ARTICLES

Optimal isotope labelling for NMR protein structure determinations

Masatsune Kainosho[1], Takuya Torizawa[1], Yuki Iwashita[1], Tsutomu Terauchi[1], Akira Mei Ono[1] & Peter Güntert[2]

Nuclear-magnetic-resonance spectroscopy can determine the three-dimensional structure of proteins in solution. However, its potential has been limited by the difficulty of interpreting NMR spectra in the presence of broadened and overlapping resonance lines and low signal-to-noise ratios. Here we present stereo-array isotope labelling (SAIL), a technique that can overcome many of these problems by applying 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. SAIL uses exclusively chemically and enzymatically synthesized amino acids for cell-free protein expression. We demonstrate for the 17-kDa protein calmodulin and the 41-kDa maltodextrin-binding protein that SAIL offers sharpened lines, spectral simplification without loss of information, and the ability to rapidly collect the structural restraints required to solve a high-quality solution structure for proteins twice as large as commonly solved by NMR. It thus makes a large class of proteins newly accessible to detailed solution structure determination.

In the past two decades, nuclear-magnetic-resonance spectroscopy has become one of the two accepted methods (along with X-ray crystallography) for determining three-dimensional structures of proteins. NMR spectroscopy provides information about the structure and dynamic properties of proteins in solution, and offers an approach for determining the three-dimensional protein structures of systems that fail to crystallize. NMR currently provides about 15% of the protein structures in the Protein Data Bank and has a role in structural genomics[1], but complete automation of the structure determination process or the structural analysis of proteins with a molecular mass greater than 25 kDa have not yet become routine with NMR. Conventional uniform labelling of proteins with ^{13}C and ^{15}N, coupled with double and triple resonance, two-dimensional to four-dimensional NMR data collection, supports the determination of NMR solution structures of proteins, in favourable cases as large as 25 kDa (refs 2, 3). Cryogenic probes are used to improve the signal-to-noise ratio, and higher-field magnets provide increased resolution and a further gain in sensitivity. However, as molecular masses increase, NMR spectra become increasingly difficult to interpret because of spectral crowding and line broadening due to fast transverse relaxation. Spin-diffusion effects decrease the ability to determine inter-proton distances from nuclear Overhauser enhancement (NOE) data. A common approach for addressing these problems is to label proteins with deuterium to simplify the spectra and to minimize spin-diffusion effects[4,5]. Data collection by transverse relaxation optimized spectroscopy (TROSY)[6] provides sharper lines for amide and aromatic groups. With these methods the global folds of a limited number of proteins larger than 30 kDa could be derived on the basis of conformational restraints for amide and selected methyl groups[7]. However, conformational data from other aliphatic or aromatic groups and thus for most of the side chains remain difficult to collect for proteins larger than 25 kDa. Consequently, by June 2005 only 1% of all NMR protein structures deposited in the Protein Data Bank were for proteins with a molecular mass of more than 25 kDa, and backbone and side-chain chemical shift assignments more than 70% complete had been recorded in the BioMagResBank (http://www.bmrb.wisc.edu) for only eight proteins larger than 25 kDa.

It has long been recognized that deuteration can be used to simplify NMR spectra[8], to obviate the need for chiral assignments[9], to facilitate the measurement of spin–spin[9] and dipolar couplings[7], and to increase resolution[10]. However, labelling patterns and the approaches for achieving them have so far been suboptimal. Random fractional deuteration methods[9] suffer from the production of

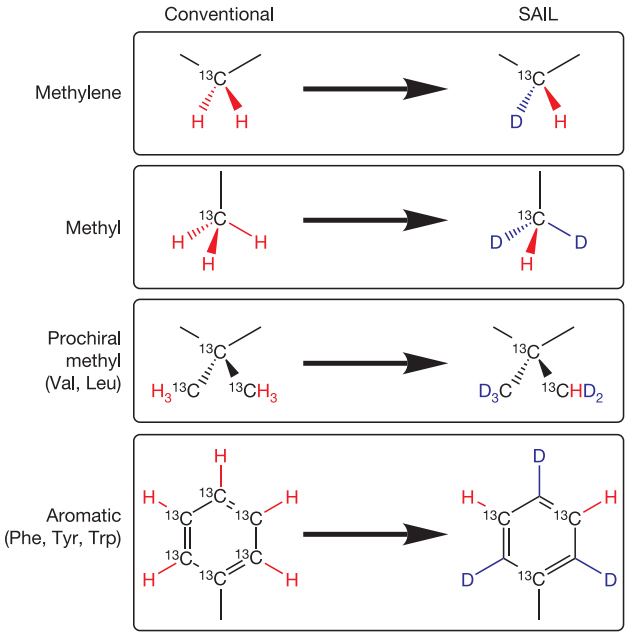


Figure 1 | SAIL amino acids. Design concepts embodied in the SAIL amino acids incorporated into CaM and MBP[13–15].

[1]CREST/JST and Graduate School of Science, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji, 192-0397, Japan. [2]Tatsuo Miyazawa Memorial Program, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, 230-0045, Japan.

Table 1 | Expected and observed features of CaM and MBP samples prepared by conventional uniform labelling and SAIL methods

Quantity	CaM		MBP	
	Uniform labelling	SAIL	Uniform labelling	SAIL
^{1}H atoms per molecule in $^{1}H_2O$	1,095	697 (64%)	2,860	1,802 (63%)
^{1}H atoms per molecule in $^{2}H_2O$*	851	453 (53%)	2,249	1,191 (53%)
Non-exchangeable side-chain ^{1}H atoms	692	305 (44%)	1,850	821 (44%)
NOE cross-peaks expected†	9,812	5,642 (58%)	17,076	9,382 (55%)
NOE cross-peaks used	–	4,576 (47%)	–	7,485 (44%)
NOE distance restraints expected‡	2,883	2,720 (94%)	4,293	4,347 (101%)
NOE distance restraints used	–	2,422 (84%)	–	3,818 (89%)

*The number was calculated by assuming that all exchangeable ^{1}H nuclei are replaced by ^{2}H.
†NOE cross-peaks are assumed to be observable between all aliphatic, aromatic, backbone amide and Asn/Gln side-chain amide ^{1}H nuclei that are less than 4.75 Å apart in the 20 conformers of the SAIL-CaM solution structure, or less than 4.2 Å apart in the 20 conformers of the SAIL-MBP solution structure.
‡Excluding duplicate restraints from symmetry-related peak pairs and NOEs for fixed distances. For uniform labelling, pseudo-atoms were assumed.

numerous isotopomer proteins with chemical shift heterogeneity and decreased signal intensities. Perdeuteration methods remove all carbon-bound protons such that much potential NOE information is lost. Methods for introducing methyl and/or aromatic protons into a perdeuterated background improve this situation, but the additional structural information is localized and unevenly distributed[4,5]. The same holds true for residue-selective[11] and segmental[12] labelling. As a consequence, the quality of three-dimensional protein structures based on these methods remains limited. As an alternative presented here, optimal labelling patterns for protein NMR can be realized by the chemical or enzymatic synthesis of amino acids[13–15] followed by *in vitro* (cell-free) protein expression[16] to build a protein exclusively from such amino acids. Cell-free methods[17,18] are crucial to making efficient use of the labelled amino acids and to the prevention of scrambling of the label, which would occur through metabolic pathways present in cell-based protein expression systems[11,19].

Stereo-array isotope labelling (SAIL)

The basic strategy of the SAIL approach is to prepare amino acids with the following features (Fig. 1): first, stereo-selective replacement of one ^{1}H in methylene groups 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, labelling of six-membered aromatic rings by alternating ^{12}C–^{2}H and ^{13}C–^{1}H moieties.

This labelling pattern preserves through-bond connectivity information needed for backbone and side chain assignments, eliminates the need for stereospecific assignments, simplifies measurements of couplings, and removes the most serious sources of spin diffusion so as to improve the accuracy of inter-proton distance measurements. Lines are sharpened both by decreasing long-range couplings and by eliminating dipolar relaxation pathways. The methyl and methylene labelling patterns simplify the analysis of the motional properties of the side chains from relaxation measurements[20]. The aromatic ring labelling strategy removes one-bond ^{13}C–^{13}C couplings, which often complicate spectra or require the use of constant time data collection methods so as to reduce spectral complexity[21].

We prepared the 20 protein-component SAIL amino acids (Supplementary Fig. 1) based on these design concepts by chemical and enzymatic syntheses[13–15]. Efficient incorporation of these amino acids into the protein of interest is achieved by an *Escherichia coli* cell-free protein synthesis system optimized for the preparation of labelled NMR samples[16]. We chose as our first targets for SAIL the 17-kDa calcium-binding protein calmodulin (CaM) from *Xenopus laevis* and the 41-kDa maltodextrin-binding protein (MBP) from *E. coli*. SAIL amino acid mixture (55 mg), prepared by following the protocol[16], yielded 5.5 mg of purified, soluble SAIL-CaM, or 5.3 mg of purified, soluble SAIL-MBP. For comparison, CaM and MBP uniformly labelled with ^{13}C and ^{15}N (UL-CaM and UL-MBP, respectively) were also prepared.

SAIL versus uniform labelling with ^{13}C and ^{15}N

On the basis of calculations (Table 1), we expected that the NMR spectra of SAIL proteins would be simpler than those of the corresponding uniformly labelled proteins. The numbers of observable protons are reduced by SAIL relative to uniform labelling. The SAIL approach reduces the number of non-exchangeable side-chain protons, which are prone to overlap but essential for defining the side-chain conformations, to less than half, and decreases the number of expected NOESY cross-peaks by 40–45%. Most of the additional NOEs from uniformly labelled proteins either involve fixed (geminal) distances or become redundant in the absence of stereospecific assignments and thus contribute to spectral overlap without furnishing independent information. In principle, stereospecific assignments could allow slightly more precise structures in the case of uniform labelling[22,23]. However, in practice only a limited number of stereospecific assignments can be made[22–24]. The BioMagResBank reports stereospecific assignments for 7% of the methylene protons and Val/Leu methyl groups, and for only 3.4% of the prochiral groups in proteins larger than 20 kDa. Therefore, in general the expected number of non-redundant, structurally relevant NOE restraints is retained with SAIL (Table 1), even if the effect of the better signal-to-noise ratio with SAIL is neglected. These numbers are corroborated by the experimental findings (Supplementary Table 1). More detailed theoretical considerations (see Methods), which take into account the enhanced signal strength and sharper lines in SAIL spectra and the fact that overlap can render peaks unidentifiable, show that in practice SAIL is expected to increase rather than decrease the number of identifiable NOE cross-peaks. The expected increase is moderate in regions without overlap but is significant in regions with strong overlap and therefore for larger proteins, for which SAIL is expected to yield two or more times the number of relevant conformational restraints than uniform labelling.

Precautions to be taken when collecting and analysing data for a SAIL protein are that deuterium decoupling should be applied during ^{13}C evolution times and that what normally are methyl and methylene groups should be treated as methine groups. ^{1}H–^{13}C CT-HSQC (constant-time heteronuclear single-quantum coherence) NMR data sets collected from SAIL and uniformly labelled proteins demonstrate the superiority of the SAIL method (Figs 2 and 3). Each ^{1}H–^{13}C pair in the protein is associated with a peak in the spectrum. Severe signal overlap in uniform labelling is alleviated in the spectrum from the SAIL proteins. The improvements, which are particularly apparent for larger proteins (Fig. 3), result from the decrease in the number of ^{1}H signals as well as from sharpening of the remaining signals.

The SAIL method also improves sensitivity. Part of the gain arises from longer ^{1}H and ^{13}C transverse relaxation times resulting from replacements of ^{1}H by ^{2}H. Reduced relaxation during magnetization transfer steps in experiments such as ^{1}H–^{13}C CT-HSQC leads to an increased signal-to-noise ratio. Reduced long-range couplings result in a further sharpening of signals. The signal intensities for methylene

groups are threefold to sevenfold higher with SAIL than with uniform labelling under the same conditions. Improvements are more pronounced for the 41-kDa SAIL-MBP protein (Fig. 3e) than for the 17-kDa SAIL-CaM (Fig. 2i). Although each observed SAIL methyl group contained only one ^{1}H in comparison with three equivalent ^{1}H with uniform labelling, equivalent signal intensities were observed (data not shown) as a result of the longer ^{1}H and ^{13}C transverse relaxation times.

SAIL protein structure determinations

All peaks benefited from improved sensitivity and resolution and allowed the signals of SAIL-CaM and SAIL-MBP to be assigned readily by established methods[25]. Side-chain assignments were determined completely from the analysis of two data sets: HCCH-TOCSY (total correlation spectroscopy) data provided connectivities among all side-chain signals, and HCCH-COSY (correlation spectroscopy) data were used to identify the spin systems. The detection and assignment of signals from aromatic rings containing alternating ^{12}C and ^{13}C nuclei (Fig. 1) is straightforward by an unconventional approach that we describe separately[26]. The expected ^{1}H, ^{13}C and ^{15}N chemical shifts could be assigned without exception for SAIL-CaM, and to 94% for SAIL-MBP, including more than 90% of the aliphatic and aromatic side-chain protons (Supplementary Table 1). Many of the shifts that could not be assigned are in the region of residues 229–241, which have been shown to interact with the bound cyclodextrin[27]. Because of conformational heterogeneity, most of the expected resonances of residues 229–241 could also not be assigned in earlier studies[7].

We obtained distance restraints for the structure calculation from three-dimensional ^{15}N- and ^{13}C-edited NOESY spectra[28] and from a two-dimensional NOESY spectrum for the aromatic region. These spectra were simplified by a decreased number of signals, as expected (Table 1). Because of lower spin diffusion, maximum NOE intensities for SAIL proteins were typically reached at mixing times 1.5-fold to 3.0-fold those for uniform labelling (data not shown). On the basis of the simplified and more quantitative NOESY spectra, we were able to obtain structures of SAIL-CaM and SAIL-MBP by means of the combined automated NOE assignment and structure calculation protocol in the program CYANA[29,30] (Supplementary Table 1). For instance, a dense network of NOEs including 949 non-redundant, long-range distance restraints was established for SAIL-MBP. Of the 3,818 non-redundant NOE distance restraints, 1,879 involve side-chain atoms beyond H^{β}. The SAIL-CaM and SAIL-MBP solution structures show good quality in terms of the agreement with the experimental data and other validation parameters.

Structures of the calcium-bound form of calmodulin have been determined previously by crystallography[31] and NMR. The NMR structure[32] was based on residual dipolar couplings (RDCs),

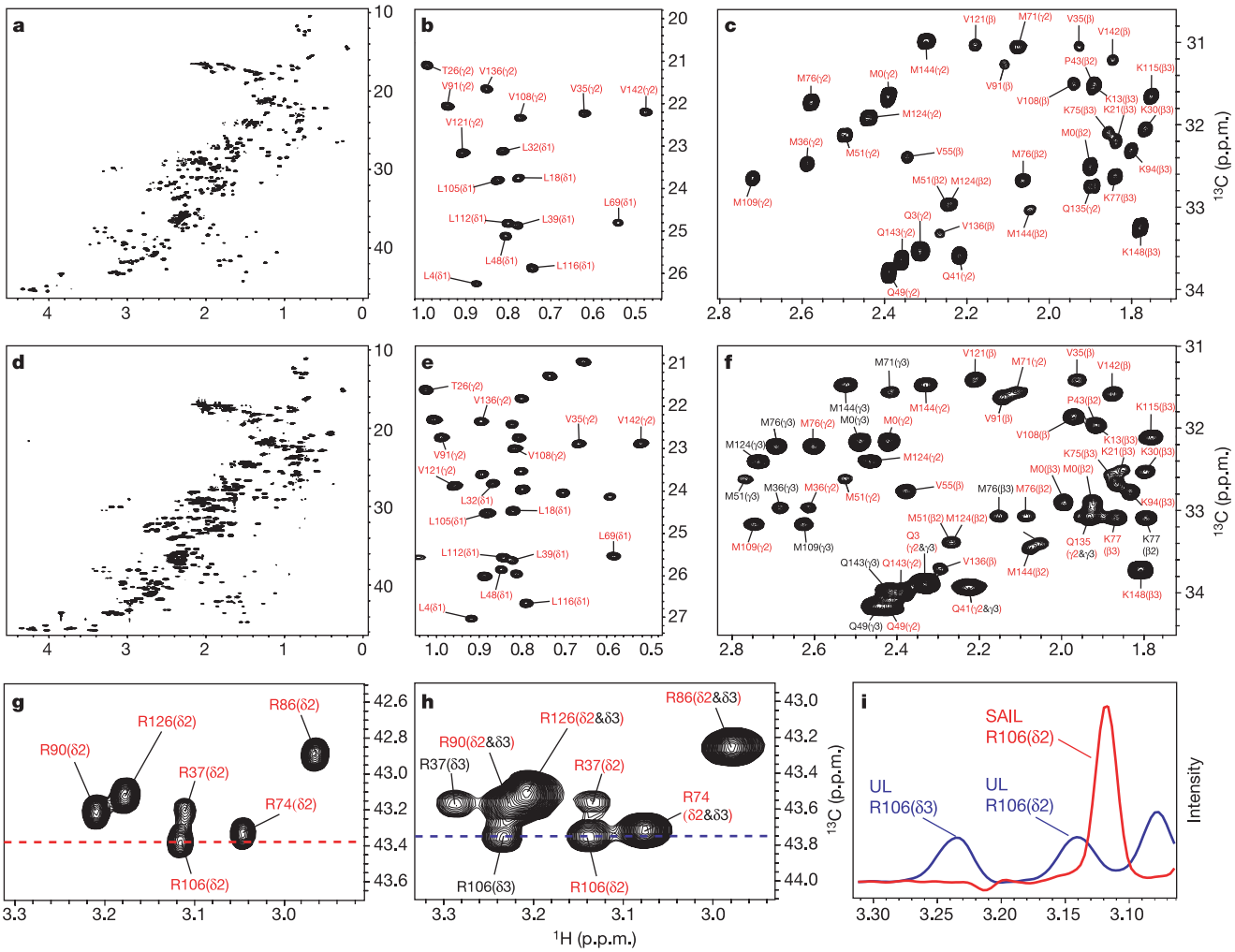


Figure 2 | ^{1}H–^{13}C CT-HSQC spectra of CaM. **a**, SAIL-CaM, aliphatic region. **b**, SAIL-CaM, methyl region. **c**, SAIL-CaM, methylene region. **d**, UL-CaM, aliphatic region. **e**, UL-CaM, methyl region. **f**, UL-CaM, methylene region. **g**, SAIL-CaM, Arg δ region. **h**, UL-CaM, Arg δ region. **i**, Cross-sections from **g** and **h**. The spectra for SAIL-CaM and UL-CaM were recorded under identical conditions and scaled for equal noise levels. Assignments are indicated by one-letter amino-acid code, residue number and atom identifier. Assignments of UL-CaM are as reported previously[37].

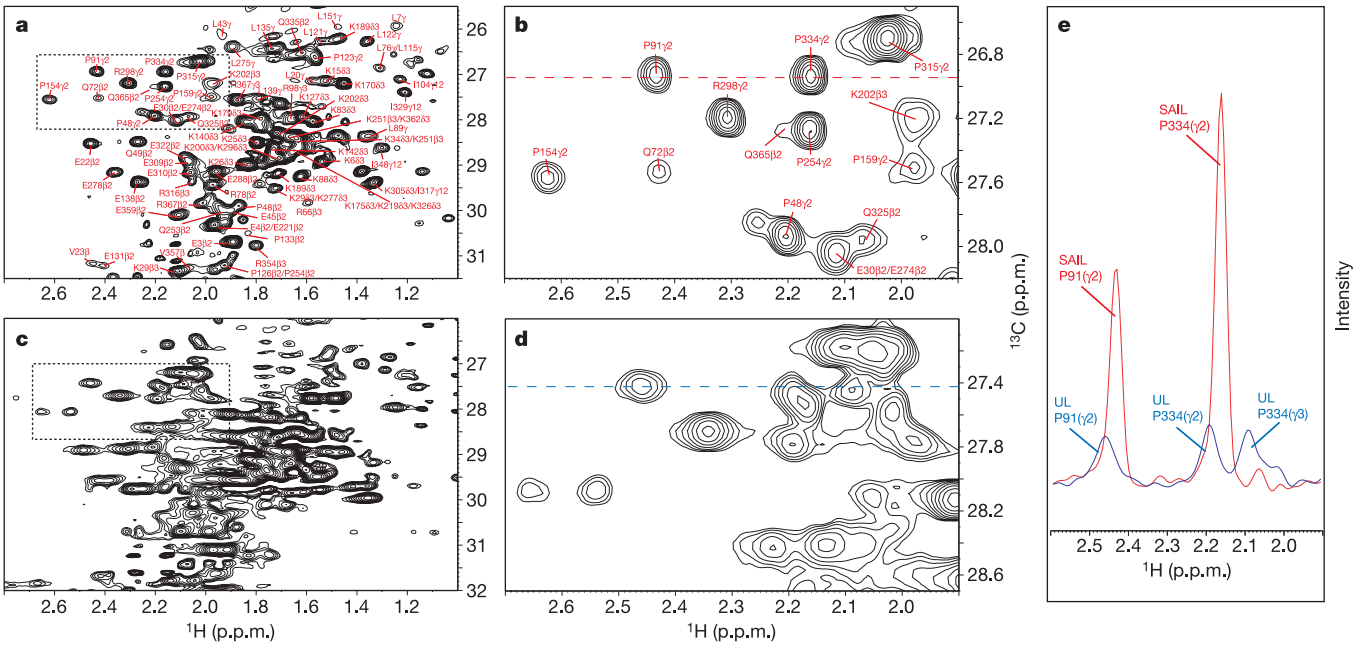


Figure 3 | ^{1}H–^{13}C CT-HSQC spectra of MBP. **a**, Aliphatic region of CDH groups in SAIL-MBP. **b**, Enlargement of the rectangular region marked in **a**. Assignments are indicated with one-letter amino-acid code, residue number and atom identifiers. **c**, **d**, Corresponding regions for UL-MBP. **e**, Cross-sections taken at the positions indicated in **b** and **d**. The spectra for SAIL-MBP and UL-MBP were recorded under identical conditions and scaled for equal noise levels.

measured mainly for the polypeptide backbone, and on the crystal structure. Our NMR structure, the first one for calcium-bound calmodulin based exclusively on NOEs, provides detailed information on the side-chain conformations. It is in close agreement with the crystal structure and the RDC-based backbone structure (Fig. 4a).

In addition, the solution structure of the 41-kDa SAIL-MBP coincides closely with the crystal structure[27] determined previously under slightly different conditions (Fig. 4b, c, and Supplementary Table 1). The 41-kDa SAIL-MBP solution structure is of similar precision and accuracy to those of smaller proteins, and the structural statistics are comparable to those commonly found in NMR structure determinations of smaller proteins. Previously[7], a global fold of MBP was determined by NMR on the basis of NOEs between amide and methyl protons, residual dipolar couplings for the polypeptide backbone, and hydrogen-bond restraints. That NMR study could provide a good determination of the global fold of the polypeptide backbone and the conformations of the methyl-containing side chains of valine, leucine and isoleucine, but the approach used cannot provide direct structural information on the other side chains. Their conformations therefore remained largely undetermined, resulting in root-mean-square deviations (RMSDs) of more than 3.8 Å for all side-chain heavy atoms of the amino-terminal and carboxy-terminal domains of MBP. The corresponding

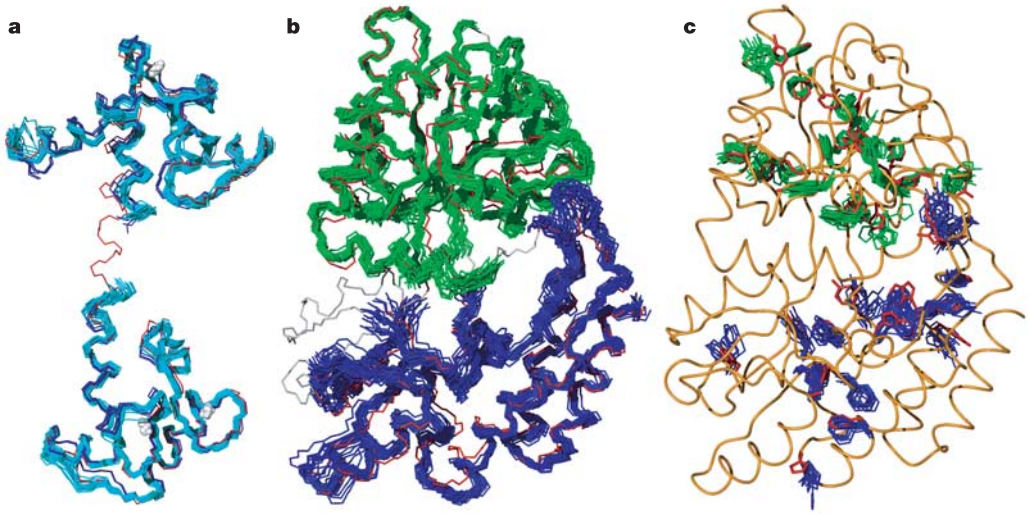


Figure 4 | CaM and MBP solution and crystal structures. **a**, SAIL-CaM (backbone in cyan, Ca^{2+} in white), CaM X-ray structure[31] (red), and three solution conformers of UL-CaM determined from residual dipolar coupling data[32] (blue). **b**, MBP solution and crystal structures[27]. Backbone of the N-terminal domain (SAIL-MBP in green, X-ray in red) and the C-terminal domain (SAIL-MBP in blue, X-ray in red). **c**, Aromatic side chains (SAIL-MBP in green and blue, X-ray in red), and backbone ribbon representation of the X-ray structure (gold). Superpositions of CaM and MBP solution conformers on the X-ray structures were performed separately for the two flexibly connected domains.

side-chain RMSDs for the SAIL-MBP structure are 2.3 Å for both domains.

Discussion

The SAIL strategy described here is expected to support high-throughput protein structure determination without loss of structural quality. The much smaller numbers of ^{1}H shifts that need to be assigned make SAIL proteins particularly amenable to automated resonance assignment, which also benefits from the complete absence of uncertainties associated with the lack of stereospecific assignments and frequent accidental chemical shift degeneracies for diastereotopic pairs in uniformly labelled proteins. SAIL improves the quality of virtually all commonly used multi-dimensional NMR spectra and can be used in conjunction with other techniques to improve the sensitivity of NMR experiments, such as TROSY[6], cryogenic probes and high-field magnets. Our structure determination of MBP shows that high-quality solution structures of proteins up to at least 40 kDa can now be solved by NMR.

METHODS

Expected numbers of identifiable NOESY peaks with SAIL and uniform labelling. We define a peak as identifiable if it is above the noise level and not overlapped with other peaks. Assuming N peaks, distributed randomly within a region of size $\boldsymbol{\Gamma}$ in a d-dimensional spectrum, the expected number of peaks that are not overlapped with other peaks is $N_1 = N(1-\gamma/\boldsymbol{\Gamma})^{N-1} \approx Ne^{-N\gamma/\boldsymbol{\Gamma}}$, where γ denotes the size of the peak region for each peak. SAIL reduces the number of ^{1}H nuclei relative to uniform labelling, and thus yields a smaller number, $\tilde{N}_0 < N$, of NOE signals with integral above the same threshold. Reduced relaxation gives rise to narrower line widths, $\Delta\tilde{\omega} < \Delta\omega$, greater peak heights, $\tilde{H} = H(\Delta\omega/\Delta\tilde{\omega})^d$, smaller peak areas, $\tilde{\gamma} = \gamma(\Delta\tilde{\omega}/\Delta\omega)^d$, and an increased number of peaks above the noise level, $\tilde{N} = \tilde{N}_0(\Delta\omega/\Delta\tilde{\omega})^{d/2}$. The latter follows from the relationship $N(H \geq H_{\min}) = N(I \geq I_{\min}) = N(r \leq r_{\max}) \propto r_{\max}^3 \propto I_{\min}^{-1/2} \propto \Delta\omega^{-d/2}$, where H, I and r are respectively the peak height, peak integral and distance for a given NOE, and $H_{\min}$, $I_{\min}$ and $r_{\max}$ are the noise level, corresponding intensity threshold and maximal NOE observation distance, respectively. The number of identifiable peaks for a SAIL protein, $\tilde{N}_1$, becomes, in terms of the corresponding number, N_1, for the uniformly labelled protein,

$$\tilde{N}_1 \approx \tilde{N}e^{-\tilde{N}\tilde{\gamma}/\boldsymbol{\Gamma}} = N_1\sqrt{\frac{\tilde{H}}{H}}\frac{\tilde{N}_0}{N}\exp\left[\left(1-\sqrt{\frac{H}{\tilde{H}}}\frac{\tilde{N}_0}{N}\right)\frac{N\gamma}{\boldsymbol{\Gamma}}\right]$$

For MBP with a SAIL-to-UL peak height ratio of $\tilde{H}/H \approx 4$–7 (Fig. 3e) and a ratio $\tilde{N}_0/N \approx 0.55$ (Table 1) for the number of NOEs below 4.2 Å, SAIL is expected to increase the number of identifiable NOEs by at least 10% for regions without overlap ($N\gamma/\boldsymbol{\Gamma} = 0$) and by more than 50% for a crowded region with $N\gamma/\boldsymbol{\Gamma} = 0.5$.

NMR spectroscopy. 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. The SAIL-MBP and UL-MBP samples each contained 0.33 mM protein, 3.3 mM β-cyclodextrin, 20 mM sodium phosphate, 3 mM NaN_3 and Complete Mini protease inhibitor mix (Roche), pH 7.2. NMR experiments were performed at 37 °C on a Bruker DRX800 spectrometer equipped with a TXI *xyz*-gradient probe. NMR spectra were analysed with the program SPARKY (University of San Francisco, California).

Structure calculation. The SAIL-CaM and SAIL-MBP structures were obtained with the program CYANA[29] with the use of automated NOE assignment[30] and torsion angle dynamics for the structure calculation[33], which was started from 100 (CaM) or 200 (MBP) conformers with random torsion angle values. The standard CYANA-simulated annealing schedule was applied with 10,000 (CaM) or 20,000 (MBP) torsion angle dynamics steps[29]. Backbone torsion angle restraints obtained from chemical shifts with the program TALOS[34] were added to the input for CYANA. Hydrogen-bond restraints were not used. The 20 conformers with the lowest final CYANA target function values were embedded in a water shell of 8 Å thickness and energy-minimized against the AMBER force field[35] with the program OPALp[36] in the presence of the NOE distance restraints as the only experimental data.

Received 8 September; accepted 14 December 2005.

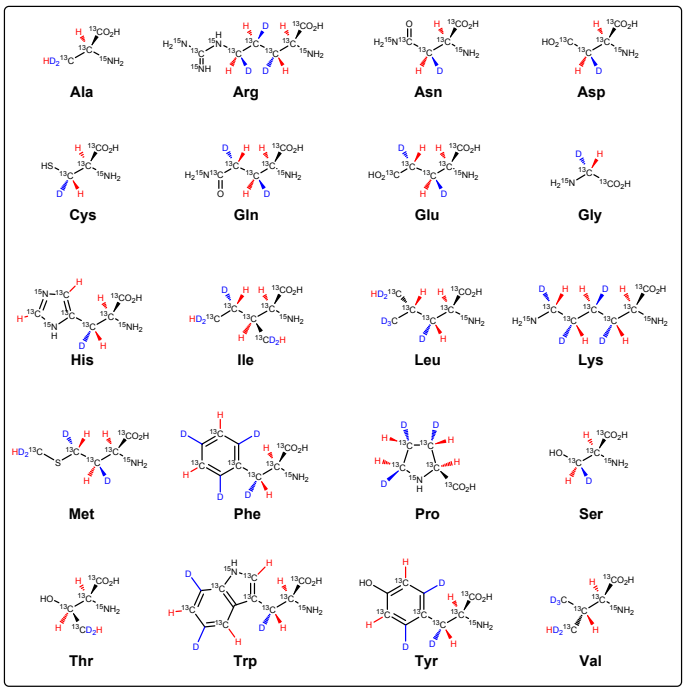
1. Kennedy, M. A., Montelione, G. T., Arrowsmith, C. H. & Markley, J. L. Role for NMR in structural genomics. *J. Struct. Funct. Genom.* **2,** 155–169 (2002).
2. Clore, G. M. & Gronenborn, A. M. Structures of larger proteins, protein–ligand and protein–DNA complexes by multidimensional heteronuclear NMR. *Prog. Biophys. Mol. Biol.* **62,** 153–184 (1994).
3. Clore, G. M. & Gronenborn, A. M. NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* **2,** 564–570 (1998).
4. Gardner, K. H. & Kay, L. E. The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 357–406 (1998).
5. Goto, N. K. & Kay, L. E. New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Opin. Struct. Biol.* **10,** 585–592 (2000).
6. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. 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 (1997).
7. Mueller, G. A. *et al.* Global folds of proteins with low densities of NOEs using residual dipolar couplings: Application to the 370-residue maltodextrin-binding protein. *J. Mol. Biol.* **300,** 197–212 (2000).
8. Markley, J. L., Putter, I. & Jardetzky, O. High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. *Science* **161,** 1249–1251 (1968).
9. LeMaster, D. M. Chiral β and random fractional deuteration for the determination of protein side-chain conformation by NMR. *FEBS Lett.* **223,** 191–196 (1987).
10. Farmer, B. T. II & Venters, R. A. NMR of perdeuterated large proteins. *Mod. Techniques Protein NMR* **16,** 75–120 (1999).
11. Arata, Y., Kato, K., Takahashi, H. & Shimada, I. Nuclear-magnetic-resonance study of antibodies—a multinuclear approach. *Methods Enzymol.* **239,** 440–464 (1994).
12. Yamazaki, T. *et al.* Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. Chem. Soc.* **120,** 5591–5592 (1998).
13. Oba, M., Kobayashi, M., Oikawa, F., Nishiyama, K. & Kainosho, M. Synthesis of ^{13}C/D doubly labelled L-leucines: Probes for conformational analysis of the leucine side-chain. *J. Org. Chem.* **66,** 5919–5922 (2001).
14. Oba, M., Terauchi, T., Miyakawa, A., Kamo, H. & Nishiyama, K. Stereoselective deuterium-labelling of diastereotopic methyl and methylene protons of L-leucine. *Tetrahedron Lett.* **39,** 1595–1598 (1998).
15. Oba, M., Terauchi, T., Miyakawa, A. & Nishiyama, K. Asymmetric synthesis of L-proline regio- and stereoselectively labelled with deuterium. *Tetrahedron Asymmetry* **10,** 937–945 (1999).
16. Torizawa, T., Shimizu, M., Taoka, M., Miyano, H. & Kainosho, M. Efficient production of isotopically labelled proteins by cell-free synthesis: A practical protocol. *J. Biomol. NMR* **30,** 311–325 (2004).
17. Kigawa, T., Muto, Y. & Yokoyama, S. Cell-free synthesis and amino acid-selective stable-isotope labeling of proteins for NMR analysis. *J. Biomol. NMR* **6,** 129–134 (1995).
18. Zubay, G. In-vitro synthesis of protein in microbial systems. *Annu. Rev. Genet.* **7,** 267–287 (1973).
19. McIntosh, L. P. & Dahlquist, F. W. Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear-magnetic-resonance spectra of proteins. *Q. Rev. Biophys.* **23,** 1–38 (1990).
20. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A. & Torchia, D. A. Pulse sequences for removal of the effects of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T_1 and T_2 values in proteins. *J. Magn. Reson.* **97,** 359–375 (1992).
21. Vuister, G. W. & Bax, A. Resolution enhancement and spectral editing of uniformly ^{13}C-enriched proteins by homonuclear broadband ^{13}C decoupling. *J. Magn. Reson.* **98,** 428–435 (1992).
22. Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. Automated stereospecific ^{1}H NMR assignments and their impact on the precision of protein structure determinations in solution. *J. Am. Chem. Soc.* **111,** 3997–4004 (1989).
23. Nilges, M., Clore, G. M. & Gronenborn, A. M. ^{1}H-NMR stereospecific assignments by conformational data-base searches. *Biopolymers* **29,** 813–822 (1990).
24. Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. 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 ^{13}C labeling. *Biochemistry* **28,** 7510–7516 (1989).
25. Cavanagh, J., Palmer, A. G., Fairbrother, W. & Skelton, N. *Protein NMR Spectroscopy: Principles and Practice* (Academic, San Diego, 1996).
26. Torizawa, T., Ono, M. A., 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).
27. Sharff, A. J., Rodseth, L. E. & Quiocho, F. A. Refined 1.8-Å structure reveals the mode of binding of β-cyclodextrin to the maltodextrin binding protein. *Biochemistry* **8,** 10553–10559 (1993).
28. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. 3-dimensional heteronuclear NMR of ^{15}N-labelled proteins. *J. Am. Chem. Soc.* **111,** 1515–1517 (1989).
29. Güntert, P. Automated NMR protein structure calculation. *Prog. NMR Spectrosc.* **43,** 105–125 (2003).

30. 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).
31. Chattopadhyaya, R., Meador, W. E., Means, A. R. & Quiocho, F. A. Calmodulin structure refined at 1.7 Å resolution. *J. Mol. Biol.* **228,** 1177–1192 (1992).
32. Chou, J. J., Li, S., Klee, C. B. & Bax, A. Solution structure of Ca^{2+}-calmodulin reveals flexible hand-like properties of its domains. *Nature Struct. Biol.* **8,** 990–997 (2001).
33. 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).
34. 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).
35. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117,** 5179–5197 (1995).
36. Koradi, R., Billeter, M. & Güntert, P. Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124,** 139–147 (2000).
37. Ikura, M. *et al.* Secondary structure and side-chain ^{1}H and ^{13}C resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR spectroscopy. *Biochemistry* **30,** 9216–9228 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Ikura and T. Yamazaki for providing the calmodulin and maltodextrin-binding protein genes, respectively, and M. Takeda for help with the preparation of figures. This work was supported by CREST/JST.

Author Information Atomic coordinates of the SAIL-CaM and SAIL-MBP structures have been deposited in the Protein Data Bank with accession codes 1X02 and 2D21, respectively. Chemical shifts have been deposited in the BioMagResBank with accession numbers 6541 and 6807. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.K. (kainosho@nmr.chem.metro-u.ac.jp).



Supplementary Figure 1 | SAIL amino acids. Chemical structures of the SAIL amino acids incorporated into calmodulin and the maltodextrin-binding protein MBP. Symbols: H (red), ^{1}H; D (blue), ^{2}H; C, ^{12}C.

Supplementary Table 1 | NMR structure statistics for SAIL-CaM and SAIL-MBP

Quantity	CaM	MBP
Completeness of chemical shift assignments, %*	100.0	94.0
Completeness of ^{1}H assignments of side-chain CH_n, %†	100.0	91.0 (90.4)
NOE upper distance bound restraints	2422	3818
Short/medium/long-range distance restraints	1233/758/431	2024/845/949
Distances restrained for complexed Ca^{2+}	24	—
Hydrogen bond restraints	0	0
Maximal distance restraint violation, Å	0.14	0.14
Distance restraint violations > 0.2 Å	0	0
AMBER energy, kcal/mol (mean + s.d.)	-7404±136	-14845±252
AMBER van der Waals energy, kcal/mol	-317±15	-1293±27
Ramachandran plot statistics, %‡	89/11/0/0	86/14/0/0
Backbone RMSD, Å§	0.52/0.51	0.68/0.80
All heavy atom RMSD, Å§	0.81/0.88	1.02/1.14
Backbone RMSD to X-ray structure[31,27], Å§	1.32/0.72	1.33/1.22
All heavy atom RMSD to X-ray structure, Å§	1.98/1.40	1.88/1.84

*Percentage of the ^{1}H, ^{13}C and ^{15}N chemical shifts of all aliphatic, aromatic, backbone amide, Asn/Gln/Trp side-chain amide nuclei that are assigned.

†Percentage of the ^{1}H chemical shifts of all side-chain CH_n groups that are assigned. For MBP, the percentage obtained if the Val, Leu, and Ile $C^{\delta1}$ methyl groups are excluded is given in parenthesis.

‡Percentage of residues in the most favoured, additionally allowed, generously allowed, and disallowed regions, respectively [Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK -NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486 (1996)].

§Calculated to the mean coordinates of the N-terminal domain (residues 5–75 in CaM; 5–112 and 260–328 in MBP) and the C-terminal domain (residues 82–146 in CaM; 124–227, 246–257 and 332-370 in MBP), respectively.