



HAL
open science

Rapid Peptides Generator: fast and efficient in silico protein digestion. RPG: a fast and efficient software to predict cleaving site of common and novel enzymes

Nicolas Maillet

► To cite this version:

Nicolas Maillet. Rapid Peptides Generator: fast and efficient in silico protein digestion. RPG: a fast and efficient software to predict cleaving site of common and novel enzymes. Jobim - Journées Ouvertes Biologie, Informatique et Mathématiques, Jul 2019, Nantes, France. hal-02612502

HAL Id: hal-02612502

<https://hal.science/hal-02612502>

Submitted on 19 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rapid Peptides Generator: fast and efficient in silico protein digestion

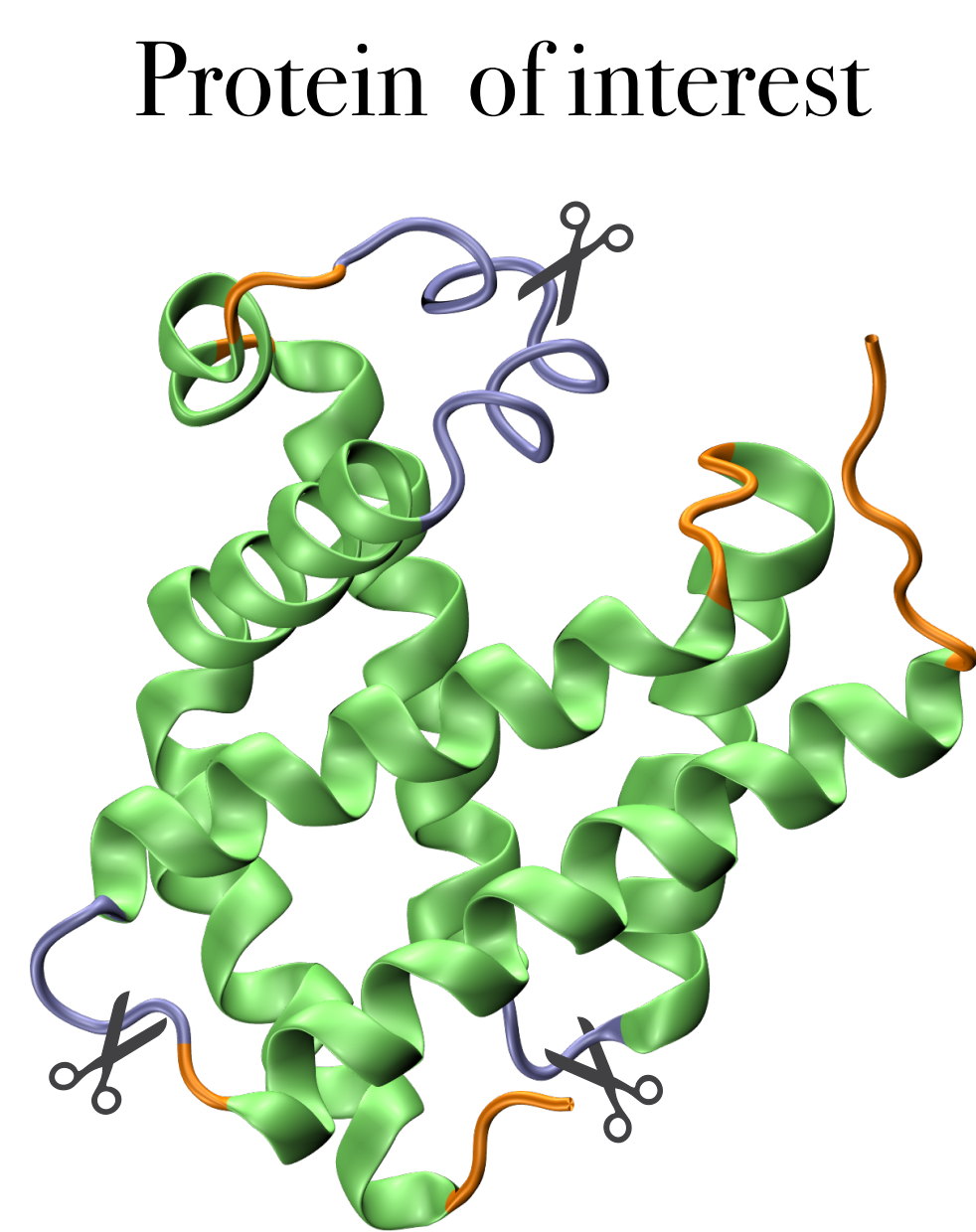
RPG: a fast and efficient software to predict cleaving site of common and novel enzymes

Nicolas Maillet – nicolas.maillet@pasteur.fr

Hub de Bioinformatique et Biostatistique – Département Biologie Computationnelle, Institut Pasteur, USR 3756 CNRS, Paris, France

Problem definition

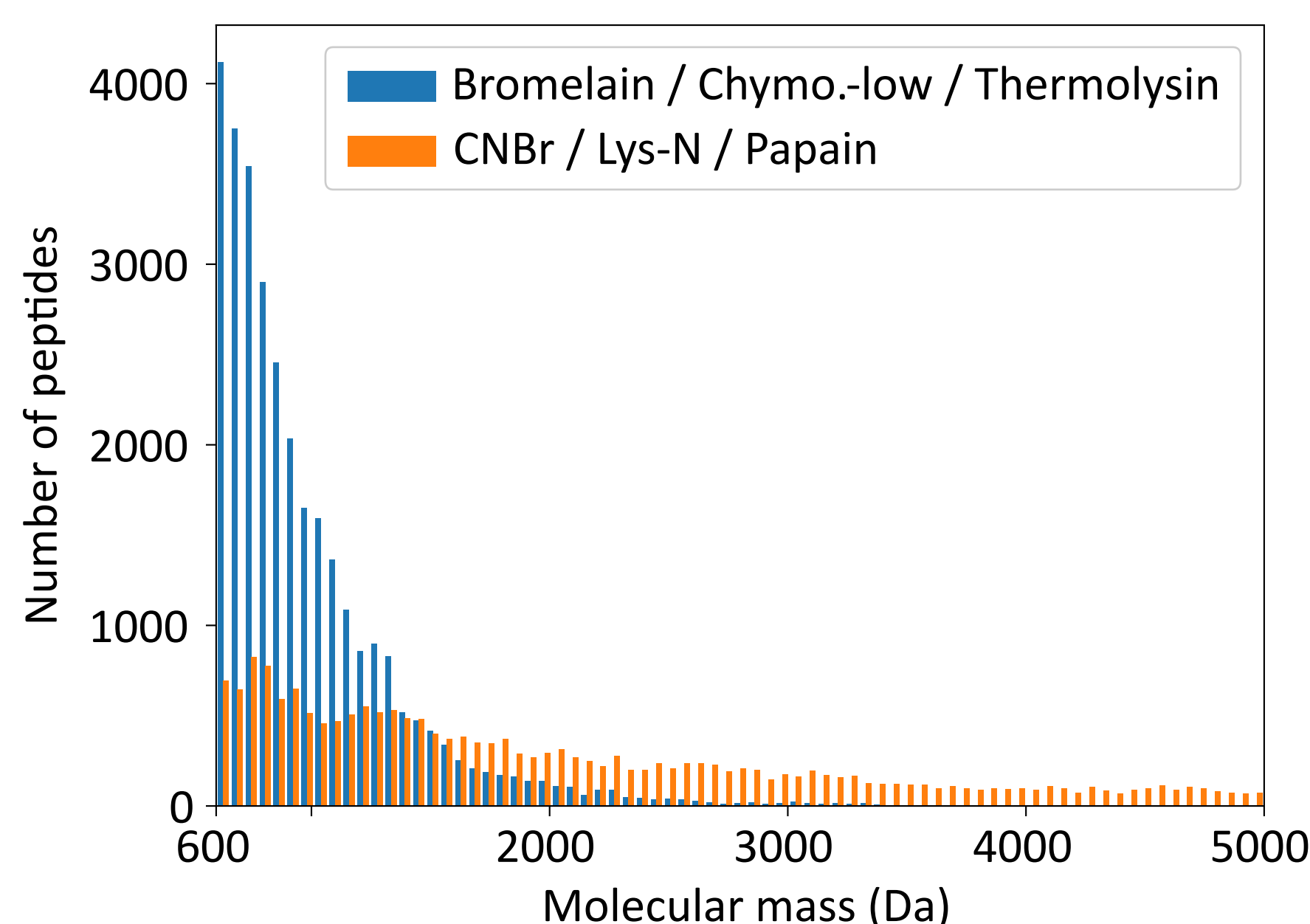
Proteolytic enzymes have been studied for more than 80 years (Neurath et al., 1999). These enzymes are widely used in industry, medicine and proteomics (Nesvizhskii et al., 2014). Yet, it is not straightforward to predict which protease combination will yield a suitable set of peptides for MS/MS.



Proteolytic digestion with several enzymes

- Set A:
- Bromelain
 - Chymotrypsin-low
 - Thermolysin
- Set B:
- CNBr
 - Lys-N
 - Papain

Result of digestion



Number of generated peptides by enzymes set A and enzymes set B in the optimal peptide size range (600–5,000Da) for Globin protein family.

Peptides example

Running RPG

Rapid Peptides Generator is a new standalone software dedicated to predict proteases-induced cleaving sites on sequences. It is primarily designed for MS/MS studies, where multiple digestions can increase the overall confidence in protein identification. Digesting 1000 proteins (V domains of immunoglobulines, mean length of 97.056 aa) with the original 42 enzymes takes 19.96 s (less than 0.5s by enzyme).

Input sequences

`rpg --input`

```
>seq1
AGGGTAFFA
>seq2
PWETGLHF
>seq3
TFLFDCSSCSGO
Fasta/Fastq
```

Enzymes selection

`rpg --list`
`rpg --enzyme 1 40 42`

- 1: Arg-C
- 2: Asp-N
- 3: BNPS-Skatole
- 4: Bromelain
- ...
- 42: Trypsin

Options

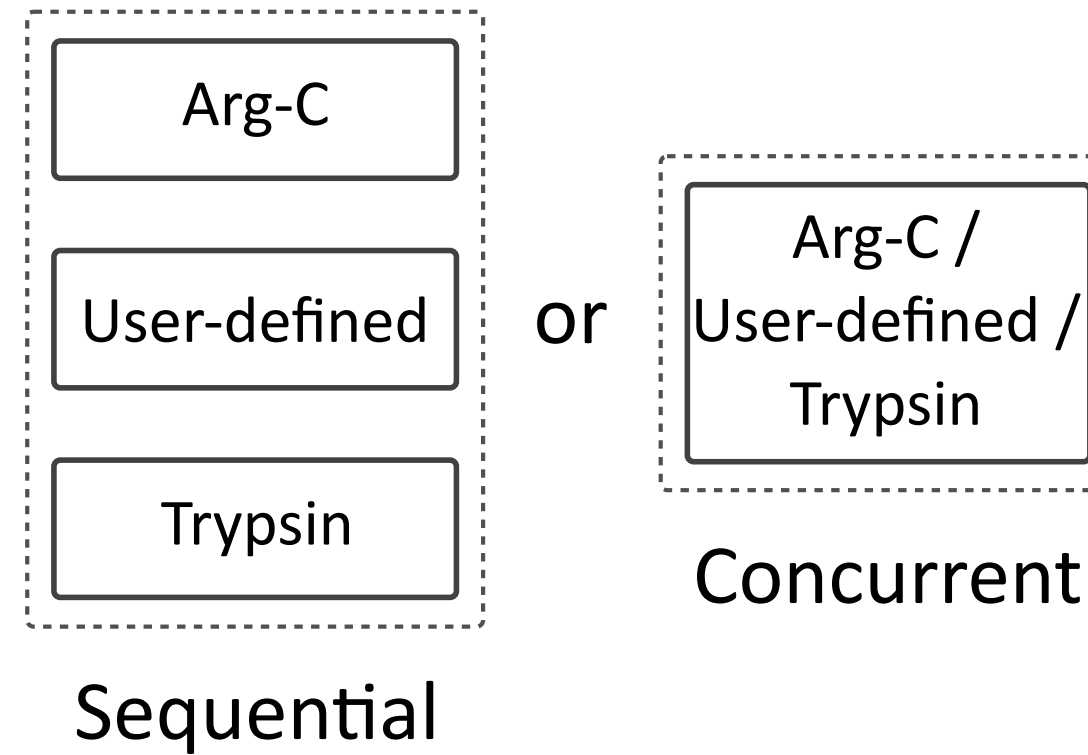
Miscleavage ratio

`--miscleavage 2 -m 18`



Digestion mode

`--digest s or c`



Output

Creation of new enzyme

The grammar of RPG is based on basic logical operators, i. e. "and", "or" and "not". This grammar allows designing complex enzymes like trypsin or thrombin, including many exceptions and different cleaving sites.

User-defined enzymes are then interpreted by RPG and included in the local installation of the software, as a Python object. It is then interpreted as a regular enzyme by RPG and can be used in combination with other enzymes.

Definition of rules

`rpg --addenzyme`

Cleave after **C** or **P**:

(C,)
(P,)

In a more synthetic way:

(C or P,)

Example of logical "and":

(C or P,)(L)

Definition of exceptions

Do not cleave after **C** if preceded by **A** two amino acids before:
(A)() (C,)

Do not cleave after **P** if preceded by **W** and followed by **G** and **G**:
(W) (P,)(G) (G)

Python-object view

```
index=0
letter=C
cut=True
pos=1
    index=-2
    letter=A
    cut=False
index=0
letter=P
cut=True
pos=1
    index=-1
    letter=W
    cut=True
        index=1
        letter=G
        cut=True
            index=2
            letter=G
            cut=False
```

Available enzymes in RPG

- `rpg --list`
- 1: Arg-C
 - 2: Asp-N
 - 3: BNPS-Skatole
 - 4: Bromelain
 - ...
 - 42: Trypsin
 - 43: NewEnzyme

Automatic on-line documentation with Sphinx:
<https://rapid-peptide-generator.readthedocs.io>

Open source, continuous integration:
<https://gitlab.pasteur.fr/nmaillet/rpg>

Available through pip:
<https://pypi.org/project/rpg/>



Availability

There is a growing interest in proteases and new or more specific enzymes (denoted as "Sequencing Grade", or SG) are developed. Depending on the company manufacturing those SG enzymes, specificity and definition can change. Hence, it is important for a user to easily adapt the software by including novel definitions of enzymes.

Rapid Peptides Generator was designed to be highly modular, both for users and developers. RPG is mainly composed of a documented API and a basic launcher.