

GOMC

User Manual
Version 2.00

Distributed by the Potoff and Schwiebert Groups
©Wayne State University

May 12, 2017

Contents

1	Tutorial Overview	3
2	Introduction	3
3	How to get the software	3
4	Platform and Software Requirements	5
4.1	Supported Operating Systems	5
4.2	Required Software Prerequisites	5
5	Highly Recommended Software Tools	7
5.1	VMD	7
5.2	Packmol	8
6	Other Useful Software Tools	8
6.1	Grace	8
6.2	Cygwin	9
7	Compiling GOMC	9
7.1	Extracting the code	9
7.2	Compiling the code	10
7.2.1	GPU code	10
7.2.2	Serial Code	11
8	Input File Formats	11
8.1	PDB	12

1 Tutorial Overview

This document will instruct a new user how to download, compile, and run the GOMC molecular simulation code. A basic understanding of statistical physics is recommended to complete this tutorial.

To demonstrate the capabilities of the code, the user is guided through the process of downloading and compiling a GOMC executable. That executable is then used to perform saturated vapor and liquid equilibria (VLE) studies on systems of pure isobutane (R600a), a branched alkane that whose application as a refrigerant/propellant is increasing.

<http://en.wikipedia.org/wiki/Isobutane>

The Transferable Potentials for Phase Equilibria (TraPPE) united atom (UA) force field is used to describe the molecular geometry constraints and the intermolecular interactions.

2 Introduction

Monte Carlo (MC) simulation is a type of simulation driven by stochastic processes. "GO" stands for GPU-Optimized; this code was intended to run optimally on modern graphics process hardware.

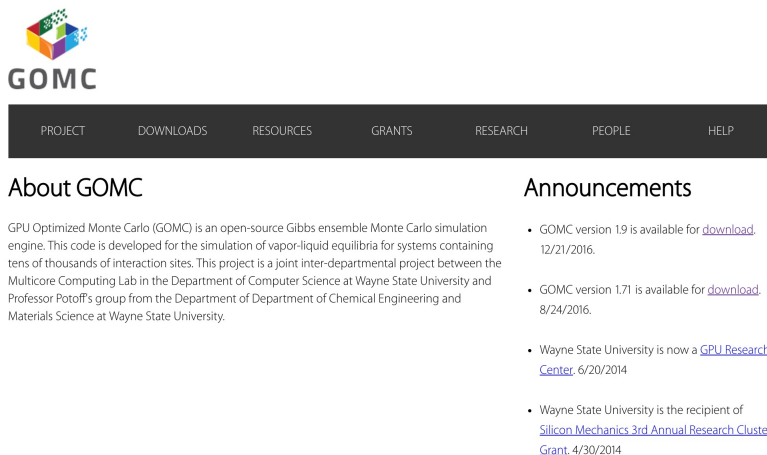
More specifically, this engine includes serial and GPU-Optimized (multi-threaded) codes designed to run Markov chain Boltzmann sampling of chemical systems – effectively sets of points defined by topological maps and interaction algorithms in a simulation box. From statistical mechanics, we know this is one way to sample phase space and model chemical systems.

GOMC currently supports canonical, isobaric-isothermal, Grand canonical, constant volume Gibbs ensemble, and constant pressure Gibbs ensemble simulations. GOMC employs widely-used simulation file types (PDB, CHARMM-style parameter file, PSF). GOMC includes configurational bias algorithms for both linear and branched charged, and none charged systems.

3 How to get the software

The latest public code builds, project logo, manual, and other resources can be obtained via the following website:

<http://gomc.eng.wayne.edu/>



GOMC

PROJECT DOWNLOADS RESOURCES GRANTS RESEARCH PEOPLE HELP

About GOMC

GPU Optimized Monte Carlo (GOMC) is an open-source Gibbs ensemble Monte Carlo simulation engine. This code is developed for the simulation of vapor-liquid equilibria for systems containing tens of thousands of interaction sites. This project is a joint inter-departmental project between the Multicore Computing Lab in the Department of Computer Science at Wayne State University and Professor Potoff's group from the Department of Chemical Engineering and Materials Science at Wayne State University.

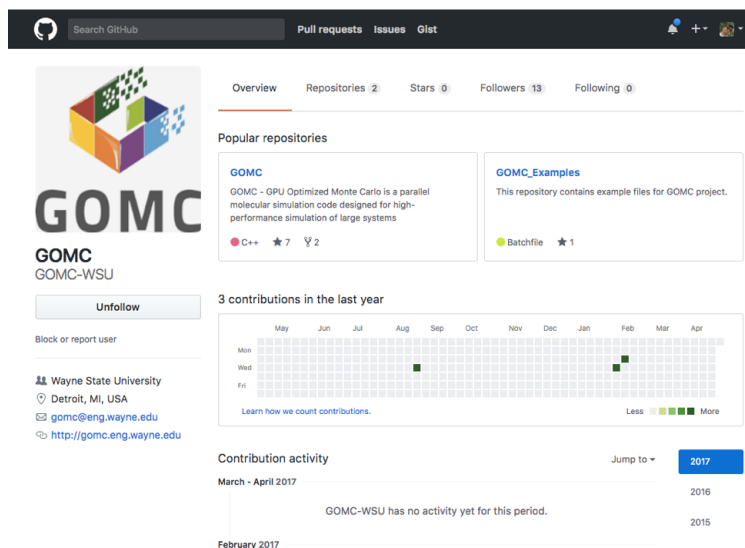
Announcements

- GOMC version 1.9 is available for [download](#). 12/21/2016.
- GOMC version 1.71 is available for [download](#). 8/24/2016.
- Wayne State University is now a [GPU Research Center](#). 6/20/2014.
- Wayne State University is the recipient of [Silicon Mechanics 3rd Annual Research Cluster Grant](#). 4/30/2014.

The code can be found under the download tab, below and to the right of the logo. When new betas (or release builds) are announced, they will replace the prior code under the downloads tab. An announcement will be posted on the front page to notify users.

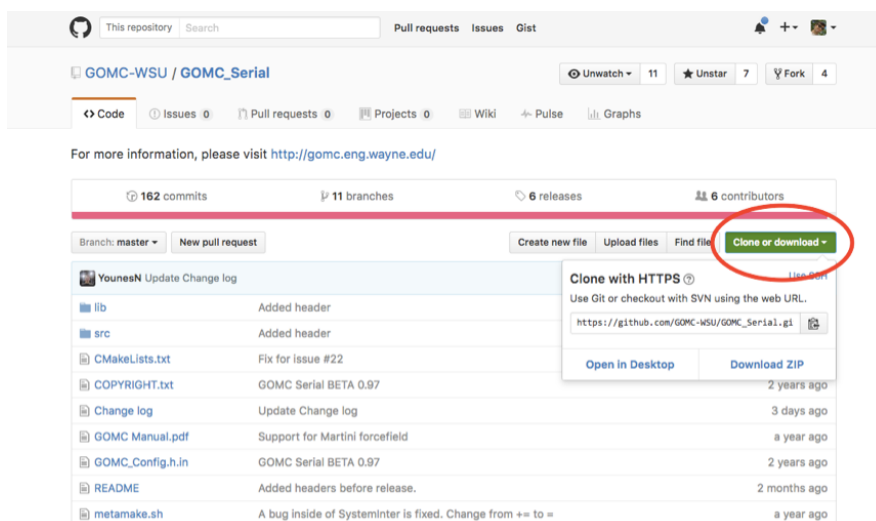
Currently, version control is handled through the GitHub repository. The posted builds in Master branch are “frozen” versions of the code that have been validated for a number of systems and ensembles. Other branches are created as a means of implementing new features. The latest updated code builds, project logo, manual, example files, and other resources can be obtained via the following GitHub repository:

<https://github.com/GOMC-WSU>



The CPU and GPU code are merged together under GOMC repository and can be found under the main page. In addition, Examples repository can be found under the main page. Under each repository, the code and manual can be downloaded by clicking on the Clone or download tab. For more information regarding GitHub, visit the following link:

<https://guides.github.com/activities/hello-world/>



4 Platform and Software Requirements

4.1 Supported Operating Systems

GOMC officially supports Windows 7, 8, and most modern distributions of Linux (see the next section). This software has the ability to compile on recent versions of OS X; however, such a platform is not officially supported.

4.2 Required Software Prerequisites

GOMC has some mild software requirements, which are widely available for Linux operating systems. Required software requirements are:

1 C++03 Compliant Compiler

1.1 Linux/OS X

1.1.1 `icc` (Intel C++ Compiler)

Type the following command in a terminal:

```
$ icc --version
```

If gives a version number 4.4 or later, you're all set. If it's older than 4.4 (released in 2009), we recommend upgrading.

In Linux, the Intel compiler will generally produce the fastest serial executables (when running on Intel Core processors).

1.1.2 `g++` (GNU GCC)

Type the following command in a terminal.

```
$ g++ --version
```

If gives a version number 4.4 or later, you're all set. If it's older than 4.4 (released in 2009), we recommend upgrading.

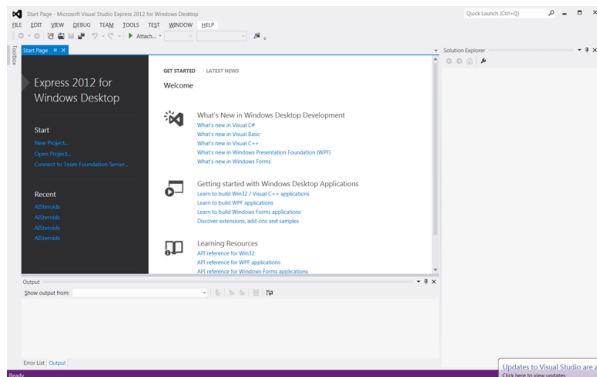
1.2 Windows

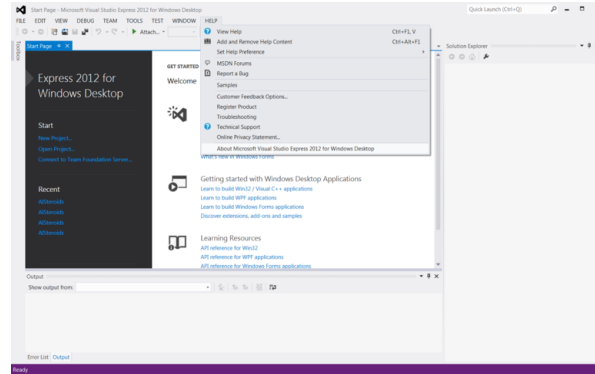
1.2.1 Visual Studio

Microsoft's Visual Studio 2010 or later is recommended.

To check the version:

Help (top tab) → *About Microsoft Visual Studio*





- 1.2.2 cmake (if compiling on Linux)
To check if cmake is installed:

```
$ which cmake
```

To check the version number:

```
$ cmake --version
```

- 1.2.3 nvcc/CUDA libs
The GPU builds of the code requires NVIDIA's CUDA 6.0 or newer:
To check if nvcc is installed:

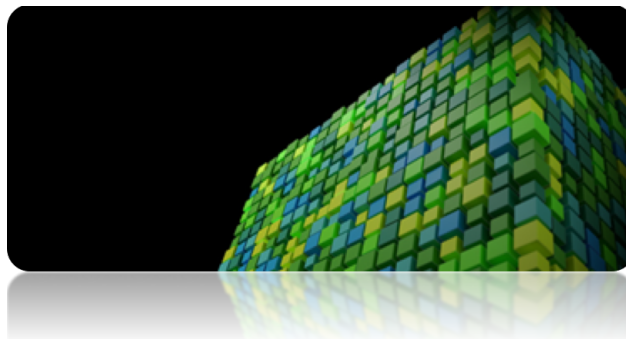
```
$ which nvcc
```

To check the version number:

```
$ nvcc --version
```

CUDA is viewed as an essential requirement, but is not used to compile the serial code, which can be compiled on systems without CUDA.
To download CUDA visit NVIDIA's webpage:

<https://developer.nvidia.com/cuda-downloads>



CUDA is required to compile the GPU executable in both Windows and Linux. Please refer to CUDA Developer webpages to select an appropriate version for the desired platform.

To install CUDA in Linux root/sudo, privileges are generally required. In Windows, administrative access is required.

5 Highly Recommended Software Tools

NOTE: *The listed programs are used in this manual and are generally considered necessary.*

5.1 VMD

VMD (Visual Molecular Dynamics) is a 3-D visualization and manipulation engine for molecular systems written in C-language. VMD is distributed and maintained by the University of Illinois at Urbana-Champaign. Its sources and binaries are free to download. It comes with a robust scripting engine, which is capable of running python and tcl scripts. More info can be found here:

<http://www.ks.uiuc.edu/Research/vmd/>

Although GOMC uses the same fundamental file types ? PDB (coordinates) and PSF (topology) as VMD, it uses some special tricks to obey certain rules of those file formats. One useful purpose of VMD is visualization of your systems.

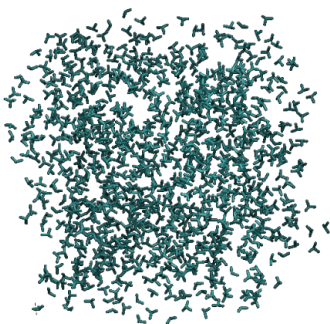


Figure 1: A system of united atom isobutane molecules

Nonetheless, the most critical part of VMD is a tool called PSFGen. PSFGen uses a tcl or python script to generate a PDB and PSF file for a system of one or more molecules. It is, perhaps, the most convenient way to generate a compliant PSF file.

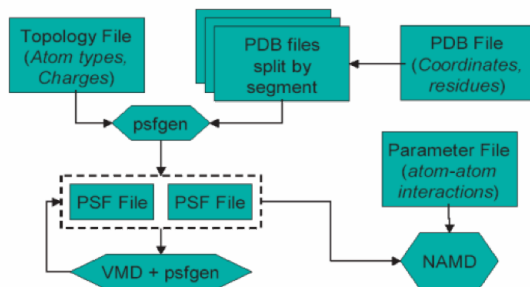


Figure 2: An overview of the PSFGen file generation process and its relationship to VMD/NAMD

To read more about PSFGen, reference:

Plugin homepage @ UIUC

<http://www.ks.uiuc.edu/Research/vmd/plugins/psfgen>

“Generating a Protein Structure File (PSF)”, part of the NAMD Tutorial from UIUC

<http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-html/node6.html>

In-Depth Overview [PDF]

<http://www.ks.uiuc.edu/Research/vmd/plugins/psfgen/ug.pdf>

5.2 Packmol

Packmol is a molecule packing tool created by José Mario Martínez, a professor of mathematics at the State University of Campinas, Brazil. It is written in Fortran and is free to download. More information is available on their homepage:

<http://www.ime.unicamp.br/~martinez/packmol>

To compile it, a Fortran language compiler is needed, such as gfortran. Many Linux distributions no longer come with Fortran compilers automatically, so this may need to be installed additionally.

Packmol allows a specified number of molecules to be packed at defined separating distances within a certain region of space. One of Packmol's limitations is that it is unaware of topology; it treats each molecule or group of molecules as a rigid set of points.

WARNING: *Another more serious limitation is that it is not aware of periodic boundary conditions (PBC). As a result, when using Packmol to pack PDBs for GOMC, it is recommended to pack to a box 1 Angstroms smaller than the simulation box size. This prevents hard overlaps over the periodic boundary.*

6 Other Useful Software Tools

6.1 Grace

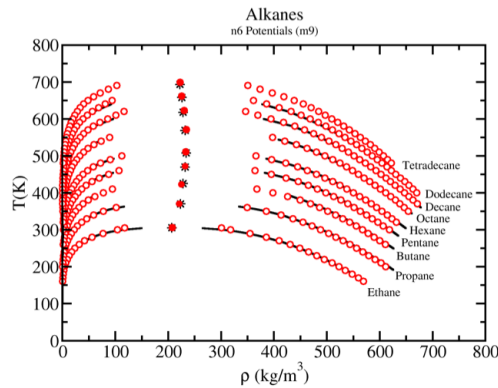
Grace is a piece of graphing software written and maintained by the Weizmann Institute of Science's Plasma Laboratory (Rehovot, Israel). Mostly used in Linux, it can also be compiled in Windows. The developers warn it may be missing some functionality.



In-depth information and the source can be found on the project page, here:

<http://plasma-gate.weizmann.ac.il/Grace>

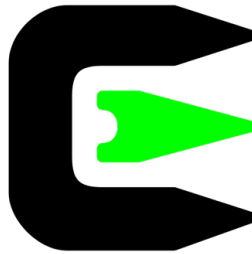
When compiled, Grace's executable in Linux is typically named “xmgrace”. This tool allows the production of high quality, precise line and dot graphs, ideal for visualizing much of the thermodynamic data from the GOMC engine. Below is an example of the results of simulations of saturated VLE densities of linear alkanes produced with Grace.



6.2 Cygwin

Cygwin is one option to assist in building and visualizing systems in Windows. It provides Microsoft Windows users with a Unix-like environment and command-line interface, and offers Windows-compatible ports of common Linux applications.

<https://cygwin.com>



The software is a free and open source, licensed under the GNU General Public License version 3. Its primary maintainers are Red Hat Inc. and NetApp. One of the most impressive abilities of Cygwin is its ability to launch a full Windows-compatible X-server Window, which allows convenient visualization of Linux app GUIs. It is compatible with the Grace graphing software. In practice, this package behaves most analogously to a Linux virtual machine in Windows.

7 Compiling GOMC

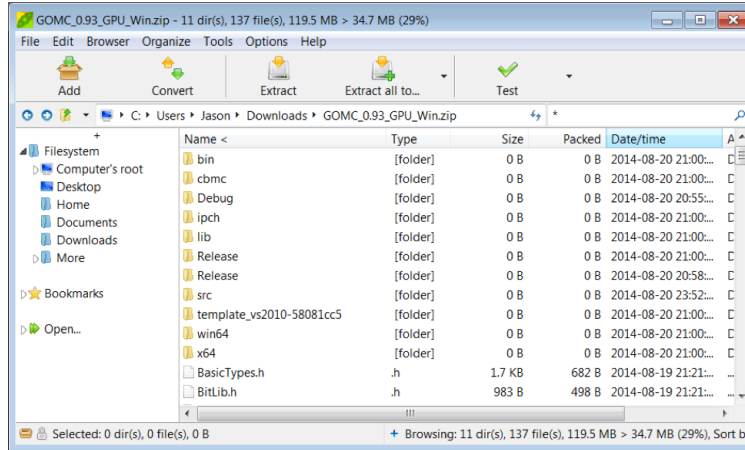
7.1 Extracting the code

GOMC is distributed as a compressed folder, containing the source and build system. To compile the code after downloading it, the first step is to extract the compressed build folder.

In Windows, the folder for the GPU code is compressed using a standard *.zip file format. To unzip simply use a utility like Peazip:

<http://peazip.sourceforge.net/>

Below is an example of what the downloaded code looks like when unzipping in Peazip.



In Linux, the GPU and Serial codes are compressed using gzip and tar (*.tar.gz). To extract, simply move to the desired folder and type in the command line:

```
$ tar -xzf <file name>.tar.gz
```

7.2 Compiling the code

7.2.1 GPU code

Compilation on Windows

Once the code is extracted, to compile it on Windows, you need to load the project into Visual Studio by opening the extracted folder and double clicking on the solution file of the desired Visual Studio version. After the solution is opened in Visual Studio, go to the “Build” menu and select “Build solution” to compile the code. You can compile either with release mode or with debug mode by selecting the desired mode from the “Solution Configuration” drop box. To run the project, simply click the run button or hit F5 on the keyboard.

Compilation on Linux

To compile the GPU code on Linux, go to the directory of the project, and type in the command line:

```
$ make
```

You can configure the “makefile” file to choose different C compilers, to select the desired compute capability, and to configure many more compilation flags.

The default compute capability is 3.0. To change the compute capability, go to the `GENCODE_FLAGS` option, and set it to one of the compute capability flags that are defined in the file.

```
# CUDA code generation flags
GENCODE_SM10    := -gencode arch=compute_10,code=sm_10
GENCODE_SM20    := -gencode arch=compute_20,code=sm_20
GENCODE_SM30    := -gencode arch=compute_30,code=sm_30
GENCODE_SM35    := -gencode arch=compute_35,code=sm_35
GENCODE_FLAGS   := $(GENCODE_SM30)
```

To run the program, run the executable “GOMC.out”. The system’s `LD_LIBRARY_PATH` will need to be configured to support CUDA (more on this later).

7.2.2 Serial Code

Compilation on Windows

See GPU “Compilation on Windows” section and follow an identical procedure for the released serial code. See README for instructions on how to use the CMake-GUI to build the configuration and solution files necessary for the Windows build.

Compilation on Linux

In Linux, the CPU code uses a simple makefile. Enter the directory and type in the command line:

```
$ make all
```

This will use the Makefile to compile a GPU-compatible executable called “GOMC.out”. To run, the system’s LD_LIBRARY_PATH will need to be configured to support CUDA (more on this later). For the serial code, which uses cmake for compilation, go to the base directory and type in the command line:

```
$ ./metamake.sh
```

This cmake script will create a directory named “bin”. Enter this directory:

```
$ cd bin
```

and type:

```
$ make
```

Four executables - GOMC.Serial.GEMC (Gibbs ensemble), GOMC.Serial.NVT (NVT ensemble), GOMC.Serial.NPT (isobaric-isothermal ensemble), and GOMC.Serial.GCMC (Grand canonical ensemble) - will be produced. By default, the distribution compiles in release mode. To compile in debug mode (if you’re using the code as a developer), open the file “CMakeCache.txt” while still in the “bin” folder. This file contains information used by cmake to build the executables. To compile in debug mode, change the value after “CMAKE_BUILD_TYPE:STRING=” from “Release” to “Debug”, and retype the command:

```
$ make
```

The output executables should now be compiled with debugger symbols. You can also swap the compiler by modifying the “CMAKE_CXX_COMPILER” variable. For more information, refer to the CMake documentation. Running GOMC in parallel using OpenMP:

To run the parallel version of CPU code, it needs to be compiled with openmp library. Open the file “CMakeCache.txt”, while still in the “bin” folder, and change the value after “CMAKE_CXX_FLAGS_RELEASE:STRING=” from “-O3 -DNDEBUG” to “-O3 -qopenmp -DNDEBUG”.

And retype the command:

```
$ make
```

8 Input File Formats

In order to run simulation in GOMC, the following files need to be provided:

- GOMC executable
- Input file “NAME.conf” (proprietary control file)

- PDB file(s)
- PSF file(s)
- Parameter file

8.1 PDB

The PDB file stores coordinates for the simulation. The file format is widely adopted.

- Protein Databank (PDB) Files (plural: PDB files)
- Open format, well-documented
- Fixed-width format (hence white space is significant)
- Up to 13.5m page views a month; up to 55.8m FTP requests per month
- Used by NAMD, GROMACS, CHARMM, ACEMD, Amber

An overview of the PDB standard can be found here:

<http://www.wwpdb.org/docs.html>

The advantage of PDB files is their ubiquity and thorough documentation. Disadvantages include limited fixed point floating precision for coordinates, unused space, and proprietary implementations creating inconsistencies.

One PDB file is required per box. For NVT ensemble simulations, one file is expected; for Gibbs and grand canonical ensemble, two files are required. GOMC recognizes the following keywords in PDB files:

- ◇ REMARK
- ◇ CRYST1
- ◇ ATOM
- ◇ END

Currently, REMARK is ignored. Formerly, it was used to store proprietary information in frames (e.g. step number). Packmol typically leaves the following remark:

REMARK	original	generated	coordinate	pdb	file
--------	----------	-----------	------------	-----	------

at the top of the file. Note that this is another example of an inconsistency with the spec. As of the PDB v3.30 specification the REMARK entry contains an identifying integer, which is supposed to occupy lines 8-10.

Main Index

REMARK 3

REMARK 0, 1, 2, 4, 5 - 299

REMARK 300 - 999

REMARKS

Overview

REMARK records present experimental details, annotations, comments, and information not included in other records. In a number of cases, REMARKs are used to expand the contents of other record types. A new level of structure is being used for some REMARK records. This is expected to facilitate searching and will assist in the conversion to a relational database.

The very first line of every set of REMARK records is used as a spacer to aid in reading.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
8 - 10	Integer	remarkNum	Remark number. It is not an error for remark n to exist in an entry when remark n-1 does not.
12 - 79	LString	empty	Left as white space in first line of each new remark.

REMARK 3

REMARK 0,1,2,4,5-299

REMARK 300-999

A file generated by Packmol has "ori" in this position. Hence you may see future codes that are incompatible with this legacy kind of remarks.

Note also that the spaces 7 and 11 are not reserved; hence, they may be used in proprietary specifications. CRYST1 can be used to store the cell dimensions, which can also be put as a tag in the proprietary control file.

<http://www.wwpdb.org/documentation/format33/sect8.html#CRYST1>

Crystallographic and Coordinate Transformation Section

This section describes the geometry of the crystallographic experiment and the coordinate system transformations.

CRYST1

Overview

The CRYST1 record presents the unit cell parameters, space group, and Z value. If the structure was not determined by crystallographic means, CRYST1 simply provides the unitary values, with an appropriate REMARK.

Record Format

COLUMNS	DATA	TYPE	FIELD	DEFINITION
1 - 6	Record name		"CRYST1"	
7 - 15	Real (9.3)		a	a (Angstroms).
16 - 24	Real (9.3)		b	b (Angstroms).
25 - 33	Real (9.3)		c	c (Angstroms).
34 - 40	Real (7.2)		alpha	alpha (degrees).
41 - 47	Real (7.2)		beta	beta (degrees).
48 - 54	Real (7.2)		gamma	gamma (degrees).
56 - 66	LString		sGroup	Space group.
67 - 70	Integer		z	Z value.

Details

- If the entry describes a structure determined by a technique other than X-ray crystallography, CRYST1 contains $a = b = c = 1.0$, $\alpha = \beta = \gamma = 90$ degrees, space group = P 1, and $Z = 1$.
- The Hermann-Mauguin space group symbol is given without parenthesis, e.g., P 43 21 2. Please note that the screw axis is described as a two digit number.
- The full International Table's Hermann-Mauguin symbol is used, e.g., P 1 21 1 instead of P 21.
- For a rhombohedral space group in the hexagonal setting, the lattice type symbol used is H.
- The Z value is the number of polymeric chains in a unit cell. In the case of heteropolymers, Z is the number of occurrences of the most populous chain.

As an example, given two chains A and B, each with a different sequence, and the space group P 2 that has two equipoints in the standard unit cell, the following table gives the correct Z value.

NOTE: Only cubic and orthogonal cells are supported in this code. The main entry in the PDB file are ATOM| entries. The keyword "ATOM" is always followed by two spaces. An entry has a number of fields.

Coordinate Section

The Coordinate Section contains the collection of atomic coordinates as well as the MODEL and ENDMDL records.

ATOM

Overview

The ATOM records present the atomic coordinates for standard amino acids and nucleotides. They also present the occupancy and temperature factor for each atom. Non-polymer chemical coordinates use the HETATM record type. The element symbol is always present on each ATOM record; charge is optional.

Changes in ATOM/HETATM records result from the standardization atom and residue nomenclature. This nomenclature is described in the Chemical Component Dictionary (<ftp://ftp.wwpdb.org/pub/pdb/data/monomers>).

Record Format

COLUMNS	DATA	TYPE	FIELD	DEFINITION
1 - 6	Record name		"ATOM "	
7 - 11	Integer		serial	Atom serial number.
13 - 16	Atom		name	Atom name.
17	Character		altLoc	Alternate location indicator.
18 - 20	Residue name		resName	Residue name.
22	Character		chainID	Chain identifier.
23 - 26	Integer		resSeq	Residue sequence number.
27	AChar		iCode	Code for insertion of residues.
31 - 38	Real (8.3)		x	Orthogonal coordinates for X in Angstroms.
39 - 46	Real (8.3)		y	Orthogonal coordinates for Y in Angstroms.
47 - 54	Real (8.3)		z	Orthogonal coordinates for Z in Angstroms.
55 - 60	Real (6.2)		occupancy	Occupancy.
61 - 66	Real (6.2)		tempFactor	Temperature factor.
77 - 78	LString(2)		element	Element symbol, right-justified.
79 - 80	LString(2)		charge	Charge on the atom.

Details

- ATOM records for proteins are listed from amino to carboxyl terminus.
- Nucleic acid residues are listed from the 5' to the 3' terminus.
- Alignment of one-letter atom name such as C starts at column 14, while two-letter atom name such as FE starts at column 13.
- Atom nomenclature begins with atom type.
- No ordering is specified for polysaccharides.
- Non-blank alphanumeric character is used for chain identifier.
- The list of ATOM records in a chain is terminated by a TER record.
- If more than one model is present in the entry, each model is delimited by MODEL and ENDMDL records.
- AltLoc is the place holder to indicate alternate conformation. The alternate conformation can be in the entire polymer chain, or several residues or partial residue (several atoms within one residue). If an atom is provided in more than one position, then a non-blank alternate location indicator must be used for each of the atomic positions. Within a residue, all atoms that are associated with each other in a given conformation are assigned the same alternate position indicator. There are two ways of representing alternate conformation- either at atom level or at residue level (see examples).
- For atoms that are in alternate sites indicated by the alternate site indicator, sorting of atoms in the ATOM/HETATM list uses the following general rules:
 - In the simple case that involves a few atoms or a few residues with alternate sites, the coordinates occur one after the other in the entry.
 - In the case of a large heterogen groups which are disordered, the atoms for each conformer are listed together.
- Alphabet letters are commonly used for insertion code. The insertion code is used when two residues have the same numbering. The combination of residue numbering and insertion code defines the unique residue.
- If the depositor provides the data, then the isotropic B value is given for the temperature factor.
- If there are neither isotropic B values from the depositor, nor anisotropic temperature factors in ANISOU, then the default value of 0.0 is used for the temperature factor.
- Columns 79 - 80 indicate any charge on the atom, e.g., 2+, 1-. In most cases, these are blank.
- For refinements with program REFMAC prior 5.5.0042 which use TLS refinement, the values of B may include only the TLS contribution to the isotropic temperature factor rather than the full isotropic value.

The key parameters are the coordinates x, y, and z. The precision is limited to eight whole decimal digits and three fractional decimal digits.

Other important entries are the residue name, atom name, and chain ID. Numbering is important primarily because it represents an inconvenience in packing/loading large systems. Revisiting the previous example,

the atom name is “C1” and residue name is “ISB”. The PSF file (next section) contains a lookup table of atoms. These contain the atom name from the PDB and the name of the atom kind in the parameter file it corresponds to. As multiple different atom names will all correspond to the same parameter, these can be viewed “atom aliases” of sorts. The chain letter (in this case ‘A’) is sometimes used when packing a number of PDBs into a single PDB file.

A few important **Notes** / **Warnings** on Undocumented PDB Format Conventions:

- While it is explicitly stated in some other sections of the PDB file, the general convention observed by most codes is to right align when padding with white space.
- Some codes (including PSFGen/VMD) use the 21st unused character to add a fourth letter to the residue (molecule name). This extension is currently supported, but is unofficial and, hence, may change in the future.
- VMD requires a constant number of ATOMS in a multi-frame PDB (multiple records terminated by “END” in a single file). To compensate for this, all atoms from all boxes in the system are written to the output PDBs of this code.
- For atoms not currently in a box, the coordinates are set to $< 0.00, 0.00, 0.00 >$
- The occupancy is commonly just set to “1.00” and is left unused by many codes. We recycle this legacy parameter by using it to denote, in our output PDBs, the box a particle is in (box 0 occupancy=0.00 ; box 1 occupancy=1.00)
- As the x, y, and z coordinates are fixed point with only three digits of precision, the energy values you get when restarting may be mildly different, particularly for bonded interactions due to roundoff in the coordinates. This will eventually be remedied by the implementation of a full-precision trajectory (e.g. DCD) file.
- The “ISB” entry in columns 73-75 is not an official part of the PDB standard. This is a proprietary entry called “Segname”, which has been embraced by NAMD and some other codes.