REDEN: Interactive Multi-Fitting Decomposition-based NMR Peak Picking Assistant

Mehdi Rahimi,[1,†] Abigail Chiu,[1,†] Andrea Estefania Lopez Giraldo,[1] Je-Hyun Yoon,[2] Woonghee Lee [1,*]

[1] Department of Chemistry, University of Colorado Denver, Denver, CO, 80204, USA

[2] Department of Oncology Science, University of Oklahoma, Oklahoma City, OK 73104

[†] Co-first authors

[*] Correspondence: woonghee.lee@ucdenver.edu

Abstract

We present a new program REDEN (Residual Decomposition of NMR peaks) designed to perform identification of peaks in NMR spectra. This integrated, cross-platform, open-source software visually assists with explicit peak picking through decomposition of NMR peaks on the frequency domain data. It provides a distinctive interactive workflow with iPick due to its integration with the POKY suite, providing users with a seamless and efficient experience. The decomposition of peaks operates in a chosen region of an NMR spectrum by multi-fitting simulated peaks with four lineshape fitting options as support, Gaussian, Lorentzian, a fast/optimized Lorentzian, and Pseudo-Voigt. Furthermore, REDEN provides a way to fine-tune for the users in two operating modes (Basic and Advanced). REDEN is pre-built in the POKY suite, which is available from https://poky.clas.ucdenver.edu.

Keywords: Peak decomposition; multi-fitting; graphical user interface; REDEN; POKY

1. Introduction

Biomolecular NMR (Nuclear Magnetic Resonance) is a versatile tool for the study of the structure, dynamics, and interactions of biological molecules such as proteins, nucleic acids, and carbohydrates. Biomolecular NMR research involves a variety of essential tasks, including resonance assignments, structure determination, dynamic characterization, ligand binding studies, protein-protein interaction investigations, and metabolomics. Among these attributes, a crucial common trait is “signal detection”, which is the ability to distinguish signal from noise. “Peak picking” represents a more specific type of signal detection that refers to the distinction of signals from each other. It can affect the overall quality and reliability of outcomes especially since overall signal-to-noise is one of major reasons causing unsatisfactory peak picking. It can make peak picking a time-consuming and difficult process, particularly in multidimensional NMR spectra. Nonetheless, advances in NMR instrumentation and software tools have improved the accuracy and efficiency of peak picking in recent years.

We previously introduced an automated peak picking program called iPick, which is based on local extrema of peaks with rigorous validation criteria accompanied with easy-to-use graphical user interfaces (GUIs) (Rahimi et al., 2021). Still, because some peaks have lower intensity or proximity compared to more intense peaks, they tend to get overshadowed, making them obscured from regular peak picking. A few approaches have been introduced to overcome this challenge like geometry based algorithm (Wurz & Guntert, 2017), line shape analysis (Waudby et al., 2016), and signal modeling (Dudley et al., 2020). Nevertheless, peak integration has shown promising results (Ahlner et al., 2013).

In recent years and based on proliferation of machine learning techniques, some approaches have used such capabilities to build systems for identifying these shoulder peaks. A deep neural network

(DNN)-based approach (Li et al., 2021) and convolutional neural network (CNN) (Klukowski et al., 2018) have been used to analyze 1D and 2D NMR spectra. Higher-dimensional NMR techniques, such as 3D NMR, commonly used in the study of biomacromolecules such as proteins and nucleic acids, remain challenging and insufficiently supported because they require more interactive user intervention with the analysis program, such as navigating different planes through the z-axis and rotating axes. Additionally, adopting these approaches may pose difficulties as each requires the installation of new software that could conflict with user's operating system or pre-existing libraries. Moreover, these approaches concentrate on scrutinizing an entire spectrum, an endeavor that can be both time-consuming and unnecessary for many signals despite certain spectral regions presenting challenges, such as peak pairs or crowded peak clusters only need to be scrutinized. These issues can constitute a significant impediment for users analyzing spectra. Since these standalone programs exist only for peak picking, other programs such as POKY must be used in parallel to assemble the actual spectrum and study its structure and dynamics (Lee et al., 2021). As a result, there is an urgent need to solve the difficulty of repeating not only the file input and output for different programs, but also the full spectrum analysis of the deep learning-based program itself.

In this situation, we introduce REDEN, one of the latest additions to the POKY suite, an integrated plugin program that confronts these issues by offering a multi-fitting approach to identifying hidden shoulder peaks in a crowded cluster. Thanks to its seamless integration with iPick in POKY, REDEN provides users with a unique interactive workflow for multidimensional NMR spectra, establishing it as a user-friendly solution for various applications such as proteomics and metabolomics. The inclusion of REDEN in the Integrative NMR platform enhances its robustness

and flexibility, thus rendering it an effective choice for studying biomolecular structures and functions (Lee et al., 2016).

2. Implementation

The REDEN software is written in Python 3 and can be run as a module for POKY. The user does not need to prepare and import separate files for operating REDEN. While running through POKY, REDEN will give the user the best experience via the cohesive integration with other preexisting tools. Since the POKY suite is a cross-platform program, the user can run REDEN on Linux, Mac, or Windows without the need for difficult installation steps. SBGrid (Morin et al., 2013) and NMRbox (Maciejewski et al., 2017) provide the POKY suite to their subscribers, and REDEN is also accessible from their service. Furthermore, a Singularity container version is available for outdated operating systems.

2.1. Initiating REDEN.

The general workflow starts with user selecting a cluster on a spectrum either zooming in (arrow keys, +/- keys, mouse wheels, etc) or by indications given in the iPick peak list window with an asterisk flag (Fig. 1A). The iPick program (two-letter-code *"ip"*) is also a pre-existing module that runs an efficient local extremum-based peak picking followed by the peak shape fitting of the user's choice from Gaussian, Lorentzian, and Pseudo-Voigt in the Integration Settings window (two-letter-code *"it"*). The fitting by iPick is performed on picked peaks, but it does not decompose multiple peaks clustered that are overshadowed and obscured. Still, it assesses the fitting quality by calculating residuals from actuals subtracted by models. When a cluster is selected (viewed) from either iPick's suggestion or user's choice, REDEN can be initiated by using the two-letter-code "*re"*. This will open the main window of REDEN (Fig. 1B) showing the selected cluster. The appropriate module will open automatically depending on whether it is a 2D or 3D NMR cluster.

2.2. Features of REDEN.

REDEN offers two modes of operation from its main window: "Basic" mode, which is the default (Fig. 1B), and "Advanced" mode, accessible by selecting the corresponding checkbox (Fig. 1D).

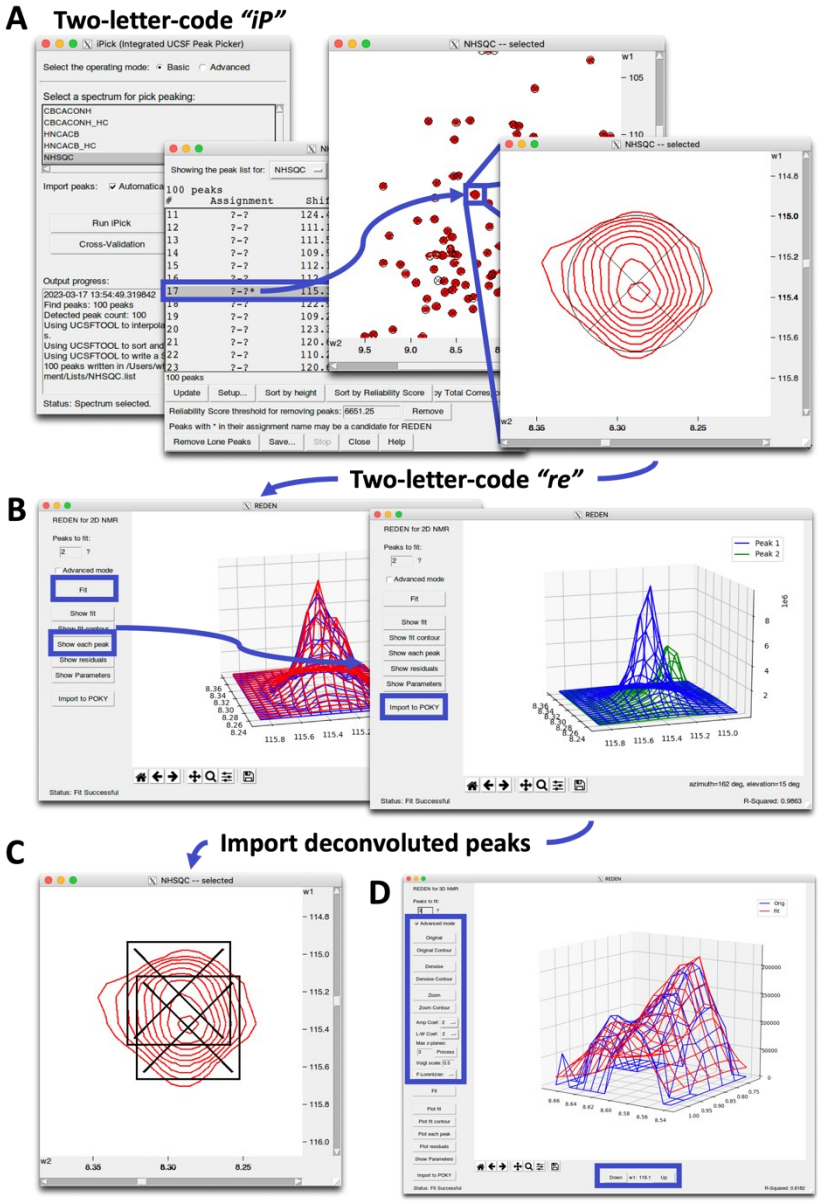

Fig. 1. The suggested workflow of REDEN GUI assisted by the iPick peak picker. (A) A cluster flagged by the iPick peak picker (two-letter-code "*iP*") is selected by the user for the further analysis using REDEN. (B) REDEN is executed (two-letter-code "*re*") and suggests two peaks for the cluster. It deconvolutes the cluster and fits two peaks in the *Basic* mode. (C) Fitted peaks can be simply imported back to POKY by clicking the "Import to POKY" button after the deconvolution process is completed. (D) A screenshot of the *Advanced* mode of 3D REDEN. The same workflow can be used to analyze a cluster in the 3D spectrum. Additional buttons for the Z-dimension navigation (*Down* and *Up*; blue box) left and right to the manual plane input box are offered for 3D REDEN.

Basic mode. In Basic mode, fitting a cluster of peaks is as simple as selecting the desired number of peaks and clicking the "Fit" button. The default estimated number of peaks is the number REDEN recognizes in that window by local maximum criterion. However, the user can easily adjust this value to fit more peaks in that cluster to see if there is a peak hidden in the cluster. In just a few seconds, the user can examine the fitting by manipulating the cluster and viewing it from different angles via mouse movements. The results can be shown in various options. The overlay of the actual data and fitting model can be summoned whenever the user clicks the “Show fit” button from the left panel. In addition, users may want to see a 2D of the fit, so REDEN provides an option to display a contour plot of the cluster and fitting by “Show fit contour”. “Show each peak” provides visuals of each of the peaks that REDEN recognizes. “Show residuals” is a residual plot of the fitting. “Show Parameters” gives the fitting parameters, peak amplitude, linewidth, skewness, and the peak center, of each identified peak. At the same time, parameters based on the spectrum can be given such as volume, fit height, line widths, and data heights. Figures of these options are available in the Supplementary document. If the fitting result is unsatisfactory, the user can repeat the process with a different number of peaks to fit. R^2 is provided to evaluate the goodness of fit. Users can see how it changes as parameters are changed and use this information to find the best fit. Once the optimal peaks have been picked by the decomposition, the peaks and their parameters can be effortlessly imported back into the spectrum via the "Import to POKY" button (Fig. 1C). The process for decomposition peaks in 3D NMR is identical to that for 2D NMR, as described earlier. The main distinction is that 3D NMR produces 4D data because it includes additional peak intensity, so visualizing the data as a 3D plot will only display a single plane. To address this, the 3D REDEN module includes supplementary buttons that enable users to navigate

the third axis of the data which is essentially a stack of 2D planes (bottom-center of Fig. 1D; blue box).

Advanced mode. REDEN's "Advanced" mode provides users with a wealth of data and extensive fine-tuning capabilities (green box in Fig. 1D). This mode offers additional buttons for displaying intermediate processing steps, including 3D and contour plots, which can be useful in defining a cluster of peaks. Given the challenge of accurately identifying a cluster when peaks are closely spaced, these intermediate plots offer insights into the processing steps that led to the result. If a mistake was made in selecting a nearby cluster, the user can adjust the viewing window and re-run REDEN as needed. Aside from the peak display options in the “Basic” mode described above, the “Advanced” mode provides a way to display different aspects of the selected data region (buttons in the green box of Fig. 1D). The user can see the intermediate steps that occurs throughout REDEN calculation. First, “Original” and “Original Contour” show the fits on the original data that the user selects. Next, “Denoise” and “Denoise Contour” shows the fits on the denoised data. When one of these buttons is pressed, REDEN performs the wavelet denoising in real time and use the cleaner data. Then, “Zoom” and “Zoom Contour” shows only the identified main cluster while everything else including other nearby clusters are hidden. Buttons with the “Contour” in the label show 2D contour plots, while the buttons without, show 3D mesh plots. REDEN uses default values for fitting in the “Basic” mode, but sometimes the cluster is so out of shape that the default values do not result in a good fit. In such cases, the user can adjust the "Amp Coef" (the coefficient of amplitude) or "L-W Coed" (the coefficient of linewidths) parameters to attempt another fit in the “Advanced” mode. Usually, just adjusting these parameters once will improve the fit of the cluster.

3D spectrum. REDEN implements the axis order defined in the selected view window of POKY for the 3D spectrum analysis. “Max z-planes” (also in blue box in the Fig. 1D) defines how many planes through z-dimension will be used. Because REDEN applies this “Max z-planes” value to navigate and capture the cluster, sometimes it is necessary to adjust it accordingly. If the cluster is not identified, not a whole, or overlapped with the other cluster, REDEN will advise the user what to do for identifying a cluster successfully. The user will need to zoom in, zoom out, or pan the spectral view extent if x- and y-dimensions need to be adjusted. If z-dimension needs to be adjusted, the user will use different value in the “Max z-planes” box.

2.3. Lineshape fitting options.

REDEN provides four lineshape fitting options: Gaussian, Lorentzian, a fast/optimized Lorentzian (called "F-Lorentzian"), and Pseudo-Voigt (Zaghloul and Ali, 2012). Users can switch between these options in the “Advanced” mode. A fitting optimization algorithm was used to minimize the difference between the lineshape simulation and the data in a time efficient manner. The Sequential Least Squares Programming (SLSQP) was chosen as the fitting optimization method due to its advantageous for a moderate number of variables and constraints (Kraft, 1988). The “F-Lorentzian” is a Lorentzian function that we formulated the multivariate_t function with NumPy’s mathematics functions, while the “Lorentzian” is formulated with SciPy’s multivariate_t function. The “Gaussian” is formulated with the SciPy’s multivariate_normal function. The “Pseudo-Voigt” is the combination of Lorentzian and Gaussian functions with the linear combination scale between them. By default, the Gaussian lineshape is used, as also employed in Basic mode, and the user can change to one of other lineshapes to obtain the most satisfactory results without hassles in the “Advanced” mode.

3. Results and Discussion

When it comes to real life application, it's highly probable that users can encounter some major clusters that may be difficult for POKY's stock fitting. Here, we show how to tackle this issue via REDEN with two examples, proteins and metabolites.

3.1. REDEN testing on synthesized metabolite spectrum.

To test out how well REDEN can perform, we synthesized a 2D ^{1}H,^{13}C-HSQC NMR spectrum of a metabolite mixture made of two standards, maslinic acid and bakuchiol found in the Biological Magnetic Resonance Bank database (BMRB) (Romero et al., 2020). Maslinic acid and bakuchiol were chosen as they have similar peak positions. The original data size of the compounds was at 1024×1024. The spectral widths were 12.991 ppm and 165.058 ppm for the ^{1}H and ^{13}C dimensions, respectively. The resolutions were 0.013 ppm/point and 0.161 ppm/point. For the best fitting results by REDEN, the scales of the spectrum were doubled to 2048×2048 using a script in the POKY Notepad titled "scale_spectrum_size_script.py" to make them 0.006 ppm/point and 0.081 ppm/point. Then, in each of the spectra, a cluster with similar positions were found as the main target. REDEN was then used to pick the peaks in those clusters. After recording their position and volumes, maslinic acid spectrum was concatenated onto bakuchiol spectrum via another script in the POKY Notepad titled "concatenate_spectra_script2.py". The spectra were collected under different conditions, we wanted to match intensity levels between them. Thus, we used the scale factor 2.4, the ratio between the highest volume of each compound. This scale factor created the most elusive combination between the maslinic acid spectrum and bakuchiol spectrum. Then, REDEN was performed using the Gaussian lineshape option on the same cluster region multiple times in hopes of identifying the same peaks and volumes found from the maslinic spectrum and bakuchiol spectrum. For troubleshooting, we would adjust the contour level, the position of the

cluster in window, and re-running REDEN multiple times to get the best decomposition prediction. Figure 2 shows the results of the REDEN testing on bakuchiol and maslinic acid.

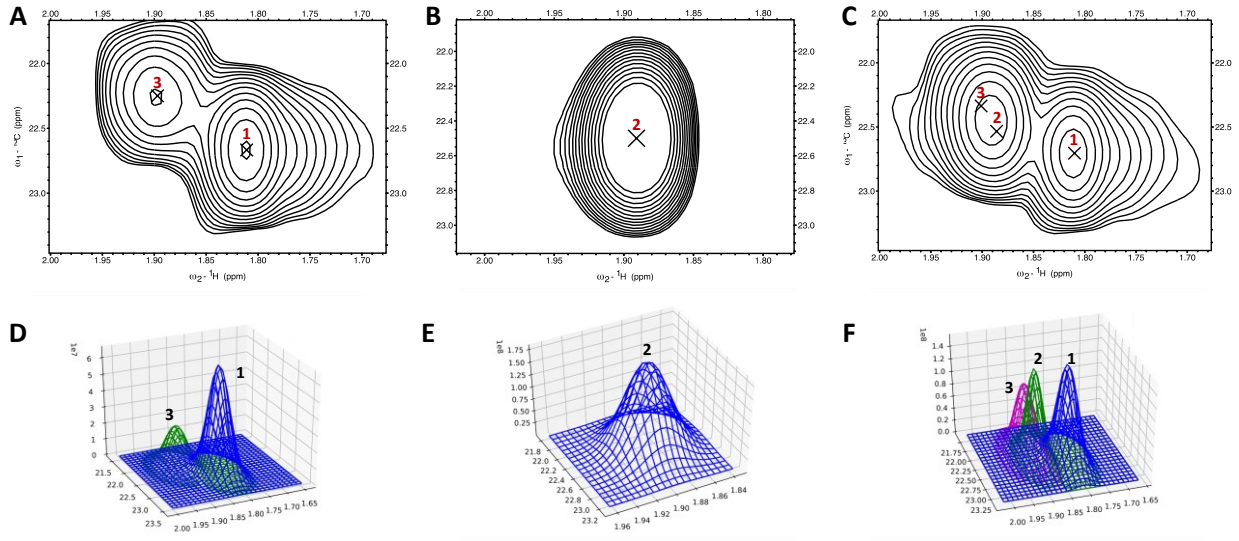

Fig. 2. Synthesizing a spectrum from bakuchiol and maslinic acid. (A) Target cluster with picked peaks in scaled spectrum of maslinic acid compound. (B) Target cluster with picked peaks in scaled spectrum of bakuchiol compound. (C) Target cluster with picked peaks in the synthesized concatenated spectrum of maslinic acid and bakuchiol. (D) REDEN's result of identifying peaks in maslinic acid spectrum (E) REDEN's result of identifying peaks in bakuchiol spectrum (F) REDEN's result of identifying peaks in the synthesized concatenated spectrum.

From the results (shown in Figure 2), REDEN exhibited an exceptional performance in effectively decomposition peaks. The r^2 value of the Gaussian fit done by REDEN for Figure 2D was 0.9779. The r^2 value of the Gaussian fit done by REDEN for Figure 2E was 0.9926. The r^2 value of the Gaussian fit done by REDEN for Figure 2F was 0.9901. Although these numbers can be used as a supplement, it is important to understand that high correlation values can also mislead to overfitting. Therefore, users must carefully consider other factors when making decisions. Unfortunately, the volumes in the standard spectra are not the same or like the volumes found in the concatenated spectrum. There are no correlation or patterns either as to why the volumes of those peaks are different. This is likely due to the nature of 2D NMR as it utilizes two resonances and two magnetizations meaning that its relaxation time is longer. Figures of each of the standard compound results can be found in the supplemental document.

3.2. REDEN testing on experimental protein spectra.

We tested REDEN on two 2D $^{1}H,^{15}N$-HSQC spectra of Nsp9 from SARS-CoV-2 that were collected on 400MHz and 700MHz Bruker NMR spectrometers. The 700MHz spectrum was from the Covid19-NMR (https://covid19-nmr.de) research partner, the Pastore group at King's College London (E et al., 2021), and we acquired the 400MHz spectrum ourselves in the similar condition. We expressed the Nsp9 protein in E. coli Bl21(DE3) using the pET28a-LIC-nsp9 plasmid (Littler et al., 2020), with cells incubated in ^{15}N M9 labeled media (McIntosh & Dahlquist, 1990) supplemented with Kanamycin at 200 rpm and 37°C until reaching an optical density of 0.9. Isopropyl. β-d-1-thiogalactopyranoside was added at a final concentration of 0.5 mM to induce expression, which was carried out at 37°C for 4 hours. Bacterial pellets were harvested, resuspended in Lysis buffer (L buffer) containing 20mM HEPES at pH 7.0, 150mM NaCl, 20mM Imidazole, 2mM $MgCl_2$, and 0.5mM TCEP, and then sonicated (20 pulses for 20 s ON at 60 W and 40 s OFF) on ice with 1mg of Lysozyme and 1mg of DNAase. The lysate was centrifugated at 10,000×g for 45 minutes, filtrated through a membrane of 0.2 um and the flowthrough was applied to a nickel affinity column (Hi-Trap chelating columns, GE Healthcare, Waukesha, WI, USA). After washing with 5 volumes of L buffer and 5 volumes of L buffer with 100mM of imidazole, the protein was eluted with 4 volumes of L buffer with 500mM Imidazole followed by His-tag removal through overnight incubation with PreScission 3C protease (Z03092 ©GenScript) at 4°C. Gel filtration (S75 16/60, GE HealthCare) was conducted using 50mM sodium phosphate pH 7.0, 150 mM NaCl, 1mM TCEP and 0.02% NaN_3. Nsp9 samples were prepared in 10% v/v D_2O deionized water (151890 Sigma-Aldrich®) at 280μM for $^{1}H,^{15}N$-HSQC NMR spectrum acquisition on a 400MHz Bruker BioSpin instrument at 298 K. NMR data were processed using

NMRPipe software (Delaglio et al., 1995), and data visualization was conducted using POKY (Lee et al., 2021).

The size of the 700Mhz spectrum was 1024×2048, and our spectrum was smaller after automatically zero-filled (128×512). Neither zero-filling nor linear prediction could improve the data quality. Therefore, we decided to apply the "scale_spectrum_size_script.py" in POKY Notepad which scaled the spectrum size 1024×1024 in the frequency domain. We used the BMRB entry number 50622 to assign spectra. The chemical shifts were loaded via the resonance tab in the POKY suite. The downloaded assignments were then transferred onto the spectra via the transfer and simulate tool (two-letter code "*ta*") in the POKY suite. We found a cluster region that could be a good example in the 400MHz spectrum as comparison to the 700MHz. Then, REDEN

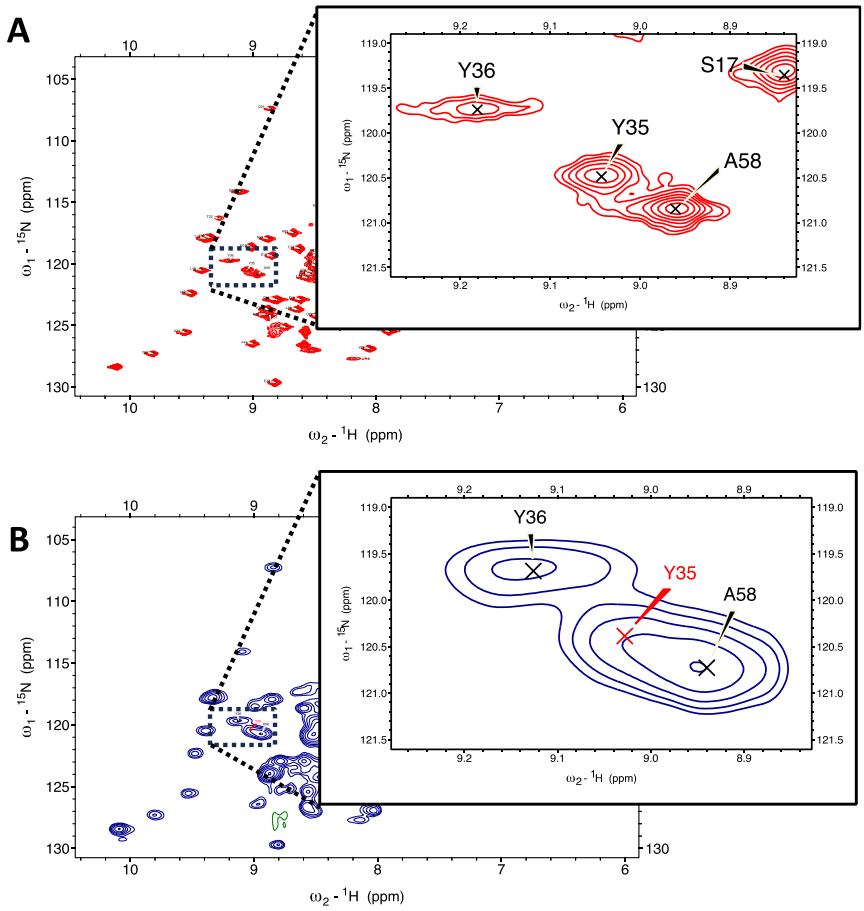

Fig. 3. NSP9 spectra taken with a 700MHz Bruker instrument with a cryo-probe and the other with a 400MHz Bruker instrument with a room temperature probe (A) The 700MHz spectrum with assignments with a cluster focused for comparison. (B) In the 400MHz spectrum, REDEN was able to accurately identify the same peaks found in the same cluster as the 700MHz spectrum. The peak in red was not identified by the simple peak picking algorithm in the POKY suite.

was run on the cluster region with the Gaussian line shape option to see if it can correctly identify the peaks that should be in the same position as the ones in the 700MHz spectrum.

Figure 3 shows the two spectra with differing quality. As a result, REDEN was able to accurately recover the peaks in that cluster which shows that low magnetization power of the NMR spectrometer can be overcome with peak decomposition. However, S17 peak was not recovered because it wasn't able to be captured by the low magnet power, temperature variations between 400 MHz and 700 MHz, sample degradation or some unknown reasons. Still, for the peaks that

were obtained, REDEN was able perform flawlessly which further shows REDEN's powerful advantage in the field of proteomics.

3.3. Notable observations while troubleshooting

In general, a cluster can be quite difficult to decompose especially when there are more than 10 peaks in the cluster. Not only does the user need to fit the whole cluster in the spectral window, but the overall process can be quite time consuming due to the large cluster size. Thus, because of this, it is possible for REDEN to give inaccurate parameter information based on the spectrum. Because REDEN only uses a small cluster region in the spectrum, high-resolution data is essential for better performance. In cases where the number of points in the spectrum is sparse, we recommend users to use resolution enhancement methods like extrapolation on time domain data such as zero-filling (Bartholdi & Ernst, 1973) and linear prediction (Ernst & Anderson, 2004), and interpolation on the frequency domain data like nearest-neighbor algorithm for upscaling as we have shown in this paper. This method can be easily accessible from POKY Notepad's "scale_spectrum_size_script.py" as mentioned above. However, as the number of time-domain points decreases, the lineshape observed in the Fourier spectrum becomes increasingly non-ideal, with increasingly prominent truncation artifacts. Therefore, upscaling method will not be able to rescue the data. It is recommended for users to run multiple times to get the best parameters with low residuals even though it may be tedious. Also, because of 2D NMR spectra are not always proportional to the concentration of samples, users should be cautious when utilizing the volume calculated by REDEN. However, once an extrapolation technique like $HSQC_0$ is implemented for raw spectra, it will be possible for REDEN to approximate the concentration based on the volume with more confidence (Hu et al., 2011). Users may get better results by running multivariate analysis methods like PCA with REDEN's decomposition power compared to traditional binning

or region-of-interest (ROI) approaches. Additionally, REDEN currently does not support full automation on a whole spectrum which may increase significant running time. However, we highly recommend that users pair REDEN with one of our automatic peak picking tools like iPick that also provides a reliability score for each peak picked. So, users only need to focus on the major clusters that are most likely to have hidden peaks that are overshadowed.

3.4. Simulating a spectrum of modeled peaks and residuals

We tested the performance of REDEN by running multiple fit numbers on the synthesized concatenated spectrum of bakuchiol and maslinic acid from the BMRB database. A range of 1 – 5 peaks were fitted on a targeted cluster in the original concatenated spectrum. Then, to visualize its residual plot, a new tool was created that makes simulated and residual spectra via the two-letter-code “mr” in the POKY suite. This tool creates two-dimensional contour simulated and

residual spectrum of the original spectrum with its fitted peaks. 3D perspective plot was a tool created to visualize those in the POKY suite via two-letter code "*3p*".

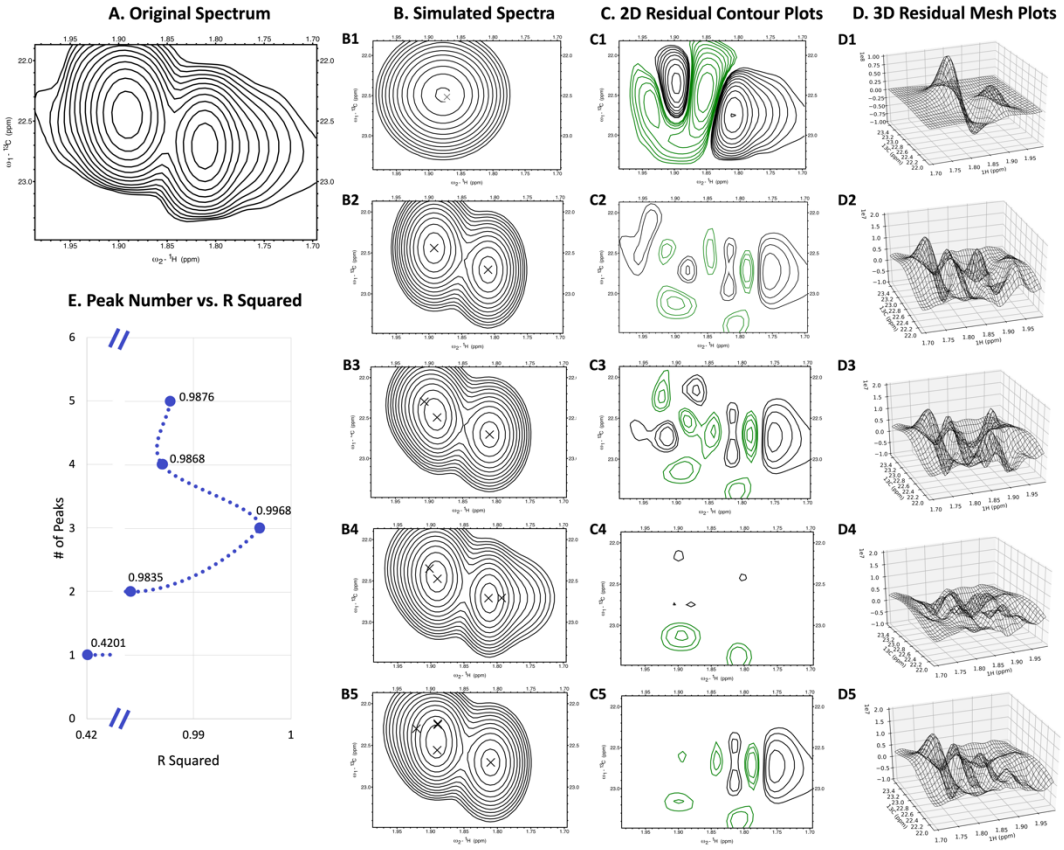

Fig. 4. Simulating a spectrum from concatenated mixture of bakuchiol and maslinic acid. (A) Target cluster of original mixture spectrum scaled to 2048×2048 (B) Target cluster of simulated mixture spectrum with peaks picked from a range of 1 – 5. (C) Target cluster of residual plot between original and simulated spectra. (D) Three-dimensional wireframe plots of residual plots. For D2-5, the Z-axis range are the same, but for D1, its Z-axis range 10 times larger. (E) Plot of each fitting and their R^2 value.

REDEN was done on the original spectrum to obtain the fitted picked peaks. Because REDEN always seeks the optimal parameters, the fit parameters from the previous fit were not used for the next fit number. The original spectrum with the fitted peaks were then used to create the two-dimensional simulated and residual contours as shown in Figure 4 as B and C. Three-dimensional wireframe plots of the residual contours were made from the 3D perspective plot tool in the POKY suite. The outcomes indicate that fitting three peaks yielded the highest R^2 value, as demonstrated in Figure 4E. This emphasizes the precision of REDEN in its predictions, rather than simply attempting to fit more peaks without proper evaluation.

Moreover, the current line shapes fall short of accurately modeling actual NMR signals perfectly, which means that the volume cannot accurately reflect the concentration of the sample. Still, we remain open to adopt where line shapes have high fidelity or when machine learning approaches have been advanced to mitigate this limitation.

4. Conclusion

Our team has created REDEN, an intuitive tool designed for visually assisting picking of shoulder peaks or smaller peaks that may be obscured by high-intensity neighboring peaks. One of the most unique and powerful features of REDEN is its ability to use subregions in POKY interactively. With this feature, users can pick even the smallest peaks that may be hidden by the neighboring high-intensity peaks, making it an invaluable preference for researchers working with complex data sets especially with metabolite spectral data. We show protein and metabolite data as examples in this manuscript because we are more interested in biomolecules, but we believe REDEN can also be applied to spectra from organic compounds and inorganic materials if lineshapes exhibit characteristics that REDEN can handle. REDEN can also be coupled with POKY's latest addition, TINTO (Two-dimensional Imaging for NMR sTrip Operation via CV; two-letter-code "*ti*" for 2D and "*sp/SP/Sp*" for 3D) (Giraldo et al., 2023). By employing TINTO for peakless strip matching via computer vision method, REDEN can aid in deciphering intricate regions during protein assignments. Subsequently, versatile assigner, represented by two-letter code "*va*", can be used for semi-automatic residue assignments in conjunction with reference views using two-code "*ir*" for supplementary referencing(Manthey et al., 2022). This interactive subregion capability sets REDEN apart from other peak picking tools and makes it an essential addition to any researcher's toolkit. This program is open-source and can be utilized as a plugin in POKY in conjunction with the iPick peak picker. The latest version of POKY already includes REDEN and iPick, and it is

freely available to non-commercial users on Windows, Linux, and macOS, including Apple Silicon CPUs (e.g., M1, M2). It is recommended to use peak picking software like iPick alongside with REDEN as shown in this paper, and we also plan to make an interface to UnidecNMR to seek greater synergies (Buchanan et al., 2022).

5. Acknowledgements

This work was supported by the National Science Foundation (Grant No. DBI-2051595 & DBI-1902076 to W.L.) and the University of Colorado Denver. J.-H.Y is supported by NIH R01AA027532 and Startup fund from University of Oklahoma. Authors appreciate Dr. Jamie Rossjohn from Monash University for providing with pET28a-LIC-nsp9 plasmid. Authors also appreciate the Annalisa Pastore group from King's College London and the Covid19-NMR (https://covid19-nmr.de) for making high quality Nsp9 data available.

Conflict of Interest: none declared.

References

Ahlner, A., Carlsson, M., Jonsson, B.-H., & Lundström, P. (2013). PINT: a software for integration of peak volumes and extraction of relaxation rates. *Journal of Biomolecular NMR*, *56*(3), 191-202. https://doi.org/10.1007/s10858-013-9737-7

Bartholdi, E., & Ernst, R. R. (1973). Fourier spectroscopy and the causality principle. *Journal of Magnetic Resonance (1969)*, *11*(1), 9-19. https://doi.org/https://doi.org/10.1016/0022-2364(73)90076-0

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, *6*(3), 277-293. https://doi.org/10.1007/BF00197809

Dudley, J. A., Park, S., MacDonald, M. E., Fetene, E., & Smith, C. A. (2020). Resolving overlapped signals with automated FitNMR analytical peak modeling. *J Magn Reson*, *318*, 106773. https://doi.org/10.1016/j.jmr.2020.106773

E, F. D., Puglisi, R., Korn, S. M., Alfano, C., Bellone, M. L., Piaz, F. D., Kelly, G., Monaca, E., Schlundt, A., Schwalbe, H., & Pastore, A. (2021). Backbone chemical shift spectral assignments of SARS coronavirus-2 non-structural protein nsp9. *Biomol NMR Assign*, *15*(2), 235-241. https://doi.org/10.1007/s12104-021-10011-0

Ernst, R. R., & Anderson, W. A. (2004). Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments*, *37*(1), 93-102. https://doi.org/10.1063/1.1719961

Giraldo, A. E. L., Werner, Z., Rahimi, M., & Lee, W. (2023). Breaking boundaries: TINTO in POKY for computer vision-based NMR walking strategies. *Journal of Biomolecular NMR*. https://doi.org/10.1007/s10858-023-00423-6

Hu, K., Westler, W. M., & Markley, J. L. (2011). Simultaneous quantification and identification of individual chemicals in metabolite mixtures by two-dimensional extrapolated time-zero (1)H-(13)C HSQC (HSQC(0)). *J Am Chem Soc*, *133*(6), 1662-1665. https://doi.org/10.1021/ja1095304

Klukowski, P., Augoff, M., Zieba, M., Drwal, M., Gonczarek, A., & Walczak, M. J. (2018). NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. *Bioinformatics*, *34*(15), 2590-2597. https://doi.org/10.1093/bioinformatics/bty134

Kraft, D. (1988). *A Software Package for Sequential Quadratic Programming*. Wiss. Berichtswesen d. DFVLR. https://books.google.com/books?id=4rKaGwAACAAJ

Lee, W., Cornilescu, G., Dashti, H., Eghbalnia, H. R., Tonelli, M., Westler, W. M., Butcher, S. E., Henzler-Wildman, K. A., & Markley, J. L. (2016). Integrative NMR for biomolecular research. *J Biomol NMR*, *64*(4), 307-332. https://doi.org/10.1007/s10858-016-0029-x

Lee, W., Rahimi, M., Lee, Y., & Chiu, A. (2021). POKY: a software suite for multidimensional NMR and 3D structure calculation of biomolecules. *Bioinformatics*, *37*(18), 3041-3042. https://doi.org/10.1093/bioinformatics/btab180

Li, D. W., Hansen, A. L., Yuan, C., Bruschweiler-Li, L., & Bruschweiler, R. (2021). DEEP picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra. *Nat Commun*, *12*(1), 5229. https://doi.org/10.1038/s41467-021-25496-5

Littler, D. R., Gully, B. S., Colson, R. N., & Rossjohn, J. (2020). Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9. *iScience*, *23*(7), 101258. https://doi.org/10.1016/j.isci.2020.101258

Maciejewski, M. W., Schuyler, A. D., Gryk, M. R., Moraru, II, Romero, P. R., Ulrich, E. L., Eghbalnia, H. R., Livny, M., Delaglio, F., & Hoch, J. C. (2017). NMRbox: A Resource for

Biomolecular NMR Computation. *Biophys J*, *112*(8), 1529-1534. https://doi.org/10.1016/j.bpj.2017.03.011

Manthey, I., Tonelli, M., Ii, L. C., Rahimi, M., Markley, J. L., & Lee, W. (2022). POKY software tools encapsulating assignment strategies for solution and solid-state protein NMR data. *Journal of Structural Biology: X*, *6*, 100073. https://doi.org/https://doi.org/10.1016/j.yjsbx.2022.100073

McIntosh, L. P., & Dahlquist, F. W. (1990). Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q Rev Biophys*, *23*(1), 1-38. https://doi.org/10.1017/s0033583500005400

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P. C., Timony, M. A., Ottaviano, M., & Sliz, P. (2013). Collaboration gets the most out of software. *Elife*, *2*, e01456. https://doi.org/10.7554/eLife.01456

Rahimi, M., Lee, Y., Markley, J. L., & Lee, W. (2021). iPick: Multiprocessing software for integrated NMR signal detection and validation. *J Magn Reson*, *328*, 106995. https://doi.org/10.1016/j.jmr.2021.106995

Romero, P. R., Kobayashi, N., Wedell, J. R., Baskaran, K., Iwata, T., Yokochi, M., Maziuk, D., Yao, H., Fujiwara, T., Kurusu, G., Ulrich, E. L., Hoch, J. C., & Markley, J. L. (2020). BioMagResBank (BMRB) as a Resource for Structural Biology. *Methods Mol Biol*, *2112*, 187-218. https://doi.org/10.1007/978-1-0716-0270-6_14

Waudby, C. A., Ramos, A., Cabrita, L. D., & Christodoulou, J. (2016). Two-Dimensional NMR Lineshape Analysis. *Sci Rep*, *6*, 24826. https://doi.org/10.1038/srep24826

Wurz, J. M., & Guntert, P. (2017). Peak picking multidimensional NMR spectra with the contour geometry based algorithm CYPICK. *J Biomol NMR*, *67*(1), 63-76. https://doi.org/10.1007/s10858-016-0084-3