Swiss Institute of Bioinformatics

SwissSidechain - Documentation

David Gfeller
Swiss Institute of Bioinformatics
1015 Lausanne

13 august 2012

david.gfeller@isb-sib.ch

Table of contents

1. Introduction

Amino acid sidechains form the building blocks of all proteins. Natural proteins are in general made of 20 naturally encoded amino acids (to which one could add the two rare amino acids pyrrolysine and selenocysteine). Some of these amino acids can be modified by post-translational modifications, so that the actual number of sidechains observed in proteins is closer to 50-60. However, this corresponds only to a very small fraction of the number of possible amino acids that can be chemically synthesized and incorporated into polypeptide chains.

Non-natural sidechains refer to amino acid sidechains that are not found (or rarely found) in natural proteins. Non-natural sidechains are powerful to add new chemistry to existing proteins or peptides and they have been used in a number of applications ranging from crystallographic and biochemical studies to protein engineering and drug design.

2. Content of the database

The SwissSidechain database contains structural and molecular data for 210 non-natural sidechains, both in L- and D-configurations, in addition to the 20 natural ones. These data are stored in files describing the chemical structure (SMILES) and 3D coordinates (PDB, MOL2) of these sidechains, files describing the physico-chemical properties of these sidechains (partial charges, LogP, bond/angle/torsion constants), as well as files describing the different possible conformations of these sidechains (rotamers). In addition, we provide tools to rapidly integrate non-natural sidechains into existing analysis (CHARMM and GROMACS) and visualization (PyMOL and UCSF Chimera) software.

2.1 File formats

2.1.1 PDB

This is the standard format for protein structure files.

```
ATOM      1  N   ABA A   1      -2.227  -0.872  -0.586  1.00  0.00           N
ATOM      2  CA  ABA A   1      -1.162   0.123  -0.638  1.00  0.00           C
ATOM      3  C   ABA A   1      -0.776   0.027  -2.022  1.00  0.00           C
ATOM      4  O   ABA A   1      -0.917  -1.009  -2.647  1.00  0.00           O
ATOM      5  CB  ABA A   1      -0.024  -0.173   0.382  1.00  0.00           C
ATOM      6  CG  ABA A   1      -0.494  -0.095   1.853  1.00  0.00           C
ATOM      7  OXT ABA A   1      -0.517   1.039  -2.648  1.00  0.00           O
ATOM      8  HA  ABA A   1      -1.577   1.111  -0.443  1.00  0.00           H
ATOM      9  HB1 ABA A   1       0.773   0.557   0.234  1.00  0.00           H
ATOM     10  HB2 ABA A   1       0.372  -1.170   0.191  1.00  0.00           H
ATOM     11  HG1 ABA A   1       0.344  -0.311   2.515  1.00  0.00           H
ATOM     12  HG2 ABA A   1      -0.873   0.905   2.060  1.00  0.00           H
ATOM     13  HG3 ABA A   1      -1.286  -0.826   2.020  1.00  0.00           H
ATOM     14  H1  ABA A   1      -2.602  -0.916   0.350  1.00  0.00           H
ATOM     15  H2  ABA A   1      -2.960  -0.613  -1.231  1.00  0.00           H
ATOM     16  H3  ABA A   1      -1.856  -1.775  -0.843  1.00  0.00           H
```

Briefly speaking, for lines starting with ATOM, column 2 gives the atom number. Column 3 gives the atom name. Column 4 gives the residue name. Column 5 gives the chain. Column 6 gives the residue number. Columns 7-9 give the x,y,z coordinates. Column 10 shows the occupancy (set to 1.00 in SwissSidechain entries). Column 11 shows the temperature factor (also called B-factor, set to 1.00 in SwissSidechain entries). The last column shows the element symbol (C,H,N,O, …).

Note that PDB format is space delimited so that columns in some cases may touch each other.

You can find much more detailed information about PDB file format at http://www.wwpdb.org/docs.html.

2.1.2 MOL2

MOL2 files are often used in chemoinformatics applications. They contain information on the 3D position and partial charges of each atom of a molecule, as well as on chemical bonds.

Lines following the "@<TRIPOS>ATOM" provide atom number, atom names, 3D coordinates, atom kind, chain, residue name and atomic partial charges information.

```
@<TRIPOS>ATOM
      1 N          -1.6280    1.1068   -0.6297 N.4       1 ABA1       -0.8386
      2 CA         -1.0161    0.0142    0.1179 C.3       1 ABA1       -0.0725
      3 C           0.0867    0.7016    0.7382 C.2       1 ABA1        0.9356
      4 O           0.5921    1.6817    0.2207 O.co2     1 ABA1       -0.8033
      5 CB         -0.6136   -1.1783   -0.7981 C.3       1 ABA1       -0.0974
```

Lines following the "@<TRIPOS>BOND" provide a list of chemical bonds. Column 1 stands for a numbering of the bond. Column 2 and 3 gives the number of the bound atoms. Column 4 indicates the nature of the bond (single, double, aromatic,…).

```
@<TRIPOS>BOND
     1    3    2 1
     2    3    4 2
     3    3    7 2
     4    2    5 1
     5    2    8 1
```

2.1.3 SMI

Simplified Molecular-Input Line-Entry System (SMILES) corresponds to a string describing unambiguously chemical molecules. This format is widely used in chemoinformatics. Example:

```
CC[C@@H]([C](=O)=O)[NH3]          ABA.pdb
```

SMILES can be translated into 2D and 3D structure files with tools such OpenBabel.

Detailed information about how to read SMILES are available at: http://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system

2.1.4 TOP

This is the CHARMM topology format for amino acids.

The line starting with RESI gives the amino acid name and total charge

```
RESI ABA          0.00  !  2-Aminobutyric acid
```

Lines starting with ATOM list the atoms of the amino acid, the atom kind as defined in the CHARMM force field and the partial charge.

```
GROUP
ATOM    N       NH1     -0.4700
ATOM    HN      H       0.3100
ATOM    CA      CT1     0.0700
ATOM    HA      HB      0.0900
GROUP
ATOM    CB      CT2     -0.1800
ATOM    HB1     HA      0.0900
ATOM    HB2     HA      0.0900
...
```

Lines starting with BOND or DOUBLE list the chemical bonds between atoms of the amino acid.

```
BOND CB CA  N  HN  N  CA C  CA  C  +N  CA HA
DOUBLE O  C
BOND    CG      CB
...
```

Lines starting with IMPR give the improper angles.

```
IMPR N -C CA HN  C CA +N O
```

Lines starting with CMAP indicate how to connect the backbone atoms to the one in the previous and next amino acid.

```
CMAP -C  N  CA  C   N  CA  C  +N
```

Lines starting with IC list the internal coordinates used to reconstruct the molecule (useful if only part of the sidechain is present in the crystal structure).

```
IC -C   CA   *N   HN    1.3551 126.4900  180.0000 115.4200  0.9996
IC -C   N    CA   C     1.3551 126.4900  180.0000 114.4400  1.5390
IC N    CA   C    +N    1.4592 114.4400  180.0000 116.8400  1.3558
IC +N   CA   *C   O     1.3558 116.8400  180.0000 122.5200  1.2297
```

Exclamation marks (!) indicate comments.

2.1.5 RTP

This is the GROMACS topology format for amino acids.

The first line gives the name of the amino acid.

```
[ ABA ] ; 2-Aminobutyric acid
```

Then all atoms are listed together with the atom type and the partial atomic charge.

```
[ atoms ]
       N       NH1     -0.4700 0
       HN      H       0.3100  1
       CA      CT1     0.0700  2
       HA      HB      0.0900  3
       CB      CT2     -0.1800 4
       HB1     HA      0.0900  5
       HB2     HA      0.0900  6
       ...
```

The chemical bonds are also listed.

```
[ bonds ]
       N       HN
       N       CA
       C       CA
       C       +N
       CA      HA
       CB      CA
       O       C
       ...
```

Finally improper angles and connections between residues for backbone atoms are listed.

2.1.6 HDB

These files are used in GROMACS to add hydrogen atoms to the heavy atoms of the sidechains.

```
ABA     4
1       1       HN      N       -C      CA
1       5       HA      CA      N       C       CB
2       6       HB      CB      CA      CG
3       4       HG      CG      CB      CA
```

The first line indicates the sidechain code and the number of heavy atoms to which hydrogen atoms are bound.
Subsequent lines indicate the kind of hydrogen atoms. Column 1 stands for the number of H atoms. Column 2 stands for the GROMACS type of hydrogen atoms (see GROMACS manual for more information, http://www.gromacs.org/Documentation/Manual). Column 3 stands for the name of hydrogen atoms. In case there are more than one hydrogen atom, a number will be added to the name (e.g., HB -> HB1 and HB2). Column 4 indicates the heavy atoms to which hydrogen atoms are bound. The remaining columns show other heavy atoms bound to the one of column 4 (these are used by GROMACS to infer where to place the hydrogen).

2.1.7 bbdep_Gfeller.lib

These files correspond to the backbone dependent rotamer libraries.

```
ABA     -180    -180    27      1       0       0       0       0.732000
ABA     -180    -180    27      2       0       0       0       0.070400
ABA     -180    -180    27      3       0       0       0       0.197600
ABA     -180    -170    14      1       0       0       0       0.608100
ABA     -180    -170    14      2       0       0       0       0.175900
ABA     -180    -170    14      3       0       0       0       0.216000
-->
60.5    0.0     0.0     0.0     1.7     0.0     0.0     0.0
-161.3  0.0     0.0     0.0     11.0    0.0     0.0     0.0
-71.8   0.0     0.0     0.0     12.4    0.0     0.0     0.0
57.9    0.0     0.0     0.0     1.6     0.0     0.0     0.0
-157.2  0.0     0.0     0.0     10.3    0.0     0.0     0.0
-70.5   0.0     0.0     0.0     12.5    0.0     0.0     0.0
```

Column 1: Residue name
Column 2: ϕ backbone dihedral angle (-5 to +5 intervals).
Column 3: ψ backbone dihedral angle (-5 to +5 intervals).
Column 4: Number of data points with the corresponding backbone dihedral angles and rotameric conformation. For non-natural sidechain in SwissSidechain this number corresponds to the number of snapshots in the MD simulations. However, this number should not be used to infer probabilities (column 9), as it does not include smoothing and renormalization.
Column 5-8: Rotameric state. Each column corresponds to the rotameric state of sidechain dihedral angles. 0's indicate that the corresponding sidechain dihedral angles do not exist. For a detailed description of the sidechain dihedral angles and their rotameric state, refer to the Supplementary material of our paper (Gfeller *et al*, 2012) (available at http://swisssidechain.ch/contacts.html).

Column 9: Final probabilities of the rotamer (i.e. $p(r_1, \ldots r_N \mid \phi, \psi)$).
Column 10-13: Mean values to the sidechain dihedral angles.
Column 14-17: Standard deviation values for the sidechain dihedral angles.

For sidechains with more than 4 freely rotating dihedral angles (6CL, HHK, HRG in the current version of SwissSidechain), additional columns are included and the following columns are shifted accordingly.

2.1.8 bbind_Gfeller.lib

These files correspond to the backbone independent rotamer libraries.

```
ABA     1       0       0       0       50020   50020
ABA     2       0       0       0       71841   71841
ABA     3       0       0       0       78139   78139
-->
 11.07   0.00    100.00  0.00    60.29   0.06
 31.96   0.00    100.00  0.00    -169.72 0.05
 56.97   0.00    100.00  0.00    -65.67  0.05
```

Column 1: Residue name
Column 2-5: Rotameric state. Each column corresponds to the rotameric state of sidechain dihedral angles. 0's indicate that the corresponding sidechain dihedral angles do not exist. For a detailed description of the sidechain dihedral angles and their rotameric state, refer to the Supplementary material of our paper (Gfeller *et al*, 2012) (available at http://swisssidechain.ch/contacts.html).
Column 6: Number of data points with the corresponding first dihedral angle rotameric state.
Column 7: Number of data points with the corresponding rotameric state (i.e. considering all sidechain dihedral angles).
Column 8: Final probabilities of the rotamer (i.e. $p(r_1, \dots r_N)$), including smoothing and renormalization.
Column 9: Standard deviation of the probability in column 8 (set to 0 in SwissSidechain).
Column 8: Final probabilities of the rotamer given the first dihedral angle rotameric state (i.e. $p(r_1, \dots r_N \mid r_1)$).
Column 9: Standard deviation of the probability in column 9 (set to 0 in SwissSidechain).
Column 12-13: Mean and standard deviation values for the first sidechain dihedral angle.
Column 14-15: Mean and standard deviation values for the second sidechain dihedral angle.
Column 16-17: Mean and standard deviation values for the third sidechain dihedral angle.
Column 18-19: Mean and standard deviation values for the fourth sidechain dihedral angle.

Columns 14 and larger are optional, depending on the number of sidechain dihedral angle.

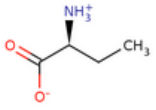
For sidechains with more than 4 freely rotating dihedral angles (6CL, HHK, HRG in the current version of SwissSidechain), additional columns are included.

2.2 Browsing options

2.2.1 Full and family-based tables

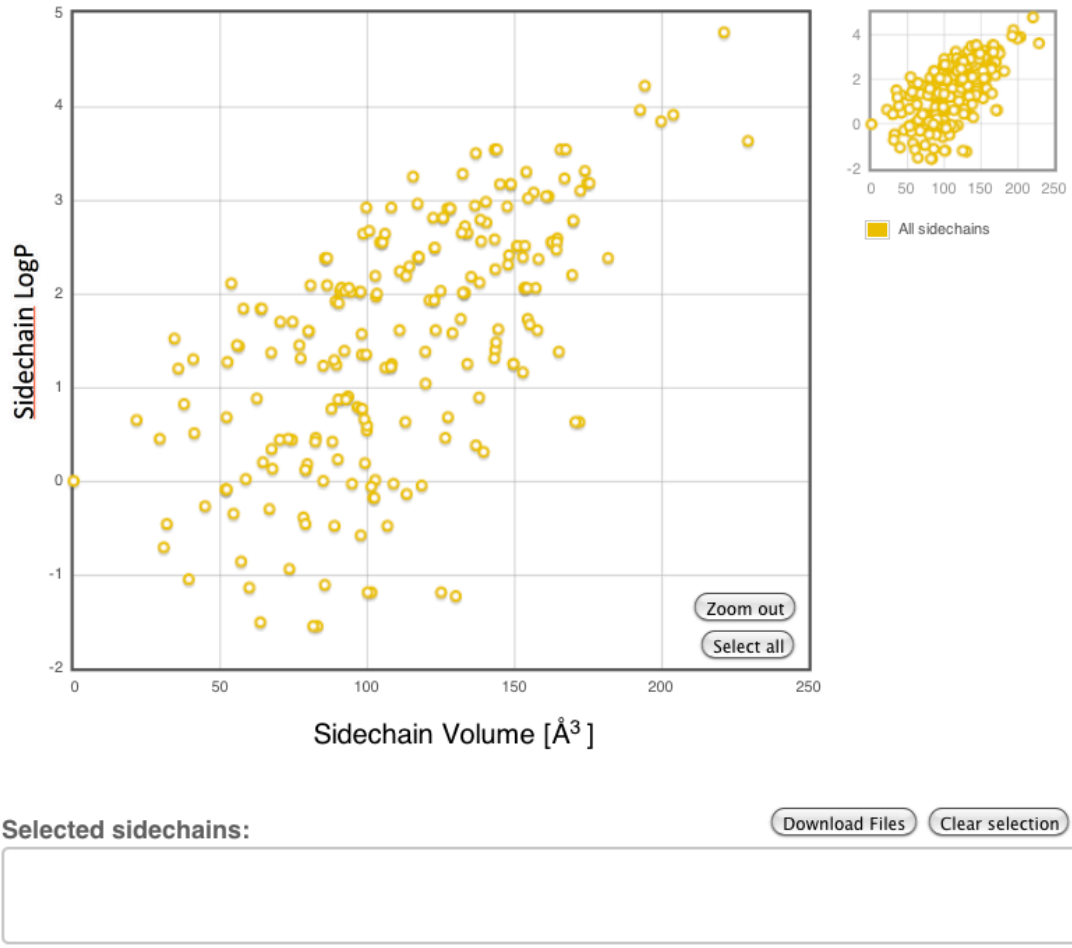
This is the first option to browse data in SwissSidechain. All sidechains are listed in a large table starting with their 3- or 4-letter code in the first column (the code of the D-form is shown in parenthesis). This code is also linked to the sidechain information page. The second column shows the full chemical name. 2D structure is shown in the third column. Colum 4 provides download links for structural files. In addition one can immediately used the sidechain as a ligand in the SwissDock docking server by following the "Dock me" link. Links to the PDB are provided in column 5. Rotamer files can be accessed in column 6. Bbdep stands for backbone-dependent rotamers, while bbind stands for backbone independent (see (Dunbrack, 2002) (Gfeller *et al*, 2012)). Topology files for CHARMM and GROMACS are available in column 7.

Example of a table entry:

ABA (DABA)	2-Aminobutyric acid		ABA.pdb ABA.mol2 ABA.smi Dock me	PDB Ligand	ABA_bbdep ABA_bbind	ABA.top ABA.rtp ABA.hdb

2.2.2 Browsing by physico-chemical properties

Instead of an alphabetical table, you can also browse SwissSidechain data in a 2D plot based on sidechain physico-chemical properties (see below).



The logP indicates the sidechain hydrophobicity. The higher the value, the more hydrophobic your sidechain will be.

3. Visualization plugins

Visualizing non-natural sidechains in protein structures is crucial to explore the structural environment of this new sidechain. This is useful to assess whether introducing a non-natural sidechain at a given place may result in favorable interactions with the surrounding or, in the contrary, will create unfavorable interactions (e.g. repulsive electrostatic interactions) or large steric clashes.

3.1 Pymol plugin

PyMOL can be accessed at http://www.pymol.org/ (license required) or freely downloaded at http://sourceforge.net/projects/pymol/.

Installation:

Automatic installation from a terminal (Mac + Linux):

- Download and unzip the file PySwissSidechain.zip in a directory of your choice.
- Go to the new PySwissSidechain directory and run the installation script (# ./install.sh).

You will be asked to give the path to your PyMOL application.

On Mac, if PyMOL is in the Applications folder, this should be something like: "/Applications/MacPyMOL.app/".

On Linux, give the path of the directory containing the pymol/ folder. This is typically the directory where you have untarred the pymol-vxxx-Linux-xxx.tar.bz2 file.

Manual installation (Mac + Linux + Windows):

- Download and unzip the file PySwissSidechain.zip in a directory of your choice.
- Copy all files from the fragments directory into PYMOL_PATH/pymol/data/chempy/fragments/.
- Create a directory PYMOL_PATH/pymol/data/chempy/SwissSidechains.
- Copy the two files from the SwissSidechains directory in this new directory.
- Copy mutagenesis1.py into PYMOL_PATH/pymol/modules/pymol/wizard/.
- Copy PySwissSidechain.py into PYMOL_PATH/pymol/modules/pmg_tk/startup/.

Note that on Windows, the PYMOL_PATH would typically look like "C:\Program Files\PyMOL" and pymol is sometimes spelled PyMOL depending on the version of PyMOL you are using.

Running:

Plugin Menu (on Mac,version 1.0 to 1.4)

Go under the Plugin->PySwissSidechain:

- Enter the residue you want to mutate as: *Object*//*Chain*/*ResNumber*/ (e.g., complex//E/82/)
- Enter the new residue code (e.g., HLEU)

Command line

Use the command:

```
Mutate Object//Chain/ResNumber/, Newres
```

For instance, to mutate residue number '82' on chain 'E' in object 'complex' into Homoleucine (HLEU), write:

```
Mutate complex//E/82/, HLEU
```

You can then choose between the different rotamers by clicking on the arrows at the lower right corner (same as standard Mutagenesis). Click 'Apply' and 'Done' to validate your mutation.

Troubleshooting:

1) I cannot install the plugin:

- Make sure all files in PySwissSidechain.zip are correctly located or install them manually.

  On Linux: Go to the directory where you installed pymol.

  - Make sure PySwissSidechain.py is found in pymol/modules/pmg_tk/startup.
  - Make sure mutagenesis1.py is found in pymol/modules/pymol/wizard/.
  - Make sure all .pkl files of the fragments/ folder are found in pymol/data/chempy/fragments.

- Make sure sc_bb_dep.pkl and sc_bb_ind.pkl are found in pymol/data/chempy/SwissSidechain/.

On MacOS: Right click on the MacPyMOL icon (most likely you'll find it in the Applications folder) and select 'Show Package content'.

- Make sure PySwissSidechain.py is found in pymol/modules/pmg_tk/startup.
- Make sure mutagenesis1.py is found in pymol/modules/pymol/wizard/.
- Make sure all .pkl files of the fragments/ folder are found in pymol/data/chempy/fragments.
- Make sure sc_bb_dep.pkl and sc_bb_ind.pkl are found in pymol/data/chempy/SwissSidechain/.

On Windows: Go to PyMOL program (most likely you'll find it in C:\Program Files\)

- Make sure PySwissSidechain.py is found in PyMOL\modules\pmg_tk\startup\.
- Make sure mutagenesis1.py is found in PyMOL\modules\pymol\wizard\.
- Make sure all .pkl files of the fragments/ folder are found in PyMOL\data\chempy\fragments\.
- Make sure sc_bb_dep.pkl and sc_bb_ind.pkl are found in PyMOL\data\chempy\SwissSidechain\.

2) I cannot mutate a residue:

- Make sure you correctly selected the residue. If your PyMOL session contains multiple objects, make sure you specify which object you want to mutate: *object//chain/residue_number/.* Pay attention to the "/" as all of them are required.
- If your structure does not include chain name (e.g., if you built a peptide with the Build/Residue/ option), leave the 'chain' empty in selecting a residue to mutate.
- Make sure you use one of the SwissSidechain three- or four-letter codes. Note that D-residues in the D_PySwissSidechain plugin have different codes than their corresponding L-form.
- On MacOS, if you use MacPyMOL.0.99 or MacPyMOL.1.5+ plugins are not automatically loaded. Therefore every time you open MacPyMOL, you should load the plugin with:

  ```
  run PYMOL_PATH/pymol/modules/pmg_tk/startup/PySwissSidechain.py
  ```

  Alternatively, you may also rename you MacPyMOL application to MacPyMOLX11Hybrid. This will automatically switch to the same graphical interface as MacPyMOL.1.4 and plugins will load automatically.
- Make sure you are running the same version of PyMOL as the one where you installed the plugin.
- You should also have Python.2.7+ installed on your machine.

3.2 UCSF Chimera plugin

UCSF Chimera (Pettersen *et al*, 2004) is developed by the University of San Francisco and is freely available at http://www.cgl.ucsf.edu/chimera/.

Installation (Chimera 1.5+):

- Create a new directory that will contain the plugin.
- Download and unzip the file SwissSidechain_chimera.zip in this new directory.
- Open Chimera and go to Favorites/Preferences.
- Select "Tools" Category and add the directory containing the SwissSidechain folder to the "Locations".
- Restart Chimera

The SwissSidechain module now should appear under Tools->Structure Editing menu.

Proceed similarly for D_SwissSidechain_chimera plugin.

Running:

Tools/Structure editing menu:

- Select the residue you want to mutate.
- Go to Tools->Structure Editing->SwissSidechain/.
- Select one of the non-natural sidechains of SwissSidechain and click 'Apply'.
- Select one of the rotamers and click Apply'.

For D-amino acids, use Tools->Structure Editing->D_SwissSidechain/

Command line:

Use the command 'swapnaa':

```
swapnaa Newres #Object:ResNumber.Chain
```

For instance, to mutate residue number 82 on chain E in object 0 into Aminobutyric acid (ABA), write:

```
swapnaa ABA #0:82.E
```

'swapnaa' automatically selects the most favorable rotamer, as predicted by Chimera.

For D-amino acids, use 'swapdaa' with the corresponding D-residue codes. Note that for D-alanine (DAL) the hydrogen atoms automatically appear in the structure. You can hide them by first selecting them (Select… Chemistry… element… H) and then hiding them (Actions… Atoms/Bonds… hide).

4. Molecular dynamics simulations

Molecular Dynamics are used to model the dynamical behavior of macromolecules. There, each atom is considered and the different forces between atoms (both bonded and nonbonded) are described in terms of physical forces.

Some of the most widely used software for molecular include CHARMM (Brooks *et al*, 2009), GROMACS (Van Der Spoel *et al*, 2005) or NAMD.

In SwissSidechain, we provide parameters and topologies compatible with CHARMM and GROMACS (using the CHARMM force field (Bjelkmar *et al*, 2010)).

Topology files describe the different atoms of the non-natural amino acids, their partial atomic charges and their bonds. Parameter files describe the force field parameters governing the interactions between atoms.

Topologies for natural sidechains are not included, as they are already available in CHARMM and GROMACS. As these software do not have data for selenium atoms, we do not provide the topology file for selenocysteine.

4.1. GROMACS

To run GROMACS with the non-natural sidechains of SwissSidechain, you need to download the SwissSidechain_gromacs.zip in the http://www.swisssidechain/MD/gromacs.html page.

Go then to the location of the charm force field. If you have installed gromacs on your machine, the path should look like.

Copy the files existing aminoacids.rtp, aminoacids.hdb, gb.itp, ffbonded.itp, ffnonbonded.itp and atomtypes.atp into a backup directory and add the ones from SwissSidechain in the gromacs/top/charmm27.ff/ directory.

Do the same with residuetypes.dat in gromacs/top/.

Simulation including only natural amino acids will give the same results as with the standard force field. However, if you have manually modified the force field or the topologies, or include other topology files, you should pay attention to possible conflicts in the atoms or residue names.

If you want to run simulation with D-amino acid, you do not need any additional force-field parameters. You can either use the code of the corresponding L-sidechain (easiest if you manually changed the chirality of a sidechain without changing its name). Or you can use the SwissSidechain topologies for D-residues (recommended if your work with protein structures as we use the same code for D-residues as they do in the PDB).

4.2. CHARMM

To run simulations with CHARMM and the SwissSidechain data, first download the top_all22_prot_swisssidechain.inp.zip and par_all22_prot_swisssidechain.inp.zip.

top_all22_prot_swisssidechain.inp contains the topologies for both natural and non-natural sidechains. You need to use this file instead of the standard top_all22_prot.inp in your CHARMM input:

```
open unit 1 card read name @lib/top_all22_prot_swisssidechain.inp
read RTF card unit 1
close unit 1
```

par_all22_prot_swisssidechain.inp contains the new parameters required to run simulation with the non-natural sidechains. You need to add these parameters to the standard parameters in your CHARMM input:

```
open unit 1 card read name @lib/par_all22_prot_swisssidechain.inp
read PARA card unit 1 append
close unit 1
```

Note the append at the end of the second line.

Simulations including only natural amino acids will give the same results as with the standard force field. However, if you have manually modified the force field or the topologies, or included other topology files, you should pay attention to possible conflicts in the atoms or residue names.

If you want to run simulation with D-amino acids, you do not need any additional force-field parameters. You can either use the code of the corresponding L-sidechain (easiest if you manually changed the chirality of a sidechain without changing its name). Or you can use the SwissSidechain topologies for D-residues (recommended if your work with protein structures as we use the same atom names and three-letter residue codes as in the PDB).

5. References

Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E (2010) Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J Chem Theory Comput* **6:** 459.

Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30:** 1545-1614.

Dunbrack RL, Jr. (2002) Rotamer libraries in the 21st century. *Curr Opin Struct Biol* **12:** 431-440.

Gfeller D, Michielin O, Zoete V (2012) Expanding molecular modeling and design tools to non-natural sidechains. *J Comput Chem* **55:** 1525-1535.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25:** 1605-1612.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* **26:** 1701-1718.