Tutorial on the use of the program `denfert` version 2.1 for low-resolution structure recovery of biological molecules from SAXS/SANS data.

July 9, 2019

1 General

The program `denfert` is implementing a simulated annealing algorithm similar to `DAMMIN` program by D. Svergun (Biophys. J. 76, 2879-2886) for the restoration of the low-resolution structure of bio-molecules from small-angle x-ray (SAXS) or small-angle neutron (SANS) scattering data. The major advantage of `denfert` is that the hydration layer around bio-molecules is taken into account by introducing a second type of beads (hydration beads) in the model. For a complete description of the used methodology please refer to the two articles Koutsioubas & Pérez (2013) Journal of Applied Crystallography 46, 1884 and Koutsioubas et al. (2016) Journal of Applied Crystallography 49, 690.

The general protocol for low-resolution structure reconstructions with `defert` includes a) the indirect Fourier transform (IFT) of the experimental data using the program `GNOM` b) passage of `GNOM` output to `denfert` together with the parameters of the reconstruction and c) optional post-reconstruction analysis of multiple runs.

2 Program installation and execution

The program is provided in the form of a single binary executable for the three major platforms (`Windows 64-bit, macOS` and `Linux`). By opening a terminal window, navigate to the directory where the binary is located and type (for example in `Linux`)

```
./denfert_linux
```

In this way the program runs in dialog mode where input parameters are entered one-by-one by the user. Alternatively a text file containing each parameter in successive lines can be passed as argument to the program. The

parameter order should be the same as in the dialog mode (see examples in following sections).

An optional additional step, is to install the programs `pymol` and `gnuplot` on the system, that aid in the visualisation of the final model and of the quality of the fit respectively.

Note that IFT calculations and the generation of input files for `denfert` require installation of the program `GNOM` (version 4.x).

3 Description of program's input and output

In the following we give a brief explanation of **input parameters** in dialog mode and also describe the output generated by the program.

* After launching the program the first requested input concerns the **mode** of the run. There are three options ([F] fast, [S] slow or [A] advanced). In fast mode, a large bead radius is selected in order to speed up calculations. In slow mode a smaller bead radius in implemented. In advanced mode the user will be asked about all the parameters of the simulated annealing reconstruction (see below) including a value for the bead radius. We advise new users to use fast and slow mode in their initial reconstructions, since their annealing parameters should work for the vast majority of reconstructions.

* Then the **type** of input data has to be specified. [X] for SAXS data or [N] for SANS data. Note that if fast or slow has been selected above, the program assumes that in the case of SANS data, the measurements concern a molecule dissolved in a heavy water buffer. If the buffer used consists of a mixture of D2O/H2O or pure H2O then the proper solvent scattering length density can be set in advanced mode.

* A **project name** is requested. All output files will begin with the provided **project name**. Previous output files with the same **project name** will be overwritten.

* Only in the case where a SANS dataset is provided as input the program asks for three parameters related to the instrumental resolution of a single wavelength SANS diffractometer. In the case of SAXS data these input parameters are bypassed, since resolution effects are negligible in modern x-ray sources.

  - **Neutron wavelength** (in Å). The default value (6 Å) represents a usual wavelength used in single wavelength SANS diffractometers. Note that if no value is given (by pressing return), then the default value is used.
  - **Neutron wavelength spread** ($\Delta\lambda/\lambda$). The default value (0.1) represents a usual wavelength spread characterising single wavelength SANS diffractometers.

- **Beam divergence** (in radians) due to the collimation of the neutron beam. The default value (0.002) represents a usual collimation setting in SANS diffractometers.

* Provide the **GNOM output filename** containing the IFT of the experimental data. The full name together with the file extension needs to be provided. The GNOM file should be present in the current working directory. Note that the angular units of the GNOM file should be in Å^{-1} .

* Provide the **maximum wavevector transfer value** $Qmax$ that should be considered during the reconstruction. The default value is equal to $0.25\,\text{Å}^{-1}$. It is advised to use use this default value in the general case, and only for particles with a small radius of gyration, a higher $Qmax$ can be set.

* **Maximum diameter** of the particle. The default value is the one estimated by the IFT performed by GNOM.

* **Particle's sld or electron density** for SANS or SAXS data respectively. Default values concern protein molecules.

* **Solvent/buffer sld or electron density** for SANS or SAXS data respectively. Default values concern heavy water (SANS) or water (SAXS) based buffers. This parameter can be set only in advanced mode.

* **Hydration layer contrast**. By default it is assumed that the hydration layer has a contrast of 10% relative to the sld or electron density of the solvent. This parameter can be set only in advanced mode.

* **Subtracted constant** in order to force Porod behaviour at higher Q, accounting for the internal density fluctuations of biomolecules. A default value is proposed based on internal calculations.

* **Number of knots** is essentially the number of interpolation points of the regularised GNOM curve that are considered during the fit. It is advised to use the default value. This parameter can be set only in advanced mode.

* **Dummy atom radius** in Å. A default value is used based on maximum diameter of the particle and the mode of run. This parameter can be set only in advanced mode.

* **Initial Annealing Temperature** with a default value 10^{-4}. This parameter can be set only in advanced mode and should be changed only in cases where in the initial steps of the annealing procedure the trial acceptance ratio is very low.

* **Annealing Schedule** with a default value 0.95. This parameter can be set only in advanced mode and dictates the rate of annealing temperature drop between subsequent annealing steps. If the annealing procedure appears to be trapped in local minima, then the value can be increased up to values close but not equal to 1.

* **Penalty weight** is the parameter that enforces the compactness of the final model. If the final model appears to be too compact or too 'loose' then penalty weight can be changed in advanced mode to a different value than the proposed one.

* **Trials per bead at each temperature** with a default value equal to 100. This parameter can be set only in advanced mode and should be changed only in cases where the annealing procedure appears to be trapped in local minima.

4 A step-by-step example reconstruction from SANS data.

* We assume that a SANS measurement file (from a protein dimer in D2O buffer) containing scattering data in three columns (Q (in Å units) / Intensity / error bar) is available. We may then use the program the program `GNOM` in order to perform the IFT calculations and obtain an output file that will be used as input for `denfert` calculations. Alternatively we may also use the programs `AUTORG` and `DATGNOM` in order to perform an automatic evaluation of the particles radius of gyration and for performing the IFT calculations, as we will showcase below.

* We first estimate the radius of gyration of the particle using `AUTORG`

```
autorg blac.dat
Rg = 21 +/- 0.12 (1%)
I(0) = 0.288 +/- 0.00039
Points 1 to 61 (61 total)
Quality:  98%
```

* Using the found $R_g = 21$ Å we use `DATGNOM` in order to perform automatically the IFT calculations.

```
datgnom blac.dat --rg 21
Random seed is:  1562573566
Dmax = 67.200000000000003 Total = 0.73363706122713113
Guinier = 21.000000000000000 Gnom = 20.878472729704605
```

The output IFT calculations are contained in the file `blac.out`.

* We then lunch `denfert` in dialog mode

```
./denfert_linux
```

* First we have to pick up the run mode. We pick up advanced mode.

```
[F] fast, [S] slow, [A] advanced mode?.............  A
```

* We then specify that we work with a SANS dataset

```
[X] SAXS data or [N] SANS data?....................  N
```

* We give a representative project name. All output files will begin with the project name.

```
Project name.....................................  betalactoglobulin
```

* We are then asked about the parameters of the SANS measurement that are related to instrument resolution. Note that when treating SAXS data, no such parameters are requested from the program. In this particular example we accept the default values by just pressing return.

```
Neutron wavelength (A) (default=6A)...................
Neutron wavelength spread (default=0.1)...............
Beam divergence (radians) (default=0.002).............
```

* We are asked to provide the file containing the IFT calculations.

```
GNOM output filename (SANS).........................blac.out
```

* We specify the maximum Q to be considered. We accept the default value of $0.25\,\text{Å}^{-1}$ by pressing return

```
Qmax (default=0.25A-1)..............................
```

* The value of the maximum particle diameter as determined by the IFT is displayed by default. We accept this value by pressing return

```
Maximum Diameter - Dmax = 67.2 A ..................
```

* Also for all remaining parameters we accept the default values

```
Particles's sld (default=3.11 10 -6 A -2).............
Solvent/buffer sld (default=6.35 10 -6 A -2)...........
Hydration layer contrast (default=0.635 10 -6 A -2)....
Calculating contribution of internal inhomogeneities...  Please
wait...
Constant to subtract from SANS data = 0.708E-04 ......
Number of knots (default= 20).........................
Dummy atom radius (default=2.412A)....................
Initial Annealing Temperature (default=0.0001)........
Annealing Schedule (default=0.95).....................
Penalty weight (default=0.600E-02).................
Trials per bead at each temperature (default=100) ....
```

* Then after displaying a summary of the input parameters, the annealing procedure starts. At each annealing step the program displays information about the current annealing temperature **Temp**, the score fuction value **Rf** with and without the compactness penalty, the looseness of the current model **loose**, the trial success ratio, the number of particle and hydration layer beads in the model, the radius of gyration and the overall volume of the particle.

```
Temp=0.100E-03 | Rf 2 =0.148E+00| Rf 2 + penalties =0.148E+00
| loose=0.796E-01
success= 405/ 853 | #beads= 852( 1144) | Rg= 19.23A | Volume =
67689A 3
```

* Also at each step of the annealing procedure the model and it's hydration layer are exported in two pdb files `_model.pdb and _hydration.pdb`. Additionaly the current fit is exported to an `ASCII` file (`_fit.dat`) with the column format (Q, experimental intensity, theoretical intensity).

* After the end of the annealing procedure (that may take from a few minutes up to several hours depending on the annealing parameters and the size of the system) the program reports the results

```
Final model parameters
Number of beads :  658
Number of Hydration beads :  562
Molecular envelope Volume without hydration layer (A 3) :  52276
Radius of Gyration model + hydration layer (A) : 21.284
Looseness Penalty :  0.140E-01
Rf :  0.124E-02

Chi against raw data:  0.196E+01
```

while the final model is exported in pdb format and the fit against the raw data in written in the file `_raw.fit` with the column format (Q, experimental intensity, experimental error, theoretical intensity). Usually in a successful run the goodness of fit Rf should be less than 10^{-2}, and loosness (loose) should be less than 0.02.

* A gnuplot (`blac_gnuplot.txt`) and a pymol (`pmscript_blac.py`) related file are written on disk in order to facilitate the visualization of the results. If `gnuplot` and `pymol` are installed on the system, using the commands

```
gnuplot blac_gnuplot.txt -
pymol pmscript_blac.py
```

we obtain the fit and the 3D reconstruction of the particles's structure as seen in figure 1.

* We may also run the program in automatic input mode by preparing an `ASCII` file containing all the parameters and passing it as an argument to the program. So consider a file with the name `blac_run.txt` with the following contents

```
A
N
betalactoglobulin
[empty line]
[empty line]
[empty line]
blac.out
[empty line]
[empty line]
[empty line]
[empty line]
[empty line]
[empty line]
[empty line]
[empty line]
[empty line]
[empty line]
[empty line]
[empty line]
```

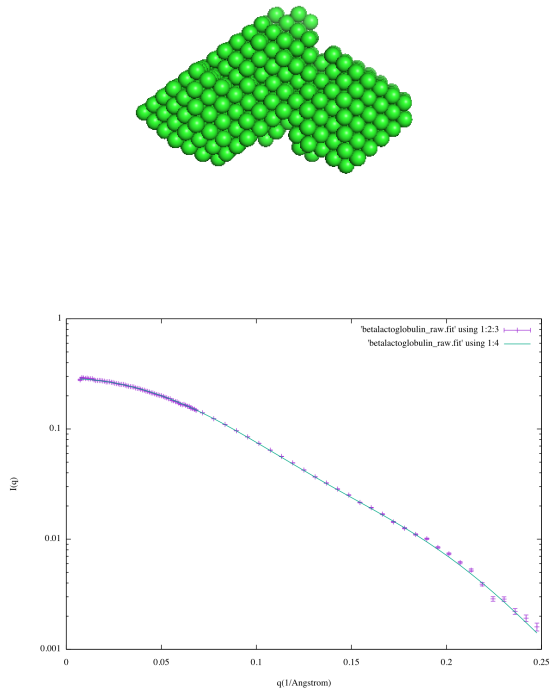

Figure 1: Fig.1 (top) reconstructed model of beta-lactoglobulin dimer (bottom) fit of the raw experimental data.

where the empty lines imply the use of the default values for each parameter [1]. Then we run the program using the command

```
./denfert_linux blac_run.txt
```

* If multiple reconstructions are executed, then the output model PDB files can be post-processed using the program `DAMAVER` in order to align the models, select the most typical one and build an averaged model.

5 Additional info

For questions related to the program, bug reports and feature requests please contact: `a.koutsioumpas[at]fz-juelich.de`.

[1] Notice that the order of parameter input is exactly the same as in dialog mode