STRFACT

Determining X-Ray Total Scattering Factors from MD Simulation Data

Carlos E. S. Bernardes

STRFACT

(Version 1.3.1)

**Determining X-Ray Total Scattering Factors from
Molecular Dynamics Simulation Data**

Carlos Bernardes
Faculty of Sciences, University of Lisbon
Lisbon 9/1/2024

** INDEX **

1. DESCRIPTION

STRFACT is a software package written in Fortran that analyzes trajectory data computed with the Molecular Dynamics programs DL_POLY, CHARMM, LAMMPS, and GROMACS. It is currently able to calculate total scattering factors for systems containing atoms of H to Cf, and radial distribution functions (RDFs). Additionally, it allows the conversion of RDF data into partial scattering factors.

If you find this software useful for your research, please cite the following references:

- C.E.S. Bernardes; *Zenodo* **2021**, http://doi.org/10.5281/zenodo.10473250.

- C.E.S. Bernardes; K. Shimizu; J.N.C. Lopes; *Comparative Structural Analyses in Four Ionic Liquid Systems: the Two Low-q Peaks of IL Structure Factor Functions*; *Mol. Simulat.* **2018**, *44*, 478-484.

Examples of works where this suite of programs was used include:

- Shimizu, K.; Bernardes, C. E.; Triolo, A.; Canongia Lopes, J. N.; *Nano-Segregation in Ionic Liquids: Scorpions and Vanishing Chains*; *Phys. Chem. Chem. Phys.* **2013**, *15*, 16256-16262.

- Shimizu, K.; Bernardes, C. E. S.; Canongia Lopes, J. N.; *Structure and Aggregation in the 1-Alkyl-3-Methylimidazolium Bis(trifluoromethylsulfonyl)imide Ionic Liquid Homologous Series*; *J. Phys. Chem. B* **2014**, *118*, 567-576.

- Bernardes, C. E. S.; Shimizu, K.; Ferreira, A. I. M. C. L.; Santos, L. M. N. B. F.; Canongia Lopes, J. N.; *Structure and Aggregation in the 1,3-dialkyl-imidazolium Bis(trifluoromethylsulfonyl)Imide Ionic Liquid Family. 2. From Single to Double Long Alkyl Side Chains*; *J. Phys. Chem. B* **2014**, *118*, 6885-6895.

The version currently available is working, however, it is not free from bugs. Any problems or questions do not hesitate to send an email to cebernardes@ciencias.ulisboa.pt. Additional information about the program and release notes can be found at the end of this manual or on the webpage: http://webpages.fc.ul.pt/~cebernardes/index.html.

STRFACT is free software, distributed under the terms of the GNU General Public License as published by the Free Software Foundation and included in the source code documentation. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY, without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. In no event, the author will be liable to you for damages, including any general, special, incidental, or consequential damages (including but not limited to arising out of the use or inability to use the program, to loss of data or data being rendered inaccurate, or losses sustained by you or third parties, or a failure of the program to operate with any other programs), even if the author has been advised of the possibility of such damages.

2. INSTALLING AND RUNNING

After download, the program should be extracted by typing:

```
$ tar -xvf STRFACT_1_3.tar
```

The program is installed from the "src" directory, using the following steps:

(1) Change to directory "src":

```
$ cd  STRFACT_1_3_1/src
```

(2) Pre-compile the "libgmxfort" and "libxdrfile" libraries using:

```
$ sudo .pre-install.sh
```

The earlier script installs the libraries in the "/usr/local" folder, a step that normally requires superuser privileges.

(3) Finally, to install the program, by typing:

```
$ make
```

This installs the program using Gfortran, and places the program executable in the folder "STRFACT_1_3_1/EXE". This directory should be added to the system PATH or the files copied to a folder already included in the PATH (e.g., /usr/local/*bin* or *~/bin*). You can also edit the makefile to install the program directly into the desired place, by changing the BINROOT line in the file "makefile".

Besides STRFACT, three additional programs will be compiled: RDF, RDFTOTAL, and RDF2STRCT. The first two, calculate radial distribution functions, based on atomic positions (RDF) or atom names (RDFTOTAL). RDF2STRCT converts the output files of the earlier programs into partial scattering factors.

The program STRFACT runs by typing:

```
$ strfact [arguments]
```

The following arguments are available:

-dcd   **[file]** Read configurations from a DCD file.

-hydro   Ignore RFDs with hydrogen atoms for a faster calculation. Consider using this option if no peaks are expected below $q < 0.6$ Å^{-1}.

-h   Help information.

-lmps   **[file]** Read trajectory from a Lammps output.

-np   **[integer]** Number of CPUs.

-psf   **[file]** Read molecular information from a CHARMM PSF file.

-q_max   **[val]** Maximum value of q (Å^{-1}) to be considered.

-top   **[file]** Read the topology file.

--version   Current version information.

-xtc   **[file]** Read the XTC trajectory file from GROMACS

-xtcdir   **[dir]** Reads all GROMACS XTC trajectory files in a folder.

After the start, the program checks the data files and retrieves the necessary information. If no trajectory and topology files are stated, STRFACT will look, by default, for the FIELD and HISTORY files of DL_POLY. After this process, the program will perform all tasks automatically. Always check if the information in the topology file was correctly retrieved. This will not be verified by the program and, if wrong, will lead to abnormal results. Thus, confirm the information printed when the program starts, as shown in the following example:

```
No of molecular species 2
Data retrieved for molecule in position 1
No of molecules N1123OH 350
No of atoms in N1123OH 20
Atoms in molecule N1123OH
1:  C1   2: H1  3:  H1   4: C1
5:  H1   6: H1  7:  H1   8: C1
…
```

3. INPUT FILES

To run, STRFACT (and the other programs), two INPUT files are needed: *i)* one containing the information about the simulated system and *ii)* a trajectory file. For the first case, by default, the FIELD file for DLPOLY is searched. In the case of CHARMM and GROMACS, in principle, this information can be retrieved from the PSF and TOP files, respectively. In the case of LAMMPS, or if a change in the atom's names is required, a topology file with the following structure can be used:

```
types          2
name         NTF
nummols      400
atoms          5
C   12.011
F   18.998
F   18.998
F   18.998
S   32.066
finish
name         mimo
nummols      400
atoms          4
N   14.007
C   12.011
N   14.007
C   12.011
finish
```

If the simulations with LAMMPS were prepared with the program DLPGEN 3.0, a file named "topology.inp" will be automatically generated during the input file creation, which can be directly used with the programs described in this manual.

In the example above, "types" refers to the number of types of molecules in the simulation box; "name" is the molecule name; "nummols" is the number of molecules in the system; "atoms" is the number of atoms in the molecule; and "finish" signals that all entries were given for the molecule.

Immediately after the "atoms" entry, the name of each atom followed by the corresponding atomic mass should be provided.

Note that: i) the order of appearance of the molecules and atoms should match that used in the trajectory file (caution, the program will not check this); ii) the initial 2 characters of the atom's name will be used to recognize the element. <u>The program is case sensitive</u>, i.e., an atom that starts with "CL" is recognized as carbon, and one starting with "Cl" is recognized as a chlorine atom. In the current version, elements from H to Cf are supported by STRFACT.

4. AUXILIARY PROGRAMS

Besides the main STRFACT program, this package comes with 3 additional programs:

1) The **RDF** program runs by typing:

```
$ rdf [arguments]
```

Computes radial distribution functions considering atomic positions or centers of mass.

2) The **RDFTOTAL** program runs as:

```
$ rdftotal [arguments]
```

Calculates radial distribution functions based on atom names. It is possible to select an element (e.g. "C" or "H"), or a specific atom name (e.g. "CT" or "HC"). CAUTION: the program also considers atoms in the same molecules.

3) The **RDF2STRC** program runs as:

```
$ rdf2strc [arguments]
```

This program uses files named "RDFOUT.txt", created using programs **RDF** and **RDFTOTAL** or files formatted with two columns (x,y), to compute partial scattering factors.

Most of the arguments in section 2 are valid with the previous programs. To check which ones are available in each case, use the argument "-h". The calculation conditions are set by the user through a questionnaire that appears in the terminal when the program runs.

5. OUTPUT FILES

A successful run of STRFACT will create two folders, "RDF" and "FAC_EST", and a file "STRCOUT.txt" with the total scattering factors calculation result. The "RDF" folder contains all individual radial distribution functions computed during the calculation. The content in this directory will also be used in case the "**–restart**" argument is called. In turn, "FAC_EST" contains all individual partial scattering factors.

The **RDF** and **RDFTOTAL** programs generate an "RDFOUT.txt" file, containing radial distribution calculation results. Regarding the second program, **intramolecular interactions may be considered**. Also, note that both programs will overwrite any "RDFOUT.txt" file in the working directory. Thus, backup previous results before running the programs again.

Finally, the **RDF2STRC** program converts the radial distribution function in two columns [r, G(r)] files, to partial scattering files. r should be in Angstrom and the first line of the file can contain a title. By default, the program looks for "RDFOUT.txt" archives, but any file can be used with the program by typing:

```
$ rdf2strc -f [FILE]
```

If no other option is given, the program will simply apply the Fourier Transform to the data. If a topology file is set as:

```
$ rdf2strc -top [FILE]
```

the information contained in the file can be used to normalize the data (see *Mol. Simulat.* **2018**, *44*, 478-484). For these calculations, the user will be enquired about the length of the squared side of the simulation box used in the MD run, the atom types considered in the RDF calculation (e.g. C, H, O, N, …), and the corresponding number of atoms, N, involved in the RDF calculation. By default, the program will multiply the number of all atoms in the topology file, n, for a given element by the corresponding number of molecules. For example, if the element is carbon, then:

$$N = \sum_i n_{carbons}(i) \times Mols(i)$$

where *i* denotes the different types of molecules in the simulation box and Mols the number of molecules of each *i*. Still, this value can be changed to match, for example, the number of atoms of a given type (e.g. CT, CA, NA, …). The program output will be written in the file "CALCOUT.txt".

All programs use Angstrom (Å) for distances and the *q* is given in $Å^{-1}$. The trajectory files of the different MD programs are converted, if necessary, by the programs without the need for user intervention.

6. ADDITIONAL NOTES

The argument –xtcdir, may be used to read all GROMACS trajectory files contained in a folder. This <u>folder should only contain trajectory files</u>. Otherwise, the program may fail to run. Each file can contain a different number of trajectory frames.

The argument –hydro allows ignoring the hydrogen atoms in the calculations with **STRFACT**. This allows to speed up the calculation but may lead to changes in peak intensity for $q < 1$ Å^{-1}. Before using this option, it is advised to check the impact of this approximation on your results.

7. RELEASE NOTES

** VERSION 1.3.1 **
(11 January 2024)

News
- Solved a problem related to the compilation of the *libxdrf* library, due to a support change in the more recent OSs (e.g., Debian-based operating systems).

** VERSION 1.2.0 **
(09 June 2021)

New Features
- Support for atomic species from hydrogen to californium.
- Improved compatibility with GROMACS input files.
- Improved compatibility with LAMMPS trajectory files.
- Improved IO.
- Support for macOS Big Sur.

Bug Fixes
- Memory leak problem corrected.
- Solved problems related to the use of triclinic simulation boxes with GROMACS.

** VERSION 1.1.0 **
(19 September 2017)

New Features
- Ability to read trajectory files from GROMACS and LAMMPS.
- New procedures to convert radial distribution functions into partial scattering factors.

- Introduction of a user manual.
- New calculation status report.
- Restart option.
- Several other tweaks and optimizations.
-

Bug Fixes

- Minor corrections.

** VERSION 1.0.1 **

(30 September 2015)

New Features

- Ability to read CHARMM output files (*.dcd and *.psf);

Bug Fixes

- Minor format corrections.

8. ACKNOWLEDGEMENTS