

# Dynamix

Ben Tatman

October 21st 2020

Dynamix is a program developed to fit dynamics models (see *Models*) to solid state NMR Relaxation data. This guide will go into what features are available, how to use them, how to format data, and also go into any caveats or issues with the model.

## Features

As of writing, Dynamix can fit the following models;

- Simple Model Free (SMF) Fits two parameters,  $S^2$  and  $\tau$ , with spectral density given as;

$$J(\omega) = \frac{(1 - S^2)\tau}{1 + (\omega\tau)^2}$$

- Simple Model Free with Temperature Dependence (SMFT) Fits three parameters,  $S^2$ ,  $\tau_0$  and  $Ea$ , where the correlation time is temperature dependent;

$$\tau(T) = \tau_0 \exp\left(\frac{Ea}{RT}\right)$$

This is then fit as in SMF.

- Extended Model Free (EMF) Fits three parameters,  $S_{\text{slow}}^2$ ,  $\tau_{\text{slow}}$ ,  $\tau_{\text{fast}}$ . The fast order parameter,  $S_{\text{fast}}^2$ , is calculated as  $S_{\text{dipolar}}^2/S_{\text{slow}}^2$ . The spectral density is given as;

$$J(\omega) = \frac{(1 - S_f^2)\tau_f}{1 + (\omega\tau_f)^2} + \frac{S_f^2(1 - S_s^2)\tau_s}{1 + (\omega\tau_s)^2}$$

- Extended Model Free with Temperature Dependence (EMFT) Five parameter model, fitting  $S_{\text{slow}}^2$ ,  $\tau_{0,\text{slow}}$ ,  $\tau_{0,\text{fast}}$ ,  $Ea_{\text{slow}}$ ,  $Ea_{\text{fast}}$ .  $S_{\text{fast}}^2$  is calculated as in EMF, and the correlation times are temperature dependent as in SMFT. The spectral density is as in EMF.
- Extended Model Free without Dipolar Approximation (DEMF) Four parameter model,  $S_{\text{slow}}^2$ ,  $S_{\text{fast}}^2$ ,  $\tau_{\text{slow}}$ ,  $\tau_{\text{fast}}$ . Spectral density as in EMF.
- Extended Model Free without Dipolar Approximation, with Temperature Dependence (DEMFT) Six parameter model, fitting  $S_{\text{slow}}^2$ ,  $S_{\text{fast}}^2$ ,  $\tau_{0,\text{slow}}$ ,  $\tau_{0,\text{fast}}$ ,  $Ea_{\text{slow}}$ ,  $Ea_{\text{fast}}$ . Spectral density as in EMF.
- Gaussian Axial Fluctuations (GAF) Eight parameter model:  $\tau_{\text{slow}}$ ,  $\tau_{\text{fast}}$ ,  $\sigma_{\text{slow}}^\alpha$ ,  $\sigma_{\text{slow}}^\beta$ ,  $\sigma_{\text{slow}}^\gamma$ ,  $\sigma_{\text{fast}}^\alpha$ ,  $\sigma_{\text{fast}}^\beta$ ,  $\sigma_{\text{fast}}^\gamma$ . These axial fluctuations are then used to derive order parameters, which are then fit to the EMF spectral density function.

$$S_{\mu\nu}^2 = \frac{4\pi}{5} \sum_{l,k,k',m,m'=-2}^2 = (-i)^{k-k'} \exp\left(-\frac{\sigma_\alpha^2(k^2 + k'^2)}{2} - \sigma_\beta^2 l^2 - \frac{\sigma_\gamma^2(m^2 + m'^2)}{2}\right) \times \\ d_{kl}^{(2)}\left(\frac{\pi}{2}\right) d_{k'l}^{(2)}\left(\frac{\pi}{2}\right) d_{mk}^{(2)}\left(\frac{\pi}{2}\right) d_{m'k'}^{(2)}\left(\frac{\pi}{2}\right) Y_{2m}(e_\mu^{pp}) Y_{2m'}^*(e_\nu^{pp})$$

Where parameters are defined as in Lienin 1998. The relaxation rates take into account dipolar contributions between N-H, N..H(rest), C-N, CA-N, C-H, C..H(rest), C-N, C-CA, as well as the anisotropic chemical shifts of nitrogen and carbon.

- Gaussian Axial Fluctuations with Temperature Dependence (GAFT) Ten parameter model. All of those in GAF, plus fast and slow activation energies. Temperature dependent time constants calculated as in SMFT, EMFT and DEMFT, angles used to calculate  $S^2$ , then fit to EMF spectral density.
- Model Free with slow Gaussian Axial Fluctuations (EGAF) Six parameter model. Implements MF order parameter on fast time scale and GAF order parameter for slow motion.
- Model Free with slow Gaussian Axial Fluctuations with Temperature Dependence (EGAFT)

These models can be fit to  $^{15}\text{N}$  and  $^{13}\text{C}$   $R_1$  and  $R_{1\rho}$  values, as well as C-C, C-H, N-H and C-N dipolar order parameters.

Passing the ‘OR\_VARY = 1’ parameter alongside a GAF model (EGAF, EGAFT, GAF, GAFT) will transform it into a variable orientation model, in which the orientation of the axes fit to the peptide plane are allowed to vary according to three rotations,  $\alpha$ ,  $\beta$ ,  $\gamma$ . This rotation is implemented as a rotation of the second order spherical harmonics in the GAF order parameter term via Wigner D matrices. In effect,

$$Y_l^{m'}(r') = \sum_{m=-l}^l l[D_{m'm}^{(l)}(R)]^* Y_l^m(r)$$

The initial orientation of the X, Y, Z axis has Z aligned along CA-CA, with the CA-N positive relative to CA-C. X is roughly parallel to the C-O bond (C->O positive) and Y is perpendicular to X and Z such that the standard X, Y, Z convention is retained (eg  $Y = Z \times X$ ). These variable models are generally termed **VGAF**, **VGAFT**, **VEGAF**, **VEGAFT**.

For each of these models, Dynamix will perform a user specified number of optimizations with random starting points using the Nelder-Mead simplex method to find an optimum. Each optimum is output into a **residue\_N.dat** file. Once complete, it will perform back calculations for each relaxation data point, outputting these into **backcalc\_N.dat** files. If one of the GAF modes is used, it will calculate effective  $S_{\text{NH}}^2$  order parameters and output these into **gaf.dat**.

If error mode is enabled, it will perform a further set of optimizations where the starting point is set to the optimized parameters. Back calculated relaxation rates are then varied within experimental error, and optimization performed. The new optimum points for each repeat in the error calculations are then used to determine standard deviations for the optimized values<sup>1</sup>.

## Compilation

To compile the program from source on Karplus, navigate to the directory above **src/** and run

```
gcc src/main.c -lm -pthread -o dynamix -O3
```

This will pull in all other required C files in the **src/** directory, load the math(s) and pthread libraries, using optimization level 3, and output the program into **dynamix**. Also contained in the **src/** directory is documentation (build using **doxygen** **Doxyfile** if not up to date) and test programs to verify the model against the MATLAB scripts.

## Data Formats

There are four data formats used to provide data to Dynamix. In all of these, beginning a line with ‘%’ marks it as a comment.

### System File

The system file contains information about the model being fit, and how to fit it. It also tells Dynamix where to find all other data. Generally I use the file extension **.dx** for these, though it is not necessary.

The first portion of the system file consists of key value pairs laid out as

---

<sup>1</sup>Note that the output will have the minimum optimized points and two standard deviations for the errors; it will not output the mean of the error calculations (unless you explicitly change the code to do so - at the moment this is line 408 of **main.c**, in which you should change **m.residues[1].parameters[i]** to **m.residues[1].errors\_mean[i]**.

KEY = VALUE

The keys are all upper case, and there must be spaces on either side of the equals.

Key	Value
MODEL	Model being fit - see Features
S2NH	File containing dipolar order parameters for N-H as defined below
S2CH	File containing dipolar order parameters for C-H
S2CC	File containing dipolar order parameters for C-C (not currently used)
S2CN	File containing dipolar order parameters for C-N
CSISON	File containing isotropic chemical shifts for $^{15}\text{N}$
CSISOC	File containing isotropic chemical shifts for $^{13}\text{C}$
N_RESIDUES	Number of residues - <b>must</b> be the same as the number of lines in each input file (or bad things may happen)
OUTPUT	Directory (eg <code>output/</code> ) to place output files into
N_ITER	Number of iterations for optimization
N_ERROR_ITER	Number of iterations to perform for error calculation
IGNORE	Residue to ignore; each residue to ignore should have its own line
OR_NH	N-H orientations
OR_NC	N-C orientations
OR_NCA	N-C $\alpha$ orientations
OR_NCSAxx/yy/zz	Nitrogen chemical shift anisotropy orientations
OR_CCAp	$^{13}\text{C}'$ - $^{13}\text{C}_{i-1}^{\alpha}$ orientation
OR_CCAc	$^{13}\text{C}'$ - $^{13}\text{C}_i^{\alpha}$ orientation
OR_CN	C-N orientation
OR_CNH	C-amide proton orientation
OR_CCSAxx/yy/zz	Carbon chemical shift anisotropy orientations
NTHREADS	Number of threads to run; generally, set to the number of processors you want to run it on
OR_VARY	Whether or not the orientation of the GAF axes relative to the peptide plane should be allowed to vary.

For example

```

MODEL = EMF
S2DIP = system/s2_dipolar.csv
CSISON = system/csisoN.csv
CSISOC = system/csisoC.csv
N_RESIDUES = 56
OUTPUT = output/
N_ITER = 200
IGNORE = 42
N_ERROR_ITER = 200
% Orientations taken from Lienin 1998
OR_NH = system/or_NH.csv
OR_NC = system/or_NC.csv
OR_NCA = system/or_NCA.csv
OR_NCSAxx = system/or_NCSAxx.csv
OR_NCSAyy = system/or_NCSAyy.csv
OR_NCSAzz = system/or_NCSAzz.csv
OR_CCAp = system/or_CCAp.csv
OR_CCAc = system/or_CCAc.csv
OR_CN = system/or_CN.csv

```

```
OR_CNH = system/or_CNH.csv
OR_CCSAxx = system/or_CCSAxx.csv
OR_CCSAyy = system/or_CCSAyy.csv
OR_CCSAzz = system/or_CCSAzz.csv
NTHREADS = 4
```

This key value header should be followed by **#RELAXATION**, and then each file containing Relaxation Data should be listed below.

```
#RELAXATION
system/15N_R1_600_50_0_271.csv
system/15N_R1_600_50_0_282.csv
...
```

## Relaxation Data

Each piece of relaxation data (eg a  $^{15}\text{N}$   $R_1$  measurement made at 300 K in 600 MHz at 50 kHz) should be placed into an individual file. This file should begin with a header.

*Note that there must be a space on either side of the equals sign*

```
FIELD = {field in MHz}
WR     = {spinning frequency in Hz}
W1     = {spin lock frequency in Hz}
TEMP   = {temperature in kelvin}
TYPE   = {15NR1, 15NR1p, 13CR1, 13CR1p}
```

This defines the ‘global’ parameters for this relaxation data. Then, a line **#DATA** denotes the start of the actual relaxation data, and should be followed by the relaxation data.

```
#DATA
1 {relaxation rate in s-1} {error in s-1 (2 standard deviations)}
2 {relaxation rate in s-1} {error in s-1 (2 standard deviations)}
...
```

The program requires a line for each residue; if you have no relaxation data for one residue, denote the relaxation rate and error with -1 instead of leaving it blank. This will tell Dynamix that the data does not exist, as opposed to it just being omitted by mistake. The program may run without this, however it will not necessarily inform you that there is insufficient data for fitting which may lead to incorrect fitting and data response.

## Orientation Data Formats

For the GAF models the orientation of each interaction vector related to the  $\alpha$ ,  $\beta$  and  $\gamma$  motional axis is required. In the form of Lienin 1998, this is taken as  $\theta$  and  $\phi$  angles. For each orientation vector there should be a file containing this data as;

```
1 {theta} {phi}
2 {theta} {phi}
...
```

This data is only needed for GAF fits, but all of it is needed. The orientations required are described below.

## Residue Data Formats

Data such as isotropic chemical shifts and dipolar order parameters should be kept in files laid out as;

```
1 {value} {error}
2 {value} {error}
...
```

## Running The Model

Once the file is setup, the model may be run as;

```
./dynamix {path to .dx file}
```

This will output the various threads being spawned as the program operates. Passing the `-e` option;

```
./dynamix {path to .dx file} -e
```

Will enable error calculation.

## Visualising Results

### Back Calculations

Using `gnuplot` the output of Dynamix can be quickly visualised. To view residue specific back calculated data, simply run (after entering the output directory)

```
> gnuplot
...
gnuplot> plot 'backcalc_N.dat' u 2:3 w points pt 7, x lw 3
```

This will produce a graph showing the calculated values (x) against experimental values (y), with a line of  $y = x$ . In order to view all back calculated points for all residues, run;

```
> cat backcalc_* > backcalc.dat
> gnuplot
...
gnuplot> plot 'backcalc.dat' u 2:3 w points pt 7, x lw 3
```

If the 'VERBOSE' key is enabled in `datatypes.c`, the back calculated files will also contain experimental information. Though beyond the scope of this brief introduction, these can be plot onto the graphs too to show where variation in relaxation rate arises (for example, I find that 700 and 850 MHz data tends to much better fit the  $y = x$  line than 600 MHz data. I don't know if this is some systematic experimental difference, or just the model validity breaking down).

### Visualising Parameters

The outputs of Dynamix may be analysed using Python files present in the `utils` directory.

- `aicbic.py` calculates Akaike, Bayesian, and adjust Akaike Information Criteria for each model<sup>2</sup>. Currently only relaxation data is used. Input is taken as `python3 aicbic.py <folder> <model> <output>`, e.g. `python3 aicbic.py gaff/ gaff gaff.csv` will output the values into `gaff.csv`, which will consist of columns `residue`, `AIC`, `BIC`, `AICc`. This output may require sanitization (e.g. removing `-1` or other error codes).
- `combine_aicbic.py` calculates AIC, BIC and AICc for comparison between multiple models. Input is taken as a list of folders, where the folder name is the model, e.g.; `python3 combine_aicbic.py gaff gaff egaff egaff`. This will produce three files, `AIC.csv`, `BIC.csv`, `AICc.csv` containing the corresponding IC values in columns ordered as given in the command. The output will additionally provide counts for how many residues were best fit for each model.
- `gen_attr.py` generates an attribute file which may be used with Chimera or ChimeraX for visualising results. Arguments are of the format `python3 gen_attr.py <filename> <model> <tag> <outputfile>` where `<filename>` is the `final.dat`, `<model>` should be the model (lower case, standard form given above e.g. `gaff` for 3D-GAF), `<tag>` should be the protein tag in Chimera (typically `#1`, though this may vary for complexes), and `<outputfile>` is an attribute file. Attributes defined include `chisq`, and others which are labelled as given in the output plots from `plot.py` (or may be seen in `models.py`).
- `plot_relax.py` generates EPS files `<model>_relax.eps` containing all relaxation data separated by conditions for easy visualisation. Arguments are taken as a list of folders containing `backcalc_*.eps` files. e.g. `python3 plot_relax.py gaff gaff egaff egaff`.
- `plot.py` generates EPS files `<model>_params.eps` containing the calculated parameters. Input is taken as a list of folders where by default the folder name is taken as the model. For non specific folder names, the parameter `-m<model>` may be passed. If error calculation was enabled, these may be plotted as error bars using the `-e` flag. For example, `python3 plot.py gaff -mgaff -e` will plot `/gaff/` as a GAF model with error bars, while `python3 plot.py egaff egaff demf demft` will produce non error bar plots of those four models.

---

<sup>2</sup>As defined in the SI for Good, D. B., Wang, S., Ward, M. E., Struppe, J., Brown, L. S., Lewandowski, J. R. & Ladizhansky, V. Conformational Dynamics of a Seven Transmembrane Helical Protein Anabaena Sensory Rhodopsin Probed by Solid-State NMR. J Am Chem Soc 136, 2833–2842 (2014).

- `gen_build.py` generates ChimeraX BILD files for GAF models. A PDB representing the protein of interest should be given in the folder, and the path to this defined on line 9. It may then be run as `python3 gen_build.py <filename> <outputfile> <v flag> <type>`, where `<filename>` and `<outputfile>` are as given in `gen_attr.py`, `<v flag>` is 1 if variable orientation is enabled<sup>3</sup>.

## Output File Formats

Note: All taus and tauf values are output in nanoseconds ( $10^{-9}$  seconds), not seconds. This is done to improve precision.

### residues\_N.dat

These residue files are generated during the optimization process. Each line refers to a separate optimization process. The first column is the residue number, the second is the minimum value of the chisq function. Plotting the second column generates a nice plot showing how well the optimization went.

The remaining columns depend on which model is in use. In particular;

- SMF: tau (ns), S2
- SMFT: tau (ns), S2, Ea
- EMF: taus (ns), S2s, tauf (ns)
- EMFT: taus (ns), S2s, tauf (ns), Eas, Eaf
- DEMF: taus (ns), S2s, tauf (ns), S2f
- DEMFT: taus (ns), S2s, tauf (ns), S2f, Eas, Eaf
- EGAF: taus (ns), tauf (ns), sAs, sBs, sGs, S2f
- EGAF: taus (ns), tauf (ns), sAs, sBs, sGs, S2f, Eas, Eaf
- GAF: taus (ns), tauf (ns), sAs, sBs, sGs, sAf, sBf, sGf
- GAFT: taus (ns), tauf (ns), sAs, sBs, sGs, sAf, sBf, sGf, Eas, Eaf

This may be useful for plotting to see how varied the individual responses are, eg how responsive the model is to one parameter. If `OR_VARY` is enabled, the final three columns will be the  $\alpha$ ,  $\beta$  and  $\gamma$  angles.

### final.dat

This contains the final fit parameters. The first column is the residue number, the second is the S2 dipolar value, then the minimum chisq value, followed by the model specific columns as above. The final column is a count of how many repeats went into error calculation (or 0 if this was not done).

### errors.dat

This file is only output if error mode is enabled. It is the same as `final.dat`, only after each parameter there is the error (2 standard deviations). eg;

```
{residue number}, {S2dipolar}, {chisq minimum}, {tau}, {tau error}, {S2}, {S2 error}
```

### gaf.dat

This file is only produced for GAF models. This is laid out as;

```
{residue number}, {tau s}, {S2 slow}, {tau f}, {S2 fast}
```

Where  $S_{\text{slow}}^2$  and  $S_{\text{fast}}^2$  are the order parameters for the N-H dipolar interaction.

### backcalc\_N.dat

These files contain back calculations. They are set out;

```
{relaxation number} {calculated R} {measured R} {error in measured R}
```

Plotting `{calculated R}` against `{measured R}` can give a useful indication of goodness of fit. To gain more insight into how different experimental parameters affect the fit, enable the `VERBOSE` flag in `datatypes.c` (eg set it to 1). This will then output a further few columns;

```
{field in MHz} {spinning frequency Hz} {spin lock frequency Hz} {temperature K}
```

---

<sup>3</sup>Not that I'm not entirely convinced I have the sense of the rotations in `gen_build.py` correct for variable orientations.