*Bioinformatics*, YYYY, 0–0

doi: 10.1093/bioinformatics/xxxxx

Advance Access Publication Date: DD Month YYYY

Manuscript Category

| Build of a model for the detection of signal peptides in proteins  Gianluca Piccolo1  1International Master in Bioinformatics, Alma Mater Studiorum - University of Bologna, Bologna  Email: gianluca.piccolo6@studio.unibo.it  **Abstract**  **Motivation:** Signal peptides are among the most common sorting signals, as they target newly syn-thesized proteins to the secretory pathway. The identification of signal peptides in protein sequences is critical to elucidate protein localization and function. Constructing a model to forecast the occurrence of signal peptides in proteins could offer fresh perspectives on the roles and interactions of proteins with limited experimental data, while potentially uncovering novel drug targets. This study presents two distinct models aimed at this objective: the initial approach utilizes a position-specific weight matrix (Von - Heijne method), while the second employs Support Vector Machine as a machine learning technique.  **Results:** The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.  **Supplementary information:** Supplementary data are available at *Bioinformatics* online. |
| --- |

# Introduction

Cells continuously produce proteins with diverse functions. These newly formed proteins must be directed either to various organelles within the cell or transported out of it. To guide this process, nascent proteins possess an inherent Signal Peptide that acts as an "address label" (Chou, 2001). During or after protein translocation, a signal peptidase cleaves the signal peptide at a specific site (Almagro Armenteros et al. 2019).

Secretory signal sequences typically are short sequences of 16 to 30 residues, comprise three structurally and potentially functionally distinct regions: a basic N-terminal region (N-region), a central hydrophobic region (H-region), and a more polar C-terminal region (C-region). The structural elements governing signal sequence cleavage appear to be located in the N- and H-regions, particularly at positions -3 and -1 relative to the cleavage site (Von Heijne, 1986). Despite sharing common characteristics, the exact sequences vary considerably among proteins.

The importance of SPs cannot be overstated: they are relevant in the production of recombinant proteins (Mergulhão *et al.*, 2005), a large number of human diseases is caused by mutations in the SPs (Jarjanazi *et al.*, 2007), they also are interesting targets of drugs (Vermeire *et al.*, 2014) and can be exploited as diagnostic biomarkers for several diseases (Dirican *et al.*, 2016).

Given the rapidly expanding number of protein sequences in databases, developing a swift and precise algorithm to identify signal sequences and predict their cleavage sites has become increasingly important (Cai, Lin, and Chou 2003). In the Gene Ontology (GO) framework, the cellular component is one of three aspects describing protein function, alongside biological process and molecular function (Carbon et al. 2021; Ashburner et al. 2000).

Understanding protein localization is crucial for identifying potential protein interactions and surface-exposed targets in drug discovery. To address this need, numerous algorithms have been developed, with SignalP being the pioneer among publicly available methods, now in its sixth iteration. The latest version can detect various types of signal peptides by leveraging a protein language model to represent the motif (Teufel et al. 2022).

In this study, we conducted a similar analysis using two distinct machine learning approaches, both trained on the same dataset. The first method employs a position-specific weight matrix, following the approach of von Heijne (1986). The second utilizes a support vector machine, inspired by the work of Cai, Lin, and Chou (2003).

These complementary approaches aim to provide insights into signal peptide prediction, offering alternative perspectives to the established SignalP algorithm. By comparing these methods, we seek to enhance our understanding of protein targeting mechanisms and contribute to the ongoing development of tools for protein localization prediction.

# Methods

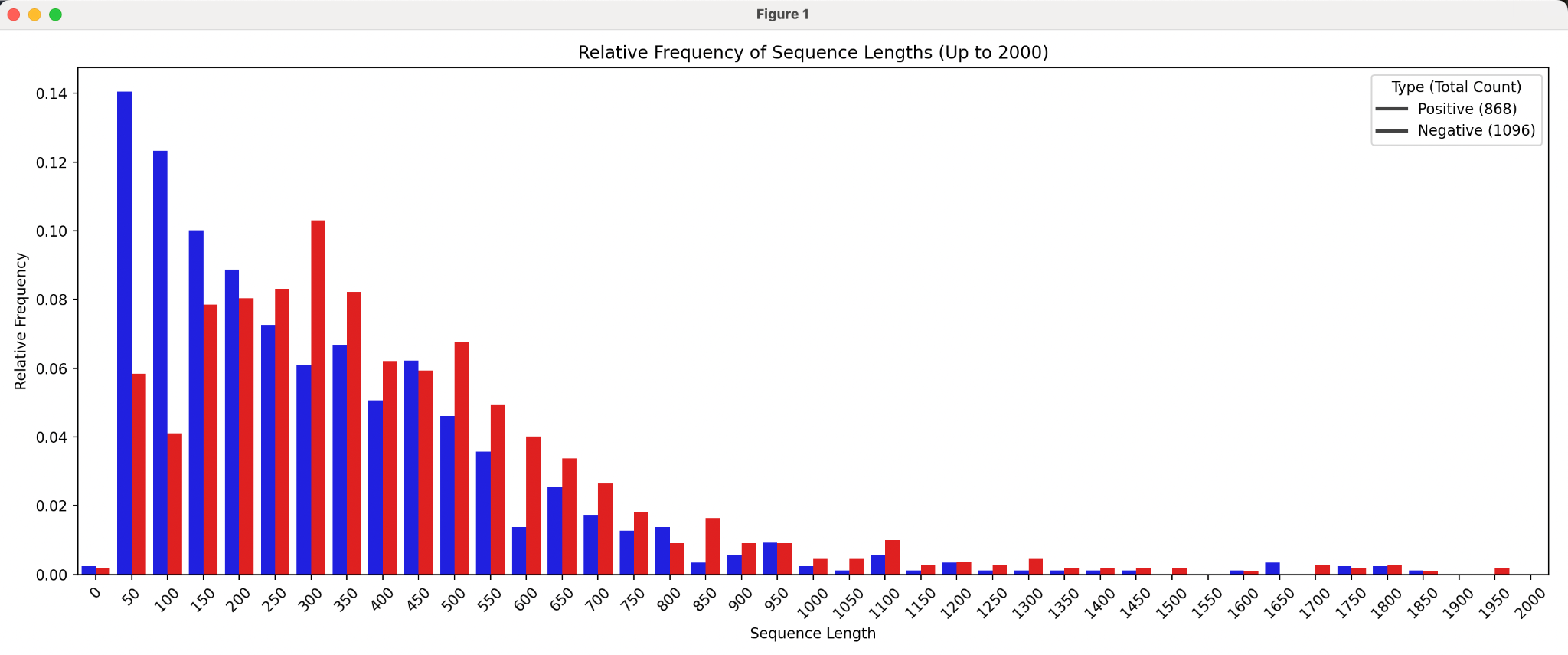
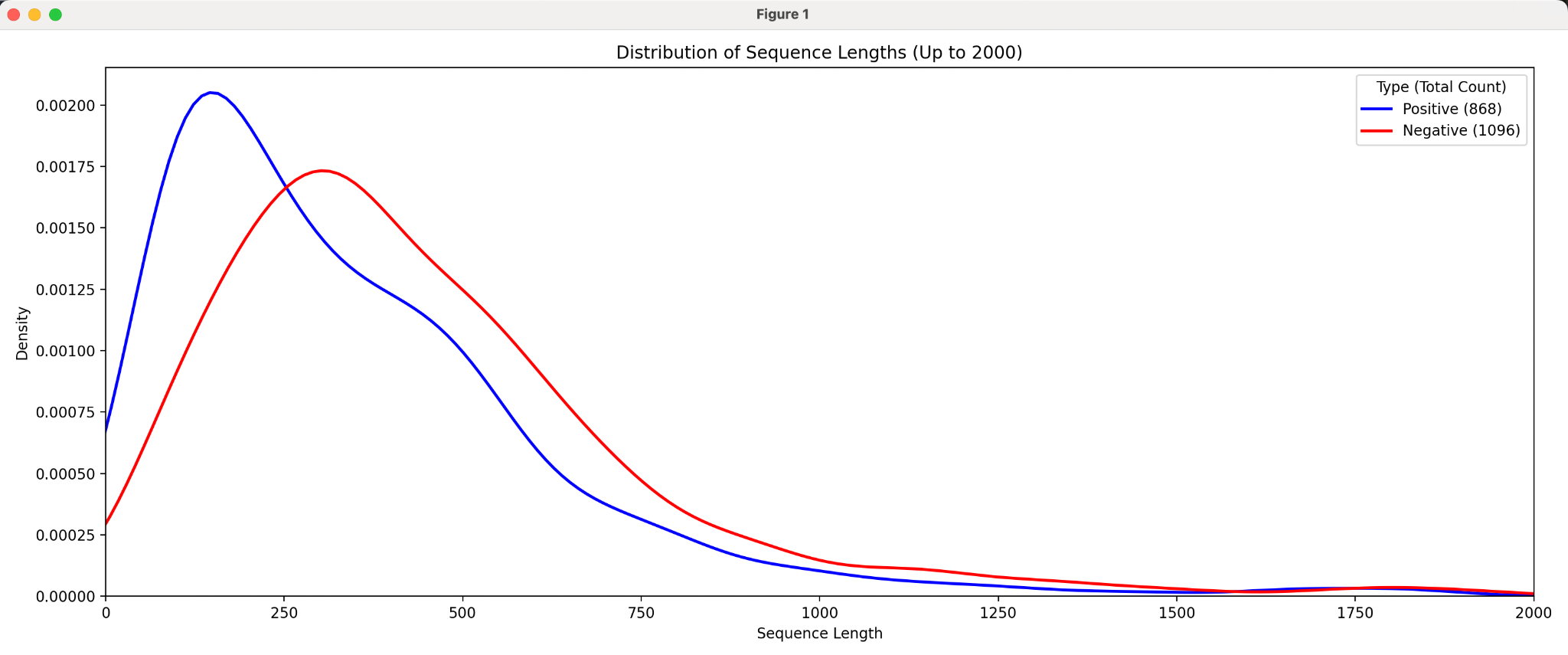
The data used in these approaches was fetched using the UniProt Knowledgebase release 2024\_05 API in Python. The dataset included only reviewed entries (UniprotKB) and SPs with experimental evidence for the cleavage site. Furthermore, proteins smaller than 40 residues or fragments were excluded.

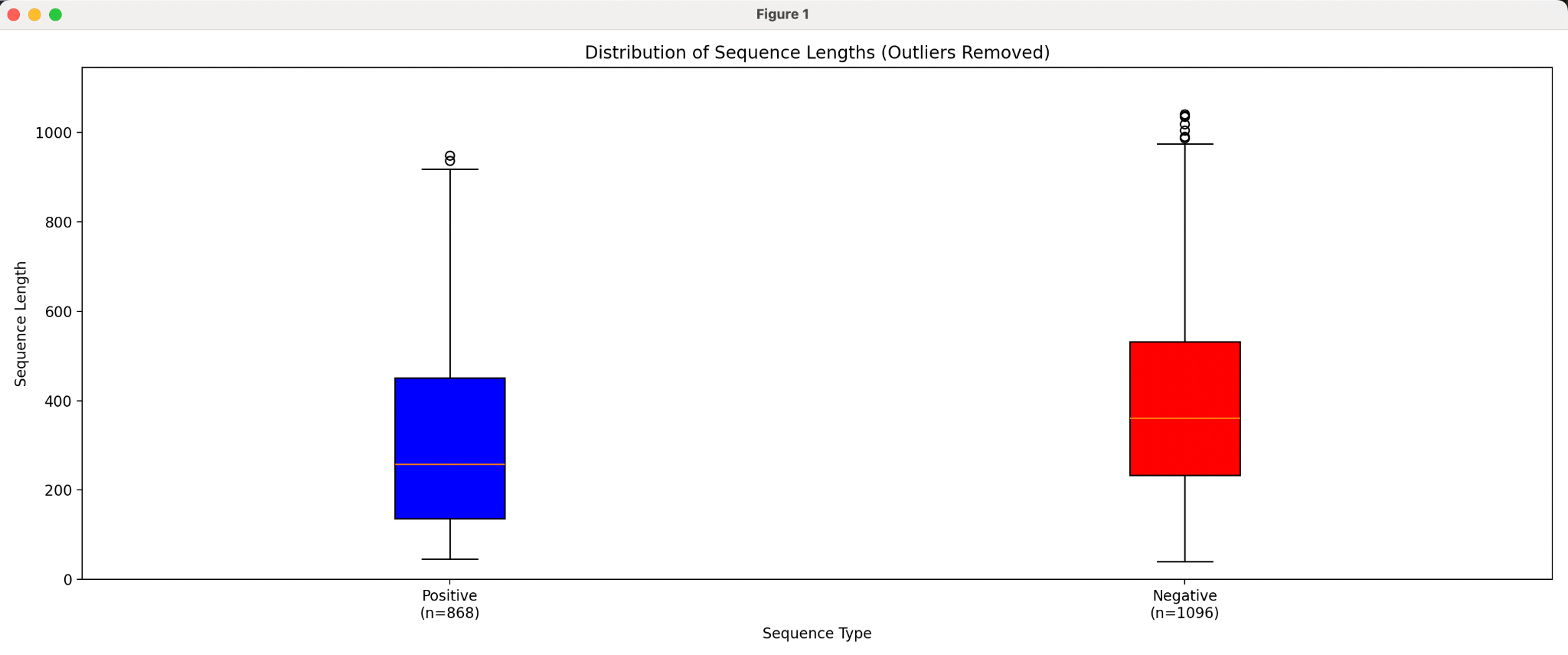
**2.1 Training dataset**

The training dataset consists of 1087 positive eukaryotic sequences (sequences with N-terminal secretory SPs) and 1371 negative eukaryotic sequences (proteins with a subcellular location annotated as cytosolic, nuclear, mitochondrial, plastid, and/or peroxisomal in Eukarya and not belonging to the secretory pathway with experimental evidence).

A data exploratory analysis on the training positive examples has been conducted, showing that the distribution is comparable to the benchmark SP length distribution (see 2.3 benchmark dataset).

Both methods that will be implemented will exploit the conservation of the average length of SPs in the training and benchmark sets.





# Results

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

## Data Structure This is Heading 2 style this is heading 2 style

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

### 3.1.1 This is heading 3 style

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

1. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
2. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
3. The quick brown fox jumps over the lazy dog.
4. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
5. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
6. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

## Unnumbered list style

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

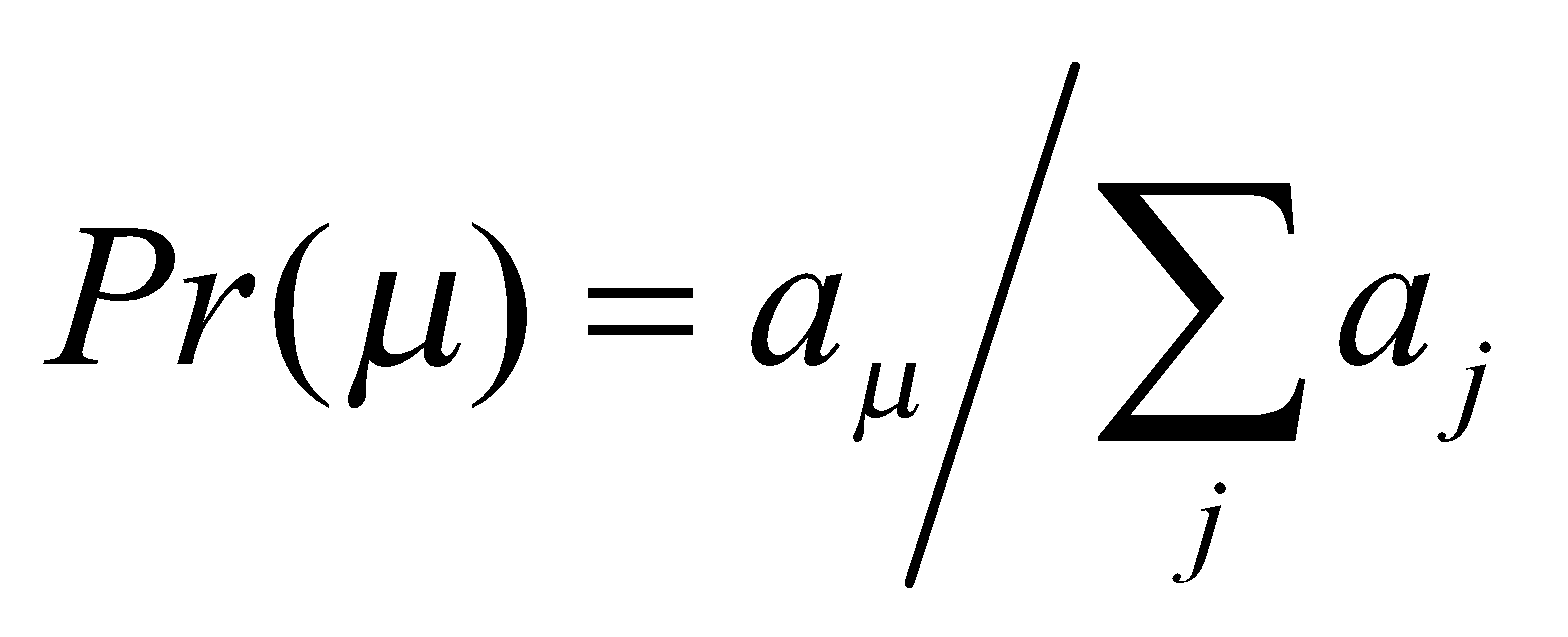
The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

 (1)

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

**Fig. 1. Relation between τ and *t*.** This example has only two continuous Steppers, S1 and S2.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

**Table 1.**Benchmark results of the cascade oscillators model

| |S| | Predicted cost | Timing | Predicted speed | Speed |
| --- | --- | --- | --- | --- |
| 1 | S219.20(100%) | 68m43s | 1.00 | 1.00 |
| 2 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 4 | 219.20(100%) | 68m43s | 1.00 | 1.00 |
| 10 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 20 | 219.20(100%) | 68m43s | 1.00 | 9.5 |

This is table foot note sample text This is table foot note sample text This is table foot note sample text

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.The quick brown fox jumps over the lazy dog.

**Acknowledgements**

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

**Funding**

This work has been supported by the …..

*Conflict of Interest:* none declared.

**References**

Alexandrescu,A. (2001) Modern C++ Design: Generic Programming and Design Patterens Applied. Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Alexandrescu,A. (2001) *Modern C++ Design: Generic Programming and Design Patterens Applied.* Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Alexandrescu,A. (2001) *Modern C++ Design: Generic Programming and Design Patterens Applied.* Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Yoo,M.S. *et al*. (2003) Oxidative stress regulated genes in nigral dopaminergic neurnol cell: correlation with the known pathology in Parkinson’s disease. *Brain Res. Mol. Brain Res.*, **110**(Suppl. 1), 76–84.

Crenshaw, B.,III, and Jones, W.B.,Jr (2003) The future of clinical cancer management: one tumor, one chip. *Bioinformatics*, doi:10.1093/bioinformatics/btn000.

Auhtor,A.B. *et al.* (2000) Chapter title. In Smith, A.C. (ed.), *Book Title*, 2nd edn. Publisher, Location, Vol. 1, pp. ???–???.

Bardet, G. (1920) Sur un syndrome d’obesite infantile avec polydactylie et retinite pigmentaire (contribution a l’etude des formes cliniques de l’obesite hypophysaire). PhD Thesis, name of institution, Paris, France.