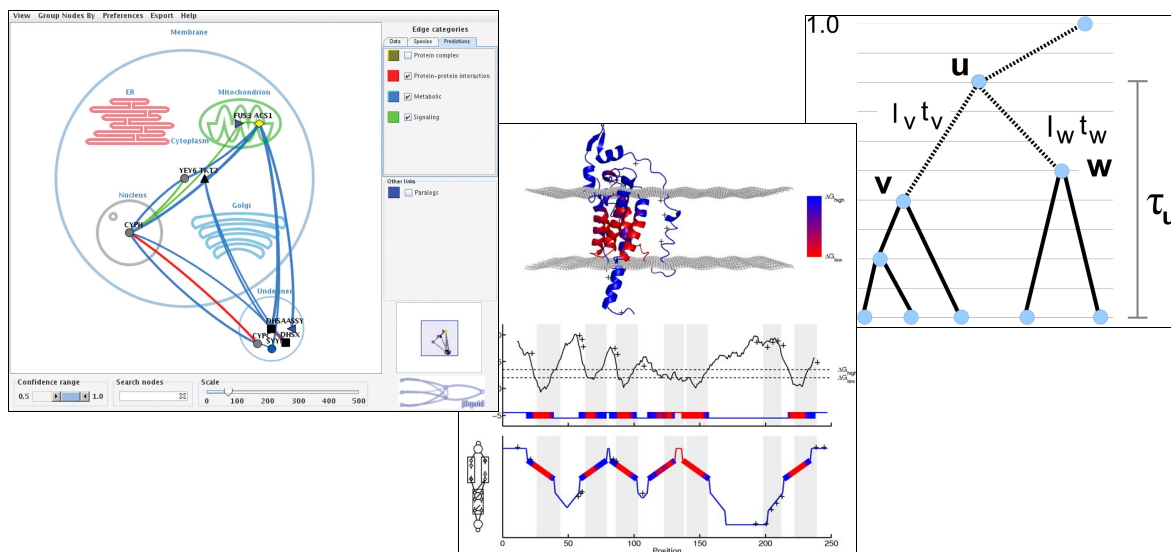


Stockholm Bioinformatics Centre Annual Report 2008



Director's summary

2008 contained many positive developments for SBC. We saw a considerable expansion of its personnel, several large grant awards, and staff promotions. In total, SBC now holds 42 members, about equally divided over the Albanova and Arrhenius sites. The ERC awarded an advanced grant to Gunnar von Heijne and a starting grant to Erik Lindahl, who also was awarded SSF's 'Framtidens forskningsledare' grant. Lars Arvestad was promoted to associate professor at CSC-KTH and Erik Lindahl is on tenure track at SU-DBB. Furthermore, Olof Emanuelsson obtained a research assistant position from VR, and will move across SBC's backyard to our neighbors BIO-KTH for in-depth bioinformatics studies of gene regulation. Many congratulations to these worthy awardees!

During 2008, five SBC PhD defenses were held: Håkan Viklund, Andreas Bernsel, Örjan Åkerborg, Diana Ekman, and Olivia Eriksson. Several new web servers were set up (PrIME-GSR, MapDP, PrIME-GEM, DASher, MultiParanoid) as well as a number of new DAS (Distributed Annotation System) services, now 10 in total.

The Master programmes that SBC participates in, the new KTH Master in Computational and Systems Biology, and the SU Master in Bioinformatics from 2007, both generated a lot of interest, but in the end enrolled far below the target number of students. There are two reasons for this: (1) there will be no Swedish bachelors looking for Master programmes until 2010, and (2) there are perhaps a hundred competing programmes around the world. Given this, it would make sense to put at least the SU the programme on hold until 2010 – this will only affect a small number of courses anyway, that were created specially for the programme.

Next year, ISMB/ECCB, the world's largest bioinformatics conference will for the first time take place in Stockholm, which should boost SBC's reputation internationally. We look forward to an exciting bioinformatics year.

Personnel during 2008:

Prof. Arne Elofsson**

Åsa Björklund	PhD student
*Diana Ekman	PhD student
*Olivia Eriksson	PhD student
Johannes Frey-Skött	PhD student
*Håkan Viklund	PhD student
Kristoffer Illergård	PhD student
Per Larsson	PhD student
Wiktor Jurkowski	Postdoc
Anni Kauko	Postdoc
Marcin Skwark	PhD student

Prof. Jens Lagergren

*Öjvind Johansson	Post-doc
Ali Tofigh	PhD student
*Örjan Åkerborg	PhD student
Hossein Farahani	PhD student
Joel Sjöstrand	PhD student

Ass. Prof. Erik Lindahl**

Anna Johansson	PhD student
Aron Hennerdal	PhD student
Pär Bjelkmar	PhD student
Jenny Falk	PhD student
Yana Vereshchaga	Post-doc
Sander Pronk	Post-doc
Arjun Ray	PhD student
Christine Schwaiger	PhD student
Rossen Apostolov	Post-doc
Samuel Murail	Post-doc
Björn Wallner	Research Associate

Prof. Gunnar von Heijne**

*Andreas Bernsel	PhD student
------------------	-------------

Prof. Erik Sonnhammer (Director of the SBC)

Andrey Alexeyenko	Post-doc
Kristoffer Forslund	PhD student
Anna Henricson	PhD student
Gabriel Östlund	PhD student
David Messina	PhD student
Oliver Frings	PhD student
Patrik Björkholm	PhD student
Sanjit Roopra	PhD student
Thomas Schmitt	PhD student

Lars Arvestad Associate professor

Bengt Sennblad	Assistant professor
Karin Julenius	Assistant professor
Olof Emanuelsson	Research associate
Erik Sjölund	System administrator

*) Left during 2008

**) Group located at Arrhenius Laboratory, Frescati

Collaboration partners

SU Molecular Biology & Functional Genomics (Prof. Marie Öhman)
SU Genetics, Microbiology and Toxicology (Prof. Ingrid Faye)
SU Wenner-Gren Institute (Prof. Christos Samakovlis)
KTH Biotechnology (Prof. Tuula Teeri, Prof. Peter Savolainen, Prof Joakim Lundeberg, and Prof. Mathias Uhlén)
KI Oncology-Pathology (Dr. Serhiy Souchelnytskyi)
KI Aging Research Center (Dr. Lars Bäckman)
KI MMK (Prof. Anna Wedell)
KI CMB (Prof. Björn Andersson)
KI MTC (Prof. Mats Wahlgren)
KI BioNut, Södertörns Högskola (Dr. Thomas Bürklin)
Uppsala University (Dr. van der Spoel)
Uppsala University (Prof. Hans Ellegren)
Göteborg University (Prof. Bengt Oxelman)
LU Clinical Genetics (Mattias Höglund)
AstraZeneca (Prof. Hugh Salter)
Bioinformatics Laboratory, BioInfoBank Institute, Poznan (Dr. Leszek Rychlewski)
Institut Pasteur, Paris (Dr. Marc Delarue)
Stanford University (Prof. Michael Levitt, Prof. Vijay S. Pande, Prof. James Trudell)
University of Wyoming (Dr. David Liberles)
McGill Centre for Bioinformatics (Dr. Mike Hallett)
University of British Columbia (Dr. Wyeth Wasserman).
Yale University, New Haven, CT. (Dr. Mark Gerstein)
University of Buffalo (Dr. Daniel Fischer)
Cornell University, Ithaca, NY. (Dr. Klaas van Wijk)
The Sanger Institute, Hinxton, UK. (Drs. Richard Durbin & Alex Bateman)
Janelia farm, VA, USA. (Dr. Sean Eddy)
University of Valencia, Spain (Dr. Gustavo Camps-Valls)
University of Rochester Medical Center (Dr. Fred Hagen)
University of Paris René Descartes (Prof. Jean-Laurent Casanova)
Max Planck Inst. für Genetik, Berlin. (Alexander Schliep)
Duke University (Dr. Joel Meyer)
EU bioinformatics network Biosapiens
EU bioinformatics network Embrace
EU bioinformatics network Genefun

Scientific publications

2008 was again a very fruitful year for SBC in terms of publications, with 29 bioinformatics papers in total. Many of them represent collaborations with experimental groups, testifying to SBC's ability to apply new computational techniques to real-world biological problems.

Based on <http://www.sbc.su.se/publications>:

Alex S Rajangam, Hongqian Yang, Tuula T Teeri, and Lars Arvestad, Evolution of a domain conserved in microtubule-associated proteins of eukaryotes, *Computational Biology and Chemistry: Advances and Applications*, 2008:1 51–69
(not in PubMed, http://www.doveoa.com/journals.php?pa=abstract&article_id=2260)

Bjorklund, A.K., Light, S., Hedin, L. and Elofsson, A. (2008) Quantitative assessment of the structural bias in protein-protein interaction assays. *Proteomics* 8 (22) : 4657-4667.

Daub, C.O. and Sonnhammer, E.L. (2008) Employing conservation of co-expression to improve functional inference. *BMC Syst Biol* 2 (1) : 81.

Rajangam, A.S., Kumar, M., Aspeborg, H., Guerriero, G., Arvestad, L., Pansri, P., Brown, C.J., Hober, S., Blomqvist, K., Divne, C., Ezcurra, I., Mellerowicz, E., Sundberg, B., Bulone, V. and Teeri, T.T. (2008) MAP20, a Microtubule-Associated Protein in the Secondary Cell Walls of Hybrid Aspen, Is a Target of the Cellulose Synthesis Inhibitor 2,6-Dichlorobenzonitrile. *Plant Physiol* 148 (3) : 1283-1294.

Moore, A.D., Bjorklund, A.K., Ekman, D., Bornberg-Bauer, E. and Elofsson, A. (2008) Arrangements in the modular evolution of proteins. *Trends Biochem Sci* 33 (9) : 444-451.

Forslund, K. and Sonnhammer, E.L. (2008) Predicting protein function from domain content. *Bioinformatics* 24 (15) : 1681-1687.

Cassel, M., Seppala, S. and von Heijne, G. (2008) Confronting fusion protein-based membrane protein topology mapping with reality: the *Escherichia coli* ClcA H⁺/Cl⁻ exchange transporter. *J Mol Biol* 381 (4) : 860-866.

Kauko, A., Illergard, K. and Elofsson, A. (2008) Coils in the membrane core are conserved and functionally important. *J Mol Biol* 380 (1) : 170-180.

Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G. and Elofsson, A. (2008) Prediction of membrane-protein topology from first principles. *Proc Natl Acad Sci U S A* 105 (20) : 7177-7181.

Viklund, H. and Elofsson, A. (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* 24 (15) : 1662-1668.

Lindahl, E.R. (2008) Molecular dynamics simulations. *Methods Mol Biol* 443: 3-23.

Klammer, M., Roopra, S. and Sonnhammer, E.L. (2008) jSquid: a Java applet for graphical online network exploration. *Bioinformatics* 24 (12) : 1467-1468.

Larsson, P., Wallner, B., Lindahl, E. and Elofsson, A. (2008) Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci* 17 (6) : 990-1002.

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q. and van Wijk, K.J. (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* 3 (4) : e1994.

Lindahl, E. and Sansom, M.S. (2008) Membrane proteins: molecular dynamics simulations. *Curr Opin Struct Biol* 18 (4) : 425-431.

Chalk, A.M. and Sonnhammer, E.L. (2008) siRNA specificity searching incorporating mismatch tolerance data. *Bioinformatics* 24 (10) : 1316-1317.

Bertaccini, E.J., Lindahl, E., Sixma, T. and Trudell, J.R. (2008) Effect of Cobratoxin Binding on the Normal Mode Vibration within Acetylcholine Binding Protein. *J Chem Inf Model* 48 (4) : 855-860.

Akerborg, O., Sennblad, B. and Lagergren, J. (2008) Birth-death prior on phylogeny and speed dating. *BMC Evol Biol* 8: 77.

Hong, J., Wei, N., Chalk, A., Wang, J., Song, Y., Yi, F., Qiao, R.P., Sonnhammer, E.L., Wahlestedt, C., Liang, Z. and Du, Q. (2008) Focusing on RISC assembly in mammalian cells. *Biochem Biophys Res Commun* 368 (3) : 703-708.

Joannin, N., Abhiman, S., Sonnhammer, E.L. and Wahlgren, M. (2008) Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. *BMC Genomics* 9: 19.

Johansson, A.C. and Lindahl, E. (Epub 2008) Correction to the article "Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations" *Proteins* (Epub ahead of print) 70 (4) : 1655-1656.

Papaloukas, C., Granseth, E., Viklund, H. and Elofsson, A. (2008) Estimating the length of transmembrane helices using Z-coordinate predictions. *Protein Sci* 17 (2) : 271-278.

Bernsel, A., Viklund, H. and Elofsson, A. (2008) Remote homology detection of integral membrane proteins using conserved sequence features. *Proteins* 71 (3) : 1387-1399.

Berglund, A.C., Sjolund, E., Ostlund, G. and Sonnhammer, E.L. (2008) InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res* 36 (Database issue) : D263-6.

Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L. and Bateman, A. (2008) The Pfam protein families database. *Nucleic Acids Res* 36 (Database issue) : D281-8.

Johansson, A.C. and Lindahl, E. (2008) Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations. *Proteins* 70 (4) : 1332-1344.

Wallner, B., and Elofsson, A. (2008) Prediction of global and local model qualities using MQAPs *Eds Bujnicki* 1: 1-1.

Hess, B., Kutzner, C. van der Spoel, D., Lindahl, E. GROMACS 4.0: Algorithms for highly efficient, load balanced, and scalable molecular simulation, *J. Chem. Theory Comp.* 4 (2), 435 (2008) (not in PubMed)

Lage K, Hansen NT, Karlberg EO, Eklund AC, Roque FS, Donahoe PK, Szallasi Z, Jensen TS, Brunak S. (2008) A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc Natl Acad Sci U S A.* 105:20870-5

Courses and workshops

Bioinformatics 6 hp DD2396 (KTH) by Lars Arvestad

Applied Bioinformatics 7.5 hp DD2397 (KTH) / DA2397 (SU) by Lars Arvestad

Algoritmisk bioinformatik 6 hp DD2450 (KTH) by Jens Lagergren

Bioinformatics 7.5 hp KB7004 (SU) by Arne Elofsson, Erik Lindahl, Gunnar von Heijne

Molecular modelling 7.5 hp KB8005 (SU) by Erik Lindahl, Arne Elofsson

Applied scientific programming 7.5 hp KB8009 (SU) by Arne Elofsson, Erik Lindahl

Comparative Genomics 7.5 hp KB8007 (SU) by Erik Sonnhammer

Structure prediction of globular and membrane proteins 7.5 hp KB8008 (SU) by Arne Elofsson, Erik Lindahl

Protein physics 7.5 hp KB8011 (SU) by Arne Elofsson, Erik Lindahl

Invited lectures and seminars

International meeting "The Unbearable Complexity of Life" (invited lecture). Tel Aviv, Israel. February 2008. by Gunnar von Heijne

Gordon conference "Ligand Recognition and Molecular Gating" (invited lecture). Ventura, USA. March 2008. by Gunnar von Heijne

International conference "Bioinformatics 2008" (invited plenary lecture). Warsaw, Poland. April 2008. by Gunnar von Heijne

INPEC Meeting 2008 (invited lecture). Turku, Finland. June 2008. by Gunnar von Heijne

Meeting of the UK, Spanish & Portuguese Biophysical societies (invited plenary lecture). Lisbon, Portugal. July 2008. by Gunnar von Heijne

FASEB Conference "Molecular Biophysics of Cellular Membranes" (invited lecture & chair). Saxton River, USA. July 2008. by Gunnar von Heijne

Horizon symposium (invited lecture). Göttingen, Germany. September 2008. by Gunnar von Heijne

EMBO Conference on Protein Translocation (invited lecture). Sainte-Maxime, France. October 2008. by Gunnar von Heijne

"EMBN workshop", Groningen Jan 2008 by Arne Elofsson

"EBI membrane workshop", Hinxton, Feb 2008 by Arne Elofsson

“Bioinformatics winter school”, Bologna Feb 2008 by Arne Elofsson

“CASP 7.5 meeting”, Madrid April 2008 by Arne Elofsson

“Network-based ranking of candidate cancer genes”, National resources group for computational and system biological studies in cancer (COMPAS), Rosenön, 17 October 2008 by Erik Sonnhammer

“Reconstructing gene networks to predict protein function”, Umeå University, 22 May 2008 by Erik Sonnhammer

“Xfam meeting”, Hinxton, April 9 2008 by Erik Sonnhammer

Systematikdagarna 2008, “PrIME - ett nytt probabilistiskt paket för genträäd, artträäd och rekoncilieringar”, SLU, Uppsala, 24 nov – 25 Nov, by Bengt Sennblad

“Annual Meeting of the Society of Molecular Biology and Evolution (SMBE)”, Barcelona, 5-8 June 2008 by Lars Arvestad

Introduction to molecular simulation and the GROMACS toolkit, Stanford Molecular Simulation Workshop, Stanford University, USA, April 7, 2008 by Erik Lindahl

High Performance computing for life science research, Erik Lindahl, Dell HPC world tour, Stockholm, May 13, 2008

Scaling and other bad ideas in high performance computing, Erik Lindahl, PDC summer school, Stockholm, Sweden, August 25, 2008

Insertion and stabilization of marginally stable residues in bilayers, Erik Lindahl, CECAM

Workshop on Membrane Protein Insertion, Lausanne, Switzerland, September 6, 2008 by Erik Lindahl

Stream computing, Erik Lindahl, CASPUR Summer School, Rome, Italy, September 12, 2008

Insertion & function of non-hydrophobic residues in membrane proteins, Erik Lindahl, Molecular approaches to membrane transport, Sigtuna, Sweden, September 13, 2008

Free energy techniques for molecular simulation, Göttingen GROMACS workshop, Max-Planck Institut Göttingen, Germany, September 25, 2008 by Erik Lindahl

Multiscale & distributed simulation for free energy calculations, Erik Lindahl, NORDFORSK, Stockholm, Sweden, December 2, 2008

Computer infrastructure

The SBC employs a very standardized computer system in which each workplace has an identically set up desktop computer. All user disk storage is done at PDC and is accessed via the AFS file system (in 5-10 Gb volumes). Heavy computation is carried out on a compute cluster also maintained by PDC, which can also access the user disks. The old SBC compute

cluster was shut down due to floor space limitations at PDC, but access was granted to the new cluster Ferlin instead. A summary of the infrastructure is listed below.

Desktop computers:

46* Intel Pentium 4 dualthread/AMD 2.8 GHz, mostly 2 Gb RAM running Centos Linux 5 or Ubuntu

Ferlin compute cluster:

In total 5440 cores, 2.66GHz CPUs on 8-CPU nodes with 8 Gb RAM (shared with SNIC).

Disk servers:

2 servers, ~8 Tb in total

Internal servers: mail, cups, life, mickey, sbcdb

Web servers:

<http://www.sbc.su.se>:

Intel Core2 Quad 2.40GHz, 4 Gb RAM, 250 Gb disk"
accessed from 9000-15000 unique IP numbers per month.

Hosted services:

- * PRIMETV: Visualize tree reconciliations
- * PrIME-GSR: A Bayesian integrated model for genes, sequences, and rates
- * MapDP: factorizing branchlengths into divergence times and rates
- * PrIME-GEM: Probabilistic orthology analysis (binaries downloadable)
- * Pmembr A threading method for membrane proteins.
- * HMMER High capacity site for use of HMMER to search SCOP or Pfam
- * ProQ A protein model quality predictor.
- * PeroxiP Predict peroxisomal proteins and Pfam domains
- * PRODIV-TMHMM Topology and reentrant predictions.
- * TMHMMfix TMHMM with optional fixing and reliability score calculation.
- * DAS Prediction of Transmembrane Regions.
- * NucPred Nuclear localization prediction.
- * DRIP-PRED Disorder/order prediction for proteins.
- * GPCPRED Contact map prediction for proteins.
- * SVMHC Prediction of MHC class I binding peptides.
- * PhylProM Phylogenetic profiles
- * OVOP automatic view generation for protein structures (source code available)
- * modhmm A modular HMM program used in PRO(DIV)-TMHMM and other studies..
- * LGscore A program to measure the similarity between proteins.
- * Palign Our alignment/threading programs.
- * ssHMM Secondary structure HMMs based on HMMER
- * LEPRO Protein modelling C++ /library.
- * TAED The Adaptive Evolution Database.
- * www.genefun.org GeneFun EU collaboration
- * www.perlgp.org PerlGP, The Open Source Perl Genetic Programming System
- * www.socbin.org Society for Bioinformatics in Northern Europe
- * prime.sbc.su.se Probabilistic Integrated Models of Evolution.
- * avdist: A tool for analyzing haplotype differences.
- * repeatalign: Binary Repeat Align server.

* RefSense: An alternative to Pubmed.

sbcweb.pdc.kth.se:

Hosted services:

helium.sbc.su.se:

2*2.80GHz Pentium 4, 2 Gb RAM, RAID 2x750 Gb disks

Hosted services:

- InParanoid.sbc.su.se A comprehensive database of orthologs and inparalogs in eukaryotes
- Pfam.sbc.su.se A comprehensive database of protein domain families.
- FunCoup.sbc.su.se Comprehensive protein networks of functional coupling.
- jSquid.sbc.su.se A java tool to visualize networks and edge scores in FunCoup.
- Sfinx.sbc.su.se Prediction of functional and structural features in proteins.
- Sfixem.sbc.su.se A java viewer for Sfinx.
- Phobius.sbc.su.se A combined transmembrane topology and signal peptide predictor.
- GPCRHMM.sbc.su.se A hidden Markov model for GPCR detection.
- Humanoid.sbc.su.se Human ortholog groups and functional shift analysis of subfamilies.
- MSA.sbc.su.se Multiple alignments and assessment of alignment accuracy.
- DASher
- MultiParanoid.sbc.su.se
- DAS services for Phobius, signalP, HMMTOP, PhD, Toppred, etc.
(<http://das.sbc.su.se:9000/das/>*)

argon.sbc.su.se:

4* 2.8 GHz Pentium 4, 4 Gb RAM; 1.25 Tb disks

fw.sbc.su.se:

Hosted services:

- SOAP Web Services for EMBRACE