So, the question remains, which of the available techniques for expression pattern analysis most accurately reflects biological reality? The best answer may be 'all of the above'. No one technique necessarily gives the whole picture. Each can provide qualitatively different types of information, and all can help in the quest to understand biological systems.

References

1. Cheng H-J, Flanagan JG: **Identification and cloning of ELF-1, a developmentally expressed ligand for the Mek4 and Sek receptor tyrosine kinases**. *Cell* 1994, **79:**157-168.
2. Cheng H-J, Nakamoto M, Bergemann AD, Flanagan JG: **Complementary gradients in expression and binding of ELF-1 and Mek4 in development of the topographic retinotectal projection map**. *Cell* 1995, **82:**371-381.
3. Flanagan JG, Cheng HJ, Feldheim D, Hattori M, Lu Q, Vanderhaeghen P: **Alkaline phosphatase fusions of ligands or receptors as in situ probes for staining of cells, tissues and embryos**. *Meth Enzymol,* in press.
4. Sobieszczuk DF, Wilkinson DG: **Masking of Eph receptors and ephrins**. *Curr Biol* 1999, **9:**R469-R470.

6. Harlow E, Lane D: *Antibodies: a laboratory manual.* Cold Spring Harbor Laboratory, New York: Cold Spring Harbor Laboratory Press; 1988.
7. Skarnes WC: **The identification of new genes: gene trapping in transgenic mice**. *Curr Opin Biotech* 1993, **4:**684-689.
8. Flanagan JG, Leder P: **The *kit* ligand: a cell surface molecule altered in Steel mutant fibroblasts.** *Cell* 1990, **63:**185-194.
9. Aruffo A, Stamenkovic I, Melnick M, Underhill CB, Seed B: **CD44 is the principal cell surface receptor for hyaluronate.** *Cell* 1990, **61:**1303-1313.
10. Gale NW, Holland SJ, Valenzuela DM, Flenniken A, Pan L, Ryan TE, Henkemeyer M, Strebhardt K, Hirai H, Wilkinson DG, *et al.*: **Eph receptors and ligands comprise two major specificity subclasses and are reciprocally compartmentalized during embryogenesis**. *Neuron* 1996, **17:**9-19.
11. Hornberger MR, Dutting D, Ciossek T, Yamada T, Handwerker C, Lang S, Weth F, Huf J, Wessel R, Logan C, *et al.*: **Modulation of EphA receptor function by coexpressed ephrinA ligands on retinal ganglion cell axons**. *Neuron* 1999, **22:**731-742.

Address: Department of Cell Biology, and Program in Neuroscience, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.
E-mail: flanagan@hms.harvard.edu

The STAS domain — a link between anion transporters and antisigma-factor antagonists

L. Aravind*† and Eugene V. Koonin†

Anion transporters of the sulfate transporter family are of major interest, as their malfunction is implicated in three human diseases: diastrophic dysplasia/achondrogenesis type IB (DTD) [1,2], Pendred's syndrome (PDS) [3] and congenital chloride diarrhoea (CLD) [4]. The CLD gene is also downregulated in intestinal adenomas and adenocarcinomas [5]. The products of these genes are distinct but related anion transporters that contain 12 transmembrane helices followed by a cytoplasmic domain at the carboxyl

a potential iodide-chloride transporter [7] and the CLD gene a chloride–$NaHCO_3^-$ exchanger [8]. Related transporters with the same domain organization, mainly involved in sulfate transport, are present in other eukaryotes and in many bacteria [9].

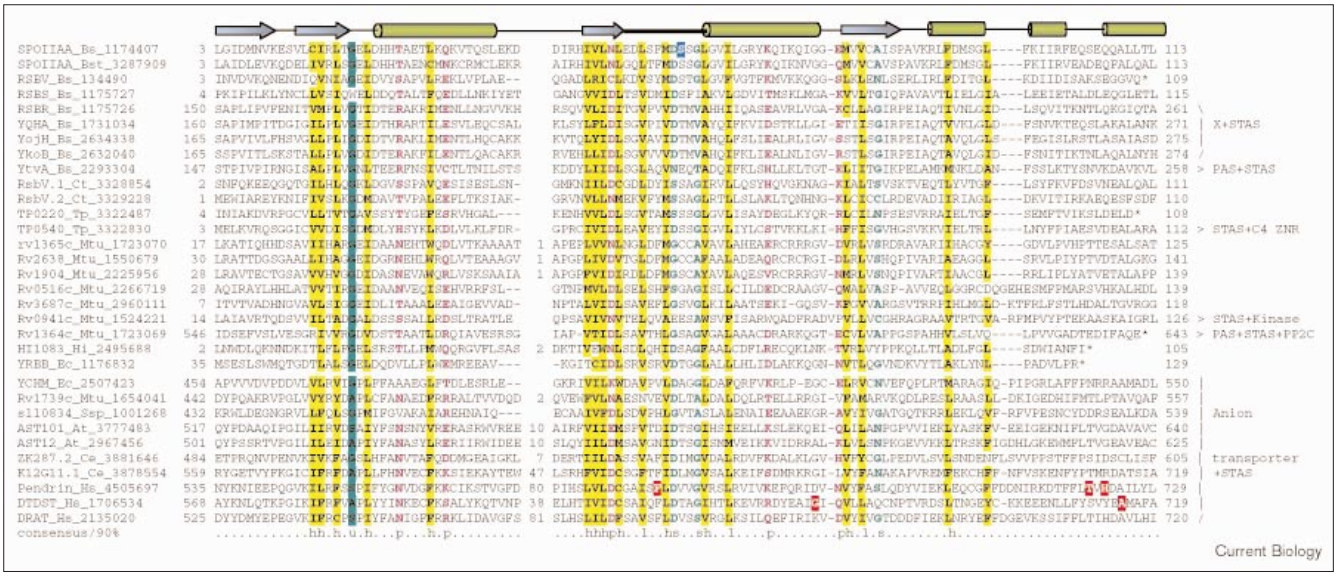
We describe here an unexpected, statistically significant similarity between the carboxy-terminal cytoplasmic domains of these transporters and the bacterial antisigma-factor antagonists (ASA) typified by *Bacillus subtilis* SPOIIAA. In a PSI-BLAST search [10,11] seeded with the SPOIIAA sequence from *Bacillus stearothermophilus*, with a profile inclusion cut-off of 0.01, the carboxy-terminal domain of the human disease-associated transporters and their eukaryotic and bacterial homologs were detected with random expectation (*E*) values of 10^{-3}–10^{-4} within five iterations. In reciprocal searches initiated with the CLD transporter carboxy-terminal domain, bacterial ASAs were detected with *E* values of 10^{-4}–10^{-6} in the third iteration.

The nuclear magnetic resonance structure of the ASA SPOIIAA [12] provides the structural framework for the emerging domain superfamily. A multiple alignment of this superfamily was constructed using the CLUSTALW program [13] and adjusted using the PSI-BLAST results (Figure 1). The alignment of the carboxy-terminal domains of anion transporters was used for secondary structure prediction-based threading through the Protein Databank (PDB) database using the PHD-TOPITS program [14]. The best hit was the PDB entry for SPOIIAA (PDB code 1AUZ), with a Z-score of 3.1, which strongly suggests a structure similar to that of SPOIIAA. Thus, ASAs and the cytoplasmic portions of anion transporters define a previously undetected, ancient, conserved domain that we named STAS after

By mapping the conserved motifs apparent from the multiple alignment (Figure 1) onto the SPOIIAA structure (Figure 2), the conservation was traced largely to the four strands that form the scaffold of the STAS domain. In addition, the turn between the two amino-terminal strands and the long loop between strand 3 and helix 2 are strongly conserved and inserts appear not to be tolerated in these elements (Figures 1,2). Most of the variability is in the loop between helix 1 and strand 3, with α-helical inserts of considerable size seen in some of the anion transporters (Figure 1). A comparison of the alignment with the tertiary structure shows that the carboxy-terminal region of the STAS domain forms a characteristic α-helical handle-like structure (Figures 1,2).

The identification of the STAS domain in the ASAs and the anion transporters provides functional clues for the regulation of anion

Figure 1



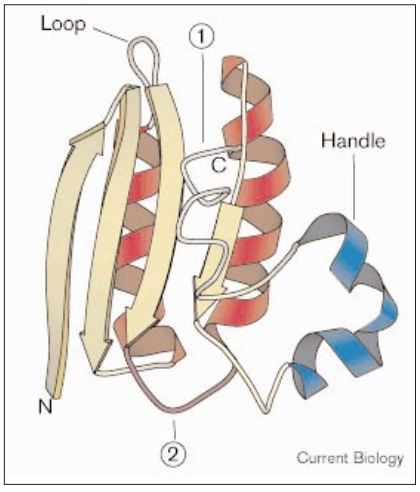
Multiple alignment of a selected set of STAS domains from antisigma-factor antagonists and anion transporters. The proteins are named according to their systematic gene name followed by the species abbreviation and the GenBank identifier number. DTDST is the sulfate channel mutated in DTD and DRAT is the down-regulated in adenoma transporter that is mutated in CLD. The residues affected by mutations in PDS and DTD as mentioned in the text are highlighted in red and the phosphorylatable serine in SPOIIAA is highlighted in blue. The coloring of conserved positions in the alignment is according to the 90% consensus: hydrophobic (h, YFWLIVMA) and aliphatic (l, LIVMA) residues highlighted in yellow, small residues (s, SAGTVPNHD) colored green, tiny residues (u, GAS) shaded green and polar residues (p, STQNEDRKH) colored pink. The secondary structure elements are derived from the structure of SPOIIAA and are indicated above the alignment; cylinders represent α-helices and arrows represent β-sheets. The thick line corresponds to the conserved loop and the broken line indicates the variable loop; in the latter, the sequences of the inserts are replaced by the number of amino acid residues (see also Figure 2). The numbers before and after the alignment indicate the positions of the first and last residue of the STAS domain in each of the aligned proteins. The carboxyl terminus of each protein is indicated by an asterisk where applicable. The domain organization of those proteins that have domains fused to STAS is indicated to the right of the alignment following brackets delineating the respective sets of proteins. All the remaining proteins contain the STAS domain alone, similar to the antisigma factor antagonists. X is an uncharacterized domain seen only in Gram-positive bacteria; the kinase domain fused to the STAS domain in a mycobacterial protein is specifically related to the antisigma factor-like serine kinases with a histidine-kinase fold; PP2C is a phosphatase domain. The species abbreviations are: Bs, *Bacillus subtilis*; Bst, *Bacillus stearothermophilus*; Ct, *Chlamydia trachomatis*; Ec, *Escherichia coli*; Hi, *Haemophilus influenzae*; Tp, *Treponema pallidum*; Mtu, *Mycobacterium tuberculosis*; Ssp, *Synechocystis species*; Ce, *Caenorhabditis elegans*; Hs, *Homo sapiens*; At, *Arabidopsis thaliana*. A more complete alignment of STAS domain sequences is available from: ftp://ncbi.nlm.nih.gov/pub/koonin/STAS.

transport. ASA-like proteins have been found in a variety of bacteria, such as Gram-positive bacteria, Actinomycetes, Cyanobacteria, chlamydiae, *Treponema* and *Thermotoga*. These proteins positively regulate sigma factors by interacting with the anti-sigma factor — a protein kinase. This kinase, in turn, phosphorylates the ASA on a serine in the conserved loop (Figure 2) and thus inactivates it; the ASA can be re-activated through dephosphorylation by a phosphatase [15]. Thus, the STAS domain is at the center of protein–protein interactions in the sigma-factor regulation network.

It has been shown that SPOIIAA binds GTP and ATP and possesses a weak NTPase activity that is abolished by phosphorylation or by mutation of the phosphorylatable serine in the conserved loop (Figure 1) [16]. The strong conservation of this loop in the STAS domains (Figure 1) suggests that this domain could possess general NTP-binding activity. The conserved loop is probably involved in phosphate-binding and the β-sheet scaffold could accommodate the rest of the NTP molecule. This mode of ligand binding resembles lipid binding by Sec14 domains, which have the same structural fold as SPOIIAA but show no detectable sequence similarity to the STAS domain [17,18]. Notably, one of the mutations in PDS has been mapped to the predicted phosphate-binding loop and probably results in its disruption (Figures 1,2).

The presence of a predicted NTP-binding domain in the cytoplasmic portions of anion transporters indicates that anion transport could be regulated by intracellular concentrations of GTP

Figure 2



Structural features of the STAS domain. The ribbon diagram was constructed from the minimized average NMR structure of SPOIIAA (PDB code 1BUZ) using the Molscript program [21]. The regions of low sequence conservation are colored grey, including the loop that accommodates large inserts in the anion transporters. The conserved loop that is implicated in nucleotide binding is colored purple. The distinctive α-helical handle – a structural feature that is seen in most members of this superfamily and shared with the Sec14 superfamily – is indicated in blue. 1, poorly conserved distorted carboxy-terminal helix; 2, conserved loop; Loop, loop with variable inserts; Handle, α-helical handle; N, amino terminus; C, carboxyl terminus.

and/or ATP. The NTPs are likely to elicit specific conformational changes in the STAS domain through binding and/or hydrolysis. The critical role of the STAS domain in anion transporters is supported by a number of mutations in PDS and DTD that map to this domain [3,19,20]. Experimental testing of these predictions, for which bacterial transporters with the same domain architecture could serve as a model, should clarify the regulation of these important transporters, which appears to be more complex than previously suspected. We detected ASA-like STAS domain proteins in some bacteria (for example *Escherichia coli*) that lack the typical sigma regulatory system and new members in Gram-positive bacteria that are fused to another ligand-binding domain — the PAS domain (Figure 1). These STAS proteins could represent new bacterial regulatory systems.

Acknowledgements

We thank Michael Yudkin for helpful discussions.

References

1. Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, *et al.*: **The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping**. *Cell* 1994, **78:**1073-1087.
2. Superti-Furga A, Hastbacka J, Wilcox WR, Cohn DH, van der Harten HJ, Rossi A, Blau N, *et al.*: **Achondrogenesis type IB is caused by mutations in the diastrophic dysplasia sulphate transporter gene**. *Nat Genet* 1996, **12:**100-102.
3. Everett LA, Glaser B, Beck JC, Idol JR, Buchs A, Heyman M, *et al.*: **Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS)**. *Nat Genet* 1997, **17:**411-422.
4. Hoglund P, Haila S, Socha J, Tomaszewski L, Saarialho-Kere U, Karjalainen-Lindsberg ML, *et al.*: **Mutations of the Down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea**. *Nat Genet* 1996, **14:**316-319.
5. Schweinfest CW, Henderson KW, Suster S, Kondoh N, Papas TS: **Identification of a colon mucosa gene that is down-regulated in colon adenomas and adenocarcinomas**. *Proc Natl Acad Sci USA* 1993, **90:**4166-4170.
6. Satoh H, Susaki M, Shukunami C, Iyama K, Negoro X, Hiraki Y: **Functional analysis of diastrophic dysplasia sulfate transporter. Its involvement in growth regulation of chondrocytes mediated by sulfated proteoglycans**. *J Biol Chem* 1998, **273:**12307-12315.
7. Scott DA, Wang R, Kreman TM, Sheffield VC, Karnishki LP: **The Pendred syndrome gene encodes a chloride-iodide transport protein**. *Nat Genet* 1999, **21:**440-443.
8. Melvin JE, Park K, Richardson L, Schultheis PJ, Shull GE: **Mouse down-regulated in adenoma (DRA) is an intestinal Cl(–)/HCO(3)(–) exchanger and is up-regulated in colon of mice lacking the NHE3 Na(+)/H(+) exchanger**. *J Biol Chem* 1999, **274:**22855-22861.
9. Smith FW, Ealing PM, Hawkesford MJ, Clarkson DT: **Plant members of a family of sulfate transporters reveal functional subtypes**. *Proc Natl Acad Sci USA* 1995, **92:**9373-9377.
10. Altschul SF, Koonin EVP: **PSI-BLAST – a tool for making discoveries in sequence databases**. *Trends Biochem Sci* 1998, **23:**444-447.
11. Altschul S, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25:**3389-3402.
12. Kovacs H, Comfort D, Lord M, Campbell ID, Yudkin MD: **Solution structure of SpoIIAA, a phosphorylatable component of the system that regulates transcription factor sigmaF of Bacillus subtilis**. *Proc Natl Acad Sci USA* 1998, **95:**5067-5071.
13. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22:**4673-4680.
14. Rost B, Schneider R, Sander C: **Protein fold recognition by prediction-based threading**. *J Mol Biol* 1997, **270:**471-480.
15. Duncan L, Alper S, Losick R: **SpoIIAA governs the release of the cell-type specific transcription factor sigma F from its anti-sigma factor SpoIIAB**. *J Mol Biol* 1996, **260:**147-164.
16. Najafi SM, Harris DA, Yudkin MD: **The SpoIIAA protein of Bacillus subtilis has GTP-binding properties**. *J Bacteriol* 1996, **178:**6632-6634.
17. Aravind L, Neuwald AF, Ponting CP: **Sec14p-like domains in NF1 and Dbl-like proteins indicate lipid regulation of Ras and Rho signaling**. *Curr Biol* 1999, **9:**R195-R197.
18. Hubbard TJ, Ailey B, Brenner SE, Murzin AG, Chothia C: **SCOP: a structural classification of proteins database**. *Nucleic Acids Res* 1999, **27:**254-256.
19. Hastbacka J, Superti-Furga A, Wilcox WR, Rimoin DL, Cohn DH, Lander ES: **Atelosteogenesis type II is caused by mutations in the diastrophic dysplasia sulfate-transporter gene (DTDST): evidence for a phenotypic series involving three chondrodysplasias**. *Am J Hum Genet* 1996, **58:**255-262.
20. Superti-Furga A, Rossi A, Steinmann B, Gitzelmann R: **A chondrodysplasia family produced by mutations in the diastrophic dysplasia sulfate transporter gene: genotype/phenotype correlations**. *Am J Med Genet* 1996, **63:**144-147.
21. Kraulis P: **A program to produce both detailed and schematic plots of proteins**. *J Appl Crystallogr* 1991, **24:**946-950.

*Department of Biology, Texas A&M University, College Station, Texas 77843, USA. †National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA.
E-mail: koonin@ncbi.nlm.nih.gov