Automated structure determination of proteins with the SAIL-FLYA NMR method

Mitsuhiro Takeda[1,2], Teppei Ikeya[1,2], Peter Güntert[2,3] & Masatsune Kainosho[1,2]

[1]Graduate School of Science, Center for Structural Biology, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan. [2]Graduate School of Science, Department of Chemistry, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji 192-0397, Japan. [3]Institute of Biophysical Chemistry and Center of Biological Magnetic Resonance, Department of Biochemistry, Chemistry and Pharmacy, University of Frankfurt, Max-von-Laue-Strasse 9, 60438 Frankfurt am Main, Germany. Correspondence should be addressed to M.K. (kainosho@nmr.chem.metro-u.ac.jp).

Published online 8 November 2007; doi:10.1038/nprot.2007.423

The labeling of proteins with stable isotopes enhances the NMR method for the determination of 3D protein structures in solution. Stereo-array isotope labeling (SAIL) provides an optimal stereospecific and regiospecific pattern of stable isotopes that yields sharpened lines, spectral simplification without loss of information, and the ability to collect rapidly and evaluate fully automatically the structural restraints required to solve a high-quality solution structure for proteins up to twice as large as those that can be analyzed using conventional methods. Here, we describe a protocol for the preparation of SAIL proteins by cell-free methods, including the preparation of S30 extract and their automated structure analysis using the FLYA algorithm and the program CYANA. Once efficient cell-free expression of the unlabeled or uniformly labeled target protein has been achieved, the NMR sample preparation of a SAIL protein can be accomplished in 3 d. A fully automated FLYA structure calculation can be completed in 1 d on a powerful computer system.

INTRODUCTION

NMR is a widely accepted method for the analysis of protein structures and dynamics in solution. However, the assignment of resonances and nuclear Overhauser effect (NOE) peaks by conventional interactive analysis is time-consuming and requires specialized knowledge in NMR, and, as the molecular weight of a protein becomes larger, the spectrum analysis of its NMR spectra becomes increasingly difficult, owing to broadened resonance lines and overlapped signals.

To overcome many of these problems, we developed the SAIL technique[1]. SAIL uses 20 chemically and enzymatically synthesized amino acids with a complete stereospecific and regiospecific pattern of stable isotopes that is optimal with regard to the quality and information content of the resulting NMR spectra as shown by Kainosho *et al.*[1]. The SAIL amino acids have the following features: first, stereo-selective replacement of one ^{1}H in each methylene group by ^{2}H; second, replacement of two ^{1}H in each methyl group by ^{2}H; third, stereo-selective modification of the prochiral methyl groups of Leu and Val such that one methyl is $-^{12}C(^{2}H)_3$ and the other is $-^{13}C^{1}H(^{2}H)_2$; and last, labeling of six-membered aromatic rings by alternating ^{12}C–^{2}H and ^{13}C–^{1}H moieties (see **Fig. 1**). The SAIL isotope pattern enables the detailed structure analysis of proteins of at least 40 kDa molecular weight. Further overlap and relaxation-optimized SAIL patterns for larger proteins can be conceived[2].

Compared to uniformly ^{13}C- and ^{15}N-labeled proteins (UL proteins), SAIL proteins have improved properties for NMR spectroscopy. The signals of SAIL proteins are sharper than those of UL proteins, owing to reduced transverse relaxation. The signal intensities for methylene groups are threefold to sevenfold higher with SAIL than with uniform labeling under the same conditions[1]. In addition, overlaps of NMR resonances are reduced because of the stereo-specific replacement of many ^{1}H by ^{2}H. In contrast to proteins that are uniformly or randomly labeled with ^{2}H, ^{13}C and ^{15}N, SAIL proteins contain many stereo-specifically placed and isotopically undiluted ^{1}H nuclei that provide an ample amount of distance restraints for the determination of the structure, including all side-chains. The SAIL method thus yields NMR spectra with a higher signal-to-noise ratio and less peak overlap than those obtained by uniform labeling. This reduces the time required for obtaining the assignments of resonances and NOE peaks.

The structure determination of SAIL proteins up to about 25 kDa can be fully automated using the recently introduced FLYA algorithm[3]. On the basis of a set of 2D and 3D NMR spectra (e.g., **Table 1**), FLYA assigns the backbone and side-chain resonances with the program GARANT[4,5] and uses the program CYANA[6–8] to calculate the structure of the target protein without human intervention solely on the basis of raw NMR spectra and the amino-acid sequence of the protein as input data, as recently shown for UL proteins with up to 15 kDa. The combined SAIL-FLYA method extends the applicability of fully automated NMR structure determination with FLYA to proteins with larger molecular weight. Here, we describe a protocol for the structure determination of SAIL proteins, including sample determination, NMR measurements, resonance assignment and structure calculation by FLYA.

Experimental design

SAIL amino acids are available from SAIL Technologies, a company that was recently established for supplying SAIL amino acids to the NMR community. To incorporate the SAIL amino acids into proteins of interest without metabolic scrambling, a cell-free system is used. The preparation of a S30 extract that is optimized for the *in vitro* synthesis of SAIL proteins is described in detail in **Box 1**, as we found that commercially available S30 extracts often contain non-labeled amino acids, which results in the incorporation of non-labeled amino acids into the protein[9]. The conditions of the cell-free expression, including the amount of amino acids and the incubation time and temperature, should be optimized in small-scale reactions before the large-scale reaction. The large-scale reaction with SAIL amino acids is started only after the successful expression of the corresponding UL protein by the identical protocol. In this protocol, the small-scale reactions are carried out with volumes of the inner and outer solution of 0.5 and 2.0 ml, respectively. The expression of the target protein is evaluated by SDS–polyacrylamide gel electrophoresis

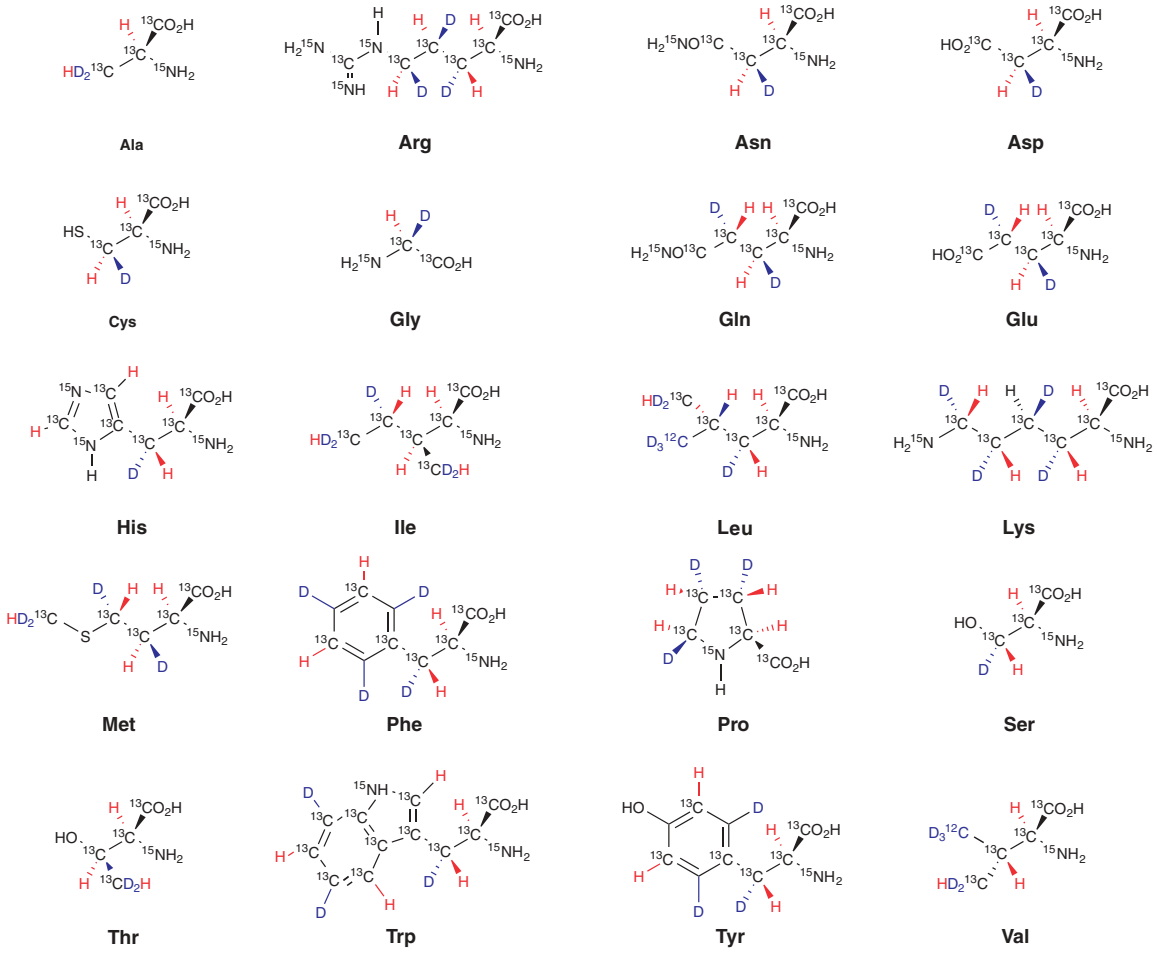

Figure 1 | Chemical structures of the SAIL amino acids. Symbols: H (red), ^{1}H; D (blue), ^{2}H; C, ^{12}C.

(SDS-PAGE) analysis. We recommend carrying out both *Escherichia coli* cellular expression and cell-free production of the target protein with uniform ^{15}N labeling. This allows for a comparison between the NMR spectra obtained from *in vivo* and *in vitro* expression. ^{1}H–^{15}N HSQC spectra from both approaches should be compared carefully to detect any possible difference between *in vivo* and *in vitro* expression. Compared with the conventional strategy using UL proteins, a limitation of the SAIL-FLYA method is the requirement to produce the protein sample from as small as possible an amount of SAIL amino acids by employing the cell-free system.

This protocol does not require a specific set of NMR spectra. Any of the common 2D and 3D spectra[10] for the assignment of the polypeptide backbone and the amino-acid side-chains can be used. In SAIL proteins, the detection and assignment of signals from aromatic rings containing alternating ^{12}C and ^{13}C nuclei (**Fig. 1**) are straightforwardly achieved using an unconventional approach[11].

This protocol provides two alternatives for the structure calculation. Fully automated structure determination with the FLYA algorithm can be used for high-quality data sets, that is, for monomeric proteins for which a highly concentrated (0.4 mM or more) and stable (2 weeks or longer) sample is available to acquire a series of 3D spectra with good signal-to-noise ratio that do not show significant conformational exchange and that do not exhibit extensive peak overlap. In the case of spectra that exhibit one or several of these

TABLE 1 | NMR experiments commonly used for the SAIL-FLYA method.

Backbone assignment	Side-chain assignment	NOE distance restraints
2D ^{1}H–^{15}N HSQC	2D ^{1}H–^{13}C CT-HSQC	3D ^{1}H–^{15}N NOESY-HSQC
2D CBCGHE[11]	2D ^{1}H–^{13}C HSQC	3D ^{1}H–^{13}C NOESY-HSQC (offset set to aliphatic region)
3D HNCA	3D H(CCCO)NH-TOCSY	3D ^{1}H–^{13}C NOESY-HSQC (offset set to aromatic region)
3D HN(CO)CA	3D (H)C(CCO)NH-TOCSY	
3D HNCACB	3D HCCH-TOCSY	
3D HNCOCACB	3D HCCH-COSY	
3D HNCO	3D HBHA(CO)NH	
3D HN(CA)CO	3D HBHANH	
	3D HBCB(CG)HE (ref. 11)	

HSQC, heteronuclear single quantum coherence; TOCSY, total correlation spectroscopy

BOX 1 | PREPARATION OF S30 EXTRACT

1. Inoculate stock of *E. coli* (A19, BL21 Star, etc.) into 10 ml of LB medium in a 50 ml tube and grow the cells overnight at 37 °C with shaking.
2. Inoculate 10 ml of the culture medium into 1 l of incomplete rich medium in a 2-liter flask.
3. Grow the cells at 37 °C with shaking to an OD_{650} of 0.7.

▲ **CRITICAL STEP** The growth rate of the cells in the culture correlates with the activity of the resulting extract.

4. Centrifuge the cells (5,000*g*, 4 °C, 10 min) and wash them with 200 ml of ice-cold S30 buffer containing 0.05% 2-mercaptoethanol by suspending moderately three times.

▲ **CRITICAL STEP** Do not allow foaming of the suspension.

■ **PAUSE POINT** Can be left overnight in the centrifuge tube at −80 °C.

5. Gently resuspend the cell pellet with 200 ml of ice-cold S30 buffer containing 0.05% (vol/vol) 2-mercaptoethanol by slurrying. Centrifuge them (5,000*g*, 4 °C, 10 min) and weigh the *E. coli* pellets. Resuspend the pellet in 1.27 ml of S30 buffer per gram of *E. coli*.
6. Disrupt the cells with the French Press at 20,000 psi (1,400 kg cm^{-2}). Add 30 µl of 1 M DTT to the lysate. Centrifuge (30,000*g*, 4 °C, 30 min) using DEPC-treated/autoclaved centrifuge tubes. Carefully remove approximately 1.4 ml of the supernatant per gram of *E. coli* without mixing with the precipitate.
7. Transfer the supernatant to the unused DEPC-treated/autoclaved tubes. Centrifuge them (30,000*g*, 4 °C, 30 min) and remove approximately 1.0 ml of the supernatant per gram of *E. coli* into a 50 ml tube.
8. Shake the tube at 37 °C for 80 min.
9. Dialyze the solution at 4 °C for 45 min against 2 liters of S30 buffer using a dialysis tube with MWCO of ~6,000–8,000. Allow a little air into the tube to float. Repeat the dialysis twice. Centrifuge (15,000*g*, 10 min, 4 °C).

! **CAUTION** The dialysis tube should be treated with DEPC-containing water and then rinsed with RNase-free water before use.

10. Fill an open column (Econo-column chromatography column 2.5 cm × 20 cm) with Sephadex G25 resin uniformly and set the column vertically in cold space. Attach an Econo-column funnel to the top end of the column. Pour the 500 ml of S30 buffer from the funnel into the column.

! **CAUTION** Uniformity of the filled resin affects resolution of gel filtration.

11. Apply the supernatant from Step 9 to a column that was pre-equilibrated at Step 10 at 4 °C. After loading the supernatant, continue to supply the funnel with the S30 buffer to maintain the flow in the column. When the first fraction reached to the bottom, start to collect 1.4 times the volume of the applied extract. Determine the first fraction judging from its colour and turbidity.

! **CAUTION** The fraction to be collected looks yellow.

12. Dialyze the eluate at 4 °C for 70–80 min against 700 ml of an equal weight mixture of PEG-8000 and S30 buffer. Before use, the PEG-S30 buffer at 4 °C should always be stirred to avoid PEG deposition. Adjust the dialysis time so as to concentrate the extract up to 0.86 times the volume. Dialyze it at 4 °C for 60 min against 2 liters of S30 buffer.
13. Transfer the extract to 1.5 ml tubes. Freeze the tubes with liquid nitrogen. Store them at −80 °C.

● **TIMING** Steps 2–4, 6–7 h; Steps 5–12, 7–8 h

deficiencies that make it difficult to correctly identify a sufficiently complete set of cross peaks, a semiautomated approach can be used that consists of interactive resonance assignment and automated evaluation of the nuclear Overhauser enhancement spectroscopy (NOESY) peak lists to provide the conformational restraints for the structure calculation. The range of applicability of the fully automated approach can be extended without resorting to time-consuming interactive resonance assignment by visually inspecting the results of peak picking before the start of the automated resonance and NOESY assignment with FLYA. The FLYA approach also benefits from redundancy in the spectral information, that is, if the same signals can be obtained from several different spectra.

MATERIALS

REAGENTS

- SAIL amino-acid mixture (SAIL Technologies) ! **CAUTION** SAIL amino-acid mixture purchased from SAIL technologies contains 10 mg of each SAIL amino acid. Tryptophan and tyrosine are less soluble in water than other amino acids.
- Diethyl pyrocarbonate (DEPC; Sigma-Aldrich, cat. no. D5758) ! **CAUTION** DEPC is a carcinogen. Handle it carefully with a sanitary glove.
- BL21 Star (DE3) (Invitrogen, cat. no. 44-0049)
- KH_2PO_4 (Wako, cat. no. 169-04245)
- K_2HPO_4 (Wako, cat. no. 164-04295)
- Bacto yeast extract (Difco, cat. no. 212750)
- Thiamine hydrochloride (Wako, cat. no. 201-00852)
- Glucose (Wako, cat. no. 041-00595)
- $Mg(OAc)_2$ (Sigma, cat. no. M-2545)
- KOAc (Wako, cat. no. 160-03175)
- DTT (Wako, cat. no. 049-08972)
- 2-Mercaptoethanol (Wako, cat. no. 137-06862)
- Dialysis tube used for preparations of the S30 extract (Spectrum, cat. no. 132655)
- Sephadex G25 medium (GE Healthcare, cat. no. 17-0033-02)
- PEG-8000 (Sigma, cat. no. P-2139)
- NH_4OAc (Wako, cat. no. 019-02835) ! **CAUTION** If amide groups of side chain for Asn and Gln are labeled with ^{15}N, use ^{15}N NH_4OAc (CIL, cat. no. NLM 177-1).
- Creatine phosphate (Wako, cat. no. 306-50523)
- RNase inhibitor (human placenta) (Takara, cat. no. 2310A)
- Creatine kinase (Roche, cat. no. 127566)
- HEPES (Dojindo; Wako, cat. no. 340-01371)
- KOH (Wako, cat. no. 168-03855)
- ATP (Sigma, cat. no. A-7699)
- CTP (Sigma, cat. no. C-1506)
- GTP (Sigma, cat. no. G-8877)
- UTP (Sigma, cat. no. U-6625)
- cAMP (Wako, cat. no. 017-16903)
- Folinic acid (Sigma, cat. no. F-7878)

• tRNA (Roche, cat. no. 109541)

EQUIPMENT

• French press (Ohtake)
• Econo-column chromatography column 2.5 cm × 20 cm (Bio-Rad, cat. no. 737-2521)
• Econo-column funnel (Bio-Rad, cat no. 731-0003)
• Float-A-lyzer (Spectra/Por cat. no. 235058)
• 800 MHz (or higher field) NMR spectrometer
• Linux computer system for the FLYA calculations (multiple processors recommended)

REAGENT SETUP

RNase-free water DEPC-treated, autoclaved water or Milli-Q PF Plus water can be used[12].

Incomplete rich medium For 1 liter of medium, combine the following: 5.6 g of KH_2PO_4, 28.9 g of K_2HPO_4, 1 g of Bacto yeast extract, 1.5 mg of thiamine, 50 ml of 40% (wt/vol) D-glucose and 10 ml of 0.1 M $Mg(OAc)_2$, glucose and $Mg(OAc)_2$ should be added after the autoclaving.

S30 buffer For 1 liter of buffer, combine the following: 10 ml of 1 M Tris-acetate (pH 8.2), 10 ml of 1.4 M $Mg(OAc)_2$, 10 ml of 6 M KOAc and 1 ml of 1 M DTT (add after autoclaving).

LM mixture For 200 ml of mixture, combine the following: 22 ml of 2 M HEPES-KOH (pH 7.5), 33.4 ml of 6 M K(OAc), 210 mg of DTT, 530 mg of ATP, 338 mg of CTP, 335 mg of GTP, 310 mg of UTP, 172 mg of cAMP, 28 mg of folinic acid, 140 mg of tRNA, 64 ml of 50% (wt/vol) PEG-8000, RNase-free water, up to 200 ml. Prepared LM mixture can be frozen at −20 °C for 1 month or more. In the case of SDS-PAGE analysis of cell-free reactant, PEG-8000, which hampers SDS-PAGE analysis, should be removed by ethanol precipitation before addition of a sample dye to the reaction solution.

EQUIPMENT SETUP

French press Thoroughly wash all parts of the instrument with tap water and then with MilliQ (twice), DEPC-containing water (once) and RNase-free water (once). This wash is done the day before use and the washed parts are chilled at 4 °C overnight.

Gel filtration column We use a column with the size of 2.5 cm × 20 cm filled with Sephadex G25 medium. The resin should be swelled with DEPC-containing water and the empty column should have been autoclaved the day before use.

Computer and software setup Although structure calculations can, in principle, be run on a computer with a single processor, we use for efficiency a cluster of Linux computers interconnected by Gigabit Ethernet. Almost ideal speedup can be obtained by running the calculations in parallel on up to as many processors as independent conformers are calculated with the program CYANA[6–8] or independent assignment runs are performed with the program GARANT[4,5].

The SAIL-adapted version of the program CYANA, including the FLYA algorithm[3], is required. In addition, an interactive NMR spectrum analysis program such as XEASY[13], NMRView[14] (http://www.onemoonscientific.com) or Sparky (T.D. Goddard and D.G. Kneller, University of California, San Francisco; http://www.cgl.ucsf.edu/home/sparky) is needed if one performs semiautomated structure determination, including manual or manually screened peak picking and conventional, interactive resonance assignment. For fully automated structure analysis with FLYA, one needs, in addition to the program CYANA, the program NMRView or AUTOPSY[15] for automated peak picking and the program GARANT for automated resonance assignment.

PROCEDURE

Preparation of SAIL proteins by the *E. coli* cell-free method

1| Prepare the reaction solution and the dialysis solution by mixing the components as shown in **Table 2**. Dissolve SAIL amino-acid mixture in water and then add it to the cell-free reaction solution. If SAIL amino acids appear insoluble in water, warm it up to 60 °C.

! CAUTION Use sanitary gloves to prevent contamination with RNases. Thaw the frozen S30 extract on ice. Prepare creatine phosphate in RNase-free water just before use. An excess of heating to SAIL amino acid may cause a racemization especially at high pH.

2| Cut outer tube of the Float-A-lyzer at an appropriate height such that inner solution in the tube is completely soaked in the dialysis solution when an inner membrane apparatus is set to the outer tube. Pour the dialysis solution into the outer tube. Place the inner membrane apparatus of the Float-A-lyzer to the outer tube and pour the reaction solution into the inner membrane. Cover the tube with Parafilm.

3| Shake the tube to allow for production of target proteins. The optimal temperature and incubation times should be determined in small-scale cell-free reactions with volumes of the reaction and dialysis solutions of 0.5 and 2.0 ml, respectively, before the large-scale reaction.

4| Retrieve the reaction solution and the dialysis solution. If the produced protein has a molecular weight smaller than molecular weight cutoff of the membrane, check the outer solution for presence of the protein.

? TROUBLESHOOTING

NMR sample preparation and measurements

5| Purify the produced protein according to purification procedures of the target protein.

! CAUTION The N-terminus of the protein produced by cell-free expression may be heterogeneous owing to incomplete deformylation by peptide deformylase. This can be overcome by using a cleavable N-terminal tag[9].

6| Transfer the prepared sample into the NMR tube.

7| Collect NMR spectra required for assignments and structural calculation.

TABLE 2 | Composition of reaction and dialysis solutions in cell-free reactions.

Stock solution	Reaction solution	Dialysis solution
RNase-free water	1,120 μl	11,608 μl
1.4 M NH_4OAc	98 μl	392 μl
0.5 M $Mg(OAc)_2$	150 μl	600 μl
SAIL amino-acid mixture (25 mM each)	200 μl	800 μl
0.645 M creatine phosphate	400 μl	1,600 μl
LM mixture	1,250 μl	5,000 μl
1 mg ml^{-1} template DNA	100 μl	—
11 mg ml^{-1} T7 RNA polymerase	45 μl	—
40 U μl^{-1} RNase inhibitor	12.5 μl	—
10 mg ml^{-1} creatine kinase	125 μl	—
S30 extract	1,500 μl	—
Total volume	5 ml	20 ml

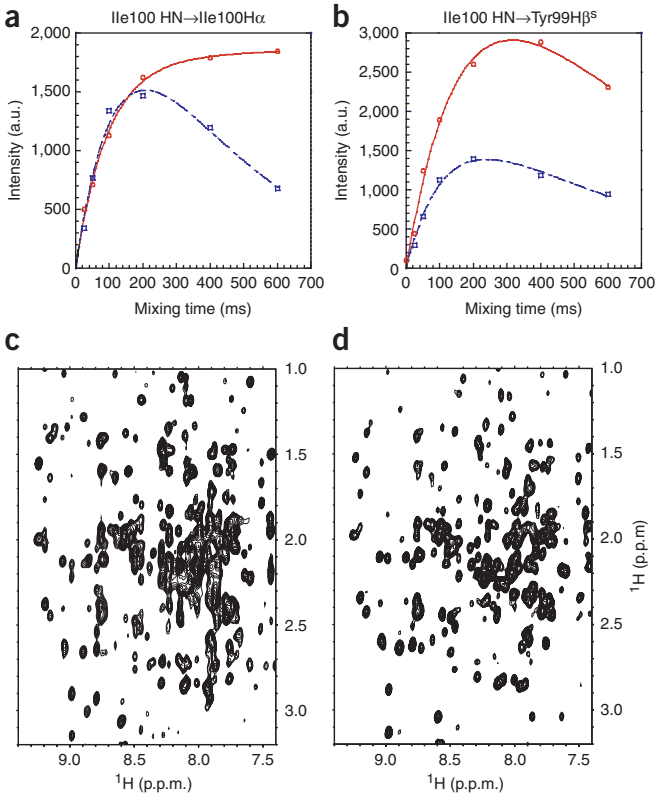

Figure 2 | Comparison of NOESY data for SAIL-CaM and UL-CaM. The SAIL-CaM and UL-CaM samples each contained 0.7 mM protein, 5 mM MES-d13 and 10 mM bis-Tris-d19 (Cambridge Isotope Laboratories), 5 mM $CaCl_2$ and 0.1 mM NaN_3, pH 6.5. Intensities of the cross peaks (**a**) Ile100HN-Ile100H$^{\alpha}$ and (**b**) Ile100HN-Tyr99H$^{\beta s}$ as a function of the mixing time in 3D NOESY-HSQC experiments performed with SAIL-CaM (red) or UL-CaM (blue). Regions from the 2D NOESY spectra for (**c**) UL-CaM and (**d**) SAIL-CaM. The NOESY spectra for SAIL-CaM and UL-CaM were obtained under identical conditions at 37 °C on a Bruker DRX800 spectrometer equipped with a TXI xyz-gradient probe and scaled to equal noise levels.

! CAUTION Deuterium decoupling should be applied during ^{13}C evolution times. Constant-time evolution is not required for observations of ^{1}H–^{13}C pair in aromatic region owing to absence of adjacent ^{13}C–^{13}C coupling in SAIL aromatic amino acids. In NOESY experiments, the optimal mixing time can be longer than for UL proteins owing to decreased spin diffusion (**Fig. 2a**).
▲ CRITICAL STEP Check that a 2D ^{1}H–^{15}N HSQC spectrum of the produced protein is identical to that of the UL-protein.
? TROUBLESHOOTING

Spectra analysis and structure determination

8| Interactive resonance assignment is followed by combined semiautomated NOE assignment and structure calculation with CYANA (option A) or fully automated structure analysis with FLYA (option B).

(A) Semiautomated structure determination

(i) Determine sequence-specific resonance assignments by established interactive methods.
! CAUTION In aliphatic moieties, the ^{13}C chemical-shift values for SAIL proteins differ from those for conventional UL proteins owing to ^{2}H isotope shifts. If the TALOS program[16] is used to obtain dihedral angle restraints, the chemical-shift values of corresponding resonances for UL proteins should be used. To facilitate this process, we are now investigating average isotope shifts of UL versus SAIL for each amino acid, which will be available in the near future.
▲ CRITICAL STEP More than 90% of the backbone and side-chain chemical shifts should be assigned[7,17].

(ii) Prepare NOESY peak lists containing the positions and volumes or intensities of the cross peaks in the NOESY spectra.
! CAUTION Use consistent chemical shift referencing in the NOESY spectra and for the chemical shifts determined in Step A(i).

(iii) Perform combined automated NOE assignment[7] and structure calculation using torsion angle dynamics-driven simulated annealing[6] with CYANA[8]. Typical calculation parameter values are given in **Table 3**. This step takes 0.2–5 h, depending on performance of the computer and protein size.
! CAUTION The resulting structure can be unreliable if the structure bundle obtained in the first cycle of combined automated NOE assignment and structure calculation has a root mean square deviation (RMSD) of more than 3 Å for the backbone atoms, excluding flexible regions[7,17].

(B) Fully automated structure determination with the FLYA algorithm

(i) Prepare the spectra and set the parameters for the automated peak-picking step within FLYA. In the present version of FLYA, the programs NMRView[14] or Autopsy[15] can be used for this step. Peak lists must be named as according to the spectrum type (**Table 1**). Alternatively, peak lists may also be prepared by interactive peak picking.
! CAUTION Use consistent chemical shift referencing over all spectra.

(ii) Set the parameters for the project-specific peak list preparation and filtering step within FLYA[3]. This step may include, for instance, the elimination of peaks on and near the diagonal or the water line, recovering the original chemical-shift coordinates for folded peaks, etc.

TABLE 3 | Parameters for SAIL-FLYA calculations.

Parameter	Typical value
Tolerance for ^{1}H chemical shift matching	0.03 p.p.m.
Tolerance for ^{13}C and ^{15}N chemical shift matching	0.5 p.p.m.
Number of independent GARANT assignment runs (FLYA)	20
Median upper distance bound for automated NOE calibration	4.2 Å
Number of conformers for structure calculations	100
Number of accepted conformers from structure calculations	20
Number of torsion angle dynamics steps per conformer	10,000

These parameter values apply both for semiautomated structure determination and for fully automated structure determination with the FLYA algorithm, where applicable.

(iii) Perform the FLYA calculation[3] within the program CYANA using an automated peak-picking program, for example, NMRView or AUTOPSY, and the automated chemical-shift assignment program GARANT as plug-ins. Important calculation parameters are given in **Table 3**. Typically, an ensemble of 10–20 initial chemical-shift assignments for every nucleus is generated by a corresponding number of independent GARANT runs. FLYA consolidates the results into a single consensus chemical-shift value for every nucleus, which is in turn used for the assignment of NOESY cross peaks with CYANA. The structure calculations are normally started from 100 initial conformers with random torsion angle values.

▲ **CRITICAL STEP** This step takes 3–48 h, depending on the performance of computer and protein size.

■ **PAUSE POINT** FLYA calculations require several hours of unattended computation time. Multiple processors can be used to almost ideally speed up the number of individual GARANT runs that are performed by FLYA.

(iv) Analyze the results of the FLYA calculation. Evaluate the FLYA reliability measures[3]. For more details, check the extent of peak picking and peak assignments in the different spectra, the results of combined automated NOESY assignment and structure calculation, and the final 3D structure with a molecular graphics program, for example, MOLMOL[18].

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 4**.

TABLE 4 | Troubleshooting table.

Problem	Possible reason	Solution
Low production in the cell-free reaction (Steps 1–4)	Problem with used reagent	Check the reagents, but most plausible case is LM mixture or S30 extract
Low labeling degree (Steps 1–4)	Non-labeled amino acids in the S30 extract	Use S30 extract free from non-labeled amino acids
Doubling of NMR signals (Step 7)	Incomplete deformylation in the cell-free reaction	Use a construct encoding the proteins with a cleavable N-terminus
No, or very few, peaks assigned by FLYA in a certain spectrum (Step 8B)	Incorrect specification of the spectral dimensions, or inconsistent chemical shift referencing	Check CYANAFORMAT specification in the peak list header or correct chemical shift referencing
No convergence in structure calculation (Step 8)	Insufficient number of NOE distance restraints	Check the extent and quality of automated NOESY peak picking

ANTICIPATED RESULTS

The efficient incorporation of SAIL amino acids into the protein of interest is crucial for the practical feasibility of this approach. In the case of calmodulin (CaM), 4 mg of purified soluble protein was obtained from 40 mg of SAIL amino-acid mixture[9]. Similar yields were observed for other proteins, including, for instance, the 41 kDa maltose-binding protein[1]. Compared with UL-CaM, SAIL-CaM yielded spectra with reduced overlap of signals and strongly improved signal-to-noise ratio[1]. In the collection of NMR spectra, we modified the existing NMR experiments for SAIL as described above. Whereas the number of short ^{1}H–^{1}H distances that give rise to NOESY cross peaks is reduced in SAIL-CaM by more than 40% relative to UL-CaM, the expected corresponding number of meaningful distance restraints remains virtually unchanged, because in the case of UL-CaM, many NOESY cross peak pairs originate from diastereotopic partners[1]. Such pairs of cross peaks provide less information than the corresponding single, stereospecifically assigned NOE in SAIL-CaM. With the resulting spectra for SAIL-CaM, a high

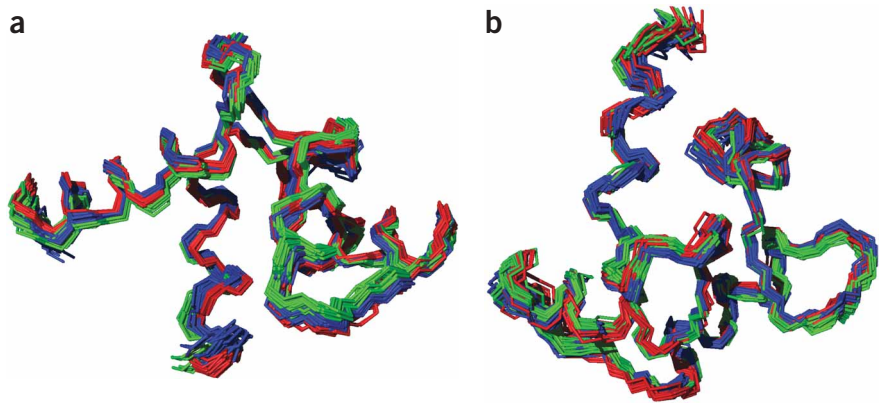

Figure 3 | Solution structure of SAIL-CaM. (**a**) N-terminal domain of residues 5–75. (**b**) C-terminal domain of residues 82–146. The structure obtained with SAIL-FLYA on the basis of an extensive set of ten 3D spectra is shown in red. The structure obtained with SAIL-FLYA from a minimal set of six 3D spectra is shown in blue. The structure that had been determined earlier by traditional, manual assignment methods is shown in green. Ten conformers are depicted for each structure bundle. The figure was prepared with the program MOLMOL[18].

degree of assignment was achieved readily. The optimal value of the NOESY mixing time was found to be larger for SAIL-CaM than that for UL-CaM owing to reduced spin diffusion (**Fig. 2a,b**). The NOESY spectra themselves were also well resolved, and peaks that gave reliable distance restraints for defining the structure of CaM were obtained without redundancy (**Fig. 2c,d**). We performed FLYA calculations using two different sets of input spectra, both of which were significantly smaller than those used for the initial FLYA calculations of UL proteins[3]. The 'extensive' set of spectra included three 2D HSQC spectra (^{15}N-HSQC, ^{13}C-HSQC, CBCGHE[11]), eight 3D through-bond spectra for chemical-shift assignment (HNCA, HN(CO)CA, CBCANH, CBCA(CO)NH, HBHANH, HBHA(CO)NH, HCCH-TOCSY, HBCBCGHE[11]) and two 3D NOESY spectra (^{15}N-edited NOESY, ^{13}C-edited NOESY). In the 'minimal' set of spectra, only four 3D through-bond spectra (HNCA, CBCA(CO)NH, HCCH-TOCSY, HBCBCGHE) were retained besides the 2D spectra and the 3D NOESY spectra. Peak picking was performed interactively with NMRView. In both cases, almost complete chemical-shift assignments and a well-converged structure of CaM were obtained (**Fig. 3**). FLYA yielded identical chemical-shift assignments as the conventional manual approach for 97.9% of all backbone and side-chain resonances, and 98.9% of the chemical shifts were assigned to the correct residue even with the minimal set of input spectra. The structures calculated with SAIL-FLYA deviated from the conventionally determined SAIL-CaM structure by 0.75–0.88 Å backbone RMSD for the structured regions of the flexibly linked N-and C-terminal domains. The remaining deviations are largely confined to surface-loop regions.

ACKNOWLEDGMENTS This work was supported by MEXT (Technology Development for Protein Analyses and Targeted Protein Research Program) and JST (CREST) to M.K. The work was also supported by a Grant-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (JSPS) and by the Volkswagen Foundation to P.G. The authors greatly thank Drs. Tsutomu Terauchi and Akira Mei Ono of Tokyo Metropolitan University for their help in preparing the SAIL amino acids.

COMPETING INTERESTS STATEMENT The authors declare competing financial interests (see the HTML version of this article for details).

Published online at http://www.natureprotocols.com
Rights and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Kainosho, M. *et al.* Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52–57 (2006).
2. Ikeya, T., Terauchi, T., Güntert, P. & Kainosho, M. Evaluation of stereo-array isotope labeling (SAIL) patterns for automated structural analysis of proteins with CYANA. *Magn. Reson. Chem.* **44**, S152–S157 (2006).
3. López-Méndez, B. & Güntert, P. Automated protein structure determination from NMR spectra. *J. Am. Chem. Soc.* **128**, 13112–13122 (2006).
4. Bartels, C., Güntert, P., Billeter, M. & Wüthrich, K. GARANT—a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. *J. Comput. Chem.* **18**, 139–149 (1997).
5. Bartels, C., Billeter, M., Güntert, P. & Wüthrich, K. Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. *J. Biomol. NMR* **7**, 207–213 (1996).
6. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
7. Herrmann, T., Güntert, P. & Wüthrich, K. 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 (2002).
8. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 (2004).
9. Torizawa, T., Shimizu, M., Taoka, M., Miyano, H. & Kainosho, M. Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol. *J. Biomol. NMR* **30**, 311–325 (2004).
10. Cavanagh, J., Fairbrother, W.J., Palmer , A.G.III, Skelton, N.J. & Rance, M. *Protein NMR Spectroscopy. Principles and Practice* (Academic Press, San Diego, CA, 2006).
11. Torizawa, T., Ono, A.M., Terauchi, T. & Kainosho, M. NMR assignment methods for the aromatic ring resonances of phenylalanine and tyrosine residues in proteins. *J. Am. Chem. Soc.* **127**, 12620–12626 (2005).
12. Huang, Y.H., Leblanc, P., Apostolou, V., Stewart, B. & Moreland, R.B. Comparison of Milli-Q PF Plus water to DEPC-treated water in the preparation and analysis of RNA. *Biotechniques* **19**, 656–661 (1995).
13. Bartels, C., Xia, T.H., Billeter, M., Güntert, P. & Wüthrich, K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **6**, 1–10 (1995).
14. Johnson, B.A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352 (2004).
15. Koradi, R., Billeter, M., Engeli, M., Güntert, P. & Wüthrich, K. Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. *J. Magn. Reson.* **135**, 288–297 (1998).
16. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
17. Jee, J. & Güntert, P. Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. *J. Struct. Funct. Genomics* **4**, 179–189 (2003).
18. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).