

3DMTRanalyzer
Operation Manual
Version “volumetricMTR_v01_008”

Outline

Table of Contents

- I. 3DMTRanalyzer - Manual and SOP 3
 - A. Important Considerations 3
 - B. Application Panels..... 3
 - C. Operation 12
 - D. 3DMTR methods..... 12
 - E. Version updates 13
 - G. Happy Patching: 14

I. 3DMTRanalyzer - Manual and SOP

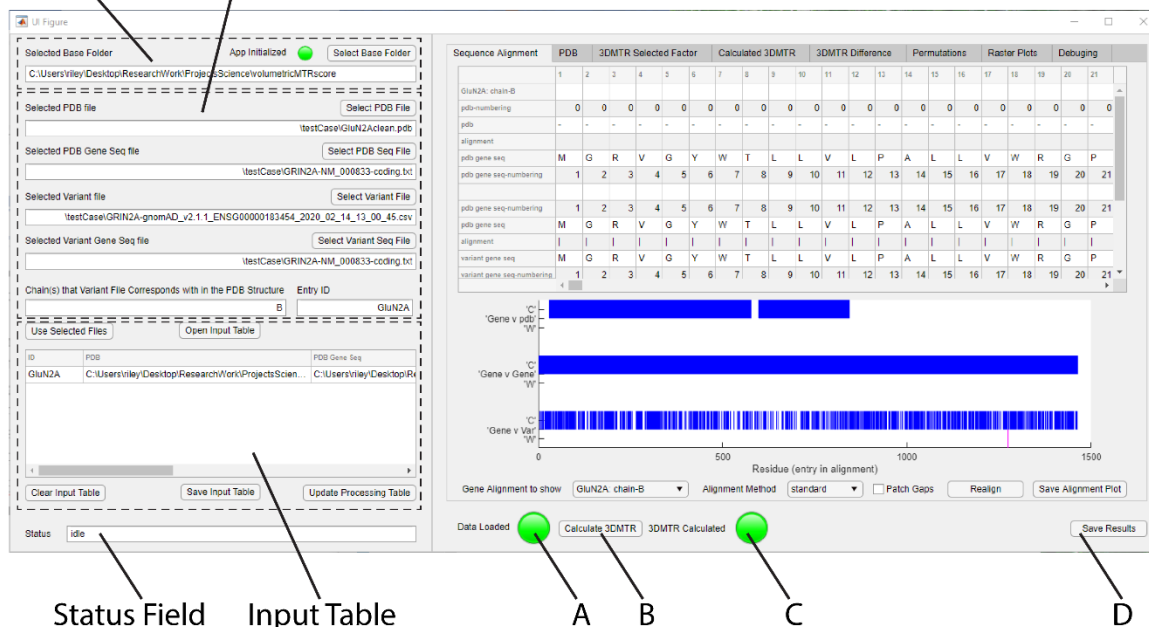
A. Important Considerations

1. This application has been tested thoroughly on NMDARs, using homology and crystal structures that contain all components to be tested. Separate .PDBs can be used if the orientation of the separate structures are appropriately orientated to one another (and is best if they all are labeled by different chain IDs). For example, the coordinates of the PDBs are used as is, so if the PDB structures are overlapping this will dramatically impact the inter-residue distances and will generate spurious results.
2. The debug tab should allow a user to save the current state of the application and can be shared to help troubleshoot any errors or concerns about the results generated by the application.

B. Application Panels

1. Left Panel, File Loading, and Main Right Panel Features

Base Folder File Input Fields



- a) Base Folder – This is where the results folder will be created and also where, when searching for files, the dialog boxes will opened up to. It is best not to change this after moving forward with any subsequent steps.
- b) File Input Fields – All of these file address fields and the chain and ID fields need to be correctly filled before adding them to the “Input Table”. The address field root folder is the selected base folder.
 - PDB File – The PDB file that you want to analyze, this can be a single protein chain or a multimeric structure. It is best that all different genes are segregated to separate chains and the residue numbering matches as closely as possible to the gene sequence that is used. It is best to remove any extra elements from the PDB (agonist, waters, ions, post-translational modifications,

multiple states, etc), its best to remove what one can in pymol and then export the molecule to a pdb of only the active state.

- PDB Gene Seq – This needs to be the coding DNA sequence of the final gene product that is represented in the PDB, this can be downloaded from www.ncbi.nlm.nih.gov/gene/, search for the gene and then select the “NCBI Reference Sequences (RefSeq)” (eg NM_000833.5), then click the “send to” drop down box, select the “coding sequences” option and download the “FASTA nucleotide”.
- Variant Gene Seq – This is listed separately, most of the time this will be the same sequence as the PDB Gene Seq, however you can attempt to analyze a close analog of the protein represented in the PDB by providing the analog’s gene sequence and its variant dataset. Obtain the same type of file as above. This could be the human gene sequence if the PDB is of a mouse analog or could be a high consensus parallel analog.
- ExAC/gnomAD Variant file – The ExAC or gnomAD exported missense and samesense (synonymous) variant dataset in a CSV file for the gene. Columns “J” and “L” (as represented in Microsoft excel) are used and requires all missense and synonymous SNVs reported.
- Chains – The letter chain identifier that the gene sequence and variant data applies to. If there are multiple chains in the PDB that all correspond to the same gene, separate each chain letter with a ‘,’ (“A,C”).
- ID – This is identifier that is applied to this gene dataset and is also used for naming the output files (all constituent names are concatenated), choose something appropriate, lacking special characters that might cause issues (eg ‘\’) but also not too long.

c) Input Table

- Use Selected Files – This takes the information set in the above fields and saves them into the table, file pointers saved in the table is what is used to load into the application when the “Update Processing Table” button is pushed.
- Open Input Table – the opens a dialog box to open a saved input table file.
- Clear Input Table – this clears the input table and clears the data loaded into the application.
- Save Input Table – this button opens a dialog box to select a file to save the current input table to a file.
- Update Processing Table – this button takes the file pointers inputted into the input table and loads all the files into the application.

d) Status Field – the activities of the application will be reported here. There may be unexpected errors if you click buttons while the application is processing data, so it is best to wait for the status field to say “idle” before clicking another button or changing a drop down box or any other application field.

e) Right Panel Buttons and Lamps

- Data Loaded Lamp – this lamp is activated when the files listed in the input table are fully loaded into the application. When this lamp is activated the data in the Alignment tab is current and reflective of the loaded data.
- Calculate 3DMTR – This button is active when the input table has been fully loaded into the application and it calculates the bulk of the analysis to enable the use of most of the rest of the analysis tabs (all except for the permutation tab, which is active but requires further analysis for the results to be extracted from the application).

- **3DMTR Calculated Lamp** – This lamp is activated after the “calculated 3DMTR” button is pushed and the MTR analysis is complete.
- **Save Results** – This button is active if the “3DMTR Calculated Lamp” is on and it exports the 3DMTR analysis along with all of the various methods of calculating the MTR and supporting data.

2. Sequence Tab

Sequence Tab

Alignment Table

Alignment Options

Alignment Pictograph

a) **Alignment Table** - here the alignments of the loaded sequences are shown. There are three alignments PDB to PDB gene seq, gene seq to variant seq, and the variant seq to the gnomAD data.

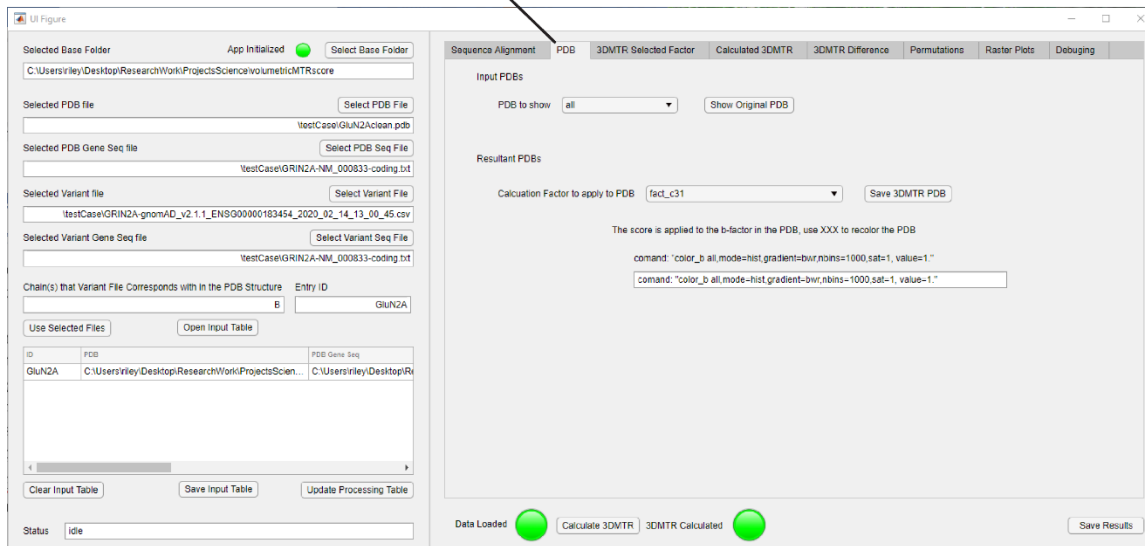
b) **Alignment Options**

- **Gene Alignment to Show** – This will be populated with all the chains of the loaded PDBs, selecting a different one will load the alignment data for that chain into the Sequence Alignment Tab.
- **Alignment Method** – There are 3 different methods that can be implemented, 2 automatic and a custom method. The two automatic methods are standard and local, standard attempts to align the sequences attempting to cover the whole of the two sequences, whereas the local method will try and minimize the gaps in the alignment. The custom method requires a txt file for each chain you wish to use a custom alignment, the txt file needs to be a two row ascii file the first row is the PDB gene sequence and the second is the variant gene seq (gaps represented by “-”).
- **Patch Gaps** – This option will fill any gaps in the pdb sequence with blank residues, clicking it will help force the gaps to be fixed whereas sometimes the aligning algorithm will scatter residues in the gaps.
- **Realign** – the button will realign the sequences with the current setting selected.
- **Save Alignment Plot** – This will save the “alignment pictograph” to a rasterized file (.png, there is also an option to save a vector based file - .svg).

c) Alignment Pictograph – The three alignment results are represented here by raster plots, the blue and upward lines represent fully matched residues, the magenta and downward lines represent mismatched residues. Gaps are represented by the lack of a line.

3. PDB Tab

PDB Tab

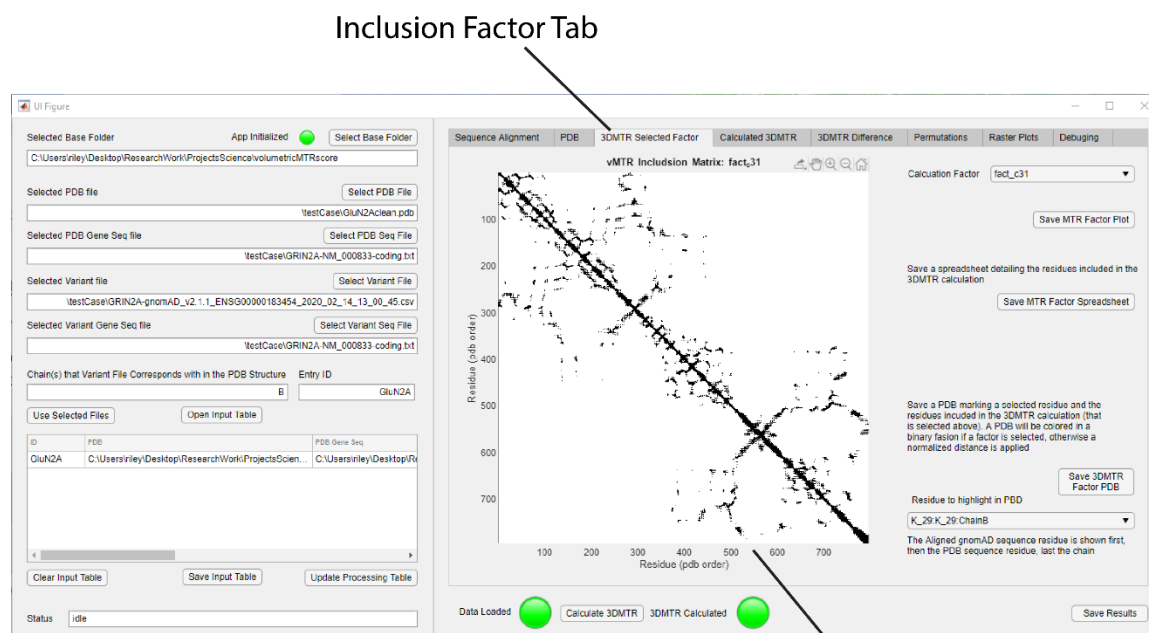


- Show Original PDB – this will display a PDB viewer window and show the PDB selected in the “PDB to show” drop down box.
- Save 3DMTR PDB – this will save the MTR result dataset to the combined PDB and save to a file. Options available are all the various 3DMTR and 1DMTR methods and also custom methods. The custom methods will take an ascii file containing a value for each residue and apply it to the PDB based on the type of the custom method selected. The score is covered to colorscale where 0 is blue, 0.5 is white, and 1 is red. For difference scores, score is covered to colorscale where -1 is blue, 0 is white, and 1 is red.
 - custom_rawNorm – the data loaded is has the minimum value set to 0 and then normalized to the largest value. (baseline subtracted and then normalized)
 - custom_4xMean – The data is baseline subtracted and then normalized, then is divided by 4x the mean score.
 - custom_1xSTD - The data is baseline subtracted and then normalized, then the maximum scale is set to the 1xStandard deviation. So the white color scale reflects the mean, the blue color scale is 1xSTD less than the mean (all scores more extreme are truncated), and the red color scale is 1xSTD more than the mean (all scores more extreme are truncated).
 - custom_2xSTD - The data is baseline subtracted and then normalized, then the maximum scale is set to the 2xStandard deviation. So the white color scale reflects the mean, the blue color scale is 2xSTD less than the mean (all scores more extreme are truncated), and the red color scale is 2xSTD more than the mean (all scores more extreme are truncated).
 - custom_3xSTD - The data is baseline subtracted and then normalized, then the maximum scale is set to the 3xStandard deviation. So the white color scale reflects the mean, the blue color scale is 3xSTD less than the mean (all scores more extreme

are truncated), and the red color scale is 3xSTD more than the mean (all scores more extreme are truncated).

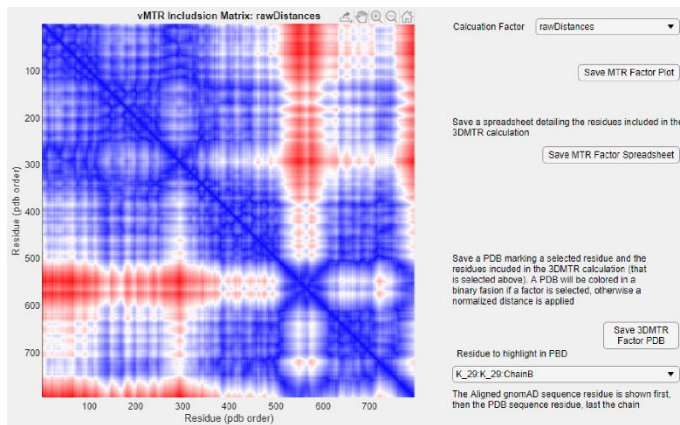
- custom_4xSTD - The data is baseline subtracted and then normalized, then the maximum scale is set to the 1xStandard deviation. So the white color scale reflects the mean, the blue color scale is 4xSTD less than the mean (all scores more extreme are truncated), and the red color scale is 4xSTD more than the mean (all scores more extreme are truncated).
- custom_[0_1] – This score assumes that the values supplied range from 0 to 1 (more extreme scores are truncated). In essence this is the same as the standard score.
- custom_[-1_1] - This score assumes that the values supplied range from -1 to 1 (more extreme scores are truncated). In essence this is the same as the standard difference score.
- custom – the data loaded is normalized to the largest value.
- Pymol command to recolor the output PDB with the loaded coloration (data stored in the b-factor attribute) - “color_b **enabled**,mode=hist,gradient=bwr,nbins=1000,sat=1, value=1.” - load the repository form “<http://pldserver1.biochem.queensu.ca/~rlc/work/pymol/>”

4. Factor Tab



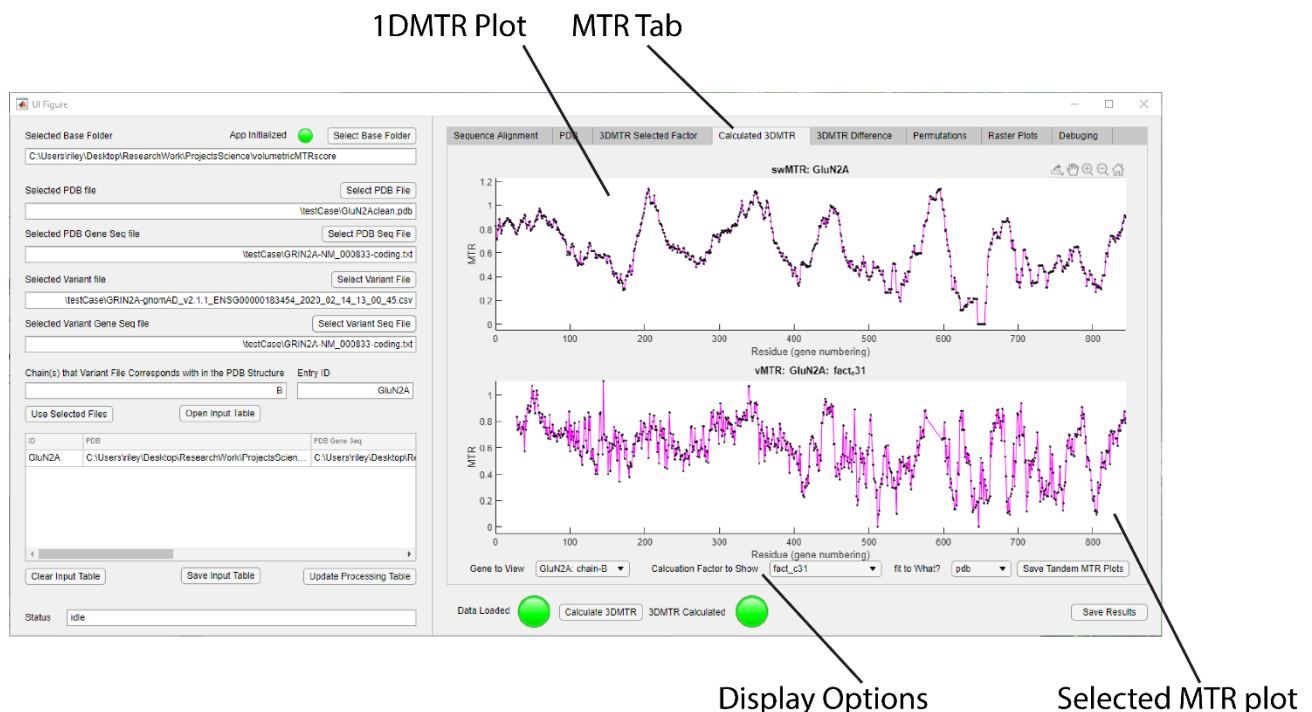
Matrix Display

- Save MTR Factor Plot – A graphical representation of the inclusion matrix that is selected in the “Calculation Factor” dropdown box. The rows represent each residue, and the columns are the included residues based on the selected factor. The black pixels represent included residues. The residue distances can also be visualized (see below), blue are closer residues and red are further residues.



- **Save MTR Factor Spreadsheet** – This saves a series of spreadsheets that include documentation of which residues are used, how many residues from each change are used, and how many times each residue is included in the factor selected by the “Calculation Factor” dropdown box. There is also an option to document the inter-residue distances for all residues, if the “rawDistances” option is selected. The certain “fact_t” factors have not been fully tested. Note, this can be a processing/time intensive task.
- **Save 3DMTR Factor PDB** – Based on the factor that is selected in the “Calculation Factor” dropdown box, a PDB is saved with the residue that is selected in the “Residue to highlight in PDB” and the other residues that are included in the selection factor, using the same color scale mechanism as above. The selected residue (or center residue) is colored Red and the included residues (the surrounding residues) are in blue. The inter-residue distances can also be exported, the blue residues are the closest and the red are the farthest.

5. MTR Plot Tab



- The top plot is the 1DMTR of the selected gene (thus reflects all the residues of the entire gene) and then the bottom is the selected MTR method (can be either 1DMTR or 3DMTR), this plot will only contain the residues represented in the PDB.
- **Gene to View** – Selects which chain to display in the two plots.

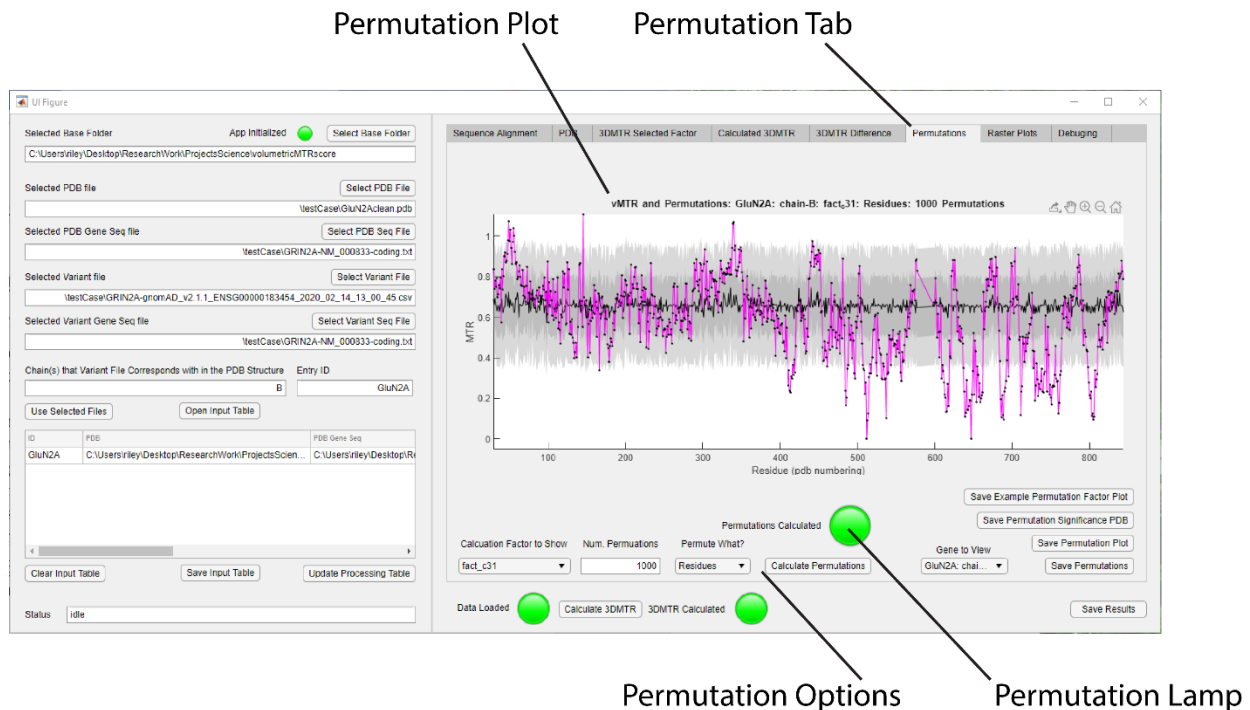
- Calculation Factor to Show – Select which factor to display in the lower plot.
- Fit to what? – Sets what the x-axis is fit to; pdb, gene, tight. PDB, set the lower limit to 0 and the upper limit to the last pbd residue. Gene, set the lower limit to 0 and the upper limit to the last residue of the gene product. Tight, , set the lower limit to the first and the upper limit to the last pbd residue. The residue numbering used to plot both of these traces is the variant gene sequence numbering, which should reflect the same residues in both plots.
- Save Tandem MTR Plot – Saves the two plots, per current settings, to a raster image and the option to save a vector file.

6. Difference Tab



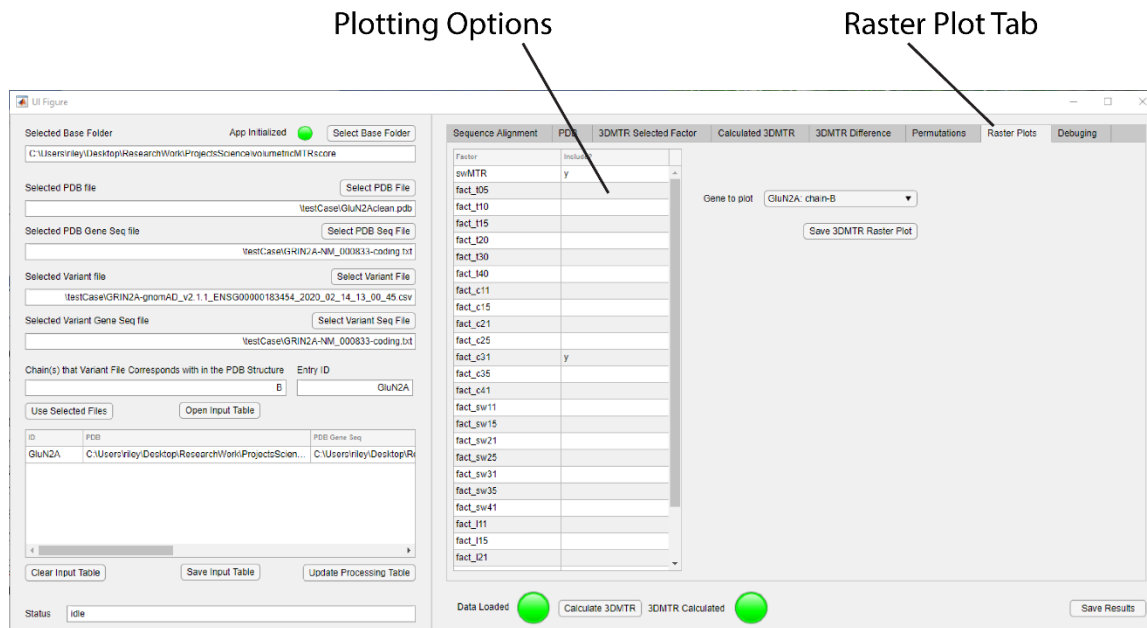
- Gene to View - Selects which chain to display in the difference plot.
- Factor 1 – a 1DMTR or 3DMTR factor to determine the difference from.
- Factor 2 – a 1DMTR or 3DMTR factor to subtract from Factor 1.
- Save Diff 3DMTR PDB - Based on the difference of factors that are selected and the gene selected, a PDB is saved (values in the plot) using no difference as white, a difference of -1 is blue, and +1 is red (more extreme cases are truncated).
- Save Diff MTR Plots - Saves the difference plot, per current settings, to a raster image and the option to save a vector file.

7. Permutation Tab



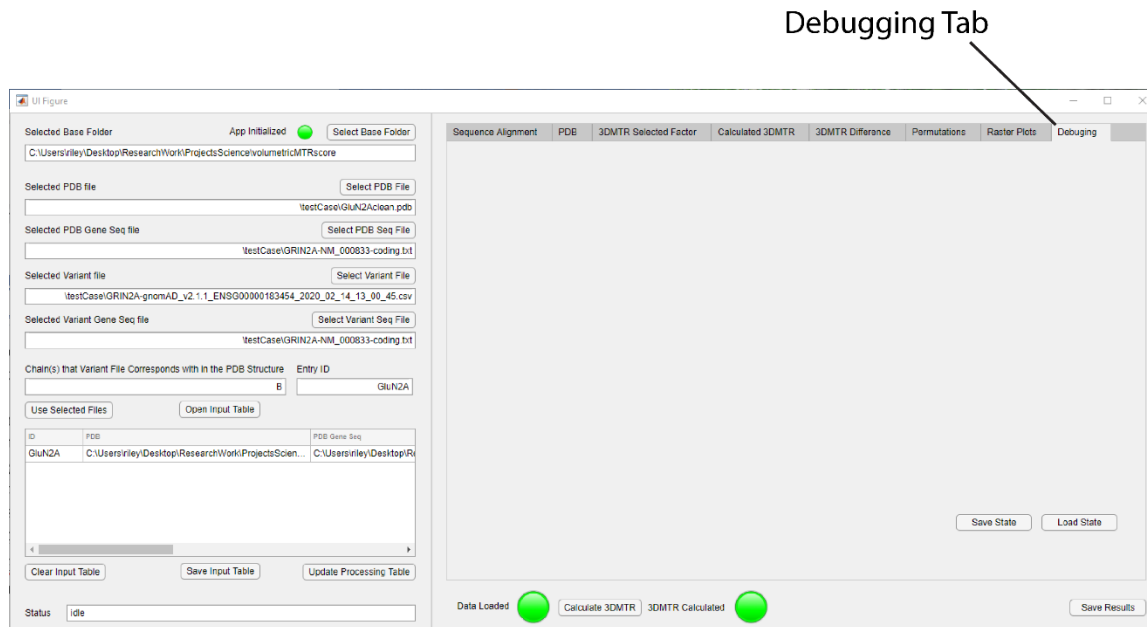
- Calculation Factor to Show - Select which factor to display in the lower plot.
- Num Permutations – number of iterations of Permutations to calculate.
- Permute What – What kind of randomization to perform, methods include Residue, Factor, and Score. Residue permutation randomizes the residue location. Factor permutation randomized the gnomAD data that is applied to each residue. Score is just a randomization of the final resulting MTR score for each residue.
- Calculate Permutation – Based on the setting, the application will start to perform the permutation analysis. Note, this can be a processing/time intensive task.
- Permutation Lamp – this becomes active when the permutation analysis is concluded.
- Gene to View - Selects which chain to display in the difference plot.
- Save Example Permutation Factor Plot – Saves the original MTR factor inclusion matrix and one instance of the randomized inclusion matrix, save a compressed image plot and the option to save a lossless pixel representation of the inclusion matrix.
- Save Permutation Significance PDB – saves a PDB of with the residues that their actual MTR score is outside 2x STD from the mean permutation value. Intolerant scores are blue and tolerant scores are Red, non-significant scores are white.
- Save Permutation Plot - Saves the permutation plot, per current settings, to a raster image and the option to save a vector file.
- Save Permutation – Saves the Permutation results to spreadsheets.

8. Raster Plot Tab



- Save 3DMTR Raster Plot – Saves the MTR raster plot representations of the MTR scores that are marked in the “Plotting Options” table and the gene selected in the dropdown box.

9. Debugging Tab



- Save State – Saves the current state of the application, can be compressed and shared help the troubleshooting process.
- Load State - Loads the saved state of the application, the app will then update a number of things to allow examination of any potential error states.

C. Operation

To start the process of analyzing a protein using this application. First, the .PDB file you wish to analyze should be cleaned of all extraneous components (ligands, waters, etc, essentially any non-protein atoms). Then download the gene sequences and gnomAD variant table. Then you can build the input table detailing all the components you wish to analyze, then click the “Update Processing Table” button. Check the status bar to see when processing has concluded and the field reads “Idle”, the “Data Loaded” lamp should also be on. Check the alignments tab to be certain that there were no issues with the inputted files and standard alignment methods. If there are issues, you can change the settings and then realign the sequences to try to resolve any issues. If all of the alignment methods fail, you can use a custom alignment. Next the push the “Calculate 3DMTR” button and allow the application to become “Idle”, the “3DMTR calculated” lamp should be on. Now most of the remaining tabs should be populated with data and can be edited to view different analyses of the protein MTR. Permutations can now be calculated, however this can be time intensive and so try out a sample permutation of 10-20 permutation before starting an extensive analysis.

1. GluN2A Test Case

A test case can be found in the GitHub depository, illustrating the results you should expect if you use the application properly. The test case uses GluN2A subunit and the *GRIN2A* gene, all the required files and the application results can be found in the depository.

D. 3DMTR methods

1. swMTR – This method applies the 1DMTR (using a window of 31 residues) to the full gene sequence. This should be the same
2. fact_tXX – This 3DMTR method uses a straight spherical cutoff threshold ($X \text{ \AA}$) to decide which residues to include in the 3DMTR analysis. As proteins have variable residue density throughout its entire structure this method can result in some calculation errors or variable number of residues included in the calculation for different portions of the protein.
 - a) Variations include: fact_t05, fact_t10, fact_t15, fact_t20, fact_t30, fact_t40
3. fact_cXX – This 3DMTR method uses X nearest neighbors in Euclidean space across all protein chains represented in the .PDB file to decide which residues to include in the 3DMTR analysis.
 - a) Variations include: fact_c11, fact_c15, fact_c21, fact_c25, fact_c31, fact_c35, fact_c41
4. fact_swXX – This 1DMTR method uses X nearest neighbors as determined by the amino acid sequence of those included in the .PDB to decide which residues to include in the 1DMTR analysis.
 - a) Variations include: fact_sw11, fact_sw15, fact_sw21, fact_sw25, fact_sw31, fact_sw35, fact_sw41

5. fact_1XX – This 3DMTR method uses X nearest neighbors in Euclidean space limited to only the residues from the same protein chain as marked in the .PDB file to decide which residues to include in the 3DMTR analysis.

a) Variations include: fact_111, fact_115, fact_121, fact_125, fact_131, fact_135, fact_141

E. Version updates

v001.008 – original published manual and application

G. Happy Analyzing

