7 Calculating contacts: `g_kuh`

7.1 Introduction

`g_kuh` is a custom written Gromacs utility contained within the Gromacs 4.5.3 or 5.0.4 source code extension, used to calculate various native contact related quantities directly from xtc files.

7.2 Example using `g_kuh`

The typical definition of Q used by the SMOG authors defines a native contact formed if the distance between two atoms is less than 1.2 times their distance in the native state. That is, for atomic configuration k,

$$Q_k = \sum_{\{ij\}} \theta(1.2 r_{ij}^{\mathrm{N}} - r_{ij}^{k}) \tag{6}$$

where $\{ij\}$ denotes the set of atom pairs forming native contacts (usually the list of interactions under [pairs] in the `.top` file), r_{ij}^{N} is the distance between the pair ij in the native state and and r_{ij}^{k} is the distance between pair ij in the configuration k. To compute Q with this definition over an entire trajectory the following options should be given to `g_kuh`:

```
user>$ $GROMACS/bin/g_kuh -f trajectory.xtc -s native.gro -n contacts.ndx -o qvals.out -noabscut -noshortcut -cut 0.2

user>$ head contacts.ndx
[contacts ]
 1 5
 1 6
 2 6
 3 7
 4 8
 5 9
 6 10
 7 11
 8 24
 8 12
```

This will give $Q(t)$ in an outfile called `qvals.out`. The program will use native.gro to define r_{ij}^{N}. With the `-noabscut` option used, the parameter to `-cut` is the cutoff ratio minus 1.

7.3 Full documentation

```
user>$ $GROMACS/bin/g_kuh -h

Option     Filename  Type         Description
------------------------------------------------------------
  -s     native.gro  Input        Structure file: gro g96 pdb tpr etc.
  -f       traj.xtc  Input        Trajectory: xtc trr trj gro g96 pdb cpt
  -n      index.ndx  Input, Opt.  Index file
 -nc   contacts.dat  Input, Opt.  Generic data file
  -g     groups.ndx  Input, Opt.  Index file
  -o      qvals.out  Output       Generic output file
  -i      qimap.out  Output, Opt. Generic output file
 -on      qatom.out  Output, Opt. Generic output file
  -r        qaa.out  Output, Opt. Generic output file
 -ri       qiaa.out  Output, Opt. Generic output file
 -rn       qres.out  Output, Opt. Generic output file
  -d    density.out  Output, Opt. Generic output file

Option       Type   Value   Description
------------------------------------------------------
-[no]h       bool   no      Print help info and quit
```

```
-[no]version bool   no      Print version info and quit
-nice        int    19      Set the nicelevel
-b           time   0       First frame (ps) to read from trajectory
-e           time   0       Last frame (ps) to read from trajectory
-dt          time   0       Only use frame when t MOD dt = first time (ps)
-cut         real   0       Contact cutoff
-[no]abscut  bool   yes     use absolute cutoff (instead of relative)
-[no]shortcut  bool yes     use cutoff also at short distances
-qiformat    enum   none    for individual contacts: none, index, pair or list
-[no]times   bool   no      print times into output file
-kappa       real   0       kappa for continuous Q
-[no]groups  bool   no      define groups
-[no]g1std   bool   yes     use built-in groups for group 1
-[no]g2std   bool   yes     use built-in groups for group 2
-[no]filtmap bool   no      filter cmap by groups
-avgdens     enum   atmatm  aggregating choice for density: atmatm, atmres,
                            resatm or resres
-resval      enum   min     aggregation method: min or avg
-weight      enum   none    weight with contacts?: none or abs
-[no]sum2res bool   no      sum up residue densities
-dfunc       enum   step    density function: step or gauss
-drad        real   0.6     radius for density
-nrad        real   0       native radius for density
-[no]exnative  bool no      exclude native neighbors
```

Required inputs are the (native) reference structure (`-s`), a trajectory (`-f`), and the contact map. The contact map can be specified either via an index file(`-n`), which includes only atom pairs, or as a data file(`-nc`), which includes on each line an atom pair and the native contact distance. If an index file (`-n`) is supplied, the native distances are computed from the reference structure. If both `-n` and `-nc` are given, the index file is used. Atom indices start from 1. Contacts are not reordered internally, it is recommended to use an ordered contact file.

A contact is counted as formed if the distance between the atom pair is within a cut-off of their native distance. The specific criterion for contact formation is controlled by the options `-cut`, `-[no]abscut`, `-[no]shortcut`. `-cut` is the (positive) cutoff value. If `-abscut` is given, it is interpreted as the maximum absolute deviation from the native value; if `-noabscut` is selected, `-cut` is taken to be the maximum relative deviation from the native distance for a formed contact. The option `-noshortcut` specifies that only deviations to larger distances than the native value are considered. If `-shortcut` is selected, a contact also becomes unformed if the pair distance becomes too short compared to the native value. (In most situations this is not desirable.)

The contact map can further be restricted to contacts between atoms in specified groups, if option `-filtmap` is selected. Only contacts between one atom in group1 and another atom in group2 will then be considered. Groups 1 and 2 can however be identical. If `-g1std` or `-g2std` are chosen, the built-in choice of groups will be offered, otherwise the available groups will be read from an index file `-g`.

For each processed frame in the trajectory, the total number of formed contacts is written to a new line in the main output file (`-o`). If the option `-times` is activated, the simulation time is written before the contact count.

Optionally, the identity of the formed contacts can be written to the output file `-i`. This output is controlled by the flag `-qiformat=[none/pair/list/index]`. `pair` gives the atom numbers of each formed contact on a new line, terminated by "`0 0`" at the end of each frame. `list` indicates the status of each contact in the system by a value of `0` or `1` for unformed and formed contacts respectively, using one line per frame. When `index` is chosen, the index number of each formed contact is written out according to the ordering used in the contact file, without separation between frames. (This legacy format needs to be processed together with the main output `-o`.)

If file `-on` is specified, the number of formed contacts for each atom is written, on one line per frame.

For the analysis of all-atom structure-based simulations, residue contacts can be determined that are defined as formed when any atom-atom contact between a residue pair is formed. If specified, the number of formed residue contacts is written to file `-r`. The identity of formed residue contacts is written to file `-ri` if `-r` is given and a `-qiformat` other than none is selected. The number of formed contacts per residue can be written to file `-rn`.

The standard parameters `-b` and `-e` are available to restrict the range of output to only part of the trajectory.

For the analysis of umbrella simulations, the continuous Q_{tanh} (analogous to V_Q in Equation 5 in Section 4.1) can be calculated instead of the normal number of formed contacts. To activate this behavior a value for the shape parameter `-kappa` (γ in Equation 4) is supplied. The other parameters that control the criterion for contact formation are then automatically reset to the values that are used in the umbrella code, and the main output file `-o` will contain the value of Q_{tanh} for each processed frame.