Update

Techniques & Applications

An engineer's view on regulation of seed development

Astrid Junker[1], Anja Hartmann[1], Falk Schreiber[1,2] and Helmut Bäumlein[1]

[1] Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben (IPK), Corrensstrasse 3, D-06466 Gatersleben, Germany
[2] Institute of Computer Science, Martin Luther University Halle-Wittenberg, Von-Seckendorff-Platz 1, D-06120 Halle, Germany

The recently proposed Systems Biology Graphical Notation (SBGN) represents a flexible system of nomenclature for the description of biological networks, comparable to the notation employed by designers of electronic circuits. It allows the uniform and unambiguous display of complex biological information. Here we present an application of SBGN to describe processes occurring during seed development in *Arabidopsis thaliana*. Representative network maps can be accessed via the open resource RIMAS web portal.

Circuit maps in biology

A uniform system of nomenclature describing the components of networks and based on a well-defined set of symbols, is well established within the engineering field, as it greatly facilitates communication efficiency and clarity. In biology, equivalent notation systems have been suggested [1–3], but have as yet failed to become accepted as a common standard due to several shortcomings. A concerted effort mounted by a major consortium of biological network specialists has now proposed the 'Systems Biology Graphical Notation' (SBGN, see Glossary and www.sbgn.org) scheme [4]. SBGN incorporates a limited number of 33 easily recognizable glyphs, and can be used to represent a broad range of biological networks, including metabolic pathways and transcriptional regulatory circuits. It can incorporate structural, non-dynamic information, and thereby simplifies further network analyses (e.g. based on data mapping). A range of supporting software tools has also been developed (www.sbgn.org). To accommodate different informational aspects of the same network and to permit network visualization at various levels of granularity, SBGN allows for the use of three independent languages: 'Process Description' (PD), 'Entity Relationship' (ER)' and 'Activity Flow' (AF). The PD view describes a sequence of processes as transitions of entities from one state to another. In contrast, the ER language focuses on the interactions of entities without considering their temporal course. Finally, the AF notation allows the conceptual visualization of information fluxes (such as modulation, inhibition or stimulation) between entities. Thus, SBGN provides a uniform basis for the representation of complex biological systems, and has the potential to be widely accepted and routinely applied.

To demonstrate the performance of SBGN, we show here an application based on the PD language, to represent much of the current understanding of the regulatory processes underpinning seed development in *Arabidopsis* (*Arabidopsis thaliana*) [5–7]. Network maps can be accessed online from the RIMAS (Regulatory Interaction Maps of *Arabidopsis* Seed Development) portal (see Box 1), which focuses on hierarchical transcriptional regulation, and hormonal and epigenetic factors. As an open resource it requires and allows extension in all conceivable directions.

A general representation of master hierarchies

The LEC1/AFL-B3 network

The LEC1/AFL-B3 network (Figure 2, http://rimas.ipk-gatersleben.de) involves at least four key regulators of seed development, namely the CCAAT-box binding factor LEAFY COTYLEDON1 (LEC1), and the three B3 domain-containing proteins ABSCISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3) and LEAFY COTYLEDON2 (LEC2). The network has been derived from a large number of genetic, molecular genetic and phenotypic analyses of relevant mutant lines [8]. LEC1 is thought to function upstream of LEC2, FUS3 and ABI3 [8,9], while LEC2 controls both FUS3 and ABI3 [8,10], as well as triggering a feedback loop acting on LEC1 [11]. Autoregulatory feedback loops ensure the homeostatic expression of *FUS3* and *ABI3* during seed maturation, when the expression of both *LEC1* and *LEC2* ceases [8,10]. We generated a LEC1/AFL-B3 network map (Figure 2a) hiding

Glossary (Key SBGN vocabulary)

AF: Activity Flow Language.
Clone marker: Multiple occurrences of an EPN within a network.
Conceptual EPN: EPNs not corresponding to specific structures in the network (e.g. source/sink, unspecified entity).
Conceptual type (ct): Supplementary information regarding the function of an EPN in a process (e.g. 'nucleic acid features' with different cts, such as gene or mRNA).
Connecting arc: Lines connecting EPNs and PNs (e.g. 'consumption' and 'production'), as well as the influences of an effector on a given process (e.g. 'stimulation', 'necessary stimulation', 'modulation' and 'inhibition').
Container node (CN): defined sets of EPNs (e.g. compartment).
Entity pool node (EPN): Pools of uniform substrates, products or effectors of processes.
ER: Entity Relationship Language.
Logical operator: Combiner of several effectors on a given process ('AND', 'OR', 'NOT').
Material EPN: EPNs specified by their structures (e.g. 'macromolecules' and 'nucleic acid features').
Material type (mt): Supplementary information regarding the chemical structure of an EPN in a process (e.g. 'nucleic acid features' of different material types, such as DNA or RNA).
PD: Process Description language.
Process node (PN): The transition of an EPN.
RIMAS: Regulatory Interaction maps of *Arabidopsis* Seed Development.
SBGN: Systems Biology Graphical Notation.
State variable: EPN label giving supplementary information regarding the configuration, physical state or chemical modification of the EPN.
Unit of information: EPN label providing supplementary information other than EPN state.

Corresponding author: Junker, A. (junkera@ipk-gatersleben.de).

Box 1. The RIMAS information portal

RIMAS (Regulatory Interaction Maps of *Arabidopsis* Seed Development) has been established as a web-based portal. Access is at http://rimas.ipk-gatersleben.de. The developers seek feedback from scientific colleagues to extend the knowledge base underpinning RIMAS. The portal provides three main features:

- Four detailed SBGN-based network maps which describe interactions between LEC1/AFL-B3 transcription factors and maturation gene promoter elements, hormonal pathways and epigenetic processes.
- Links to PubMed literature database (http://www.ncbi.nlm.nih.gov/pubmed/), corresponding TAIR gene entries (http://www.arabidopsis.org) and to prior reviews covering seed development in *Arabidopsis*.
- The ready export of maps in common exchange formats (such as the Graph Modelling Language, GML), allowing for the modification of network layouts to suit users' requirements, and further application using tools such as VANTED [23] and others.

a number of mechanistic details (e.g. marginally significant sub-processes such as transcription and translation). The molecular interactions between LEC1/AFL-B3 factors (e.g. direct versus indirect regulation) are not yet understood in every detail. The LEC1/AFL-B3 network exemplifies the basic structure and function of SBGN representations based on the PD language (Figure 1); this language was chosen because it reflects mechanistic processes well, while at the same time being flexible enough to accommodate whatever level of detail has been exposed by a specific set of experiments. Figure 1 displays PD glyphs as used in the exemplary maps which will be presented in the following sections. Corresponding collections and examples of ER and AF notations can be found at www.sbgn.org.

By analogy with biochemical reactions, regulatory processes can be defined by three major components: one or

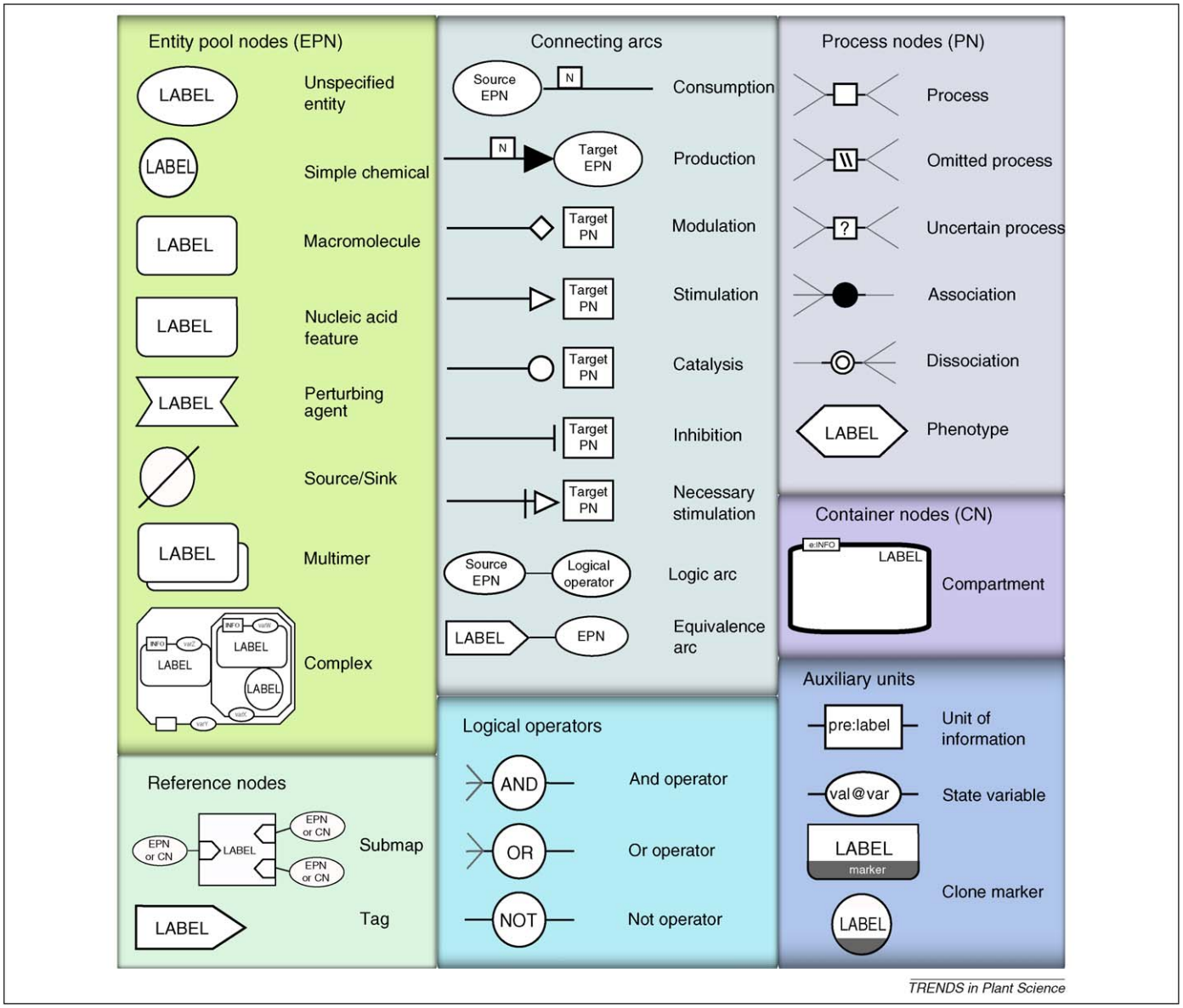

Figure 1. The SBGN Process Description (PD) notation. This overview represents all SBGN glyphs employed by the PD language. Glyphs are classified into the following categories: 'Entity Pool Nodes (EPN)', 'Process Nodes (PN)' and 'Connecting Arcs' as basic elements of a process; 'Reference Nodes' and 'Container nodes (CN)' as means of encapsulation of EPNs, 'Auxiliary Units' as annotations of EPNs and 'Logical Operators' as means of combining effects of different EPNs. (For description see Glossary).

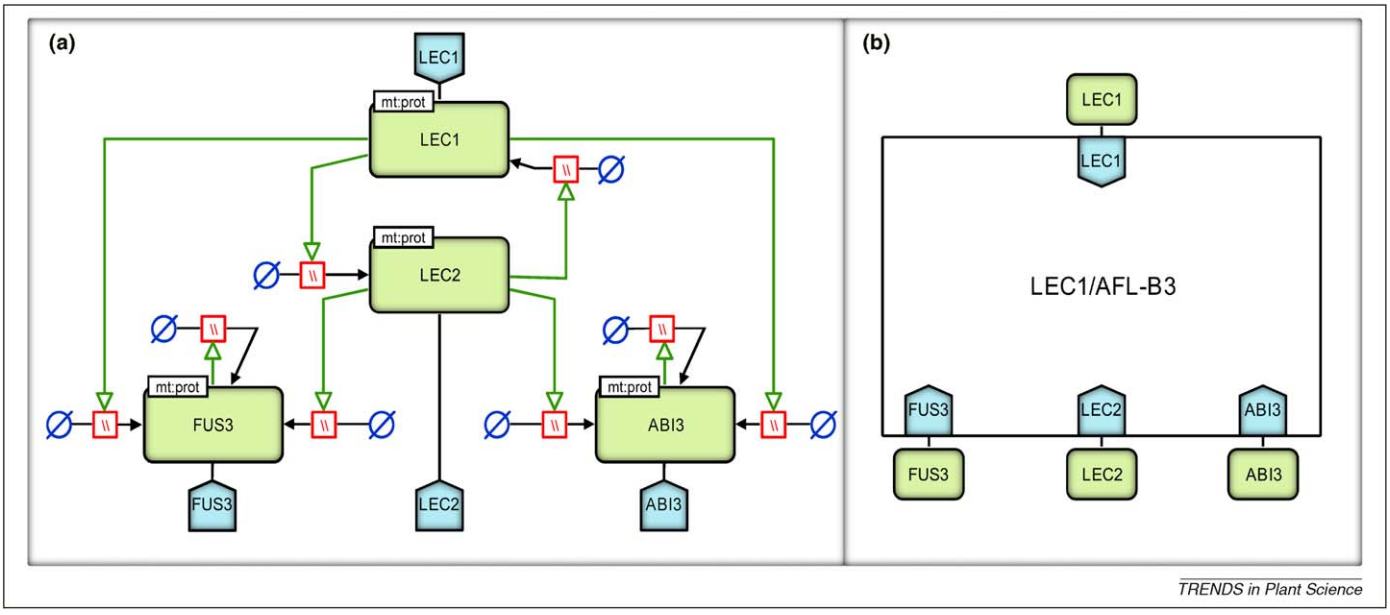

Figure 2. The LEC1/AFL-B3 network. **(a)** LEC1, LEC2, FUS3 and ABI3 represent four master regulators of seed development in *Arabidopsis*, with LEC1 and LEC2 each controlling the respective three remaining factors. FUS3 and ABI3 operate as auto-regulatory feedback loops. The LEC1/AFL-B3 proteins stimulate the formation of a particular factor from an unspecified source. Green filled – 'macromolecules' (in this case proteins); Blue – 'Unspecified source' (i.e. all substrates required for the synthesis of a particular protein); Green – connecting arc 'stimulation'; Red – process node 'omitted process' (represents sub-processes such as transcription, translation and the action of intermediate factors which do not require to be specified or are unknown); Blue filled – 'tags' (represent input/output terminals of a network related to the 'submap' glyph). **(b)** Depending on the level of detail required, it is possible to encapsulate all processes of a network using the 'submap' glyph. Process nodes and connecting arcs are hidden and only input/output terminals are represented.

more sources (substrates), a process node and one or more products. Both substrates and products are 'entity pool nodes' (EPNs), which could be either material entities (such as a macromolecule or a nucleic acid feature) or conceptual entities (an unspecified source) (Figure 1). An 'unspecified source' glyph (Figure 2a, blue) is particularly useful in the representation of regulatory processes, which require a large number of substrates such as trinucleotides for transcription, and activated amino acids for translation. The use of a single symbol combining all the non-specified substrates together represents a convenient strategy, in continuance of which non-relevant sub-processes can be hidden by assuming an 'omitted process' node (Figure 2a, red). Where auto-regulatory loops are involved (such as in *ABI3* and *FUS3* expression), these omitted processes would include transcription and translation, since both the substrates and products of these processes are proteins which can be represented by a 'macromolecule' glyph (Figure 1). Macromolecule EPNs (Figure 2a, green filled) can be associated with additional units of information, as indicated by supplementary labelling (see 'Auxiliary units' in Figure 1). Thus, the chemical structure of an EPN can be indicated by the provision of additional information, in particular the 'material type' (mt) of these macromolecules. Transcription factor EPNs, which also act as effectors of other processes, are linked to their target process nodes by connecting arcs (Figures 1 and 2). The 'stimulation arc' (Figure 2a, green) represents a positive effect of an EPN on the flux of the target process. For example, LEC1 stimulates omitted processes which result in the formation of the ABI3 protein. Other connecting arcs ('modulation', 'inhibition', 'catalysis' and 'necessary stimulation') are used to describe the influence of an EPN(s) on a target process(es) (Figure 1). To further simplify the LEC1/AFL-B3 overview map, some processes can be encapsulated by the use of a 'submap' glyph (see 'Reference nodes' in Figure 1). The submap only displays input and output terminals for all four LEC1/AFL-B3 proteins (Figure 2a and b, blue filled). These have to be defined in the original map by connecting 'tag' glyphs (Figure 2a, blue filled) with the same label as the EPN via 'equivalence arcs' (Figure 1). The 'submap' glyph thus provides a suitable tool to integrate partially repetitive components into larger networks.

A detailed representation of varying levels of knowledge

LEC1/AFL-B3 factors in maturation gene control

RIMAS comprises two maps summarizing our current understanding of the LEC1/AFL-B3 regulon (http://rimas.ipk-gatersleben.de). Candidate genes have been divided into those associated with the acquisition of desiccation tolerance, dormancy and storage compound accumulation; phytohormone signalling was focused mainly on gibberellic and abscisic acid, and epigenetic pathways were included. The maps are based on experimental data derived from a combination of single and multi-gene analyses, as well as from chromatin-immunoprecipitation experiments giving the final proof for direct gene regulation. The flexibility of SBGN is illustrated in three independent examples (Figure 3).

Example #1

ABI3 complexes with the bZIP transcription factor ABI5, before being able to interact with the *AtEM6* promoter [12–15] (Figure 3a). This direct regulation is represented as a complex between the *upstream* regulators ABI3 and ABI5 and the *downstream AtEM6* promoter (Figure 3a, green). *Upstream* regulator nodes are macromolecules (proteins)

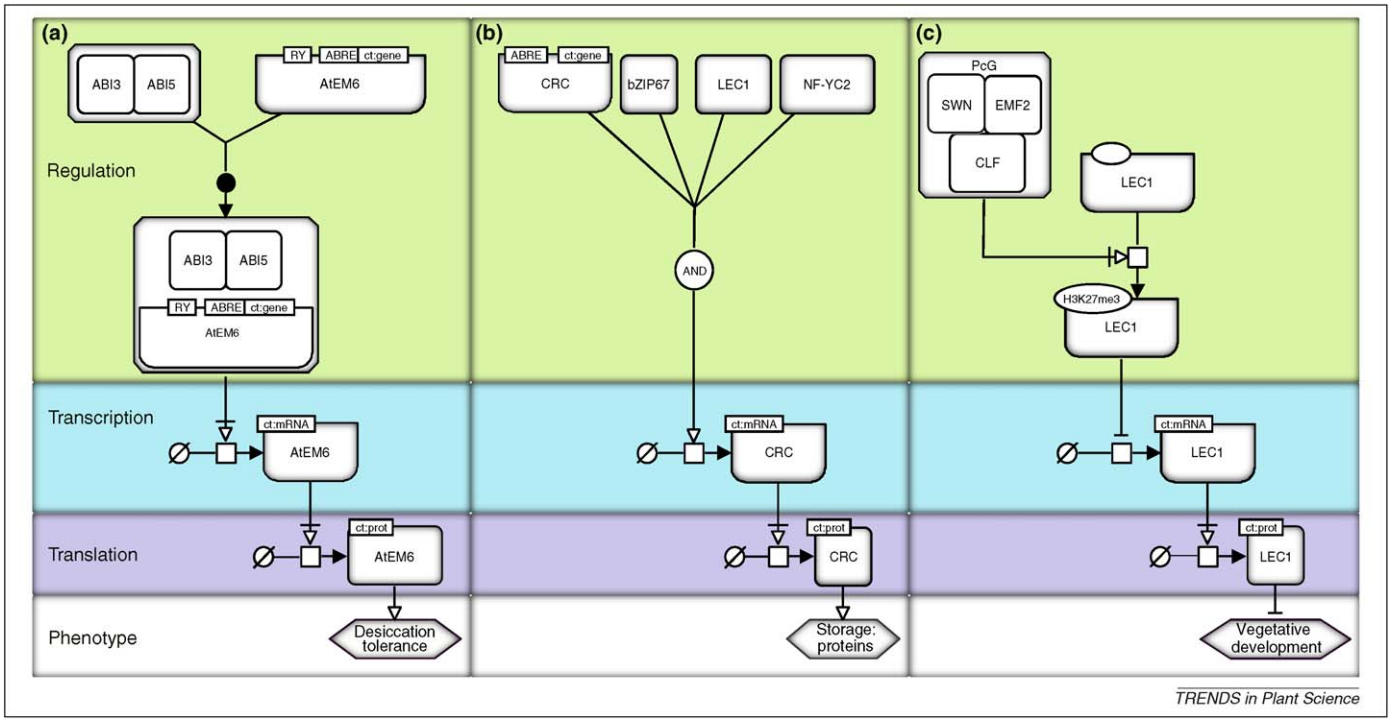

Figure 3. Representation of varying knowledge levels and regulatory mechanisms connected to LEC1/AFL-B3 factors. A regulatory process includes 'regulation' (green), 'transcription' (blue), 'translation' (purple) and 'phenotype' (white). **(a)** Direct regulation of *AtEM6* by ABI3 and ABI5, which complex with the *AtEM6* sequence, and drive the transcription and translation of the AtEM6 protein (associated with the acquisition of desiccation tolerance). **(b)** LEC1, bZIP67 and NF-YC2, and the *CRC* together drive the transcription of *CRC*. The *CRC* product is an important seed storage protein in *Arabidopsis*. **(c)** A complex of the PcG proteins EMF2, CLF and SWN is responsible for trimethylation of histone H3 (H3K27me3) at the *LEC1* locus thereby preventing transcription of *LEC1*. Thus, the LEC1-mediated inhibition of vegetative development is released and the plant life cycle is able to proceed.

whereas the *downstream* gene promoter is given as a 'nucleic acid feature'. The function of EPNs of the same type (here the 'nucleic acid features') might differ in the context of a given process. This is indicated using auxiliary units with the information about the 'conceptual type' of this EPN which could be 'ct:gene' or 'ct:mRNA'. Further information with respect to transcription factor binding sites (TFBS) can also be given by supplementary EPN labels. B3 domain proteins such as ABI3 bind the RY motif [16], whereas bZIP factors such as ABI5 recognize the abscisic acid response element (ABRE) [13]. The interaction (represented by the 'association' glyph) between the complex of the two regulators and the TFBS present in a given promoter triggers transcription (Figure 3a, blue). The connecting arc 'necessary stimulation' indicates that stimulation by the regulator–target-complex is necessary for transcription to proceed. The *AtEM6* transcript is represented by a 'nucleic acid feature' with the 'ct:mRNA' which triggers its own translation (Figure 3a, purple). Finally, the 'phenotype' glyph is used to indicate that the AtEM6 protein contributes to the acquisition of desiccation tolerance (Figure 3a, white).

Example #2

This example concerns the activation of the *cruciferin* (*CRC*) gene promoter by LEC1, bZIP67 and NF-YC2 [17]. Transient assays have shown that expression of these three transcription factors activates various promoters, and the assumed heterotrimeric complex has been suggested for the regulation of *CRC* [17], even though no supportive experimental data exist about the interaction of LEC1 and bZIP67. *CRC* expression is upregulated in parallel with *LEC1* expression [9,18]. This level of understanding of a regulatory process can be indicated by the inclusion of the logical operator 'AND' (Figure 3b, green), which reflects the involvement of *upstream* regulators such as LEC1, bZIP67 and NF-YC2 'AND' the *CRC* sequence itself in the stimulation of *CRC* transcription (Figure 2b, blue). This scenario contrasts with the gene/transcription factor complexes implicated in direct gene regulation (Figure 3a). The use of the logical operator 'AND' does not exclude intermediate factors from participating in the regulatory cascade. Therefore, the role of LEC1, bZIP67 and NF-YC2 is depicted as 'stimulation' (Figure 3b) rather than 'necessary stimulation'. By analogy with example #1, the *CRC* transcript triggers its own translation (Figure 3b, purple) and the resulting gene product constitutes one of the major storage proteins in *Arabidopsis* seeds (Figure 3b, white).

A representation of regulatory mechanisms beyond transcriptional regulation

The epigenetic control of LEC1/AFL-B3 factors

Besides transcription-factor-mediated regulation, gene regulation can also be influenced by epigenetic chromatin remodelling, or by the presence of small RNAs which act to degrade transcripts in a sequence-specific manner. RIMAS includes a fourth map which depicts the action of *upstream* regulators on LEC1/AFL-B3 factors (http://rimas.ipk-gatersleben.de). Most of these factors suppress the seed transcriptional program during vegetative plant development, which is necessarily required in order to complete the plant life cycle. The regulation of *LEC1* transcription by the chromatin remodelling factor PICKLE (PKL) exemplifies

chromatin-based regulation of transcription. PKL is a negative regulator of *LEC1* [18], and decreased levels of trimethylation of the lysine residues present at position 27 in the histone H3 (H3K27me3) at the *LEC1* locus were found in the *pkl* mutant [19]. PKL repression is mediated by Polycomb group proteins (PcG) which share some similarity with components of the *Drosophila melanogaster* Polycomb Repressive Complex 2 (PRC2). The PcG genes *SWINGER* (*SWN*) and *EMBRYONIC FLOWER2* (*EMF2*) are direct targets of PKL [20]. The complex of SWN, EMF2 and possibly CURLY LEAF (CLF), then trimethylates the critical lysine residue of histone H3 [20,21], and the corresponding methylated and demethylated states of the *LEC1* locus are depicted using the auxiliary unit 'state variable' at the EPN border (Figure 3c). The 'state variable' contains information relating to the configuration and/or chemical modification of the EPN. Because H3K27me3 suppresses transcription, the methylated *LEC1* locus inhibits transcription of *LEC1*, and subsequently translation, leading to the formation of the LEC1 protein (Figure 3c). *LEC1* expression is incompatible with vegetative development [22] which, in order to proceed, requires the active suppression of this gene by factors such as PKL and PcG proteins. Other regulatory processes which depend on DNA modification (such as acetylation) can be visualized in analogous fashion.

Conclusions

SBGN offers an efficient, flexible and clear nomenclature system for biological networks. To demonstrate its performance, we have applied it to illustrate major regulatory processes involved in embryogenesis and seed formation. Our intention is to use the RIMAS web portal as a means of demonstrating the general utility of SBGN, and we specifically invite the seed biology research community to treat RIMAS as a platform for discussion and exchange. We plan to advance and update RIMAS on a regular basis to improve its functionality as an information resource and a showcase for SBGN.

Acknowledgements

We thank Christian Klukas and Tobias Czauderna for their advice regarding the use of VANTED [23] and SBGN-ED tools (http://vanted.ipk-gatersleben.de/addons/sbgn-ed) which were used for map drawing, Björn Junker for his helpful suggestions, and Gudrun Mönke for her critical reading of the manuscript.

References

1 Kohn, K.W. (2001) Molecular interaction maps as information organizers and simulation guides. *Chaos* 11, 84–97
2 Kohn, K.W. *et al.* (2006) Molecular interaction maps of bioregulatory networks: A general rubric for systems biology. *Mol. Biol. Cell* 17, 1–13
3 Kitano, H. (2003) A graphical notation for biochemical networks. *BioSilico* 1, 169–176
4 Le Novere, N. *et al.* (2009) The Systems Biology Graphical Notation. *Nat. Biotechnol.* 27, 735–741
5 Gutierrez, L. *et al.* (2007) Combined networks regulating seed maturation. *Trends Plant Sci.* 12, 294–300
6 Santos-Mendoza, M. *et al.* (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant J.* 54, 608–620
7 Verdier, J. and Thompson, R.D. (2008) Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.* 49, 1263–1271
8 To, A. *et al.* (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell* 18, 1642–1651
9 Kagaya, Y. *et al.* (2005) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 46, 399–406
10 Kroj, T. *et al.* (2003) Regulation of storage protein gene expression in Arabidopsis. *Development* 130, 6065–6073
11 Stone, S.L. *et al.* (2008) Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3151–3156
12 Nakashima, K. *et al.* (2006) Transcriptional regulation of ABI3-and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of Arabidopsis. *Plant Mol. Biol.* 60, 51–68
13 Carles, C. *et al.* (2002) Regulation of *Arabidopsis thaliana Em* genes: role of ABI5. *Plant J.* 30, 373–383
14 Nakamura, S. *et al.* (2001) Physical interactions between ABA response loci of Arabidopsis. *Plant J.* 26, 627–635
15 Bies-Etheve, N. *et al.* (1999) Importance of the B2 domain of the Arabidopsis ABI3 protein for Em and 2S albumin gene regulation. *Plant Mol. Biol.* 40, 1045–1054
16 Mönke, G. *et al.* (2004) Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta* 219, 158–166
17 Yamamoto, A. *et al.* (2009) Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J.* 58, 843–856
18 Rider, S.D. *et al.* (2003) Coordinate repression of regulators of embryonic identity by PICKLE during germination in Arabidopsis. *Plant J.* 35, 33–43
19 Zhang, H. *et al.* (2008) The CHD3 remodeler PICKLE promotes trimethylation of histone H3 lysine 27. *J. Biol. Chem.* 283, 22637–22648
20 Aichinger, E. *et al.* (2009) CHD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. *PLOS Genet.* 5, e1000605
21 Chanvivattana, Y. *et al.* (2004) Interaction of polycomb-group proteins controlling flowering in Arabidopsis. *Development* 131, 5263–5276
22 Lotan, T. *et al.* (1998) Arabidopsis *LEAFY COTYLEDON1* is sufficient to induce embryo development in vegetative cells. *Cell* 93, 1195–1205
23 Junker, B.H. *et al.* (2006) VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 7, 109

1360-1385/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tplants.2010.03.005 Trends in Plant Science 15 (2010) 303–307