Creating a Protocol for RunAbout to Assigning HSQC Spectra of PDZ-1 Domain of the Protein ZO-1

Joshua Taylor
Advisor: Ariane Jansma

Abstract

PDZ domains are found in most signaling proteins for viruses, plants, bacteria, and animals. They help to recruit other proteins to build systems, and localize signaling complexes. Viral proteins have been known to target PDZ domains through PDZ binding domains. In order to identify where and how these viral proteins attach to the PDZ domains on the zonula occludens-1 (ZO-1) protein, a ^{15}N-^{1}H HSQC spectrum was taken of PDZ from ZO-1 to map out the amino acids in the protein. This project sought to create a working and efficient protocol for using the program RunAbout to assign amino acid information to 2-dimensional NMR spectra for the PDZ domain of the protein ZO-1. The results allow not only further study of the binding of ZO-1 to proteins expressed by flaviviruses, but also offer a platform for future structural analyses when bound with other viral proteins.

Introduction

PDZ (PSD-95/DLG/zonula occludens-1) domains are modular protein-binding motifs that are found in viruses, plants, bacteria, and animals. The name PDZ comes from the first three proteins that were identified to contain these domains: PSD-95 (signals at the postsynaptic density), DLG (the *Drosophila melanogaster* Discs large protein), and ZO-1 (zonula occludens-1 which maintains the epithelial polarity) (1). PDZ domains are extremely diverse, but all retain the same general structures. They all contain 5-6 β-sheets and 1-2 α-helices. PDZ-1 from ZO-1 is roughly 104 amino acids long and contains 1 α-helices and 5 β-sheets (Figure 1).

PDZ domains are commonly found on protein complexes responsible for scaffolding complexes, cytoskeleton structure, cell polarity, and cell signaling. Many of these proteins are found or associated with the cell membrane (2). These domains act as recognition sites for protein-protein interactions that play a key role in the organization of diverse cell signaling assemblies, specifically recognizing short C-terminal peptides or internal peptides that are structurally similar to the C-terminus (1). They organize multiprotein signaling complexes to allow cell-to-cell communication. They can achieve this function in a few ways, specifically through cell polarity, transporting and targeting proteins.

PDZ domains have been found to be especially important in organizing tight junctions (2). Tight junctions are cell-to-cell connections that join together to form a barrier which stops the movement of ions, macromolecules, and cells across the endothelia and epithelia (3). The most relevant protein, in terms of PDZ domains, to the localization of tight junctions are the zonula occludens proteins, one of which is zonula occludens-1 (ZO-1). The ZO-1 protein contains three different PDZ domains. It has been found that when the PDZ1-3 are knocked out, the ZO-1 polypeptides are unable to assist in the creation of the tight junctions (4).

Because PDZ domains are heavily involved in organizing many signaling and structural properties of cells, they are common targets for pathogenic viruses. Many viruses (hepatitis B and C, human T cell leukaemia virus type 1,

Kaposi sarcoma herpesvirus, human immunodeficiency virus, high-risk human papillomaviruses and adenovirus type 9, tick borne encephalitis virus, dengue virus) encode in their genome the ability to target specific cells containing PDZ domains (5). The viruses use a similar short peptide PDZ binding motif (PBM) to bind to the PDZ domains. They target a common variant of PDZ domains that are involved in cell polarity and tight junctions. More specifically, dengue virus nonstructural protein 5 (DENV-NS5) primarily targets the PDZ domains of ZO-1 (6).

Knowing the structure of this PDZ domain, allows us to accurately identify where flaviviruses bind to and subsequently inhibit these structures. Therefore, it is imperative to accurately and efficiently assign the HSQC spectra received from past experiments. In order to do this, a working and efficient protocol for using the RunAbout program was developed.

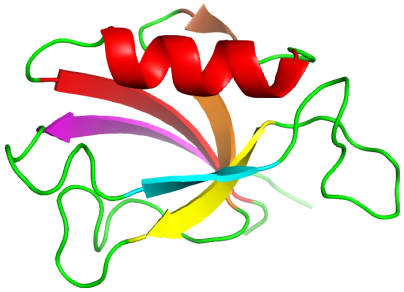

Figure 1. The structure of PDZ1 of ZO-1. PDZ1 contains 5 β-sheets and 1 α-helix.

Material and Methods

Creating a Project in NMRViewJ

In order to analyze the past NMR experiments a project in NMRViewJ by OneMoon Scientific must be created. To do this, open up NMRViewJ, click on "File," then "Projects," and "New." This project was named "PDZ010821." *NOTE* save your project after every step taken, the program has a tendency to crash. This is a link to a very quick and rough tutorial by NMRView: https://nmrfx.org/tutorials/nmrviewj.

Opening Datasets

Once the project is created, we must bring in the five NMR experiments that are relevant to our project: HSQC, HNCACB, HNCACO, HNCO, and CBCACONH. This is done via the "Datasets" tab. Click on the "Datasets" tab and then "Open Datasets." This prompts you to find the NMR experiments, these files end in a ".nv" file format.

Formatting Datasets

Once all the files have been brought into your project we must format each dataset so that we can pick peaks efficiently later. Formatting datasets is an objective process that looks different for each experiment that you have. However, the same ideas are applied to each dataset. First we must draw each dataset. In order to do this, click on "Datasets" and then "Datasets Table." This will bring up a table with all your datasets that are opened in this project. To draw the experiments, click on the experiment you want to format and then click the draw button. This should produce a graph with peaks on it. In order to efficiently and correctly pick the right amount of peaks, we need to set a reasonable level for our dataset. To do this, click on "Attributes" on the top left of the graph, also depicted as a gear. This will bring up another menu where you can set the level of the dataset (insert figure). This is where setting the level becomes objective. We want a low enough level to see all the relevant peaks, but not too low as to be muddled in the noise of the spectrometer. (insert figure of too much noise and "right" level). To change the level of the dataset, there are three arrows you can click, "Auto," "Higher," and "Lower." There is also a numerical way to change the level by entering a number under the "Level" box in the attributes window. Once an adequate level is chosen for

that dataset, we need to view all the layers of the dataset. Click on the "View" tab and then on the "Z." This will bring up a menu, click "Full." This will show all the layers for that given experiment. Repeat this process for each 3D dataset. For the HNCACB there will be negative peaks that you need to enable. To do this, in the attributes window click on the box next to the "-Color" box and click "Refresh." This will show all the negative and positive peaks in the graph.

Creating Peak Lists for Datasets

Now that the datasets are formatted correctly, peaks can be picked. Click on the "PeakPick" tab in the attributes window. Then type in a relevant name for the peak list, change the "Thickness" to three and click the yellow "Pick" box. After a few seconds in the box labeled "# picked," the amount of peaks will be shown. Repeat this process for each of your datasets.

Loading a Sequence File

We also need to add the sequence of our protein into NMRView. In order to do this, we must first use a text editor to write out each amino acid in its three letter abbreviation; each amino acid gets its own line. Save the file as a ".seq" file and bring it into NMRView by clicking on "Molecule," then "Read/Write Topology," and then "Sequence File." This will prompt you to choose a file, simply use the one just created. To check that this worked, click on "Assign," then "Sequence," then "Options," and then "Apply." This should show your full sequence once apply is clicked.

Setting Up the Parameters for RunAbout

Now that the datasets have been formatted and the peaks have been picked, it is time to use RunAbout. In order to access RunAbout, click on the "Analysis" tab and then click on "RunAbout." This will bring up the RunAbout window on the "Parameters" tab. Click on the "Setup Peak Lists…" which will bring up the peak list window. For our experiments, the HSQC will be the "RefList." If the peak lists were named the same as the experiments, meaning, for the HNCO experiment, the peak list is called "hnco," click on the "Guess Lists" button. This will automatically fill in the lists that have been created and assign them to the corresponding experiments. However, if this was not the naming convention, manually go through and select each experiment from the list on the left and match them with the corresponding peak list created for that dataset. The "RefList" should be "hsqc." Once the correct experiments have been selected with their corresponding peak lists, click "Status" to refresh the window and then close the window. Now the tolerances need to be set. This is achieved by clicking on the "Auto" button to the left of the "Set" button. This automatically changes the tolerances based on the code written for RunAbout that accurately chooses the tolerance levels based on the given experiment.

Setting up the View for RunAbout

Now that the parameters for RunAbout have been set, the View needs to be set. To do this, click on the "View" tab in the RunAbout window. In this section, make sure that the active boxes of the "cb" and "c" are checked. For the "Full" column make sure all the boxes are checked. Then confirm that the "N Wings On" is checked as well. To assure that everything is set up correctly, click on the "Mode" button and then click on "Edit Peaks." If done correctly, there should be two rows of four boxes.

Editing Peaks via RunAbout

There are two ways to edit the peaks for the experiments: manually (safe) and automatically (risky). To manually edit the peaks for each given peak on the HSQC, use the "Edit Peaks" mode and then click on "Helm." In the helm

view, there are arrows that move through the associated peaks in the reference list, the HSQC, that are associated with the 3D experiments. When manually editing the peaks, identify for each peak on the HSQC if there are peaks in the 3D datasets that were either picked that shouldn't have been or if there are peaks that weren't picked that should have been. This part is objective since the level at which the peaks are chosen is set by the researcher. To automatically edit the peaks, click on the "Actions" tab next to the "Helm" tab. From this menu, click on the "Filter Peak List" drop down bar and click on the HSQC peak list. Once that is selected, click "Filter." This filters all the peaks from the 3D datasets against the HSQC. This will remove or add any peaks that are far from the HSQC or closely linked to it respectively. Once that is down, press the "Compress/Degap" button.

Editing Clusters via RunAbout

Now that the peaks have been properly edited, it is time to edit the clusters. Clusters are the groups of peaks from each 3D experiment that correlate to the HSQC based on N, H, and C ppm. To create these clusters, click on the "Actions" tab and then click on the "Cluster Peaks" button. This will cluster all the relevant peaks together. Change the mode to "Edit Clusters (HN-C)" and then go back to the helm view. This will bring a different view than the "Edit Peaks" mode. (insert figure). This mode will show the peaks in the cluster and also identify any peaks that the program believes should be added to this cluster, shown by yellow arrows. It will also show any peaks that it believes should not be a part of the cluster based on the amount of peaks that are already present in the cluster, shown by a red arrow pointing at the peak. In order to add peaks to a cluster, click on the box with the yellow arrow(s) and press the "z" key on the keyboard. This will zoom into that box and allow you to add peaks to the cluster by hovering the cursor over the desired peak and pressing "a" on the keyboard. This will successfully add a peak to the cluster. Deleting peaks is accomplished in the same way, but by pressing "d" instead of "a."

Editing Links via RunAbout

Now that the clusters have been created, it is time to create links between each of the clusters. The editing links mode finds clusters that correspond to residues which are adjacent to each other in the sequence. Change the mode to "Edit Links." This will bring up a new layout that allows you to link the clusters together (insert figure). In this mode it shows what cluster it is on and what the possible linkable clusters are to that cluster. There are two windows to show this, a left and a right window. The left shows the cluster that would be before the cluster being viewed and the right shows the cluster following the current cluster. In each window there are five columns. The first column is the cluster number that the current cluster being viewed is most closely related to. The second column shows the "score" that this possible link received based on the number of carbons that matched up between the i and i-1 for the residues. The third column is the number of carbons that matched between the two residues. The fourth column is the cluster number that is most closely related to the cluster in the first column. Since this is in the left window, the first column represents the residue that is most likely before the cluster we are viewing, making the cluster number in the fourth column the most likely cluster to come after the cluster in the first column. In an ideal world, the cluster in the fourth column will match the cluster that is currently being viewed in the helm. The fifth column has the possibility to show three letters: R, A, and V. The R stands for a reciprocal match. This is the idea that the cluster in the fourth column will match the cluster that is currently being viewed in the

helm. If this is true, there will be an R present in the fifth column. A stands for availability of the cluster. If this cluster has not already been linked to another, then it will show an A, whereas if it has already been linked there will be no A present in the column. The final letter, V, stands for viable. This is present if the link that is being made is viable, meaning that the chemical shifts of the fragment that is being created once the clusters are linked, matches the expected chemical shifts for the given amino acids in the protein sequence. If there is no place in the sequence where the combination of those chemical shifts is present, then the V will not appear making it not a viable link. In order to make a confident link, all three letters should be present with a high score in the second column and the max number of carbons in the third column. In order to make a link, click the green check mark. Now look in the window on the right to make a connection with the cluster in front of the cluster that is currently being viewed. Continue this until most of the sequence is linked together.

Results and Discussion

Results

Roughly 75% of the protein has been linked in some way; meaning, in total 76 amino acids have been linked together in some combination, but not all 76 amino acids form one continuous link. This leaves 27 residues that have not been linked to create the entirety of the sequence.

Discussion

This study created a working and efficient protocol for assigning the 15N-1H-HSQC of PDZ 1 from ZO-1. Although the entirety of the HSQC spectra was not assigned, great progress was made having roughly 75% of the sequence linked in some fashion. This project gives a good starting point to finish the assignment so that further studying of PDZ1 from ZO-1 can take place. With the fully assigned spectra, we will be able to analyze the binding mechanics of the flaviviruses TBEV and DENV with their PBMs.

References

1. Baruch Z. Harris and Wendell A. Lim Mechanism and role of PDZ domains in signaling complex assembly. *Journal of Cell Science, 114,* (2001): 3219-3231

2. Alan S. Fanning, Ming F. Lye, James M. Anderson, and Arnon Lavie Domain Swapping within PDZ2 IS Responsible for Dimerization of ZO Proteins. *Journal of Biological Chemistry, 282,* (2007): 37710-37716

3. Anderon, J. M., Van Itallie, C. M., and Fanning, A. S. (2004) *Curr. Opin. Cell Biol.* 16, 140-145

4. Umeda, K., Ikenouchi, J., KAtahira, S., and Furuse, M. (2006) *Cell 126,* 741-754

5. Ronald T. Javier and Andrew P. Rice Emerging Theme: Cellular PDZ Proteins as Common Target of Pathogenic Viruses. *Journal of Virology 85,* (2011): 11544-11556

6. Karin Ellencrona, Asim Syed, and MAgnus Johansson Flavivirus NS5 associates with host-cell proteins zonula occludens-1 (ZO-1) and regulating synaptic membrane exocytosis-2 (RIMS2) via an internal PDZ binding mechanism. *Biol. Chem 390,* (2009): 319-323