NUCS: Mixed scaling protocol

Sonja M. Schwarzl and Stefan Fischer

June 17-2008, version 1.2

1 Combining global and targeted scaling

In many QM/MM calculations, the QM region is surrounded by a flexible MM region, which in turn is surrounded by a region in which protein atoms are kept fixed. In such simulations it is necessary to adequately represent the electrostatic potential at the QM and flexible regions, whereas it is not necessary to correctly represent the electrostatic potential at the fixed region. Thus, the following points should be satisfied:

1. The interaction energy between the QM region and any given protein group calculated with scaled charges should be close to the Poisson-Boltzmann interaction energy. This ensures that the electrostatic potential at the positions of the QM atoms is adequately represented and thus the electrostatic QM/MM interactions are appropriately captured.

2. The interaction energies between two group within the flexible region calculated with scaled charges should be close to the Poisson-Boltzmann interaction energies. This ensures that the electrostatic interaction between any two mobile MM atoms is captured.

3. The interaction energy between a group in the flexible region and *all* other groups calculated with scaled charges should be close to the Poisson-Boltzmann interaction energy. This ensures that the overall electrostatic potential at the position of a mobile MM atom is adequately captured.

This is achieved by deriving scaling factors in a three-step procedure:

1. Determination of global scaling factors for groups in the flexible and QM regions.

2. Determination of targeted scaling factors for groups in the fixed region.

3. Validation of the scaling factors.

2 Getting the global scaling factors

The scripts for the determination of global scaling factors for the groups within the flexible and QM regions are almost identical to the scripts described in NUCS_global_protocol. The only difference is that the user must define a target region in *settings.str*. In QM/MM applications this target region should comprise both the QM and the flexible regions.

Avoid choosing a target region that has a large net charge, since this may lead to very unfavorable interactions between the charged members of the target region, thus to excessive scaling of these groups. If possible, choose a set of residues which have approx. as many positive as negative charges, so that they compensate each other.

This target region is increased in *target.str* (streamed from *initialize.str*) so as to ensure that the boundary between the target and the nontarget regions is along the group definitions used for the calculation of NUCS factors. All charges outside the target region are set to zero. Thus, scaling factors are calculated only for groups within the target region, and only interactions between groups belonging to the target region are considered.

The scaling factors for groups in the target region must be calculated, counterscaled, and verified as described in NUCS_global_protocol. The counterscaled scaling factors are provided in a CHARMM coordinate file in which the WMAIN column is set to the values of the scaling factors for atoms belonging to the target region and to zero for atoms outside the target region.

3 Getting the targeted scaling factors

3.1 Basic considerations

Scaling factors for groups outside the target region are determined according to

$$\lambda_I = \frac{|E^{PB}_{I,target}|}{|E^{vac}_{I,target}|}, \tag{1}$$

where $E^{PB}_{I,target}$ is the Poisson-Boltzmann interaction energy between the charges of the group I and all charges within the target region, and $E^{vac}_{I,target}$ is the corresponding Coulomb interaction energy calculated using unscaled charges on the atoms of group I and scaled charges on the atoms within the target region.

This definition of scaling factors is comparable to the definition of the group:group screening factors ϵ_{IJ} in the global charge scaling routines. However, here, the whole screening not captured by the global screening factors on the atoms of the target region shall be assigned to the group I.

WARNING: The scale-factors obtained from the "targeted" scripts are the inverse of the scale-factors from the "global" scripts. The targeted scaling factors

(Eq. 1) MULTIPLY the charges, whereas the global scaling factors DIVIDE the charges !!!

Eq. 1 uses absolute values for the interaction energies in order to compensate for the change in sign which frequently occurs if the interaction energies in solution and/or in vacuum are small. It also occurs that $|E^{vac}_{I,target}|$ is smaller than $|E^{PB}_{I,target}|$, leading to scaling factors larger than 1. Thus, charges would be increased instead of decreased, which is undesired. Since this usually occurs only when the interaction energies are small, the scaling factors are set to 1.0 in such cases.

3.2 Usage

The CHARMM scripts to determine targeted scaling factors in the non-target region are located in *./scripts/targeted/scaling/*

The input file to be piped into CHARMM is *interaction.inp*.

The user only modifies *generate.str* and *settings1.str*. *generate.str* should be identical to the file used in the determination of global scaling factors. The settings provided in *settings2.str* should be identical to the settings used in the determination of global scaling factors. The differences are the following:

- The name of the CHARMM coordinate file with scaling factors for atoms in the target region must be given in $INPUTFAC$.
- The name of the output CHARMM coordinate file which will contain the scaling factors for atoms outside the target region must be given in $OUTPUT$.
- A filename must be provided in $OUTPUT2$, into which the electrostatic interaction energies between the groups outside the target region and the target region calculated with Poisson-Boltzmann and with a Coulomb potential, the distance between the group and the target region, and the scaling factor are given

Finally, *combine_all_factors.inp* allows to concatenate the global scaling factors determined for the target region and the targeted scaling factors determined for the non-target region.

3.3 Details of Implementation

interaction.inp streams *generate.str*, *settings1/2.str*, and *initialize.str*. In *initialize.str*, *defi19.str* or *defi22.str* and *target.str* are streamed. After an energy call to initialize the energy function, the electrostatic potential due to the atoms of the target region is determined by numerical solution of the Poisson-Boltzmann equation in *pbeq.str*. The vector $SCA9$ contains the resulting potential at the

positions of all atoms. *loop_inter.str* then loops over all residues and streams *calculation.str*. In *calculation.str*, *vacuum.str* (returns $E^{vac}_{I,target}$) and *solution.str* (returns $E^{PB}_{I,target}$) are streamed for all groups that are outside the target region. Then, the scaling factor is determined and stored to $SCA6$. Moreover, the distance between the group and the target region is computed. The group identification, $E^{vac}_{I,target}$, $E^{PB}_{I,target}$, the distance, and the scaling factor are written to $OUTPUT2$.

4 Optional: validating the mixed scaling factors

CHARMM input scripts for the validation of the scaling factors according to the three aspects given in Section 1 have been prepared. In all sets of scripts only *settings.str* and *generate.str* must be modified by the user according to the settings used in the determination of the scaling factors. Additional settings are described below.

In all cases, interaction energies are computed with Poisson-Boltzmann electrostatics to provide reference values. The corresponding interaction energies are then determined using a Coulomb potential with either scaled or unscaled charges, allowing for a comparison.

1. **Potential in QM region.** The input file to be piped into CHARMM is *interactions_with_qm.inp*. This streams *generate.str*, *settings.str*, and *initialize.str*. In *settings.str* the user must specify the QM region, the file name of the CHARMM coordinate file with scaling factors for all groups in WMAIN, and the output file name. The electrostatic potential due to the atoms of the QM region is determined in *pbeq.str* and returned in $SCA9$. *loop_screen.str* loops over *all* groups and streams *calculation.str*. *calculation.str* streams *solution.str* that return the Poisson-Boltzmann interaction energy, calculates the Coulomb interaction energy with scaled and unscaled charges, determined the distance between the group and the QM region, and write the output.

2. **Potential in flexible region.** The input file to be piped into CHARMM is *interactions.inp*. This streams *generate.str*, *settings.str*, and *initialize.str*. In *settings.str* the user must specify the target and QM regions, the file name of the CHARMM coordinate file with scaling factors for all groups in WMAIN, and the output file name. *loop.str* loops over all groups and streams *calcener.str* for groups that are in the target region but not in the QM region. The electrostatic potential due to the atoms of the group is determined in *pbeq.str* and returned in $SCA9$. Then, *calculation.str* is streamed that streams *solution.str* that return the Poisson-Boltzmann interaction energy between the group and all other groups, calculates the

Coulomb interaction energy with scaled and unscaled charges and writes the output.

3. **Group:group interactions in flexible region.** The input file to be piped into CHARMM is *interactions.inp*. This streams *generate.str*, *settings.str*, and *initialize.str*. In *settings.str* the user must specify the target and QM regions, the file name of the CHARMM coordinate file with scaling factors for all groups in WMAIN, and the output file name. *loop.str* loops over all groups and streams *calcener.str* for groups that are in the target region but not in the QM region. The electrostatic potential due to the atoms of the group is determined in *pbeq.str* and returned in $SCA9$. Then, *loop2.str* loops over all groups and streams *calculation.str* for groups that are in the target region but not in the QM region. *calculation.str* streams *solution.str* that returns the Poisson-Boltzmann interaction energy between the two groups, calculates the Coulomb interaction energy with scaled and unscaled charges, determines the distance between the two groups and writes the output.