Tutorial on MMM and MMMx

G. Jeschke*[1]

[1]ETH Zürich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland

September 1, 2023

1 Introduction

Multiscale Modelling of Macromolecules (MMM) is a Matlab program with graphical user interface that supports *in silico* spin labelling, full-fledged protein and nucleic acid visualization with special features for spin labels, and simulation and analysis of DEER (PELDOR) data in the context of spin-labelled biomacromolecules.

MMMx is the successor, script-based, program that is recommended for all tasks that do not require interactive visualization of protein structures, such as *in silico* spin labelling site scans, prediction of distance distributions between spin labels or between atoms in ensemble structures, integrative modelling and ensemble modelling based on distance distribution restraints, and analysis of ensemble structures. MMMx provides only limited visualization features, but can generate visualization scripts for MMM. If MMM is open in the same Matlab instance during running an MMMx visualization script, visualization can be automatically executed in MMM and figures can be saved automatically.

MMM is still supported regarding bug fixes and minor adaptations to user requests, including new rotamer libraries for spin labels. New visualization features that require ingteractivity or that are very closely related to existing MMM visualization are implemented in MMM. New modelling functionality and any ensemble analysis functionality is implemented exclusively in MMMx.

*corresponding author, gjeschke@ethz.ch

2 Input and output concept

2.1 MMM

MMM inputs structures as PDB files, including direct download from the PDB. Once the PDB no longer supports download in the original PDB format, an mmCIF reader with minimal required functionality will be added. MMM outputs structures in PDB format or in its proprietary Matlab-based format. MMM imports DEER data in Bruker or ASCII format and distance distributions in DeerAnalysis output format.

2.2 MMMx

MMMx inputs structures and ensembles as PDB files, including direct download from the PDB, the protein ensemble databasew (PED), the AlphaFold protein structure database, and Zenodo datasets. Ensembles can feature weights (populations) for conformers. MMMx outputs structures and ensembles as PDB files or in its internal Matlab-based format.

MMMx outputs metadata or analysis results as comma-separated value (CSV) files and figures in any Matlab-supported output format (PDF or PNG recommended). Log files are human-readable text files. MMMx imports DEER data with background fit and distance distributions as CSV files. Further formats for such data will be added upon request.

The concept of MMMx is geared to modelling and analysis pipelines that can be implemented either fully in MMMx or can use other tools at any section of the pipeline.

3 Third-party software concept

MMMx accesses third-party software for complex tasks in modelling or analysis where well documented tools exist can be called from a Matlab program. Most MMMx tasks can be run in the absence of third-party software, but some functionality is the missing. We recommend to install at least SCWRL4 for protein sidegroup rotamer modelling. For integrative modelling that uses small-angle scaterring restraints (SAXS and SANS curves), the ATSAS suite is required. Structure optimization in an MMMx pipeline requires Yasara Structure, which requires a paid license even for academic research. Alternatively, structure optimization cen be performed outisde MMMx and the optimized structures can be fed to the next step of the MMMx pipeline.

Nice add-ons for specialized tasks are MUSCLE (sequence alignment,

license-free) and MSMS (solvent-accessible surfaces for adding a coarse-grained lipid bilayer model).

For automatic calls from within MMMx, all third-party software must be on the Matlab path. As an exception, the access path to SCWRL4 can also be specified within MMMx scripts.

4 Help and documentation

4.1 MMM

MMM has help files that can be read with a web browser or the Matlab help browser. MMM help files are no longer updated.

4.2 MMMx

MMMx documentation is accessible on the internet at mmmx.info. MMMx documentation is continuously updated when features change or new features are added. Documentation updates may lag implementation by a few weeks.

5 Download

5.1 MMM

For MMM, not all definition and data files are supplied on GitHub. Hence, we recommend initial installation from epr.ethz.ch/software. However, the current version including the newest bug fixes is more conveniently accessed via GitHub at github.com/gjeschke/MMM.

5.2 MMMx

MMMx is best downloaded via GitHub at github.com/gjeschke/MMMx. We recommend to specify the GitHub commit and download date when reporting on research that uses MMMx. Unlike MMM, MMMx can be compiled into a Windows executable that does not require a Matlab license or Matlab toolbox licenses. Note that this executable, downloadable at epr.ethz.ch/software.html, is only updated upon request.

6 MMM: Spin labelling T4 Lysozyme

6.1 Structure download

Open MMM and select the menu item `File/New from PDB/web`. Load structure `2LZM` from the PDB server. In the command line, type [select 72,131] and enter. Now select the menu item `EPR/Site scan/selected residues`. Click `OK` in the next two windows that are brought up to confirm default settings. When prompted for saving the site scan analysis, click `Save` and in the next window `Yes`. Your windows web browser will open and display the results of *in silico* spin labeling. Select the menu item `EPR/Attach precomputed rotamers`. Click `OK` in the window that is brought up. Your protein is now spin-labelled, but the labels are not yet displayed.

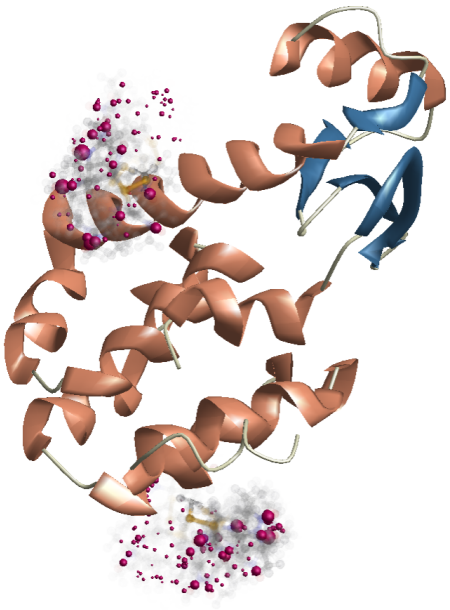

Figure S1: Visualization of spin labels at residues 72 and 131 of T4 Lysozyme (PDB: 2lzm.

6.2 Visualization

In the command line, type [show 72,131 label] and enter. Rotamer clouds will now be displayed. If you wish to see the backbones of the rotamers with transparency-encoded population, in the command line type `show 72,131.: ball&stick`. The results is shown in Fig. S1.

7 MMMx: Distance distributions in T4 Lysozyme

On the Matlab prompt, type `MMMx demo_distributions_T4L`. After some time, distance distributions will be displayed in Matlab windows. As an example, the distribution between labelled sites 72 and 131 is shown in Figure S2. The system text editor will open and display the log file.

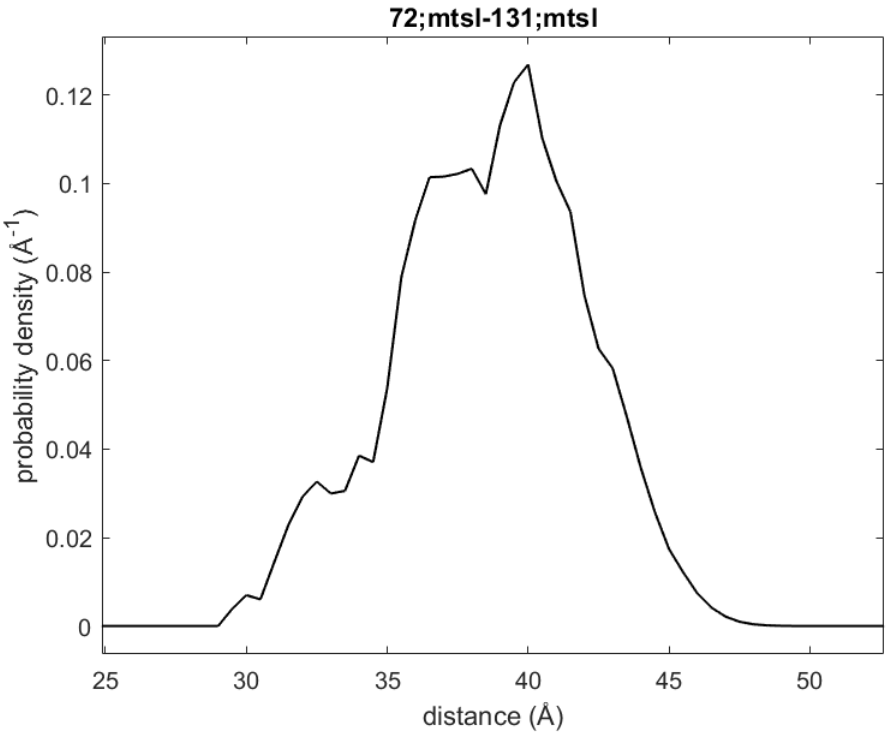

Figure S2: Simulated distance distribution between methanethiosulfonate spin labels at sites 72 and 131 of T4 Lysozyme (PDB: 2lzm).

8 MMMx: Localization of a spin label and visualization

At the Matlab command prompt, first open MMM. Then type `MMMx T4L_localize_131` and enter. After some time, display in MMM will start. When it is finished, the system editor will open the logfile. The logfile contains information on localization with respect to the PDB coordinates. The visualization looks as in Figure S3, if an old DSSP executable is on your Matlab path.

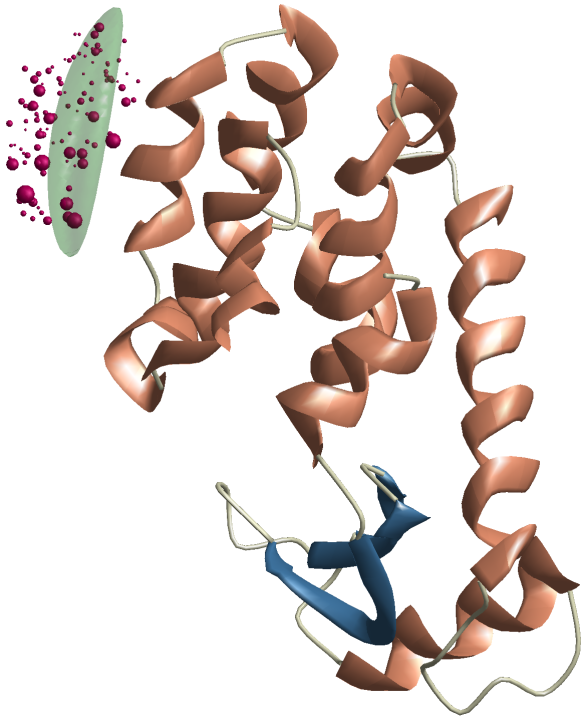

Figure S3: Localization of a spin label at site 131 in T4 lysozyme (PDB: 2lzm) via distances to other spin labels.

9 MMMx: Experiment design for RigiFlex

In this part, we generate an MMMx script template for integrative ensemble structure determination of a protein that contains two folded domains joined by a flexible linker, as well as flexible N-terminal and C-terminal domains. To that end, we use an AlphaFold prediction of the protein structure that MMMx directly downloads from the AlphaFold Protein Structure Database. All we need is the UniProt identifier of the peptide sequence, which in our case is Q07955 for human SRSF1. The MMMx script for generating a MMMx RigiFlex script is rather simple:

```
!ExperimentDesign
    RigiFlex Q07955
.ExperimentDesign
```

MMMx first displays the predicted aligned error matrix that is part of the AlphaFold output and assigns the folded domains (Figure S4). Then it optimizes the positions of three reference sites in the folded domains and

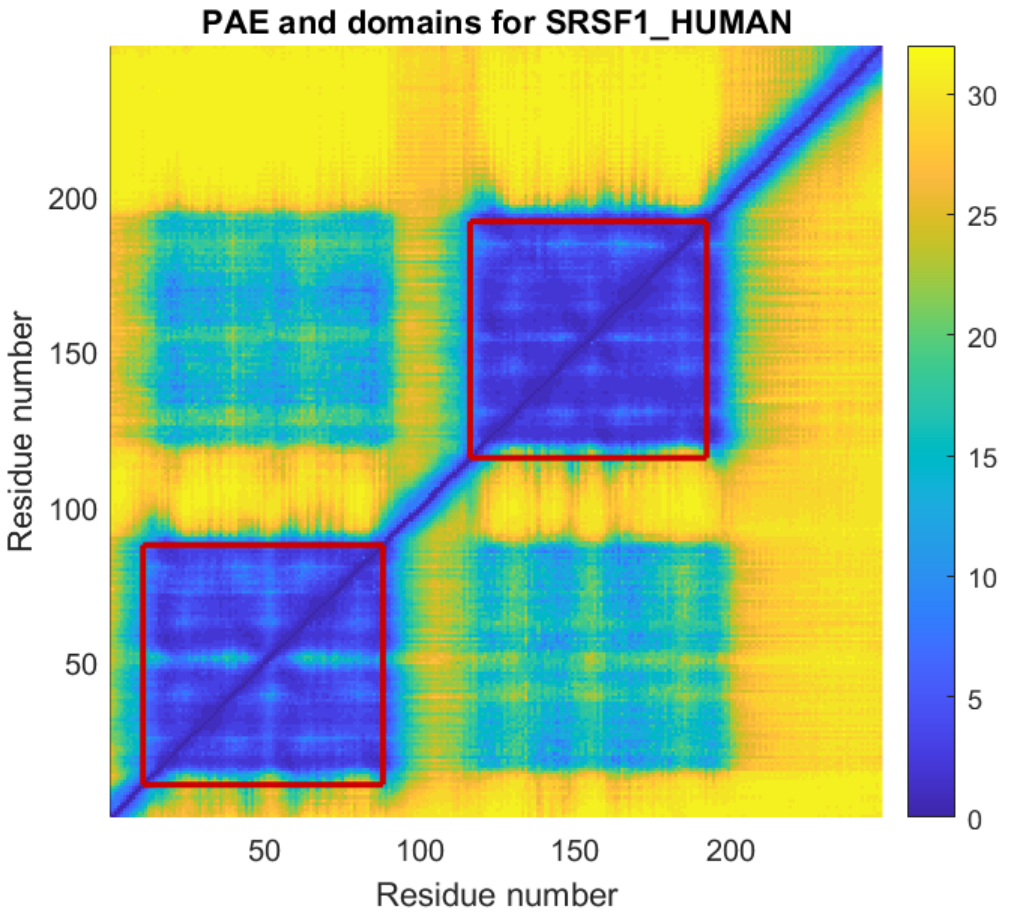

Figure S4: Predicted aligned error for the AlphaFold model of SRSF1 (blue to yellow colours) and assignment of folded domains by MMMx (red squares).

writes a rigid-body PDB file that is needed by the RigiFlex run. Finally, it writes the script sections for the Rigi module call and the three calls for the Flex modules. The task is performed by typing `MMMx SRSF1_RigiFlex` at the Matlab prompt and enter.

This file needs to be amended by experimental restraints. Currently, MMMx supports distance distributions between label sites or atoms (also label-to-atom), small-angle scattering data (SAXS and SANS), and NMR paramagnetic relaxation enhancement (PRE) data.

In a practical case, one may want to adapt the RigiFlex script to other information that exists on the protein and for convenience in spin labelling. This will be shortly discussed for SRSF1, where we did an experimental study with RigiFlex for a cosntruct devoid of the terminal domains.

10 MMMx: Ensemble analysis

In this part, we analyse an ensemble structure for measles virus nucleoproten (400-525) that was submitted by Martin Blackledge's group to the protein ensemble database (PED). This ensemble is based solely on NMR restraints. WE are interested here in what MMMx ensemble analysis can recognize in an ensemble structure without having access to the experimental restraints. The taks is run by typing `analysis_MeV` at the Matlab prompt and enter.

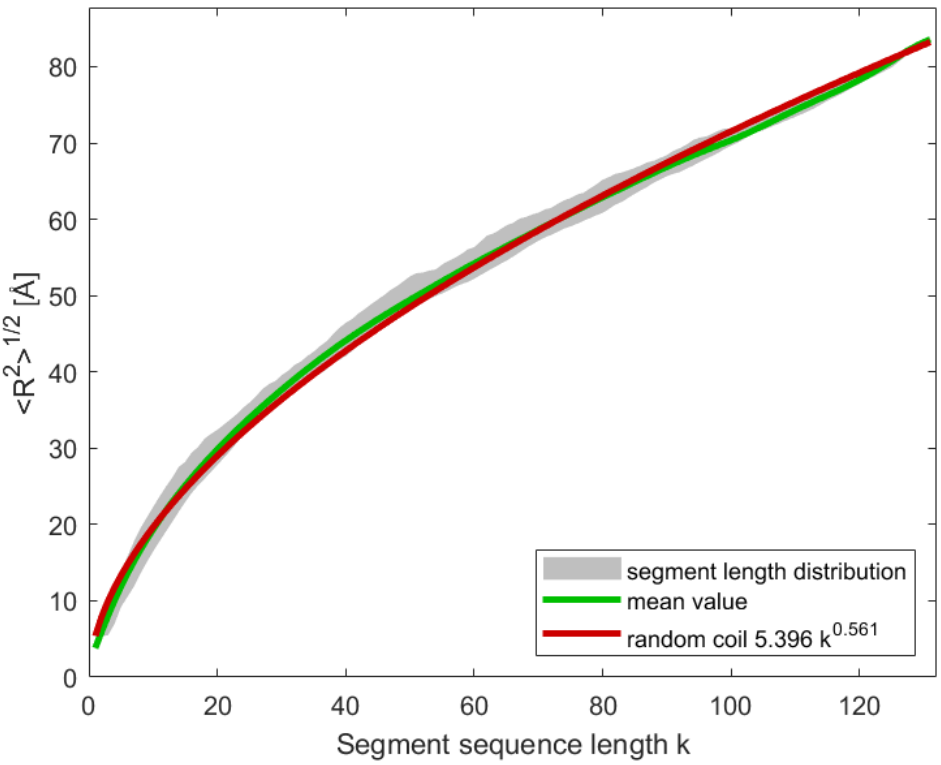

Figure S5: Agreement of the ensemble of measles virus nucleoprotein (PED00020) with a random-coil polymer model.

One output is shown in Figure S5, where root mean square end-to-end distances of all sections of the protein are compared to the expectation fro a random-coil polymer.