signalHsmm: prediction of malarial signal peptides

Michał Burdukiewicz[1]*, Piotr Sobczyk[2], Paweł Błażej[1], Paweł Mackiewicz[1]

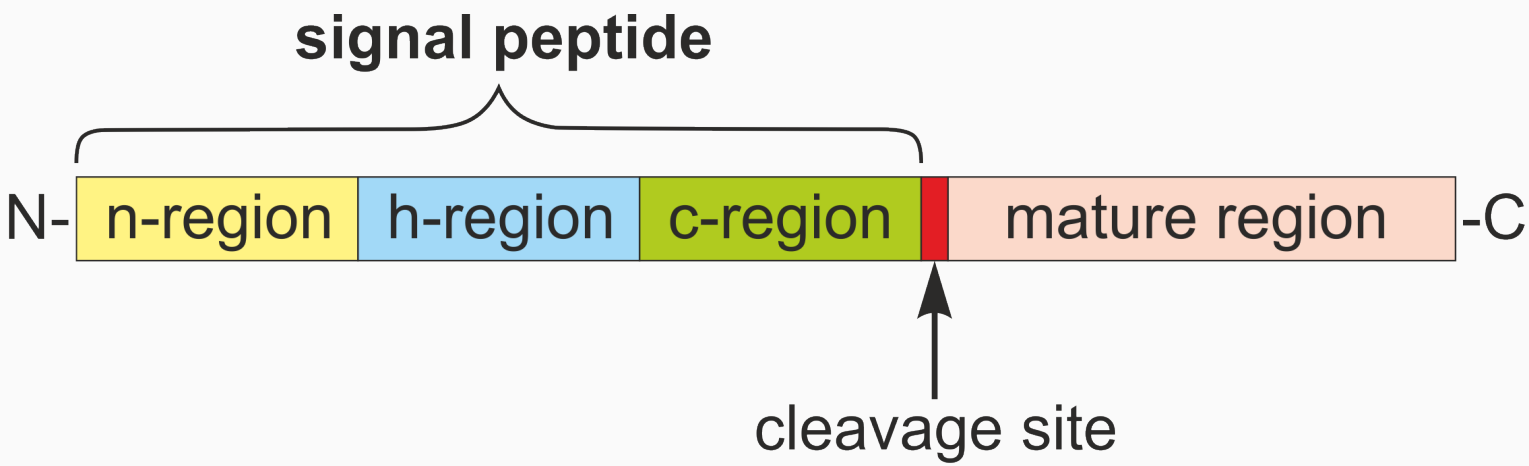
[1]University of Wrocław, Department of Genomics,
[2]Wrocław University of Science and Technology, Faculty of Pure and Applied Mathematics,

Signal peptides

Secretory signal peptides:

- are short (20-30 residues) N-terminal amino acid sequences,
- direct a protein to the endomembrane system and next to the extracellular localization,
- are universal enough to direct properly proteins in different secretory systems; artifically introduced bacterial signal peptides can guide proteins in mammals (Nagano and Masuda, 2014) and plants (Moeller et al., 2009),
- tag hormons, immune system proteins, structural proteins, and metabolic enzymes.

Architecture



Signal peptides possess three distinct domains with variable length and characteristic amino acid composition (Hegde and Bernstein, 2006):

- n-region: mostly basic residues (Nielsen and Krogh, 1998),
- h-region: strongly hydrophobic residues (Nielsen and Krogh, 1998),
- c-region: a few polar, uncharged residues (Jain et al., 1994).

Malarial signal peptides

Apicoplast

Four membrane-bounded plastid responsible for several biochemical pathways including the biosynthesis of fatty acids, isoprenoids, haem and iron–sulfur clusters.

Apicoplast

The signal peptide is required for targeting the apicoplast protein to the endomembrane system, whereas the transit peptide is required to traffic the protein to the apicoplast.

TUTAJ FIG2.

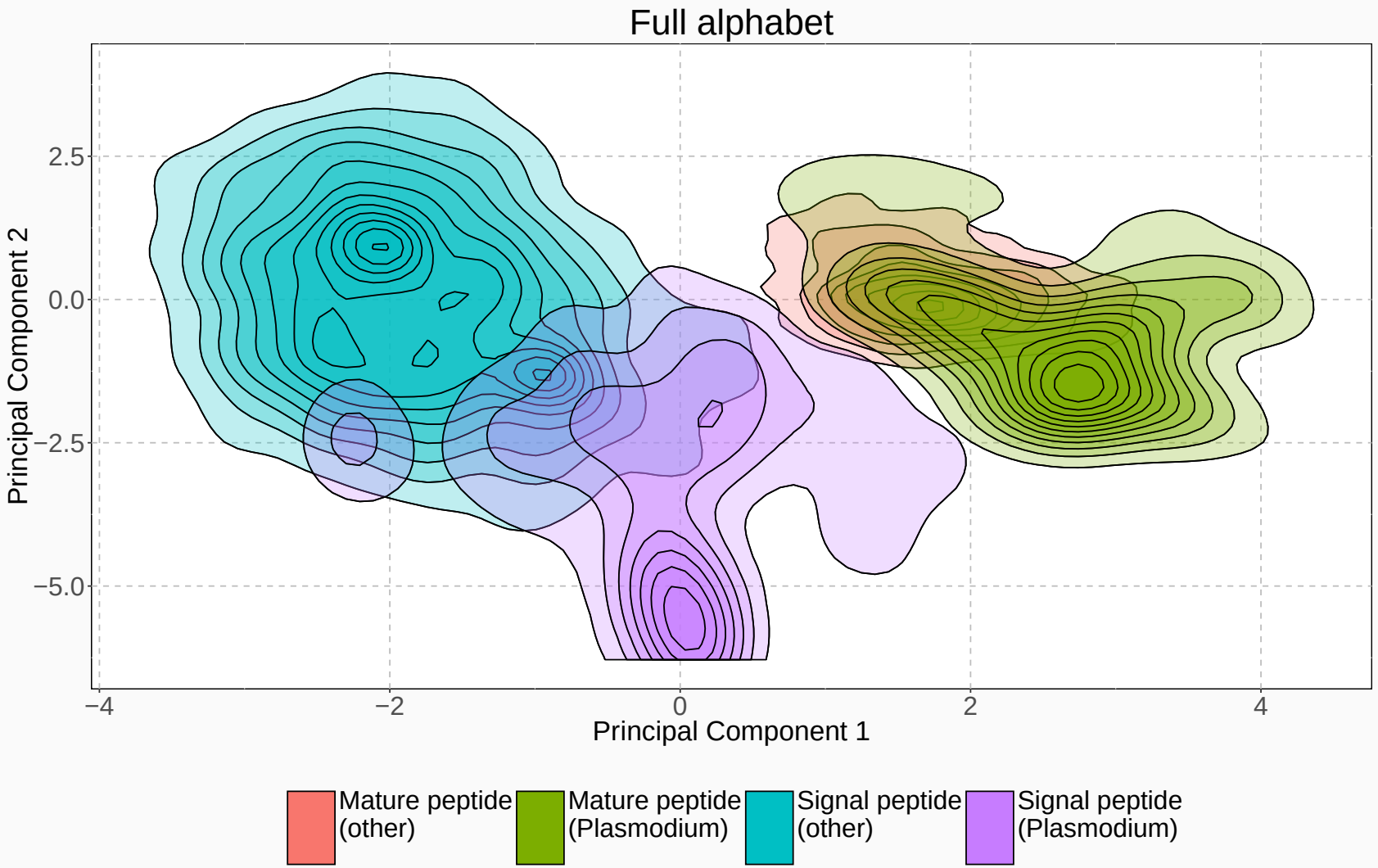
Heavy adenine-thymine bias of parasitic genomes alters amino acid composition of malarial signal peptides.

Even nonbiological sequences can be effective signal peptides provided they fulfill general requirements.

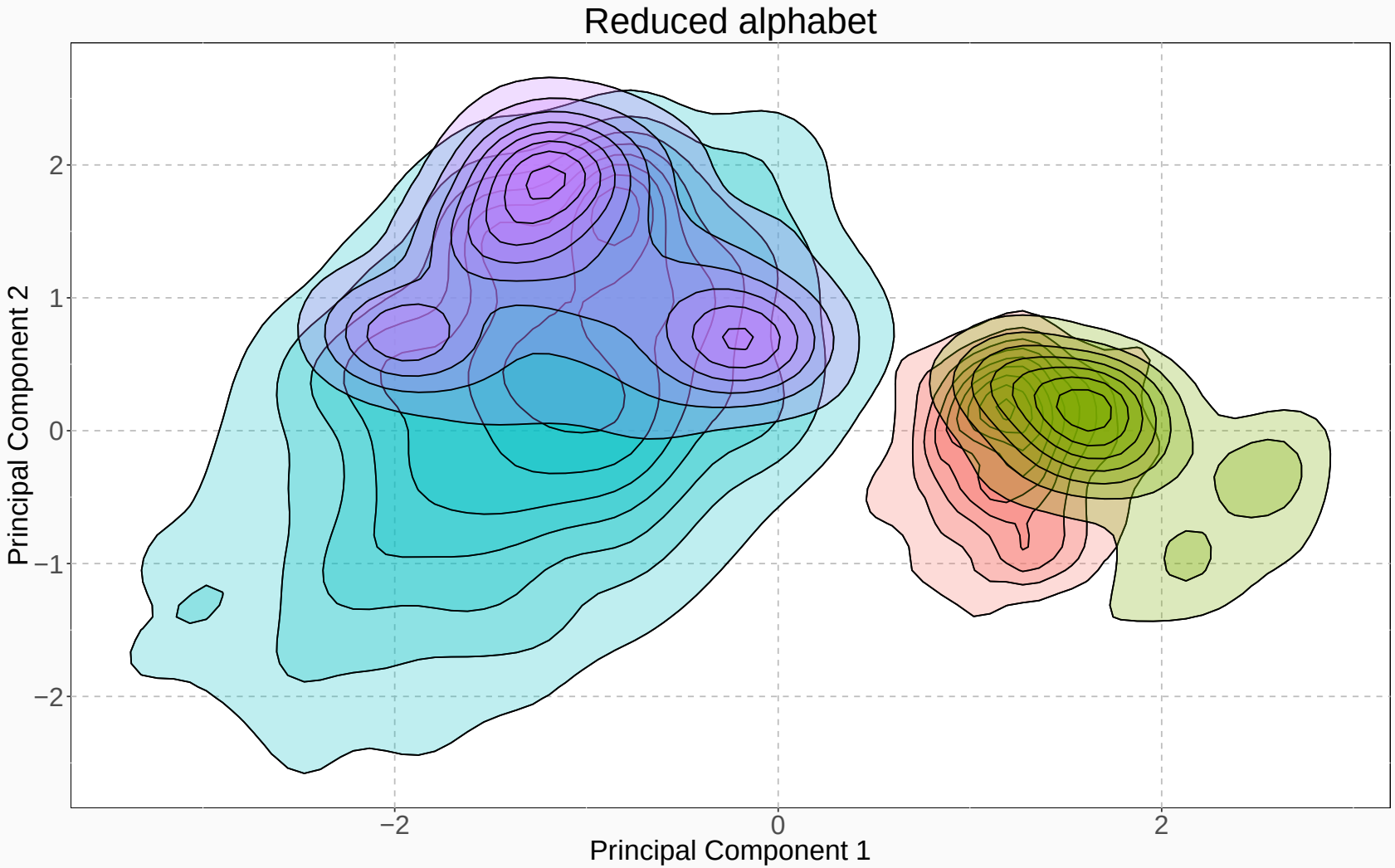
NH_2-SKINNYSLINKYKINKYTHING-COOH - targets apicoplast.

NH_2-ITWILLNEVERTARGETPLASTID-COOH - does not target apicoplast.

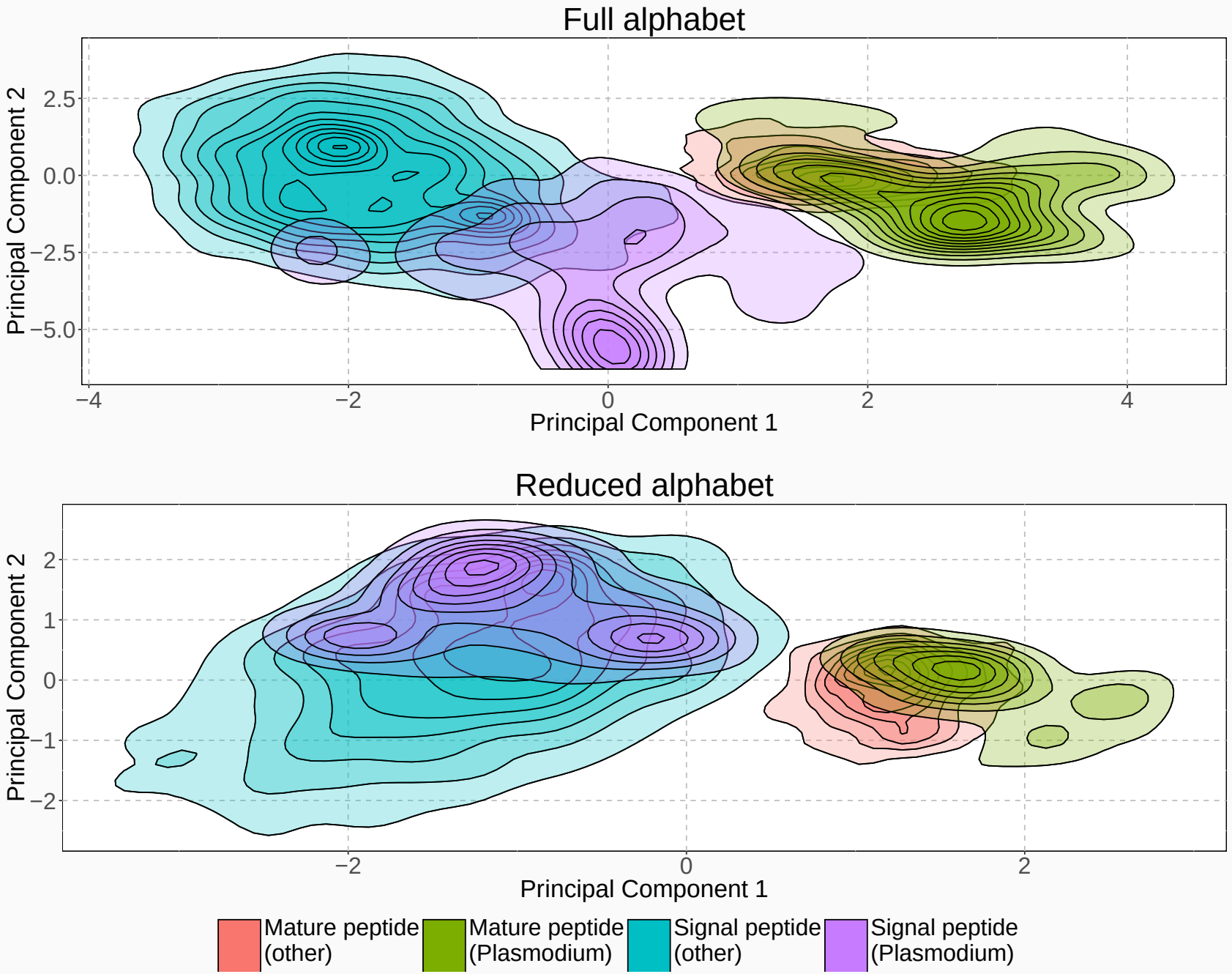
Results


Full alphabet
Principal Component 2
2.5
0.0
−2.5
−5.0
−4
−2
0
2
4
Principal Component 1
Mature peptide (other)
Mature peptide (Plasmodium)
Signal peptide (other)
Signal peptide (Plasmodium)

Group	Amino acids
1	D, E, H, K, N, Q, R
2	G, P, S, T, Y
3	F, I, L, M, V, W
4	A, C



Signal peptides, after the reduction of the amino acid alphabet, group together despite their different origins.


Full alphabet
Principal Component 2
2.5
0.0
−2.5
−5.0
−4
−2
0
2
4
Principal Component 1
Reduced alphabet
Principal Component 2
2
1
0
−1
−2
−2
0
2
Principal Component 1
Mature peptide (other)
Mature peptide (Plasmodium)
Signal peptide (other)
Signal peptide (Plasmodium)

Benchmark data set: 51 proteins with signal peptide and 211 proteins without signal peptide from members of *Plasmodiidae*.

	MCC	AUC
signalP 4.1 (no tm) (Petersen et al., 2011)	0.6872	0.8667
signalP 4.1 (tm) (Petersen et al., 2011)	0.6196	0.7951
signalP 3.0 (NN) (Bendtsen et al., 2004)	0.7220	0.8938
signalP 3.0 (HMM) (Bendtsen et al., 2004)	0.5553	0.7734
Phobius (Käll et al., 2004)	0.5895	0.7880
signalHsmm-2010	0.7409	0.9262
signalHsmm-2010 (hom. 50%)	**0.7621**	**0.9384**
signalHsmm-2010 (raw aa)	0.6853	0.8718
signalHsmm-1987	0.7271	0.9063
signalHsmm-1987 (hom. 50%)	0.7194	0.9090
signalHsmm-1987 (raw aa)	0.6350	0.8350

signalHsmm1987: trained on data set of 496 eukaryotic proteins with signal peptides added before year 1987.

signalHsmm2010: trained on data set of 3676 eukaryotic proteins with signal peptides added before year 2010.

Acknowledgements and funding

This research was partially funded by the KNOW Consortium and National Science Center (2015/17/N/NZ2/01845).

- Paweł Mackiewicz.
- **biogram** package (`https://cran.r-project.org/package=biogram`):
  - Piotr Sobczyk,
  - Chris Lauber.

References

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: Signalp 3.0. *Journal of Molecular Biology*, 340(4):783 – 795.

Hegde, R. S. and Bernstein, H. D. (2006). The surprising complexity of signal sequences. *Trends in Biochemical Sciences*, 31(10):563–571.

References II

Jain, R. G., Rusch, S. L., and Kendall, D. A. (1994). Signal peptide cleavage regions. functional limits on length and topological implications. *The Journal of Biological Chemistry*, 269(23):16305–16310.

Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338(5):1027–1036.

Moeller, L., Gan, Q., and Wang, K. (2009). A bacterial signal peptide is functional in plants and directs proteins to the secretory pathway. *Journal of Experimental Botany*, 60(12):3337–3352.

References III

Nagano, R. and Masuda, K. (2014). Establishment of a signal peptide with cross-species compatibility for functional antibody expression in both escherichia coli and chinese hamster ovary cells. *Biochemical and Biophysical Research Communications*, 447(4):655 – 659.

Nielsen, H. and Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden markov model. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 6:122–130.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10):785–786.