

31.03.2006 г.–31.05.2006 г.

**Вариант 8**  
**ИНСТРУКЦИЯ ПО УСТАНОВКЕ СЕМЕЙСТВА СИЛОВЫХ ПОЛЕЙ AMBER**  
**ДЛЯ ПРОГРАММНОГО ПАКЕТА ПО МОЛЕКУЛЯРНОЙ ДИНАМИКЕ GROMACS 3.3**

*В. В. Колосов (sv.kolosov@gmail.com), К. О. Одынец*

Использовали инструкцию по установке семейства силовых полей (СП) AMBER для пакета GROMACS 3.3, взятую на веб-сайте авторов: E.J. Sorin, S. Park, 2005:

<http://folding.stanford.edu/ffamber/index.html>

В связи с выходом новой версии GROMACS 3.3.1 от 05 апреля 2006 г. возможны некоторые изменения в данной инструкции.

На веб-сайте пакета GROMACS желательно посмотреть страницу, посвященную силовым полям:

[http://www.gromacs.org/topologies/force\\_fields.php](http://www.gromacs.org/topologies/force_fields.php)

GROMACS 3.3 и СП AMBER уже установлены на вычислительном кластере КНУ имени Тараса Шевченко:

<http://www.cluster.univ.kiev.ua>

и кластерах института кибернетики НАН Украины:

<http://cluster.icyb.kiev.ua/>

Литература

1. Sorin E. J., Pande V. S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations // Biophys. J.–2005.–**88**.–4.–P.2472–2493.

1. Архивный файл и его распаковка

Загрузить с сайта <http://folding.stanford.edu/ffamber/index.html> архивный файл **ffamber\_v3.3-doc.tar.gz** с данными по шести разным СП AMBER размером 6328320 байт. После распаковки архива получится каталог **/ffamber\_v3.3/**, в котором находится 7 папок и 14 файлов. Среди них находится 6 папок с СП AMBER-94, AMBER-96, AMBER-99, AMBER-GS, AMBER-GS-S и AMBER-99ф; одна папка со статьями и копией инструкции, расположенной на сайте; файлы координат и топологии для различных типов воды, созданные для СП AMBER и несколько системных файлов пакета GROMACS, оптимизированных с условием добавленных силовых полей.

2. Копирование файлов

Предполагается, что каталог **/top/** с файлами топологии инсталлированного пакета GROMACS 3.3 имеет путь **/usr/local/gromacs\_3.3/share/gromacs/top/**.

В этот каталог **/top** нужно скопировать из каталога **/ffamber\_v3.3/** файл **aminoacids-NA.dat** (нужно предварительно его переименовать на просто **aminoacids.dat** – это нужно для работы с нуклеиновыми кислотами и их комплексами с белками) и файл **vdwradii.dat** так, чтобы они заменили исходные файлы GROMACS с теми же названиями. В файле **aminoacids.dat** содержатся имена аминокислотных остатков и нуклеотидов, которые описаны топологией СП AMBER. В файле **vdwradii.dat** содержатся аппроксимированные величины Ван-дер-Ваальсовых радиусов некоторых атомов.

Судя по статьям из каталога **/ffamber\_v3.3/README/pdfs/**, которые описывают шесть различных СП семейства AMBER, последним, универсальным и наиболее успешным является СП AMBER-99ф (читается “фи”), которое обозначается как **ffamber99p**. Нужно скопировать в тот же каталог **/top/** (в одну кучу) все 8 файлов этого СП из каталога **/ffamber\_v3.3/ffamber99p/**.

Для других СП семейства AMBER нужно аналогично копировать по 8 файлов из соответствующих каталогов.

3. Редактирование файла FF.dat

В верхней строке файла **FF.dat** из каталога **/top/** нужно увеличить число, обозначающее общее количество установленных силовых полей с 11 на 17. Далее в конце этого файла нужно добавить 6

новых строк с именами и описаниями добавленных новых СП AMBER по примеру имеющихся строк.

#### Исходный текст файла **FF.dat**

```
11
ffG43a1    GROMOS96 43a1 force field
ffG43b1    GROMOS96 43b1 vacuum force field
ffG43a2    GROMOS96 43a2 force field (improved alkane dihedrals)
ffG45a3    GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
ffG53a5    GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
ffG53a6    GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
ffop1saa   OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
ffgmx      [DEPRECATED] Gromacs force field (see manual)
ffgmx2     [DEPRECATED] Gromacs force field with hydrogens for NMR
ffencadv   Encad all-atom force field, using scaled-down vacuum charges
ffencads   Encad all-atom force field, using full solvent charges
```

#### Отредактированный текст файла **FF.dat**

```
17
ffG43a1    GROMOS96 43a1 force field
...
ffencads   Encad all-atom force field, using full solvent charges
ffamber94  AMBER94 Cornell protein/nucleic forcefield
ffamber96  AMBER96 Kollman et. al.
ffamber99  AMBER99 Wang protein/nucleic acid forcefield
ffamberGS  AMBERGS Garcia and Sanbonmatsu
ffamberGSs AMBERGSs Garcia and Nymeyer
ffamber99p AMBER99p protein/nucleic forcefield Sorin and Pande
```

#### 4. Изменение переменных окружения

Перейти в каталог **/bin/** пакета GROMACS, найти исполняемый файл **GMXRC** и выполнить команду:

```
>source GMXRC
```

Эта команда нужна для изменения некоторых переменных окружения в GROMACS. Она срабатывает, но не выводит никакого комментария.

#### 5. Топология кластеров молекул воды

В каталог **/top/** нужно скопировать 5 пар файлов координат и топологии для 5-ти разных моделей кластера молекул растворителя (воды), созданных с учетом особенностей СП AMBER:

```
ffamber_tip3p.gro
ffamber_tip3p.itp
ffamber_tip3p_heavy.gro
ffamber_tip3p_heavy.itp
ffamber_tip4p.gro
ffamber_tip4p.itp
ffamber_tip4pEW.gro
ffamber_tip4pEW.itp
ffamber_tip5p.gro
ffamber_tip5p.itp
```

#### 6. Топология ионов $Na^+$ и $Cl^-$

Для правильного подключения ионов  $Na^+$  и  $Cl^-$  нужно использовать именно те их обозначения, которые приведены в файле **ffamber\*.rtp** из соответствующего СП AMBER. Для обеспечения этого имеется 2 варианта – это результат поиска по e-мэйл архиву веб-сайта GROMACS <http://www.gromacs.org/search/index.php> и переписки с Е.Ж. Сорин, Д.Б. Ковальским, V. Venditti и еще некоторыми людьми, которые сталкивались с данной проблемой.

1-й вариант (более предпочтительный) Добавить дополнительные секции в конец файла **ions.itp** с топологией ионов в каталоге **/top/** пакета GROMACS, описывающие параметры ионов в СП AMBER. Секция для описания топологии иона  $Na^+$  в СП AMBER-99ф была любезно предоставлена Vincenzo Venditti ([vincenzovenditti@hotmail.com](mailto:vincenzovenditti@hotmail.com)):

```
#ifdef _FF_AMBER99p
[ moleculetype ]
; molname      nrexcl
```

```

NA+          1
[ atoms ]
; id at      type      res nr residu name  at name      cg nr  charge mass
1      amber99_31  1      NA+          NA           1          1      22.98977
#endif

```

Секцию для топологии иона  $\text{Cl}^-$  добавлять аналогично.

```

#ifdef _FF_AMBER99p
[ moleculetype ]
; molname    nrexcl
CL-          1
[ atoms ]
; id at      type      res nr residu name  at name      cg nr  charge mass
1      amber99_30  1      CL-          CL           1          -1     35.45000
#endif

```

2-й вариант (менее радикальный) Скопировать в папку по конкретному проекту новые файлы **na+.itp** и **cl-.itp**, описывающие топологию ионов  $\text{Na}^+$  и  $\text{Cl}^-$  в СП AMBER-99ф. В качестве текста файла **na+.itp** должно быть:

```

[ moleculetype ]
; molname    nrexcl
NA+          1
[ atoms ]
; id at      type      res nr residu name  at name      cg nr  charge mass
1      amber99_31  1      NA+          NA           1          1      22.98977

```

а в файле **cl-.itp**

```

[ moleculetype ]
; molname    nrexcl
CL-          1
[ atoms ]
; id at      type      res nr residu name  at name      cg nr  charge mass
1      amber99_30  1      CL-          CL           1          -1     35.45000

```

В файле топологии **\*.top** подготавливаемой нами системы для каждого конкретного проекта в секции подключения топологии для ионов необходимо вместо ссылки на основной файл ионов GROMACS (строка **#include "ions.itp"**) поставить ссылки на новые файлы для используемых ионов  $\text{Na}^+$  и  $\text{Cl}^-$ . Необходимо изменить:

```

; Include generic topology for ions
#include "ions.itp"
на:
; Include generic topology for ions
#include "na+.itp"
#include "cl-.itp"

```

Соответственно в команде **genion** и **\*.mdp** файлах использовать имена ионов  $\text{NA}^+$  и  $\text{CL}^-$ .

## 7. Добавление ссылки на кластеры воды

При запуске команды **genbox** в опции **-cs** вместо модели кластера воды **spc216.gro** необходимо использовать, например, **ffamber\_tip3p.gro**, или другой из приведенных выше 5-ти вариантов.

Также нужно отредактировать секцию растворителя в файле топологии конкретного проекта – подставить в него ссылку на файл **ffamber\_tip3p.itp** вместо имеющегося по умолчанию файла **spc.itp** – заменить:

```

#include "spc.itp"
на:
#include "ffamber_tip3p.itp"

```

## 8. Переименовка аминокислотных остатков и нуклеотидов

В СП AMBER приняты немного другие обозначения для аминокислотных остатков белков и нуклеотидов нуклеиновых кислот. Для того, чтобы понять, как корректно переименовывать аминокислотные остатки и нуклеотиды в координатном **pdb**-файле очень важно посмотреть те файлы

примеров, которые идут вместе с инструкцией.

Для белков нужно переименовать некоторые внутренние аминокислотные остатки, чтобы перевести их в ионизированное состояние – по схеме: ASP/ASH, GLU/GLH, LYS/LYP, CYS/(CYN или CYM), HIS/(HID, HIE или HIP). Также нужно переименовать NH<sub>2</sub>-концевые остатки по принципу ALA/NALA и COOH-концевые остатки по принципу ALA/CALA. Все подобные замены легко провести через функцию “заменить все” во многих текстовых редакторах.

Для нуклеиновых кислот для РНК нужно переименовать внутренние нуклеотиды по принципу: A/RA, C/RC, G/RG и U/RU. Также нужно переименовать 5'-концевые нуклеотиды на RA5, RC5, RG5 или RU5, а 3'-концевые нуклеотиды на RA3, RC3, RG3 или RU3.

Для 5'-концевых нуклеотидов с 5'-фосфатной группой отсутствуют необходимые готовые топологии, и их еще нужно будет дополнительно прописывать по примеру из оригинальных файлов СП AMBER.

## 9. Переименовка атомов

Для некоторых атомов придется переименовывать имена атомов согласно тому, как они названы в файле топологии **\*.rtp**. Нужно найти в нем секцию, отвечающую за соответствующий аминокислотный остаток или нуклеотид и посмотреть, как там именуются атомы. Потом соответственно заменить имена атомов в координатном **\*.pdb** файле.

Для молекул РНК в координатном файле **\*.pdb** необходимо вместо 5'-концевой фосфатной группы подставить “заглушку” из терминального атома водорода в строку с координатами атома фосфора, т.е. заменить строки:

```
...O1P...
...O2P...
... P...
на:
...H5T...
```

## Фрагмент инструкции к методике E.J. Sorin, S. Park, 2005

### Installation & Testing

- (1) Install the desired [GROMACS](#) distribution (v3.1.4, v3.2.1, or v3.3).
- (2) Download the appropriate ffamber ports (.tar.gz) with or without pdf documentation from the table below, being sure that the version number you choose matches the version of GROMACS you are using.
- (3) Unzip/untar the downloaded tar.gz file.
- (4) Copy aminoacids.dat and vdwradii.dat to the "top" directory in your gromacs distribution (you should see force field files there, such as ffoplsaa.\*). If you plan on simulating nucleic acids, refer to the [note for nucleic acids in aminoacids.dat](#) below.
- (5) Files for each force field are located in a separate subdirectory, such as ffamber94/ for the Cornell potential. Copy the desired ffamber\* files to the top directory in your gromacs distribution.
- (6) Increment the number at the top of the "top/FF.dat" file by 1 for each AMBER port you'll install (so that it matches the total number of forcefields available in the "top" directory).
- (7) Add lines like the following to the "top/FF.dat" file. These are used by pdb2gmx to allow you to identify the desired FF and field 1 must match the ffamber\* filename prefixes, whereas the following fields can be user-defined:  
 ffamber94 AMBER94 Cornell protein/nucleic forcefield  
 ffamber99 AMBER99 Wang protein/nucleic acid forcefield  
 ffamber99p AMBER99p protein/nucleic forcefield
- (8) Locate the GMXRC in your GROMACS distribution and run `source GMXRC`.
- (9) Run `pdb2gmx -H14 -f any.pdb` with any pdb to verify that these force fields are now seen by GROMACS. Working example .pdb files are available below.  
[Villin headpeice](#) (protein) [Dickerson dodecamer](#) (DNA) [GCAA tetraloop hairpin](#) (RNA)
- (10) Modify residue names in your pdb file(s) as described below to properly generate .gro and .top files using pdb2gmx (the files above are ready to use).

### Important notes on using the AMBER ports:

This list includes only the most important information needed to properly use our AMBER ports inside the GROMACS suite. Non-vital information is compiled in the [FAQ](#) listed above. Information on using GROMACS as well as FAQs, tutorials, and a user forum are available at <http://www.gromacs.org/>.

- (1) **(!) -H14 flag in gromacs:** You must use the -H14 flag when running the GROMACS v3.1.4 version of pdb2gmx to get \*ALL\* hydrogen-hydrogen 1-4 interactions that should be present in AMBER topologies. This has been corrected in v3.2.1 and the -H14 flag is no longer necessary.
- (2) **(!) Residue Nomenclature:** Residues in the AMBER ports are named according to their position in the sequence (i.e. terminal, non-terminal, monomer) following standard AMBER conventions. For this reason, it may be necessary to rename residues in the .pdb file you will import beforehand. Please note that all residues are named in the residue topology files (i.e. ffamber\*.rtp), so if you are unsure of the correct residue name to use, you should be able to find it there. The .rtp files are ordered as follows: water models (TIP), ions & urea (URE), peptide terminal capping residues (ie. ACE, NH2, NMe), non-terminal amino acids (i.e. TYR, ALA), non-terminal acidic amino acids (i.e. ASH, GLH, etc.), C-terminal amino acids (i.e. CALA, CGLY), N-terminal amino acids (i.e. NALA, NGLY), and all nucleic acid residues. Nucleic acids listed at the end of each .rtp follow the following order for each residue type: DNA is first, followed by RNA, in the order 5'-term, 3'-term, non-terminal, and monomer. The three .pdb files above are examples of how pdb files should be modified. Residues in the ffamber ports have been named as follows:
  - (a) **Non-terminal amino and nucleic acid residues** follow standard AMBER naming conventions. To avoid confusion between GROMACS and AMBER conventions, we have omitted the redundant HIS residue, leaving HID, HIE, HIP, and terminal versions of these topologies. Additionally, due to the automated changing of certain residue names by pdb2gmx, the LYS and CYS residues have been renamed LYP (Lysine plus) and CYN (Cysteine neutral, compared to AMBER residue CYM = Cysteine minus).
  - (b) **C- and N-terminal amino acids** include a C or N prefix respectively, so C-terminal ALA is CALA and N-terminal PHE is NPHE. As with non-terminal versions, the LYS and CYS terminal residues are listed as NLYP,CLYP and NCYN,CCYN.
  - (c) **Nucleic acid residues** come in four flavors. All residue names include XY, where X = D or R for DNA or RNA respectively, and Y = first letter of the nucleotide name. A suffix (XYZ) is added for monomers (Z=N), 5'-terminal (Z=5), and 3'-terminal (Z=3) residues. For example, 3'-term DNA Cytosine = "DC3", 5'-term RNA Cytosine = "RC5", non-terminal DNA Cytosine = "DC", and lone RNA Cytosine monomer = "RCN".
- (3) **(!) Atom Nomenclature:** Once you have modified the residue names in your pdb file according to the rules above, pdb2gmx may report a fatal error like "Fatal error: Atom AA in residue XYZ not found in rtp entry with NN atoms

while sorting atoms." This may or may not be caused by the use of an aminoacids.dat file in the "top" directory of your GROMACS distribution that does not include the name of the residue that is causing the problem (XYZ in this case). This can be easily fixed by modifying the name of the problematic atom in your .pdb file to match what is shown for residue XYZ in the .rtp file.

(4) **(!) Water nomenclature:** residues in pre-solvated pdb files should be named as shown here to use the correct water model in the desired amber port: TIP3P ("HOH" or "T3P"), heavy TIP3P ("T3H"), TIP4P ("T4P"), TIP4P-Ew ("T4E"), and TIP5P ("T5P"). This usage allows any model to be used without modifying the force field (.rtp) files. Note, however, that for the default "HOH" listing the TIP3P model will be assumed. Also note that because GROMACS supports SOL=HOH water definitions, importing a solvated PDB with names other than HOH or SOL may cause the solvent molecules to be treated as part of the biopolymer rather than listing SOL molecules at the end of the topology, which has to be modified by hand after the fact.

(5) **(!) Specifying water models in .top files:** If not using the default SPC water model in GROMACS, make sure you list the correct ffamber\_tip\*.itp file in your gromacs topology file! TIP3P heavy water (H's four times heavier, difference subtracted from oxygen) is defined in ffamber\_tip3p\_heavy.itp, and the TIP4P-Ewald model uses ffamber\_tip4pEW.itp. All other models are specified in the ffamber\_tipXp.itp files, where X = {3,4,5}. Water boxes are also present, with the .gro file extension.

(6) **(!) For nucleic acid simulations:** Amino acid residues have been added to aminoacids.dat, which (i) allows easy use of GROMACS analysis tools and (ii) aids GROMACS in matching atoms in a pdb with those in the residue topology (.rtp) file. However, for nucleic acids this also often causes pdb2gmx to replace an H atom in the first residue of all nucleic acid chains with an incorrect H atom, resulting in non-neutral charge. The correct atom is generally replaced with an atom of type amberXX\_25 (hydroxyl H), as pdb2gmx treats it as a terminal hydrogen. For this reason, we provide both the standard [aminoacids.dat](#) (without nucleic acids) and a second version with nucleic acid residues included ([aminoacids-NA.dat](#)). You can run [gcaa.pdb](#) with both versions of aminoacids.dat by copying either file into your "top" directory before each trial and diffing the resulting top files to see this difference. There are three ways to handle this:

**(a)** Include all nucleic acid names in aminoacids.dat and correct the [ atoms ] section of the resulting .top file by hand/script after running pdb2gmx (both the atom type and the charge).

**(b)** Remove nucleotide names from aminoacids.dat before importing your pdb file and add them back afterward to use standard GROMACS tools.

**(c)** Do not include nucleic acid names in aminoacids.dat and make an index file which specifies the DNA/RNA atoms to use standard GROMACS tools.

Note that if you choose options **(b)** or **(c)** above, you may need to modify the atom names in your .pdb file to agree with the residue listings in the .rtp file due to issue (3) above.

(7) **(!) Using AMBER ions:** Our AMBER ports include [common ion definitions](#), which are listed in the ffamber\*.rtp files (just below the TIP water models). This allows the AMBER ports to be used without modification or use of the GROMACS ions.itp file. At the moment these include Cl<sup>-</sup>, IB<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Rb<sup>+</sup>, Cs<sup>+</sup>, Li<sup>+</sup>, Ca<sup>2+</sup>, Mg<sup>2+</sup>, Zn<sup>2+</sup>, Sr<sup>2+</sup>, and Ba<sup>2+</sup>. If your pdb file has ions present and pdb2gmx does not properly convert those ions, please check the atom and residue name of your ions and [rename them if necessary to agree with the .rtp file](#). If you are using ion-related GROMACS tools, such as genion, you will need to enter the AMBER ion definition to the ions.itp file in the "top" directory of the GROMACS distribution.