How do I run multiple proteins at once? p_batch and pro_batch

p_batch is a program included in MCCE, under the folder MCCE_bin. It accepts a directory containing ".pdb" files, and runs MCCE with identical settings on each protein file.

```
p_batch -h

usage: p_batch [-h] [-custom script_path] input_path

p_batch accepts a directory containing PDB files, and executes identical MCCE runs in
directories named after each respective PDB file.

positional arguments:
  input_path           Path to a .pdb file

options:
  -h, --help           show this help message and exit
  -custom script_path  Give a shell script with custom instructions. If not defined, a default
script will be created and used.
```

By default, default_script.sh is created by p_batch at run time, if it does not exist. The default scripts calls for steps 1-4 of MCCE to be run at level 1, assumes a dielectric constant of 8 with NGPB as the Poisson Bolztmann solver, and sets entropy control on.

Another tool called "pro_batch" is available. This behaves very similarly to p_batch, though the default instruction script pro_batch creates has more features for heavy users, including the ability to use a scheduler. Both p_batch and pro_batch can accept custom shell scripts with the "-custom" flag. Learn about creating and editing shell scripts here.

```
pro_batch -h
```

```
usage: pro_batch [-h] [-custom script_path] [--sbatch | --no-sbatch] input_path

pro_batch accepts a directory containing PDB files, and executes identical MCCE runs in
directories named after each respective PDB file. pro_batch creates
a high-level shell script. The user then edits the shell script to their liking, and executes
it with the -custom flag.

positional arguments:
  input_path            Path to a directory containing PDB files.

options:
  -h, --help            show this help message and exit
  -custom script_path   Give a shell script with custom instructions. If not defined, a
default script will be created and used.
  --sbatch, --no-sbatch
                        Turn on this flag to use a scheduler. (default: False)
```

Let's look at an example of how p_batch can be used. First, create a directory, and fill it with protein files. For this example, assume a directory called "source_files" containing a couple PDB files.

```
user@example:/pro_batch_testing$ ls source_files
1ans.pdb  4pti.pdb
```

Now, let's use p_batch.

```
p_batch source_files

New book.txt created.  You can remove protein files to be run by editing book.txt if desired,
and resume by running p_batch again.

These proteins will be run:

4pti
1ans
Pre-existing directories for these proteins will be emptied and replaced with information from
the new run.
Run MCCE with the current settings? (yes/y/no)
```

Typing "yes" or "y" will start the default MCCE process in each directory (unless the -custom flag is used to choose an alternate script).

```
Processing source_files/4pti.pdb...
Processing source_files/1ans.pdb...

Bash script is being executed in each directory. Double check processes are running with
command 'top', or 'mcce_stat'. mcce_stat also updates book.txt to reflect completed runs.
```

mcce_stat

Reviewing multiple protein runs can be cumbersome. To aid the user, "mcce_stat" is included. The directories created by p_batch or pro_batch will contain the output files of each MCCE run. mcce_stat checks of these directories for "signal" files to check how each run is progressing, as of mcce_stat's runtime.

```
mcce_stat

To see when PDB was run, reference mcce_timing.log in running directory. (pro_batch)

Completion: 7.50%
Directory step1         step2         step3     step4  Status
1ans      Exists        Exists
4pti      Exists
```

When all four steps are completed, the pKas of the selected protein will be available to see in "pK.out".

book.txt

Before beginning the MCCE run, the program creates a file named "book.txt". book.txt is a text file containing a list names of proteins that can be edited to "turn off" proteins. This is what our new book.txt file looks like:

```
4pti
1ans
```

If we wanted to turn off both of these, we could edit book.txt to:

```
4pti   x
1ans   c
```

Where "x" means exclusion, and "c" means completed. Attempting to run p_batch now results in the following message:

```
p_batch source_files/

book.txt found! Protein files identified in book.txt:

4pti   x
1ans   c

No runnable proteins found. Remove 'c' or 'x' from lines in book.txt to make them runnable,
add protein files to run, or delete book.txt to re-do all runs.

Aborting MCCE...
```

In this way, the user can choose any desired subset of proteins from the source directory's protein set.

Revision #3
Created 12 August 2025 18:27:50 by Jared
Updated 12 August 2025 19:34:40 by Jared