

Table 1: A list of initial features for prediction of blood-secreted proteins. (The number in parentheses represents the vector dimension of each property)

Type of properties	Features (dimension)	Sources
General sequence features	Amino acid composition (20), sequence length (1), di-peptides composition (400)	locally calculated
	Normalized Moreau-Broto autocorrelation (240), Moran autocorrelation (240), Geary autocorrelation (240), Sequence order(160), Pseudo amino acid composition (50)	Profeat[1]: http://jing.cz3.nus.edu.sg/cgi-bin/prof/prof.cgi
Physicochemical properties	Hydrophobicity (21), normalized Van der Waals volume (21), polarity (21), polarizability (21), charge (21), secondary structure (21) and solvent accessibility (21)	locally computed with three descriptors: composition (C), transition (T), and distribution (D)[2]
	Solubility (1), unfoldability (1), disorder regions (3), global charge (1) and hydrophobicity (1)	PROSO[3] : http://webclu.bio.wzw.tum.de:8080/proso/ Phobius[4]: http://phobius.sbc.su.se/
Structural properties	Secondary structural content (4), shape (Radius Gyration) (1)	SSCP[5]: http://coot.embl.de/SSCP/ Radius Gyration : http://www.scfbio-iitd.res.in/software/proteomics/rg.jsp
Domains and motifs	Signal peptide (1), transmembrane domains (alpha helix and beta barrel) (5), Glycosylation (both N-linked and O-linked) (4), Twin-arginine signal peptides motif (TAT) (1)	SignalP[6]: http://www.cbs.dtu.dk/services/SignalP/ ; TMB-Hunt[7]: http://bmbpcu36.leeds.ac.uk/~andy/betaBarrel/AACompPred/aaTMB_Hunt.cgi ; NetOgly[8]: http://www.cbs.dtu.dk/services/NetOGlyc/ ; NetNgly: http://www.cbs.dtu.dk/services/NetNGlyc/ ; TatP[9]: http://www.cbs.dtu.dk/services/TatP/

- Li, Z.R., et al., *PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence*. Nucleic Acids Res, 2006. **34**(Web Server issue): p. W32-7.
- Cui, J., et al., *Computer prediction of allergen proteins from sequence-derived protein structural and physicochemical properties*. Mol Immunol, 2007. **44**(4): p. 514-20.
- Brown, J.M. and A.J. Giaccia, *The unique physiology of solid tumors: opportunities (and problems) for cancer therapy*. Cancer Res, 1998. **58**(7): p. 1408-16.
- Kall, L., A. Krogh, and E.L. Sonnhammer, *An HMM posterior decoder for sequence feature prediction that includes homology information*. Bioinformatics, 2005. **21** Suppl 1: p. i251-7.
- Eisenhaber, F., et al., *Prediction of secondary structural content of proteins from their amino acid composition alone. I. New analytic vector decomposition methods*. Proteins, 1996. **25**(2): p. 157-68.
- Bendtsen, J.D., et al., *Improved prediction of signal peptides: SignalP 3.0*. J Mol Biol, 2004. **340**(4): p. 783-95.

7. Garrow, A.G., A. Agnew, and D.R. Westhead, *TMB-Hunt: a web server to screen sequence sets for transmembrane beta-barrel proteins*. *Nucleic Acids Res*, 2005. **33**(Web Server issue): p. W188-92.
8. Julenius, K., et al., *Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites*. *Glycobiology*, 2005. **15**(2): p. 153-64.
9. Bendtsen, J.D., et al., *Prediction of twin-arginine signal peptides*. *BMC Bioinformatics*, 2005. **6**: p. 167.