The mutant constructs with the requisite terminal single-stranded overhangs were generated by using a modified overlap extension protocol (see the Supporting Information). The PCR products were used directly as templates for expression. Mutagenesis thus added little to the time required for sample preparation, and many mutations were easily made in parallel. The  $^{15}\text{N}$ -HSQC and  $^{13}\text{C}$ -HSQC spectra were recorded by using an experiment with simultaneous evolution of  $^{15}\text{N}$  and  $^{13}\text{C}$  resonances in the indirect frequency dimension.<sup>[11]</sup> In this way, 15 of the 16 isoleucine amide cross-peaks and 11 of the 22 isoleucine methyl cross-peaks were assigned (most of the isoleucine methyl peaks are unresolved, see Figure 4b). The same amide cross-peak was assigned to Ile 123 and Ile 165 of NS3 because mutation of either residue led to its disappearance (see Figure 4a and the Supporting Information). As both residues are close in the three-dimensional structure,<sup>[9]</sup> mutations at either site could affect the cross-peak, resulting in an ambiguous assignment.



**Figure 4.**  $^{15}\text{N}$ - and  $^{13}\text{C}$ -HSQC spectra of a  $200 \mu\text{M}$  solution of  $^{15}\text{N}/^{13}\text{C}$ -Ile-labeled dengue virus NS2B/NS3 protease at  $25^\circ\text{C}$ , pH 6.9. NS2B resonances are marked with a star. a) Amide region of the I139V mutant. b) Methyl region of the I78V mutant. (The HSQC spectra of these two mutants were indistinguishable from those of the wild-type protein in the amide and methyl region, respectively.)

The resonance assignments obtained for the NS2B/NS3 protease attribute the intense and narrow cross-peaks to those parts of the protein for which no electron density was observed in the crystal structure,<sup>[9]</sup> indicating that these segments are highly mobile in solution. Furthermore, these data show that all structured parts of the protein are affected by line broadening. Nonetheless, the assigned cross-peaks are now available for NMR-spectroscopy-based ligand screening.

In conclusion, PCR amplification of DNA templates for high-yield protein synthesis in cell-free systems coupled with NMR spectroscopic analysis opens many attractive applications in structural biology, including protein production from complementary DNA (cDNA), optimization of domain boundaries and protein solubility, production of mutants, and introduction of stop codons for site-specific incorporation of non-natural amino acids.<sup>[12]</sup> Without the need for large-scale plasmid preparation, cloning, or protein purification, these applications have become highly practical.

### Experimental Section

**PCR and cell-free protein expression:** PCR was performed by using Vent polymerase (New England Biolabs), which has 3'→5' proof-reading activity and produces mostly blunt-ended PCR products as required for proper overhang generation. Further details are provided in the Supporting Information. PCR products were purified by using the QiaQuick PCR purification kit (Qiagen). *E. coli* cell-free coupled transcription/translation reactions were carried out as described,<sup>[5]</sup> with the exception that the plasmid templates were replaced by PCR-generated DNA templates by using 10 µg DNA per mL of cell-free reaction mixture. <sup>15</sup>N/<sup>13</sup>C-labeled isoleucine was from Cambridge Isotope Laboratories.

**NMR spectroscopy:** Dengue virus NS2B/NS3 protease was prepared from the CF40.gly.NS3pro construct used for the crystal structure determination,<sup>[9]</sup> except that it was preceded by the sequence MASMTG. Following cell-free synthesis in a reaction mixture (1 mL; typical protein yield 1.4 mg), the samples were

dialyzed against NMR buffer solution (20 mM Tris-HCl, 50 mM NaCl; pH 6.9; Tris = tris(hydroxymethyl)aminomethane) and concentrated to 200 µL by using a Centricon-10 ultrafilter (Amicon). 10% D<sub>2</sub>O was added and NMR spectra recorded in 3-mm NMR tubes by using a Bruker Avance 800 MHz NMR spectrometer with a cryoprobe. Simultaneous <sup>15</sup>N/<sup>13</sup>C HSQC spectra<sup>[11]</sup> were recorded in 2.5 h per spectrum by using  $t_{1\max} = 32$  ms and  $t_{2\max} = 93$  ms.

Received: December 27, 2006

Published online: March 22, 2007

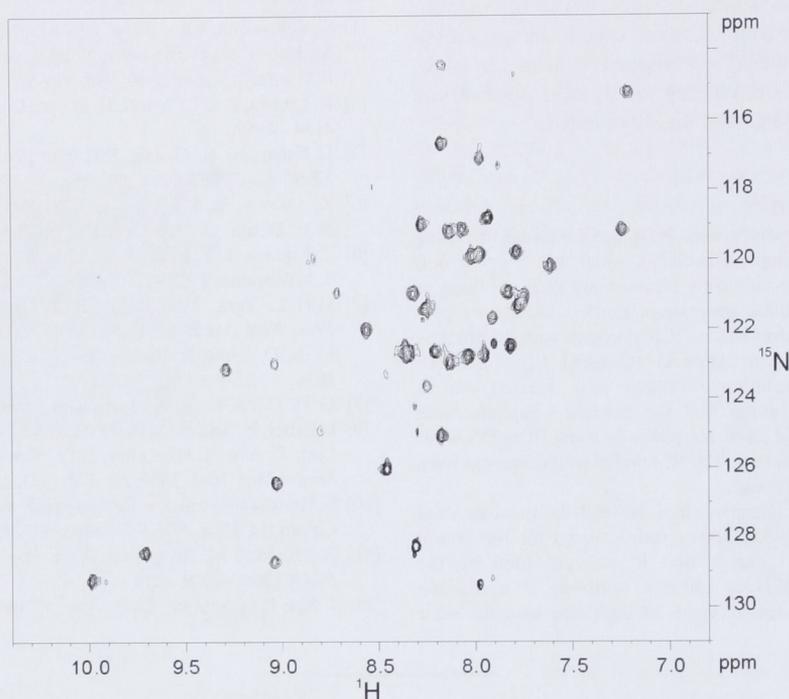
**Keywords:** cell-free protein synthesis · high-throughput methods · NMR spectroscopy · polymerase chain reaction

- [1] S. Yokoyama, *Curr. Opin. Chem. Biol.* **2003**, *7*, 39–43.
- [2] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. M. Ono, P. Güntert, *Nature* **2006**, *440*, 52–57.
- [3] K. Ozawa, P. S. C. Wu, N. E. Dixon, G. Otting, *FEBS J.* **2006**, *273*, 4154–4159.
- [4] L. Guignard, K. Ozawa, S. E. Pursglove, G. Otting, N. E. Dixon, *FEBS Lett.* **2002**, *524*, 159–162.
- [5] K. Ozawa, M. J. Headlam, P. M. Schaeffer, B. R. Henderson, N. E. Dixon, G. Otting, *Eur. J. Biochem.* **2004**, *271*, 4084–4093.
- [6] T. Kigawa, T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata, S. Yokoyama, *FEBS Lett.* **1999**, *442*, 15–19.
- [7] a) H. L. Yang, L. Ivashkiv, H. Z. Chen, G. Zubay, M. Cashel, *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 7029–7033; b) S. A. Lesley, M. A. D. Brow, R. R. Burgess, *J. Biol. Chem.* **1991**, *266*, 2632–2638.
- [8] G. H. Hahn, D. M. Kim, *Anal. Biochem.* **2006**, *355*, 151–153.
- [9] P. Erbel, N. Schiering, A. D'Arcy, M. Renatus, M. Kroemer, S. P. Lim, Z. Yin, T. H. Keller, S. G. Vasudevan, U. Hommel, *Nat. Struct. Mol. Biol.* **2006**, *13*, 372–373.
- [10] S. Melino, S. Fucito, A. Campagna, F. Wrubl, A. Gamarnik, D. O. Cicero, M. Paci, *FEBS J.* **2006**, *273*, 3650–3662.
- [11] R. Boelens, M. Burgering, R. H. Fogh, R. Kaptein, *J. Biomol. NMR* **1994**, *4*, 201–213.
- [12] J. Xie, P. G. Schultz, *Curr. Opin. Chem. Biol.* **2005**, *9*, 548–554.

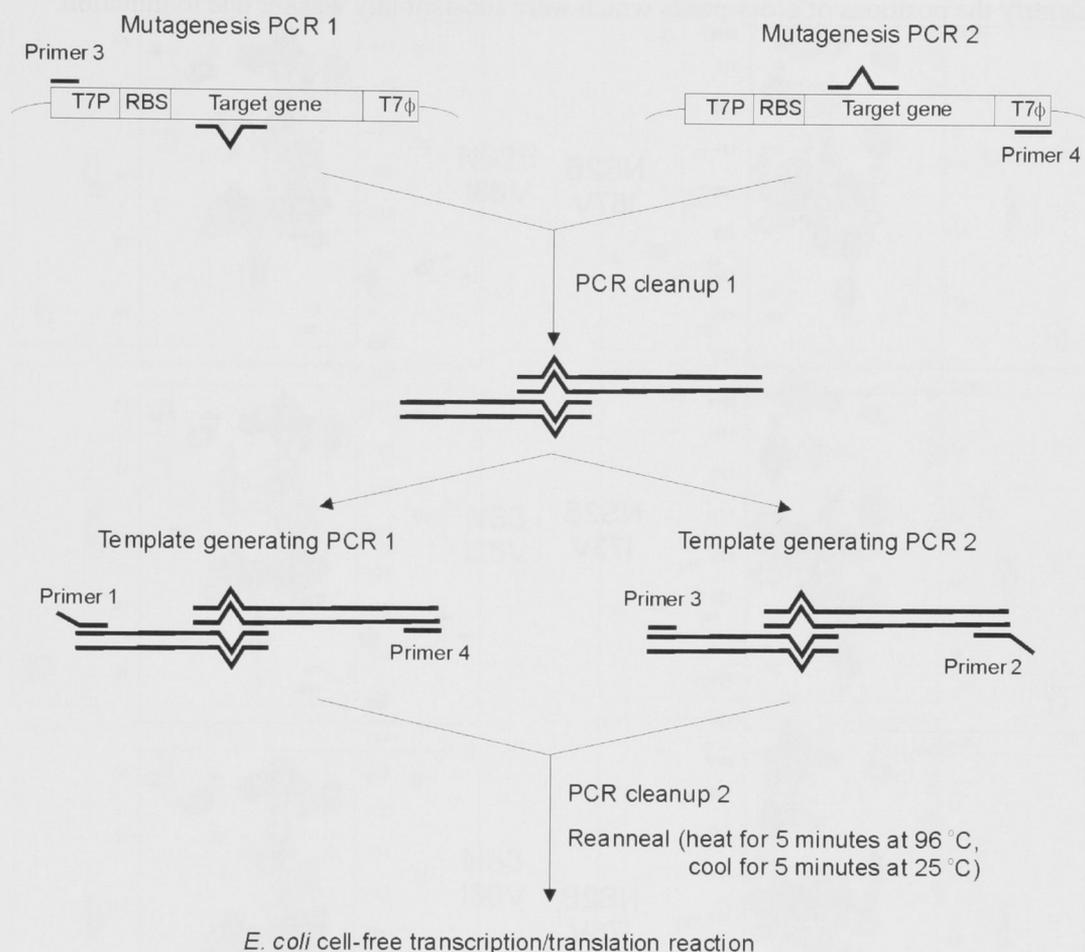
## Supporting Information

### Cell-free transcription/translation from PCR-amplified DNA for high-throughput NMR studies

Peter S. C. Wu, Kiyoshi Ozawa, Siew Pheng Lim, Subhash G. Vasudevan, Nicholas E. Dixon, and Gottfried Otting



**Figure S1.**  $^{15}\text{N}$ -HSQC spectrum of  $\tau_{\text{C14}}$  produced by cell-free protein synthesis with  $^{15}\text{N}$ -labeled Ala, Lys, Arg, Phe, Gln, Met, Cys and Trp. The DNA template used for transcription was generated in two PCR steps from the  $\tau_{\text{C14}}$  gene in a  $\lambda$  promoter vector (pSJ1308),<sup>[1]</sup> where the first step added the T7  $\phi$ 10 promoter and T7 terminator to the ends of the  $\tau_{\text{C14}}$  gene. Based on pET vector sequences,<sup>[2]</sup> the following primers were designed. Forward primer: 5'-TCGATCCCGCGAAATTAATACGACTCACTATAGG **GAGACCACAACGGTTCCCTCTAGAAATAATTTTGTTTAAAC** (bold characters identify the sequence complementary to pSJ1308 upstream of the ribosome binding site. Red characters identify the stem loop structure. The T7 promoter sequence is marked by underlining. Blue identifies a 15 bp upstream sequence for improved transcription.); reverse primer: 5'-CAAAAACCCCTCAAGACCCGTTTAGAGGCCCAAGGGGTT **ATGTTACTGAATATTATTATC** (Bold – sequence complementary to  $\tau_{\text{C14}}$ ; underlined – T7 terminator). The second PCR step was performed as illustrated in Figure 1. The spectrum shows the same cross-peaks as Figure S4, where the protein was prepared directly using a T7-promoter plasmid as template. The  $^{15}\text{N}$ -labeled amino acids in this spectrum (Trp, Arg, Cys, Phe, Ala, Lys, Gln, Met) and in the spectra of Figures 4, S3 and S4 were provided at the concentrations reported in ref. 3.



**Figure S2.** Modified overlap extension protocol for generation of site-directed mutant templates suitable for direct expression in *E. coli* cell-free transcription/translation systems. Primers 1-4 are those reported in Figure 2. Only a single additional PCR step is required to generate a site-directed mutation. NMR spectra of the expressed mutants can be recorded in the cell-free reaction mixture without protein purification. Following PCR amplification, purification of the template can be achieved by use of the QiaQuick PCR purification kit (Qiagen). Purification after the first step determines the background of wild-type DNA. Using the purification kit, 5-15% wild-type protein was detected by NMR spectroscopy, whereas purification by agarose gel electrophoresis reduced the background of wild-type DNA to less than 2% (Figure S4). The conditions of the PCR were chosen as per the manufacturer's recommendation (New England Biolabs). The cycling parameters of the mutagenesis PCR were: 94 °C (2 min), 25 x [94 °C (40 s), 54 °C (40 s), 74 °C (60 s)], 74 °C (2 min). The cycling parameters of the template generating PCR were: 94 °C (2 min), 30 x [94 °C (40 s), 54 °C (40 s), 74 °C (60 s)], 74 °C (2 min).

**Figure S3.**  $^{15}\text{N}$ -HSQC (left panel) and  $^{13}\text{C}$ -HSQC spectra (right panel) of each of the 16 Ile→Val mutants of the dengue virus NS2B/NS3 protease at pH 6.9 and 25 °C. Arrows identify the positions of cross-peaks which were substantially weaker due to mutation.<sup>a)</sup>

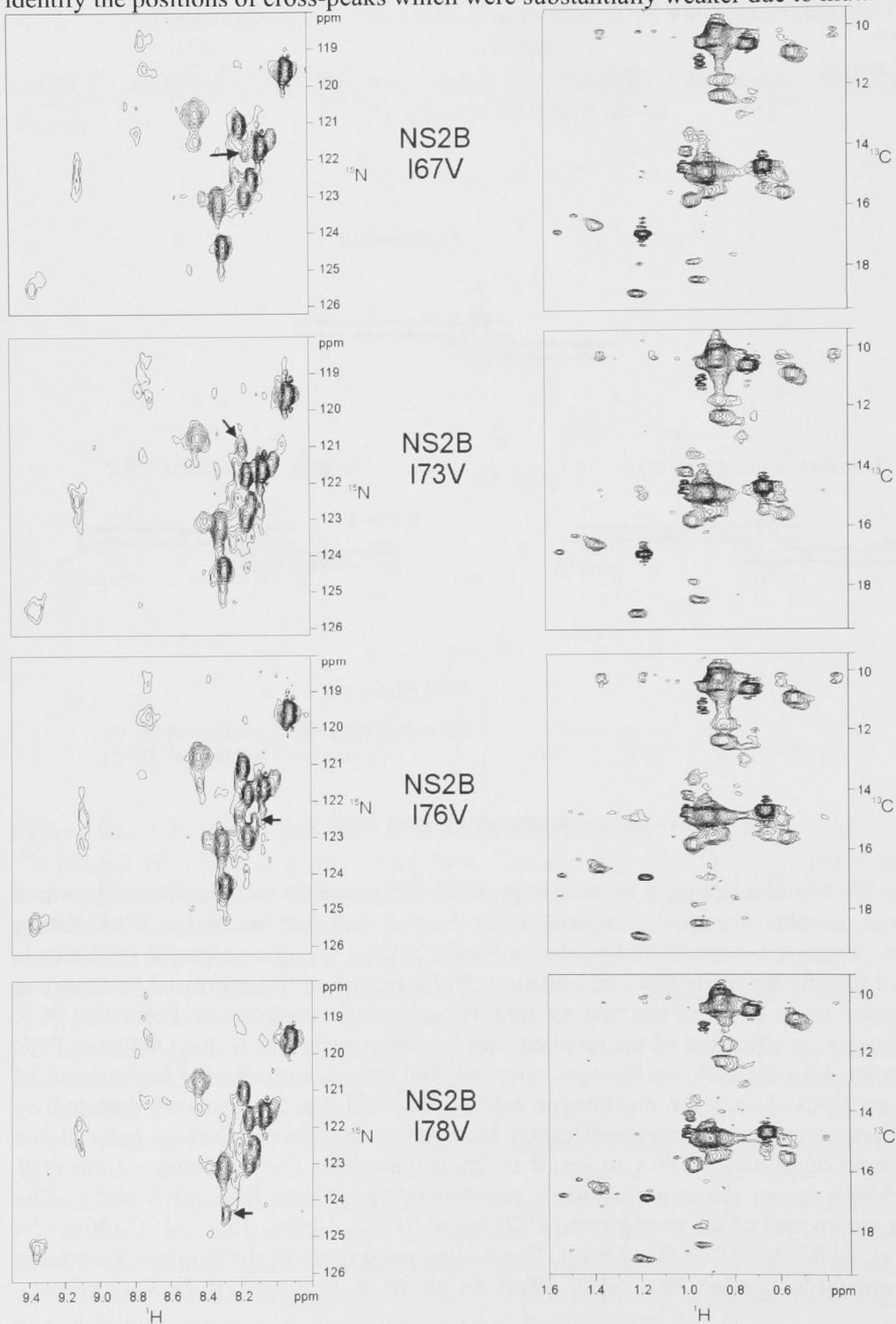


Figure S3, continued

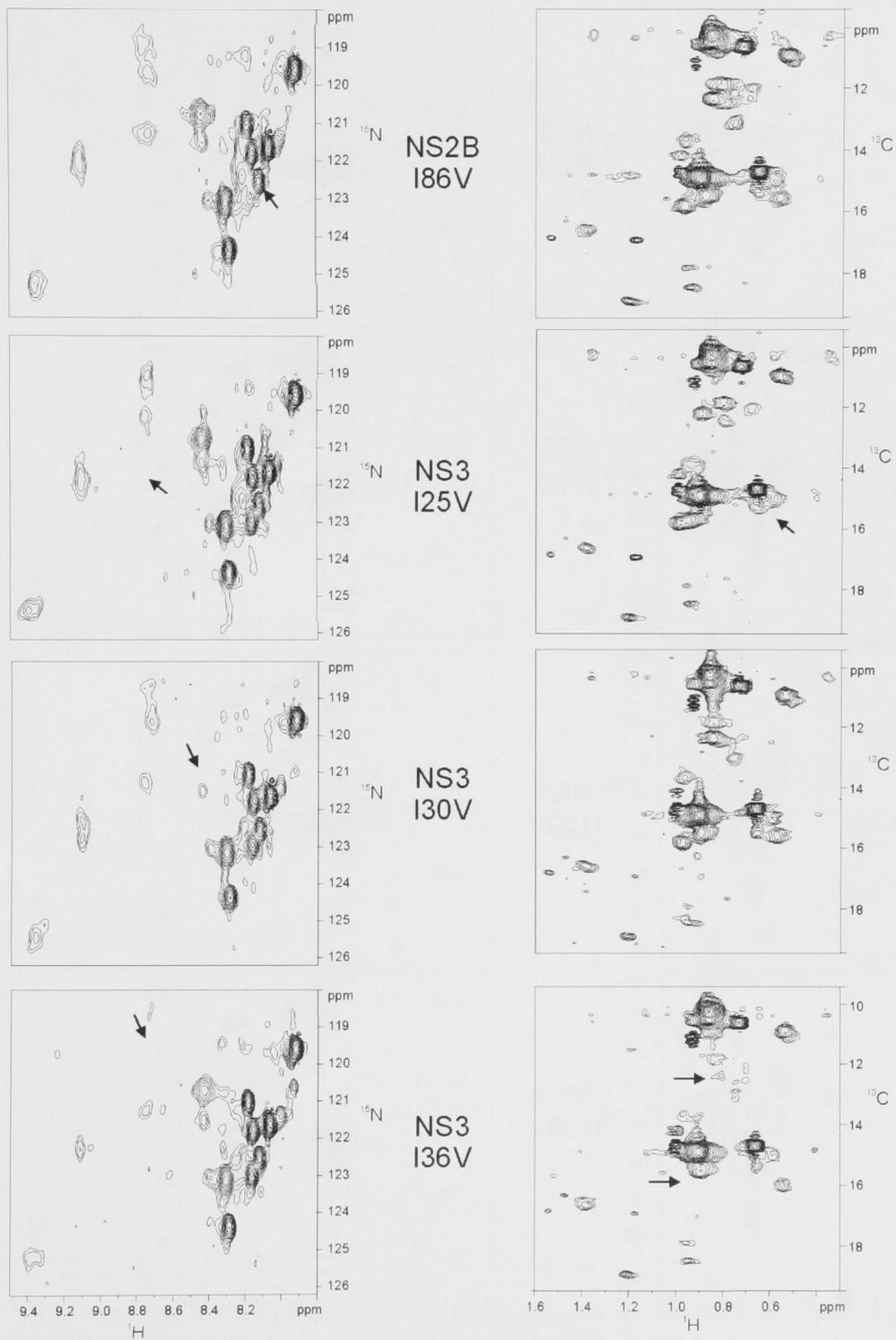


Figure S3, continued

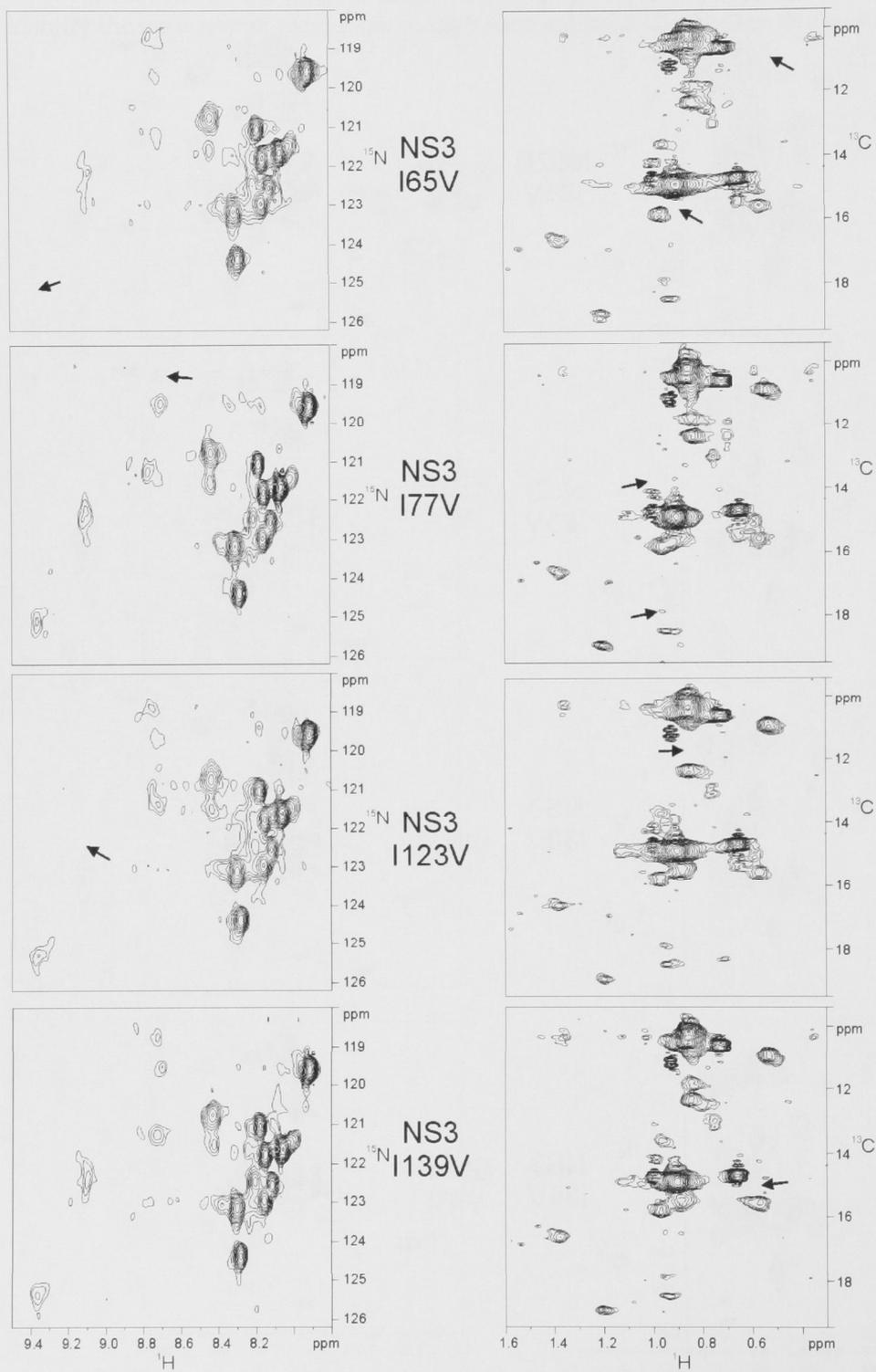
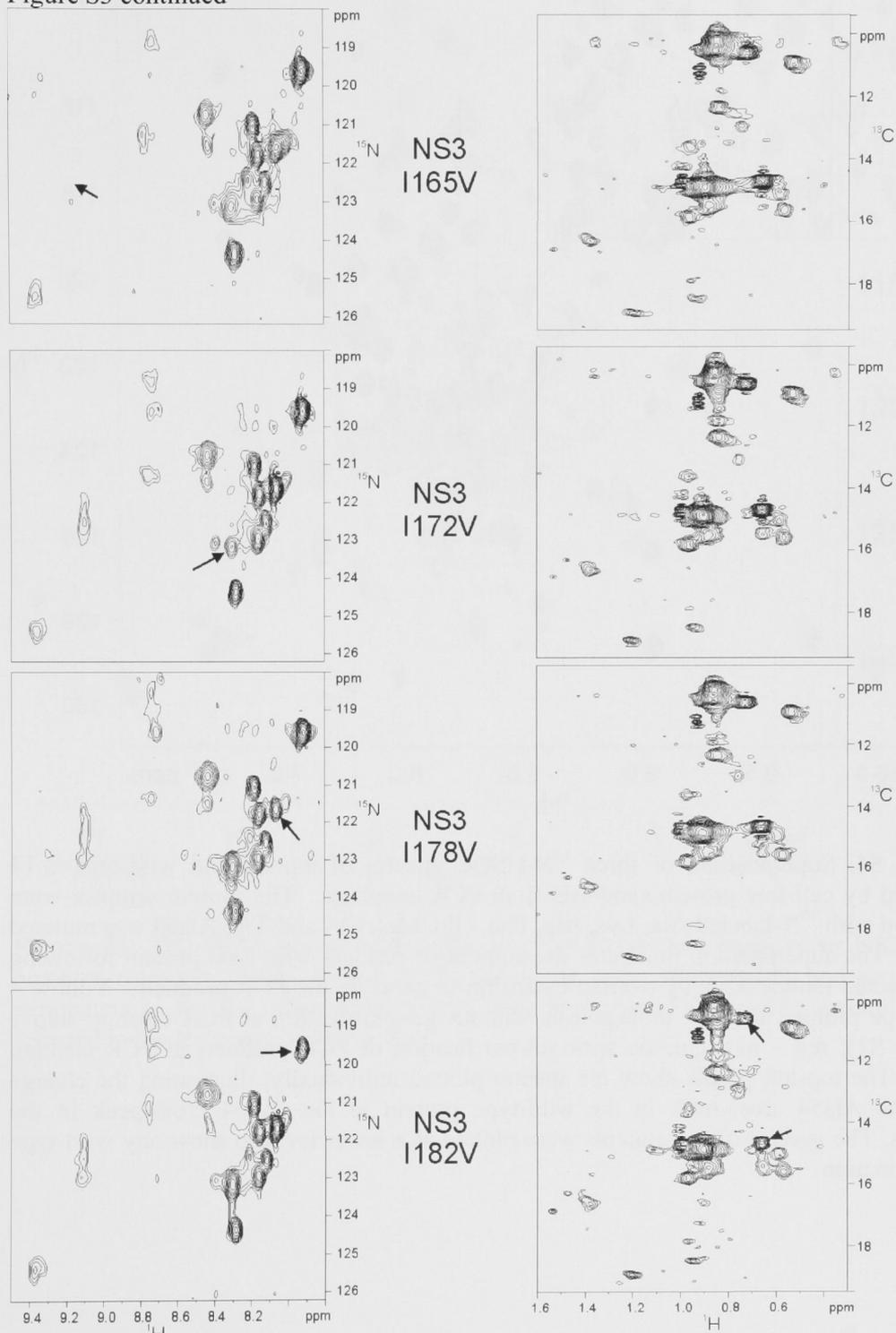
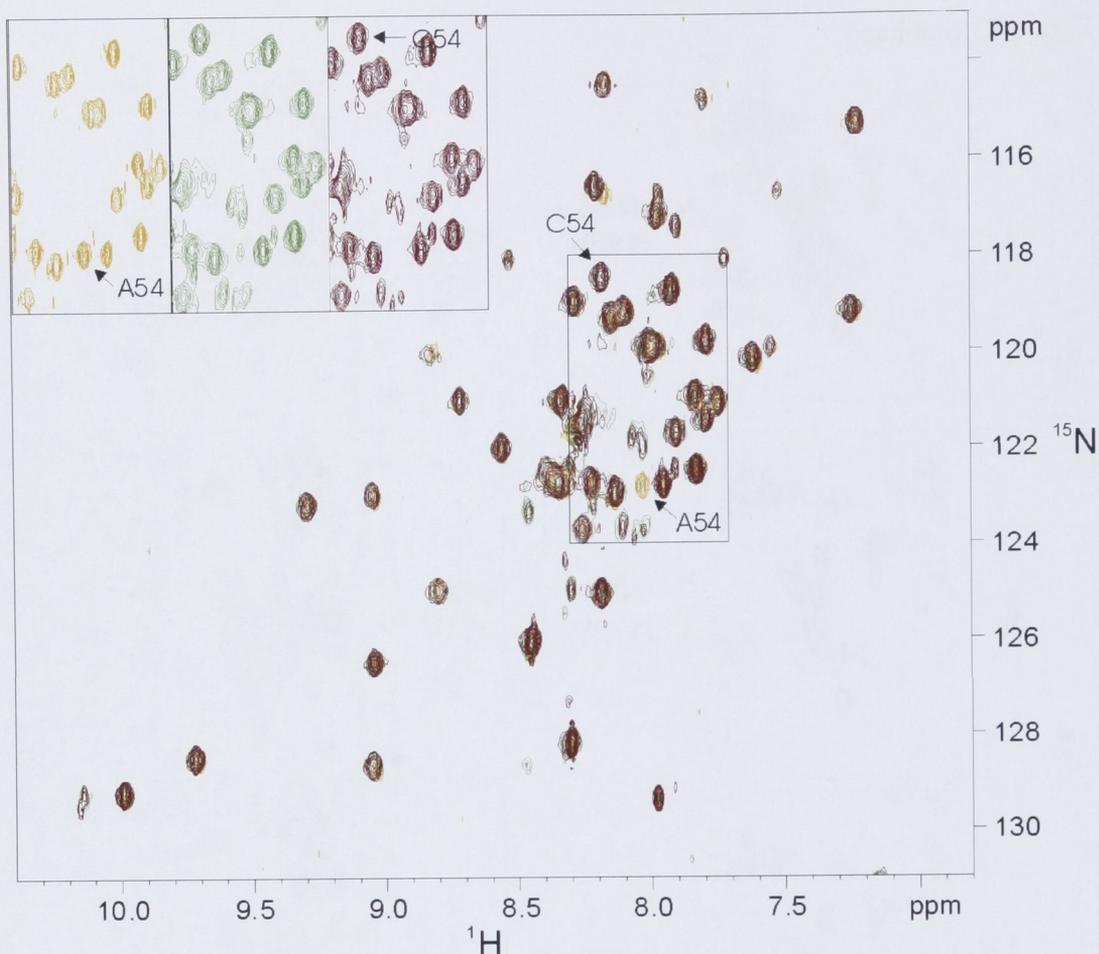


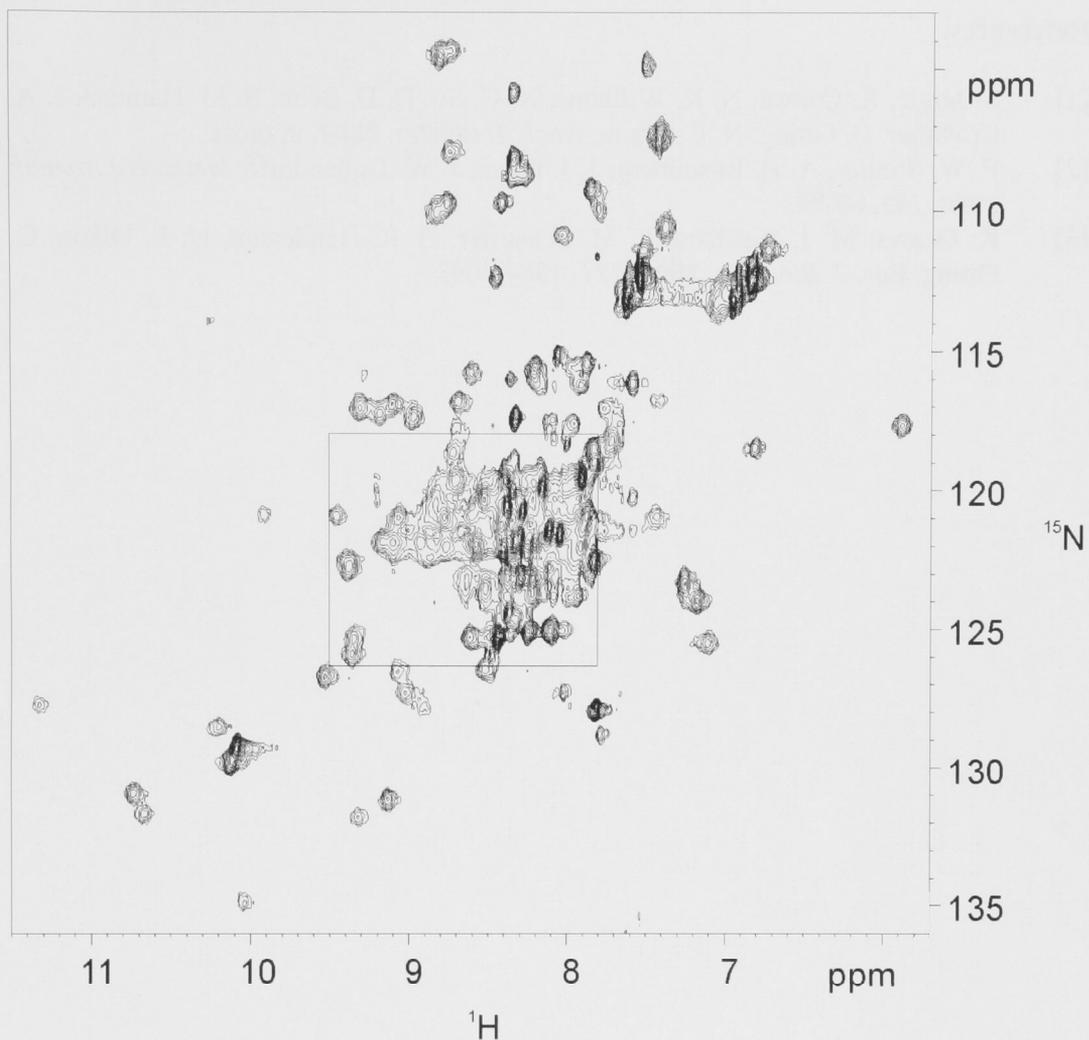
Figure S3 continued



<sup>a)</sup> The two most intense peaks in the  $^{13}\text{C}$ -HSQC spectra contain the unresolved cross-peaks of  $\text{C}^{\gamma}\text{H}_3$  and  $\text{C}^{\delta}\text{H}_3$  resonances of highly mobile Ile residues.



**Figure S4.** Superposition of three  $^{15}\text{N}$ -HSQC spectra of mutant and wild-type  $\tau_{\text{C14}}$  prepared by cell-free protein synthesis from PCR templates. The protein samples were prepared with  $^{15}\text{N}$ -labeled Ala, Lys, Arg, Phe, Gln, Met, Cys and Trp. Ala54 was mutated to Cys. The superposition illustrates the amount of residual wild-type protein following site-directed mutagenesis by overlap extension to generate the PCR products. Yellow – wild type protein; green – mutagenesis without gel-purification at PCR cleanup step 1 (Figure S1); red – mutagenesis with gel-purification of PCR products at PCR cleanup step 1. The top-left panels show the spectra plotted individually, illustrating the change from the Ala54 cross-peak in the wild-type protein to the Cys54 cross-peak in the mutants. The spectra of the mutants were plotted at a lower level to show any wild-type contamination.



**Figure S5.**  $^{15}\text{N}$ -HSQC spectrum of uniformly  $^{15}\text{N}/^{13}\text{C}$  labeled dengue virus NS2B/NS3 protease recorded at pH 7.1 and 25 °C, using the same spectrometer and acquisition parameters as for the other NMR spectra. The rectangle identifies the spectral region selected in the left panel of Figure S3.

## References:

- [1] S. Jergic, K. Ozawa, N. K. Williams, X.-C. Su, D. D. Scott, S. M. Hamdan, J. A. Crowther, G. Otting, N. E. Dixon, *Nucl. Acids Res.* **2007**, in press
- [2] F. W. Studier, A. H. Rosenberg, J. J. Dunn, J. W. Dubendorff, *Methods Enzymol.* **1990**, *185*, 60-89.
- [3] K. Ozawa, M. J. Headlam, P. M. Schaeffer, B. R. Henderson, N. E. Dixon, G. Otting, *Eur. J. Biochem.* **2004**, *271*, 4084-4093.