FOCUS: a program for analyzing molecular dynamics simulations, featuring digital signal-processing techniques

D. J. Osguthorpe and P. Dauber-Osguthorpe

Molecular Graphics Unit and School of Chemistry, University of Bath, UK

FOCUS is a program for analyzing molecular dynamics simulations. It enables the researcher to monitor structural and energetic properties during the trajectory, and to calculate the corresponding statistical averages, correlation functions and Fourier transforms. In addition to these conventional methods, the program also utilizes novel methods based on digital signal-processing techniques to characterize the various motions. The characteristic frequencies in the system are revealed by the frequency distribution function g(ν)*, which is calculated from the Fourier transform of the atomic coordinates. A filtering technique is employed to remove uninteresting motion (e.g., high-frequency bond stretching) while retaining and focusing on important motion (e.g., low-frequency conformational motion). The filtering technique enables fast display of slow events without getting a blurry or jittery picture due to the high-frequency motions. Another new way for analyzing the motion is by extracting "characteristic modes" and associated frequencies. This yields a pictorial description of the oscillatory motions in a manner analogous to normal mode analysis.*

Keywords: molecular mechanics, molecular dynamics, Fourier transforms, characteristic modes, frequency distribution

INTRODUCTION

In recent years the theoretical study of the dynamic behavior of molecules has expanded rapidly. Molecular dynamics simulations have been applied to many molecular systems, from small organic molecules to large assemblies of molecules.[1,2] Molecular dynamics simulations result in a complex and comprehensive description of the molecular motion. To gain insight into the dynamic behavior of molecules it is important to be able to identify different types of motion and to characterize them. The program FOCUS (Finally One Can Understand Simulations)[3] was designed as a tool for the analysis and interpretation of molecular dynamics simulations. This program incorporates newly developed techniques aimed at characterizing the motion of the system (along with the conventional statistical analysis of the trajectories revealing the average properties of the system).

One of the most common methods for examining the results of a molecular dynamics simulations is visual inspection of the trajectory on a graphics display system. However, visual analysis of conformational motion is complicated by the long time scale of the motion and the hindering nature of the superimposed high-frequency motion. A fast display rate causes blurry pictures, while skipping frames results in jerky motion. The filtering technique enables one to separate low-frequency conformational movement (and the associated energies) from the high-frequency "jittering" of bond stretching and angle bends.[4–6]

The characteristic frequencies of motion are revealed by the frequency distribution $g(\nu)$ which is obtained from the Fourier transform of the atomic trajectories.[5] We also have developed a method that allows extraction of the characteristic modes of motion from the molecular dynamics trajectory and thus classification of the modes of motion in an analogous way to normal mode analysis.[7]

The tasks the program can perform currently are:

- Calculate the frequency distribution of the system $g(\nu)$, and the thermodynamic properties (entropy and free energy).
- Filter trajectories (Cartesian coordinates, internal coordinates and energies).
- Extract characteristic modes.
- Calculate trajectories of structural properties (i.e., bonds, angles, torsions, nonbonded interatomic distances, etc.).
- Create trajectories of temperature, total energy, kinetic and potential energy, and potential energy components.
- Calculate averages and standard deviations of structural and energy properties.

Color Plates for this article are on page 164.

Address reprint requests to Dr. Osguthorpe at the Molecular Graphics Unit and School of Chemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK.

Received 19 September 1991; accepted 24 September 1991

- Create "snapshots" of transient structures along the trajectory, by writing sets of coordinates for specific time steps, for further analysis (e.g., display, minimize, plot, etc.).
- Calculate average coordinates and standard deviations (for comparison with experimental structure and temperature factors).
- Calculate correlation functions for the various properties.
- Calculate the Fourier transform of the various properties.
- Plot structural and energy trajectories, correlation functions, Fourier transforms, and $g(\nu)$.

METHODS

Ordering molecular dynamics trajectories

In molecular dynamics simulations, at each time step all the coordinates of the system are updated to create the next time step. At constant intervals the coordinates of all atoms (and the energies and atomic velocities) are written to a file to create the "general history," or trajectory file. Thus, a trajectory file consists of a set of atomic coordinates for a time step, followed by a set of atomic coordinates for the next time step, and so on. For many applications, a format in which all the time steps for one coordinate are given sequentially is essential or at least more efficient. In particular, any analysis involving the calculation of the Fourier transform of a property requires simultaneously the values of this property for the complete time interval. Thus, the first task of the program is usually to reorder the trajectory file produced by the molecular dynamics simulation program into the appropriate format. This is achieved by reading the trajectory file into an internal buffer, and writing a new file with reordered coordinates. The new file is stored on disk as a direct-access file (DA file).

The same process is applied to other properties, such as internal coordinates or energies. For energies this involves storing all the energy components in an internal buffer and writing them to a DA file. For internal coordinates, the Cartesian coordinates of each time step are read in, and the required internal coordinates are calculated and stored in the internal buffer. This is repeated for all time steps, and the full trajectory of each internal coordinate is written to a DA file.

Time averages and standard deviations

The statistical average and standard deviation of a property P are given by

$$\begin{aligned} \langle p \rangle &= 1/N \sum p(t_k) \\ \langle p^2 \rangle &= 1/N \sum p^2(t_k) \\ \sigma &= (\langle p^2 \rangle - \langle p^2 \rangle)^{1/2} \end{aligned} \tag{1}$$

where $p(t_k)$ is the value of the property at time t_k; $\langle p \rangle$ and $\langle p^2 \rangle$ are the time averages of the property and the square of the property respectively; and σ is the standard deviation.

Correlation functions and Fourier transforms

The cross-correlation function $c(t)$ of two properties $p(t)$ and $q(t)$ is defined by

$$c(t) = \int_{-\infty}^{\infty} p(\tau)q(t + \tau)\, d\tau \tag{2}$$

The Fourier transform of the correlation function is

$$\begin{aligned} C(\nu) &= \int_{-\infty}^{\infty} c(t)e^{-j\nu t}\, dt \\ &= P^*(\nu)Q(\nu) \end{aligned} \tag{3}$$

where P and Q are the Fourier transforms of p and q.

The auto-correlation function of a property and the corresponding Fourier transform are obtained from the above by using $q = p$.

The calculation of the correlation functions is a time-consuming task; it is much more efficient to calculate the corresponding Fourier transforms directly, and to obtain the correlation functions by applying an inverse Fourier transform to the resulting function.[8]

Frequency distribution and thermodynamic properties

The frequency distribution function $g(\nu)$ gives the number of characteristic motions in the system having a frequency ν. This function is obtained from the trajectories of atomic coordinates:[5]

$$\begin{aligned} g(\nu) &= (1/kT)\nu^2 \sum_i H_i^2(\nu) \\ H_i(\nu) &= \int_{-\infty}^{\infty} q_i(t)e^{-j\nu t}\, dt \end{aligned} \tag{4}$$

where $q_i(t)$ are the mass-weighted coordinates, and $H_i(\nu)$ are the corresponding Fourier transforms. The thermodynamic properties can be obtained from $g(\nu)$ using the Einstein equations:[9]

$$\begin{aligned} E_0 &= 1/2 \int h\nu g(\nu)\, d\nu \\ E &= E_0 + \int [g(\nu)h\nu/(e^{h\nu/kT} - 1)]\, d\nu \\ A &= E_0 + \int g(\nu)kT \ln(1 - e^{-h\nu/kT})\, d\nu \\ S &= (E - A)/T \end{aligned} \tag{5}$$

where E_0, E, A and S are the zero-point energy, the vibrational enthalpy, the free energy and the entropy, respectively.

Filtering molecular dynamics trajectories

A new method that we have developed enables us to carry out a fully flexible molecular dynamics simulation, and to focus on the motions that are of interest (in particular the

low-frequency conformational modes) in the analysis.[4,5] This method is based on digital signal-processing techniques, in which filtering is used to remove "noise" from an electronic signal.[10] Here the individual atomic trajectories are treated in an analogous way to electronic signals. The technique involves three steps:

(1) Fourier transforming each atomic trajectory to the frequency domain:

$$H_i(\nu) = \int_{-\infty}^{\infty} x_i(t) e^{-j\nu t} dt \tag{6a}$$

(2) Applying a filtering function to remove the high-frequency components:

$$F(\nu) = 1 \qquad \nu_{\min} < \nu < \nu_{\max}$$

$$F(\nu) = 0 \qquad \nu < \nu_{\min};\ \nu > \nu_{\max} \tag{6b}$$

$$H_i'(\nu) = H_i(\nu)F(\nu)$$

(3) Inverse Fourier transforming back to the time domain:

$$x_i'(t) = \int_{-\infty}^{\infty} H_i'(\nu) e^{j\nu t}\, d\nu \tag{6c}$$

Extracting normal modes

The Fourier transform of a coordinate gives the amplitude of oscillation of this coordinate for each frequency. The set of amplitudes of all coordinates for a specific frequency defines the "characteristic mode" of motion at that frequency, in analogy to normal mode analysis. Thus, by partitioning the Fourier transform of the Cartesian coordinates into frequency regions the characteristic modes of motion in the molecular dynamics trajectory can be obtained.[7]

PROGRAM DESCRIPTION

The program is written in FORTRAN 77. It runs both under VMS (microVAX II and VAX workstation) and Unix (Stardent Titan) operating systems.

Task execution strategy and data manipulation

As mentioned above, the first task in the analysis will usually be to rearrange the information in the general history file. FOCUS achieves this objective by writing a set of selected properties to a DA file. In principle, the program needs to store all the coordinates (or any other property) for all time steps in an internal buffer. The size of this buffer would have to be $3*N_{atom}*N_{time}$, which for large systems or long simulations may be prohibitively large. Thus the reordering is usually carried out in a few cycles where each cycle orders all coordinates in a segment of the time.

The properties on the DA file can be processed easily to produce statistical information, or to calculate related properties (e.g., filtered properties), which can be written to a new DA file.

Input and output files

The program requires a standard input file (logical unit 5) in which the user defines the required operations (see examples in the Appendix), and produces a standard output file (logical unit 6) that contains a log of the job's progress and error/warning messages. Jobs that request plotting produce a "plot file" (logical unit 4), which contains the plotting code for a calcomp system. These three files have to be assigned explicitly in a control file (a ".com" file on VAX/VMS). All other file names are defined in the standard input file.

For each property two files are generated—a DA file containing the trajectories and an information file containing relevant data, such as step size, initial step number, etc. Information about the file name for each property is maintained in a general information file.

Program capabilities and requirements

Most important tasks of the program require reading from or writing to a DA file. For large systems this may require a significant amount of memory, to process the information, and of disk space, to store the DA file. A reduction in memory can be achieved by processing small time periods at a time; however, this results in longer processing time.

Although generating the trajectories can be quite demanding in CPU time the analysis stage is much less demanding. For the N-acetylalanine-N'-methylamide molecule a 133-ps simulation took ≈33 hours on a microVAX II, and generated ≈90 Mbytes of history files (coordinates, velocity and energies). Reordering the coordinates of the whole trajectory took 14 minutes and generated a file of 8.7 Mbytes. Filtering the coordinates required 25 minutes.

The current dimensions of the program are given in Table 1.

Command language

Each task the program has to perform is defined by a command and a set of associated parameters. The command name and the values of the various parameters are defined by a free format string. The values of parameters can be numeric, strings or tokens. Some parameters can have lists of numbers or strings as values. Many commands can be given in one file. The corresponding tasks will be executed in the order the commands are given.

Table 1. Program dimensions

Number of atoms	3600
Number of bonds	3000
Number of angles	3000
Number of torsions	3000
Number of steps	32768
Internal buffer	1000000

APPLICATIONS

Frequency distribution

A useful first step in characterizing the motion during a molecular dynamics simulation is the calculation of the frequency distribution $g(\nu)$. See Equation (4). An example of this function for the molecule acetamide is given in Figure 1. Each peak corresponds to a characteristic motion. The nature of the motion corresponding to a peak can sometimes be inferred from the frequency (e.g., frequencies at $\approx$3000 cm^{-1} correspond to CH stretches). A more detailed understanding can result only from applying appropriate filters to the coordinates and monitoring the motion on a graphics display or by extracting the modes.

Filtering coordinates and energy

A molecular dynamics trajectory of N-acetylalanine-N′-methylamide was carried out for 133 ps. During the simulation the molecule undergoes several conformational transitions between an α_R and a C_7^{eq} conformation. To study the conformational motion in detail we applied a low-pass filter with ν_{max} = 70 cm^{-1} to the coordinates and a filter with ν_{max} = 140 to the energy. Torsion and valence angles where calculated from the Cartesian coordinates of the original and filtered trajectory. Figure 2 compares the filtered and unfiltered trajectories of ϕ, ψ, the $C_\alpha CO$ angle and the total potential energy. This figure demonstrates that the filtering process retained the changes in ϕ and ψ characteristic of the conformational changes. Moreover, filtering also revealed the accompanying changes in valence angles and in the potential energy, which were difficult to detect in the original trajectory due to the superimposed large amplitude high-frequency fluctuations. In Color Plate 1 we displayed two series of snapshots of the molecule along one of the conformational transitions in the original and filtered trajectory. In general the two series are very similar, but in the original trajectory there are many local distortions while the filtered transition appears smooth. This difference reflects one of the major advantages of filtering—it is possible to display "frames" at large intervals without "jerkiness" and thus faster display of slow events is possible.

Extracting characteristic Modes of Motion

We have extracted the 21 characteristic modes of acetamide from a molecular dynamics simulation. The set of normal modes obtained by performing a classical normal mode analysis and the set extracted from the molecular dynamics simulation are compared in Color Plates 2a and 2b, respectively. The frequencies and vectors of most modes are nearly identical. A quantitative comparison was performed by calculating the correlation (dot product) C_{ij} between the two sets of vectors. In general $C_{ij} = \delta_{ij}$. However, some intermixing of the CH bends and of the amide bond torsion with the NH out of plane has occurred in the dynamics, as indicated by $C_{ij} \neq 0$.

This type of representation of characteristic modes is particularly useful when trying to reduce the space and time multidimensional characteristics of motion into a two-dimensional representation. It provides a means of describing the motion without the use of interactive graphics display.

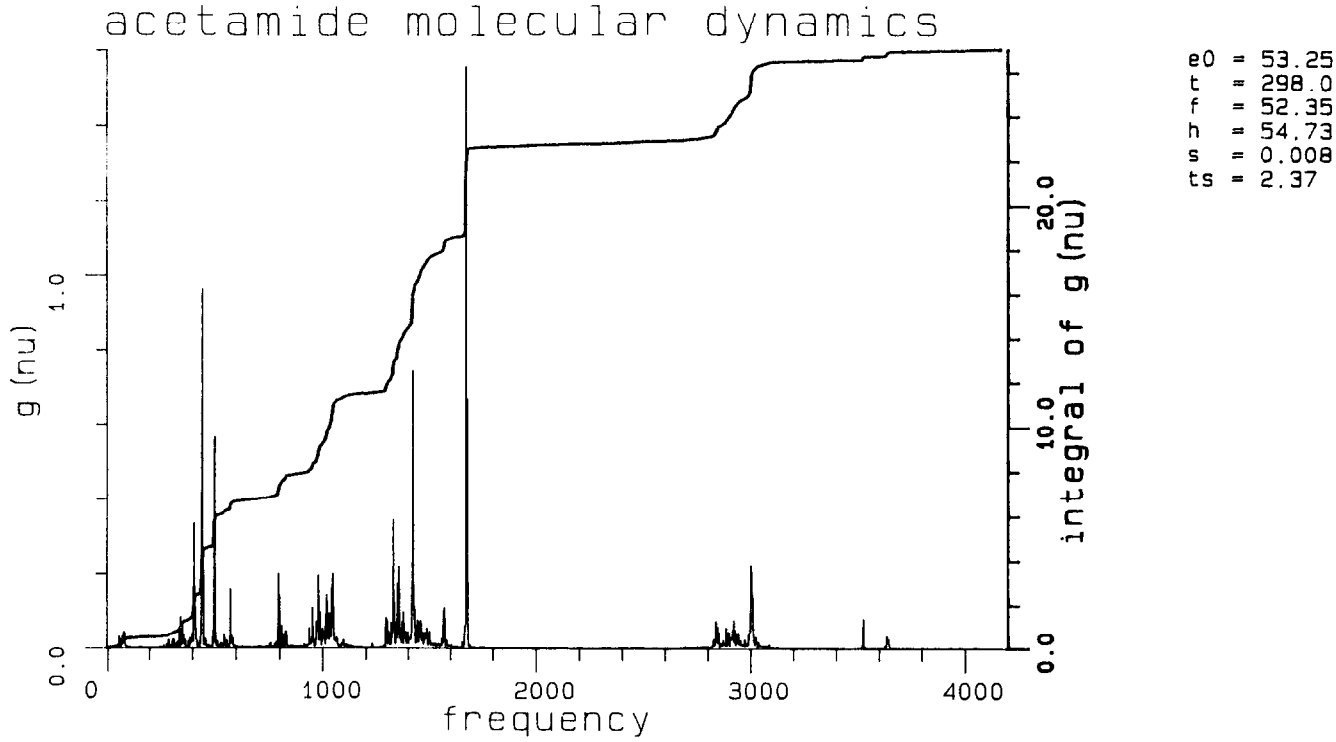

Figure 1. Frequency distribution $g(\nu)$ in a molecular dynamics trajectory of acetamide.

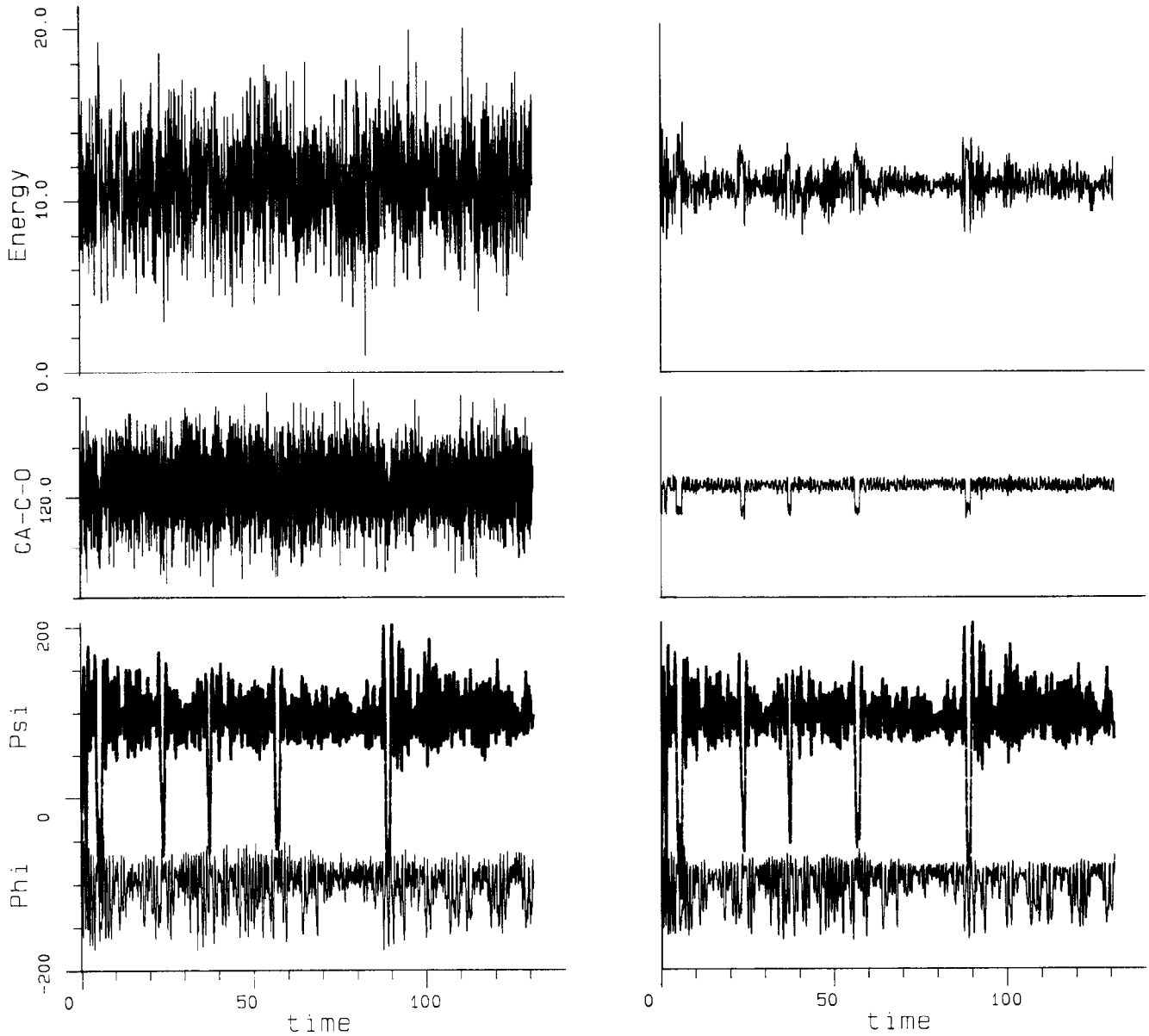

Figure 2. Trajectories of ϕ, ψ, the valence angle C^{α}—C=O and the total potential energy of N-acetylalanine-N'-methylamide. The original trajectories are given on the left and the filtered trajectories are on the right. (Angles are in deg and time is in ps).

CONCLUSIONS

The program FOCUS enables the user to carry out conventional analyses of molecular dynamics trajectories, such as calculating and plotting structural or energy properties and the associated averages, standard deviations, correlation functions and Fourier transforms. It also utilizes new techniques to calculate frequency distributions, filter out unwanted motion and extract characteristic modes of motion. The filtering method enables the user to focus on motions of interest, in particular low-frequency conformational motion. Extracting typical modes provides pictorial descriptions of all the motions during the dynamics. These techniques are very flexible and can be applied to the molecular system as a whole or to a subset of atoms of interest. Fourier transform calculations are very fast and therefore the analysis stage is not very demanding in computational resources. The program is therefore a very powerful tool for enhancing the understanding of molecular motion.

REFERENCES

1 McCammon, J.A. and Harvey, S.C. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge (1987)

2 Allen, M.P. and Tildesley, D.J. *Computer Simulation of Liquids*. Clarendon, Oxford (1987)

3 The program is distributed by Oxford Molecular Ltd., Magdalen Centre, Oxford Science Park, Sandford-on-Thames, Oxford OX4 4GA, UK
4 Sessions, R.B., Dauber-Osguthorpe, P. and Osguthorpe, D.J. *J. Mol. Biol.* 1989, **210**, 617–634
5 Dauber-Osguthorpe, P. and Osguthorpe, D.J. *J. Amer. Chem. Soc.* 1990, **112**, 7921–7935
6 Dauber-Osguthorpe, P. and Osguthorpe, D.J. *Biochemistry*. 1990, **29**, 8223–8228
7 Dauber-Osguthorpe, P. and Osguthorpe, D.J. work in progress
8 Brigham, E.O. *The Fast Fourier Transform and its Applications*. Prentice-Hall, Englewood Cliffs (1988)
9 Hill, T.L. *An Introduction to Statistical Thermodynamics*. Addison-Wesley, Reading (1960)
10 Oppenheim, A.V. and Schafer, R.W. *Digital Signal Processing*. Prentice-Hall International, London (1975)

APPENDIX–Sample Inputs

(1) Create a DA file with ordered coordinates, filter the coordinates (ν_{max} = 70) and recreate a history file of filtered coordinates.

```
create_info info_file_name = alamd
ordmd DA_property_name = coord *
  number_DA_steps_per_record = 100 maximum_DA_steps = 33000 *
  DA_property_title = ''coordinates for all atoms'' *
  history_file_name = ''ala.his;1'',



        ''ala.his;17''

filter DA_input_property_name = coor *
  DA_output_property_name = flt35 *
  number_DA_steps = 100 maximum_DA_steps = 33000 *
  DA_property_title = ''filtered coordinates, f = 35, df = 70'' *
  first_md = 1 *
  number_md_steps = 131072 *
  filter_centre_frequency = 35.0 filter_range = 70.0
da2gen DA_property_name = flt35 *
  first_md_step = 33300 *
  last_md_step = 39000 *
  old_history_file_name = ''ala.his'' *
  history_file_name = ''ala_f35.his''
endda
```

(2) Create a DA file with trajectories of torsion angles and plot some of them.

```
set_info info_file_name = alamd
do_torsion_trajectory *
  DA_property_name = torsions *
  number_DA_steps_per_record = 100 maximum_DA_steps = 33000 *
  DA_property_title = ''All torsions'' *
  torsion_list_file_name = ''torsions.lst'' *
  history_file_name = ''ala.his;1'',



        ''ala.his;17''

plot_property *
  list_property = torsions list_sub = 12, 18 *
  y_axis_size = 7.0 x_axis_size = 10.0 *
  y_ title = ''Phi     Psi'' plot_title = ''Alanine'' *
  first_md = 4 last_md = 131072 *
  skip_plot_steps = 20
endda
```

(3) Calculate $g(\nu)$ and thermodynamic properties.

```
set_info info_file_name = acetamd_md
gnu DA_property_name = coor_all *
  gnu_file = ''acetamd.gnu'' *
  history_file_name = ''acetamd.his'' *
  first_md = 60000 *
  number_md_steps = 16384
endda
```

(4) Extract characteristic modes of motion and frequencies.

```
set_info info_file_name = acetamd_md
extract_normal_modes DA_property_name = coor_all *
  first_md = 60000 *
  number_md_steps = 16384 *
  normal_mode_out_file = ''acetamd.frq'' *
  normal_mode_def_file = ''acetamd.fdef'' *
  history_file_name = ''acetamd.his''
endda
```