BMC Biology

RESEARCH ARTICLE　　　　　　　　　　　　　　　　　　　　　**Open Access**

Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity

Kathrin Schrick[1,2*], Michael Bruno[3], Aashima Khosla[1], Paige N Cox[1], Sara A Marlatt[2], Remigio A Roque[2], Henry C Nguyen[2], Cuiwen He[2], Michael P Snyder[3], Daljit Singh[4] and Gitanjali Yadav[4]

Abstract

Background: Steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domains were first identified from mammalian proteins that bind lipid/sterol ligands via a hydrophobic pocket. In plants, predicted START domains are predominantly found in homeodomain leucine zipper (HD-Zip) transcription factors that are master regulators of cell-type differentiation in development. Here we utilized studies of *Arabidopsis* in parallel with heterologous expression of START domains in yeast to investigate the hypothesis that START domains are versatile ligand-binding motifs that can modulate transcription factor activity.

Results: Our results show that deletion of the START domain from *Arabidopsis* Glabra2 (GL2), a representative HD-Zip transcription factor involved in differentiation of the epidermis, results in a complete loss-of-function phenotype, although the protein is correctly localized to the nucleus. Despite low sequence similarly, the mammalian START domain from StAR can functionally replace the HD-Zip-derived START domain. Embedding the START domain within a synthetic transcription factor in yeast, we found that several mammalian START domains from StAR, MLN64 and PCTP stimulated transcription factor activity, as did START domains from two *Arabidopsis* HD-Zip transcription factors. Mutation of ligand-binding residues within StAR START reduced this activity, consistent with the yeast assay monitoring ligand-binding. The D182L missense mutation in StAR START was shown to affect GL2 transcription factor activity in maintenance of the leaf trichome cell fate. Analysis of *in vivo* protein–metabolite interactions by mass spectrometry provided direct evidence for analogous lipid-binding activity in mammalian and plant START domains in the yeast system. Structural modeling predicted similar sized ligand-binding cavities of a subset of plant START domains in comparison to mammalian counterparts.

Conclusions: The START domain is required for transcription factor activity in HD-Zip proteins from plants, although it is not strictly necessary for the protein's nuclear localization. START domains from both mammals and plants are modular in that they can bind lipid ligands to regulate transcription factor function in a yeast system. The data provide evidence for an evolutionarily conserved mechanism by which lipid metabolites can orchestrate transcription. We propose a model in which the START domain is used by both plants and mammals to regulate transcription factor activity.

Keywords: Transcription, Steroidogenic acute regulatory related lipid transfer, START, StAR, Homeodomain, HD-Zip, Glabra2, Yeast, *Arabidopsis*, Mouse

* Correspondence: kschrick@ksu.edu
[1]Division of Biology, Kansas State University, Manhattan, KS 66506, USA
[2]Keck Graduate Institute of Applied Life Sciences, Claremont, CA 91711, USA
Full list of author information is available at the end of the article

Background

Steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) is an evolutionarily conserved module of approximately 200 amino acids implicated in lipid/sterol binding [1]. The prototype for the START domain is found in mammalian StAR proteins that bind cholesterol for the initiation of steroidogenesis [2]. The START domain subfamily belongs to an expansive clan of α/β helix-grip-fold structures that is given the name 'SRPBCC (START/RHO_alpha_C/PITP/Bet_v1/CoxG/CalC) superfamily' in the NCBI Conserved Domains Database, named for six of its major subfamilies. Another subfamily termed PYR/PYL/RCAR-like includes the newly identified abscisic acid receptors from plants [3,4]. Members of the SRPBCC superfamily, also referred to in the literature as the Bet_v1-like [5] or START [6] superfamily, exhibit the common property of a deep hydrophobic ligand-binding pocket. The focus of the present study is on proteins carrying domains of the START subfamily, and we restrict the term 'START' to the subfamily of proteins that share significant sequence similarity to mammalian StAR, as described in [7]. START domains vary in size and configuration and occur primarily in animals and plants as well as some species from other taxa. While the StAR protein contains the START domain as the major motif, START domains are modular in that they are found in combination with other functional domains. Ligand-binding by the START domain in a multidomain protein may regulate the activities of other domains such as Rho-GAP and thioesterase domains that occur in human START domain-containing proteins.

Experimental evidence for START domains as ligand-binding motifs derives largely from mammalian proteins [8,9]. The human genome encodes 15 START domain-containing proteins, several of which are implicated in disease [10]. Human StAR/STARD1 is associated with an inherited disorder known as congenital lipoid adrenal hyperplasia (CAH) [11]. Identified ligands for mammalian proteins include sterols such as cholesterol and hydroxycholesterol, and other lipids such as phospholipids and sphingolipids. More recently bile acids were shown to bind human STARD5 [12], expanding the repertoire of possible interactors. Mammalian START domain crystal structures are available, revealing central antiparallel β-sheets and a carboxy-terminal α-helix forming a hydrophobic cavity to accommodate a single ligand [13-17]. Ligand specificity is thought to be conferred by the configuration of amino acids that form the inner lining of the cavity. Amino acid changes that perturb START domain function were uncovered in defective StAR genes from CAH patients, and specific amino acids are predicted to affect cholesterol binding [16]. The co-crystal of phosphatidylcholine transfer protein (PCTP)/STARD2 bound to phosphatidylcholine (PtCho) identified multiple residues that contact the ligand in the hydrophobic pocket [14].

The structures of CERT/STARD11 in its apo-form and in complex with ceramides of variable acyl chain lengths indicate contact points between amino acids and molecular features of sphingolipids [17].

START domain-containing proteins are abundant in the plant kingdom, where the majority are members of a plant-specific homeodomain leucine zipper (HD-Zip) transcription factor family [7]. *Arabidopsis* contains 21 HD-Zip START domain-containing transcription factors of the class III and IV subfamilies. Genetic analysis indicates key roles in cell differentiation and patterning in development, and several family members exhibit striking mutant phenotypes. The class III HD-Zip family contains five proteins implicated in vasculature, meristem initiation and/or organ polarity [18]. The larger class IV HD-Zip family comprises 16 members involved in cell fate determination [19], and includes *Arabidopsis Meristem Layer 1* (*ATML1*) and its close relative *Protodermal Factor 2* (*PDF2*), which are required for epidermal cell fate of the shoot [20,21], and *Glabra2* (*GL2*), which is required for the specification of epidermal cells in the shoot [22], root [23] and seed [24].

Although START domains are amplified in HD-Zip transcription factors, the function of the plant domains in lipid/sterol binding has not been verified, nor has the crystal structure been solved for any plant-derived START-subfamily domain to date. Since HD-Zip proteins are transcription factors, one hypothesis is that the START domain, by binding lipid/sterol ligands, controls transcription analogously to steroid hormone receptors in metazoans [1,7]. One advantage of the proposed mechanism is that the metabolic state of a cell as reflected in lipid/sterol synthesis could dictate cell-type specific mRNA expression. Such a mechanism would be of broad interest since many organisms contain START domain proteins whose functions have not yet been uncovered.

In this study, the overall aim was to investigate the role of the START domain in transcription. We demonstrate that the START domain is required for transcription factor activity of a representative HD-Zip protein, GL2, in *Arabidopsis*. A domain swap experiment indicated that although START domains in mammals and plants exhibit marginal sequence similarity, their function appears to be conserved. We utilized a yeast system to assess the roles of START domains in the context of transcription, and to compare the behavior of known ligand-binding START domains from mammals to 'orphan' START domains from plants. With this approach, ligand-binding activity for several mammalian domains with known ligands was reproduced in yeast. START domains from both plants and mammals were shown to stimulate transcription factor activity when placed in a yeast synthetic transcription factor. Immunoisolation of the synthetic transcription factors followed by mass spectrometry indicated that START domains from two HD-Zip transcription factors participate in protein–metabolite
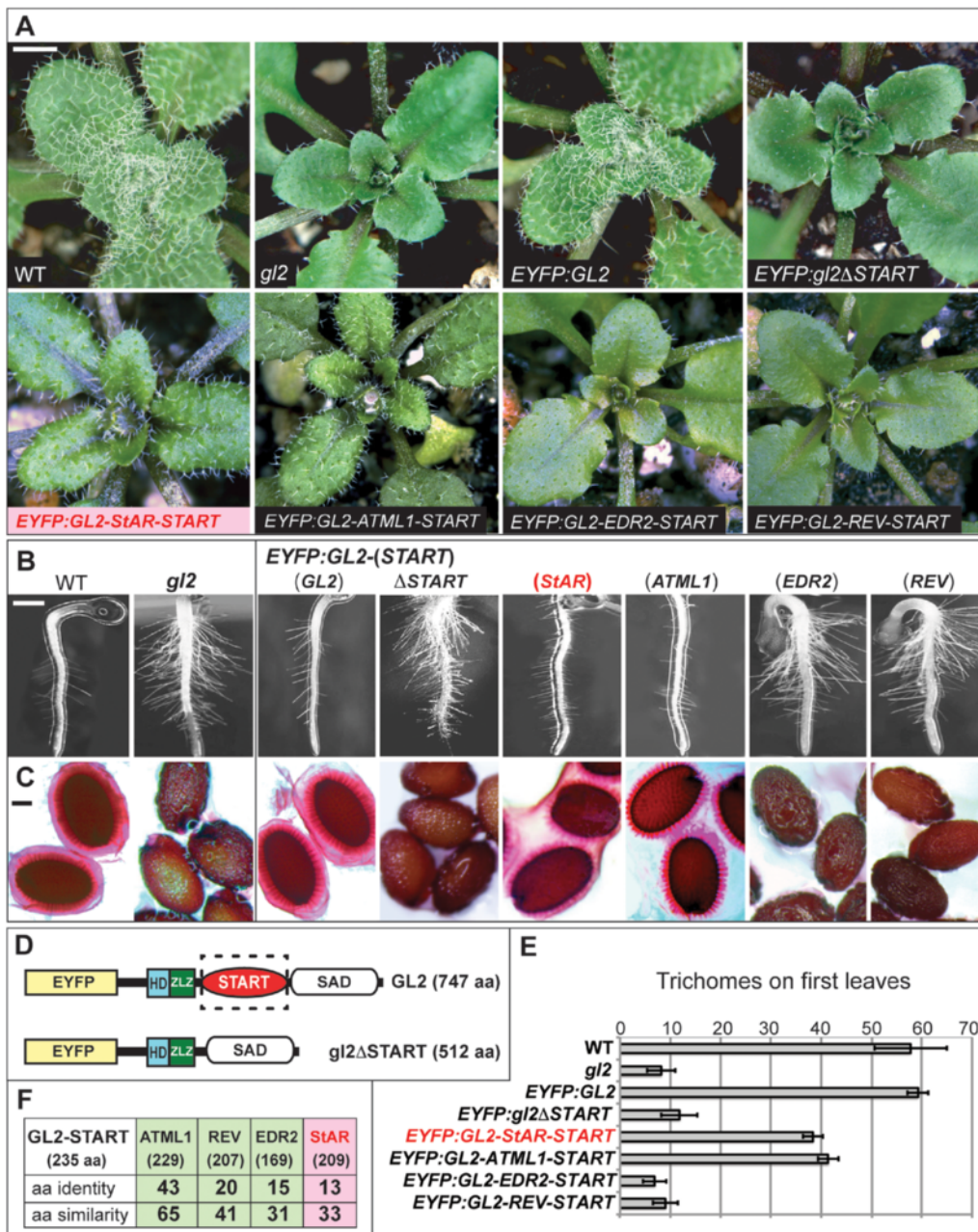
Figure 1 Function of the START domain in HD-Zip transcription factor GL2 from *Arabidopsis*. (A) Rosettes exhibiting leaf trichomes in wild-type (WT), and reduction of trichomes in the *gl2* null mutant. Scale bar: 2 mm. **(B)** Roots were germinated on 0.8% agar medium and imaged after 3 to 5 days. *gl2* mutant exhibits excessive root hairs in comparison to WT. Scale bar: 200 μm. **(C)** Mucilage of WT seeds was stained with ruthenium red after imbibition. *gl2* mutants lack mucilage layer. Scale bar: 20 μm. **(A, B, C)** *EYFP:GL2* displays complete rescue of the mutant phenotype while *EYFP: GL2-StAR-START* (pink) and *EYFP:GL2-ATML1-START* exhibit partial rescue. The phenotypes of *EYFP:gl2*ΔSTART, EYFP:GL2-EDR2-START and *EYFP:GL2-REV-START* are indistinguishable from *gl2*. **(D)** Schematic diagrams of the EYFP:GL2 and EYFP:gl2ΔSTART proteins depicting enhanced yellow fluorescence protein (EYFP), homeodomain (HD), Zip-Loop-Zipper (ZLZ) and START-adjacent domain (SAD). For specific constructs, the START domain was deleted or replaced with a heterologous START domain. **(E)** Trichomes were counted on the first leaves of 1-week-old plants grown on soil. Error bars indicate standard deviations for $n = 20$. **(F)** Amino acid similarities and identities of *Arabidopsis* ATML1, REV and EDR2 START domains (green) and mouse StAR START domain (pink) in comparison to the *Arabidopsis* GL2 START domain. Amino acid sizes of corresponding START domains are indicated in parentheses. aa, amino acid; ATML1, *Arabidopsis thaliana* Meristem Layer 1; EDR2, Enhanced Disease Resistance 2; EYFP, enhanced yellow fluorescent protein; GL2, Glabra2; HD, homeodomain; REV, Revoluta; SAD, START adjacent domain; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; WT, wild type.

interactions as do the mammalian counterparts. We provide experimental evidence for a model in which ligand-binding to the START domain regulates transcription factor activity, linking metabolism to gene expression.

Results

Deletion of the START domain from HD-Zip transcription factor GL2 results in loss of activity

To probe the function of the START domain within HD-Zip transcription factors we chose *Arabidopsis GL2*

for analysis. The *GL2* gene product is dispensable for viability, but *gl2* null mutants exhibit distinct phenotypes in differentiation of the epidermis, including defects in leaf trichome development [22] (Figure 1A; Additional file 1: Figure S1A), excessive root hair formation [23] (Figure 1B) and lack of seed mucilage production [24] (Figure 1C). We deleted the START domain from a GL2 construct in which the cDNA sequence was translationally fused to the enhanced yellow fluorescent protein (EYFP) tag at its amino-terminus (Figure 1D), and transformed *gl2* plants
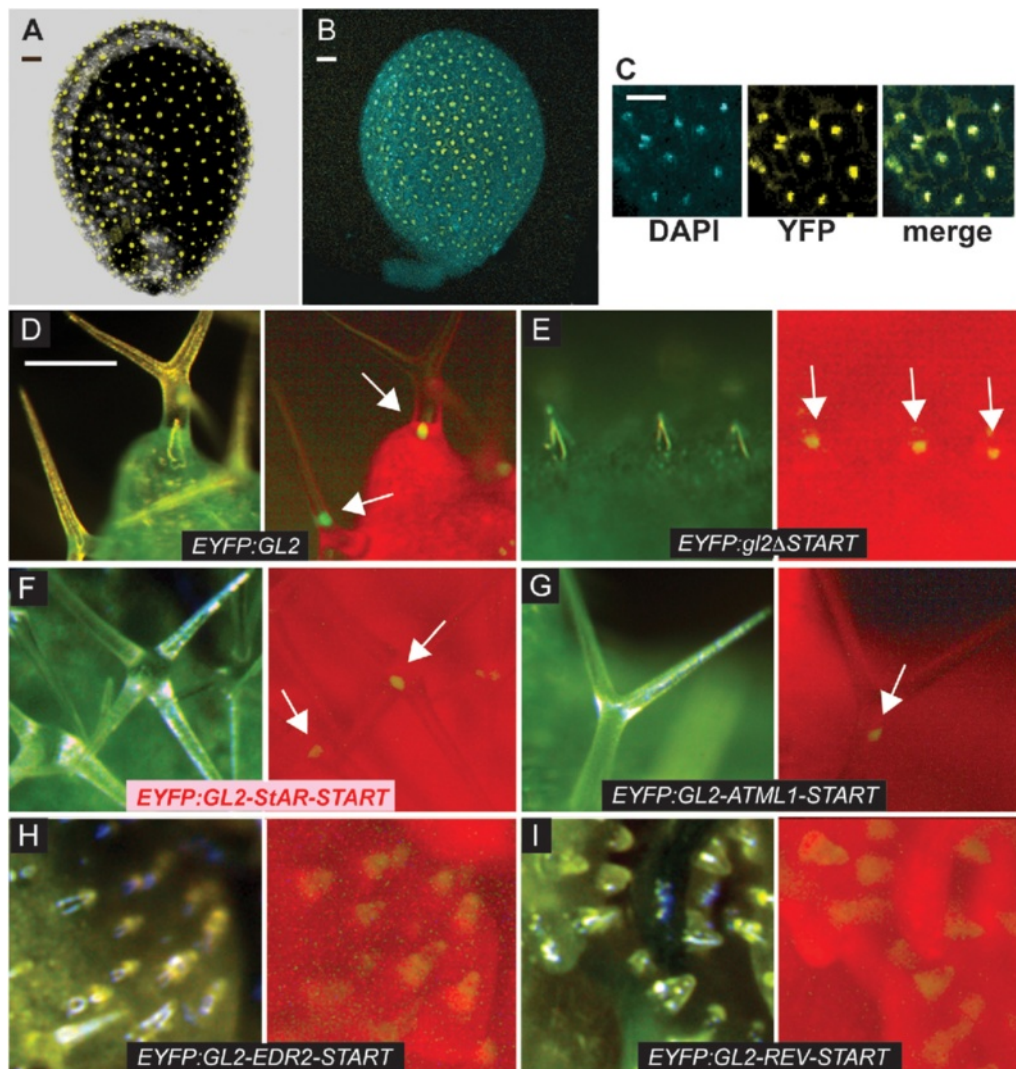

Figure 2 Nuclear localization of HD-Zip transcription factor GL2. Confocal laser scanning images show *Arabidopsis* ovules expressing **(A)** wild-type EYFP:GL2 and **(B)** EYFP:gl2ΔSTART proteins, indicating nuclear localization in immature mucilage secretory cells. **(C)** Nuclear stain DAPI (blue), EYFP (yellow) and merge (green) display nuclear localization of EYFP:gl2ΔSTART from **(B)**. **(A, B, C)** Scale bar: 20 μm. Live imaging of leaf trichomes indicates nuclear localization of **(D)** EYFP:GL2 and **(E)** EYFP:gl2ΔSTART, as well as EYFP:GL2-START domain swap proteins containing the **(F)** mouse StAR (pink) or **(G)** ATML1 START domains. Domain swaps with **(H)** EDR2 or **(I)** REV START domains result in diffuse expression in trichomes. Light (left; green, chlorophyll) and matching fluorescence (right; red, chlorophyll) images are shown for each construct. Scale bar: 100 μm. ATML1, Arabidopsis thaliana Meristem Layer 1; DAPI, 4',6-diamidino-2-phenylindole; EDR2, Enhanced Disease Resistance 2; EYFP, enhanced yellow fluorescent protein; GL2, Glabra2; REV, Revoluta; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; YFP, yellow fluorescent protein.

to examine complementation of the mutant phenotypes. The *EYFP:GL2* transgene was expressed under the native *GL2* promoter (*ProGL2*). Whereas the wild-type *ProGL2: EYFP:GL2* construct rescued all three mutant phenotypes regarding leaves, roots and seeds, the *ProGL2: EYFP:gl2ΔSTART* construct resulted in null phenotypes indistinguishable from the *gl2* loss-of-function mutant (Figure 1A,B,C,E). Despite the inability of the *EYFP: gl2ΔSTART* transgene to confer phenotypic complementation, we observed nuclear localization in ovules and trichomes, similar to that for the wild-type *EYFP:GL2* transgene (Figure 2A–E).

The START domain from mouse StAR functionally replaces the endogenous START domain from GL2

The 209 amino acid START domain from the mouse StAR domain shares 33% similarity and 13% identity with the 235 amino acid START domain from *Arabidopsis* GL2 (Figure 1F). We used a domain swap experiment to test the ability of this mammalian StAR START domain to replace the *Arabidopsis* START domain functionally in the EYFP:GL2 protein cassette (Figure 1D). Among the T2 transformants of *gl2*, we characterized several independent lines that exhibited partial rescue of the trichome defects and selected one of these lines for detailed analysis. In addition to partial complementation of the trichome defects (Figure 1A,E and Additional file 1: Figure S1), these lines also exhibited partial rescue of the root hair patterning (Figure 1B) and seed mucilage defects (Figure 1C) of *gl2* null mutants. The results indicate that despite relatively low sequence similarity (33%), a mammalian START domain is able to replace the plant START domain from GL2 functionally to a similar level as that of a related class IV HD-Zip family member, ATML1 [21] (Figure 1A–E and Additional file 1: Figure S1). START domain swaps with mouse StAR or *Arabidopsis* ATML1 both resulted in nuclear localization of the EYFP:GL2 protein (Figure 2F,G). The function of the START domain in GL2 was dependent on the encoded domain sequence, since replacement with two additional *Arabidopsis*-derived START domains, namely from Enhanced Disease Resistance 2 (EDR2) [25] and from class III HD-Zip member Revoluta (REV) [26], were not found to rescue the *gl2* mutant phenotype (Figure 1A,B,C,E), nor did they display nuclear localization (Figure 2H,I).

A yeast assay for START domain function in transcription

We developed an assay to monitor the activity of the START domain within a synthetic transcription factor using heterologous expression in the yeast *Saccharomyces cerevisiae*. The START domain-coding region from various proteins (Figure 3A) was fused between an amino-terminal GAL4 DNA binding domain (DBD) and a carboxy-terminal VP16 activation domain (AD) (Figure 3B).

The GAL4-VP16 synthetic transcription factor is driven by the *Mitochondrial Ribosomal Protein 7* (*MRP7*) constitutive promoter, resulting in very low background levels of reporter gene expression. Activities of GAL4-DBD:START: VP16-AD (GSV) synthetic transcription factors were monitored by their ability to activate the *LacZ* reporter (Figure 3C). The yeast strain harbored an *erg6* mutation to facilitate permeability and uptake of small molecules. In a GAL4-DBD:hER:VP16-AD (GEV) construct in which the human estrogen receptor steroid binding domain (hER) is placed between GAL4-DBD and VP16-AD domains, the level of reporter activity, quantified in β-galactosidase (β-Gal) units, was reduced to very low levels. However, these levels were increased approximately 25-fold by the addition of estradiol (Figure 3D), consistent with a previous report demonstrating that hER is responsive to its biological ligand estradiol when expressed in a GEV transcription factor in yeast [27]. A GAL4 DNA binding domain:VP16 activation domain (GV) control construct in which the GAL4-DBD was fused directly to VP16-AD was tested and shown to display an approximately sixfold lower level of activity (Figure 3D).

START domains from mammalian proteins stimulate transcription factor activity in yeast

We studied the function of the START domain in the synthetic transcription factor with a mammalian StAR protein that is known to bind cholesterol. When the START domain from mouse StAR was placed between the GAL4-DBD and VP16-AD domains, we observed significantly elevated levels of β-Gal activity in the GSV plasmid in comparison to control plasmids, such as GV without the START domain (approximately threefold and approximately sixfold increases; Figure 3D,E), and GEV-expressing yeast in the absence of estradiol (approximately 10- and approximately 20-fold increases; Figure 3D,E).

To examine the generality of this ability to stimulate transcription factor function, we tested the START domains from two other mammalian proteins, human metastatic lymph node 64 (MLN64/STARD3) and human PCTP. MLN64 is closely related to StAR and has been shown to bind cholesterol [28], while the more distantly related PCTP binds PtCho [14]. Both MLN64 and PCTP behaved similarly to StAR in the ability to stimulate transcription factor activity in the yeast assay (Figure 3E).

START domains from several *Arabidopsis* proteins also boost transcription factor activity

Based on amino acid similarity to the mammalian counterparts, the *Arabidopsis* genome encodes 21 HD-Zip transcription factors that contain START domains out of a total of 35 START domain-containing proteins [7]. Then 25 START domains from *Arabidopsis* were tested in the yeast assay and compared with the reporter gene
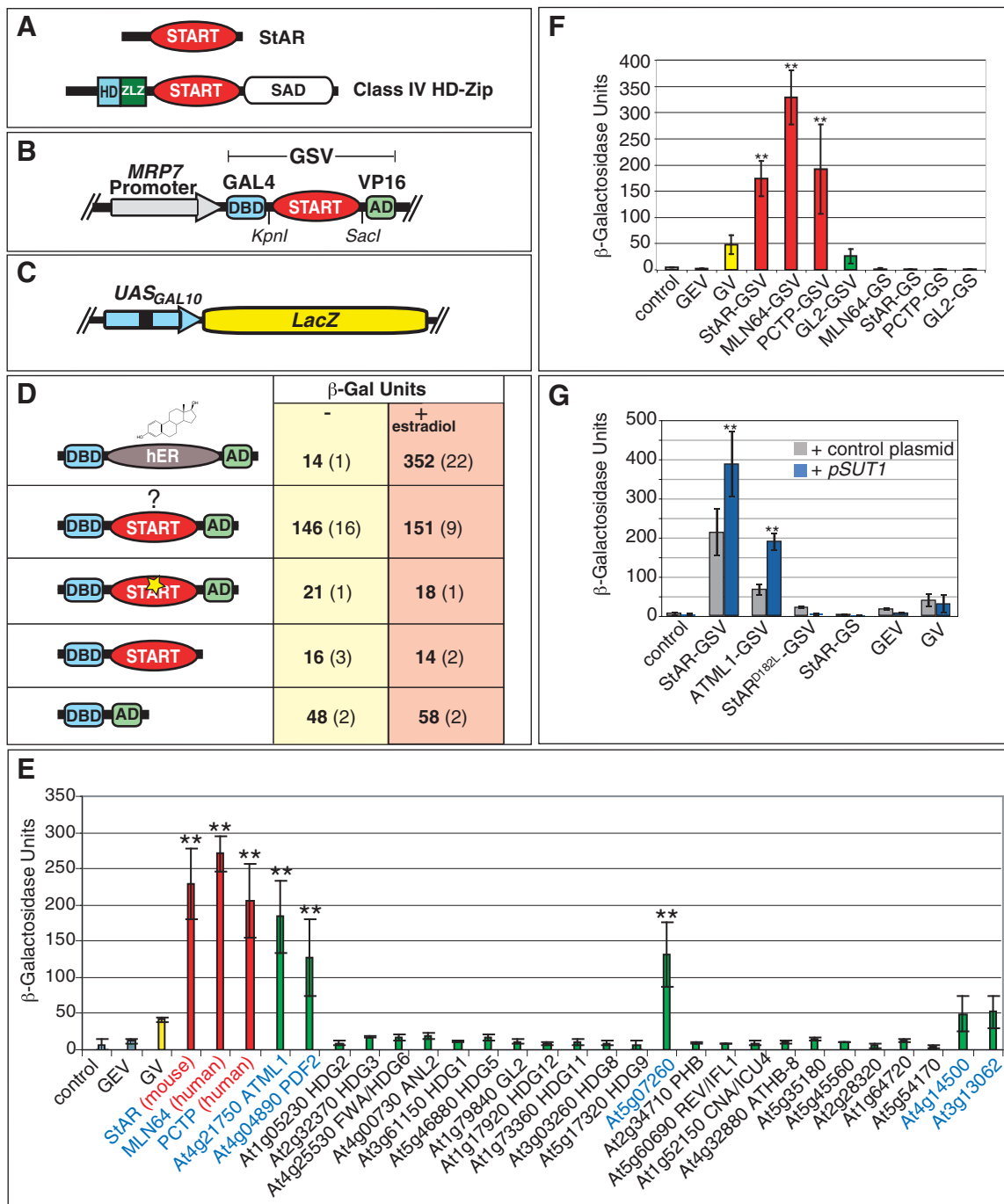
Figure 3 (See legend on next page.)

(See figure on previous page.)

Figure 3 The START domain stimulates transcription factor activity in yeast. (A) Mammalian StAR and a plant-derived HD-Zip transcription factor. **(B)** GSV synthetic transcription factor. **(C)** GAL4-DBD binding to UAS from GAL10 drives a *LacZ* reporter. **(D)** The START domain stimulates activity of the GV transcription factor. A representative experiment is shown for the START domain derived from mouse StAR. When hER is placed between GAL4-DBD and VP16-AD (GEV), treatment with estradiol (10 μg/ml) leads to increased *LacZ* activity. START domain activity is unchanged by estradiol treatment. START domain missense mutation (StAR^{D182L}) (yellow star) results in loss of activity. Removal of VP16-AD similarly results in loss of activity, and GV that lacks the START domain results in a low level of activity. **(E)** Three Mammalian START domains (red) and at several *Arabidopsis* START domains (green) exhibit significant activity levels in comparison to one or more controls: GV (yellow), GEV (grey) and the vector control (pRS314) (grey). **(F)** The START domain is not a transcriptional AD. Removal of VP16-AD resulted in GS constructs that lack activity. **(G)** Over-expression of *SUT1* results in elevated activity. START domains from mouse StAR or *Arabidopsis* ATML1 exhibit an increase in GSV activity when co-expressed with the *pSUT1* (blue). Asterisks indicate a significant difference over control plasmid (two-tailed *t*-test, $P < 0.0005$). Other constructs that do not contain a functional START domain, i.e. StAR^{D182L}, GEV and GV, do not display increased activity. In all graphs, error bars indicate standard deviations for two independent transformants in four trials. AD, activation domain; ATML1, *Arabidopsis thaliana* Meristem Layer 1; DBD, DNA binding domain; GEV, Gal4 DNA binding domain:estrogen receptor: VP16 activation domain; GSV, Gal4 DNA binding domain: START domain: VP16 activation domain; GV, GAL4 DNA binding domain: VP16 activation domain; HD, homeodomain; hER, human estrogen receptor; LacZ, gene for β-galactosidase; MRP7, Mitochondrial Ribosomal Protein 7; SAD, START adjacent domain; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; SUT1, Sterol Uptake 1; Zip, leucine zipper; ZLZ, Zip-Loop-Zipper (a plant-specific leucine zipper); β-Gal, β-galactosidase.

activities of three mammalian START domains: StAR, MLN64 and PCTP (Figure 3E). Although the majority of the *Arabidopsis* START domains assayed resulted in little or no reporter activity, a group of five START domains conferred enhanced transcription factor activity similar to that of the three mammalian proteins tested.

The START domains from class IV HD-Zip transcription factors ATML1 and PDF2 exhibited similar levels of activity (4.4- and 3.1-fold over GV control, and 25- and 17-fold over vector control, respectively; Figure 3E). These proteins display a high degree of sequence conservation (90% and 95% amino acid identity and similarity, respectively) and have been shown to be functionally redundant *in vivo* [20]. Similarly, At5g07260, a HD-Zip-related protein of unknown function that lacks the HD domain, exhibited activity. In addition, two PCTP-like proteins, At4g14500 and At3g13062, displayed weak activities that were not significantly different from the GV control but were increased over other plant-derived START domains (Figure 3E).

START domains do not behave as classical activation domains

One possible explanation for the increased transcription factor activity conferred by the START domain is that it acts as a classical transcriptional AD. We examined whether the START domain functions as an AD in combination with GAL4-DBD by deletion of VP16-AD from several GSV constructs. The resulting GAL4-DBD:START (GS) constructs containing mammalian START domains from StAR, MLN64, PCTP or *Arabidopsis* GL2, exhibited a lack of reporter activity in the yeast assay, comparable to that of the empty vector pRS314, about tenfold lower than that of the GV control (Figure 3C,F). Consistent with these data, when the START domain from *Arabidopsis* PDF2 or GL2 was expressed as a bait in a yeast two-hybrid assay, there was no auto-activation of reporter genes although expression of the myc-tagged GS proteins were detected by Western blot (AK and KS, in preparation). We conclude

that the START domain is not a transcriptional activator in the classical sense, since it was found to boost the activity of the transcription factor only when VP16-AD was also present.

Over-expression of a positive regulator of sterol biosynthesis increases reporter activity of GSV constructs containing active START domains

One possible interpretation to explain the effect of the START domain on the synthetic transcription factor is that endogenous levels of lipid/sterol metabolites in yeast act as ligands to regulate levels of activity. Over-expression of the yeast transcriptional regulator *Sterol Uptake 1* (*SUT1*) gene elevates sterol biosynthesis activity [29]. High levels of *SUT1* likely results not only in an increase in sterols, but may also alter the expression profiles of other lipid metabolites needed for membrane synthesis. We found that co-expression of this multi-copy plasmid carrying the *SUT1* gene under control of the strong *PMA1* promoter results in increased activity of the *LacZ* reporter (approximately 2- to 12-fold) in GSV-expressing yeast (Figure 3G). The reporter activity is only increased for those constructs that contain a 'functional' START domain that exhibits activity in yeast, since constructs containing mutations in the START domain do not yield significant levels of transcriptional activity (see below). These data suggest that *SUT1* over-expression acts directly or indirectly on START domains within the GSV constructs.

Over-expression of the *SUT1* gene has also been shown to enhance sterol uptake [29]. However, exogenous addition of cholesterol to the medium did not increase the activity of the *LacZ* reporter in *SUT1* over-expressing cells with GSV constructs containing the START domain from StAR (Additional file 1: Figure S2). With *SUT1* over-expressed, the reporter activity for GSV-StAR did not change with a range of cholesterol concentrations from 0 μM to 50 μM, perhaps due to saturation of ligand-binding.
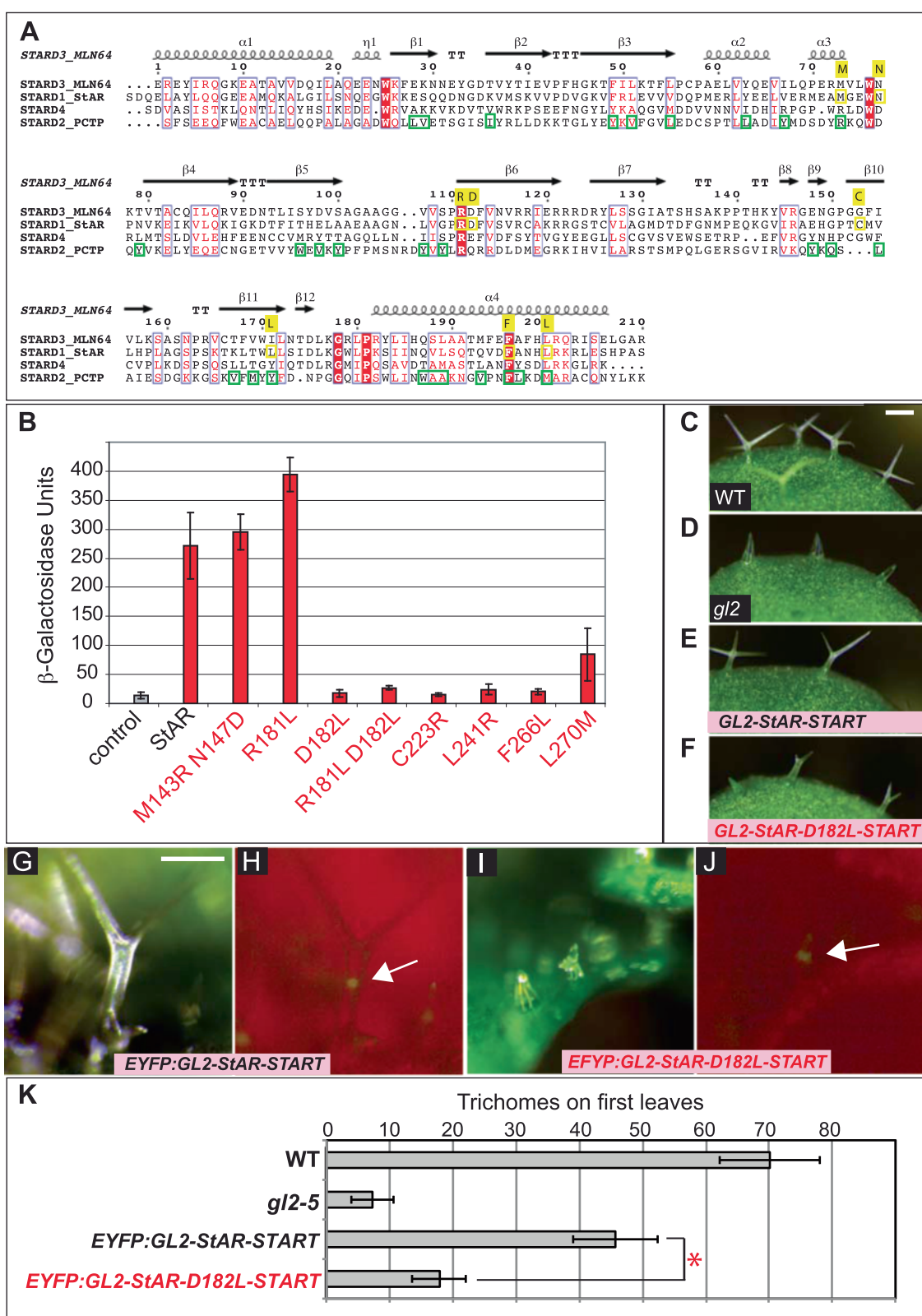
Figure 4 (See legend on next page.)

(See figure on previous page.)

Figure 4 Mutational analysis of amino acids required for StAR START domain activity in yeast assay. (A) Structural alignment of START domains from MLN64, StAR, STARD4 and PCTP with MLN64 as a reference. Consensus α-helices and β-sheets are displayed with ESPript/ENDscript. Amino acids targeted for site-directed mutagenesis of StAR START are indicated in yellow. Ligand contact points (green), as derived from the PCTP-PtCho co-crystal [14], are shown for PCTP. **(B)** Site-directed mutagenesis of the START domain from mammalian StAR. Activities are indicated for GSV constructs that contain missense mutations in the START domain. The M143R;N147D and R181L mutants display wild-type and slightly elevated activities, respectively. L270M renders partial activity, while the other five mutants exhibit low or no activity. Uninduced GEV served as the control. Error bars indicate standard deviations for two independent transformants in four trials. **(C-F)** Trichomes are shown from second leaves of *Arabidopsis* seedlings. Scale bar: 100 μm. **(C)** Wild-type (WT), **(D)** *gl2* mutant and *gl2* lines transformed with **(E)** *ProGL2:EYFP:GL2-StAR-START* and **(F)** *ProGL2:EYFP:GL2-StAR-D182L-START*, in which the START domain of StAR contains the D182L missense mutation. Live imaging of leaf trichomes indicates nuclear localization of **(H)** EYFP:GL2-StAR-START and **(J)** EYFP:GL2-StAR-D182L-START (white arrows). **(G, I)** Light (green, chlorophyll) and **(H, J)** matching fluorescence (red, chlorophyll). Scale bar: 100 μm. **(K)** Quantification of trichomes on first leaves. Error bars indicate standard deviations for $n = 20$. The asterisk marks a significant difference between EYFP:GL2-StAR-START and EYFP:GL2-StAR-D182L-START (two-tailed *t*-test, $P < 0.00001$). EYFP, enhanced yellow fluorescent protein; GEV, Gal4 DNA binding domain:estrogen receptor:VP16 activation domain; GL2, Glabra2; MLN64, Metastatic Lymph Node 64; PCTP, phosphatidylcholine transfer protein; PtCho, phosphatidylcholine; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; WT, wild type.

Site-directed mutagenesis of StAR indicates requirement for ligand-binding

Functional analysis of specific amino acids within the START domain of the mammalian StAR protein was carried out to test further the idea that the START domain is a ligand-binding module within the GSV transcription factor. Using site-directed mutagenesis, missense mutants were constructed on the basis of conservation and alignment with known binding sites in PCTP (Figure 4A). Among the eight mutants tested, two were double mutants. The M143R;N147D double mutant has two amino acids altered that are conserved among the mammalian cholesterol-binding START domains. Altering the M/N pair to R/D is predicted to change the specificity of the StAR START domain from cholesterol-binding to zwitterion lipid-binding, making it more similar to that of PCTP [16]. Similarly, the L270M mutant is predicted to alter ligand specificity from cholesterol-binding to phospholipid-binding. The other missense mutants remove or create additional charged residues within the hydrophobic ligand-binding cavity. The L241R mutation disrupts a hydrophobic cluster within a PCTP binding region in the 11th β-sheet (β11) of the predicted structure. Likewise, the F266L mutation affects the charge of a residue that is both highly conserved and is a predicted ligand contact site. R181L, a mutation that corresponds to a substitution in human patients exhibiting lipoid CAH, affects the charge of a highly-conserved residue within β6 of the START domain that contains other residues implicated in direct ligand contact. In the R181L;D182L double mutant affecting the same region, two adjacent charged residues are changed to leucines. C223R, like R181L and D182L, affects the charge of the cavity, although this residue is also not predicted to be in direct contact with the ligand [14]. However, an adjacent residue, Met at position 224, is mutated (M224T) in a subset of lipoid CAH patients [16].

The activities of the mutants were compared to wild-type StAR (Figure 4B). Although most of the mutations in StAR reduce or completely abolish function, the R181L mutant, which is predicted to affect lipid transfer but not binding [30], displayed wild-type levels of activity. The two mutants, L270M and M143R;N147D, that were predicted to affect specificity, resulted in wild-type and partial activities, respectively.

To examine whether a mutant version of the StAR START domain affects transcription factor activity of the HD-Zip GL2 protein *in planta*, a *ProGL2:EYFP:GL2-StAR-START* transgene harboring the D182L missense mutation was constructed and transgenic lines were analyzed in comparison to the wild-type counterpart. Among several independent lines for which expression of the transgene was verified, fewer branched trichomes were observed compared to a *GL2* transgene containing the wild-type StAR-START domain (Figure 4C–F; Additional file 1: Figure S3). Nuclear localization of the EYFP:GL2-StAR-D182L-START transcription factor indicated that the mutant protein is expressed and properly localized (Figure 4G–J). Quantification of trichomes on first leaves revealed significantly reduced function for the GL2-StAR protein containing the D182L mutation compared to the wild-type version (Figure 4K). These results indicate that a functional ligand-binding START domain is required for GL2 transcription factor activity.

Protein levels of GSV transcription factors and the relation with activity levels

One possibility for explaining the ability of the START domain to stimulate transcription factor activity is that it may stabilize protein levels when it is ligand-bound. We utilized the yeast system to further probe the mechanism by which the START domain functions to enhance transcription factor activity, by examining protein levels for a representative subset of the GSV synthetic transcription factors. Because we were not successful in detecting the GSV proteins with commercial antibodies against GAL4 or VP16, likely due to very low expression levels from the *MRP7* promoter, the GSV transcription factors were fused to the yeast-enhanced green fluorescent protein (yEGFP3)

in plasmid pUG35 [31] (Additional file 1: Figure S4A). Green fluorescent protein (GFP) expression levels were quantified by flow cytometry (Additional file 1: Figure S4B,C). Each of the GSV:yEGFP3 constructs exhibited a GFP level above the negative control (Additional file 1: Figure S4B), and the mean values of fluorescent signals were similar to each other (mean range: 36 to 49) in contrast to that of the control (mean: 12) (Additional file 1: Figure S4C). These data indicate that each of the GSV:yEGFP fusion proteins is properly expressed in yeast.

We assayed the GSV-yEGFP3 constructs for reporter gene activity in the yeast assay (Additional file 1: Figure S4D). Levels of expression, as monitored by percentage GFP, do not strictly correlate with transcription factor activity levels, since START domains that lacked activity (StAR^{C223R} and StAR^{F266L}) displayed similar percentage GFP levels (0.12 and 0.10) as START domains that displayed activity (ATML1 and PDF2) (0.11). However, the three GSV:yEYFP3 constructs that displayed the lowest percentage (0.03) of GFP cells (*Arabidopsis* GL2, mouse StAR^{D182L} and StAR^{L241R}), also displayed low levels of transcription factor activity. These results suggest that the ligand-binding activity of a given START domain may result in increased protein stability.

Identification of metabolites bound to START domains in the yeast system

To investigate the molecular nature of the ligands underlying activation of the GSV transcription factors, we used a modified version of a metabolite–protein isolation protocol that was developed for yeast proteins [32]. For these experiments, we chose a subset of GSV transcription factors carrying START domains from mouse StAR, human PCTP, *Arabidopsis* HD-Zip transcription factors PDF2 and GL2, as well as an *Arabidopsis* PCTP-like protein At3g13062, and tagged these GSV proteins with an immunoglobulin G (IgG)-binding domain [33] (Figure 5A). As negative controls, we included the D182L mutant version of the START domain from mouse StAR that exhibits reduced activity in both yeast and plants (Figure 4) and the GV fusion that lacks the START domain altogether.

Small metabolites bound to the tagged proteins were identified by affinity protein purification followed by mass spectrometry. We focused on the metabolites significantly enriched in the GSV transcription factors relative to the GV control. A shortlist of candidate masses were queried against the Human Metabolome Database (HMDB) and METLIN to identify and assign a metabolite name to the enriched features at a high or low confidence level (Additional file 2: Table S1). Purified proteins corresponding to each sample were examined by SDS-PAGE, which indicated expression of GSV-IgG (and GV-IgG control) proteins of expected sizes (Figure 5B). In parallel, the GSV-IgG constructs were tested for activity in the yeast assay

(Figure 5C), and shown to behave in a similar manner as the original GSV constructs (Figures 3E and 4B). While GSV transcription factors containing START domains from mammalian StAR and PCTP, as well as *Arabidopsis* PDF2 and At3g13062 displayed activity levels that were increased over the GV control, the START domains from the StAR^{D182L} mutant and *Arabidopsis* GL2 displayed low levels of activity.

Hierarchical clustering of the mass spectrometry data revealed several distinct clusters of protein–metabolite interactions (Figure 5D). The horizontal dendrogram places the GSV-StAR transcription factor in a separate cluster from the missense mutant StAR^{D182L}. The enriched metabolites of StAR^{D182L}, which clustered with the *Arabidopsis* At3g13062 START domain, are distinguished by an abundance of several distinct species of diacylglycerols. The remaining proteins are grouped by the horizontal clustering dendrogram such that both *Arabidopsis* PDF2 and GL2 START domains have more similar bound-metabolite preferences, with human PCTP being more closely related to this group than the mouse StAR domain. In addition to several metabolites that were shared with StAR (e.g. lanosterol, squalene, PtCho, diacylglycerol and LysoPtCho), PCTP was shown to bind reproducibly to several types of lipids including sphinganine, sphingosine, behenic acid and linoleamide. The vertical dendrogram reveals a major metabolite cluster comprising StAR, PCTP, PDF2 and GL2 that contains an overrepresentation of LysoPtCho. Metabolites common to PCTP, StAR and the mutant StAR^{D182L} form a distinct cluster, which includes some phospholipids, such as PtCho and LysoPtCho, as well as several sterols including four cholesterol esters (CEs). This cluster also includes two major precursors of the cholesterol biosynthesis pathway, lanosterol and squalene. The mutant StAR^{D182L} is significantly less enriched for CE, PtCho, LysoPtCho, squalene and lanosterol within this cluster.

Additional file 2: Table S1 provides a detailed list of significant features that were enriched or depleted for individual GSV transcription factors. Although the overall chemical nature of the bound metabolites are diverse in class, glycerolipids and phospholipids together comprise the majority of interacting metabolites identified. We generated a comprehensive protein–metabolite network (Additional file 1: Figure S5A) from enrichment data in Additional file 2: Table S1 for all six START domains tested. To compare and contrast metabolite binding preferences better, we constructed sub-networks (Additional file 1: Figure S5B,C and Figure 6A) from a more stringent filtering, and included only high confidence and validated metabolites. The data support a hypothesis that the START domains from plant and mammalian origin can interact with a common set of metabolic ligands (Figure 6).

Figure 5 (See legend on next page.)

(See figure on previous page.)

Figure 5 Expression and activity of tagged GSV transcription factors and dendrogram from protein–metabolite immunoisolation.
(A) GSV translational fusion to a triple tag (6x-HIS (white), HA-epitope (dark blue), Protease 3C cleavage site (orange), and IgG binding domain ZZ (yellow)) in pBG1805 [33]. **(B)** Immunoisolation of tagged proteins with IgG Sepharose beads followed by SDS-PAGE. One of four biological replicates is shown. Y258 URA3+ (lane 1) served as the negative control, while GSV proteins are indicated by red asterisks. The GV fusion lacking the START domain is marked by a black asterisk. Arrowheads indicate IgG heavy and light chains. **(C)** Tagged GSV constructs were expressed in *erg6* cells containing the *LacZ* reporter and transcriptional activity was measured in β-Gal units. Double asterisks indicate a significant increase in activity over the GV control (two-tailed *t*-test, *P* < 0.03). Error bars indicate standard deviations for two independent transformants in four trials. **(D)** Dendrogram from protein–metabolite immunoisolation. Two-way hierarchical clustering was performed on the metabolite enrichment data relative to the GV control with values expressed as the logarithm base 2 (ratio) of normalized metabolite intensities. Horizontal clusters represent START domain clusters based on protein-bound metabolite enrichment. Vertical clusters show groups of protein-bound metabolites clustered by START domain. Metabolite names marked by asterisks were validated by mass spectrometry, matching exact mass and retention time to a known standard. CE Metabolites of the same chemical subclass were grouped and given a superscript number to distinguish members that bear distinct masses and retention times. aa, amino acids; AD, activation domain; DBD, DNA binding domain; GL2, Glabra2; GSV, Gal4 DNA binding domain:START domain:VP16 activation domain; GV, GAL4 DNA binding domain:VP16 activation domain; IgG, immunoglobulin G; PCTP, phosphatidylcholine transfer protein; PDF2, Protodermal Factor 2; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer.

Structural modeling of *Arabidopsis* START domain ligand-binding cavities

From their X-ray crystal structures, the ligand-binding cavity volumes of mammalian START domains from MLN64 (848 ± 107 Å^3) [16] and PCTP (approximately 882 Å^3) [14] are highly similar although they bind different hydrophobic ligands. Cholesterol, the ligand of MLN64, occupies a 741 Å^3 volume, while PtCho, the ligand of PCTP, occupies 797 Å^3. Based on our domain swap experiments in *Arabidopsis* (Figures 1 and 2) and GSV transcription factor activities in the yeast assay (Figure 3E), and binding activities in the protein–metabolite immunoisolations (Figure 5), *Arabidopsis* START domains from class IV HD-Zip transcription factors are predicted to exhibit similar ligand-binding cavities as their mammalian counterparts.

Using available three-dimensional coordinates for crystal structures of mammalian START domains and threading-based fold prediction, the closest structural homolog for each of the *Arabidopsis* START domains tested in the yeast assay was determined to be the mouse STARD4 domain (Figure 7A). Three-dimensional homology models were constructed and the corresponding hydrophobic tunnels were assessed for size, volume and identities of the cavity lining residues. The modeling results indicated that the hydrophobic tunnels of the *Arabidopsis* START domains fall into three categories of volume sizes. Plant START cavities were found to be either small in size (Figure 7B; volumes less than 400 A^3), medium (Figure 7C; volumes between 400 and 700 Å^3) or large (Figure 7D; volumes >700 Å^3). The five START domains showing activity in the yeast assay fall into the large volume category. For example, START domains from HD-Zip transcription factors ATML1 and PDF2, which exhibit activity in the yeast assay, have predicted volumes of 770 and 720 Å^3, respectively. In contrast, START domains from other HD-Zip transcription factors, such as GL2 and Corona (CAN), which had little or no activity in the yeast assay, fall into the small volume category with tunnels of 284 Å^3 and 320 Å^3, respectively. Among the plant START domains

that exhibited a negative result, three others are predicted to have similarly small cavities.

Detailed examination of the predicted cavity-lining amino acid residues in each of these cases revealed that certain side-chain residues are oriented into the space within the original tunnel, thereby blocking the passage and restricting the cavity size (Figure 7E). Some frequent blocking positions correspond to N147, A178 and G200 (StAR numbering), all of which are replaced by residues with much larger side chains in *Arabidopsis* START domains, such as Arg, Trp/Tyr/Phe/Gln/Asn and Ile/Leu, respectively. Most of the domains that lack activity in the yeast assay correspond to volumes that fall into the medium category, while two domains from HD-Zip transcription factors (HDG2 and FWA/HDG6) have cavities of the large volume category. In the START domain from HDG2, the opening of the tunnel is slightly shifted, such that a potential ligand would be predicted to encounter steric hindrance in entering the hydrophobic tunnel. Thus, steric factors may explain why some of the plant-derived START domains fail to exhibit activity in the yeast assay.

Discussion
A role for the START domain in the regulation of transcription

Here we chose GL2, a representative family member from *Arabidopsis*, to investigate the significance of the START domain in HD-Zip transcription factors. Our results indicate that the START domain is required for transcription factor activity (Figures 1 and 2). Strikingly, the GL2 START domain could be functionally replaced by the mouse StAR START domain, a mammalian counterpart. The GL2 START domain could similarly be replaced by the START domain from ATML1, a transcription factor from the same class IV HD-Zip subfamily that shares a relatively high degree of amino acid similarity, but other *Arabidopsis* START domains of lower similarity (EDR2 and REV; Figure 1F) could not functionally replace the
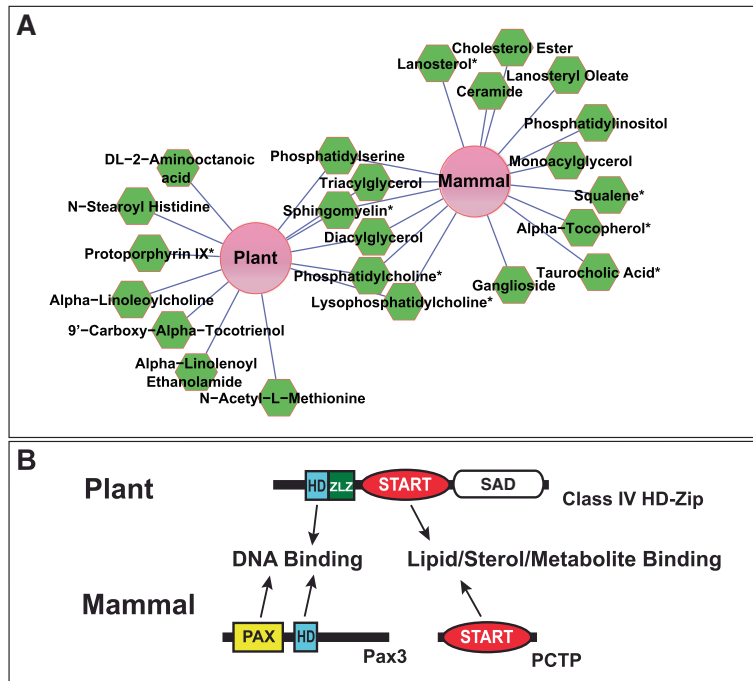
Figure 6 Protein–metabolite interaction network of mammalian and *Arabidopsis* START domains. (A) Cytoscape was used to visualize the normalized protein–metabolite enrichment data in an edge-weighted interaction network where metabolite interactions held in common by the *Arabidopsis* START domains were grouped to produce a single node labeled 'Plant', and common interactions for mouse and human proteins (StAR, StAR^{D182L*} and PCTP) were grouped to produce a 'Mammal' node. Distances between protein and metabolite nodes reflect the interaction strengths based on the magnitude of the fold-change; the shorter the edge the more enriched the metabolite. The network was filtered for interactions with a greater than tenfold change in enrichment relative to the GV control and only high confidence metabolite assignments were included. If a node had multiple interactions with the same chemical sub-class of metabolite (e.g. PtCho) these interactions were combined and weighted to give one interaction. Asterisks designate metabolites that were validated by mass spectrometry, matching exact mass and retention time to a known standard. **(B)** Schematic illustrating how START domains may modulate transcription factor activity in plants and mammals. In plants, the START domain is found in HD-Zip transcription factors that contain an HD DNA binding domain. In mammals, the START domain and the DNA binding domains are in two separate proteins. A physical interaction between the START protein PCTP and Pax3 transcription factor comprising both paired box (PAX) and HD DNA binding domains was been reported [34]. GV, GAL4 DNA binding domain:VP16 activation domain; HD, homeodomain; PCTP, phosphatidylcholine transfer protein; PtCho, phosphatidylcholine; SAD, START adjacent domain; StAR, steroidogenic acute regulatory protein; START, StAR-related lipid transfer; Zip, leucine zipper; ZLZ, Zip-Loop-Zipper (a plant-specific leucine zipper).

GL2 START domain, indicating that a specific sequence or structural feature is critical.

The ability of the mouse StAR START domain to confer partial complementation may reflect the presence of the appropriate ligand (i.e. cholesterol) in GL2-expressing cells, whereas the ligands of EDR2 or REV START, which are unknown to date, are potentially not present in these cells. The conclusion that the mouse StAR START domain participates in ligand-binding in plant cells is further supported by our finding that the D182L mutation results in reduced GL2 activity (Figure 4). Alternatively, there is a function unrelated to ligand-binding that the START domains of StAR and ATML1 can fulfill. Although our structural modeling data suggest different cavity volumes for StAR and GL2 (as well as ATML1 and GL2), the protein–metabolite immunoisolation data from yeast indicate an overlap in binding partners between GL2 and StAR, as well as between GL2 and PDF2, the closest relative of ATML1. It is noteworthy

that the EDR2 and REV START domain swaps are expressed but are not localized in nuclei (Figure 2H,I). One possible explanation is that these START domains yield improperly folded transcription factors, which interfere with transport to the nucleus. However, since the GL2 protein lacking the START domain is found in the nucleus, we conclude that the START domain is not strictly required for correct subcellular localization.

START domains stimulate transcription factor activity via ligand-binding

It was hypothesized that HD-Zip transcription factors from plants are regulated via lipid/sterol ligands that bind to their START domains [1,7]. The yeast assay described here tests the function and ligand-binding properties of START domains within a synthetic transcription factor, and indicates that START domains of mammalian or plant origin can regulate activity in a similar fashion (Figure 3). The
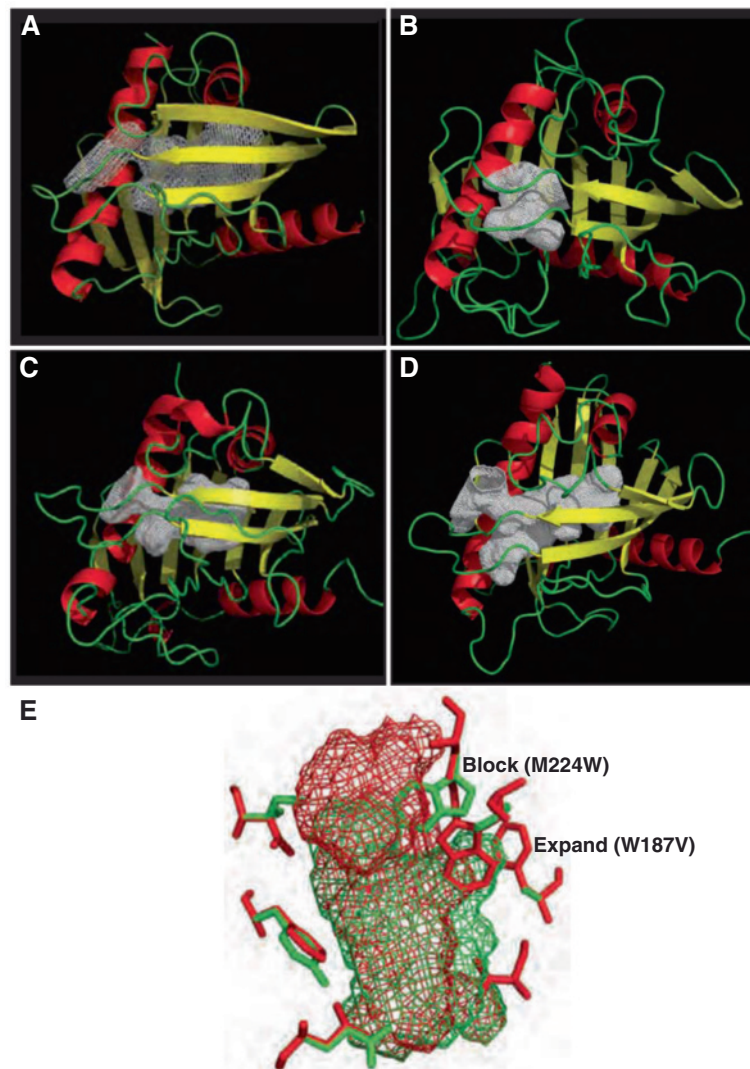
Figure 7 Comparative structural modeling of plant START domains. Ribbon diagrams of three-dimensional models comprised β-sheets (yellow), α-helices (red) and loops (green), showing three categories of cavity volume. The hydrophobic tunnels are represented as a white mesh. **(A)** Crystal structure of the mammalian START domain from STARD4. **(B, C, D)** Examples of small, medium and large cavity volumes, respectively, from modeled *Arabidopsis* START domains. **(E)** Superimposition of cavities from a mammalian domain (red mesh) and a model of an *Arabidopsis* domain (green mesh) to illustrate the amino acid residues that are predicted to change the dimensions of the cavity by blockage or expansion in several plant START domains. A Trp residue (green, marked 'block') in the plant structure corresponding to M224 (StAR numbering) results in shortening of the cavity length. In contrast, the mammalian cavity is restricted by W187 (red, marked 'expand'), which corresponds to a smaller Val residue in the plant START domain, resulting in an increase in width at the center of the cavity.

START domains do not behave as classical ADs in the absence of VP16 (Figure 3F), and based on our site-directed mutagenesis, protein–metabolite immunoisolation experiments and structural modeling, we conclude that the likely mode of action is via ligand-binding. Since several START domain-containing proteins from humans have been found in the nucleus [8,9], our findings support a possible role for the START domain in regulating gene expression in both plants and animals.

Relevant to our results, mammalian PCTP was previously shown to interact physically with HD transcription factor Pax3 and the two proteins are co-localized in the nuclei of mouse embryos [34] (Figure 6B). The previous study showed that PCTP co-activated the transcriptional activity of Pax3 in tissue culture cells, and it was postulated that the physical interaction of PCTP with Pax3 positively regulates its transcriptional activity. Perhaps the interaction between PCTP and Pax3 is analogous to the function of the START domain in plant HD-Zip transcription factors (Figure 6B). While the previous study examined the function of the START domain in an interacting protein [34], here we demonstrate the

regulatory role of the START domain as an integral part of the transcription factor.

We probed three mammalian START domains and 25 *Arabidopsis* START domains for the ability to stimulate transcription factor activity in yeast (Figure 3E). While all three mammalian START domains exhibited activity in comparison to the control, only 5/25 of the *Arabidopsis* START domains displayed similar levels of activity. Using protein–metabolite immunoisolation experiments (Additional file 2: Table S1, Figures 5 and 6, and Additional file 1: Figure S5), we tested the idea that lipids/sterols present in yeast bind the corresponding START domains, resulting in the observed activity. We reasoned that START domains that lack activity might require plant-specific metabolites not found in yeast. The *Arabidopsis* GL2-START domain, which displays little or no activity in yeast, was found to bind to a large number of metabolites. The range of metabolites bound to the START domains from *Arabidopsis* PDF2 (showing activity) and GL2 (lacking activity) were overlapping but not identical, leaving the possibility that those specific to GL2 might behave as negative regulators. In general, protein–metabolite interactions may serve to: (a) activate or negatively regulate proteins, (b) stabilize protein levels, and/or (c) facilitate protein–protein interactions [35]. In addition to activating or negatively regulating transcription factor function, our experiments with the yEGFP3 fusions provide evidence that ligand-binding stabilizes protein levels of the GSV transcription factors (Additional file 1: Figure S4). In addition, ligand-binding may also contribute to correct protein folding.

Site-directed mutagenesis of the START domain from mouse StAR was utilized in conjunction with the yeast assay to characterize START domain function further within the synthetic transcription factor. Our results indicate that transcription factor activity correlates with predicted ligand-binding capacity of the START domain (Figure 4). From the protein–metabolite immunoisolation experiments, comparison of the wild-type START domain from StAR versus the mutant StAR^{D182L} reveals that many more metabolites (including CEs) bind to the wild-type StAR, correlating high levels of activity with ligand-binding in the yeast assay. Consistent with the interpretation that our yeast assay measures ligand-binding, the StAR^{R181L} mutation, which displays wild-type levels of activity, was reported not to affect ligand-binding affinity *in vitro* [36]. In transfected mammalian cells, this missense mutation has an adverse effect on cholesterol transfer and the corresponding conserved Arg may be required for a conformational change that allows release of bound ligand. The StAR mutants that are predicted to change the specificity of ligand-binding had little (StAR^{L270D}) or no affect (StAR$^{M143R;N147D}$) on the activity of the START domain in the yeast assay. This finding was consistent with the results from our protein–metabolite immunoisolation

experiments, in which a range of lipid metabolites including PtCho and sterol-related molecules were found associated with the tagged GSV transcription factor containing the START domain from StAR (Additional file 2: Table S1 and Figures 5 and 6A). We propose that interaction of START domains with accessory proteins, via tissue-specific expression and proper subcellular compartmentalization, are likely to contribute to specificity in binding of selected metabolite ligands during the lifespan of the protein within the cells of the host organism.

Immunoisolation experiments reveal novel protein–metabolite interactions for START domains

The GSV protein–metabolite immunoisolation experiments performed in yeast uncovered potential lipid ligands for mammalian StAR and PCTP that had previously been shown to bind these START domains, namely cholesterol (eg. CE) and PtCho, respectively (Additional file 2: Table S1, Figure 5 and Additional file 1: Figure S5). In addition, novel metabolite–protein associations were found, including known modulators of lipids and sterols in mammals, such as sphinganine, behenic acid and LysoPtCho. Precursors and intermediates of the sterol biosynthesis pathway, including lanosterol and squalene, were also significantly enriched for the START domain from StAR.

There is mounting evidence for a broad range of potential binding partners for a given START domain. In a previous study, it was shown that StAR can stimulate the transfer of both cholesterol and β-sitosterol, a plant sterol, from liposomes to mitochondria [37]. A structural study showed that in addition to ceramides, the START family member ceramide transfer protein (CERT) binds diacylglycerol [17], another metabolite that was uncovered here. It is noteworthy that key lipid metabolites in low-density lipoprotein (LDL) transport, namely CE and LysoPtCho, which are converted from cholesterol and PtCho by lecithin-cholesterol acyl transferase (LCAT), an enzymatic reaction that is critical for LDL transport in the cell, were found to be associated with several START domains that we tested. The identified START domain interactor sphinganine is a known cholesterol transport inhibitor that blocks LDL-induced formation of CE, the product of LCAT activity [38]. Behenic acid, which was found enriched only in PCTP and not the other START domains, has been reported to increase cholesterol levels in humans [39]. Taken together, our findings support the premise that START domains are central players in metabolic regulatory networks in the cell.

Conclusions

The START domain has been characterized primarily as a lipid/sterol-binding domain in mammals, but its function in plant HD-Zip transcription factors was previously not known. Here we show that the START domain is

required for transcription factor activity of GL2, a representative member of the family of class IV HD-Zip transcription factors in plants. We report evidence that START domains from the plant kingdom can act as ligand-binding modules, similar to their mammalian counterparts, and that they can modulate transcription factor activity. The protein–metabolite immunoisolation experiments performed in yeast generated a list of small molecules that are candidates for binding the *Arabidopsis* START domains from HD-Zip transcription factors within the physiological context of plant growth and development. In addition, the unexpected protein–metabolite interactions that were detected for the two mammalian START domains tested, StAR and PCTP, may represent additional small molecule binding partners that regulate these START domains in human cells. In summary, the results presented here provide new insights into the molecular mechanisms underlying START domain function in eukaryotic cells, and reveal novel protein–metabolite interactions that may play critical roles in modulating transcription factor activity in both animals and plants.

Methods

Plants, growth conditions, plant constructs and phenotypic assays

The wild-type *Arabidopsis thaliana* ecotype used was Columbia (Col). The *gl2* null allele was *gl2-5* [40]. Seeds were stratified at 4°C for 3 to 5 days and grown on soil mix containing Supersoil (McLellan Co, Marysville, OH), vermiculite and perlite (McConkey Co, Sumner, WA) in 4:3:2 ratio at 23°C under continuous light. The *GL2* cDNA was sub-cloned into pBluescript via *Kpn*I and *Sal*I restriction sites to serve as a template for the deletion construct. Phusion® Site-Directed Mutagenesis Kit (Thermo Fisher Scientific, Waltham, MA) was used to generate the START domain deletion, following the manufacturer's protocol. After sequence confirmation, the *GL2* START deletion cassette was restriction digested with *Kpn*I and *Sal*I, and ligated into the binary vector SR54 (*ProGL2:EYFP:GL2*), which is derived from pCAMBIA1300. Construction of *ProGL2:EYFP:GL2* will be described elsewhere (AK *et al.,* manuscript in preparation). For domain swap constructs, the START domain sequences were amplified using PCR from corresponding cDNA clones using Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA). The REV sequence contained mutations in the microRNA binding site as in [41]. Gene-specific primers with 15-bp extensions homologous to vector ends were used to amplify the insert followed by In-Fusion HD cloning reactions (Clontech, Mountain View, CA) using a PCR-amplified pBluescript plasmid harboring the GL2 cDNA with flanking homologous ends. For the mutant version of GL2-StAR-START, site-directed mutagenesis was performed as described

below for the yeast constructs. The reconstructed GL2 cDNAs were cloned into the *Sal*I/*Kpn*I sites of binary vector SR54 carrying *ProGL2:EYFP:gl2-START∆*. The constructs were transformed into *gl2-5* plants by floral dip transformation using *Agrobacterium* strain GV3101 (pMP90) [42]. Transgenic plants were selected with 50 µg/ml hygromycin B. Segregation patterns for EYFP expression were observed in T2 progeny of >30 independent lines. T3 homozygous lines with stable EYFP expression were confirmed by PCR and assayed for leaf trichomes, root hairs and seed coat mucilage. For mucilage analysis, seeds were stained in a 0.25% aqueous solution of ruthenium red (Sigma, St. Louis, MO) for 2 h and replaced with deionized water prior to observation. Seeds, rosettes and roots were imaged with a Leica M125 fluorescence stereo microscope, and images were captured using a Leica DFC295 digital camera and Leica Application Suite software version 4.1. Ovules were excised from siliques, stained with 1 µg/ml DAPI (Sigma, St. Louis, MO) and imaged with Zeiss LSM 510 meta and chameleon modules with standard settings for visualization of yellow fluorescent protein and DAPI fluorophores. Oligonucleotide sequences used to make DNA constructs are listed in Additional file 1: Table S2.

Yeast strains and growth media

The genotype of the *Saccharomyces cerevisiae* wild-type strain was *MAT*α *leu2-3,112:: UAS$_{GAL10}$-lacZ-LEU2 his3∆ gal4∆ ura3-52.* The genotype of *erg6* mutant strain was: *MAT*α *erg6-∆1 leu2-3,112::UAS$_{GAL10}$-lacZ-LEU2 his3∆ gal4∆ ura3-52.* These strains were constructed from a cross between YGY13 [27] and CJY004 [43]. For protein expression, Y258 (*MAT*a *pep4-3, his4-580, ura3-52, leu2-3,112*) cells were used. Y258 with an integrated *URA3* gene from the pRS426 vector was a negative control for ligand-binding studies [32]. Growth media were made according to standard protocols. DNA constructs were introduced into yeast using a standard lithium acetate transformation method [44].

Yeast plasmids and DNA constructs

For each GSV construct, the START domain sequence was amplified using PCR from the corresponding cDNA clone or an *Arabidopsis* cDNA library [45] and ligated in-frame into the *Kpn*I/*Sac*I restriction sites of pGEV-HIS3 [27,46]. cDNA clones of the following genes were obtained from the RIKEN *Arabidopsis* full-length clones collection [47]: HDG2 (At1g05230), ANL2 (At4g00730), HDG1 (At3g61150), HDG12 (At1g17920), HDG11 (At1g73360), HDG8 (At3g03260), At5g07260, PHB (At2g34710), CNA (At1g52150), ATHB-8 (At4g32880), At5g35180, At5g45560, At2g28320, At1g64720, At5g54170, At4g14500 and At3g13062. When cDNA inserts had internal *Kpn*I and/or *Sac*I sites, site-directed mutagenesis and primer

extension PCR modified from [48] were used to change the nucleotide sequences without altering the amino acid. To construct GS translational fusions from GSV plasmids, VP16-AD was removed by restriction digestion with *Sac*I, followed by ligation. The empty vector control was pRS314. The GV control construct lacking the START domain was made using site-directed mutagenesis and primer extension PCR of the GSV-StAR plasmid to insert *Nru*I sites at the junctions of its START domain followed by restriction digestion with *Nru*I, and re-circularization. To construct the control vector for the *SUT1* over-expression experiments, *pSUT1* (pNF1) [29] was restriction digested with *Not*I to remove the *SUT1* gene, followed by ligation. To construct translational C-terminal fusions to the yEGFP3 [31], the coding regions for the GSV transcription factors were cloned into the pUG35 plasmid using *Bam*HI/*Eco*RI restriction sites. Site-directed mutagenesis followed by primer extension PCR was used to insert an ATG codon in the polylinker of pUG35 5′ to the *Bam*HI site. For protein expression in conjunction with the identification of *in vivo* ligands, the coding regions for GSV transcription factors containing START domains from StAR, StAR^{R182L}, PCTP, GL2, PDF2 and At3g13062, as well as the GV control were cloned into pENTR^{TM}-D-TOPO$^{®}$ (Life Technologies, Carlsbad, CA) and then moved into pBG1805 [33] by Gateway$^{®}$ LR clonase II (Life Technologies, Carlsbad, CA). Oligonucleotides used for plasmid construction, site-directed mutagenesis and sequence confirmation are listed in Additional file 1: Table S2.

Quantitative β-galactosidase assay

Yeast cultures were grown to exponential phase in 24-well plates (2 ml media per well) at 30°C and 300 rpm using a thermoshaker. OD_{655} measurements were taken prior to centrifugation. Cell pellets were washed and suspended in 100 µl Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4 \bullet 7H_2O$) followed by two freeze-thaw cycles, and addition of 700 µl Z-buffer containing 50 µM β-mercaptoethanol. Then 160 µl ortho-Nitrophenyl-beta-galactoside (4 mg/l) was added to each sample sequentially at 15-s intervals, followed by incubation in a 1,200 rpm thermoshaker at 30°C. Samples were monitored for the development of a yellow color, after which the reaction was stopped by addition of 400 µl of 1 M $NaCO_3$. A_{415} measurements of the supernatants were taken and β-Gal activity was calculated in Miller units [49].

Site-directed mutagenesis of the START domain of StAR

START domains from ATML1, PDF2, GL2, MLN64/STARD3, StAR/STARD1 and PCTP/STARD2 were aligned using ClustalW [50] and Jalview [51]. Site-directed mutagenesis of mStAR-GSV was performed using a one-step PCR-based method [48]. Reaction mixtures contained 25 ng template DNA, 135 ng of each oligonucleotide, 1 µl of 10 mM deoxynucleotide triphosphates, 1 µl *PfuUltra*™ High-Fidelity DNA Polymerase (Agilent Technologies, Santa Clara, CA), 5 µl buffer supplied with the polymerase, and H_2O in 50 µl. PCR conditions were: 95°C denaturation for 1 min; 16 cycles of 95°C for 50 s, 60°C for 50 s, and 68°C for 1 min/kb; extension at 68°C for 7 min. PCR products were digested with 30 U of *Dpn*I for 3 h at 37°C, followed by ethanol precipitation and transformation of XL10-Gold® Ultracompetent Cells (Agilent Technologies, Santa Clara, CA).

Identification of *in vivo* ligands: affinity purification and metabolite extraction

Y258 *ura3-52* yeast cultures expressing pBG1805 constructs were grown in raffinose until they reached an OD_{600} of 0.5 to 1.0. The Y258 strain with an integrated *URA3* selectable marker that was not transformed with the tagged construct served as a minus protein control for non-specific binding. Galactose was added to induce gene expression for 6 to 8 h at 30°C, and cell pellets were collected from 300 ml cultures. Protein expression was monitored after purification from lysates with IgG Sepharose 6 (GE Healthcare, Little Chalfont, United Kingdom) using SDS-PAGE. For the identification of protein-bound ligands, proteins were isolated from lysates using rabbit IgG-crosslinked M-270 epoxy Dynabeads (Life Technologies, Carlsbad, CA) as described [32] with some modifications. After 2 hr incubation at 4°C, the beads were washed in 500 mM NH_4Ac followed by 50 mM NH_4Ac for 5 min each. Using an improved metabolite extraction method (MB and MPS, unpublished data), Dynabeads were briefly (1 min) incubated in 15 µl 95% methanol/5% H_2O with acetic acid added to give a final pH of 4.5, then neutralized by 45 µl of 95% methanol/5% H_2O with ammonium hydroxide and incubated for 8 min at room temperature. The extract was aspirated, stored on ice, and 30 µl fresh 100% methanol was added to the beads with incubation for 10 min at room temperature. Extracts were combined prior to mass spectrometry. To examine protein levels, the beads were boiled in 30 µl 2× SDS sample buffer for 10 min and 15 µl of the supernatant was analyzed by SDS-PAGE.

Mass spectrometry and data analysis

A Thermo LTQ Orbitrap Velos mass spectrometer equipped with a heated electrospray or atmospheric-pressure chemical ionization source was used for ionization and detection of the samples following separation using ultra performance liquid chromatography. A phenyl column was tested with cholesterol first and a gradient method was developed to screen in both positive and negative ionization modes using heated electrospray ionization with optimized parameters in a methanol/water 0.1% formic acid mobile phase system. Extracts were additionally screened using

the same chromatography method with an atmospheric-pressure chemical ionization source and optimized parameters. For data analysis, raw files were converted, peaks centroided and parameters optimized for peak identification in MZmine. The data, collected from three technical replicates in each of four biological replicates (12 data points per peak), were median normalized. A two-tailed *t*-test was performed relative to the Y258 *URA3* control to filter out non-essential features and noise. The normalized data representing each sample and GV control were aligned by mass and retention time using the SIMA algorithm [52], and peak intensities were adjusted by normalizing for different protein levels (Additional file 2: Table S1). Any peak not present in at least 80% of all replicates and not enriched to greater than fourfold relative to the GV control was excluded from further analysis. Both HMDB and METLIN databases were queried with the remaining M/Z values to identify metabolites within a 2-ppm mass tolerance. Unknown features were removed. Features that yielded database hits were compared to the GV control to calculate fold-change enrichment and to facilitate further filtering. Analysis of known chemical standards permitted validation of metabolites identified initially by accurate mass, and of the 27 tested, all 27 matched. Of the 109 features that were significantly enriched, 65 had a high confidence metabolite assignment based upon querying the determined mass against the HMDB and METLIN databases (Additional file 2: Table S1). Hierarchical clustering was performed with Cluster 3.0 [53] and visualized with Treeview [54]. Cytoscape [55] was employed to generate protein–metabolite networks using the edge-weighted force-directed layout based on interaction weight. The high confidence and standard matched protein–metabolite interactions were assigned the identifier EBI-9687416 and deposited to the IntAct database [PMID:22121220] within the International Molecular Exchange (IMEx) Consortium. The IntAct dataset represents a subset of the total interaction data with corresponding unique PubChem identifiers, as liquid chromatography–mass spectrometry was unable to distinguish certain molecules, e.g. PC(20:4(5Z,8Z,11Z,14Z)/21:0) versus PC(21:0/20:4(5Z,8Z,11Z,14Z)).

Three-dimensional structural data analysis

The closest structural homolog for each *Arabidopsis* START domain was identified by fold recognition or threading based predictions using PSIPRED [56,57] with 'Fold Recognition' as the prediction method [57,58]. Three-dimensional coordinates for available crystal structures of START domains, either unbound (mouse STARD4 (Protein Database (PDB) 1JSS) and mouse MLN64 (PDB 1EM2)) or ligand-bound (human PCTP (PDB 1LN1)) were obtained from Brookhaven Protein Structure Database. To construct atomic-resolution models of *Arabidopsis* START domains, the closest structural homologs identified by threading were

used as templates for model building in alignments using MODELLER 9v2 [59]. Structure validation for homology models was carried out with PROCHECK [60] and WHAT IF [61]. For the models that passed structure validation tests, tunnel architectures were examined using VOIDOO [62] with a probe radius of 1.4 Å. The remaining cavities were rendered by MAPMAN [56] and visualized in PyMOL Molecular Graphics System V.1.3 (Schrödinger, LLC, New York, NY) on an Intel Dual Core iMAC installed with OS X version 10.6.5 (2.4 GHz, 4 GB RAM). VOIDOO identified a large number of cavities, of which ambiguous and false positives were filtered out by manual visualization of cavities in the context of predicted cavity lining residues.

Additional files

Abbreviations
AD: activation domain; ANL1: Anthocyanless 1; ATML1: *Arabidopsis thaliana* Meristem Layer 1; Bet_v1: major pollen allergen of *Betula verrucosa* (white birch); bp: base pairs; CAH: congenital lipoid adrenal hyperplasia; CalC: calicheamicin gene cluster protein; CE: cholesterol ester; CERT: ceramide transfer protein; CNA: Corona; CoxG: carbon monoxide dehydrogenase subunit G; DAPI: 4′,6-diamidino-2-phenylindole; DBD: DNA binding domain; EDR2: Enhanced Disease Resistance 2; EYFP: enhanced yellow fluorescent protein; GEV: Gal4 DNA binding domain:estrogen receptor: VP16 activation domain; GFP: green fluorescent protein; GL2: Glabra2; GS: Gal4 DNA binding domain: START domain; GSV: Gal4 DNA binding domain: START domain: VP16 activation domain; GV: GAL4 DNA binding domain: VP16 activation domain; HD: homeodomain; HDG: homeodomain glabrous; hER: human estrogen receptor; HMDB: Human Metabolome Database; IgG: immunoglobulin G; kb: kilobase;

LacZ: gene for β-galactosidase; LCAT: lecithin-cholesterol acyl transferase; LDL: low-density lipoprotein; MLN64: Metastatic Lymph Node 64; MRP7: Mitochondrial Ribosomal Protein 7; PCR: polymerase chain reaction; PCTP: phosphatidylcholine transfer protein; PDB: protein database; PDF2: Protodermal Factor 2; PHB: Phabulosa; PITP: phosphatidylinositol transfer protein; PtCho: phosphatidylcholine; PYL: PYR-like; PYR: pyrabactin resistance; RCAR: regulatory components of abscisic acid receptors; REV: Revoluta; RHO_alpha_C: alpha oxygenase subunit of the Rieske-type non-heme iron aromatic ring-hydroxylating oxgenases; SAD: START adjacent domain; SRPBCC: START/RHO_alpha_C/PITP/Bet v1/CoxG/CalC; StAR: steroidogenic acute regulatory protein; START: StAR-related lipid transfer; SUT1: Sterol Uptake 1; WT: wild type; yEGFP3: yeast-enhanced green fluorescent protein; Zip: leucine zipper; ZLZ: Zip-Loop-Zipper (a plant-specific leucine zipper); β-Gal: β-galactosidase.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
KS conceived of the study, coordinated and performed experiments, and prepared the manuscript. MB and MPS developed and conducted the protein–metabolite immunoisolation work along with the associated data analysis. AK performed and analyzed the domain swap and site-directed-mutagenesis experiments in *Arabidopsis*. PNC, SAM and CH made constructs and performed yeast assays. SAM constructed the START domain deletion for expression in *Arabidopsis*. RAR conducted the mutagenesis experiments with StAR. HCN constructed and analyzed yEGFP3 constructs. DS and GY performed computational modeling and data analysis. Each author contributed to and approved of the final manuscript.

Author details
[1]Division of Biology, Kansas State University, Manhattan, KS 66506, USA. [2]Keck Graduate Institute of Applied Life Sciences, Claremont, CA 91711, USA. [3]Department of Genetics, Stanford University, Stanford, CA 94305, USA. [4]National Institute of Plant Genome Research, New Delhi, India.

References
1. Ponting CP, Aravind L: **START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins.** *Trends Biochem Sci* 1999, **24**:130–132.
2. Lavigne P, Najmanivich R, Lehoux JG: **Mammalian StAR-related lipid transfer (START) domains with specificity for cholesterol: structural conservation and mechanism of reversible binding.** *Subcell Biochem* 2010, **51**:425–437.
3. Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E: **Regulators of PP2C phosphatase activity function as abscisic acid sensors.** *Science* 2009, **324**:1064–1068.
4. Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF, Alfred SE, Bonetta D, Finkelstein R, Provart NJ, Desveaux D, Rodriguez PL, McCourt P, Zhu JK, Schroeder JI, Volkman BF,

Cutler SR: **Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins.** *Science* 2009, **324**:1068–1071.
5. Radauer C, Lackner P, Breiteneder H: **The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands.** *BMC Evol Biol* 2008, **8**:286.
6. Iyer LM, Koonin EV, Aravind L: **Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily.** *Proteins* 2001, **43**:134–144.
7. Schrick K, Nguyen D, Karlowski WM, Mayer KF: **START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors.** *Genome Biol* 2004, **5**:R41.
8. Alpy F, Tomasetto C: **Give lipids a START: the StAR-related lipid transfer (START) domain in mammals.** *J Cell Sci* 2005, **118**:2791–2801.
9. Alpy F, Legueux F, Bianchetti L, Tomasetto C: **START domain-containing proteins: a review of their role in lipid transport and exchange.** *Med Sci (Paris)* 2009, **25**:181–191.
10. Clark BJ: **The mammalian START domain protein family in lipid transport in health and disease.** *J Endocrinol* 2012, **212**:257–275.
11. Bhangoo A, Anhalt H, Ten S, King SR: **Phenotypic variations in lipoid congenital adrenal hyperplasia.** *Pediatr Endocrinol Rev* 2006, **3**:258–271.
12. Letourneau D, Lorin A, Lefebvre A, Frappier V, Gaudreault F, Najmanovich R, Lavigne P, LeHoux JG: **StAR-related lipid transfer domain protein 5 binds primary bile acids.** *J Lipid Research* 2012, **53**:2677–2689.
13. Thorsell AG, Lee WH, Persson C, Siponen MI, Nilsson M, Busam RD, Kotenyova T, Schuler H, Lehtio L: **Comparative structural analysis of lipid binding START domains.** *PLoS One* 2011, **6**:e19521.
14. Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, Rajashankar KR, Cohen DE: **Structure of human phosphatidylcholine transfer protein in complex with its ligand.** *Nat Struct Biol* 2002, **9**:507–511.
15. Romanowski MJ, Soccio RE, Breslow JL, Burley SK: **Crystal structure of the *Mus musculus* cholesterol-regulated START protein 4 (StarD4) containing a StAR-related lipid transfer domain.** *Proc Natl Acad Sci USA* 2002, **99**:6949–6954.
16. Tsujishita Y, Hurley JH: **Structure and lipid transport mechanism of a StAR-related domain.** *Nat Struct Biol* 2000, **7**:408–414.
17. Kudo N, Kumagai K, Tomishige N, Yamaji T, Wakatsuki S, Nishijima M, Hanada K, Kato R: **Structural basis for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide.** *Proc Natl Acad Sci USA* 2008, **105**:488–493.
18. Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE: **Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development.** *Plant Cell* 2005, **17**:61–76.
19. Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T: **Characterization of the class IV homeodomain-leucine zipper gene family in *Arabidopsis*.** *Plant Physiol* 2006, **141**:1363–1375.
20. Abe M, Katsumata H, Komeda Y, Takahashi T: **Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*.** *Development* 2003, **130**:635–643.
21. Lu P, Porat R, Nadeau JA, O'Neill SD: **Identification of a meristem L1 layer-specific gene in *Arabidopsis* that is expressed during embryonic pattern formation and defines a new class of homeobox genes.** *Plant Cell* 1996, **8**:2155–2168.
22. Rerie WG, Feldmann KA, Marks MD: **The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in *Arabidopsis*.** *Genes Dev* 1994, **8**:1388–1399.
23. Di Cristina M, Sessa G, Dolan L, Linstead P, Baima S, Ruberti I, Morelli G: **The *Arabidopsis* Athb-10 (GLABRA2) is an HD-Zip protein required for regulation of root hair development.** *Plant J* 1996, **10**:393–402.
24. Western TL, Burn J, Tan WL, Skinner DJ, Martin-McCaffrey L, Moffatt BA, Haughn GW: **Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in *Arabidopsis*.** *Plant Physiol* 2001, **127**:998–1011.
25. Tang D, Christiansen KM, Innes RW: **Regulation of plant disease resistance, stress responses, cell death, and ethylene signaling in *Arabidopsis* by the EDR1 protein kinase.** *Plant Physiol* 2005, **138**:1018–1026.
26. Talbert PB, Adler HT, Parks DW, Comai L: **The REVOLUTA gene is necessary for apical meristem development and for limiting cell divisions in the leaves and stems of *Arabidopsis thaliana*.** *Development* 1995, **121**:2723–2735.
27. Gao CY, Pinkham JL: **Tightly regulated, beta-estradiol dose-dependent expression system for yeast.** *Biotechniques* 2000, **29**:1226–1231.

28. Alpy F, Tomasetto C: **MLN64 and MENTHO, two mediators of endosomal cholesterol transport.** *Biochem Soc Trans* 2006, **34**:343–345.

29. Ness F, Bourot S, Regnacq M, Spagnoli R, Berges T, Karst F: **SUT1 is a putative Zn[II]2Cys6-transcription factor whose upregulation enhances both sterol uptake and synthesis in aerobically growing *Saccharomyces cerevisiae* cells.** *Eur J Biochem* 2001, **268**:1585–1595.

30. Murcia M, Faraldo-Gomez JD, Maxfield FR, Roux B: **Modeling the structure of the StART domains of MLN64 and StAR proteins in complex with cholesterol.** *J Lipid Research* 2006, **47**:2614–2630.

31. Cormack BP, Bertram G, Egerton M, Gow NA, Falkow S, Brown AJ: **Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in *Candida albicans*.** *Microbiology* 1997, **143**:303–311.

32. Li X, Gianoulis TA, Yip KY, Gerstein M, Snyder M: **Extensive *in vivo* metabolite-protein interactions revealed by large-scale systematic analyses.** *Cell* 2010, **143**:639–650.

33. Gelperin DM, White MA, Wilkinson ML, Kon Y, Kung LA, Wise KJ, Lopez-Hoyo N, Jiang L, Piccirillo S, Yu H, Gerstein M, Dumont ME, Phizicky EM, Snyder M, Grayhack EJ: **Biochemical and genetic analysis of the yeast proteome with a movable ORF collection.** *Genes Dev* 2005, **19**:2816–2826.

34. Kanno K, Wu MK, Agate DS, Fanelli BJ, Wagle N, Scapa EF, Ukomadu C, Cohen DE: **Interacting proteins dictate function of the minimal START domain phosphatidylcholine transfer protein/StarD2.** *J Biol Chem* 2007, **282**:30728–30736.

35. Li X, Snyder M: **Metabolites as global regulators: a new view of protein regulation: systematic investigation of metabolite-protein interactions may help bridge the gap between genome-wide association studies and small molecule screening studies.** *BioEssays* 2011, **33**:485–489.

36. Baker BY, Epand RF, Epand RM, Miller WL: **Cholesterol binding does not predict activity of the steroidogenic acute regulatory protein, StAR.** *J Biol Chem* 2007, **282**:10223–10232.

37. Kallen CB, Billheimer JT, Summers SA, Stayrook SE, Lewis M, Strauss JF: **Steroidogenic acute regulatory protein (StAR) is a sterol transfer protein.** *J Biol Chem* 1998, **273**:26285–26288.

38. Underwood KW, Andemariam B, McWilliams GL, Liscum L: **Quantitative analysis of hydrophobic amine inhibition of intracellular cholesterol transport.** *J Lipid Res* 1996, **37**:1556–1568.

39. Cater NB, Denke MA: **Behenic acid is a cholesterol-raising saturated fatty acid in humans.** *Am J Clin Nutr* 2001, **73**:41–44.

40. Khosla A, Paper JM, Boehler AP, Bradley AM, Neumann TR, Schrick K: **HD-Zip proteins GL2 and HDG11 have redundant functions in *Arabidopsis* trichomes, and GL2 activates a positive feedback loop via MYB23.** *Plant Cell* 2014, **26**:2184–2200.

41. Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF, Bowman JL: **Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes.** *Curr Biol* 2003, **13**:1768–1774.

42. Clough SJ, Bent AF: **Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*.** *Plant J* 1998, **16**:735–743.

43. Jackson CL, Kepes F: **BFR1, a multicopy suppressor of brefeldin A-induced lethality, is implicated in secretion and nuclear segregation in *Saccharomyces cerevisiae*.** *Genetics* 1994, **137**:423–437.

44. Gietz RD, Woods RA: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** *Methods Enzymol* 2002, **350**:87–96.

45. Grebe M, Gadea J, Steinmann T, Kientz M, Rahfeld JU, Salchert K, Koncz C, Jurgens G: **A conserved domain of the *Arabidopsis* GNOM protein mediates subunit interaction and cyclophilin 5 binding.** *Plant Cell* 2000, **12**:343–356.

46. Louvion JF, Havaux-Copf B, Picard D: **Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast.** *Gene* 1993, **131**:129–134.

47. Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K, Shinagawa A, Shinozaki K: **Functional annotation of a full-length *Arabidopsis* cDNA collection.** *Science* 2002, **296**:141–145.

48. Scott SP, Teh A, Peng C, Lavin MF: **One-step site-directed mutagenesis of ATM cDNA in large (20 kb) plasmid constructs.** *Hum Mutat* 2002, **20**:323.

49. Miller JH: **Assay of beta-galactosidase.** In *Experiments in Molecular Genetics*. Cold Spring Harbor, New York: CSH Laboratory Press; 1972:352–355.

50. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31**:3497–3500.

51. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20**:426–427.

52. Voss B, Hanselmann M, Renard BY, Lindner MS, Kothe U, Kirchner M, Hamprecht FA: **SIMA: simultaneous multiple alignment of LC/MS peak lists.** *Bioinformatics* 2011, **27**:987–993.

53. de Hoon MJ, Imoto S, Nolan J, Miyano S: **Open source clustering software.** *Bioinformatics* 2004, **20**:1453–1454.

54. Page RD: **TreeView: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12**:357–358.

55. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, et al: **Integration of biological networks and gene expression data using Cytoscape.** *Nat Protoc* 2007, **2**:2366–2382.

56. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT: **Protein structure prediction servers at University College London.** *Nucleic Acids Res* 2005, **33**:W36–W38.

57. Jones DT: **Protein secondary structure prediction based on position-specific scoring matrices.** *J Mol Biol* 1999, **292**:195–202.

58. McGuffin LJ, Jones DT: **Improvement of the GenTHREADER method for genomic fold recognition.** *Bioinformatics* 2003, **19**:874–881.

59. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.** *J Mol Biol* 1993, **234**:779–815.

60. Laskowski RA, Macarthur MW, Moss DS, Thornton JM: **Procheck - a program to check the stereochemical quality of protein structures.** *J Appl Crystallogr* 1993, **26**:283–291.

61. Vriend G: **WHAT IF: a molecular modeling and drug design program.** *J Mol Graph* 1990, **8**:52–56.

62. Kleywegt GJ, Jones TA: **Detection, delineation, measurement and display of cavities in macromolecular structures.** *Acta Crystallogr D Biol Crystallogr* 1994, **50**:178–185.