eNergiome and N-TerPred: novel tools to improve the prediction of plastidic and mitochondrial mature N-termini,

W Bienvenut, P-A Charbit, J-P Scarpelli, T. Meinnel, C. Giglione

To cite this version:

W Bienvenut, P-A Charbit, J-P Scarpelli, T. Meinnel, C. Giglione. eNergiome and N-TerPred: novel tools to improve the prediction of plastidic and mitochondrial mature N-termini,. The FEBS Congress 2018, Jul 2018, Prague, Czech Republic. hal-02381731

HAL Id: hal-02381731
https://hal.archives-ouvertes.fr/hal-02381731
Submitted on 26 Nov 2019

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.
**eNergiome and N-TerPred:**

novel tools to improve the prediction of plastidic and mitochondrial mature N-termini,

**W V Bienvenut, P-A Charbit, J-P Scarpelli, T Meinnel, C Giglione,**

Institute for Integrative Biology of the Cell, Paris-Saclay University, France

---

**Project overview**

Protein N-terminal maturation is essential for protein activity, sub-cellular location and half-life...

Huge interest to know the exact status and position of proteins’ N-termini but...

**Main issues:**

- Experimental data are not available for all proteins...
- Prediction tools are available but not always reliable...

**Objectives:**

- Collecting experimental data : the eNergiomeDB.
- Data validation: based on data redundancy and manual validation
- Datamining for transit peptide cleavage site prediction.

**Deliverable:**

- N-TerPred tool suite:
  - protein mature N-termini and N-terminal transit peptide prediction tool

---

**Sample preparation and data processing [1-2]**

- **eNergiomeDB overview**
  - More than 10000 proteins/entries:
    - 3500 N: salsola, 6000 A. thaliana, 700 S. lycoopersicum, 500 E. coli
  - More than 8500 distinct N-termini:
  - 4500 proteins (N-terminus (Pos 1-2) from which 2761/1274 N/M mitochondrial proteins and 1076 plastidic proteins
  - 4000 downstream mature N-termini including: 280/400 Mitochondrial N-term: (x: At) 230 plastidic (At: N-term)
  - 1700 N-termini used for prediction tool training:
    - 1230 at protein N-term (Pos 1&2)
    - 4500 at protein N-term (Pos 1-2)
  - 2300 N-term quantified for Acetylation yield:
    - 1230 at protein N-term (Pos 1&2)
  - 1144 downstream of the protein N-term (Pos >2)

**Data reprocessing:**

- Data collected from repositories e.g., PXD0002069 & PXD0002690 [3-5]

**EnCOUNTer xml export file**

- Literature data
  - (manual annotation)

---

**Plastidic localisation prediction**

Randomized dataset using 1400 proteins per training round (700 plastid/700 cytoplasmic candidates)

70% Train
30% Test

**Transit Peptide length Prediction Results**

2 levels of confidence defined: Top-Pred & Extended *

<table>
<thead>
<tr>
<th>N-TerPred</th>
<th>ChloroP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>N-termini data</td>
<td>Hits</td>
</tr>
<tr>
<td>Total hits</td>
<td>371</td>
</tr>
<tr>
<td>Top Pred</td>
<td>315</td>
</tr>
<tr>
<td>True</td>
<td>56</td>
</tr>
<tr>
<td>False</td>
<td>338</td>
</tr>
<tr>
<td>Extended Pred</td>
<td>58</td>
</tr>
</tbody>
</table>

---

**Conclusion:**

- **eNergiomeDB:**
  - Manually Curated N-terminome data

**N-TerPred toolbox:**

- A powerful prediction suite for
  - Subcellular localization
  - Transit Peptide length (Mitochondrial/Plastidic)
  - N-term Met excision
  - N-term Acetylation

**Better N-TerPred reliability**

for transit peptide prediction vs. TargetP

cTP prediction: 91% vs. 63%
Subcellular localisation: 95% vs. 91%

**Combined predictions:**

88% vs. 75%

---

**References:**

3. SilProNAQ/enCOUNTer applications:
   - J Lindser et al. Nat Commun. 2015, 6: 7640

---