Bachelor's Thesis

submitted in partial fulfillment of the
requirements for the course "Applied Computer Science"

My Title

Robin William Hundt

Institute of Computer Science

Bachelor's and Master's Theses
of the Center for Computational Sciences
at the Georg-August-Universität Göttingen

09. May 2020

Georg-August-Universität Göttingen
Institute of Computer Science

Goldschmidtstraße 7
37077 Göttingen
Germany

☎ +49 (551) 39-172000
FAX +49 (551) 39-14403
✉ office@informatik.uni-goettingen.de
🌍 www.informatik.uni-goettingen.de

First Supervisor: Prof. Dr. Burkhard Morgenstern
Second Supervisor: Dr. Peter Meinicke

Abstract

Here comes the abstract...

Contents

Chapter 1

Introduction

Chapter 2

Basics

2.1   Multiple sequence alignment

Chapter 3

Prior Work

Chapter 4

Algorithm

In this chapter, the analysis of ...

Chapter 5

Implementation

In this chapter, the implementation of ...

Chapter 6

Evaluation

6.1 BAliBASE 3

The third version of the BAliBASE benchmark protein alignment database has been released in 2005 and is widely employed for the comparison of multiple alignment programs [1,2].

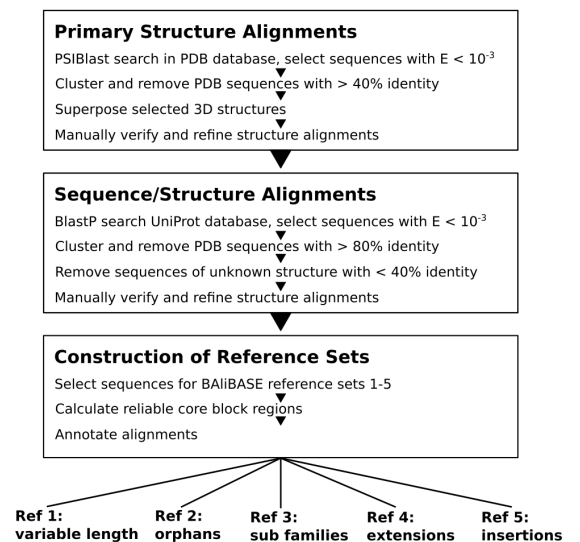## 6.2 Sum-of-pairs and column score

6.3 MAFFT

6.4 Results

Primary Structure Alignments

PSIBlast search in PDB database, select sequences with $E < 10^{-3}$

Cluster and remove PDB sequences with > 40% identity

Superpose selected 3D structures

Manually verify and refine structure alignments

Sequence/Structure Alignments

BlastP search UniProt database, select sequences with $E < 10^{-3}$

Cluster and remove PDB sequences with > 80% identity

Remove sequences of unknown structure with < 40% identity

Manually verify and refine structure alignments

Construction of Reference Sets

Select sequences for BAliBASE reference sets 1-5

Calculate reliable core block regions

Annotate alignments

Ref 1: variable length   **Ref 2:** orphans   **Ref 3:** sub families   **Ref 4:** extensions   **Ref 5:** insertions

Figure 6.1: Flow chart showing the semi automatic process used to establish the reference sets

Chapter 7

Conclusion

7.0.1 Further work

Bibliography

[1] J. D. Thompson, P. Koehl, R. Ripp, and O. Poch, "Balibase 3.0: latest developments of the multiple sequence alignment benchmark," *Proteins: Structure, Function, and Bioinformatics*, vol. 61, no. 1, pp. 127–136, 2005.

[2] D. J. Russell, *Multiple Sequence Alignment Methods -*, softcover reprint of the original 1st ed. 2014 ed.   unbekannt: Humana Press, 2016.