

# Investigation into the Amino Acid Characteristic in Intrinsically Disordered Proteins

Samantha C. Patterson, Kaylan Schilling, Grant Bowman PhD.

Abstract:

*Caulobacter crescentus*, which belongs to the class of Alphaproteobacteria, has a highly organized cytoplasm that includes complex macromolecular structures at the cell poles. These polar complexes are assembled through the activity of a scaffolding protein, known as the Polar Organizing Protein Z (PopZ). PopZ interacts with other proteins via an N-terminal domain that is mostly intrinsically disordered, and contains, an abundance of proline, glutamate, and aspartate (P,E, and D) residues. Surprisingly, scrambling the order of amino acids in the PED region does not affect PopZ function, suggesting that binding affinity and specificity are not dependent on interactions with individual amino acids within this low-complexity region. The biochemical qualities of P,E and D promote intrinsic disorder, but they are not the only types of amino acids that are found in unstructured protein domains. By making targeted changes in the PopZ sequence, we are asking if binding affinity and specificity are dependent on qualities associated specifically with P,E, and D (for example, negative charge), or if any form of intrinsic disorder is sufficient to support a functional PopZ binding domain. One change we are making is to substitute all of the prolines for serine. While this will certainly affect the hydrophobicity of the protein, our *in silico* analyses predict that the change will preserve intrinsic disorder. Another change will be to substitute the aspartate and glutamate residue for lysine, thus reversing the electrostatic charge with an amino acid that promotes intrinsic disorder.

Introduction:

Protein structure has four levels of organization. The primary level is the sequence of protein monomer amino acids and have what is called the N-terminus and the C-terminus. Secondary structure organization happens when the amino acids interact with one another to form either alpha-helix structures or beta-sheet structures. The secondary structures are formed from chemical interactions between the amino acids. Tertiary structure comes from the amino acids forming bonds that cause the folding of the protein into the 3-D structure of the protein. Quaternary structure happens when more than one protein forms a polyprotein. The best example is hemoglobin that has four proteins that make up the poly protein and the quaternary structure is the interactions between the protein chains that keep it as a polyprotein. These are the more traditional forms of structure.

Proteins with a different structure those that do not have structure until they bind a ligand; these are called intrinsically disordered proteins. These proteins appear to have no order beyond the primary amino acid sequence until they bind their ligand and can have a different form every time they binds. The primary amino acid sequence can influence whether or not a protein is intrinsically disordered or not. There have been studies looking into what amino acids are present in the primary sequence area of intrinsically disordered proteins throughout all domains of life. The amino acids found in these areas of proteins were proline, serine, glutamate, glutamine, lysine, alanine, and glycine. Aspartate threonine and arginine are all amino acids that can still be found in these regions but less frequently than the others.

Intrinsically disordered have the ability to bind multiple ligands in the same region and structured proteins have different binding areas to bind different ligands. This trait makes intrinsically disordered proteins good candidates for organizing cell polarity. The alphaproteobacteria *Caulobacter crescentus* is shown to divide asymmetrically so that one cell has a stalk that allows it to adhere to surfaces. Another cell that is called a swarmer cell and has a flagellum is also produced. An intrinsically disordered protein that helps with the organization that allows for this kind of division is called Polar Organizing Protein Z (PopZ).

PopZ has a specific domain that is called the PED region for the presence of the amino acids proline, aspartate, and glutamate and these make up the intrinsically disordered region of the protein. There is a structured alpha helix region that is called the Molecular Recognition Feature (MoRF) which helps bind other proteins. PopZ is capable of forming a scaffold network at the poles of the cell, which recruits PopZ's multiple binding partners to the poles. The protein's C-terminus allows for the formation of the scaffold and localizing PopZ to the pole. PopZ is capable of binding multiple partners because of its PED region. Our investigation examined the amino acid characteristics of the PED region. Creating constructs with different amino acids we observed whether or not they were capable of binding with known binding partners of the wild type PopZ.

#### Materials and Methods:

The constructs made for this investigation had PED region changes as follows: A scrambled PED sequence, a half scrambled PED sequence, a PED only composed of glycine and serine residues, and a PED only composed of proline and alanine residues. All constructs having the rest of the

protein with a wild type characteristic. To make these constructs, the computational system Serial Cloner was used to make the codon changes necessary for the glycine/serine construct and the proline/alanine construct. We made the scrambled and half scrambled and scrambled constructs by taking the amino acid sequence for the PED region from Serial Cloner and placing it into an online amino acid scrambler system. These scrambled amino acid sequences were then placed back in Serial Cloner to get a nucleotide sequence. The amino acid sequences of each construct were plugged into an online disorder prediction system to give an idea to how the disorder of the region would be affected. The sequences were then taken from Serial Cloner and submitted to a company that could produce and provide oligo nucleotide sequences. The oligo sequence would just be of the PED constructs. These oligos were then placed into a plasmid backbone that contained the other regions of the PopZ nucleotide sequence as well as fusing the protein to the fluorescent protein mCherry. The oligos were placed into the plasmid using a restriction digest isothermal cloning method. This plasmid along with another plasmid that has either ChpT or ParB fused to Green Fluorescent Protein (GFP) were transformed into *Escherichia coli* cells. ChpT and ParB are known binding partners for wild type PopZ. This is a dual plasmid expression system so that there is a different fluorescent protein tag for each protein of interest. The transformed cells were then grown overnight in 2 mL of Liquid Broth (LB) and this culture was diluted and grown for 2 hours. After 2 hours, they were checked for proper growth by observing the OD of the liquid culture. Then, 500 microliters of the culture were placed into a new tube and given the proper inducers for expression of the two plasmids (IPTG for PopZ plasmid and Arabinose for binding partner plasmid). The levels of inducers would have to be modified for each construct to provide optimal expression conditions. One microliter of the cells were then placed on an agarose gel on a slide for viewing under a fluorescent microscope. An image of each construct was taken and controls were used. The controls were wild type PopZ strain, Arabinose induction only, and deletion of the PED region strain.

#### Results and Conclusions:

In summary, all of the constructs were still able to bind with the two candidate binding partners. The foci were observed to be overlapping at the cell pole (Figures at end of paper). The overlap would give the appearance of a yellow foci and when the channels were turned off for each color signal, it showed that both green and red signals were present in the same place. The

computational data showed that the scrambled and proline/alanine constructs still have high disorder profiles compared to the wild type. The glycine/serine showed the largest decrease in the disorder profile of the PED region and half scrambled was partially reduced in disorder in the PED region. The computational data showed the glycine/serine and half scrambled constructs would be the least likely to bind to the candidates because those two have less disorder than the rest. However, in the images it was shown that these constructs still bound their candidate proteins. These results were unexpected because the idea is that the reason PopZ is capable of binding other proteins is because of this disorder. The expected result was that the green foci would either be completely eliminated or reduced and more diffuse green signal would be seen throughout the cell. The conclusions that could be drawn from this is that there is something else about the PED region that is allowing it to continue to bind to ChpT and ParB and co-localize them both to the pole.

#### Discussion:

Based on the computational data, there could possibly be another reason these constructs are working. Another assay can be done by making new constructs where all the prolines are changed to serines and another one that has the glutamate and aspartate residues replaced with lysine. The second construct can give some insight into how charge affects the binding. The glycine/serine and proline/alanine constructs do not have any charged residues but if instead of removing the charge, we change the kind of charge, this could possibly stop PopZ from being able to bind its partners. Computational analysis of these proposed constructs has already been done and shown a significant decrease in disorder for the construct with all the prolines being changed to serines in the PED region. There is still a relatively high disorder profile for the aspartates and glutamates changed to lysine (still charged just positively). The computational data is shown as supplementary figures. Further investigation needs to be done to understand how the intrinsic disorder is affecting PopZ's ability to bind its partners. The other thing to consider is that computational analysis, while helpful, does not necessarily imitate biology, which can be unpredictable and difficult to get expected results based on computational data. The importance in understanding this intrinsically disordered polar organizing protein is because it is important to human health. First, it can lead to a better understanding of how bacteria organize their cells. Some of the bacteria that form biofilms use those biofilms to adhere to human cells and cause

infections. Being able to disrupt this could help treat infections with bacteria that form these stocks. Also, there is formation of a flagella with the help of this kind of organization which is present in many infectious bacteria and being able to prevent movement could also prevent infection spread. Second, there is a protein in humans that is similar to PopZ in that it is a polar organizing protein and has an intrinsically disordered region, this protein is p53. The human p53 protein is known as a tumor suppressor protein and mutations in it can cause cancer. Understanding intrinsically disordered proteins can have applications to human health and cell biology since they seem to have a role in cell organization.

Figure Presentation:

Figure 1:

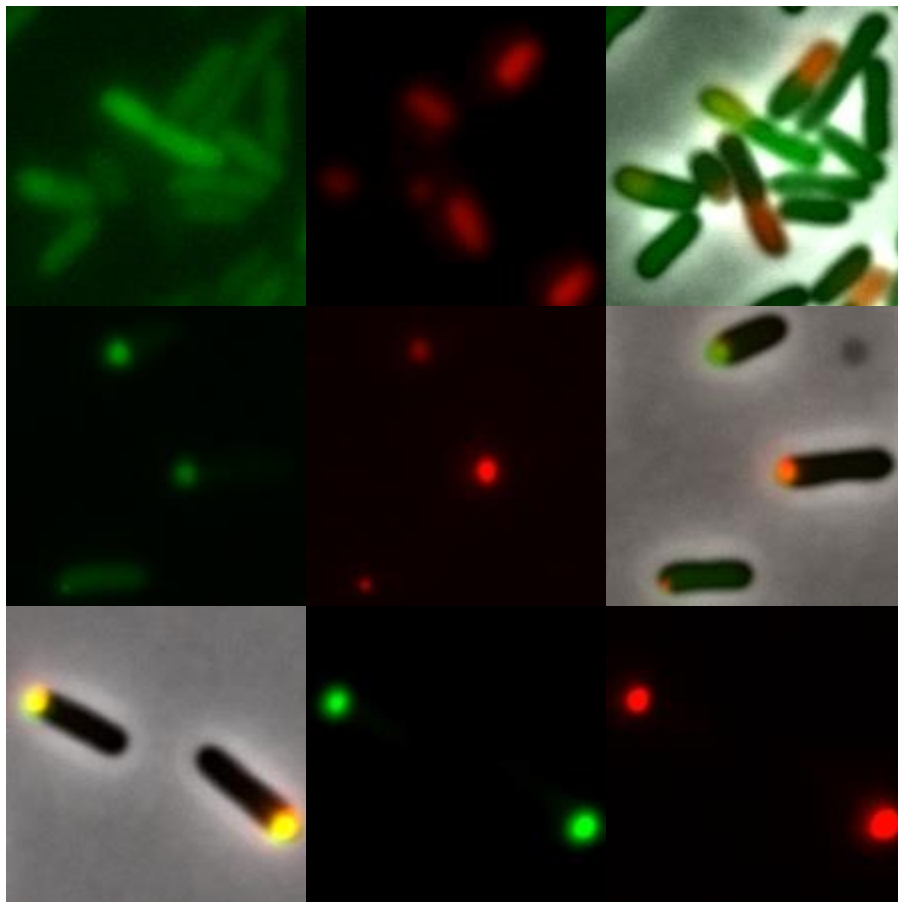


Figure 2:

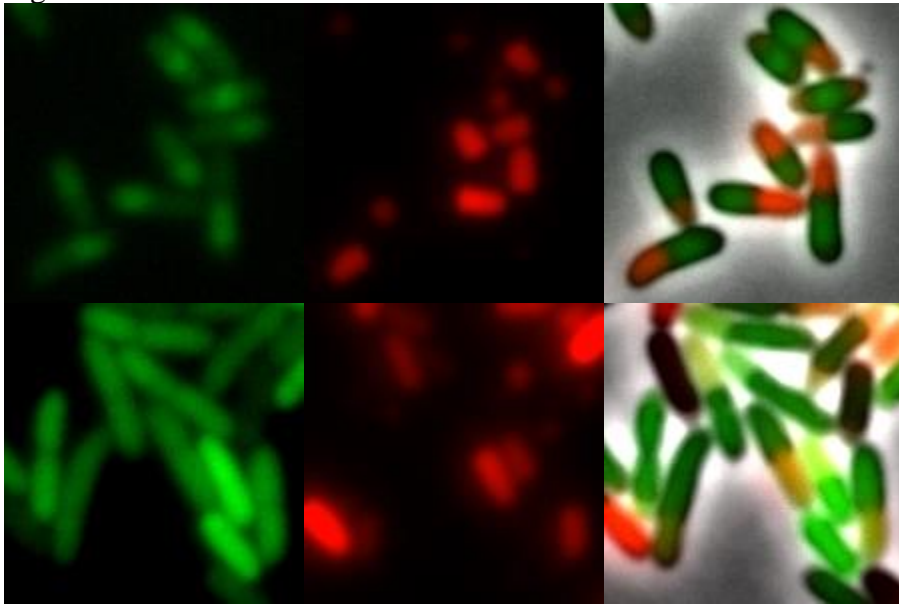


Figure 3:

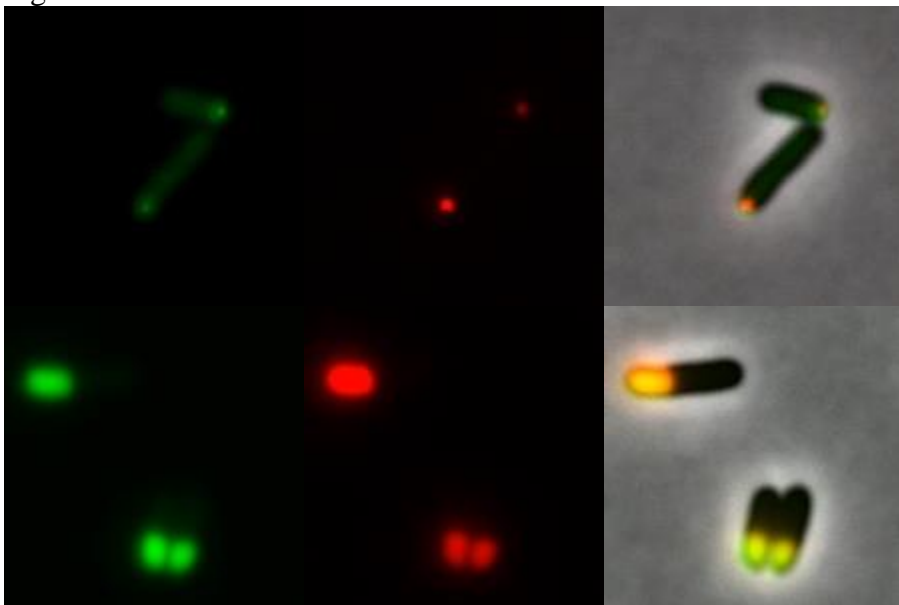


Figure 4:

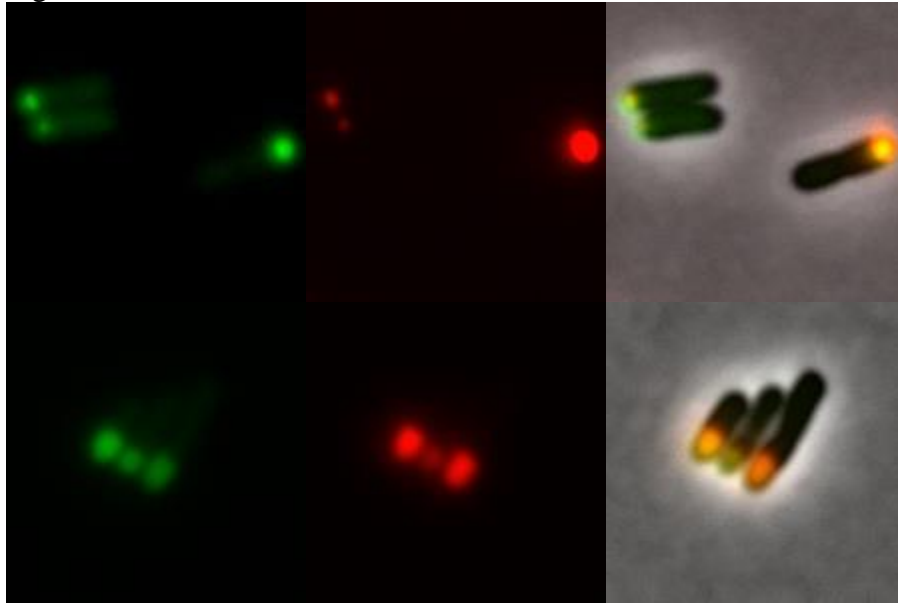


Figure. 5:

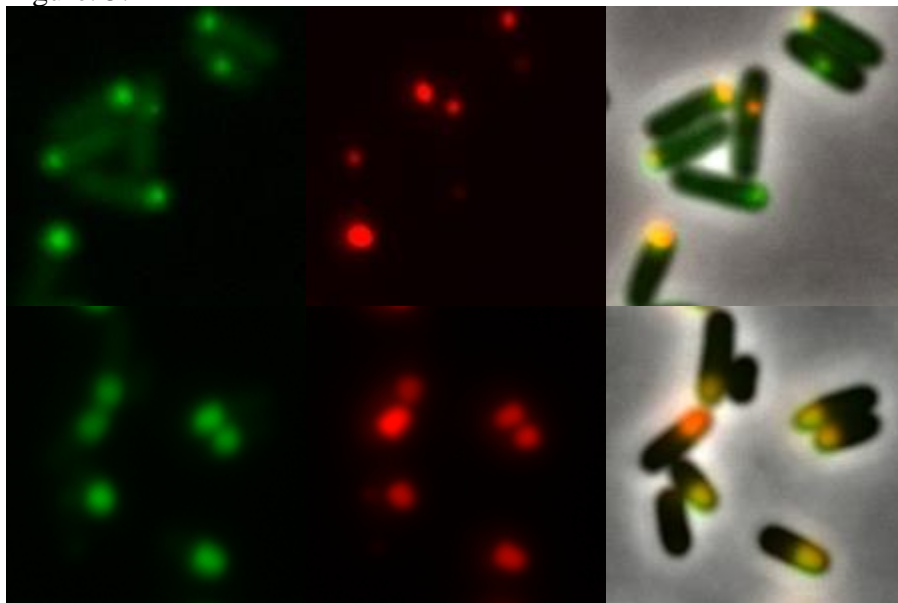
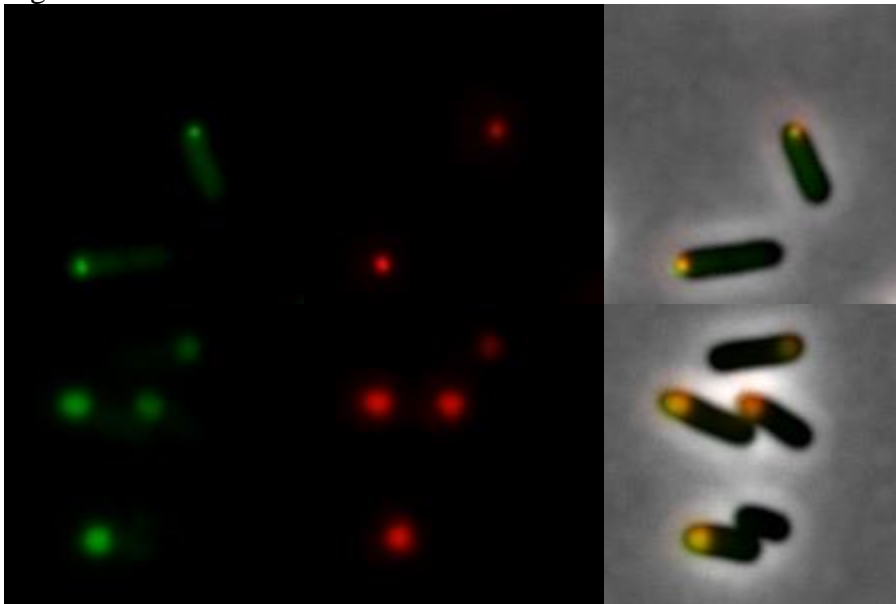
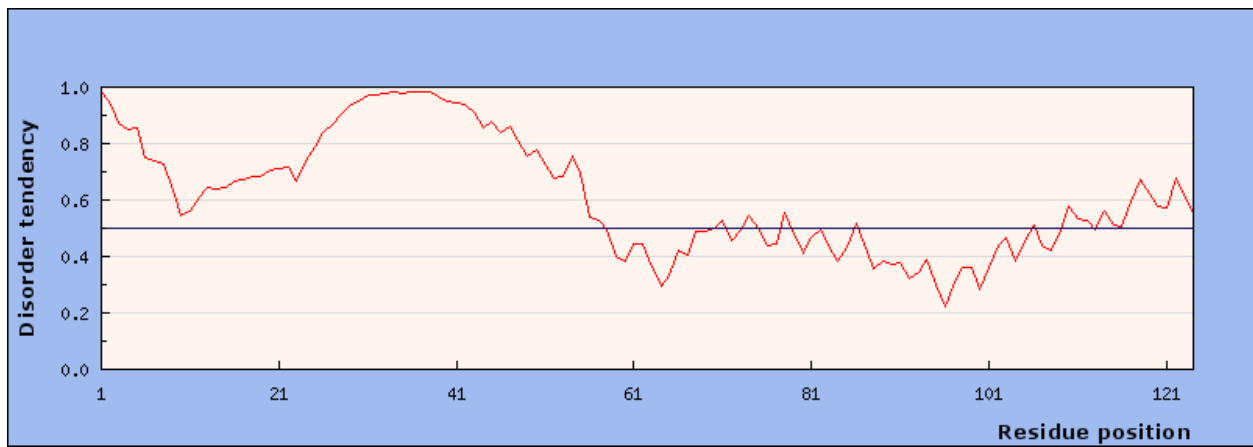


Figure 6:

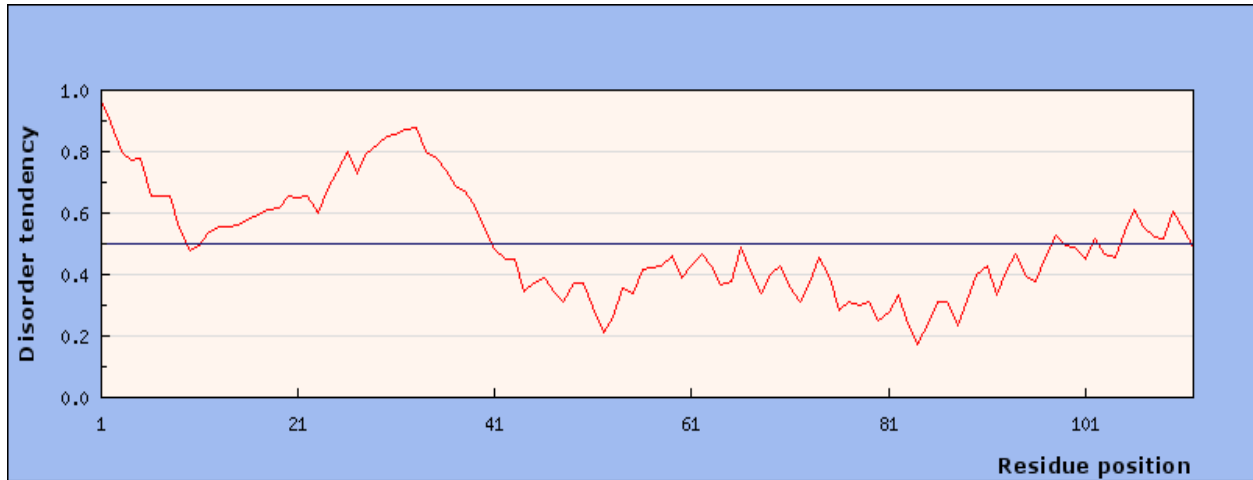


Supplementary Figure 1:





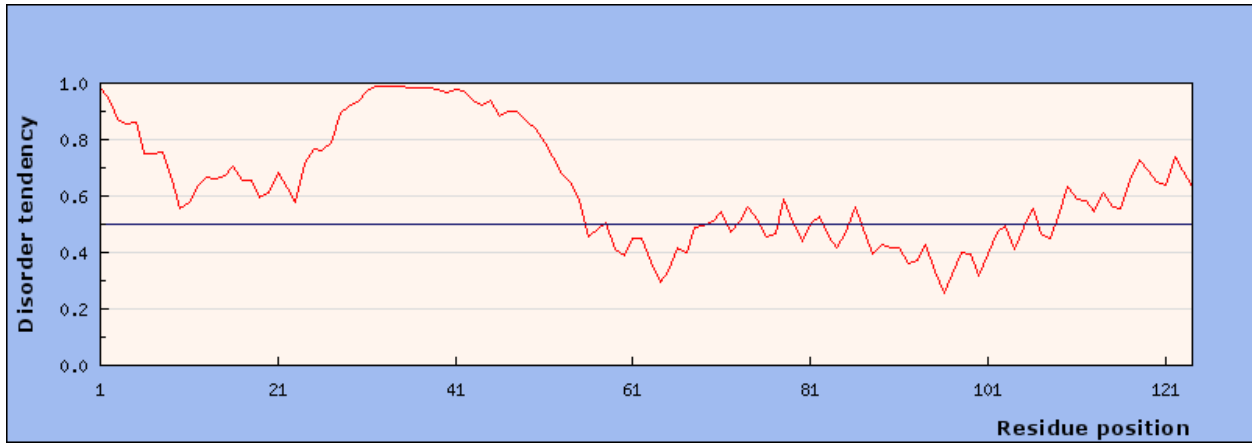
S Figure 2:



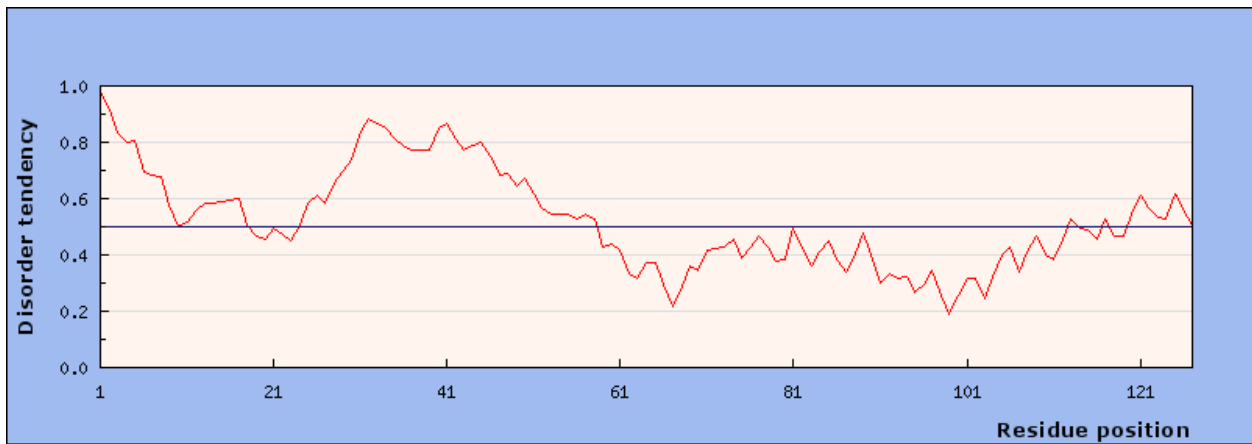
S Figure 3:



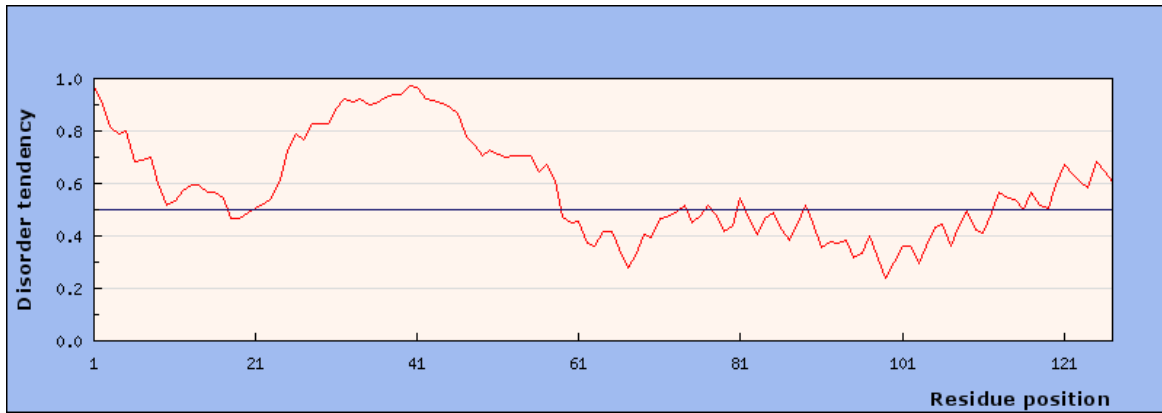
S Figure 4:



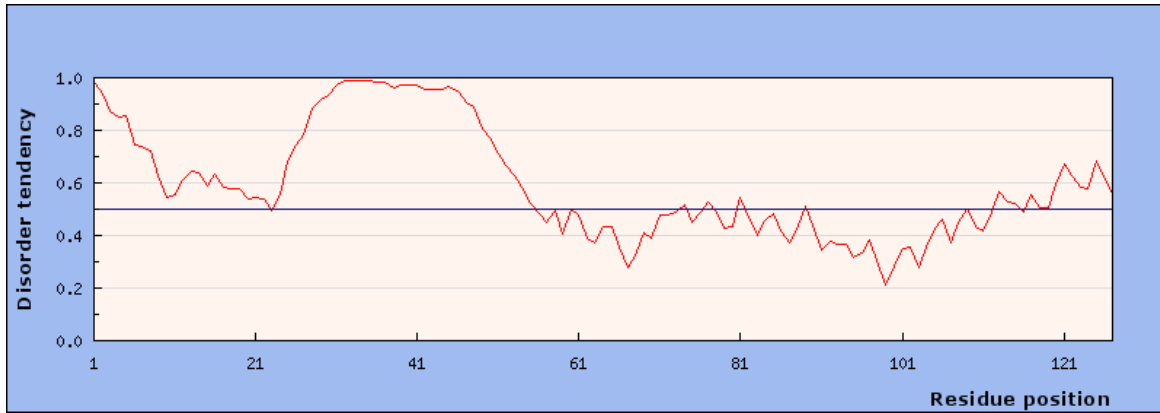
S Figure 5:



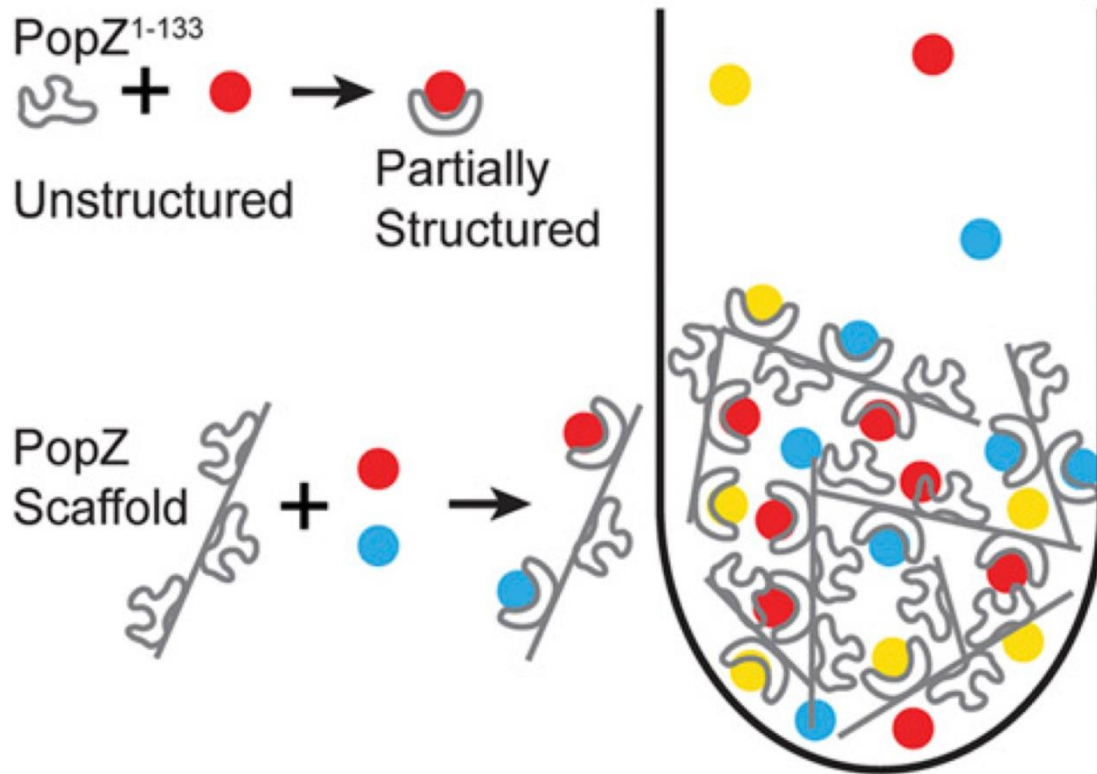
S Figure 6:



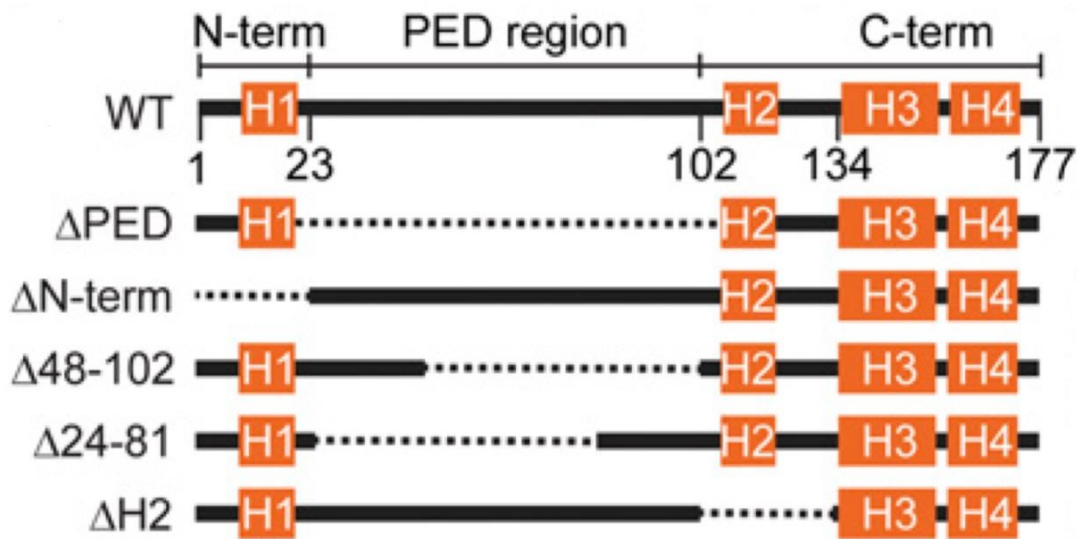
S Figure 7:



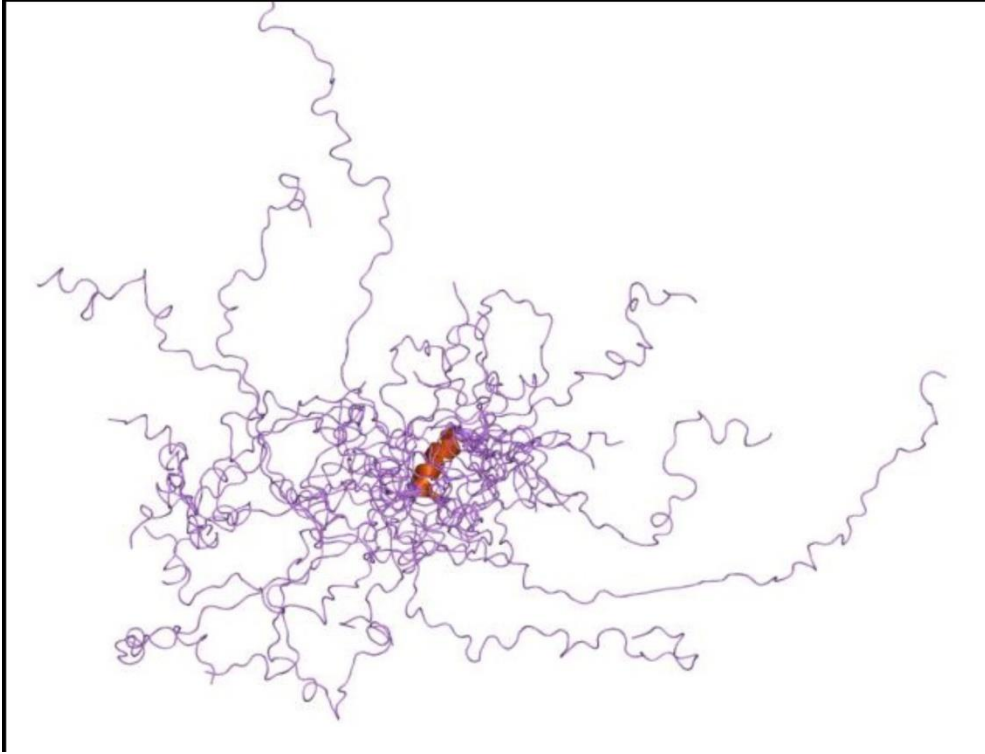
S Figure 8:



S Figure 9:



S Figure 10:



S Figure 11:

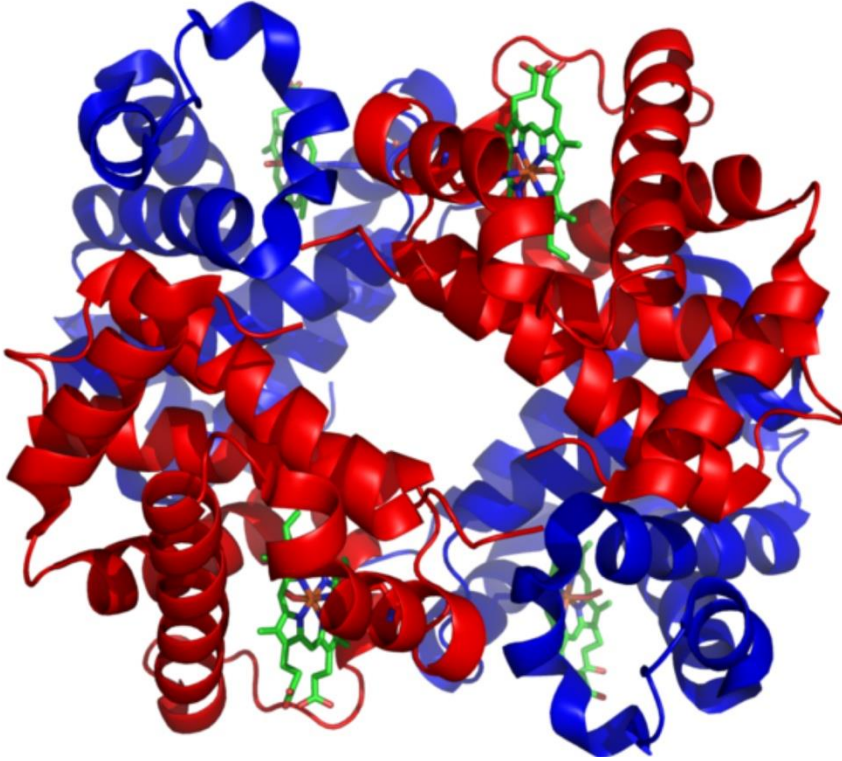


Figure Legend:

Figure 1:

The panels from left to right: (top) A strain with a PopZ construct that has had the PED region completely deleted from the protein and the candidate protein ChpT. The first panel is the GFP only channel being shown meaning the ChpT signal is the only one being shown and the green is shown to be diffuse throughout the cell and not localized without the PED region of PopZ. The second panel is the mCherry, the only channel being shown where PopZ is in the cell and the red is localized to the poles of the cell. The third panel shows a merge of both channels and the red and green are shown not to be overlapping much, so PopZ is not recruiting ChpT to the pole of the cell. (Middle) Wild type PopZ co-expressed with ChpT. The first panel is the GFP only channel showing where ChpT is in the cell and it is shown to be localizing to the pole of the cell. The second panel is the mCherry only channel showing where PopZ is in the cell and it is localized to the pole of the cell. The third panel is the merge of mCherry and GFP channels and shows that PopZ is localized to the pole and is recruiting ChpT to the pole within the presence of the PED region. (Bottom) Wild type PopZ co-expressed with ParB. The first panel is GFP only

channel showing where ParB is in the cell and the green is localized to the pole of the cell. The second panel is mCherry only channel showing where PopZ is in the cell and the red is localized to the pole of the cell. The third panel is the merge of mCherry and GFP and the yellow dot is the overlap of the two protein signals and both signals have been localized to the pole within the presence of the PED region.

Figure 2:

The panels from left to right: Controls that were used for the experiment have the N-terminal of PopZ deleted. (Top) the first panel is the GFP channel only showing where the ParB is in the cell and the green is diffuse throughout the cell therefore ParB is not localized to the pole. The second panel is the mCherry channel only showing where PopZ is in the cell and the red is localized to the pole. The third panel is the merge showing that there is not an extensive overlap of the red and the green signals indicating that PopZ is not co-localizing with ParB. (Bottom) The first panel is the GFP channel only showing where the ChpT is in the cell and the green is diffuse throughout the cell. The second panel is the mCherry only channel showing where the PopZ is in the cell and the red is localized to the pole of the cell. The third panel is the merge showing that there is little overlap of the red and green signals demonstrating the PopZ is not co-localizing with ChpT.

Figure 3:

The panels from left to right: The scrambled PED PopZ construct. (Top) The first panel is the GFP channel only showing where the ParB is in the cell and the green has a small foci localized to the pole of the cell. The second panel is the mCherry only showing where PopZ is in the cell and the red has a small foci localized at the pole of the cell. The third panel is the merge showing that there is overlap of the green and red as a yellow foci at the pole of the cell. The overlap shows that the scrambled PED region is still capable of binding and recruiting ParB to the pole of the cell. (Bottom) The first panel is the GFP channel only showing where ChpT is in the cell and the green has a foci at the pole of the cell. The second panel is the mCherry channel only showing where PopZ is in the cell and the red has a foci at the pole of the cell. The third panel is the merge showing a yellow foci at the pole of the cell indicating that PopZ is also recruiting and co-localizing with ChpT.

Figure 4:

The panels from left to right: The half scrambled PED PopZ construct. (Top) The first panel is the GFP only channel showing where ParB is in the cell and the green has a foci at the pole of the cell. The second panel is mCherry only showing where PopZ is in the cell and the red has a foci at the pole of the cell. The third panel is the merge showing an overlap of the green and red channels with the yellow foci at the poles, indicating that PopZ is still recruiting and co-localizing with ParB at the pole of the cell. (Bottom) The first panel is GFP only showing where ChpT is in the cell and the green has a foci at the pole of the cell. The second panel is the mCherry only channel showing where PopZ is in the cell and there is a red foci at the pole of the cell. The third panel is the merge showing a yellow foci at the pole of the cell indicating that PopZ is still recruiting and co-localizing with ChpT to the pole of the cell.

Figure 5:

The panels from left to right: The PED region having only Glycine and Serine residues. (Top) The first panel is the GFP only channel showing where ParB is in the cell and there is a green foci at the pole of the cell. The second panel is the mCherry channel only showing where PopZ is in the cell and there is a red foci at the pole of the cell. The third panel is the merge showing a yellow foci where the red and green signal have overlapped indicating that PopZ is still recruiting and co-localizing with ParB. (Bottom) The first panel is the GFP channel only showing where ChpT is in the cell and there is a green foci in the pole of the cell. The second panel is the mCherry only channel showing where PopZ is in the cell and there is a red foci at the pole of the cell. The third panel is the merge showing a yellow foci at the pole of the cell from the overlap of the red and green signals indicating that PopZ is still recruiting and co-localizing with ChpT.

Figure 6:

The panels from left to right: The PED region having only Proline and Alanine residues in the region. (Top) The first panel is the GFP only channel showing where ParB is in the cell and there is a green foci at the pole of the cell. The second panel is the mCherry only channel showing where PopZ is in the cell and there is a red foci at the pole of the cell. The third panel is the merge showing a yellow foci from the overlap of the green and red channels indicating that PopZ is still recruiting and co-localizing with ParB. (Bottom) The first panel is the GFP only channel showing where ChpT is in the cell and there is a green foci at the pole of the cell. The second panel is the mCherry only channel showing where PopZ is in the cell and there is a red foci at the



pole of the cell. The third panel is the merge showing a yellow foci where the red and green signals overlap indicating that PopZ is still recruiting and co-localizing with ChpT.

Supplementary Figure 1: The graph predicting the intrinsic disorder of the scrambled PED PopZ construct. The x-axis is the residue position in the protein and the y-axis is the disorder ranking with the line in the center being the division between disorder (the top portion of the graph) and order. The PED region is amino acids 24-50 and there is a high point in this graph predicting that this region of the protein will be disordered even though it is scrambled. The prediction matches the results that were shown in the images of the construct.

Supplementary Figure 2: The graph predicting the intrinsic disorder of the half scrambled PED PopZ construct. The PED region of the construct is predicted to have a less disordered PED region than the scrambled did. The prediction was not matched to the results shown in the images of this construct. The disorder dropped in the graph but the PopZ construct still bound with the candidate proteins.

Supplementary Figure 3: The graph prediction the intrinsic disorder of the glycine/serine PED PopZ construct. The PED region of the construct is predicted to be even less disordered than the half scrambled construct. The prediction does not match the results shown in the images of this construct as this construct was predicted to bind even less with the candidate proteins.

Supplementary Figure 4: The graph predicting the intrinsic disorder of the proline/alanine PED PopZ construct. The PED region of the construct is predicted to be disordered, similar to the scrambled construct. The prediction matches the results shown in the images for the construct as the PopZ binds with the candidate proteins.

Supplementary Figure 5: The graph predicting the intrinsic disorder of one of the possible future constructs. The PED region has the glutamate and aspartate residues, which are negatively charged, replaced with lysine, which is positively charged. The prediction for disorder has gone down on this construct suggesting that it could perhaps prevent PopZ from binding with the candidate proteins.

Supplementary Figure 6: The graph predicting the intrinsic disorder of another possible future construct. The PED region has had all of the proline residues replaced with serine residues. The change will take a highly disordered amino acid and replace it with a very low scoring disorder amino acid. The prediction is even lower for disorder on this construct suggesting that this PopZ will be less likely to bind with the candidate proteins.

Supplementary Figure 7: The reference graph for all the other predictions. The wild type PopZ shows the highly disordered PED region. All of the constructs have disorder predictions lower than the WT.

Supplementary Figure 8: (Supplied by Holmes et al.) A diagram showing how PopZ works at the pole of *Caulobacter crescentus* cells. The diagram shows how PopZ works as a scaffold to bind multiple partners and organize them at the pole of the cell. The diagram also shows how PopZ does not have a structure until it binds its partner.

Supplementary Figure 9: (Supplied by Holmes et al.) A map of the regions of PopZ. The 23-101 amino acid area is the PED region in the wild type. The H1, H2, H3, H4, labels are for the presence of helices in the more ordered regions of PopZ that have functions in recognition, pole localization, and scaffold formation. The following maps show where certain regions were deleted to investigate PopZ's regions for function.

Supplementary Figure 10: (Supplied from Wikipedia) An image demonstrating how intrinsically disordered proteins exist before binding to partners. The proteins have no true structure and exist as what looks like string in the image. The ribbon structures are more ordered regions of the proteins. These proteins usually have some order to them. Disorder is usually associated with misfolded proteins, which can cause disease in humans and gaining a better understanding of proteins that naturally exist without order can lead to a knowledge of how the systems for misfolding proteins work.

Supplementary Figure 11: (Supplied by Wikipedia). An image of the human structured protein hemoglobin. This protein has all four of the traditional levels of structure and demonstrates the difference between the two proteins in appearance alone. The comparison of this protein with intrinsically disordered proteins supports the argument for form fitting function as hemoglobin has the very specific function of moving oxygen and carbon dioxide in the blood where disordered proteins have even more binding partners. Disordered hub proteins are important for cell organization.

#### References:

1. (Main source)A Holmes, Joshua & Follett, Shelby & Wang, Haibi & P Meadows, Christopher & Varga, Krisztina & R Bowman, Grant. (2016). *Caulobacter* PopZ forms an

intrinsically disordered hub in organizing bacterial cell poles. *Proceedings of the National Academy of Sciences*. 113. 10.1073/pnas.1602380113.

2. (Secondary Source) Theillet, F., Kalmar, L., Tompa, P., Han, K., Selenko, P., Dunker, A. K., . . . Uversky, V. N. (2013). The alphabet of intrinsic disorder. *Intrinsically Disordered Proteins*, 1(1). doi:10.4161/idp.24360
3. (Secondary Source) Bowman, G. R., Perez, A. M., Ptacin, J. L., Ighodaro, E., Foltastogniew, E., Comolli, L. R., & Shapiro, L. (2013). Oligomerization and higher-order assembly contribute to sub-cellular localization of a bacterial scaffold. *Molecular Microbiology*, 90(4), 776-795. doi:10.1111/mmi.12398
4. (Secondary Source) Bowman, G. R., Comolli, L. R., Gaietta, G. M., Fero, M., Hong, S., Jones, Y., . . . Shapiro, L. (2010). *Caulobacter*PopZ forms a polar subdomain dictating sequential changes in pole composition and function. *Molecular Microbiology*, 76(1), 173-189. doi:10.1111/j.1365-2958.2010.07088.x
5. (Basic information source) Nelson, David L., Albert L. Lehninger, and Michael M. Cox. *Lehninger principles of biochemistry*. Macmillan, 2013.