Brownmove-Setup-Documentation

Tihamér Geyer

Zentrum für Bioinformatik, Universität des Saarlandes, D–66041 Saarbrücken, Germany
tihamer.geyer@bioinformatik.uni-saarland.de

V1.3 — Oct. 30, 2009

Covering Brownmove V1.0 of Oct 2009

The Structure of a Simulation

Each simulation is performed inside a so-called "cup" (remember the "cup of hot tea" used to drive the "Heart of Gold"?), a container for the walls that encloses the actual simulation volume and a number of "proteins", which are the moving objects.

A "protein" is a hierarchical structure defining a bead-spring model of a polymer or a biological protein. The individual beads move and interact with other beads independently. A protein thus contains at least one rigid gestalt object, resembling the "beads". Each gestalt again contains a number of shapes, which model the vdW surface of this rigid unit, the effective charges, the hooks to which the springs are attached, and a shape for the hydrodynamic interactions. This structure is shown in the following figure. The GeomShape object is special as it is not involved in collecting forces from the interactions, but converts the total force and torque into the respective displacements. Consequently, this shape must always be present, whereas the other four shapes only have to be present if the corresponding interaction are used in the model.
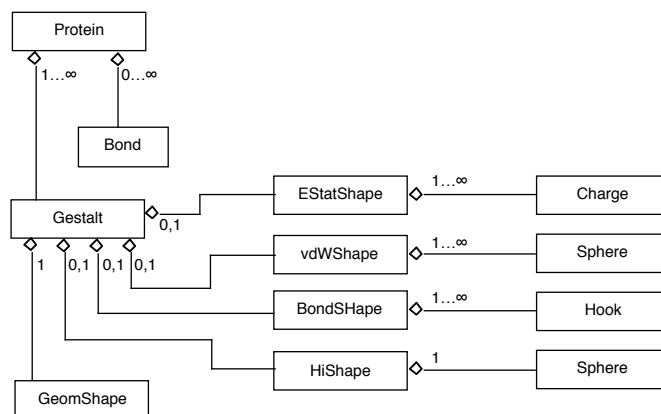

Interactions

The "natural" units of brownmove are one nanometer for distances, one picosecond for time intervals, one kDa for masses, and one kJ/Mol = 9.9278×10^{-4} kDa nm^2 ps^{-1} for energies. Charges are given in elementary charges. Thus, the vacuum dielectric constant $\varepsilon_0 = 5.728 \times 10^{-4}$ e^2 (kJ/mol)$^{-1}$ nm^{-1}. Forces are given in 1 kJ/ mol nm^{-1} = 1.67 pN. Densities are in particles per nm^3, which means that 1M = 0.6022 nm^{-3}.

Some more constants in the brownmove units: Boltzmann constant: $k_B = 8.3145 \times 10^{-3}$ kJ/mol K^{-1}; viscosity of water at 20°C (293 K): $\eta_0 = 0.532$ kDa nm^{-1} ps^{-1}. Consequently, translational diffusion ceofficients are given in nm^2 ps^{-1}.

Electrostatic Interactions

Electrostatic interactions are modelled as shielded point charges a la Debye-Hückel. The interaction energy between two point charges q_i and q_k (on different gestalt objects) is

$$E_{ik} = \frac{1}{4\pi\epsilon\epsilon_0} \, q_i q_k \, \frac{\exp[-\kappa(r_{ik} - B_{ik})]}{(1 + \kappa B_{ik})r_{ik}}$$
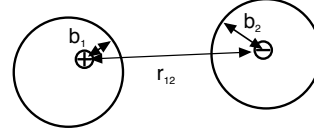

Here, ε is relative dielectric constant of water and ε0 the vacuum dielectric constant. The shielding from the ions is captured in the inverse Debye length $\kappa = 1 / l_D$. Mostly, the charges are embedded inside the protein, where no counter ions shield the interactions. $B_{ik} = b_i + b_k$ is the sum of the the effective burial depths of the two charges, b_i and b_k, respectively.

If no value is specified, brownmove uses a typical physiological ion strength of 0.9 M as a default, i.e., l_D = 1.01 nm.

van-der-Waals Interactions

In brownmove, the van-der-Waals interactions denote the effective short ranged repulsions and attractions between the protein surfaces. The vdw spheres can also be used to model hydrophobic interactions. For this, an r^{-12}, an r^{-6}, and an r^{-3} term are used for these phenomenological interaction between two of the spheres with a distance $r = r_{12} - (r_1 + r_2)$ between their closest points:

$$E = C_{12}\left(\frac{r_0}{r + dr}\right)^{12} + C_6\left(\frac{r_0}{r + dr}\right)^{6} + C_3\left(\frac{r_0}{r + dr}\right)^{3}$$
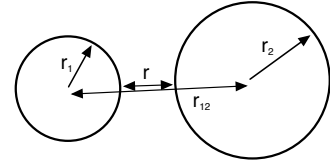
Usually, C_{12} would be positive and C_6 negative. As usual, r_0 determines the width of the potential (well). The potential is defined such that when the two vdW spheres touch, i.e., at $r = 0$, the repulsive potential has a positive value of $V_0 = 1\,k_BT$. For the usual repulsive Lennard-Jones-6-12 interaction, i.e., with, $C_{12} > 0$ and $C_3 = 0$, this leads to

$$dr = r_0 \left[\frac{2C_{12}}{\sqrt{C_6^2 + 4C_{12}V_0} - C_6}\right]^{1/6}$$

The same formula (with a different V_0) can be used to determine a minimal distance R_{min} (which is actually negative) from which on the interaction is linearized for numerical stability. If the potential should be linearized for, e.g., V ≥ 5 kT, then $R_{min} = dr(V_0 = 5kT) - dr(V_0 = 1kT)$, The last parameter is a maximal cutoff radius R_{max}. For distances $r > R_{max}$ the evaluation of the interparticle force is skipped and a zero returned. To avoid numerical artefacts, R_{max} should be larger than the extent of the potential.

To allow for different interaction parameters between different pairs of particles, each vdW sphere has an index, named the "vdW color". At setup, for any occuring pair of colors a set of interaction parameters {r_0, dr, C_{12}, C_6, C_3, R_{min}, R_{max}} has to be specified. These parameters are globally visible in the simulation.

For vdW interactions between particles and walls, the same form of the interaction potential is used with the same global vdW color indices.

Hydrogen Bonds

A hydrogen bond is modelled by an inverted Gaussian potential between the donor hydrogen and the acceptor atom. The angle between the donor-hydrogen direction and the hydrogen-acceptor direction is considered via a cosine-term.

$$E = E_0 \exp\left[-\frac{(r - r_0)^2}{2d^2}\right] \cos(\phi)$$
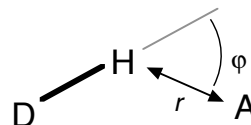
Bonds

Bonds are harmonic interaction potentials between the "hook-up points" in the bondShape objects with

$$E_{ik} = \frac{k_{ik}^2}{2}\left(r_{ik} - L_{ik}\right)^2$$

For any pair of hooks, a spring constant k_{ik} and a rest length L_{ik} have to specified within the scope of the protein. A bond can not be broken.

Hydrodynamic Interactions

The hydrodynamic interactions are modelled with the truncated expansion algorithm as described in [T. Geyer and U. Winter, "An $O(N^2)$ approximation for hydrodynamic interactions in Brownian dynamics simulations", *J. Chem. Phys.*, **130** (2009) 114905]. For the translational interactions, the modified Rotne-Prager-Yamakawa tensor of Garcia de la Torre and Bloomfield [*Biopolymers* **16** (1977) 1747] is used which allows for particles of different sizes. Note that this form of the RPY tensor diverges with increasing particle overlap.

By default, only translational HI are used. To enable (yet the not so thoroughly tested) rotational HI, set the flag -DHI_WITH_ROTATION in the file config.mk during compilation.

External Forces

External position dependent forces can be attached via hook-up points in an ExternalShape object. Currently, a constant force, a harmonic potential, and a shear force are implemented. The actual force is the product of the (position dependent) external force $F(r)$ and a "weight" α, such that the same external potential can excert different forces on different hook-up points. For a gravitational force, α would be the mass, while an α proportional to the particle radius leads to the same drift velocities of different sized particles.

The shear force is explained below. Note that the force f and the normal n do not have to be normalized or even orthogonal to each other (though this would be the usual case).

$$\vec{F}(r) = \alpha \left(\vec{n} \cdot (\vec{r} - \vec{R}_0)\right) \vec{f}$$
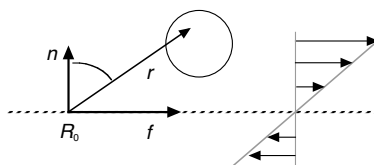
Propagation algorithms

Brownmove used a Langevin Dynamics (LD) propagation algorithm formulated in terms of forces as described in [U. Winter and T. Geyer, "Coarse Grained Simulations of a Small Peptide: Effects of Finite Damping and Hydrodynamic Interactions", *J. Chem. Phys.* **131**, (2009) 104102]. This formulation allows to continuously blend from the LD to the standard Ermak-McCammon BD propagation by setting the particle mass to zero.

Setup Configuration File

The simulation package "brownmove" is more of a many-particle BD and LD library, but can also be used as a black-box simulation tool. For this, many interesting simulation scenarios can be defined through configuration files as described below. At least two files are required: one global configuration and one file that defines the proteins. Often, there is a third file describing the simulation box (environment).

General Stuff

This central configuration setup file contains global information about the simulation like length, timestep, output interval, general interaction parameters, or output scripts as well as definitions of the simulation box and the proteins.

Configuration lines start with a case-sensitive keyword followed by the required data. Any text beyond the last required data is ignored and can be used for comments (even without a leading '#'). Keywords and data are separated by whitespace, i.e., any number of spaces or tabs. Leading whitespace is ignored, indentation can be used freely to increase readability. Empty lines or lines starting with a '#' are considered comments and are skipped.

The first section usually defines the timestep and the simulation duration in timesteps. The output interval is defined in units of timesteps. There is no specific order of these three lines, but they all have to be present.

```
# parameters for the simulation
timestep    100.0               # [ps]
runsteps    5000000             # length of sim in timesteps
outsteps    100                 # output every .. timesteps
```

Some of the global parameters can also be changed from their defaults. Currently, these are the solvent temperature (defaulting to 293 K = 20°C), the inverse Debye length $\kappa = l_D^{-1}$, which defaults to 0.99 nm^{-1}, i.e., a physiological ion concentration of 0.09 M, and the relative dielectric constant of water, $\varepsilon_{H2O} = 78$. For simulations with explicit ions, where no implicit shielding of the Coulomb interactions occurs, set $\kappa = 0$.

```
# solvent temperature in Kelvin
# temperature      293
# shielding parameter for Coulomb interactions: kappa = 1/lD
# kappa            0.99
# relative dielectric constant of water
# eps_H2O          78.0
```

Then, there are some optional keywords to change the simulation's behavior. The HI cutoff is currently an untested experimental setting and should not be used.

```
# The following keyword prevents the final dump of the proteins after the sim
```

```
noProteinDump
```

When you want to continue a simulation from the final dump, you can use the keyword "offsetsteps" to start it from a given timepoint. When offsetsteps is set, the initial particle positions are not put out so that the output files can be concatenated directly.

```
# start simulation at a later time, e.g., to continue a too short run
# offsetsteps     20000

# if this parameter is present,
# the random number generator is seeded with this fixed value
# randomseed      1998

# define a global cut-off for hydrodynamic interactions
# hicutoff  20.0          # [nm]
```

The output of the simulation is piped into an analyzer script, where a certain pre-processing or sorting of the sometimes vast amounts of data can be performed. This script is defined with the "analysis" keyword. It can be as simple as "cat", which dumps the output to the console, or a pipe of scripts and programs doing a sophisticated on-the-fly analysis. Here are some examples. Note that anything after the keyword and the first following whitespace is passed on to the shell within single quotes. Consequently, in this line there should not be any comments at the end of the command.

```
# analyzer-script
# analysis    cat
# analysis    cat > fish.txt
analysis      grep Punkt | cut -f4-6 > fish.txt
# analysis    grep timeline > anzahl_x-20.txt
```

Global Interaction Definitions

The first set of interaction parameters concern the van-der-Waals interaction. Each vdWShape of a protein is labelled with a vdW-color (starting with 0) according to its type. Here the respective parameters are defined for all potential pairs of vdW interactions. Undefined color pairs are initialized with zero, i.e., no interaction. In the following example, a single vdw color will be used.

```
# number of different vdW colors to be used
vdwarraysize       1
# vdwparms  clr1 clr2   C_12      C_6  C_3  R_0  DR        Rmin         Rmax
vdwparms    0    0      9.744594  0.0  0.0  1.0  1.122562  -0.14088552  10.0
```

The hydrogen bonds interaction parameters may also be changed from their default values with the following definitions:

```
# hbond_depth    25.0    # depth of the potential well in kJ/mol
# hbond_length   0.2     # H-acceptor distance for minimal interaction
# hbond_width    0.05    # width of the inverted-Gaussian potential
# hbond_cutoff   1.5     # cut-off for hbond calculation
```

External (position dependent) forces are defined with the externalForce keyword, a label, the type, and a number of contect specific figures. Currently, the type may be "constant" for a position independent vectorial force, "harmonic", which is defined by the position of its minimum and the spring constants in x, y, and z direction, or "shear" with its position and the two vectors as explained above. The simulation does not verify that the two vectors are orthogonal, i.e., skewed force fields may be defined.

These definitions of the external forces must preceed the protein definitions, because currently the external interactions on the particles are set up during protein assembly.

```
# definition of external forces
# single vector for constant force
externalForce   towardsX  constant   10.0 0.0 0.0
# position of min and spring constants for harmonic
externalForce   Well      harmonic   0.0 0.0 0.0   10.0 100.0 1000.0
# origin and two directions for shear
externalForce   Drift     shear      0.0 0.0 0.0  1.0 0.0 0.0   0.0 100.0 0.0
```

Protein Definitions

The next part of the configuration file defines the proteins used in the simulation. Each protein type is defined on a line of its own starting with the keyword "protein", followed by a label for further reference, the filename containing the definition (see below for their format) and the respective place in this file — a protein definition file may contain multiple protein definitions. These definitions must preceed any defintions of the simulation box.

```
# protein definitions: label  file  number (starting with 1)
protein     singleBead  chargedBead.browndef  1
```

The labels assigned to the proteins are used within the general setup process only. They are used to place proteins directly into the box. Reservoirs, which are then attached to the walls, are defined as follows with the keyword, a unique label, the (initial) density, and an inverse volume.

```
# Reservoirs of constant density
# reservoir label      initial density    inv. Vol
reservoir   PunktRein    4e-4                0.0
reservoir   PunktRaus    0.0                 1.0
```

The inverse volume corresponds to the change in the reservoir density when one particle is added or taken out. In the above example PunktRein is an inifinite reservoir with a constant density, whereas PunktRaus can be used to directly count the particles in it. A reservoir with a volume of 10 x 10 x 10 nm^3 consequently has an inverse volume of `0.001`.

Simulation Box

If one or a few freely diffusing polymers in an unbounded volume are simulated, this part can be omitted. Otherwise the simulation box will be bounded by walls, which can be simple reflecting walls, reservoirs, or walls with a gestalt. A special class of reflecting walls are periodic walls, which displace the particles by one box length upon contact. Such a quasi periodic setup can be defined directly with the following line:

```
# (quasi-)periodic scenery
# with the boxsizes in X, Y, and Z directions
periodicBox      0.0  20.0  20.0    # each in [nm]
```

If any of the dimensions is ≤ 0, no periodicity in that direction is set up. The above 2D box thus has walls at y = ±10 nm and z = ±10 nm, but none bounding the x-direction.

With periodicBox an additional flag is set that tells the shapes to consider an extra copy of the interaction partner one box length away on either side. A periodicBox should thus be larger than a reasonable cutoff for the interactions. Any other scenery is defined in an extra file and given to the parser as

```
# the scenery: walls, etc - only one scene file allowed
box          box.browndef
```

The format of this file is explained further down. Note that periodicBox may be followed by a box definition to, e.g., complete a 2D periodic setup.

Initial Protein Positions

The last part of the global setup file consists of optional lines that directly place proteins into the simulation box. In the following example the simulation is seeded with four of the above defined proteins with the label singleBead. The seven numbers define the initial position: three positions (x, y, z), a rotation angle (in degrees) and a unit vector defining the rotation axis.

```
# start sim with these proteins in place
# (use labels from above)
startWith   singleBead        -4.0 -4.0 -4.0    0.0 1.0 0.0 0.0
startWith   singleBead         4.0 -4.0 -4.0    0.0 1.0 0.0 0.0
startWith   singleBead        -4.0  4.0 -4.0    0.0 1.0 0.0 0.0
startWith   singleBead         4.0  4.0 -4.0    0.0 1.0 0.0 0.0
```

Output of Reservoir Densities and Protein Numbers

Reservoir densities and protein numbers can be printed into an additional output file, which has to be defined via the output keyword. Then, reservoir densities and particle numbers can be given:

```
# output of reservoir densities
outputfile   untenraus.txt
density      UntenRaus
number       Punkt
```

The reservoir names are the labels used above and the protein names are the ones that finally appear in the output file, i.e., the ones in the protein definition file. Note that first all densities are given in the output and then all particle numbers.

Protein Definition File

A protein definition file resembles the hierarchical structure of the proteins. Each section starts with a certain keyword and ends with a corresponding "End" tag. Indentation is optional, but may increase readability.

The outermost level starts with the Protein keyword and a label for the protein. This label is then used in the output file to denote the type of protein. This line is directly followed by the definition of the origin of the protein's coordinate system and an effective CM diffusion coefficient needed for the injectors of the reservoirs.

```
Protein Polymer
   position 0.0 0.0 0.0  0.0 1.0 0.0 0.0
   # eff. translational CM diff. coeff., required for molecule injector
   D0 1e-4  # [nm^2/ps]
```

Bonds

If the protein contains harmonic bonds between the individual Gestalt objects, these are defined next. Each bond is defined with a label for further reference within the scope of its protein, the length, and the spring constant in units of kT/nm^2.

```
Bonds Federn     # the Bonds-label is a actually a dummy argument :-)
   # Label length spring-const
   B12a   0.38    12000
   B12b   0.555   7000
   B13    0.8     180
   :
End Bonds    # saying "Bonds" is optional
```

Gestalt Objects

Now, the gestalten, which represent the physical objects, can be defined. Analogous to the protein definition, they start with the opening tag plus a label, their position in the protein's ccordinate system, and the parameters for diffusion. These consist of the translational and rotational diffusion coefficients (in units of nm^2/ps and ps^{-1}, resp.), the mass m in kDa, and an effective radius a used to calculate the rotational moment of inertia $L = 2/5\ m\ a^2$. When the mass is set to zero, a concentional overdamped Brownian propagator is used, while a Langevin propagator is used for $m > 0$. The required damping parameters are determined from the mass and the (globally set) viscosity.

```
Gestalt SER1
   position 1.266 1.187 -0.675  0.0 1.0 0.0 0.0

   # D0 D0_trans  D0_rot   mass      radius
   D0   9.42E-4   0.01361  0.10509   0.23
```

Inside the gestalt, the various shapes can be defined. All of them are optional. When the above Gestalt section is closed with an End tag, a diffusing particle without any interactions is defined. Every shape gets a label which is used for diagnostic information during the setup and later when the proteins are dumped.

The BondShape references the bonds defined above and defines where they are attached relative to the gestalt's center.

```
BondShape SER1-Hooks
   # position              Bond label
   -0.0323 0.0023 0.0030   B12a
   0.0000 0.0000 0.0000    B12b
   :
End
```

The VdwShape defines one or more vdW spheres, i.e., their position relative to the gestalt's origin, their radius, and their vdW color. Again, any text after End is ignored.

```
VdwShape SER1-vdw
   # pos              r      color
   0.00 0.10 0.00    0.371   0
End
```

The charges are defined analoguously. Now, they are characterized by their charge (in unit charges) and the shielding radius.

```
EstatShape SER1-E
    # position            charge radius
    -0.016 -0.132 0.069    0.51  0.2
    -0.109 -0.213 0.072   -0.51  0.2
    0.107  0.159 -0.124   -0.66  0.177
End
```

Hydrogen bonds can be defined between a donor hydrogen atom and an acceptor atom (note that this is different from the usual definition of an H shared between two N or C atoms). Also the angle between the donor-hydrogen bond and the hydrogen-acceptor vector is considered. A donor is defined via the positions of the hydrogen and of the donor atom (or any point on the line defined by these two). From these two a normalized vector pointing from the donor to the hydrogen is constructed. For an acceptor, only the position of the acceptor atom is required. For a weak donor (acceptor) the multiplier value should be smaller than one, while a value larger than one indicates a strong donor (acceptor).

```
HBondShape HB
    # keyword    donor position    hydrogen position  scaling factor
    donor        1.8 0.0 0.0       2.0 0.0 0.0        1.0
    # keyword    acceptor position  scaling factor
    acceptor     -1.4 0.7 0.5       1.1
End
```

The HI shape may only define a single HI sphere, which is fixed around the gestalt's center. Therefore, the only parameter is the radius.

```
HiShape SER1-HI
    0.17
End
```

After the closing End tag for the gestalt, more gestalten within the same protein may follow. A closing End then finalizes the protein.

Scene Definition File

The scene definition file specifies the walls around the simulation box. Its structure is similar to the protein definition files (actually, the same parser is used for both). A wall definition starts with the Wall keyword plus a label, followed by the wall's position. A wall is initially created in the y-z-plane at $x = 0$, its normal thus points into the positive x-direction. The final position and orientation of the wall has to be given relative to this initial setting. Note that walls are mathematical planes, i.e., they are infinitely large.

```
# bottom of the box in the y-z-plane at x = -20
Wall bottom
    position  -20.0 0.0 0.0  0.0 1.0 0.0 0.0
```

If this wall shall be a basic reflecting wall, only the closing End tag has to be added. Here we assume that the wall interacts with the diffusing particles via a vdW potential. We then need a gestalt with its position relative to the wall and a VdwShape. For a wall, the normal of the vdW surface is the same as for the wall itself, thus only the relative displacement along this normal and the color have to be specified. However, a wall may also contain vdW spheres, which are defined as in a protein.

```
Gestalt bottom
    position 0.0 0.0 0.0  0.0 1.0 0.0 0.0
```

```
        VdwShape vdwUnten
            # type:wall   offset   color
            wall          -0.3      0
            # also possible: a vdW sphere attached to the wall
            # position    radius   color
            0.0 0.0 0.0   4.0       0
        End
    End
```

This gestalt may also contain charges in the same way as a protein (not shown here).

A vdW shape specific for a wall is a so called "block" with a centered cylindrical hole along the x-axis. It is defined via its position including orientation, the extension into the three directions, the radius of the cylindrical hole, and its vdW color. The following code thus defines a quadratic slab of area 60 x 60 nm2 with a thickness of 4 nm and four holes at (y, z) = (±15, ±15) nm. This slab is centered at the origin and not rotated. When the side walls of the simulation box are placed at y = ± 30 nm and z = ± 30 nm, this slab is a barrer with four holes. If the side walls are further away, the proteins may even circumvent the slab.

```
Wall Mitte
    position 0.0 0.0 0.0   0.0 1.0 0.0 0.0

    Gestalt Block
        position 0.0 0.0 0.0   0.0 1.0 0.0 0.0

        VdwShape vdwBlock
            # block   x y z            lx ly lz       cyl_radius   color
            block    0.0 -15.0 -15.0   2.0 15.0 15.0    10.0       0
            block    0.0 -15.0  15.0   2.0 15.0 15.0    10.0       0
            block    0.0  15.0 -15.0   2.0 15.0 15.0    10.0       0
            block    0.0  15.0  15.0   2.0 15.0 15.0    10.0       0
        End
    End
End
```

To allow for particle exchange and interfacing to reservoirs, injectors and acceptors are used. They are defined with the respective keywords and a label. They both require a definition of a rectangular mask with the mask's center and how far it extends from there. These coordinates are defined relative the wall's internal coordinate system. Then, one (or more) molecules are assigned to injector and acceptor via the labels from the protein configuration file. A top wall in a cubic simulation box centered around the origin and interfacing to a constant density region at $x > 20$ nm would look like the following:

```
# top: y-z-plane at x = +20, normal pointing towards -x
Wall Oben
    # move to (20, 0, 0) and rotate by 180 degrees around y-axis
    position  20.0 0.0 0.0   180.0 0.0 1.0 0.0

    Injector Rein
        # Mask  Center     extension (+-ly +-lz)
        Mask    0.0  0.0    20.0   20.0     # mask area is 40x40 nm2 = 1600 nm2
        # Molekule  Name    Reservoir
        Molecule    Punkt   PunktRein
    End

    Acceptor Raus
        # Mask  Center  extension (+-ly +-lz)
```

```
        Mask  0.0  0.0   20.0  20.0
        # Molekule  Name  Reservoir
        Molecule     Punkt PunktRaus
    End
End
```

Output File Format

The raw output of the simulation consists of a time line with the actual simulation time and the number of gestalten in the simulation. This line is followed by one line per gestalt, which lists the label from the protein definition file, the numerical ID of the protein (assigned at insertion into the cup), and the gestalt label from the protein definition. These are followed by the gestalt's position and orientation (rotation angle plus vector for the rotation axis). A typical output at a given timestep with two proteins, a NA and a CL ion, would like this:

```
timeline    4.000    2
CL    1    CL_Bead    0.899 0.175 -0.177    0.000 1.000 0.000 0.000
NA    2    NA_Bead    -1.029 0.130 0.137    0.000 1.000 0.000 0.000
```