membranes

MDPI

Communication

ssPINE: Probabilistic Algorithm for Automated Chemical Shift Assignment of Solid-State NMR Data from Complex Protein Systems

Adilakshmi Dwarasala [1], Mehdi Rahimi [1], John L. Markley [1,2] and Woonghee Lee [1,*]

[1] Department of Chemistry, University of Colorado Denver, Denver, CO 80217, USA
[2] Department of Biochemistry, University of Wisconsin-Madison, WI 53706, USA
* Correspondence: woonghee.lee@ucdenver.edu

Abstract: The heightened dipolar interactions in solids render solid-state NMR (ssNMR) spectra more difficult to interpret than solution NMR spectra. On the other hand, ssNMR does not suffer from severe molecular weight limitations of solution NMR. In recent years, ssNMR has undergone rapid technological developments that have enabled structure-function studies of increasingly larger biomolecules including membrane proteins. Current methodology includes stable isotope labeling schemes, non-uniform sampling with spectral reconstruction, faster magic angle spinning, and innovative pulse sequences that capture different types of interactions among spins. However, computational tools for the analysis of complex ssNMR data from of membrane proteins and other challenging protein systems have lagged behind those for solution NMR. Before a structure can be determined, thousands of signals from individual types of multidimensional ssNMR spectra of samples, which may have differing isotopic composition, must be recognized, correlated, categorized, and eventually assigned to atoms in the chemical structure. To address these tedious steps, we have developed an automated algorithm for ssNMR spectra called "ssPINE". The ssPINE software accepts the sequence of the protein plus peak lists from a variety of ssNMR experiments as input and offers automated backbone and side-chain assignments. The alpha version of ssPINE, which we describe here, is freely available through a web submission form at https://poky.clas.ucdenver.edu/ssPINE.

Keywords: solid-state NMR, membrane proteins, assignment, automation, ssPINE, MAS-NMR

Citation: Adilakshmi D.; Rahimi M.; Markley J.L.; Lee W. ssPINE: Probabilistic Algorithm for Automated Chemical Shift Assignment of Solid-State NMR Data from Complex Protein Systems. *Membranes* **2021**, *11*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

1. Introduction

NMR spectroscopy is one of the major biophysical methods, along with X-ray crystallography [1, 2] and cryo-electron microscopy [3], for determining structures of biomolecules. NMR is used to study structure-function

relationships of membrane proteins and large macromolecular assemblies [4] along with their interactions with small molecules [5] as an approach to drug discovery [6].

Both solution and solid-state NMR techniques provide important information about the structures and dynamics of membrane proteins [7, 8]. Solid-state NMR (ssNMR) with magic angle spinning (MAS) has advantages over solution NMR for studies of large and immobilized proteins [9, 10]. Anisotropic nuclear spin interaction information from ssNMR can be extremely useful for structure determination and dynamics [11, 12]. The orientation of regions of membrane proteins can be extracted from ssNMR spectra of mechanically or magnetically aligned membranes [13]. The broad lines and low resolution of ssNMR spectra resulting from anisotropy can be overcome in part by ultra high MAS, cross-polarization, refined pulse sequences [14], and non-uniform sampling (NUS). Ultra high-field NMR spectrometers operating at 1.1 GHz and 1.2 GHz are improving the resolution and sensitivity of ssNMR spectra of membrane proteins and their complexes. The above-mentioned methods are enabling the collection of improved spectral data, but manual analysis of the data to obtain chemical shift assignments and structural constraints is tedious because thousands of signals need to be analyzed, correlated, and labeled.

Software technology has reduced the burden of analyzing data from solution NMR studies of biomolecules. Available web-based resources provide automated and semi-automated algorithms for determining different parameters of biomolecules and their structure [15–18]. We recently developed an updated version of the assignment engine PINE [19], I-PINE (Integrative Probabilistic Interaction Network of Evidence) [20], which utilizes a Bayesian-based probabilistic interaction network. I-PINE supports a larger range of NMR experiments and integrates real-time statistical analysis of the PACSY database [21]. The I-PINE web server produces higher assignment coverage and accuracy than PINE, and supports structure determinations based on chemical shift assignments. The POKY suite includes iPick [22], for peak picking and cross-validation of peaks from different spectra, I-PINE and PINE-SPARKY.2 [23], a user-friendly graphical user interface (GUI) for submitting, importing, and validating the data [24].

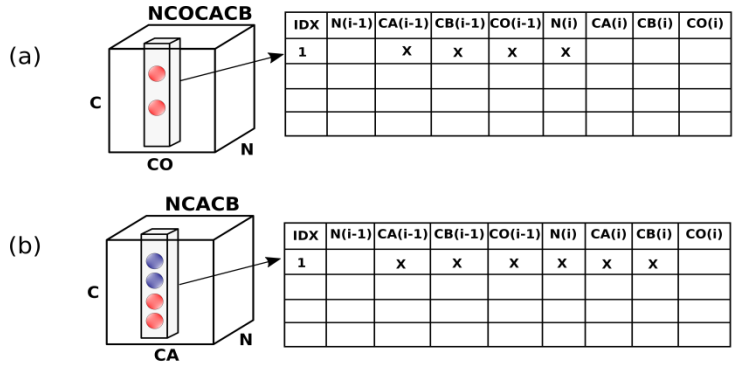
For ssNMR data, PISA-SPARKY [25], a plugin for the assignment program, NMRFAM-SPARKY [26], supports the analysis of data from oriented samples [27]. PISA-SPARKY, along with its features are now included in the POKY suite. Recently, the Veglia group introduced "a one-shot approach" called PHORONESIS that generates up to ten 3D ^{1}H-detected ssNMR spectra [28]. They used the I-PINE webserver to analyze the spectra, and found that the yield of sequential assignments was similar to that for solution NMR data. The Hunter Moseley and Chad Rienstra groups developed an ssNMR version of AutoAssign and demonstrated its ability to assign ssNMR data from the small protein, GB1 [29]. Software returned 84.1% correct assignments. The ssFLYA algorithm, which was introduced by Schmidt and colleagues [30] and currently is available only for commercial users, yielded 88-87 % and 77-90% correctness on protein microcrystals and amyloids.

Here, we describe ssPINE (solid-state PINE), a software package that is designed to handle the challenging features of ssNMR data from membrane proteins and other complex protein systems. ssPINE accepts, as input, 2D and 3D ssNMR data and gives, as output, chemical shift assignments and their probabilistic correctness. We have evaluated the performance of ssPINE with data from GB1 and with additional protein NMR data from the BMRB database [31]. The alpha version of ssPINE is freely available through a web server utility at https://poky.clas.ucdenver.edu/ssPINE.

2. Materials and Methods

2.1. ssPINE algorithm

As its first step, ssPINE generates spin system matrices [32] as shown in Figure 1. The main difference between the I-PINE and ssPINE algorithms is in their approach to comparing peaks from different experiments. I-PINE uses N_i and H_i in root experiments to find correlated signals ($CA/CB/CO_{i-1}$, $CA/CB/CO_i$) in different experiments and to establish di-peptide arrays {$CA/CB/CO_{i-1}$ N_i $CA/CB/CO_i$}. Then, it establishes a vector [CA_{i-1}, CB_{i-1}, CO_{i-1}, N_i, CA_i, CB_i, CO_i] and compares it to [CA_{i-1}, CB_{i-1}, CO_{i-1}] and [CA_i, CB_i, CO_i] from the other di-peptide arrays. Finally, I-PINE compares [CA_{i-1}, CB_{i-1}, CO_{i-1}] to [CA_i, CB_i, CO_i] in all vectors. By contrast, ssPINE uses CO_{i-1}, N_i, and CA_i in root experiments to find di-peptide signals (CX_{i-1}, CX_i). If an experiment is providing CO_{i-1}, N_i, and CA_i but in a single peak is not provided (e.g. CANCO), ssPINE combines information from different experiments, such as NCOCACB and NCACB, to obtain these correlations. Unlike I-PINE, ssPINE generates each spin system matrix by iterated spectral resolution steps (tolerances) using inter-residue connectivities optimized by a probability approach. This is a basic and important component of the ssNMR algorithm, because it helps to overcome the variable spectral resolution of ssNMR spectra. ssPINE calculates the quality of the data at each step until it reaches the point where there is no further improvement in the spin system matrix. If the quality of the data is above a threshold value, as determined by the number of spin systems and correlations between spin systems is identified compared to the numbers expected, then the process continues to the pentapeptide generation step. Otherwise, the process terminates and informs the user that more information is required. The pentapeptide generation step, which assigns signals to atoms in sequences of five amino acid residues, finds the best marginal probabilities by using the belief propagation algorithm [33] to evaluate relations between spin systems . This step includes the identification of secondary structural elements, the evaluation of possible referencing errors, and the continued assignment of backbone spin systems until convergence is reached or, alternatively, until the specified number of iterations has occurred. The last step utilizes the Bayesian network model of PINE and I-PINE to assign side chain signals [16, 20]. See the Supporting Information for a detailed description of the ssPINE algorithm.



IDX	N(i-1)	CA(i-1)	CB(i-1)	CO(i-1)	N(i)	CA(i)	CB(i)	CO(i)
1		X	X	X	X			

IDX	N(i-1)	CA(i-1)	CB(i-1)	CO(i-1)	N(i)	CA(i)	CB(i)	CO(i)
1		X	X	X	X	X	X	

IDX	N(i-1)	CA(i-1)	CB(i-1)	CO(i-1)	N(i)	CA(i)	CB(i)	CO(i)
1		X	X	X	X	X	X	
2		X	X	X	X	X	X	
3		X	X	X	X	X	X	
4		X	X	X	X	X	X	
5		X	X	X	X	X	X	
6		X	X	X	X	X	X	
7		X	X	X	X	X	X	
_		_	_	_	_	_	_	
_		_	_	_	_	_	_	
_		_	_	_	_	_	_	

Figure 1. Spin system matrix assembly in ssPINE. (a) Peaks from the strip of the NCOCACB experiment containing the CA(i-1) and CB(i-1) resonances are inserted in the row of the table corresponding to the i^{th} residue (IDX). Note that the peak is selected from the CO(i-1) and N(i) in root information. (b) Similarly, peaks from the strip of the NCACB experiment containing the CA(i) and CB(i) resonances are inserted in the row of the table corresponding to the i^{th} residue. N(i) and CA(i) in root information are used to select a peak from NCACB. (c) The process is repeated for all residues in the peptide sequence.

Table 1. ssNMR experiments supported by ssPINE with their dimensionality and connectivity profiles. CX(i) represents carbon A, B, D, E, G, or H atoms of the i^{th} residue; N(i) represents the nitrogen atom of the i^{th} residue; and CO(i-1) represents the carbon atom of the carboxyl group of the preceding residue. The minimum set of experiments needed is indicated by asterisks.

Experiment	Dimension	Profile
CC*	2D	CX/O(i)-CX/O(i)
NCA*	2D	N(i)-CA(i)
NCACB	2D	N(i)-CA/B(i)
NCO*	2D	N(i)-CO(i-1)
NCACO	3D	N(i)-CA(i)-CO(i)
NCACB	3D	N(i)-CA(i)-CA/B(i)
NCACX*	3D	N(i)-CA(i)-CX(i)
NCOCX*	3D	N(i)-CO(i-1)-CX/C(i-1)
NCOCA	3D	N(i)-CO(i-1)-CA(i-1)
NCOCACB	3D	N(i)-CO(i-1)-CA/B(i-1)
CANCO	3D	CA(i)-N(i)-CO(i-1)
CANCOCX*	3D	CA(i)-N(i)-CX/O(i-1)
CANCOCA	3D	CA(i)-N(i)-CA/O(i-1)
CANCOCACB	3D	CA(i)-N(i)-CO/A/B(i-1)

*Minimum experiments to run ssPINE.

2.2. Input files

As with PINE and I-PINE, peak lists (either raw or refined) and sequence files are used as inputs to ssPINE. Supported solid-state NMR experiments and their profiles are shown in Table 1. The minimum set of peak lists for assignments are those from 2D-CC, 2D-NCA, 2D-NCO, 3D-NCACX, 3D-NCOCX, and CAN(CO)CX ssNMR

12
117 experiments. Data from additional ssNMR experiments can be added to improve the accuracy and completeness of
118 the results.
119
120 2.2.1. *Preparation of peak lists*
121 Several peak list formats are accepted: Sparky (UCSF-/NMRFAM-SPARKY or POKY) with the *.list* file extension
122 prepared in the peak list window (two-letter-code "lt" with the *Data Heights* option turned on), XEASY with the *.peaks*
123 file extension [34], nmrDraw with the *.ft2* file extension, NMRView with the *.xpk* file extension, and I-PINE with the
124 *.txt* file extension. The file extension in the file name should match its actual format. Other programs can generate the
125 Sparky format, which is one of the most common file formats in the field. For example, CARA has the
126 *WriteSparkyPeakList.lua* script, and CCPNMR v2 has the *Format Converter* program [35, 36]. The POKY suite contains
127 multiple options for generating peak lists; of these, one of the easiest approaches is iPick. With iPick, the user simply
128 selects one or more spectra from the session and clicks on the "Run iPick" button. After peak lists have been generated
129 for each spectrum, the "Peak List" window opens, and by clicking on the "Save" button, the user can designate the
130 names for the peak lists. Peak lists can be refined by hand or by software to remove noise or other spurious peaks.
131
132 2.2.2. *Protein sequence*
133 ssPINE accepts peptide sequences in either one- or three-letter amino acid code as ASCII text files. Sequences
134 submitted in RTF (Rich Text Format; .rtf), ODT (OpenDocument Text; .odt), or DOCX (Office Open XML; .docx) are
135 automatically converted to ASCII.
136
137 *2.3. Output files*
138 The ssPINE output consists of several files: (1) The list of ssNMR experiments used. (2) A bar graph indicating
139 the assignment probabilities of each residue in the protein (Figure 2). (3) Separate files with the backbone and
140 sidechain chemical shift assignments of each residue of the protein in NMR-STAR 2.1 and 3.1 formats. (4) Sparky
141 format assignment labels and frequencies. (5) Protein secondary structure prediction by PECAN (Protein Energetic
142 Conformational Analysis from NMR chemical shifts) [37]. (6) Chemical shift referencing errors in each experiment, as
143 detected by LACS (Linear Analysis of Chemical Shifts) [38], which is used in redefining offsets during the assignment
144 iteration. (It is recommended that the user use these values to correct the offset for each peak list when a job is
145 resubmitted. This will reduce the computational time and improve the assignment accuracy.)
146
147 *2.4. Data used in developing ssPINE*
148 2.4.1. Data from *GB1*
149 In the early stages of developing ssPINE, we used unpublished ssNMR data from the uniformly $^{13}C/^{15}N$ labeled
150 small (56 residue, 6.2 kDa) protein GB1 generously provided by Chad Rienstra's group. GB1, which is the
151 streptococcal B1 immunoglobulin-binding domain of protein G20, has been used frequently as a standard sample in
152 the development of NMR technology. We prepared both unrefined and refined peak lists from raw data from the
153 following ssNMR experiments: 2D-CC, 2D-NCA, 2D-NCACB, 2D-NCO, 3D-NCACB, 3D-NCACX, 3D-NCACO, 3D-
154 CANCO, 3D-CANCOCX, 3D-NCOCA, 3D-NCOCACB, and 3D-NCOCX. We prepared unrefined peak lists
155 automatically with the iPick peak picking tool of POKY (two-letter-code *iP*). Subsequently, we created refined peak
156 lists by using the cross-validation tool of iPick to weed out noise and non-sequential signals.
157
158 2.4.2. *Other protein NMR data*
159 Additional data from the PACSY database [21] were used in refining and optimizing the ssPINE algorithm.
160 PACSY is a relational database that contains post-processed information from BMRB [39] and PDB [40]. Data from 82
161 proteins, including both large (181 residues) and small (26 residues) proteins, were included (SI Table 1). Most data
162 sets were from solution NMR, because few ssNMR entries in the BMRB contain complete assignments. We created
163 synthetic peak lists for 2D-CC, 2D-NCA, 2D-NCACB, 2D-NCO, 3D-NCACB, 3D-NCACX, 3D-NCACO, 3D-CANCO,
164 3D-CANCOCX, 3D-NCOCA, 3D-NCOCACB, and 3D-NCOCX ssNMR spectra of these proteins. For more controlled
165 evaluation, we only regarded sequential cross peaks.

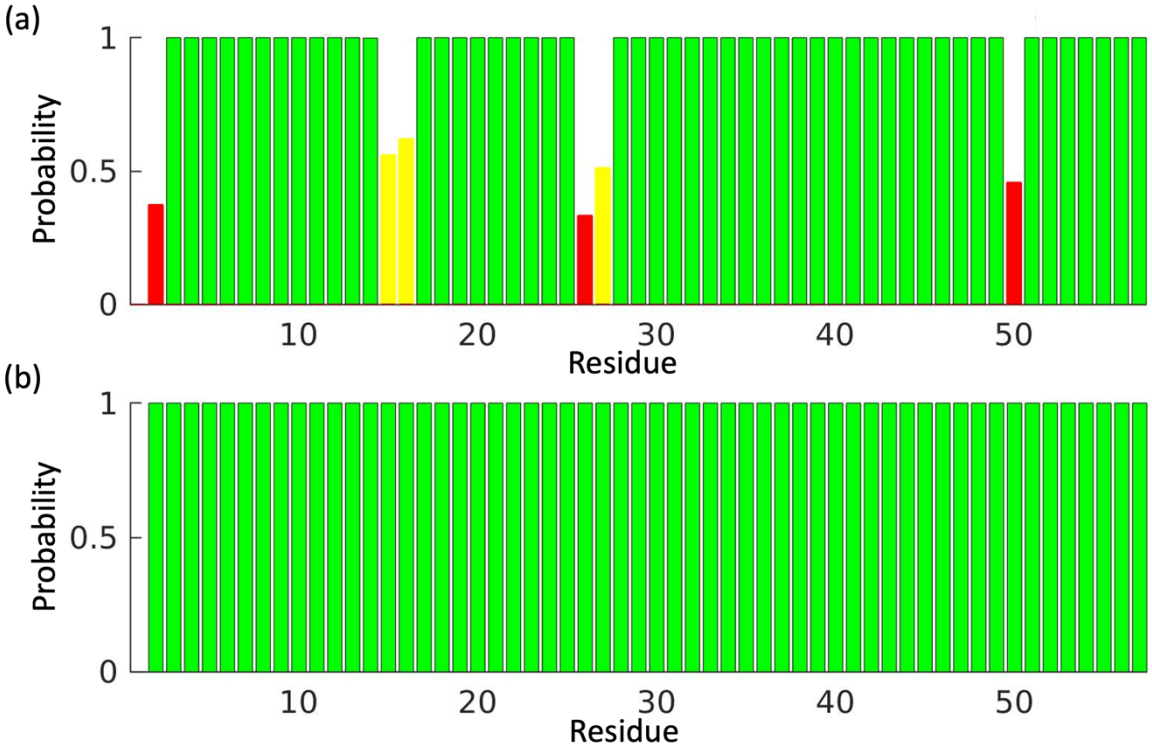
166 2.5. *ssPINE web server*
167 We utilized multiple technologies in implementing the ssPINE algorithm as a web server. Programs written in
168 Perl, Python, and shell scripting handle various parts of the task. A web-facing server hosts a form that the user can
169 fill out with their information: the amino acid sequence file and the peak lists from specified 2D and 3D solid-state
170 NMR experiments. By clicking the "Submit" button, this information is validated and sent to a processing server.
171 After the automated backbone and sidechain assignments are completed, the result is sent back to the user's email
172 address. From there, the user can download all the result files. The actual running time is determined by the size of
173 the protein and the complexity of the problem including peak list quality provided by the user, but jobs usually
174 require less than one hour. The ssPINE web server is hosted at the University of Colorado Denver and is accessible at:
175 https://poky.clas.ucdenver.edu/ssPINE. No login or signup is required, and the server is open to all researchers at
176 no cost and processes submissions in the order in which they are received.

15

177 3. **Results**

178 We evaluated the results with GB1 in terms of their *completeness* and *correctness*. "Completeness" is the number
179 of automatically assigned chemical shifts by ssPINE divided by the number of assignments for GB1 derived from our
180 manual assignment of the ssNMR data. "Correctness" is the number of correct assignments made by ssPINE divided
181 by the number of manual assignments. Because ssPINE provides multiple assignment candidates with associated
182 probabilities, only the assignment candidate with the highest probability is used in the evaluation of completeness
183 and correctness.

184 With the unrefined peak lists of GB1 as input, ssPINE yielded 100% (219/219) completeness and an average of
185 97.26% (213/219) correctness for the backbone chemical shift assignments (Figure 2a). With the refined peak lists of
186 GB1, ssPINE yielded 100% (219/219) completeness and 100% (219/219) accuracy (Figure 2b).



187

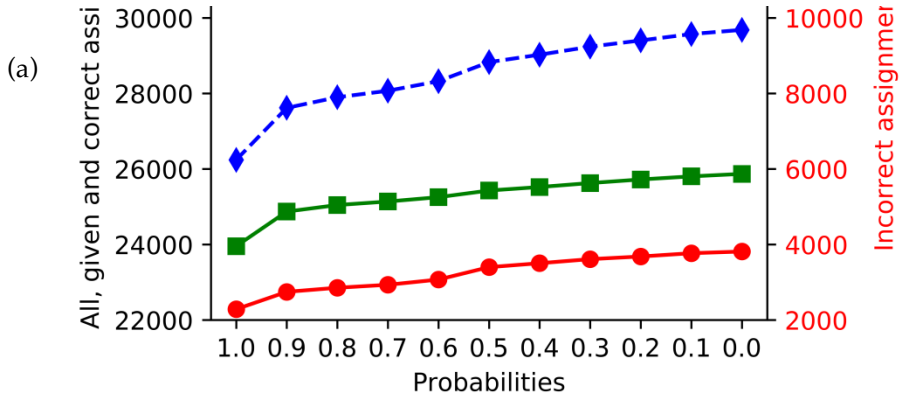
188 **Figure 2.** Bar graphs indicating the correct assignment probability (*p*) for each residue of GBI resulting from ssPINE analysis. Green
189 indicates *p greater than* 0.99; cyan indicates $p = 0.85 - 0.99$; yellow indicates $p = 0.5 - 0.84$; red indicates *p less than* 0.5; and gray
190 indicates no assignment (not seen with these test sets). (a) Unrefined GBI data as input. (b) Refined GBI data as input.

191 We also tested ssPINE algorithm with synthetic peak lists from other proteins whose assigned chemical shifts
192 had been deposited in BMRB (see section 2.4.2.). These BMRB assignments are assumed to be correct and were used in
193 evaluating the correctness of the ssPINE results. The numbers of BMRB and ssPINE assignments were used,
194 respectively, as the denominator and numerator in the completeness calculation. The number of valid ssPINE
195 assignments ("given" assignments) at the different probability cutoffs were used as the denominator in the
196 correctness calculation.

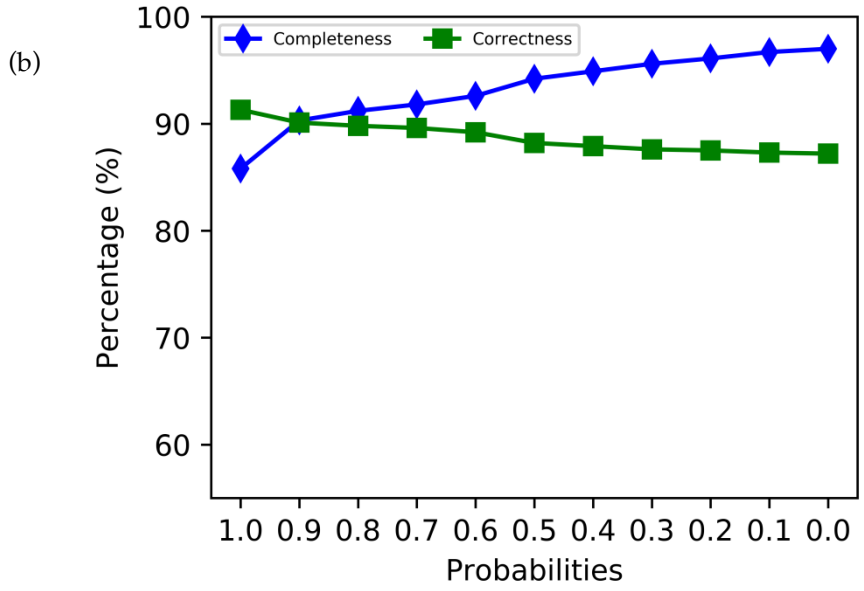
197 The total number of assignment candidates returned by ssPINE are plotted as a function of their probability
198 scores in Figure 3a. They are shown as "correct", "incorrect", "given" (sum of correct and incorrect), and "all". The
199 "all" category includes "given" plus invalid assignments, namely those with scores below the probability cutoff.

200 The correctness and completeness parameters for all assignment candidates with the highest probability for each
201 protein are plotted with respect to their probability in Figure 3b. Correctness decreased moderately as a function of
202 lower probability. The fact that it remained above 85% means that more than 85% of the given chemical shift values
203 were assigned correctly. Overall, completeness ranged between 85% and 97%. Completeness increased abruptly
204 between 1.0 and 0.9 probability and then more gradually to 0.0 probability. Plots of percentages of completeness
205 versus correctness for each BMRB entry at each probability are given in SI Figure 2.

18



206



207
208 **Figure 3.** Results from ssPINE analysis of synthetic ssNMR data as averages for the 82 proteins studied. (a)
209 Chemical shift assignment probabilities returned by ssPINE for all assignment candidates (x-axis) versus assignment
210 type (y-axis). All (dashed black) and, given (dashed blue) and correct (solid green) assignments are represented by the
211 numbers on the left side, whereas the incorrect assignments (solid red) are represented by the numbers on the right
212 side. (b) Data from the assignment candidate for each protein with the highest assignment probability. Completeness
213 (solid blue) and correctness (solid green) are plotted as a function of that assignment probability.
214
215 The unrefined GB1 peak lists led to a few incorrect $^{13}C^{\alpha}$ assignments (Figure 2a) because false signals picked by
216 automated peak picking algorithm were close to the BMRB average chemical shift value. Manual refinement of the
217 peak lists alleviated this problem by removing false-positive peaks, adding unpicked peaks, and resolving overlaps.
218 Of the 82 synthetic sets of peak lists analyzed by ssPINE, only three yielded assignment correctness below 70%
219 with probability cutoff 0.5. These are denoted by red circle in SI Figure 2 and by red text in SI Table 1. One of the
220 poorest scoring data sets (completeness = 84.5% (474/561); correctness = 66.9% (317/474)) corresponded to BMRB
221 entry 15716 (the AlgE6R1 subunit from the *Azotobacter vinelandii* Mannuronan C5-epimerase), a 153 amino acid
222 protein containing 27 glycine residues with many overlapping peaks in the carbon alpha region (~45ppm).

223 **4. Discussion**

224 In this report, we have introduced the ssPINE algorithm, for the automated analysis and assignment of solid-
225 state
226 NMR data from membrane proteins and other difficult proteins systems. ssPINE builds on the technology of our I-
227 PINE web server for solution NMR data, which serves several thousand jobs annually. We have adapted the I-PINE
228 algorithm to account for the challenging features of ssNMR data from these systems. These include broader lines,
229 extensive inter-residue dipolar interactions, and 2/3D ssNMR experiments that yield a variety of connectivities. As
230 with I-PINE, ssPINE accepts as input the amino acid sequence of the protein and raw or refined peak lists from a

21
231 variety of NMR experiments (Table 1). The output of ssPINE includes peak assignments and their probabilities. We
232 have tested and refined the implementation of the ssPINE algorithm with the excellent set of ssNMR data from the
233 small protein GB1. We also used as input to ssPINE a set of synthetic peak lists that simulated ssNMR data from 82
234 other proteins of various size generated from solution NMR data deposited in BMRB. As shown above, the choice of
235 probability cutoff is an important factor in maximizing correct assignments. In solution NMR, the recommended
236 probability cutoff for I-PINE is 0.5, because it leads to a higher probability of correct assignments [20]. With ssNMR
237 data, a cutoff of 0.6 appears to provide optimal completeness and assignment correctness. Glycine residues are harder
238 to assign because they lack the CB signals that ssPINE uses to evaluate connectivities. Proteins that contain a high
239 glycine content (e.g., BMRB entry 15716) are particularly problematic because ssPINE has difficulty distinguishing
240 among the several glycine candidates.
241 Currently, the user can use the ssPINE extension in POKY (two-letter-code *EP*) to generate and submit peak lists
242 from the web browser to the ssPINE webserver. The user can use the *Convert (ss)I-PINE outputs to POKY plugin* in
243 POKY (two-letter-code *ip*) to convert the assigned chemical shift table file from ssPINE to the POKY resonance list file
244 with the chosen probability cutoff. Finally, the POKY Notepad (two-letter-code *Pn*) can be used to propagate assigned
245 peaks onto ssNMR spectra: this is enabled by the script, *Simulate SSNMR peaks with assignments labels (predict-and-*
246 *confirm)*.
247 The analysis of ssNMR data from membrane proteins is highly challenging. ssPINE offers a promising approach
248 to resolving the chemical, structural, and dynamic information contained in these spectra. Information of this kind is
249 crucial to understanding the mechanisms underlying membrane transport, energy transfers, and signaling. We
250 encourage feedback from users of ssPINE, particularly those analyzing ssNMR spectra of membrane proteins, as a
251 means for guiding its further development. Our immediate goals with ssPINE are to incorporate information from
252 strategies commonly used in NMR spectroscopy of membrane proteins, including mutational analysis, ^{19}F labeling,
253 and/or selective isotopic labeling.
254 Longer-term plans are to develop and release a program (ssPINE-POKY) that will include a graphical user
255 interface analogous to that in PINE-SPARKY.2 for solution NMR. In addition, we envision an "integrative" version of
256 ssPINE that will increase assignment correctness and completeness by implementing adaptive probability density
257 functions that incorporate machine learning (ML) based chemical shift and structure prediction methods, and will
258 provide comprehensive visualization of structural and dynamic information from ssNMR data, analogous to that
259 afforded by I-PINE for solution NMR data.
260

261 ## 5. Web server availability

262 The usage of the webserver is described in section 2.5. The web server for ssPINE is freely accessible at
263 https://poky.clas.ucdenver.edu/ssPINE.
264

265 **Supplementary Materials:** The following are available online at www.mdpi.com/xxx/s1

266 **Author Contributions:** Conceptualization, W.L.; methodology, W.L.; software, A.D., M.R. and W.L.; validation, A.D.; formal
267 analysis, A.D.; resources, M.R.; data curation, A.D.; writing—original draft preparation, A.D., M.R. and W.L.; writing—review and
268 editing, J.L.M.; supervision, W.L.; funding acquisition, W.L. and J.L.M. All authors have read and agreed to the published version
269 of the manuscript.

270 **Funding:** This work has been supported by the NIH P41GM103399 (to J.L.M.), the NSF DBI-1902076 and DBI-2051595 (to W.L.),
271 and new faculty research support from the University of Colorado Denver (to W.L.).

272 **Institutional Review Board Statement:** Not applicable.

273 **Acknowledgments:** We thank Chad Rienstra's group, Biochemistry Department, University of Wisconsin-Madison for providing
274 the GB1 experimental data.

275 **Conflicts of Interest:** The authors declare no conflict of interest.

276

277 ## References

278 1. Ilari, A.; Savino, C. Protein Structure Determination by X-Ray Crystallography. Bioinformatics 2008, 63–87.
279 2. Fujiyoshi, Y. The Structural Study of Membrane Proteins by Electron Crystallography. Adv. Biophys. 1998, 35, 25–80.
280 3. Allen O áHill, H. Immobilization of Small Proteins in Carbon Nanotubes: High-Resolution Transmission Electron Microscopy
281 Study and Catalytic Activity. J. Chem. Soc. Chem. Commun. 1995, 1803–1804.
282 4. Wüthrich, K. The Way to NMR Structures of Proteins. Nat. Struct. Biol. 2001, 8, 923–925.
283 5. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. Science
284 1996, 274, 1531–1534.
285 6. Karsisiotis, A.I. NMR Studies of Inhibitor Binding to Metallo-B-Lactamases, University of Leicester, 2008.
286 7. Kempf, J.G.; Loria, J.P. Protein Dynamics from Solution NMR. Cell Biochem. Biophys. 2002, 37, 187–211.
287 8. Kovermann, M.; Rogne, P.; Wolf-Watz, M. Protein Dynamics and Function from Solution State NMR Spectroscopy. Q. Rev.
288 Biophys. 2016, 49.
289 9. Castellani, F.; Van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. Structure of a Protein Determined by Solid-
290 State Magic-Angle-Spinning NMR Spectroscopy. Nature 2002, 420, 99–102.

24

10. Shahid, S.A.; Bardiaux, B.; Franks, W.T.; Krabben, L.; Habeck, M.; van Rossum, B.-J.; Linke, D. Membrane-Protein Structure Determination by Solid-State NMR Spectroscopy of Microcrystals. Nat. Methods 2012, 9, 1212.
11. Hong, M. Solid-State NMR Determination of 13Cα Chemical Shift Anisotropies for the Identification of Protein Secondary Structure. J. Am. Chem. Soc. 2000, 122, 3762–3770.
12. Reif, B. Ultra-High Resolution in MAS Solid-State NMR of Perdeuterated Proteins: Implications for Structure and Dynamics. J. Magn. Reson. 2012, 216, 1–12.
13. De Angelis, A.A.; Jones, D.H.; Grant, C.V.; Park, S.H.; Mesleh, M.F.; Opella, S.J. NMR Experiments on Aligned Samples of Membrane Proteins. Methods Enzymol. 2005, 394, 350–382.
14. Wi, S.; Frydman, L. An Efficient, Robust New Scheme for Establishing Broadband Homonuclear Correlations in Biomolecular Solid State NMR. ChemPhysChem 2020, 21, 284–294.
15. Kobayashi, N.; Iwahara, J.; Koshiba, S.; Tomizawa, T.; Tochio, N.; Güntert, P.; Kigawa, T.; Yokoyama, S. KUJIRA, a Package of Integrated Modules for Systematic and Interactive Analysis of NMR Data Directed to High-Throughput NMR Structure Studies. J. Biomol. NMR 2007, 39, 31–52.
16. Bahrami, A.; Assadi, A.H.; Markley, J.L.; Eghbalnia, H.R. Probabilistic Interaction Network of Evidence Algorithm and Its Application to Complete Labeling of Peak Lists from Protein NMR Spectroscopy. PLoS Comput Biol 2009, 5, e1000307.
17. Lee, W.; Kim, J.H.; Westler, W.M.; Markley, J.L. PONDEROSA, an Automated 3D-NOESY Peak Picking Program, Enables Automated Protein Structure Determination. Bioinformatics 2011, 27, 1727–1728.
18. Lee, W.; Stark, J.L.; Markley, J.L. PONDEROSA-C/S: Client–Server Based Software Package for Automated Protein 3D Structure Determination. J. Biomol. NMR 2014, 60, 73–75.
19. Lee, W.; Westler, W.M.; Bahrami, A.; Eghbalnia, H.R.; Markley, J.L. PINE-SPARKY: Graphical Interface for Evaluating Automated Probabilistic Peak Assignments in Protein NMR Spectroscopy. Bioinformatics 2009, 25, 2085–2087.
20. Lee, W.; Bahrami, A.; Dashti, H.T.; Eghbalnia, H.R.; Tonelli, M.; Westler, W.M.; Markley, J.L. I-PINE Web Server: An Integrative Probabilistic NMR Assignment System for Proteins. J. Biomol. NMR 2019, 73, 213–222.
21. Lee, W.; Yu, W.; Kim, S.; Chang, I.; Lee, W.; Markley, J.L. PACSY, a Relational Database Management System for Protein Structure and Chemical Shift Analysis. J. Biomol. NMR 2012, 54, 169–179.
22. Rahimi, M.; Lee, Y.; Markley, J.L.; Lee, W. iPick: Multiprocessing Software for Integrated NMR Signal Detection and Validation. J. Magn. Reson. 2021, 328, 106995.
23. Lee, W.; Markley, J.L. PINE-SPARKY. 2 for Automated NMR-Based Protein Structure Research. Bioinformatics 2018, 34, 1586–1588.
24. Lee, W.; Rahimi, M.; Lee, Y.; Chiu, A. POKY: A Software Suite for Multidimensional NMR and 3D Structure Calculation of Biomolecules. Bioinformatics 2021.
25. Weber, D.K.; Wang, S.; Markley, J.L.; Veglia, G.; Lee, W. PISA-SPARKY: An Interactive SPARKY Plugin to Analyze Oriented Solid-State NMR Spectra of Helical Membrane Proteins. Bioinformatics 2020, 36, 2915–2916.
26. Lee, W.; Tonelli, M.; Markley, J.L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. Bioinformatics 2015, 31, 1325–1327.
27. Opella, S.J.; Marassi, F.M. Structure Determination of Membrane Proteins by NMR Spectroscopy. Chem. Rev. 2004, 104, 3587–3606.
28. Gopinath, T.; Manu, V.S.; Weber, D.K.; Veglia, G. PHRONESIS: A One−Shot Approach for Sequential Assignment of Protein Resonances by Ultrafast MAS Solid−State NMR Spectroscopy. ChemPhysChem 2022, doi:10.1002/cphc.202200127.
29. Moseley, H.N.; Sperling, L.J.; Rienstra, C.M. Automated Protein Resonance Assignments of Magic Angle Spinning Solid-State NMR Spectra of B1 Immunoglobulin Binding Domain of Protein G (GB1). J. Biomol. NMR 2010, 48, 123–128.
30. Schmidt, E.; Gath, J.; Habenstein, B.; Ravotti, F.; Székely, K.; Huber, M.; Buchner, L.; Böckmann, A.; Meier, B.H.; Güntert, P. Automated Solid-State NMR Resonance Assignment of Protein Microcrystals and Amyloids. J Biomol NMR 2013, 56, 243–254, doi:10.1007/s10858-013-9742-x.
31. Zhou, D.H.; Nieuwkoop, A.J.; Berthold, D.A.; Comellas, G.; Sperling, L.J.; Tang, M.; Shah, G.J.; Brea, E.J.; Lemkau, L.R.; Rienstra, C.M. Solid-State NMR Analysis of Membrane Proteins and Protein Aggregates by Proton Detected Spectroscopy. J. Biomol. NMR 2012, 54, 291–305.
32. Gopinath, T.; Veglia, G. Multiple Acquisition of Magic Angle Spinning Solid-State NMR Experiments Using One Receiver: Application to Microcrystalline and Membrane Protein Preparations. J. Magn. Reson. 2015, 253, 143–153.
33. Yedidia, J.S.; Freeman, W.T.; Weiss, Y. Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms. IEEE Trans. Inf. Theory 2005, 51, 2282–2312.
34. Bartels, C.; Xia, T.; Billeter, M.; Güntert, P.; Wüthrich, K. The Program XEASY for Computer-Supported NMR Spectral Analysis of Biological Macromolecules. J. Biomol. NMR 1995, 6, 1–10.
35. Keller, R.L.J. Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer Aided Resonance Assignment, ETH Zurich, 2005.
36. Vranken, W.F.; Boucher, W.; Stevens, T.J.; Fogh, R.H.; Pajon, A.; Llinas, M.; Ulrich, E.L.; Markley, J.L.; Ionides, J.; Laue, E.D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. Proteins Struct. Funct. Bioinforma. 2005, 59, 687–696.
37. Eghbalnia, H.R.; Wang, L.; Bahrami, A.; Assadi, A.; Markley, J.L. Protein Energetic Conformational Analysis from NMR Chemical Shifts (PECAN) and Its Use in Determining Secondary Structural Elements. J. Biomol. NMR 2005, 32, 71–81, doi:10.1007/s10858-005-5705-1.
38. Wang, L.; Eghbalnia, H.R.; Bahrami, A.; Markley, J.L. Linear Analysis of Carbon-13 Chemical Shift Differences and Its Application to the Detection and Correction of Errors in Referencing and Spin System Identifications. J. Biomol. NMR 2005, 32, 13–22.
39. Ulrich, E.L.; Akutsu, H.; Doreleijers, J.F.; Harano, Y.; Ioannidis, Y.E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z. BioMagResBank Nucleic Acids Res. 36. D402–D408 2008.
40. Berman, H.; Henrick, K.; Nakamura, H.; Markley, J.L. The Worldwide Protein Data Bank (WwPDB): Ensuring a Single, Uniform Archive of PDB Data. *Nucleic Acids Res.* **2007**, *35*, D301–D303.

359
360
361