

---

# Accuracy of secondary protein structure prediction tools for chromoproteins and fluorescent proteins

---

Anonymous Author(s)  
Affiliation  
Address  
email

## Abstract

6 abstract...

## 7 1 Introduction

### 8 1.1 Secondary protein structure prediction

9 This project aims to look at methods to predict secondary protein structure. Protein structure  
10 prediction is a major field of study and is a problem that takes massive computational power to solve.  
11 There are two main approaches looking from a biochemical point of view. The first is isolated the  
12 protein, crystallizing it, and performing crystal chromatography to figure out the structure. This  
13 structure is relaxed into its hypothesized structure via molecular dynamics. I have previously done  
14 work on molecular dynamic methods, but now I would like to look at it from the other direction. One  
15 of the most plentiful and easy to obtain biological data is DNA sequence. From the DNA sequence of  
16 a coding region there are simply rules to propose a great starting point for the protein's amino acid  
17 sequence. The problem of predicting the 3D structure from an amino acid sequence is extremely hard.  
18 I will reduce this problem to simpler features. My aim is to look at how we can use the amino acid  
19 sequence, the primary structure, to deduce secondary structure components like beta sheets, alpha  
20 helices, and coils.

### 21 1.2 Fluorescent proteins and chromoproteins

22 The style files for NIPS and other conference information are available on the World Wide Web at

23 <http://www.nips.cc/>

24 The file `nips_2016.pdf` contains these instructions and illustrates the various formatting require-  
25 ments your NIPS paper must satisfy.

26 The only supported style file for NIPS 2016 is `nips_2016.sty`, rewritten for  $\text{\LaTeX 2}_{\epsilon}$ . **Previous**  
27 **style files for  $\text{\LaTeX 2.09}$ , Microsoft Word, and RTF are no longer supported!**

28 The new  $\text{\LaTeX}$  style file contains two optional arguments: `final`, which creates a camera-ready copy,  
29 and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

30 At submission time, please omit the `final` option. This will anonymize your submission and add  
31 line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be  
32 removed during generation of camera-ready copies.

33 The file `nips_2016.tex` may be used as a “shell” for writing your paper. All you have to do is  
34 replace the author, title, abstract, and text of the paper with your own.

35 The formatting instructions contained in these style files are summarized in Sections 2, 3, and 4  
36 below.

## 37 **2 General formatting instructions**

38 The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long.  
39 The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points.  
40 Times New Roman is the preferred typeface throughout, and will be selected for you by default.  
41 Paragraphs are separated by  $\frac{1}{2}$  line space (5.5 points), with no indentation.

42 The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal  
43 rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow  $\frac{1}{4}$  inch  
44 space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the  
45 page.

46 For the final version, authors’ names are set in boldface, and each name is centered above the  
47 corresponding address. The lead author’s name is to be listed first (left-most), and the co-authors’  
48 names (if different address) are set to follow. If there is only one co-author, list both author and  
49 co-author side by side.

50 Please pay special attention to the instructions in Section 4 regarding figures, tables, acknowledgments,  
51 and references.

## 52 **3 Headings: first level**

53 All headings should be lower case (except for first word and proper nouns), flush left, and bold.

54 First-level headings should be in 12-point type.

### 55 **3.1 Headings: second level**

56 Second-level headings should be in 10-point type.

#### 57 **3.1.1 Headings: third level**

58 Third-level headings should be in 10-point type.

59 **Paragraphs** There is also a `\paragraph` command available, which sets the heading in bold, flush  
60 left, and inline with the text, with the heading followed by 1 em of space.

## 61 **4 Citations, figures, tables, references**

62 These instructions apply to everyone.

### 63 **4.1 Citations within the text**

64 The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as  
65 long as you maintain internal consistency. As to the format of the references themselves, any style is  
66 acceptable as long as it is used consistently.

67 The documentation for `natbib` may be found at

68 <http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf>

69 Of note is the command `\citet`, which produces citations appropriate for use in inline text. For  
70 example,

71 `\citet{hasselmo}` investigated\dotso

72 produces

73 Hasselmo, et al. (1995) investigated...

74 If you wish to load the `natbib` package with options, you may add the following before loading the  
75 `nips_2016` package:

76 `\PassOptionsToPackage{options}{natbib}`

77 If `natbib` clashes with another package you load, you can add the optional argument `nonatbib`  
78 when loading the style file:

79 `\usepackage[nonatbib]{nips_2016}`

80 As submission is double blind, refer to your own published work in the third person. That is, use “In  
81 the previous work of Jones et al. [4],” not “In our previous work [4].” If you cite your other papers  
82 that are not widely available (e.g., a journal paper under review), use anonymous author names in the  
83 citation, e.g., an author of the form “A. Anonymous.”

## 84 4.2 Footnotes

85 Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number<sup>1</sup>  
86 in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote  
87 with a horizontal rule of 2 inches (12 picas).

88 Note that footnotes are properly typeset *after* punctuation marks.<sup>2</sup>

## 89 4.3 Figures

90 All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction.  
91 The figure number and caption always appear after the figure. Place one line space before the figure  
92 caption and one line space after the figure. The figure caption should be lower case (except for first  
93 word and proper nouns); figures are numbered consecutively.

94 You may use color figures. However, it is best for the figure captions and the paper body to be legible  
if the paper is printed in either black/white or in color.

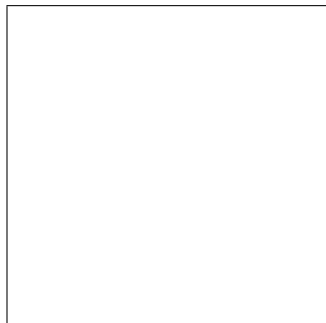


Figure 1: Sample figure caption.

95

---

<sup>1</sup>Sample of the first footnote.

<sup>2</sup>As in this example.

Table 1: Sample table title

Part		
Name	Description	Size ( $\mu\text{m}$ )
Dendrite	Input terminal	$\sim 100$
Axon	Output terminal	$\sim 10$
Soma	Cell body	up to $10^6$

#### 96 4.4 Tables

97 All tables must be centered, neat, clean and legible. The table number and title always appear before  
98 the table. See Table 1.

99 Place one line space before the table title, one line space after the table title, and one line space after  
100 the table. The table title must be lower case (except for first word and proper nouns); tables are  
101 numbered consecutively.

102 Note that publication-quality tables *do not contain vertical rules*. We strongly suggest the use of the  
103 booktabs package, which allows for typesetting high-quality, professional tables:

104 <https://www.ctan.org/pkg/booktabs>

105 This package was used to typeset Table 1.

#### 106 5 Final instructions

107 Do not change any aspects of the formatting parameters in the style files. In particular, do not modify  
108 the width or length of the rectangle the text should fit into, and do not change font sizes (except  
109 perhaps in the **References** section; see below). Please note that pages should be numbered.

#### 110 6 Preparing PDF files

111 Please prepare submission files with paper size “US Letter,” and not, for example, “A4.”

112 Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or  
113 Embedded TrueType fonts. Here are a few instructions to achieve this.

- 114 • You should directly generate PDF files using `pdflatex`.
- 115 • You can check which fonts a PDF files uses. In Acrobat Reader, select the menu  
116 Files>Document Properties>Fonts and select Show All Fonts. You can also use the program  
117 `pdf fonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- 118 • The IEEE has recommendations for generating PDF files whose fonts are also ac-  
119 ceptable for NIPS. Please see [http://www.emfield.org/icuwb2010/downloads/](http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf)  
120 `IEEE-PDF-SpecV32.pdf`
- 121 • `xfig` “patterned” shapes are implemented with bitmap fonts. Use “solid” shapes instead.
- 122 • The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS  
123 Fonts:

124 `\usepackage{amsfonts}`

125 followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for  $\mathbb{R}$ ,  $\mathbb{N}$  or  $\mathbb{C}$ . You can also  
126 use the following workaround for reals, natural and complex:

```
127 \newcommand{\RR}{\mathbb{R}} %real numbers
128 \newcommand{\Nat}{\mathbb{N}} %natural numbers
129 \newcommand{\CC}{\mathbb{C}} %complex numbers
```

130 Note that `amsfonts` is automatically loaded by the `amssymb` package.

131 If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

## 6.1 Margins in L<sup>A</sup>T<sub>E</sub>X

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...  
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (<http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf>)

A number of width problems arise when L<sup>A</sup>T<sub>E</sub>X cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

## Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

## References

- [1] Touw, W.G. & Baakman, C. & Black, B. & te Beek, T.A.H. & Kreiger, E. & Joosten, R.P. & Vriend, G. (2015) A series of PBD related databases for everyday needs. for connectionist rule extraction. *Nucleic Acids Research*, 43(Database issue): D364-D368.
- [2] Kabsch W. & Sander C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features *Biopolymers*, 983 22 2577-2637. PMID: 6667333; UI: 84128824.
- [3] Cambria, A. (2009) Hidropathy Clustering Assisted Methods. <http://www.acbrc.org/hcam.html>
- [4] Garnier, J. & Gibrat, J.F. & Robson, B. (1996) GOR Method for Predicting Protein Secondary Structure from Amino Acid Sequence. *Method in Enzymology*. 266:540-53.
- [5] Geourjon C. & Deleage G. (1994) SOPM: a self-optimimization mehtod for protein secondary structure prediction. *Protein Engineering*. 7(2):157-64.
- [6] Karypis, G. (2006) YASSPP: Better Kernels and Coding Schemes Lead to Improvements in Protein Secondary Structure Prediction. *Proteins*. 64:575-86.
- [7] Singh, M. (2001) Predicting Protein Secondary and Supersecondary Structure. Princeton University. CRC Press.
- [8] Sormanni, P. & Camilloni, C. & Fariselli, P. & Vendruscolo, M. (2015) The s2D Method: Simultaneous Sequence-Based Prediction of the Statistical Populations of Ordered and Disordered Regions in Proteins. *J. Mol. Biol.* 427: 982-96.
- [9] Sen, T.Z. & Jernigan, R.L. & Garnier, J. & Kloczkowski, A. (2005) GOR V server for protein secondary structure prediction. *Bioinformatics* Jun 1; 21(11): 2787-2788.