pdb2gmx finds OXT in non-terminal non-standard amino acid

Description
Created an attachment (id=539)
Coordinate file, residuetypes.dat, force field files

As described on the users' list:
http://lists.gromacs.org/pipermail/gmx-users/2010-September/054178.html

Files for reproducing the problem (coordinate file, residuetypes.dat, and force field files) are attached.

Associated revisions
Revision 9ad93f07 - 06/24/2014 12:41 AM - Erik Lindahl
Enable terminus-specific atom translation

By adding protein-nterm and protein-cterm classes for atom name translation in xlateat.dat we now avoid replacing names such as O2 in non-standard amino acids. This patch also corrects a cosmetic issue in the number of residues claimed to be found in each chain by pdb2gmx.

Fixes #567.

Change-Id: I5956991e3ff3e5e1202330a857fd773f9e9b6945

History
#1 - 09/21/2010 03:52 PM - Justin Lemkul
Debug information from pdb2gmx indicates that the problematic atom is O2, which in "chain_A0.pdb" (debug output) is assigned an "OXT" name:

```
... ATOM 2553  C2  CRO A 331  -10.726  6.753  -22.143  1.00  0.00           C
ATOM 2554  OXT CRO A 331  -11.108  5.574  -21.819  1.00  0.00           O
ATOM 2555  N3  CRO A 331  -10.810  7.255  -23.355  1.00  0.00           N
ATOM 2556  CA3 CRO A 331  -11.435  6.545  -24.488  1.00  0.00           C
ATOM 2557  C  CRO A 331  -10.492  6.111  -25.580  1.00  0.00           C
ATOM 2558  O  CRO A 331  -10.993  5.596  -26.570  1.00  0.00           O
...```
Replacing the "O2" atom name with something else, like "OA" has fixed the problem. It seems, though, that something in the code is telling pdb2gmx to replace any instance of "O2" with "OXT" regardless of whether the residue is a terminus or not. I would think that it would be preferable to allow the user to define whatever names are desired to the atoms, and as long as they match the .rtp entry, should not be altered. It's just odd because pdb2gmx correctly identifies the N- and C-termini:

Identified residue LYS1 as a starting terminus.
Identified residue LYS454 as an ending terminus.

Hence why it is very odd that anything is being replaced at residues that are not LYS1 or LYS454.

---

The reason is that xlateat.dat has the line:
protein O2 OXT

Not sure how this should take into account whether it is a termini. Any ideas of how make this mapping more specific?

---

The translation that caused a problem for me was terminus-specific, but in xlateat.dat can be applied to any protein residue. Other translations (like for ILE, HOH, etc) are limited to a specific residue. Could something like this be used for termini? For example, the line could read:

terminus O2 OXT

Since pdb2gmx identifies and prints out the starting and ending termini due to the new chain decision structure, I would think this could be possible, but I don't know how laborious it would be. Maybe it doesn't even really need fixing, but it seems to me that the purpose of xlateat.dat is to make the atom names conform to the .rtp entry, the latter of which should take precedence. In my case, I named an atom O2, in both the .rtp and .pdb files, but then xlateat.dat "un-helped" me by changing it. Maybe a check could be implemented such that xlateat.dat names are only used when no match can be found in the .rtp entry first? That way, anything user-defined would override whatever assumptions Gromacs is making.
I agree xlateat.dat should be done better but I don't think it qualifies as bug. So I decrease it to priority low.


Gerrit received a related patchset '1' for Issue #567.
Uploader: Erik Lindahl (erik@kth.se)
Change-Id: I5956991e3ff3e5e1202330a857fd773f9e9b6945
Gerrit URL: [https://gerrit.gromacs.org/3670](https://gerrit.gromacs.org/3670)

Uploader: Justin Lemkul
File: files.tgz
Size: 119 KB
Date: 09/20/2010