

1. MHC-peptide-TCR modelling

1.1 使用说明

```
1  Usage: Pipeline_modeller_update2021.pl -c <mhc class>
2                                     -a <mhc alpha chain>
3                                     -b <mhc beta chain>
4                                     -p <peptide>
5                                     [-A] <tcr alpha chain>
6                                     [-B] <tcr beta chain>
7                                     -u <userDir>
8                                     -n <model number>
9  DESCRIPTION:
10     -c  the input MHC class type must be "MHC-I" or "MHC-II"!
11     -a  input the sequence file (.fasta) or the allele name (eg:HLA-A*01:01,
12         HLA-DPA1*01:03) of mhc alpha chain.
13     -b  input the sequence file (.fasta) or the allele name (eg:b2m, HLA-
14         DPB1*01:01) of mhc beta chain.
15     -p  input the sequence file (.fasta) of peptide.
16     [-A] input the sequence file (.fasta) of tcr alpha chain.
17     [-B] input the sequence file (.fasta) of tcr beta chain.
18     -u  input the output direction.
19     -n  input the output model number.
```

1. 功能概述: 根据输入的MHC、peptide以及TCR的序列构建复合体结构模型, 结构建模方法基于modeller 9.25。

2. 输入项描述:

-c: 输入的MHC序列所属类型, 包括MHC class I和MHC class II两类可选, 输入值必须为 **MHC-I** 或 **MHC-II** ;

-a: MHC分子的alpha chain sequence, 可以是包含序列信息的fasta文件, 也可以是alpha chain的等位基因名称, 输入的基因名称格式如 `HLA-A*01:01`, `HLA-DPA1*01:03`。程序目前可识别的基因名称见下述两个文件:

① `./program/getTopAlleles/Top_MHCI_alleles.csv` 文件中存放的为HLA class I每个基因座上的Top10等位基因, 由 `./program/getTopAlleles/getTopMHCIAAlleles.py` 生成。

② `./program/getTopAlleles/Top_MHCII_alleles.csv` 文件中存放的为HLA class II每个基因座上的Top10等位基因, 由 `./program/getTopAlleles/getTopMHCIIAlleles.py` 生成。

说明: 也可以将目前已知的所有等位基因都列举出来供选择。

-b: MHC分子的beta chain sequence, 可以是包含序列信息的fasta文件, 也可以是beta chain的等位基因名称, 输入的基因名称格式如 `b2m`, `HLA-DPB1*01:01`。程序目前可识别的基因名称见下述两个文件:

① `./program/getTopAlleles/b2m.csv` 文件中存放的为HLA class I中的Beta-2-microglobulin。

② `./program/getTopAlleles/Top_MHCII_alleles.csv` 文件中存放的为HLA class II每个基因座上的Top10等位基因, 由 `./program/getTopAlleles/getTopMHCIIAlleles.py` 生成。

-p: 多肽的序列信息, 输入为包含序列信息的fasta文件。

-A: 可选, TCR的alpha chain sequence, 输入为包含序列信息的fasta文件。

-B: 可选, TCR的beta chain sequence, 输入为包含序列信息的fasta文件。

-u: 用户自定义的输出文件夹。

-n: 输出多少个模型。

3. 待解决问题:

TCR目前还没有实现利用基因名称作为输入项。

4. 运行示例 (输出位于 `model_build` 文件夹中):

输入基因名称

```
1 ./scripts/Pipeline_modeller_update2021.pl -c MHC-I -a HLA-A*01:01 -b b2m -p
  ./example/pep_example_sequence.fasta -u MHCI_test -n 1
```

```
1 ./scripts/Pipeline_modeller_update2021.pl -c MHC-II -a HLA-DRA*01:01 -b HLA-  
DRB1*01:01 -p ./example/pep_example_sequence.fasta -u MHCII_test -n 1
```

输入序列文件

```
1 ./scripts/Pipeline_modeller_update2021.pl -c MHC-I -a  
./example/mhcα_example_sequence.fasta -b ./example/mhcβ_example_sequence.fasta -p  
./example/pep_example_sequence.fasta -u MHCI_test_1 -n 1
```

添加TCR序列

```
1 ./scripts/Pipeline_modeller_update2021.pl -c MHC-I -a  
./example/mhcα_example_sequence.fasta -b ./example/mhcβ_example_sequence.fasta -p  
./example/pep_example_sequence.fasta -A ./example/tcrα_example_sequence.fasta -B  
./example/tcrβ_example_sequence.fasta -u MHCI_test_2 -n 1
```