J Biomol NMR (2009) 44:261–272
DOI 10.1007/s10858-009-9339-6

ARTICLE

# Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system

**Teppei Ikeya · Mitsuhiro Takeda · Hitoshi Yoshida · Tsutomu Terauchi · Jun-Goo Jee · Masatsune Kainosho · Peter Güntert**

Received: 30 March 2009 / Accepted: 24 June 2009 / Published online: 14 July 2009

**Abstract** Stereo-array isotope labeling (SAIL) has been combined with the fully automated NMR structure determination algorithm FLYA to determine the three-dimensional structure of the protein ubiquitin from different sets of input NMR spectra. SAIL provides a complete stereo- and regio-specific pattern of stable isotopes that results in sharper resonance lines and reduced signal overlap, without information loss. Here we show that as a result of the superior quality of the SAIL NMR spectra, reliable, fully automated analyses of the NMR spectra and structure calculations are possible using fewer input spectra than with conventional uniformly $^{13}C/^{15}N$-labeled proteins. FLYA calculations with SAIL ubiquitin, using a single three-dimensional "through-bond" spectrum (and 2D HSQC spectra) in addition to the $^{13}C$-edited and $^{15}N$-edited NOESY spectra for conformational restraints, yielded structures with an accuracy of 0.83–1.15 Å for the backbone RMSD to the conventionally determined solution structure of SAIL ubiquitin. NMR structures can thus be determined almost exclusively from the NOESY spectra that yield the conformational restraints, without the need to record many spectra only for determining intermediate, auxiliary data of the chemical shift assignments. The FLYA calculations for this report resulted in 252 ubiquitin structure bundles, obtained with different input data but identical structure calculation and refinement methods. These structures cover the entire range from highly accurate structures to seriously, but not trivially, wrong structures, and thus constitute a valuable database for the substantiation of structure validation methods.

**Keywords** Automated structure determination · CYANA · FLYA · SAIL · Structure validation

## Introduction

Strategies for protein structure analysis by NMR have recently seen renewed interest (Billeter et al. 2008; Williamson and Craven 2009). The complete automation of protein structure determination is one of the challenges of biomolecular NMR spectroscopy that has, despite early optimism (Pfändler et al. 1985), proved difficult to achieve. The unavoidable imperfections of experimental NMR spectra, and the intrinsic ambiguity of peak assignments that results from the limited accuracy of frequency measurements, turn the tractable problem of finding the chemical shift assignments from ideal spectra into a formidably difficult one under realistic conditions. A variety of automated algorithms tackling different parts of NMR protein structure analysis have been developed and reviewed (Baran et al. 2004; Gronwald and Kalbitzer 2004;

T. Ikeya · P. Güntert (✉)
Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
e-mail: guentert@em.uni-frankfurt.de

T. Ikeya · P. Güntert
Frankfurt Institute for Advanced Studies, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

T. Ikeya · M. Takeda · H. Yoshida · T. Terauchi · J.-G. Jee · M. Kainosho · P. Güntert
Graduate School of Science, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan

M. Takeda · M. Kainosho (✉)
Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
e-mail: kainosho@nmr.chem.metro-u.ac.jp

Güntert 2009; Williamson and Craven 2009). Recently a purely computational algorithm (FLYA) was published that is capable of determining three-dimensional (3D) protein structures on the basis of uninterpreted spectra without manual interventions (López-Méndez and Güntert 2006).

Automated protein structure determination by NMR benefits from any method that improves or simplifies the spectra, notably by stable isotope labeling, spectrometer hardware with higher signal-to-noise and frequency resolution, and optimized heteronuclear multidimensional experiments. A powerful approach is stereo-array isotope labeling (SAIL), which simultaneously achieves a four to sevenfold increase in signal-to-noise, sharper resonance lines, and a 40–60% reduction in the number of signals, without sacrificing essential information about the backbones and the side chains of all amino acid residue types (Kainosho et al. 2006). The resulting reduction of spectral overlap, the higher accuracy of the frequency determination, the complete stereospecific assignment, and the measurement of longer $^1$H–$^1$H distances made it possible to determine the high-quality solution structures of proteins larger than 30 kDa (Kainosho et al. 2006; Takeda et al. 2008). The SAIL technique uses amino acids with a complete stereospecific and regiospecific pattern of stable isotopes that is optimized with regard to the quality and information content of the resulting NMR spectra. The 20 protein-constituting amino acids are prepared by chemical and enzymatic syntheses such that in all methylene groups, one $^1$H is stereo-selectively replaced by $^2$H, in all single methyl groups, two $^1$H are replaced by $^2$H, and in the prochiral methyl groups of Leu and Val, one methyl is stereo-selectively $–{}^{12}C(^2H)_3$ and the other is $–{}^{13}C^1H(^2H)_2$. In six-membered aromatic rings, the $^{12}$C–$^2$H and $^{13}$C–$^1$H moieties alternate with each other (Kainosho et al. 2006; Torizawa et al. 2005). SAIL amino acids for the production of protein NMR samples are commercially available from SAIL Technologies, Inc. (www.sail-technologies.com), and an efficient cell-free protein expression system has been established that is suitable for the large-scale synthesis of SAIL proteins without scrambling of the isotope labels (Torizawa et al. 2004).

Guided by the ongoing assignment process, an experienced spectroscopist can often identify crucial peaks with virtual certainty and, if necessary, make an assignment on the basis of a single, uniquely identified peak. Fully automated methods for the resonance assignment of proteins, on the other hand, generally have a lower reliability of peak identification than a spectroscopist who visually inspects the spectra. A sufficient level of redundancy, e.g., the availability of multiple peaks for a given atom, is therefore required for the successful operation of fully automated methods. This can be achieved by recording a set of spectra that provide complementary information for the assignment of a given atom or group of atoms, such that the algorithm can determine their resonance assignments from a variety of data, without relying on the certain identification of any specific peak (Bartels et al. 1997). The 3D structures of three uniformly labeled 12–16 kDa proteins were determined with the fully automated FLYA algorithm, using 13–14 3D NMR spectra for each protein (López-Méndez and Güntert 2006). However, a considerable amount of measurement time, about 3 weeks, was required for each protein to record these spectra. We therefore investigated whether the FLYA algorithm can also be used with smaller sets of spectra (Scott et al. 2006). For the prototypical Fes SH2 domain protein, correct structures could be obtained from as few as five 3D spectra. A further reduction of the input data to three 3D spectra resulted in distorted structures with about 3 Å RMSD to the reference structure.

The combination of SAIL and FLYA (Takeda et al. 2007) is expected to facilitate the automated process, especially for larger, less soluble, or otherwise difficult proteins. In this report, we applied the SAIL-FLYA method to various sets of spectra recorded with a low-concentration sample of the SAIL protein ubiquitin. The purpose of our study is not the determination of a new protein structure or the application of SAIL to large proteins but to give a proof of principle that SAIL enables the fully automated structure determination of proteins using much smaller sets of input spectra than are necessary with conventional uniformly labeled proteins. Further, we characterized more than 250 SAIL-FLYA ubiquitin structures with different accuracies by common validation parameters, and showed that an overall validation $Z$ score can distinguish between correct and incorrect structures with high fidelity.

## Materials and methods

### Sample preparation

The SAIL ubiquitin protein sample was produced in an *E. coli* cell-free synthesis system optimized for the preparation of labeled NMR samples (Takeda et al. 2007; Torizawa et al. 2004), using 50 mg of the SAIL amino acid mixture (SAIL Technologies). The lysate was cleared by centrifugation at 27,000*g* for 20 min. After boiling at 350 K, SAIL ubiquitin was purified by ion exchange chromatography on a DE52 column with 50 mM sodium acetate buffer (pH 6.0) and on a MonoS 5/5 column with 50 mM sodium acetate buffer (pH 4.8) and 0.5 M NaCl, followed by gel filtration chromatography on a Superdex 75 column with 50 mM sodium phosphate buffer (pH 6.8) and 200 mM NaCl. The protein was concentrated to 0.1 or 0.4 mM and was dissolved in 90% $^1H_2O$, 10% $^2H_2O$, 10 mM sodium phosphate buffer (pH 6.6) and 0.01% $NaN_3$ for the NMR measurements.

## NMR spectroscopy

NMR data were measured at 310 K on a Bruker DRX 600 spectrometer equipped with a cryogenic probe (Table 1) and were processed with the programs NMRPipe (Delaglio et al. 1995) and Azara (www.bio.cam.ac.uk/azara). Spectra were recorded at a protein concentration of 0.1 mM except for the H(CCCO)NH, HCCH-TOCSY and NOESY spectra, for which a concentration of 0.4 mM was used. The mixing time for the NOESY spectra was 100 ms. For comparison with the automated assignment results, the spectra were also manually analyzed using the program ANSIG 3.3 (Kraulis 1989; Kraulis et al. 1994).

## FLYA calculations

The FLYA algorithm (López-Méndez and Güntert 2006) used as input data only the protein sequence and the multidimensional NMR spectra. Peaks were identified in the multidimensional NMR spectra using the automated peak picking algorithm of NMRView (Johnson 2004; Johnson and Blevins 1994), and peak lists were prepared by CYANA (Güntert 2003; Güntert et al. 1997). Depending on the spectra, the preparation included unfolding aliased signals, systematic correction of chemical shift referencing, and removal of peaks near the diagonal or water line. The peak lists resulting from this step remained invariable throughout the rest of the procedure. An ensemble of initial chemical shift assignments was obtained by multiple runs of a modified version of the GARANT algorithm (Bartels et al. 1996, 1997) with different seed values for the random number generator (Malmodin et al. 2003). The peak position tolerance was set to 0.03 ppm for the $^{1}H$ dimension and to 0.3 ppm for the $^{13}C$ and $^{15}N$ dimensions. These initial chemical shift assignments were consolidated by CYANA into a single consensus chemical shift list. Torsion angle restraints were produced by the program TALOS (Cornilescu et al. 1999), on the basis of the consensus chemical shifts. Hydrogen bond restraints were not applied. The consensus chemical shift list, the amino acid sequence, and the unassigned NOE peak lists were used as input data for combined automated NOE assignment (Herrmann et al. 2002) and structure calculation by torsion

**Table 1** Acquisition parameters of the multidimensional NMR spectra recorded for SAIL ubiquitin and sets of spectra used for FLYA calculations

| Spectrum | Points[a] | Width[b] (kHz) | Peaks | | FLYA run | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Expected[c] | Observed | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 2D spectra | | | | | | | | | | | | | | | | | | |
| $[^{15}N,^{1}H]$–HSQC | 30 | 7.1, 2.1 | 90 | 80 | y | y | y | y | y | y | y | y | y | y | y | y | y | y |
| $[^{13}C,^{1}H]$–HSQC | 128 | 7.1, 10.6 | 243 | 249 | y | y | y | y | y | y | y | y | y | y | y | y | y | y |
| HB(CBCG)HE | 10 | 7.1, 2.1 | 5 | 5 | y | y | y | y | y | y | y | y | y | y | y | y | y | y |
| 3D spectra for backbone assignment | | | | | | | | | | | | | | | | | | |
| HNCO | 36, 100 | 7.5, 1.0, 3.0 | 73 | 91 | y | | | | | | | | | | | y | | |
| HN(CA)CO | 36, 100 | 7.5, 1.0, 3.0 | 146 | 110 | y | | | | | | | | | | y | | | |
| CBCANH | 36, 120 | 7.5, 1.0, 10.6 | 281 | 262 | y | y | y | y | | y | | | | | | | | |
| CBCA(CO)NH | 36, 120 | 7.5, 1.0, 10.6 | 141 | 133 | y | y | y | y | y | | y | | | | | | | |
| 3D spectra for side chain assignment | | | | | | | | | | | | | | | | | | |
| HBHA(CO)NH | 31, 18 | 7.5, 1.0, 3.6 | 141 | 256 | y | y | y | y | | | | y | | | | | | |
| (H)CC(CO)NH | 60, 14 | 7.5, 1.0, 1.0 | 232 | 237 | y | y | y | y | y | | | | y | | | | | |
| H(CCCO)NH | 40, 14 | 7.5, 1.0, 4.2 | 232 | 353 | y | y | y | y | y | | | | | y | | | | |
| HCCH-COSY | 60, 60 | 7.5, 10.5, 3.6 | 575 | 361 | y | y | | | | | | | | | | | y | |
| HCCH-TOCSY | 60, 60 | 7.5, 10.5, 3.6 | 881 | 1,017 | y | y | y | | | | | | | | | | | y |
| 3D spectra for conformational restraints | | | | | | | | | | | | | | | | | | |
| $^{15}N$-edited NOESY | 218, 36 | 7.5, 1.0, 6.3 | 2,051 | 1,979 | y | y | y | y | y | y | y | y | y | y | y | y | y | y |
| $^{13}C$-edited NOESY | 218, 50 | 7.5, 3.5, 6.3 | 5,575 | 4,553 | y | y | y | y | y | y | y | y | y | y | y | y | y | y |

[a] Points: complex time domain data points in the indirect dimensions. The number for the first indirect dimension refers to $^{15}N$, if present, or $^{13}C$ otherwise. The second number refers to $^{1}H$, if present, or $^{13}C$ otherwise. In all 3D spectra, 512 complex time domain data points were recorded in the directly detected $^{1}H$ dimension

[b] Spectral widths in the directly detected dimension, and in the indirectly detected dimension(s)

[c] Number of cross peaks expected under ideal conditions, based on the knowledge of the magnetization transfer pathways for each experiment. In the case of NOESY spectra, the expected peaks correspond to $^{1}H$–$^{1}H$ distances shorter than 4.5 Å in the reference structure

angle dynamics (Güntert et al. 1997). Upper distance limits were derived from the NOESY peak intensities, according to an inverse sixth power law for the volume–distance relationship, and were confined to the range 2.4–5.2 Å. Structure calculations were started from 500 conformers with random torsion angles, a number five times greater than that of the default, to minimize the possible influence of erroneous chemical shift assignments, particularly in reduced data sets. Seven cycles of combined automated NOE assignment and structure calculation by simulated annealing in torsion angle space and a final structure calculation using only unambiguously assigned distance restraints were run. The complete calculation comprised three stages. In stage I, the chemical shifts and the protein structures were generated de novo. In the following stages II and III, the structures generated in the preceding stage were used as additional input for the determination of chemical shift assignments. The 20 final CYANA conformers with the lowest target function values were subjected to restrained energy minimization in explicit solvent against the AMBER force field (Cornell et al. 1995; Ponder and Case 2003), using the program OPALp (Koradi et al. 2000; Luginbühl et al. 1996). The entire procedure was driven by the NMR structure calculation program CYANA, which was also used for parallelization of all of the time-consuming steps. Calculations were performed simultaneously on 20 processors of a Linux cluster system with Intel quad-core 2.4 GHz processors. For each set of NMR spectra, three runs were conducted using different seed values for the random number generators. The results of these three runs were averaged and are presented in the "Results and discussion" section.

The original FLYA algorithm (López-Méndez and Güntert 2006) was adapted for work with SAIL proteins (Takeda et al. 2007), notably within the GARANT program (Bartels et al. 1996, 1997). Parameters were optimized, and the contribution $P_R$ of a NOESY peak to the GARANT scoring function in stages II and III, when a 3D structure was available, was set to

$$P_R = w_{\text{total}}(P_d + P_s + P_r + P_\omega + P_\delta),$$

with the contribution $P_d$ for the agreement with the input structure (see below), and the standard GARANT terms $P_s$ for the given spectrum type, $P_r$ for the density of peaks in a spectrum, $P_\omega$ for the agreement with the general chemical shift statistics, and $P_\delta$ for the deviation of the chemical shifts among the peaks involving the same atom(s) in all spectra. The contribution for the agreement with the input structure of the upper distance limit $u$ associated with a NOESY peak is given by

$$P_d = w_{\text{NOE}} \exp\left(-\frac{1}{2}\left(\frac{\max(d_{\min} - u, 0)}{\sigma}\right)^2\right),$$

where $w_{\text{NOE}}$ is a weighting factor, $d_{\min}$ is the minimal distance within the conformers of a structure bundle, and the parameter $\sigma = 0.5$ Å indicates the size of a "significant" violation.

### Structure analysis and structure validation

The program MOLMOL (Koradi et al. 1996) was used to visualize 3D structures. CYANA was used to obtain statistics on target function values, restraint violations, etc., and to compute RMSD values to the mean coordinates of a structure bundle for superpositions of the backbone atoms N, C and C′ or the heavy atoms for the structured region of the protein, residues 1–72. The single RMSD value between the two sets of mean coordinates was used to quantify the deviation of one structure bundle from another. Conformational energies were calculated with OPALp (Koradi et al. 2000; Luginbühl et al. 1996) using the AMBER (Cornell et al. 1995; Ponder and Case 2003) force field.

The following validation parameters were computed for all structure bundles: (1) The logarithm of the backbone RMSD to the mean coordinates. (2) The AMBER potential energy (Cornell et al. 1995; Ponder and Case 2003). (3) The percentage of residues in the most favored region of the Ramachandran plot, defined by the program Procheck (Laskowski et al. 1996; Morris et al. 1992). (4) The Verify3D score (Bowie et al. 1991; Lüthy et al. 1992). (5) The packing, the Ramachandran plot appearance, the $\chi^1/\chi^2$ rotamer normality and the backbone conformation quality scores of the Whatcheck program (Hooft et al. 1996). (6) The LGscore and Maxsub score of the ProQ program (Wallner and Elofsson 2003). (7) The score of the ProSa 2003 program (Sippl 1993). In addition, we calculated an overall $Z$ score from the principal scores $S_1, \ldots, S_7$ of the aforementioned validation programs, defined by

$$Z = \sum_{i=1}^{7} \frac{S_i - \bar{S}_i}{\sigma(S_i)}.$$

The sign of each of the scores $S_i$ was chosen such that a better structure has a lower score. The packing score was selected as the principle score of the Whatcheck program, and the LGscore as the principle score of ProQ.

## Results and discussion

### Spectra sets and peak picking

The NMR spectra collected for SAIL ubiquitin are listed in Table 1, along with the experiments included in the 14 different subsets of these spectra that were used as input for the SAIL-FLYA structure calculations with "full" and reduced data sets. The two 2D HSQC spectra (Fig. 1) and the 2D HB(CBCG)HE spectrum for the assignment of aromatic resonances in SAIL proteins (Torizawa et al. 2005), which can be measured quickly, and the two 3D NOESY spectra that are required for obtaining the conformational restraints for the structural calculation, were used in all FLYA runs as a basic spectra set. For Run 1, the full set of three 2D and thirteen 3D spectra was employed. Runs 2–5 were done with progressively reduced sets of spectra for the backbone and side chain chemical shift assignments. Runs 6–14 were performed with "minimal" data sets that included only the basic spectra set and a single additional 3D spectrum. Automatic peak picking was always performed over the complete spectrum, excluding only two narrow bands along the water line and along the diagonal. No other spectral regions or individual peaks were interactively excluded from peak picking. Automated peak picking yielded between 75 and 125% of the expected number of peaks, except for HBHA(CO)NH (182%), H(CCCO)NH (152%), and HCCH-COSY (62%) (Table 1).

## Chemical shift assignments

As a reference we manually made chemical shift assignments that were complete for the $^{1}$H backbone amide and aliphatic protons and their directly bound $^{13}$C and $^{15}$N nuclei, except for the amide groups of Met 1 and Gly 53, and the side-chain methyl group of Met 1.

Table 2 summarizes the chemical shift assignments and the structural statistics for the SAIL-FLYA Runs 1–14. The chemical shift assignments were classified into three categories. The category 'all' includes all assignable $^{1}$H, $^{13}$C and, $^{15}$N atoms, the category 'backbone and $\beta$CH$_n$' includes the $^{1}$H, $^{13}$C, and $^{15}$N atoms in the protein backbone along with the H$^{\beta}$ and C$^{\beta}$ atoms, and the category 'other CH$_n$' includes the remaining side chain $^{1}$H, $^{13}$C, and $^{15}$N atoms, except H$^{\beta}$ and C$^{\beta}$. Runs 1, 10, and 11 utilized HNCO and/or HN(CA)CO experiments for the assignment of the backbone carbonyl carbons, C′. The number of assigned nuclei was higher for these runs than for those that did not include experiments to assign the C′ chemical shifts. The accuracy of the chemical shift assignment was evaluated in terms of the percentages of chemical shifts that are either within the tolerance of 0.03 ppm for $^{1}$H and 0.3 ppm for $^{13}$C and $^{15}$N, equal to those made by conventional assignment ('equal'), different from the reference assignment ('different'), or that do not agree within the tolerance with the reference assignment of any atom in the same residue ('wrong'). The latter type of assignment error can potentially lead to a serious distortion of the resulting structure, unless the subsequent NOESY assignment algorithm is able to discard the erroneous assignment. $^{13}$C and $^{15}$N atoms not bound to $^{1}$H were excluded when counting the 'wrong' assignments, because they have no influence on the NOE distance restraints and thus on the 3D structure. The percentages of equal and different peaks do not necessarily add up to 100%, because only the shifts of nuclei that were assigned simultaneously by both methods could be compared.

The fully automated approach with the full spectra set (Run 1) yielded the most complete and correct chemical shift assignments, which equaled the reference assignments in 96.5% of all assignable chemical shifts, 97.9% of the backbone and $\beta$CH$_n$ chemical shifts, and 93.3% of the outer side chain chemical shifts beyond the $\beta$ position (Table 2). Among the chemical shifts assigned by FLYA, 2.3% did not agree with the reference assignment of any atom in the same residue. With a decreasing number of input spectra, the percentages of equal chemical shift assignments also decreased slightly, to 96.2–93.5% for Runs 2–5 with a reduced number of spectra, and to 93.7–86.5% for Runs 6–14 with minimal data sets. Similarly, the percentages of wrong chemical shift assignments by FLYA that did not agree with the reference assignment of any atom in the same residue increased, to 2.7–3.4% for Runs 2–5, and to 3.8–10.1% for Runs 6–14. The lowest quality of the chemical shift assignments was observed when through-bond information was provided only by HNCO or HN(CA)CO spectra for the assignment of the backbone carbonyl carbon, C′. Relatively low assignment accuracy also resulted when H(CCCO)NH or HCCH-TOCSY were the only through-bond spectra used. In all other cases, the percentage of correctly assigned peaks decreased only slightly by 0.3–4.1% compared to the result obtained with the full set of spectra.

## NOE assignments and structure calculations

The number of assigned NOESY cross peaks varied between 1,644 and 1,757, and the number of long-range distance restraints ranged between 231 and 272 among Runs 1–14, without a trend for larger numbers of conformational restraints for the runs that included more spectra (Table 2). All final structures yielded average CYANA target function values below 0.5 Å$^2$, except for Run 14 with a final target function value of 1.11 Å$^2$ (Table 2). This indicates that in the case of SAIL proteins, the CYANA algorithm for automated NOESY assignment works robustly with the chemical shift lists obtained from the automated sequence-specific assignment step in FLYA, despite the aforementioned variations in the correctness of the chemical shift assignments.

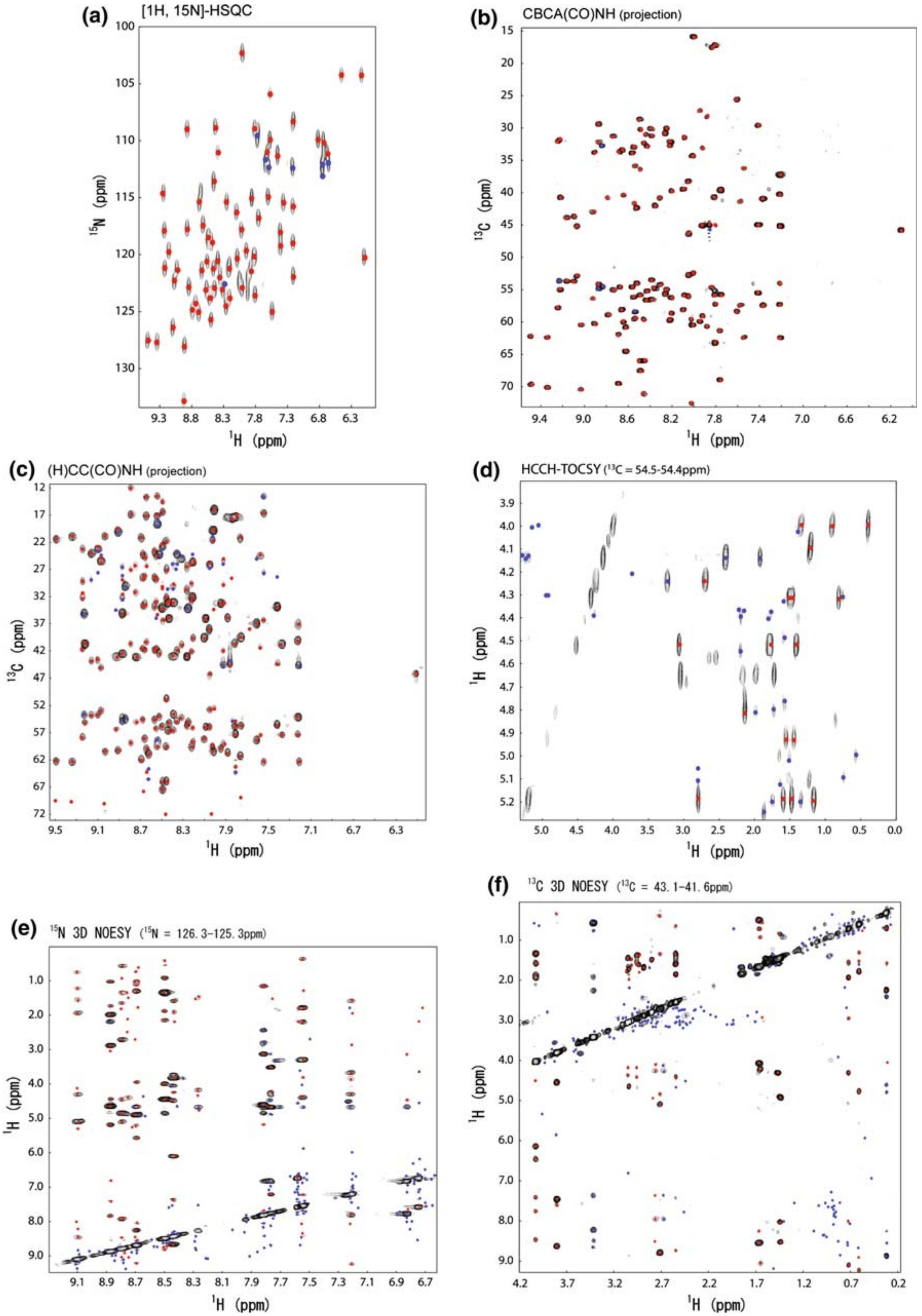

**Fig. 1** Assigned peaks (*red*) and unassigned peaks (*blue*) in Run 1 (see Table 1) of the SAIL-FLYA structure calculation of ubiquitin. **a** [$^{1}$H, $^{15}$N]-HSQC. **b** CBCA(CO)NH. **c** (H)CC(CO)NH. **d** HCCH-TOCSY. **e** $^{15}$N 3D NOESY. **f** $^{13}$C 3D NOESY. The 3D CBCA(CO)NH and (H)CC(CO)NH spectra are projected along the $^{15}$N dimension

**Table 2** Statistics of the ubiquitin chemical shift assignments and structures determined using the SAIL-FLYA algorithm with different sets of experimental NMR spectra, as defined in Table 1

| Spectrum | FLYA run | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| All chemical shift assignments | | | | | | | | | | | | | | |
| Assigned $^{1}H$, $^{13}C$, $^{15}N$ nuclei | 739 | 664 | 664 | 665 | 665 | 665 | 664 | 664 | 665 | 664 | 738 | 737 | 665 | 664 |
| Equal (%) | 96.5 | 96.2 | 95.5 | 93.5 | 93.8 | 92.4 | 92.5 | 92.3 | 92.8 | 90.3 | 88.7 | 86.5 | 93.7 | 89.6 |
| Different (%) | 2.8 | 3.2 | 3.9 | 5.9 | 5.6 | 7.0 | 6.9 | 7.1 | 6.6 | 9.1 | 10.6 | 13.0 | 5.7 | 9.8 |
| Wrong (%) | 2.3 | 2.7 | 2.6 | 3.1 | 3.4 | 4.7 | 5.2 | 4.7 | 3.8 | 5.9 | 7.9 | 10.1 | 4.5 | 7.6 |
| Backbone and $\beta CH_n$ chemical shift assignments | | | | | | | | | | | | | | |
| Assigned $^{1}H$, $^{13}C$, $^{15}N$ nuclei | 512 | 438 | 438 | 438 | 438 | 438 | 438 | 438 | 438 | 438 | 512 | 511 | 438 | 438 |
| Equal (%) | 97.9 | 98.0 | 97.4 | 97.4 | 96.8 | 96.0 | 96.6 | 96.5 | 96.2 | 92.9 | 91.1 | 89.5 | 95.1 | 90.9 |
| Different (%) | 1.6 | 1.5 | 2.1 | 2.1 | 2.7 | 3.5 | 2.9 | 3.0 | 3.4 | 6.6 | 8.3 | 10.1 | 4.4 | 8.7 |
| Wrong (%) | 1.6 | 1.5 | 1.9 | 1.8 | 2.1 | 2.8 | 2.6 | 2.1 | 2.5 | 5.3 | 7.5 | 9.2 | 4.2 | 7.4 |
| Other $CH_n$ chemical shift assignments | | | | | | | | | | | | | | |
| Assigned $^{1}H$, $^{13}C$, $^{15}N$ nuclei | 219 | 218 | 218 | 219 | 218 | 219 | 218 | 218 | 219 | 218 | 218 | 218 | 219 | 218 |
| Equal (%) | 93.3 | 92.5 | 91.6 | 87.4 | 89.0 | 87.1 | 86.3 | 86.1 | 88.3 | 87.0 | 85.2 | 81.4 | 90.7 | 87.4 |
| Different (%) | 5.8 | 6.6 | 7.5 | 11.7 | 10.1 | 12.0 | 12.8 | 13.0 | 10.8 | 12.1 | 13.9 | 17.7 | 8.4 | 11.7 |
| Wrong (%) | 4.1 | 5.0 | 4.1 | 5.2 | 5.4 | 7.2 | 8.7 | 8.2 | 5.0 | 6.3 | 7.6 | 10.7 | 5.3 | 7.8 |
| Structural statistics | | | | | | | | | | | | | | |
| Assigned NOESY cross peaks | 1,713 | 1,713 | 1,730 | 1,708 | 1,732 | 1,757 | 1,736 | 1,750 | 1,753 | 1,748 | 1,736 | 1,644 | 1,742 | 1,674 |
| Long-range ($\lvert i - j \rvert \geq 5$) distance restraints | 262 | 264 | 268 | 258 | 265 | 262 | 267 | 273 | 273 | 272 | 267 | 231 | 257 | 265 |
| CYANA target function (Å$^{2}$) | 0.12 | 0.37 | 0.83 | 0.33 | 0.38 | 0.45 | 0.35 | 0.32 | 0.44 | 0.49 | 0.44 | 0.28 | 0.33 | 1.11 |
| AMBER energy (kcal/mol) | −3,157 | −3,099 | −3,196 | −3,190 | −3,137 | −3,211 | −3,249 | −3,175 | −3,138 | −3,260 | −3,152 | −3,161 | −3,225 | −3,154 |
| Backbone RMSD to mean (Å) | 0.38 | 0.38 | 0.37 | 0.32 | 0.31 | 0.29 | 0.32 | 0.32 | 0.30 | 0.27 | 0.28 | 0.33 | 0.29 | 0.37 |
| All heavy atom RMSD to mean (Å) | 0.73 | 0.73 | 0.75 | 0.67 | 0.68 | 0.65 | 0.69 | 0.66 | 0.67 | 0.61 | 0.63 | 0.68 | 0.64 | 0.73 |
| Backbone RMSD to reference (Å) | 0.96 | 0.88 | 0.82 | 1.09 | 1.02 | 0.83 | 0.91 | 1.15 | 0.83 | 1.01 | 1.07 | 1.11 | 1.01 | 1.01 |
| All heavy atom RMSD to reference (Å) | 1.50 | 1.39 | 1.28 | 1.52 | 1.49 | 1.31 | 1.31 | 1.61 | 1.26 | 1.43 | 1.46 | 1.60 | 1.48 | 1.57 |

Chemical shift assignments were classified as “equal” when they coincided, within tolerances 0.03 ppm for $^{1}H$ and 0.3 ppm for $^{13}C/^{15}N$, with the corresponding shift from the reference assignment, as “different” when they differed by more than the tolerance from the value from the reference assignment, and as “wrong” when they did not match any conventionally assigned shift within the same residue. RMSD values were computed for the structured region of residues 1–72

## Structure accuracy

The 3D structures that resulted from the SAIL-FLYA calculations exhibited similar accuracy for all runs, indicating convergence to a correct high-quality structure for all 14 sets of input spectra (Fig. 2). The RMSD values to the reference structure obtained with conventional manual assignment varied in Runs 1–14 between 0.83 and 1.15 Å for the backbone atoms and 1.31–1.61 Å for all heavy atoms of the structured region, residues 1–72 of ubiquitin (Table 2). These deviations are comparable to the backbone RMSD values of up to 0.8 Å that could be observed in a series of conventional CYANA calculations with manually assigned chemical shifts and identical input data, except for different randomized initial structures. The results of the SAIL-FLYA calculations show that it is possible with SAIL to obtain high-quality structures of the protein ubiquitin with a minimal number of input spectra. The choice of the single triple resonance spectrum that was used as input in addition to the 2D spectra and the NOESY spectra was not crucial, as indicated by the virtually identical RMSD deviations from the reference structure for Runs 6–14.

## Impact of through-bond aromatic assignments

A particular advantage of the SAIL technology for the automation of protein structure analysis is that it enables the simple and reliable assignment of the aromatic side-chains, which typically form many long-range NOEs that are important to define the tertiary structure of the protein. The HB(CBCG)HE spectrum (Torizawa et al. 2005) correlates the $H^{\varepsilon}$ atoms of phenylalanine and tyrosine with the $H^{\beta}$ atoms by through-bond magnetization transfer independently from the intrinsically less reliable through-space correlations of the conventional approach (Wüthrich 1986). (The $H^{\delta}$ and $H^{\zeta}$ atoms are replaced by $^{2}H$ in SAIL proteins.) To evaluate the impact of the HB(CBCG)HE spectrum on the fully automatic SAIL-FLYA structure calculations, we repeated the aforementioned Runs 1–14 without using the HB(CBCG)HE spectrum as input. The results showed that the exclusion of the HB(CBCG)HE spectrum generally did not decrease the overall percentages of correctly assigned chemical shifts, but increased the number of erroneous assignments of $H^{\varepsilon}$ atoms of phenylalanine and tyrosine. The structures obtained with and without the use of the HB(CBCG)HE spectrum were of

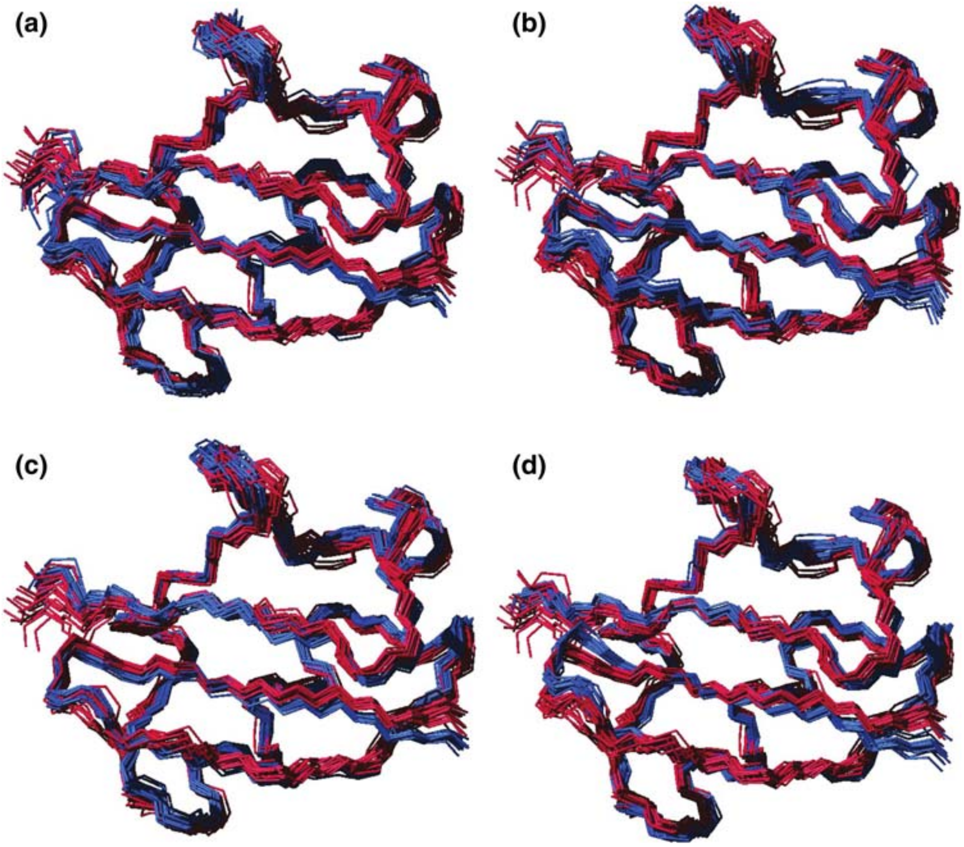

**Fig. 2** Ubiquitin structures obtained in SAIL-FLYA calculations (*blue*) superimposed on the conventionally determined reference structure of SAIL ubiquitin (*red*). **a** Run 1. **b** Run 3. **c** Run 7. **d** Run 14. See Tables 1 and 2 for details

similar accuracy except for Runs 12 and 5, for which the use of the HB(CBCG)HE spectrum decreased the backbone RMSD deviations from the reference structure from 3.35 to 1.11 Å and 1.44 to 1.02 Å, respectively. Considering that the 2D HB(CBCG)HE spectrum required only a small fraction of the total measurement time, its use is advisable for obtaining high-quality structures with SAIL-FLYA. Ubiquitin contains two phenylalanine residues and one tyrosine residue. A stronger impact of the HB(CBCG)HE spectrum is to be expected for proteins containing more aromatic residues than ubiquitin.

## Validation

The validation of NMR protein structures is an important issue that is complicated by the absence of a quality parameter that measures the deviation of the 3D structure from the original, un-interpreted data from the NMR spectrometer, in a similar way as the (free) *R*-factor in X-ray crystallography (Brünger 1992). As a partial remedy, many validation parameters have been proposed that measure the consistency of the 3D structure with either physico-chemical principles (conformational energies, hydrogen bonding, etc.), derived NMR data such as peak intensities or conformational restraints, or its "normality" in the context of the many protein 3D structures that are available from the Protein Data Bank (Spronk et al. 2004).

Recognizing erroneous NMR protein structures (Nabuurs et al. 2006) is particularly important for fully automated approaches that perform the complete structure analysis without manual checks and corrections. Fully automated approaches must be capable of detecting and discarding artifact peaks in order to deal with the imperfections of experimental NMR spectra. Such automated "noise removal" carries, in principle, the danger of converging to a wrong structure by excluding part of the experimental data (Herrmann et al. 2002; Linge et al. 2001). This did not happen with the final structures of any of the SAIL-FLYA runs of this paper, as indicated by their small RMSD deviations from the independently determined reference structure. Nevertheless, the complete set of 3D structures obtained in the course of the calculations for this paper provided a valuable data set for substantiating the validation approaches. Since the FLYA algorithm involves three stages (López-Méndez and Güntert 2006), and Runs 1–14 were each performed three times, with and without using the HB(CBCG)HE spectrum, the data set included a total 14 × 2 × 3 × 3 = 252 ubiquitin structure bundles of 20 conformers each. These structures were of widely varying quality, in terms of their deviation from the reference structure, but all were folded and free of trivial errors, such as incorrect bond lengths, bond angles or chiralities, or severe steric overlap (Hooft et al. 1996; Schultze and Feigon 1997). Since they were calculated and energy-refined in the same way, they provide a consistent set of structures to evaluate the power of various commonly used validation parameters (see "Methods") to distinguish between correct and erroneous structures. The database of the 252 ubiquitin structure bundles is available from www.sailnmr.org.

Figure 3 shows plots against the (logarithm of the) backbone RMSD deviation from the reference structure for seven validation parameters: The (logarithm of the) backbone RMSD to the mean coordinates, the AMBER potential energy (Cornell et al. 1995; Ponder and Case 2003), the percentage of residues in the most favored region of the Ramachandran plot defined by the program Procheck (Laskowski et al. 1996; Morris et al. 1992), the Verify3D score (Bowie et al. 1991; Lüthy et al. 1992), the packing score of the Whatcheck program (Hooft et al. 1996), the LGscore of the ProQ program (Wallner and Elofsson 2003), and the score of the ProSa 2003 program (Sippl 1993). In addition, an overall *Z* score was computed from the seven individual validation parameters, as described in the "Methods" section. All of these validation parameters correlated to a certain extent with the accuracy of the structure, as indicated by the correlation coefficients in Table 3. However, there were considerable differences. Only the Verify3D score, the Prosa2003 score, and the overall *Z* score showed a stronger correlation than the RMSD value to the mean coordinates of the structure bundle, which is generally considered as a bad quality measure because it describes the precision, rather than the accuracy, of a structure bundle. The strongest correlation, with a correlation coefficient of 0.91, was obtained, as expected, for the overall *Z* score, which incorporates contributions from all seven individual scores. The overall *Z* score could detect structures that are seriously wrong (backbone RMSD to the reference >2.0 Å) with high reliability: A cutoff for the *Z* score that was fulfilled by 99% of the structures with RMSDs to the reference below 2.0 Å included only a single structure with an RMSD above 2.0 Å (2.27 Å). Given the whole data set of 252 structures, the validation parameters were thus capable of distinguishing between correct and seriously wrong structures. On the other hand, it was difficult to detect slight deviations from the reference structure using the validation parameters of Fig. 3, as indicated by the low correlation coefficients that were obtained, if only the essentially correct structures with RMSD deviations from the reference structure of less than 2.0 Å were included in the analysis (Table 3). None of the individual validation scores showed a correlation coefficient above 0.5 for this reduced set of essentially correct structures. Only the overall *Z* score yielded a correlation coefficient of 0.66. This better result for the overall *Z* score reflects the fact that the

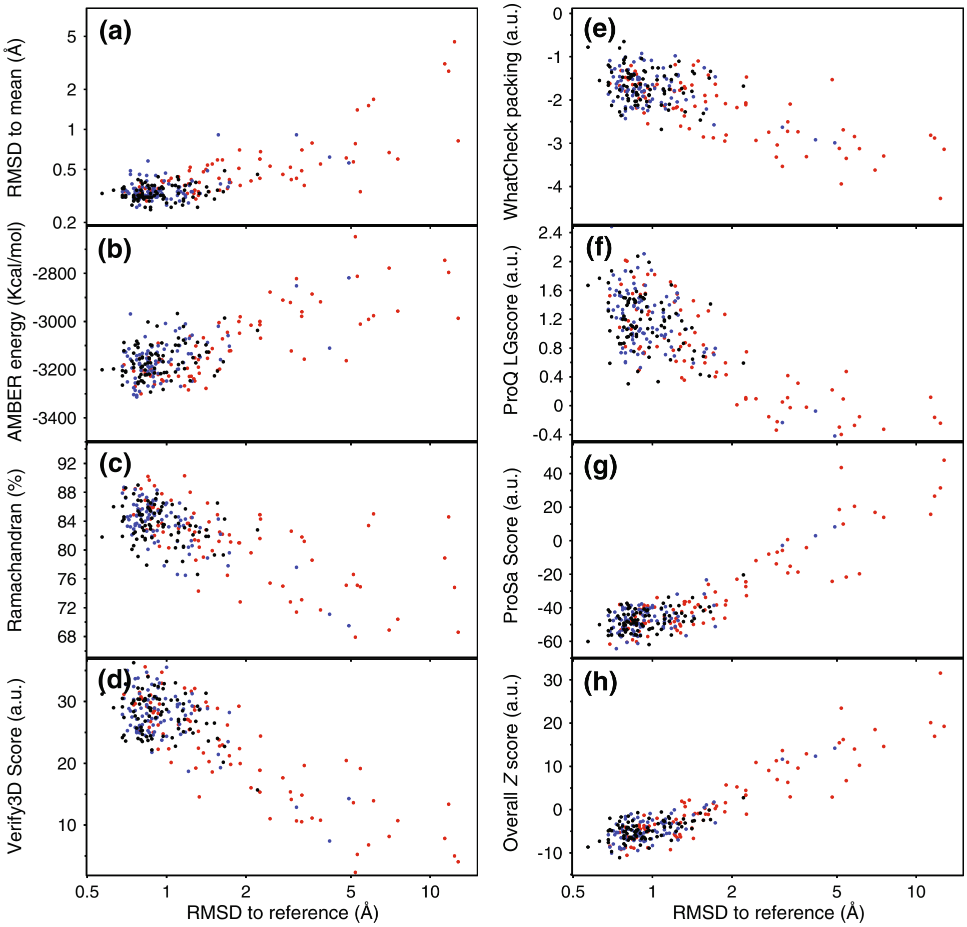

**Fig. 3** Validation scores of 252 SAIL ubiquitin structure bundles plotted against the backbone RMSD deviation from the reference structure on a logarithmic scale. **a** Backbone RMSD to the mean coordinates. **b** AMBER potential energy (Cornell et al. 1995; Ponder and Case 2003). **c** Percentage of residues in the most favored region of the Ramachandran plot, as defined by the program Procheck (Laskowski et al. 1996; Morris et al. 1992). **d** Verify3D score (Bowie et al. 1991; Lüthy et al. 1992). **e** Packing score of the Whatcheck program (Hooft et al. 1996). **f** LGscore of the ProQ program (Wallner and Elofsson 2003). **g** Score of the ProSa 2003 program (Sippl 1993). **h** Overall *Z* score, computed from the seven individual validation parameters, as described in the "Methods" section. Structures obtained in stages I, II, and III of the FLYA algorithm are represented by *red*, *blue*, and *black dots*, respectively

individual validation parameters evaluate complementary aspects of the structural quality.

## Conclusions

This study showed that it is possible to determine high quality structures of SAIL proteins by fully automatic structure calculation with FLYA, using a minimal number of spectra. These results suggest that besides the structure analysis of proteins larger than 40 kDa the SAIL method may have another application in the fully automated structure analysis of smaller proteins, provided that the future production cost for SAIL proteins can be lowered scaled-up quantities and improved syntheses of SAIL amino acids (Terauchi et al. 2008) Typically, with uniformly $^{13}C/^{15}N$-labeled proteins, a considerable number (up to 10) of 3D spectra are recorded that serve the sequence-specific assignment, but do not yield conformational restraints. It would therefore be advantageous for NMR structure analysis to reduce the measurement of such spectra as far as possible. SAIL, in conjunction with the FLYA algorithm, represents a significant step in this

**Table 3** Correlation coefficients between various structure quality scores and the logarithm of the backbone RMSD to the reference structure for the structured region of residues 1–72, calculated either over all structure bundles or only for the structure bundles with less than 2 Å RMSD to the reference structure

| Structure quality score | All structures | Structures with RMSD < 2 Å |
|---|---|---|
| Backbone RMSD to mean | 0.77 | 0.38 |
| AMBER energy | 0.74 | 0.36 |
| Procheck | | |
| Ramachandran most favored | 0.66 | 0.46 |
| Verify3D score | 0.81 | 0.37 |
| Whatcheck | | |
| Packing | 0.69 | 0.31 |
| Ramachandran | 0.43 | 0.36 |
| Rotamer | 0.31 | 0.11 |
| Backbone | 0.43 | 0.32 |
| ProQ score | | |
| Lgscore | 0.75 | 0.40 |
| MaxSub | 0.65 | 0.41 |
| Prosa2003 score | 0.88 | 0.44 |
| Overall *Z* score | 0.91 | 0.66 |

direction. A comparison of the results in this report with the analogous FLYA calculations performed for the uniformly $^{13}$C/$^{15}$N-labeled protein Fes SH2 (Scott et al. 2006) indicates a decisive improvement by SAIL-FLYA, since the former FLYA structure calculations for the uniformly labeled protein were unable to provide accurate structures when only a single additional 3D spectrum was used together with the indispensable NOESY spectra. For the future application of SAIL-FLYA fully automated structure determination with minimal spectra sets to larger proteins, improvements of the FLYA algorithm and of the quality of the NMR spectra for larger proteins may be necessary. The reduction of assignment ambiguities by the use of spectra with higher (effective) dimensionality and better resolution that can be achieved by projection and non-linear sampling techniques is particularly promising in this respect (Hiller et al. 2005; Kupče and Freeman 2008; Luan et al. 2005; Malmodin and Billeter 2005; Sakakibara et al. 2009; Szyperski and Atreya 2006). Furthermore the SAIL patterns may be overlap and relaxation optimized for large proteins (Ikeya et al. 2006).

**Acknowledgments** We thank Dr. Akira Mei Ono for help in preparing the SAIL amino acids. We gratefully acknowledge financial support by a Grant-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (JSPS), and from the CREST program of the Japan Science and Technology Agency (JST), the Technology Development for Protein Analyses and Targeted Protein Research Program of the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Lichtenberg program of the Volkswagen Foundation.

## References

Baran MC, Huang YJ, Moseley HNB, Montelione GT (2004) Automated analysis of protein NMR assignments and structures. Chem Rev 104:3541–3555

Bartels C, Billeter M, Güntert P, Wüthrich K (1996) Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. J Biomol NMR 7:207–213

Bartels C, Güntert P, Billeter M, Wüthrich K (1997) GARANT—a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. J Comput Chem 18:139–149

Billeter M, Wagner G, Wüthrich K (2008) Solution NMR structure determination of proteins revisited. J Biomol NMR 42:155–158

Bowie JU, Lüthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known 3-dimensional structure. Science 253:164–170

Brünger AT (1992) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355:472–475

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe—a multidimensional spectral processing system based on Unix pipes. J Biomol NMR 6:277–293

Gronwald W, Kalbitzer HR (2004) Automated structure determination of proteins by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 44:33–96

Güntert P (2003) Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43:105–125

Güntert P (2009) Automated structure determination from NMR spectra. Eur Biophys J 38:129–143

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

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

Hiller S, Fiorito F, Wüthrich K, Wider G (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA 102: 10876–10881

Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272

Ikeya T, Terauchi T, Güntert P, Kainosho M (2006) Evaluation of stereo-array isotope labeling (SAIL) patterns for automated structural analysis of proteins with CYANA. Magn Reson Chem 44:S152–S157

Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Meth Mol Biol 278:313–352

Johnson BA, Blevins RA (1994) NMR view—a computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614

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

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

Koradi R, Billeter M, Güntert P (2000) Point-centered domain decomposition for parallel molecular dynamics simulation. Comput Phys Commun 124:139–147

Kraulis PJ (1989) ANSIG: a program for the assignment of protein $^{1}$H 2D NMR spectra by interactive computer graphics. J Magn Reson 84:627–633

Kraulis PJ, Domaille PJ, Campbell-Burk SL, Van Aken T, Laue ED (1994) Solution structure and dynamics of Ras p21-GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. Biochemistry 33:3515–3531

Kupče E, Freeman R (2008) Hyperdimensional NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 52:22–30

Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486

Linge JP, O'Donoghue SI, Nilges M (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods Enzymol 339:71–90

López-Méndez B, Güntert P (2006) Automated protein structure determination from NMR spectra. J Am Chem Soc 128:13112–13122

Luan T, Jaravine V, Yee A, Arrowsmith CH, Orekhov VY (2005) Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. J Biomol NMR 33:1–14

Luginbühl P, Güntert P, Billeter M, Wüthrich K (1996) The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J Biomol NMR 8:136–146

Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with 3-dimensional profiles. Nature 356:83–85

Malmodin D, Billeter M (2005) High-throughput analysis of protein NMR spectra. Prog Nucl Magn Reson Spectrosc 46:109–129

Malmodin D, Papavoine CHM, Billeter M (2003) Fully automated sequence-specific resonance assignments of heteronuclear protein spectra. J Biomol NMR 27:69–79

Morris AL, Macarthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins 12:345–364

Nabuurs SB, Spronk CAEM, Vuister GW, Vriend G (2006) Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. PLoS Comput Biol 2:71–79

Pfändler P, Bodenhausen G, Meier BU, Ernst RR (1985) Toward automated assignment of nuclear magnetic resonance spectra—pattern recognition in two-dimensional correlation spectra. Anal Chem 57:2510–2516

Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Protein Chem 66:27–85

Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Wälchli M, Smith BO, Shirakawa M, Güntert P, Ito Y (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. Nature 458:102–105

Schultze P, Feigon J (1997) Chirality errors in nucleic acid structures. Nature 387:668

Scott A, López-Méndez B, Güntert P (2006) Fully automated structure determinations of the Fes SH2 domain using different sets of NMR spectra. Magn Reson Chem 44:S83–S88

Sippl MJ (1993) Recognition of errors in 3-dimensional structures of proteins. Proteins 17:355–362

Spronk C, Nabuurs SB, Krieger E, Vriend G, Vuister GW (2004) Validation of protein structures derived by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 45:315–337

Szyperski T, Atreya HS (2006) Principles and applications of GFT projection NMR spectroscopy. Magn Reson Chem 44:S51–S60

Takeda M, Ikeya T, Güntert P, Kainosho M (2007) Automated structure determination of proteins with the SAIL-FLYA NMR method. Nat Protoc 2:2896–2902

Takeda M, Sugimori N, Torizawa T, Terauchi T, Ono AM, Yagi H, Yamaguchi Y, Kato K, Ikeya T, Jee J, Güntert P, Aceti DJ, Markley JL, Kainosho M (2008) Structure of the putative 32 kDa myrosinase-binding protein from Arabidopsis (At3g164-50.1) determined by SAIL-NMR. FEBS J 275:5873–5884

Terauchi T, Kobayashi K, Okuma K, Oba M, Nishiyama K, Kainosho M (2008) Stereoselective synthesis of triply isotope-labeled Ser, Cys, and Ala: amino acids for stereoarray isotope labeling technology. Org Lett 10:2785–2787

Torizawa T, Shimizu M, Taoka M, Miyano H, Kainosho M (2004) Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol. J Biomol NMR 30:311–325

Torizawa T, Ono AM, Terauchi T, Kainosho M (2005) NMR assignment methods for the aromatic ring resonances of phenylalanine and tyrosine residues in proteins. J Am Chem Soc 127:12620–12626

Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12:1073–1086

Williamson MP, Craven CJ (2009) Automated protein structure calculation from NMR data. J Biomol NMR 43:131–143

Wüthrich K (1986) NMR of proteins and nucleic acids. Wiley, New York