Selective isotope labeling strategy and computational interpretation of spectra for protein NMR analyses

To cite this article: Takuma Kasai and Takanori Kigawa 2018 J. Phys.: Conf. Ser. 1036 012007

View the article online for updates and enhancements.

Related content

- Application of laser fluorimetry for determining the influence of a single amino-acid substitution on the individual photophysical parameters of a fluorescent form of a fluorescent protein mRFP1
  A A Banishev, E P Vrzheshch and E A Shirshin

- Nanoparticle Doped In-Cell Retarder for Low Operating Voltage in Transflective Liquid Crystal Displays
  Gak Seok Lee, Jeong Hyun Lee, Jae Chang Kim et al.

- Is $^{40}\text{Ca}$ A Stable Isotope?
  Ching Cheng-rui, Ho Tso-hsiu and Zhao Wei-qin
Selective isotope labeling strategy and computational interpretation of spectra for protein NMR analyses

Takuma Kasai\textsuperscript{1,2,4} and Takanori Kigawa\textsuperscript{1,3,4}

\textsuperscript{1} RIKEN Quantitative Biology Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
\textsuperscript{2} PRESTO, JST, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan
\textsuperscript{3} School of Computing, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8503, Japan

E-mails: takuma.kasai@riken.jp, kigawa@riken.jp

Abstract. Signal assignment is a mandatory step for the site-specific analyses of proteins by NMR. In challenging cases, such as low solubility target proteins or in-cell NMR, amino-acid selective stable isotope labeling facilitates the main-chain assignment, since it can determine the amino-acid type of each signal, thus providing complementary information to that obtained by standard three-dimensional triple resonance experiments. Inspired by information science methodologies, we developed a combinatorial selective labeling strategy, named stable isotope encoding (SiCode), to enable amino-acid typing even from noisy NMR spectra, with small numbers of labeled samples. The high noise-resistance of SiCode is achieved by both optimization of the isotope labeling pattern according to the information distance between amino acids and model-based computational retrieval of the amino-acid information from spectra, using replica exchange Monte Carlo computation. We demonstrate amino-acid typing with simulated low signal-to-noise-ratio spectra, with sample concentrations as low as micromolar order. Since the main-chain signal assignment is often a bottleneck process for various amide-signal-based NMR analyses of challenging target proteins, SiCode will facilitate and accelerate such analyses.

1. Introduction

Nuclear magnetic resonance (NMR) is a common method for analyses of protein structures, dynamics, and interactions at atomic resolution. The protein sample is usually stable-isotope labeled, to obtain rich information about the targets using the NMR-detectable nuclei, \(^1\text{H}\), \(^{13}\text{C}\), and \(^{15}\text{N}\), and to perform various measurements that rely on stable isotope enrichment. Two-dimensional \(^1\text{H}^{15}\text{N}\) spectroscopy is one of the basic methods for protein analyses. It uses amide moieties as protons to obtain residue-specific information about the target protein. For example, it includes chemical shift perturbation, transferred cross saturation, \(^{15}\text{N}\) relaxation, and paramagnetic effects. To perform such analyses, the assignment of the amide signals is required beforehand. The sequential assignment technique with triple resonance experiments is the standard procedure to perform the assignment \cite{1, 2}. However, under challenging conditions, this is difficult to accomplish because some of the triple resonance experiments are insensitive. For example, an insufficient protein concentration caused by low

\textsuperscript{4} To whom any correspondence should be addressed.
solubility or productivity of the protein leads to a low signal-to-noise ratio (SNR) of the spectra. The SNR is further reduced for large proteins, aggregated proteins, and/or proteins in crowded conditions, such as cellular environments, due to the relaxation enhancement of the magnetizations.

Amino-acid selective stable isotope labeling provides useful information for assignments even under such challenging conditions [3], since it enables amino-acid typing of each amide signal. For single selective labeling, the amino-acid typing of the residue can be performed by acquiring the standard $^1$H-$^{15}$N correlation spectrum, such as $^{15}$N-HSQC, with 19 labeled samples corresponding to all amino acids except for proline. For dual selective labeling, each sample is both $^{15}$N-labeled with a specific amino acid and $^{13}$C-labeled with another specific amino acid [4]. With the two-dimensional HN(CO) measurement, which is the most sensitive method among the triple resonance experiments, one can perform amino-acid typing of the preceding residue (residue i-1) as well as the residue including the amide (residue i), since the HN(CO) signal is only observed if both the carbonyl carbon at residue i-1 and the amide nitrogen at residue i are isotopically labeled. This greatly narrows down the assignment possibilities. For instance, if an amino-acid pair that appears only once in the amino acid sequence is labeled, then only a single peak is observed, and thus the corresponding amino-acid residue is assigned without further NMR experiments [5]. However, for the amino-acid typing of all of the signals, dual selective labeling requires tens or hundreds of labeled samples, corresponding to the number of amino acid pairs appearing in the sequence.

To reduce the required number of labeled samples, various combinatorial selective labeling methods have been developed [6], in which each amino acid corresponds to a combination of labeled samples, rather than a single sample. The methods were designed for various objectives, including the discrimination of 16 amino acids with only 5 samples [7]; the discrimination of amino acids that have close chemical shifts at the alpha and beta carbons [8, 9]; the design of a labeling pattern specialized for a specific amino-acid sequence to achieve maximal unique assignments [10-14]; the use of the alpha carbon as well as the carbonyl carbon to discriminate amino acids with various combinations of triple resonance experiments [15, 16]; the avoidance of signal overlap considering amino-acid frequency [17-19]; the labeling of a specific amino-acid set for the assignment of the transmembrane region of membrane proteins [16, 20, 21]; and the selective inverse labeling, or selective unlabeling, which is compatible with various protein production methods, including in vivo expression [9, 22].

In this paper, we describe a selective labeling strategy, facilitated by information science and computational analysis techniques, which enables amino-acid typing with a small number of labeled samples or with low-SNR spectra.

2. Stable isotope encoding (SiCode)

We developed a selective labeling strategy, named stable isotope encoding (SiCode), based on the unique point-of-view that the amino-acid information is encoded in labeling patterns and decoded from observed NMR spectra, by analogy to digital communications (figure 1a) [23]. To reduce the required number of labeled samples based on this point-of-view, the information content (entropy) per labeled sample should be increased. Therefore we introduced fractional stable isotope labeling to achieve ternary labeling levels; namely, 50%, 75%, or 100% for $^{15}$N and 0%, 50%, or 100% for $^{13}$C, while the other combinatorial labeling methods used binary labeling levels. An example of the labeling pattern is shown in figure 1b. Using this labeling scheme, we successfully discriminated all 19 non-proline amino acids with as few as 3 labeled samples [23].

The main-chain assignment of easy targets, such as small, highly-soluble, globular, and monodispersed proteins, can be achieved solely with the sequential assignment method. As mentioned, various selective labeling methods are used when the sequential assignment is difficult to accomplish. Therefore, selective labeling methods are needed to deal with the low-SNR spectra of challenging targets. To achieve the high noise-tolerance of SiCode, we improved both the encoding and decoding processes, thanks to SiCode’s inherent affinity to information science methodologies.
Figure 1. Stable isotope encoding (SiCode). (a) The basic concept of SiCode. In digital communication, a sender who wishes to send a message, in this case the letter “A”, converts the message to binary digits with a predefined codeword table. The binary digits are transmitted to a receiver, who converts them back to a letter using the identical codeword table. Amino-acid selective isotope labeling can be regarded as a similar process. In this case, the message is an amino acid type, and the codeword is an isotope labeling pattern. A receiver can retrieve the amino-acid information from the acquired NMR spectra. (b) An example of the labeling pattern of SiCode, which enables the discrimination of 19 amino acids with only 3 samples. The -C and -N labeling ratios are shown as percentages. The ternary digits that correspond to each amino acid are shown in the right column. Figures are reprinted from ref. [23].

For encoding, we introduced an optimization of the isotope labeling pattern [23]. In general, a larger information distance between two amino acids reduces misjudgements between the amino acids under noisy circumstances. Thus, according to the coding theory, we defined the minimal amino acid distance among all combinations of amino acids. The minimal distance should be maximized to obtain a “good” labeling pattern, which enables accurate amino acid discrimination even with noisy NMR spectra. We used the Euclidian distance as the information distance, because the thermal Gaussian noise is dominant for low-SNR NMR spectra [23]. We achieved the global optimization with the simulated annealing computation for the given number of amino-acids and the given number of labeled samples [23]. For example, figure 2a shows the optimized labeling pattern to discriminate all 19 non-proline amino acids with 3 labeled samples. This pattern is identical to the pattern shown in figure 1b, indicating that the pattern using the ternary three digits was the best for 19 amino acids. However, as shown in figure 2b, depending on the combination of the number of amino acids and the number of labeled samples, we can see various fractional labeling levels, such as 79.0% or 50.4%, in the optimized labeling patterns. To achieve such complicated labeling patterns, we utilized a cell-free protein synthesis system with suppressed amino-acid metabolism [24-26].

For decoding, we introduced a model-based analysis to obtain information from the low-SNR spectra [27]. Previously, we analyzed spectra based on the peak intensity; namely, at first we obtain the peak intensities from the spectra, then back-calculate the labeling ratio, and finally convert it to the amino-acid information by finding the nearest one in the labeling pattern table [23]. Obviously, this approach depends on the accuracy of the peak intensity, which may be reduced when the SNR of the spectra is low and/or the signal overlaps with another signal. In contrast, the model-based approach does not begin with obtaining the peak intensity. In this approach, we first make a list of all possible models; i.e., amino acid pairs representing residues i-1 and i, from the known amino-acid sequence of the target protein. We can calculate the intensity pattern for each model, using the known labeling pattern. Subsequently, to obtain the amino-acid information, we fit the intensity patterns to the
observed spectra and select the best-fitting model with the minimal residual error. As we use the model function with explanatory variables, including the continuous variables (signal amplitude, chemical shifts, and peak widths) and the discrete variable (the model), we employed the replica exchange Monte Carlo (REMC) computation to achieve the global optimization of the variables in a high-dimensional parameter space [27]. This approach can be used for simultaneous peak deconvolution and analysis, by fitting the spectra to the sum of multiple model functions [27].

Figure 2. Optimized labeling pattern for (a) 19 amino acids (AAs) with 3 labeled samples and (b) 18 amino acids with 4 labeled samples. The $^{13}C$ and $^{15}N$ labeling ratios are shown as percentages. Amino acids are shown as numbers, because there are numerous possibilities for the assignment of amino acids to codewords, although there are some technical considerations [23]. The amino acid assignments for the Ubiquitin mutant protein used in this work are shown in the second column of (b).

3. Evaluation with experimental data and low-SNR simulation data

3.1. High-SNR experimental data
At first, we tried amino-acid typing with the high-SNR spectra of 1 mM Ubiquitin mutant protein. We adopted the labeling pattern for 18 amino acids with the 4 samples shown in figure 2b, because the protein lacks cysteine and tryptophan residues. Ubiquitin mutant protein is 84 amino-acids long and contains 3 proline residues. Four main-chain signals (S-5 in the N-terminal linker region, F4, E24, and G53) were not observed under our experimental conditions. In total, 76 main-chain signals were observed. We picked 94 signals from the HSQC spectrum, which included 75 main-chain amide signals and 19 side-chain amide signals (arginine, asparagine, and glutamine). One of the picked main-chain signal was actually two overlapped signals, G10 and T22. Excluding this overlapped signal, correct amino-acid typing was achieved for all of the other 74 signals at both residues i-1 and i (table 1). Assuming that we could determine the overlapping by other experiments, such as three-dimensional triple resonance experiments, we successfully deconvoluted and correctly obtained the amino-acid information (table 1).

3.2. Low-SNR simulation data
To simulate low-SNR data, such as those obtained from low-solubility proteins, we added artificial noise to the experimental data. The noise amplitude was adjusted so that the final noise level in the spectra was 100-fold (for simulation 1) or 200-fold (for simulation 2), as compared to the original
experimental spectra. Therefore, these data sets emulate measurements with 10 µM and 5 µM protein solutions for simulations 1 and 2, respectively. For simulation 1, the accuracy of the amino-acid typing was the same as that obtained with the experimental data (table 1), suggesting that SiCode can deal with proteins with solubilities as low as 10 µM with sufficient reliability. For simulation 2, the accuracy dropped down close to 90%, regardless of the overlapped signal exclusion (table 1 and figure 3).

Table 1. Accuracy of amino-acid typing for the main-chain signals of Ubiquitin mutant protein.

| Signals                      | Criteria of correctness | Experimental (1 mM) | Simulation 1 (10 µM equivalent) | Simulation 2 (5 µM equivalent) |
|------------------------------|--------------------------|---------------------|---------------------------------|-------------------------------|
| Excluding overlap           | residues i-1, i         | 74 (100%)           | 74 (100%)                       | 69 (93%)                      |
|                             | residue i                | 74 (100%)           | 74 (100%)                       | 70 (95%)                      |
| Including overlap           | residues i-1, i         | 76 (100%)           | 76 (100%)                       | 69 (91%)                      |
|                             | residue i                | 76 (100%)           | 76 (100%)                       | 70 (92%)                      |

- The numbers of correct signals are shown. The percentages of correct signals are shown in parentheses.
- Target signals are 74 observed main-chain signals, excluding the overlapped G10 and T22.
- Target signals are all of the 76 observed main-chain signals.
- Signals with correct amino-acid types at both residues i-1 and i are counted.
- Signals with the correct amino-acid type at residue i are counted.

Figure 3. Amino acid typing with “simulation 2” data. The results are shown on the $^{15}$N-HSQC spectra of sample 1. Letters inside and outside the parentheses indicate the decoded amino acid information of residues i-1 and i, respectively. Blue letters indicate correct typing at both residues i-1 and i. Magenta letters indicate correct at residue i, but incorrect at residue i-1. Red letters indicate incorrect at residue i. The red asterisk denotes the overlapped signals of G10 and T22.
Figure 4 shows some examples of model fitting. Since the (N)I61 signal (the amide signal of the residue I61, which is preceded by N60) from the experimental 1 mM data has a sufficiently high SNR (figure 4a), the correct decoding is easy. In figure 4b, the same signal from “simulation 2” data (equivalent to 5 µM) is obscured by noise. In this case, we are unable to obtain the precise signal intensity from each spectrum. However, we successfully obtained the correct amino-acid information by the model-based fitting, thanks to the optimized labeling pattern to secure sufficient information distances between amino acids. The high-SNR (D)Y59 signal in the experimental spectra (figure 4c) is similarly buried in the noise in the “simulation 2” data (figure 4d). In this case, unfortunately, the best fit model to explain the spectra is (D)T, rather than the correct (D)Y. At the sampling steps of the lowest temperature replica of REMC, 45.4% (D)T, 37.2 % (D)Y, and 17.4% (D)G were sampled, showing that the correct model was the second likeliest candidate. For “simulation 2”, the decoding was not perfect but still maintained high accuracy (over 90%). The sampled frequency of the model candidates also includes information that is helpful for the assignments to some extent, depending on the purpose of using SiCode.

![Figure 4](image_url)

**Figure 4.** Spectral fitting for model-based decoding. Cross sections of the I61 (a and b) and Y59 (c and d) signals from the experimental spectra (a and c) and the “simulation 2” spectra (b and d) are shown. Black and red lines are observed spectra and fitted functions, respectively.

4. Conclusion

In this report, we described a selective stable isotope labeling strategy, SiCode, which is based on the concept that the selective labeling can be regarded as “encoding and decoding” processes. Improvements in both the encoding and decoding processes, inspired by information science methodologies, increased the robustness of amino-acid discrimination against noise. We demonstrated the noise tolerance of SiCode with the simulated low-SNR spectra corresponding to that of a low-solubility (5 to 10 µM) target protein. The total NMR measurement time for this demonstration was 56 hours (2.3 days). As we expect that N-fold lower concentration samples with N²-fold longer experiment times will result in the same SNR, amino acid typing of even a several micromolar protein solution can be accomplished in a reasonable amount of time. Increasing the sample number above 4 may further improve the accuracy of amino acid discrimination, even with a fixed total NMR measurement time [23]. Therefore, we conclude that SiCode will facilitate the NMR analyses of low-solubility proteins.

In addition to low-solubility proteins, in-cell NMR is another applicable target of SiCode. In in-cell NMR, the SNR is much worse than that in in vitro NMR, due to the insufficient cellular protein concentration, the relaxation enhancement caused by macromolecular crowding, the heterogeneity of the sample, and the limited measurement time to ensure cell viability. These are the reasons why the successful reports of main-chain signal assignment solely with in-cell NMR spectra have been limited.
to a few small proteins so far [28-31]. The sequential assignment method using three-dimensional triple resonance measurements offers the information of connectivity to the neighboring residue and of the amino-acid type. However by the relaxation enhancement, the connectivity information is often lost due to insensitivity of some spectra. The information of amino-acid type is not complete because it relies on the chemical shift values of $\alpha$ and $\beta$ carbons, which are degenerated in some amino acids. In contrast, the amino acid information of SiCode is derived not from the chemical shift values but from the isotope labeling pattern. In consequence, the information from the three-dimensional triple resonance measurements and SiCode are complementary. Therefore we are sure that SiCode will also facilitate analyses of in-cell NMR spectra by the amino-acid type information.

In contrast to the other combinatorial selective labeling methods, SiCode spectra require computational interpretation, because the amino acid information represented by the signal intensities is difficult to analyze visually. Such computation actually enabled the interpretation of low-SNR spectra, by accumulating information about coincidences between the observed intensities and the model intensities among all of the samples. With various computational methodologies, SiCode will expand the applicable targets and situations of protein NMR analyses.

5. Materials and methods

The amino-acid sequence of the human Ubiquitin mutant protein used in this work is the same as residues 1 to 76 of UniProt P62979, except for three mutations (L8A, I44A, and V70A) and additional residues (GSSGSSG at the N-terminus and D at the C-terminus). The residue numbers of the N-terminal additional residues are indicated with the non-positive numbers in the main text. Four samples were prepared according to the pattern shown in figure 2b, with the cell-free protein synthesis system [23-26]. To achieve the desired labeling ratios, we used mixtures of different isotope-enriched amino acids. For example, amino acid I (isoleucine) of sample 1 showed on figure 2b should be 79.0% $^{13}$C- and 89.5% $^{15}$N-labeled. In this case we used a mixture of $^{13}$C/$^{15}$N-enriched isoleucine, $^{15}$N-enriched isoleucine, and naturally abundant isoleucine in the ratio of 79.0%, 10.5%, and 10.5%. In the cell-free protein synthesis reaction, these three types of isoleucine are randomly incorporated to the protein, resulting in the desired isotopically labeled sample. The synthesized proteins were purified by chromatography on a HisTrap affinity column (GE Healthcare) and a HiTrap SP cation exchange column (GE Healthcare). The purified proteins were concentrated to 1 mM solutions in NMR buffer (50 mM sodium phosphate buffer, pH 6.0, 10% $^2$H$_2$O) with VIVASPIN 2 MWCO 5,000 concentrators (Sartorius). Two-dimensional $^{15}$N-HSQC and HN(CO) spectra were measured with a 700 MHz Avance III spectrometer (Bruker) equipped with a TCI cryoprobe (Bruker) at 298 K. The spectral width and the time domain data points of the $^{15}$N dimension were 33.0 ppm and 128 complex points, respectively, for both experiments. The numbers of scans were 8 and 32 for $^{15}$N-HSQC and HN(CO), respectively. To simulate low-SNR spectra, random Gaussian white noise was added to each data point in the raw time-domain data. The noise level of the data was estimated from regions without signals in the processed frequency-domain spectra [27]. The amino-acid information was decoded from the spectra, as described previously [23, 27]. For evaluation purposes, the main-chain assignment was independently performed with the standard triple resonance experiments, using a uniformly $^{13}$C/$^{15}$N-labeled sample.

Acknowledgements

We thank the lab members at RIKEN for assistance with sample preparation and for valuable comments. We also thank Dr. Kenji Nagata, Prof. Masato Okada, Prof. Toshiyuki Tanaka, and Prof. Shiro Ikeda for helpful discussions. This work was financially supported in part by JSPS KAKENHI (16K18508, 26650027, 25120003), JST PRESTO (JPMJPR1674), and JST CREST (JPMJCR13M3).

References

[1] Ikura M Kay L E and Bax A 1990 Biochemistry 29 4659–67
[2] Kay L E, Ikura M, Tschudin R and Bax A 1990 J. Magn. Reson. 89 496–514
[3] Hembram D S, Haremaki T, Hamatsu J, Inoue J, Kamoshida H, Ikeya T, Mishima M, Mikawa T, Hayashi N, Shirakawa M and Ito Y 2013 *Biochem. Biophys. Res. Commun.* **438** 653–9

[4] Kainosho M and Tsuji T 1982 *Biochemistry* **21** 6273–9

[5] Yabuki T, Kigawa T, Dohmae N, Takio K, Terada T, Ito Y, Laue E D, Cooper J A, Kainosho M and Yokoyama S 1998 *J. Biomol. NMR* **11** 295–306

[6] Jaipuria G, Krishnarjuna B, Mondal S, Dubey A and Atreya H S 2012 *Adv. Exp. Med. Biol.* **992** 95–118

[7] Parker M J, Aulton-Jones M, Hounslow A M and Craven C J 2004 *J. Am. Chem. Soc.* **126** 5020–1

[8] Shi J, Pelton J G, Cho H S and Wemmer D E 2004 *J. Biomol. NMR* **28** 235–47

[9] Kainosho M and Tsuji T 1982 *Biochemistry* **21** 6273–9

[10] Yabuki T, Kigawa T, Dohmae N, Takio K, Terada T, Ito Y, Laue E D, Cooper J A, Kainosho M and Yokoyama S 1998 *J. Biomol. NMR* **11** 295–306

[11] Hefke F, Bagaria A, Reckel S, Ullrich S J, Dötsch V, Glaubitz C and Güntert P 2011 *J. Biomol. NMR* **49** 75–84

[12] Maslennikov I and Choe S 2013 *Curr. Opin. Struct. Biol.* **23** 555–62

[13] Staunton D, Schlinskert R, Zanetti G, Colebrook S A and Campbell I D 2006 *Magn. Reson. Chem.* **44** S2–9

[14] Xun Y, Tremouilhilac P, Carraher C, Helhaus C, Ozawa K, Otting G, Dixon N E, Leippe M, Grotzinger J, Dingley A J and Kralicek A V 2009 *Protein Expr. Purif.* **68** 22–7

[15] Hein C, Löhr F, Schwarz D and Dötsch V 2017 *Biopolymers* **107** e23013

[16] Löhback F, Reckel S, Karbyshev M, Connolly P J, Abdul-Manan N, Bernhard F, Moore J M and Dötsch V 2012 *J. Biomol. NMR* **52** 197–210

[17] Sweredsø M J, Donovan K J, Nguyen B D, Shaka A J and Baldi P 2007 *Bioinformatics* **23** 2829–35

[18] Ozawa K, Wu P S, Dixon N E and Otting G 2006 *FEBS J.* **273** 4154–9

[19] Wu P S, Ozawa K, Jergic S, Su X C, Dixon N E and Otting G 2006 *J. Biomol. NMR* **34** 13–21

[20] Sobhanifar S, Reckel S, Junge F, Schwarz D, Kai L, Karbyshev M, Löhr F, Bernhard F and Dötsch V 2010 *J. Biomol. NMR* **46** 33–43

[21] Trbović N, Klammt C, Koglin A, Löhr F, Bernhard F and Dötsch V 2005 *J. Am. Chem. Soc.* **127** 13504–5

[22] Hiroaki H, Umetsu Y, Nabeshima Y, Hoshi M and Kohda D 2011 *J. Struct. Funct. Genomics* **12** 167–74

[23] Kasai T, Koshiba S, Yokoyama J and Kigawa T 2015 *J. Biomol. NMR* **63** 213–21

[24] Kigawa T, Yabuki T, Yoshida Y, Tsutsui M, Ito Y, Shibata T and Yokoyama S 1999 *FEBS Lett.* **442** 15–9

[25] Kigawa T, Yabuki T, Matsuda N, Matsuda T, Nakajima R, Tanaka A and Yokoyama S 2004 *J. Struct. Funct. Genomics* **5** 63–8

[26] Yokoyama J, Matsuda T, Koshiba S, Tochio N and Kigawa T 2011 *Anal. Biochem.* **411** 223–9

[27] Kasai T, Nagata K, Okada M and Kigawa T 2016 *J. Phys. Conf. Ser.* **699** 012003

[28] Reardon P N and Spicer L D 2005 *J. Am. Chem. Soc.* **127** 10848–9

[29] Säkkakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Wächchli M, Smith B O, Shirakawa M, Güntert P and Ito Y 2009 *Nature* **458** 102–5

[30] Hamatsu J, O’Donovan D, Tanaka T, Shirai T, Hourai Y, Mikawa T, Ikeya T, Mishima M, Boucher W, Smith B O, Laue E D, Shirakawa M and Ito Y 2013 *J. Am. Chem. Soc.* **135** 1688–91

[31] Ikeya T, Hanashima T, Hosoya S, Shimazaki M, Ikeda S, Mishima M, Güntert P and Ito Y 2016 *Sci. Rep.* **6** 38312
