

Research

Computational prediction of membrane-tethered transcription factors

Joel Zupicich*, Steven E Brenner† and William C Skarnes*

Addresses: *Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA 94720-3200, USA. †Department of Plant and Microbial Biology, University of California at Berkeley, Berkeley, CA 94720-3102, USA.

Correspondence: Joel Zupicich. E-mail: joelz@socrates.berkeley.edu

Published: 14 November 2001

Received: 5 October 2001

Genome Biology 2001, **2**(12):research0050.1-0050.6

Accepted: 15 October 2001

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/12/research/0050>

© 2001 Zupicich et al., licensee BioMed Central Ltd
(Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

Background: Sequestration of transcription factors in the membrane is emerging as an important mechanism for the regulation of gene expression. A handful of membrane-spanning transcription factors has been previously identified whose access to the nucleus is regulated by proteolytic cleavage from the membrane. To investigate the existence of other transmembrane transcription factors, we analyzed computationally all proteins in SWISS-PROT/TrEMBL for the combined presence of a DNA-binding domain and a transmembrane segment.

Results: Using Pfam hidden Markov models and four transmembrane-prediction programs, we identified with high confidence 76 membrane-spanning transcription factors in SWISS-PROT/TrEMBL. Analysis of the distribution of two proteins predicted by our method, MTJ1 and DMRT2, confirmed their localization to intracellular membrane compartments. Furthermore, elimination of the predicted transmembrane segment led to nuclear localization for each of these proteins.

Conclusions: Our analysis uncovered a wealth of predicted membrane-spanning transcription factors that are structurally and taxonomically diverse, 56 of which lack experimental annotation. Seventy-five of the proteins are modular in structure, suggesting that a single proteolysis may be sufficient to liberate a DNA-binding domain from the membrane. This study provides grounds for investigations into the stimuli and mechanisms that release this intriguing class of transcription factors from membranes.

Background

A critical step in regulating many transcriptional responses is the import of transcription factors from the cytosol to the nucleus. Many transcription factors are held outside the nucleus in a complex with cytosolic proteins or with membrane receptors, and translocate to the nucleus in response to various stimuli [1]. Alternatively, transcription factors may be inserted directly into the membrane, thereby preventing their access to the nucleus. A handful of such proteins has been shown to be released from membranes by a process known as regulated

intramembrane proteolysis (RIP) [2]. This process is best understood for SREBP-1 and SREBP-2, two basic leucine zipper (bZIP) transcription factors that normally reside in the membrane of the endoplasmic reticulum and Golgi apparatus. When cellular sterol levels dip, SREBPs are liberated from the membrane in a two-step mechanism involving the action of Site-1 protease, a site-specific protease that cleaves the protein within the Golgi lumen, followed by Site-2 protease, an integral membrane protease, that cleaves a membrane-spanning helix. Once liberated from the membrane, transport to the nucleus

enables these transcription factors to initiate expression of genes involved in cholesterol uptake and biosynthesis [2].

Several more examples of membrane-tethered transcriptional regulators have recently been identified by biochemical means, notably ATF6 [3], G13 [4], CadC [5], ToxR [6], Lzip (Luman [7]), Notch [8] and SPT23 [9]. All appear to undergo proteolytic cleavages to release a fragment that is targeted to DNA or the nucleus, but may use different proteases. For example, ATF6 uses the same proteolytic machinery, as do SREBPs [10], whereas Notch is cleaved by different proteases [11]. Tumor necrosis factor (TNF α)-converting enzyme catalyzes the cleavage of the extracellular domain of Notch, followed by presenilin/gamma-secretase-like activity to liberate the intracellular fragment [12]. Thus, the release of some membrane-bound nuclear proteins involves regulated cleavages in the luminal or extracellular space, followed by a cleavage by an integral membrane protease to release an active fragment.

Using conventional biochemistry, the identification of transmembrane transcription factors (TMTFs) can be easily overlooked. For example, transcription factors are generally assumed to be soluble proteins and, consequently, membrane fractions are often discarded during purification. Moreover, the nuclear form of the protein may be rapidly degraded and thus difficult to detect, as is the case for SREBPs [13]. Lastly, the subcellular distributions of transcription factors are often not examined. Cell-fractionation studies of other transcription factors show smaller-molecular-weight forms of these proteins enriched in the nucleus, suggestive of a cleavage event [14,15]. We thus investigated the prevalence of transmembrane transcription factors using computational tools to search for membrane-spanning proteins that contain conserved DNA-binding domains.

Results and discussion

Computational analysis of protein databases reveals a large number of predicted transmembrane transcription factors

We used Pfam [16] hidden Markov models for 53 DNA-binding domains (see Materials and methods) to search all proteins in SWISS-PROT/TrEMBL [17] and SwissPfam protein databases. The 9,261 proteins identified by our

search are presumed members of DNA-binding protein families, and most are expected to be transcription factors. These proteins were then scored for the presence of one or more transmembrane segments using prediction programs PHDhtm [18], TMHMM [19], HMMTOP [20] and PSORTII [21]. Only those proteins containing membrane-spanning helices predicted by at least three of the four programs were deemed significant in our analysis. By these stringent criteria, 76 proteins from 20 organisms and one virus were identified as putative TMTFs (Figure 1).

Our analysis predicted a surprisingly large and diverse set of membrane-tethered DNA-binding proteins. Seventeen of the 53 DNA-binding domains chosen for this analysis were represented in the final set of TMTFs. Of these, the most abundant is the zf-C4 (zinc-finger type C4) nuclear hormone receptor DNA-binding domain, found in 14 proteins in *Caenorhabditis elegans* and avian erythroblastosis virus. TMTFs in *Arabidopsis* were the most diverse, and were associated with eight different DNA-binding domains. All but two proteins have DNA-binding domains that could be separated from the rest of the protein by a single hypothetical cleavage event, if singly predicted transmembrane segments are discounted (Figure 1). DNA-binding domains were also frequently juxtaposed to bipartite nuclear localization signals, suggesting that transmembrane and DNA-binding domains in TMTFs are modular. Thus, the overall topology of these proteins is consistent with other known TMTFs. *C. elegans* has an impressive 25 predicted TMTFs, suggesting that RIP may be particularly important in the regulation of transcriptional responses in the worm. Interestingly, 56 of the 76 identified proteins lack any experimental annotation.

We deliberately used a stringent method to increase the likelihood of identifying only *bona fide* TMTFs and, as expected, most experimentally known TMTFs were detected by our analysis, including CadC [5], Lzip [7], ToxR [6] and all SWISS-PROT/TrEMBL orthologs of SREBP-1 and SREBP-2. Also found were several well-characterized proteins whose predicted membrane insertion had not been recognized. For example, the human doublesex-related protein DMRT2, *Drosophila* B-H2 (BarH2) protein, *C. elegans* UNC-86, and mouse OASIS protein are predicted TMTFs. Two known TMTFs, ATF6 and SPT23, did not satisfy our minimum

Figure 1 (see the figure on the next page)

The domain structure of predicted TMTFs is shown. Pfam-predicted DNA-binding domains, transmembrane segments and bipartite nuclear localization signals are shown for linear protein models and identified by SWISS-PROT/TrEMBL accession number. The total number of proteins predicted for each species is given. Colored icons represent various DNA-binding domains. Predicted transmembrane segments for each program are represented by a filled box. Protein lengths are drawn approximately to scale; positions of domains are approximate. Arrows in MTJ1 and DMRT2 indicate sites for truncated protein localization experiments shown in Figure 2. The scale of proteins O80659 and Q9SGP0 is reduced by half. Orthologs of predicted TMTFs not shown are: Luman (Q9UE77 *Homo sapiens*), SREBP-1 (Q60416 *Cricetulus griseus*, Q9WTN3 *Mus musculus*, P56720 *Rattus norvegicus*, Q9XX00 *Caenorhabditis elegans*), SREBP-2 (Q9UH04 *H. sapiens*, Q60429 *C. griseus*), and AFLR Reg (P43651 *Aspergillus parasiticus*). Open reading frames (ORFs) for O65420, O43989, Q17928 were extended using additional nucleotide sequence available in the NCBI database (indicated by stippled rectangles).

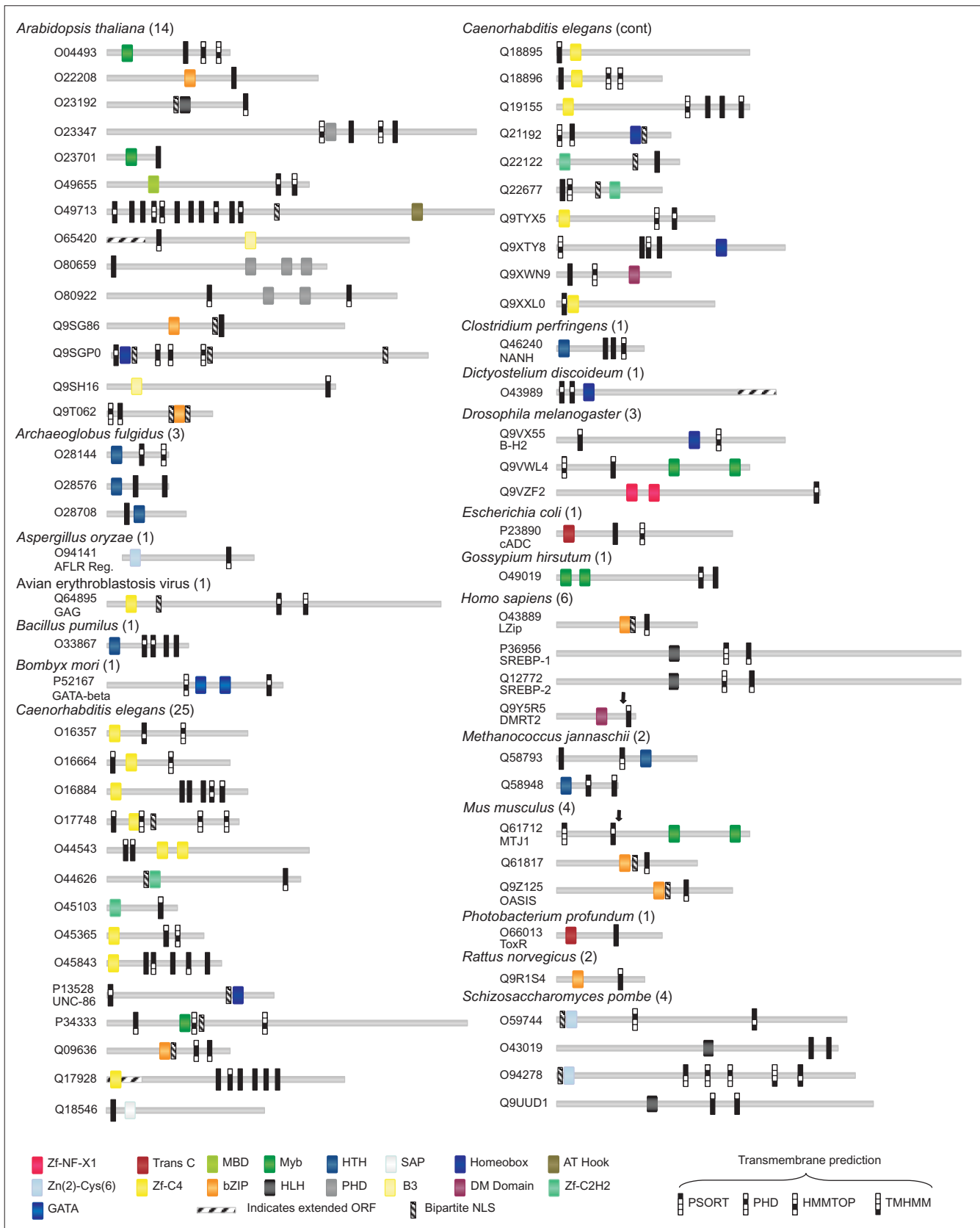


Figure 1 (see the legend on the previous page)

criteria. The transmembrane helix of ATF6 was predicted by only two programs: PSORT and HMMTOP. The immunoglobulin DNA-binding domain (TIG) of SPT23 is found in both cell-surface proteins as well as transcription factors and was therefore excluded from the set of DNA-binding domains. These results indicate that reducing the stringency of our prediction method will expand the number of predicted TMTFs.

TMTFs translocate to the nucleus on deletion of the predicted transmembrane helix

In some cases we found data in the literature to support our computational predictions. For example, cell-fractionation studies using antibodies directed at the carboxyl terminus of the chaperonin MTJ1 showed that the full-length (62 kDa) protein exists in microsomes, whereas a smaller 42 kDa form of the protein is found in the nucleus [15]. The 42 kDa species was hypothesized to represent a product of internal translation. Because MTJ1 contains putative Myb DNA-binding domains within the carboxy-terminal half of the protein, we re-examined the subcellular localization of carboxy-terminal-tagged MTJ1 in COS-7 cells (Figure 2a). Our results show clearly that full-length MTJ1 is normally associated with the endoplasmic reticulum. In contrast, a truncated form MTJ1 Δ (approximately 40 kDa in size) lacking the transmembrane segment accumulates in the nucleus. Therefore, we propose that the 42 kDa nuclear form of MTJ1 observed in cells arises by cleavage of MTJ1 from the membrane, rather than from aberrant translation of the mRNA.

DMRT2, a human homolog of *C. elegans mab-3*, was identified in our analysis as having a carboxy-terminal transmembrane segment (Figure 1). *mab-3* encodes a transcription factor known for its role in sex determination in worms [22]. *DMRT2* has gained recent attention as a candidate gene for sex-reversal phenotypes in humans [23]. To verify our prediction that *DMRT2* is a membrane-tethered transcription factor, we examined the subcellular localization of full-length and truncated forms of *DMRT2* in COS-7 cells (Figure 2b). Full-length *DMRT2* is localized primarily, but not exclusively, to vesicles outside the nucleus. A carboxy-terminal truncation containing the DNA-binding domain is, however, concentrated almost entirely in the nucleus. These results are consistent with the idea that *DMRT2* is cleaved from the membrane to produce a nuclear fragment. Interestingly, transformer protein TRA-2A, an indirect activator of MAB-3, has been identified recently as a membrane-tethered nuclear protein [24,25]. Thus, RIP may be a conserved mechanism common to sex determination in humans and worms.

Conclusions

We have used computational methods to investigate the prevalence of membrane-tethered transcription factors. The identification of 76 predicted TMTFs by our method, and the supporting cell biology, indicate that membrane-tethering may be a common mechanism for regulating transcriptional

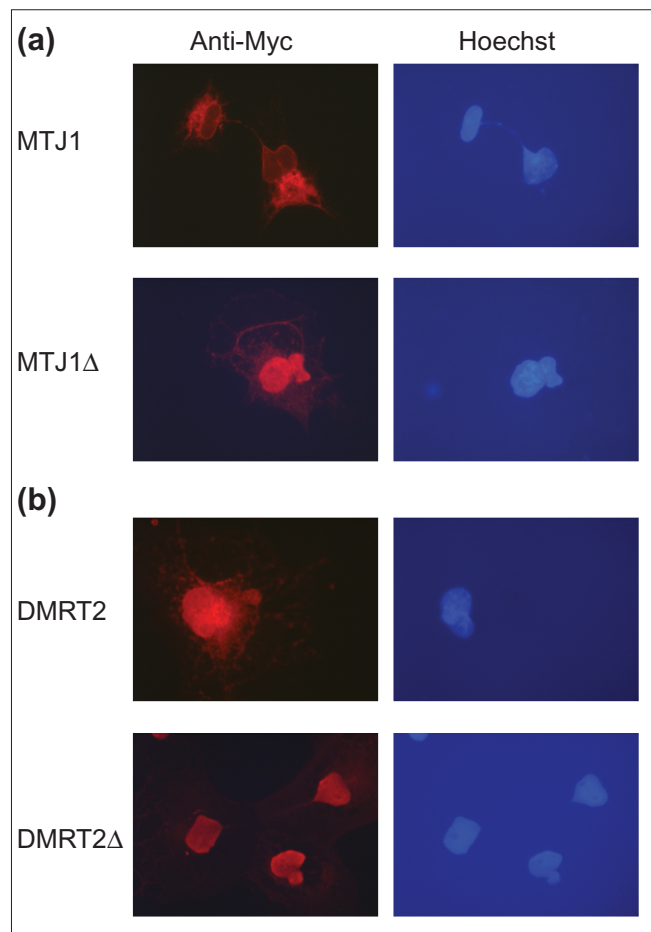


Figure 2
Subcellular localizations of predicted and truncated TMTFs in COS-7 cells were detected using anti-Myc antibodies. Full-length proteins are localized to intracellular membrane compartments, but truncated forms (Δ) lacking predicted transmembrane segments accumulate in the nucleus. Nuclei are stained with Hoechst. (a) mouse MTJ1; (b) human DMRT2.

responses. As stringent criteria were used to identify transmembrane segments and DNA-binding domains, we believe that the actual number of TMTFs is likely to be much larger. Compared to other signal transduction mechanisms, tethering transcription factors in the membrane provides an expeditious route to the nucleus in response to stimuli that must be communicated across a membrane. Our understanding of this process will be enhanced as more TMTFs are studied and the signals for membrane cleavage and their proteases are discovered.

Materials and methods

Computational analysis

Pfam [16] hidden Markov models for 53 DNA-binding domains (see DNA-binding domains below) were used to

search proteins in SWISS-PROT/TrEMBL (October 2000 release; 388,909 proteins) with p -value ≤ 0.0019 (0.01/53). SwissPfam proteins identified as having any of the 53 domains were also included in our analysis. The resulting 9,261 proteins were then analyzed for the presence of transmembrane helices. Default parameters were used for HMMTOP [20], PHDhtm [18], and TMHMM (version 2 [19]). A higher stringency (-5.0) than default was used for PSORT II (ALOM2 [21]). Transmembrane segments predicted by individual programs were considered overlapping if ten or more amino acids were shared by each segment. Proteins containing transmembrane helices predicted by at least three of the four programs were included in the final set. Bipartite nuclear localization signals were identified using PSORT II. Three predicted TMTFs were discounted as false-positives on the basis of partial or complete overlap of transmembrane helices with other Pfam domains (O01612, O23045 and Q13771).

DNA-binding domains

The following Pfam models for DNA-binding domains were used (abbreviated as in Pfam): 7 kDa DNA-binding; AP2-domain; ARID; ASNC trans reg; AT hook; Arg repressor; B3; BAH; BRO; Bac DNA-binding; basic; bZIP; CBFN NFYA; CSD; CUT; copper-fist; DM-domain; E2F TDP; fork head; GATA; HALZ; HLH; homeobox; HSF DNA-binding; HTH 3; HTH 4; HTH 5; IRF; LexA DNA-binding; MBD; MetJ; Myb DNA-binding; MutS N; Myc-LZ; PHD; RFX DNA-binding; RHD; Runt; SAP; sigma70; SRF-TF; STAT; sigma54 factors; sigma70 ECF; T-box; TBP; yeast DNA-binding; Trans reg C; zf (zinc finger)-C2H2; zf-C2HC; zf-C4; zf-NF-X1; Zn-clus.

Plasmid constructs

Full-length DMRT2 and MTJ1 Δ were generated by PCR using Pfu polymerase (Stratagene) and cloned directionally into *Bam*HI/*Xba*I sites of pCDNA3 (Invitrogen). Truncated MTJ1, in which an ATG (methionine) was added immediately before amino acid 171 (Q61712), was amplified from expressed sequence tag (EST) AI790297 (Incyte Genomics) and a Myc tag was added at the carboxyl terminus. MTJ1 Δ -forward primer: 5'-CGCGGATCCGCGATGGAAAAGCAACTGGATGAACTG-3'. MTJ1 Δ -reverse primer: 5'-GCTCTAGAGCTACAGGTCCTCCTCCGAGATGAGTTTCTGTTCCATGCTT-TTAGCCTGCTTTTCTT-3'. The ATG in bold indicates the translation start site of truncated MTJ1. Full-length MTJ1 was prepared by digesting clone AI790297 with *Xho*I, blunting ends, then digesting with *Eco*RI. This fragment was then cloned into pcDNA3-MTJ1 Δ , which was digested with *Bam*HI, blunt-ended, and digested with *Eco*RI. Full-length and truncated DMRT2 (at amino acid 180; Q9Y5R5) were amplified from EST AI985131 (Incyte), and a Myc tag was added at the amino terminus. DMRT2-forward primer: 5'-CGCGGATCCGCGATGGAACAGAACTCATCTCGGAGGAGGACCTGATGGCCGACCCGAGG-3'. DMRT2-reverse primer: 5'-GCTCTAGAGCTAAAGATGGTTTCATTATGTAC-3'. DMRT2 Δ -reverse primer: 5'-GCTCTAGAGTCAGGCTCTGACTTGCCTCTG-3'.

Cell culture and immunocytochemistry

Standard DEAE transfections [26] of plasmids were done in COS-7 cells (ATCC) and grown in 10% FBS/DMEM. Cells were fixed 72 h post-transfection in 3% PFA in PBS and Myc tags were detected with mouse anti-Myc antibodies (NeoMarkers, Fremont, CA) and Texas-Red-X goat anti-mouse antibodies (Molecular Probes, Eugene, OR) using standard procedures. Nuclei were counterstained with Hoechst 33258. Photomicrographs were taken on a Zeiss Axiophot.

Acknowledgements

We thank J. Rine and O. Kelly for critical comments on the manuscript, and D. He for assembling overlapping domains. This work was supported by the NIH (S.E.B. and W.C.S.). S.E.B. and W.C.S. are Searle Scholars.

References

- Kaffman A, O'Shea EK: **Regulation of nuclear localization: a key to a door.** *Annu Rev Cell Dev Biol* 1999, **15**:291-339.
- Brown MS, Ye J, Rawson RB, Goldstein JL: **Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans.** *Cell* 2000, **100**:391-398.
- Haze K, Yoshida H, Yanagi H, Yura T, Mori K: **Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress.** *Mol Biol Cell* 1999, **10**:3787-3799.
- Haze K, Okada T, Yoshida H, Yanagi H, Yura T, Negishi M, Mori K: **Identification of the G13 (cAMP-response-element-binding protein-related protein) gene product related to activating transcription factor 6 as a transcriptional activator of the mammalian unfolded protein response.** *Biochem J* 2001, **355**:19-28.
- Dell CL, Neely MN, Olson ER: **Altered pH and lysine signalling mutants of cadC, a gene encoding a membrane-bound transcriptional activator of the Escherichia coli cadBA operon.** *Mol Microbiol* 1994, **14**:7-16.
- Krukonis ES, Yu RR, Dirita VJ: **The Vibrio cholerae ToxR/TcpP/ToxT virulence cascade: distinct roles for two membrane-localized transcriptional activators on a single promoter.** *Mol Microbiol* 2000, **38**:67-84.
- Lu R, Misra V: **Potential role for human, the cellular homologue of herpes simplex virus VP16 (alpha gene trans-inducing factor), in herpesvirus latency.** *J Virol* 2000, **74**:934-943.
- Blaumueller CM, Qi H, Zagouras P, Artavanis-Tsakonas S: **Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane.** *Cell* 1997, **90**:281-291.
- Hoppe T, Matuschewski K, Rape M, Schlenker S, Ulrich HD, Jentsch S: **Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing.** *Cell* 2000, **102**:577-586.
- Ye J, Rawson RB, Komuro R, Chen X, Dave UP, Prywes R, Brown MS, Goldstein JL: **ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs.** *Mol Cell* 2000, **6**:1355-1364.
- Artavanis-Tsakonas S, Rand MD, Lake RJ: **Notch signaling: cell fate control and signal integration in development.** *Science* 1999, **284**:770-776.
- Weinmaster G: **Notch signal transduction: a real rip and more.** *Curr Opin Genet Dev* 2000, **10**:363-369.
- Wang X, Sato R, Brown MS, Hua X, Goldstein JL: **SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis.** *Cell* 1994, **77**:53-62.
- Skeiky YA, Drevet JR, Swevers L, Iatrou K: **Protein phosphorylation and control of chorion gene activation through temporal mobilization of a promoter DNA binding factor from the cytoplasm into the nucleus.** *J Biol Chem* 1994, **269**:12196-12203.
- Brightman SE, Blatch GL, Zetter BR: **Isolation of a mouse cDNA encoding MTJ1, a new murine member of the Dnaj family of proteins.** *Gene* 1995, **153**:249-254.
- Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL: **The Pfam protein families database.** *Nucleic Acids Res* 2000, **28**:263-266.

17. Bairoch A, Apweiler R: **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Res* 2000, **28**:45-48.
18. Rost B, Fariselli P, Casadio R: **Topology prediction for helical transmembrane proteins at 86% accuracy.** *Protein Sci* 1996, **5**:1704-1718.
19. Sonnhammer EL, von Heijne G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences.** *Proc Int Conf Intell Syst Mol Biol* 1998, **6**:175-182.
20. Tusnady GE, Simon I: **Principles governing amino acid composition of integral membrane proteins: application to topology prediction.** *J Mol Biol* 1998, **283**:489-506.
21. Nakai K, Horton P: **PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization.** *Trends Biochem Sci* 1999, **24**:34-36.
22. Raymond CS, Shamu CE, Shen MM, Seifert KJ, Hirsch B, Hodgkin J, Zarkower D: **Evidence for evolutionary conservation of sex-determining genes.** *Nature* 1998, **391**:691-695.
23. Raymond CS, Parker ED, Kettlewell JR, Brown LG, Page DC, Kusz K, Jaruzelska J, Reinberg Y, Flejter WL, Bardwell VJ, *et al.*: **A region of human chromosome 9p required for testis development contains two genes related to known sexual regulators.** *Hum Mol Genet* 1999, **8**:989-996.
24. Lum DH, Kuwabara PE, Zarkower D, Spence AM: **Direct protein-protein interaction between the intracellular domain of TRA-2 and the transcription factor TRA-1A modulates feminizing activity in *C. elegans*.** *Genes Dev* 2000, **14**:3153-3165.
25. Yi W, Ross JM, Zarkower D: ***mab-3* is a direct *tra-1* target gene regulating diverse aspects of *C. elegans* male sexual development and behavior.** *Development* 2000, **127**:4469-4480.
26. Ausubel FM: *Current Protocols in Molecular Biology*. New York: Greene Publishing, Associates/Wiley-Interscience; 1988.