Research

START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors

Kathrin Schrick*, Diana Nguyen*, Wojciech M Karlowski† and Klaus FX Mayer†

Addresses: *Keck Graduate Institute of Applied Life Sciences, 535 Watson Drive, Claremont, CA 91711, USA. †Munich Information Center for Protein Sequences, Institute for Bioinformatics, GSF National Research Center for Environment and Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany.

Correspondence: Kathrin Schrick. E-mail: Kathrin_Schrick@kgi.edu

Abstract

Background: In animals, steroid hormones regulate gene expression by binding to nuclear receptors. Plants lack genes for nuclear receptors, yet genetic evidence from Arabidopsis suggests developmental roles for lipids/sterols analogous to those in animals. In contrast to nuclear receptors, the lipid/sterol-binding StAR-related lipid transfer (START) protein domains are conserved, making them candidates for involvement in both animal and plant lipid/sterol signal transduction.

Results: We surveyed putative START domains from the genomes of Arabidopsis, rice, animals, protists and bacteria. START domains are more common in plants than in animals and in plants are primarily found within homeodomain (HD) transcription factors. The largest subfamily of HD-START proteins is characterized by an HD amino-terminal to a plant-specific leucine zipper with an internal loop, whereas in a smaller subfamily the HD precedes a classic leucine zipper. The START domains in plant HD-START proteins are not closely related to those of animals, implying collateral evolution to accommodate organism-specific lipids/sterols. Using crystal structures of mammalian START proteins, we show structural conservation of the mammalian phosphatidylcholine transfer protein (PCTP) START domain in plants, consistent with a common role in lipid transport and metabolism. We also describe putative START-domain proteins from bacteria and unicellular protists.

Conclusions: The majority of START domains in plants belong to a novel class of putative lipid/sterol-binding transcription factors, the HD-START family, which is conserved across the plant kingdom. HD-START proteins are confined to plants, suggesting a mechanism by which lipid/sterol ligands can directly modulate transcription in plants.

Background

The StAR-related lipid transfer (START) domain, named after the mammalian 30 kDa steroidogenic acute regulatory (StAR) protein that binds and transfers cholesterol to the inner mitochondrial membrane [1], is defined as a motif of around 200 amino acids implicated in lipid/sterol binding.
Ligands have been demonstrated for a small number of START-domain proteins from animals. The mammalian STAR and metastatic lymph node 64 (MLN64) proteins both bind cholesterol [3], the phosphatidylcholine transfer protein (PCTP) binds phosphatidylcholine [4], and the carotenoid-binding protein (CBP1) from silkworm binds the carotenoid lutein [5]. In addition, a splicing variant of the human Goodpasture antigen-binding protein (GPBP) called CERT was recently shown to transport ceramide via its START domain [6].

The structure of the START domain has been solved by X-ray crystallography for three mammalian proteins: PCTP [4], MLN64 [3] and StarD4 [7]. On the basis of the structural data, START is classified as a member of the helix-grip fold superfamily, also termed Birch Pollen Allergen v1 (Bet v1)-like, which is ubiquitous among cellular organisms [8]. Iyer et al. [8] used the term ‘START superfamily’ as synonymous with the helix-grip fold superfamily. Here we use the nomenclature established in the Protein Data Bank (PDB) [9] and Structural Classification of Proteins (SCOP) [10] databases, restricting the use of the acronym ‘START’ to members of the family that are distinguished by significant amino-acid sequence similarity to the mammalian cholesterol-binding STAR protein. Members of the START family are predicted to bind lipids or sterols [2,11], whereas other members of the helix-grip fold superfamily are implicated in interactions with a wide variety of metabolites and other molecules such as polyketide antibiotics, RNA or antigens [8].

The presence of START domains in evolutionarily distant species such as animals and plants suggests a conserved mechanism for interaction of proteins with lipids/sterols [2]. In mammalian proteins such as STAR or PCTP, the START domain functions in transport and metabolism of a sterol or phospholipid, respectively. START domains are also found in various multidomain proteins implicated in signal transduction [2], suggesting a regulatory role for START-domain proteins involving lipid/sterol binding.

To investigate the evolutionary distribution of the START domains in plants in comparison to other cellular organisms and to study their association with other functional domains, we applied a BLASTP search to identify putative START-containing protein sequences (see Materials and methods). We focused our study on proteins from the sequenced genomes of Arabidopsis thaliana (Table 1), rice (Table 2), humans, Drosophila melanogaster and Caenorhabditis elegans, as well as Dictyostelium discoideum (Table 3), in addition to sequences from bacteria and unicellular protists (Table 4). CBP1 from the silkworm Bombyx mori was also included in our analysis (Table 3). Figure 1 presents a phylogenetic tree comparing the START domains from the plant Arabidopsis to those from the animal, bacterial and protist kingdoms.

Results and discussion

Evolution of START domains in multicellular organisms

Our findings show that START domain-containing proteins are amplified in plant genomes (Arabidopsis and rice) relative to animal genomes (Figures 1,2). Arabidopsis and rice contain 35 and 29 START proteins each, whereas the human and mouse genomes contain 15 each [11], and C. elegans and D. melanogaster encode seven and four, respectively. In comparison, bacterial and protist genomes appear to encode a maximum of two START proteins (see below).

START-domain minimal proteins comprising the START domain only, as well as START proteins containing additional sequence of unknown or known function appear to be conserved across plants, animals, bacteria and protists (Tables 1,2,3,4, Figure 1). However, only in plants, animals and multicellular protists (D. discoideum) are START domains found in association with domains having established functions in signal transduction or transcriptional control, consistent with the idea that START evolved as a regulatory domain in multicellular eukaryotes. The cellular slime mold D. discoideum, which progresses from unicellular to multicellular developmental stages, contains an unusual START-domain protein [8] which has so far not been found in any other organism: FbxA/CheaterA (ChtA), an F-Box/WD40 repeat-containing protein [12,13]. FbxA/ChtA is thought to encode a component of an SCF E3 ubiquitin ligase implicated in cyclic AMP metabolism and histidine kinase signaling during development [14]. Mutant analysis shows that FbxA/ChtA function is required to generate the multicellular differentiated stalk fate [12].

Functional domains that were found associated with START in animals include pleckstrin homology (PH), sterile alpha motif (SAM), Rho-type GTPase-activating protein (RhoGAP), and 4-hydroxybenzoate thioesterase (4HBT) (Table 3), consistent with a previous report [11]. The RhoGAP-START configuration is absent from plants, but is conserved across the animal kingdom from mammals to insects and nematodes. The RhoGAP-START combination in addition to an amino-terminal SAM domain is apparent only in proteins from humans, mouse, and rat, indicating that SAM-RhoGAP-START proteins are specific to mammals. Similarly, the 4HBT-START combination, also referred to as the acyl-CoA thioesterase subfamily [11], is found exclusively in proteins from humans, mouse and rat, and therefore seems to have evolved in the mammalian lineage.

In humans, about half of the START domain-containing proteins (6/13) are multidomain proteins, whereas in Arabidopsis and rice approximately three-quarters (26/35; 22/29) of START proteins contain an additional domain. The largest proportion of Arabidopsis and rice multidomain START proteins (21/26; 17/22) contain a homeodomain (HD), while a smaller group of proteins (4/26; 4/22) contain a PH domain together with a recently identified domain of unknown
function 1336 (DUF1336) motif. In addition, a single START-DUF1336 protein of about the same size, but lacking strong sequence similarity to PH at its amino terminus, is present in both Arabidopsis and rice. It is striking that the sequence of the START domain correlates with the type of START protein, an indication that evolutionary speciation through duplication and subsequent sequence evolution of START domains took place after initial manifestation of novel protein architecture by domain shuffling.

The position of the START domain in proteins larger than 300 amino acids varies between plant, animal and protist kingdoms. For example, in human proteins, START is always near the carboxy terminus (1-55 amino acids from the end) (Table 3). In plant proteins, however, the START domain is not strictly confined to the carboxy terminus. In both Arabidopsis and rice the START domain can be positioned as much as approximately 470 amino acids from the carboxy terminus (HD-ZIP START proteins: Tables 1,2). Moreover, in a subset...
of plant proteins (PH-START DUF1336 proteins), the START domain is positioned centrally between two different domains. However, defined functional domains are typically amino terminal of the START domain in both animals and plants. By contrast, in the sole example of a START-domain protein in \textit{D. discoideum}, FbxA/ChtA, the START domain is present at the amino terminus, with F-Box and WD\textsubscript{40} domains positioned after it.

**HD-START transcription factors are unique to plants**

The START-domain proteins from \textit{Arabidopsis} were classified into seven subfamilies according to their structures and

### Table 1

START-domain-containing proteins from \textit{Arabidopsis}

| Accession code | Locus      | Other names | Structure | Size (aa) | START position | Chr. | Transmembrane segments |
|----------------|------------|-------------|-----------|-----------|----------------|------|------------------------|
| NP_172015     | At1g05230  |             | HD ZLZ START | 721       | 243-466        | 1    | -                      |
| NP_564041     | At1g17920  |             | HD ZLZ START | 687       | 207-438        | 1    | -                      |
| NP_174724     | At1g34650  |             | HD ZLZ START | 708       | 221-454        | 1    | -                      |
| NP_177479     | At1g73360  |             | HD ZLZ START | 722       | 228-458        | 1    | -                      |
| NP_565223     | At1g79840  | GL2         | HD ZLZ START | 747       | 253-487        | 1    | -                      |
| NP_180796     | At2g32370  |             | HD ZLZ START | 721       | 245-472        | 2    | -                      |
| NP_186976     | At3g03260  |             | HD ZLZ START | 699       | 205-436        | 3    | -                      |
| NP_191674     | At3g61150  |             | HD ZLZ START | 808       | 312-539        | 3    | -                      |
| NP_567183     | At4g00730  | ANL2, AHDP  | HD ZLZ START | 802       | 317-544        | 4    | -                      |
| NP_567274     | At4g04890  | PDF2        | HD ZLZ START | 743       | 245-474        | 4    | -                      |
| NP_193506     | At4g17710  |             | HD ZLZ START | 709       | 230-464        | 4    | -                      |
| NP_193906     | At4g21750  | ATML1       | HD ZLZ START | 762       | 254-482        | 4    | -                      |
| NP_567722     | At4g25530  | FWA         | HD ZLZ START | 686       | 207-435        | 4    | -                      |
| NP_197234     | At5g17320  |             | HD ZLZ START | 718       | 235-462        | 5    | -                      |
| NP_199499     | At5g46880  |             | HD ZLZ START | 820       | 315-549        | 5    | -                      |
| NP_200030     | At5g32170  |             | HD ZLZ START | 682       | 220-427        | 5    | -                      |
| NP_174337     | At1g30490  | PHV, ATHB-9 | HD ZIP START | 841       | 162-375        | 1    | -                      |
| NP_175627     | At1g52150  |             | HD ZIP START | 836       | 152-366        | 1    | -                      |
| NP_181018     | At2g34710  | PHB, ATHB-14 | HD ZIP START | 852       | 166-383        | 2    | -                      |
| NP_195014     | At4g32880  | ATHB-8      | HD ZIP START | 833       | 151-369        | 4    | -                      |
| NP_200877     | At5g60690  | REV, IFL1   | HD ZIP START | 842       | 153-366        | 5    | -                      |
| NP_180399     | At2g28320  |             | PH START DUF1336 | 737   | 171-364        | 2    | -                      |
| NP_191040     | At3g54800  |             | PH START DUF1336 | 733   | 176-370        | 3    | -                      |
| NP_193639     | At4g19040  |             | PH START DUF1336 | 718   | 176-392        | 4    | -                      |
| NP_199369     | At5g45560  |             | PH START DUF1336 | 719   | 176-392        | 5    | -                      |
| NP_568526     | At5g35180  |             | START DUF1336 | 778       | 240-437        | 5    | -                      |
| NP_564705     | At1g53960  |             | START       | 403       | 83-289         | 1    | (i21-43o)              |
| NP_176653     | At1g64720  |             | START       | 385       | 88-297         | 1    | (o20-42i)              |
| NP_850574     | At3g13062  |             | START       | 411       | 84-295         | 3    | (i28-50o)              |
| NP_566722     | At3g23080  |             | START       | 419       | 115-329        | 3    | (i20-42o)              |
| NP_567433     | At4g14500  |             | START       | 433       | 136-345        | 4    | (i21-43o401-423i)     |
| NP_194422     | At4g26920  |             | START       | 461       | 67-228         | 4    | -                      |
| NP_196343     | At5g07260  |             | START       | 541       | 99-296         | 5    | -                      |
| NP_199791     | At5g49800  |             | START       | 242       | 26-217         | 5    | -                      |
| NP_568805     | At5g54170  |             | START       | 449       | 123-337        | 5    | 2 (i7-28o402-424i)    |

GenBank accession codes, locus and other names, structure, total size in amino acids (aa), and position of the START domain are listed. Chr., chromosome number indicates map position. Numbers of predicted transmembrane segments followed by the amino-acid positions separated by ‘i’ if the loop is on the inside or ‘o’ if it is on the outside (in parentheses) are indicated. All proteins are represented by ESTs or cDNA clones.
sizes (Figure 2a). The majority of START domains are found in transcription factors of the HD family. HDs are DNA-binding motifs involved in the transcriptional regulation of key developmental processes in eukaryotes. However, only within the plant kingdom do HD transcription factors also contain START domains (Figure 1). Among around 90 HD family members in Arabidopsis [15], approximately one-quarter (21) contain a START domain. All HD-START proteins contain a putative leucine zipper, a dimerization motif that is not found in HD proteins from animals or yeast. Nuclear localization has been demonstrated for two HD-START proteins: GLABRA2 (GL2) [16] and REVOLUTA/INTERFASCICULAR

Table 2

| indice sequence | japonica ortholog | japonica locus, ID | Other names | Structure | Size (aa) | START position | Chr. | Transmembrane segments | Rice EST/ cDNA | Plant EST |
|-----------------|-------------------|--------------------|-------------|----------|----------|----------------|------|------------------------|---------------|-----------|
| Os002227.2      | NP_915741         | Os01w51311         | OSTF1       | HD ZLZ START | 700*     | 206-430*      | 1*  | -                      | Y (1)         | -         |
| Os014526.1      | BAB92357          | GL2                | HD ZLZ START | 779*     | 270-506*   | 1*  | -                      | Y (1)         | -         |
| Os007627.1      | BAC77158          | ROC5               | HD ZLZ START | 790*     | 294-533*   | 2*  | Y (2)                  | Y (4)         | -         |
| Os000127.7      | CAE02251          | -                  | HD ZLZ START | 851*     | 309-583*   | 4*  | Y (1)                  | Y (8)         | -         |
| Os000666.5      | CAE04753          | ROC2               | HD ZLZ START | 781*     | 284-515*   | 4*  | Y (2)                  | Y (14)        | -         |
| Os017902.1      | -                 | -                  | HD ZLZ START | 616      | 231-471    | 6   | Y (1)                  | Y (1)         | -         |
| Os007245.1      | -                 | -                  | HD ZLZ START | 749      | 259-492    | 8   | Y (15)                 | -             | -         |
| Os010085.1      | BAD01388          | OSJNb0075018       | OCL3        | HD ZLZ START | 786*     | 248-497*      | 8*  | Y (1)                  | -             | -         |
| Os030338.1      | BAB85750          | ROC1               | HD ZLZ START | 784*     | 291-520*   | 8*  | -                      | Y             | -         |
| Os042017.1      | -                 | -                  | HD ZLZ START | 662      | 340-566    | 9   | -                      | Y (1)         | -         |
| Os009778.1      | BAC77156          | ID207863           | ROC3        | HD ZLZ START | 879*     | 338-579*      | 10* | Y (1)                  | Y (4)         | -         |
| Os009778.1      | CAD41424          | ROC4               | HD ZIP START | 806*     | 300-500    | 4*  | Y (1)                  | -             | -         |
| Os000679.2      | BAB92205          | B1015E06           | HD ZIP START | 898*     | 170-413*   | 1*  | Y (15)                 | -             | -         |
| Os003709.4      | AAR04340          | -                  | HD ZIP START | 839*     | 157-370    | 3*  | -                      | Y (2)         | Y (35)    |
| Os007653.1      | AAPS4299          | -                  | HD ZIP START | 840*     | 159-372    | 10* | Y (2)                  | Y (29)        | -         |
| Os008720.4      | -                 | ID213030           | HD ZIP START | 855*     | 170-383    | 12* | Y (38)                 | -             | -         |
| Os006159.2      | AAG43283          | ID214133*          | HD ZIP START | 859*     | 173-386    | 12* | Y (50)                 | -             | -         |
| Os006334.4      | BAD07818          | OJ1435_F07         | PH START DUF1336 | 804*   | 256-453    | 2*  | Y (1)                  | -             | -         |
| Os000253.7      | BAC22213          | Os06w10955         | PH START DUF1336 | 674*   | 210-328    | 6*  | Y (4)                  | -             | -         |
| Os018163.1      | AAPS4082          | ID208089*          | PH START DUF1336 | 705*   | 209-381    | 10* | Y (1)                  | -             | -         |
| -                | AAPS4296          | -                  | PH START DUF1336 | 773*   | 204-398    | 10* | Y (6)*                 | -             | -         |
| Os003769.1      | BAD09877          | -                  | START DUF1336 | 763*     | 228-429    | 8*  | Y (2)                  | Y (2)         | -         |
| Os002751.1      | -                 | ID215312*          | START       | 435*     | 125-312*   | 2   | Y (1)                  | Y (1)         | Y (5)     |
| Os002915.3      | BAD07966          | AP005304           | START       | 419      | 106-327    | 2*  | Y (2)                  | Y (23)        | -         |
| Os005790.2      | CAE01295          | OSJNb0020P07       | START       | 400*     | 90-306*    | 4*  | Y (2)                  | Y (24)        | -         |
| Os007997.2      | -                 | -                  | START       | 366      | 56-164     | 6   | Y (2)                  | -             | -         |
| Os009194.1      | BAC83004          | CPS                | START       | 394*     | 90-300*    | 7*  | Y (1)                  | Y (10)        | -         |
| Os064970.1      | BAC20079*         | Os07w00256*        | START       | 252*     | 71-173*    | 7*  | Y (1)                  | -             | -         |
| Os091586.1      | -                 | -                  | START       | 199      | 30-191     | -   | -                      | -             | -         |

The sequence code for each indica protein is shown together with the accession number (GenBank), locus (MOsDB), and/or identification number (KOME rice full-length cDNA) of the putative japonica ortholog. The structure, total size in amino acids (aa), and position of the START domain are listed. Chr., chromosome number indicates map position. *The japonica ortholog was used for sequence analysis. †Information for both indica and japonica proteins was available for mapping. ‡Partial protein sequence having homology to HD-START proteins. Numbers of predicted transmembrane segments followed by the amino-acid positions, separated by ‘i’ if the loop is on the inside or ‘o’ if it is on the outside (in parentheses), are indicated. The availability of rice and/or plant EST and/or cDNA clones is indicated by a ‘Y’, and the number of independent matching cloned transcribed sequences is given in parentheses.
FIBERLESS1 (REV/IFL1) [17]. Furthermore, canonical DNA-binding sites are reported for GL2 [18] and two other HD-START transcription factors, A. thaliana MERISTEM LAYER1 (ATML1) [19], and PROTODERMAL FACTOR2 (PDF2) [20].

A similar spectrum of START domain-containing proteins is found in Arabidopsis and rice, suggesting their origin in a common ancestor (Figure 2b). The size of the rice genome (450 Mb) is roughly four times that of Arabidopsis (120 Mb). Despite a twofold difference between the total number of

---

Table 3

| Accession code | Locus | Other names | Organism | Structure | Size (aa) | START domain | Transmembrane segments |
|----------------|-------|-------------|----------|-----------|-----------|---------------|------------------------|
| P49675         | StarD1| Star        | Homo sapiens | START     | 285       | 69-281       | -                      |
| Q9UKL6         | StarD2| PCTP        | Homo sapiens | START     | 214       | 7-213        | -                      |
| CAA56489       | StarD3| MLN64       | Homo sapiens | START     | 445       | 235-444     | 4 (o52-74|916o123-145i153-169o) |
| Q99J5          | StarD4| CRSP        | Homo sapiens | START     | 224       | 21-223      | -                      |
| Q9NSY2         | StarD5| -           | Homo sapiens | START     | 213       | 1-213       | -                      |
| P59095         | StarD6| -           | Homo sapiens | START     | 220       | 39-206      | -                      |
| Q9NQZ5         | StarD7| GTT1        | Homo sapiens | START     | 295       | 62-250      | -                      |
| BAA11506       | StarD8| KIAA0189    | Homo sapiens | SAM RhoGAP START | 1132   | 927-1129    | -                      |
| Q9P2P6         | StarD9| KIAA1300    | Homo sapiens | START     | 1820      | 1628-1813   | -                      |
| Q9Y365         | StarD10| SDCCAG28   | Homo sapiens | START     | 291       | 26-226      | -                      |
| Q9Y5P4*        | StarD11| GPBP*      | Homo sapiens | PH Ser-Rich START | 624*   | 395-618     | -                      |
| AAR26717*      | StarD11| CERT*      | Homo sapiens | PH Ser-Rich START | 598*   | 365-589     | -                      |
| Q96Q81         | StarD12| DLC-I      | Homo sapiens | SAM RhoGAP START | 1091  | 879-1084    | -                      |
| AAQ72971       | StarD13| RhoGAP     | Homo sapiens | SAM RhoGAP START | 1113  | 900-1104    | -                      |
| QBWYW0        | StarD14| CACH       | Homo sapiens | 4HBT 4HBT START | 555     | 342-545     | -                      |
| Q8WXI4         | StarD15| THEA       | Homo sapiens | 4HBT 4HBT START | 607     | 378-590     | -                      |
| NP_609644      | CG6565| LDO5321p   | Drosophila melanogaster | START     | 425       | 186-381     | -                      |
| AAR19767       | CG3522| Start†     | Drosophila melanogaster | START     | 583       | 262-362     | 4 (o59-81|102-124o128-150i162-179o) |
| NP_648199      | CG7207| GH07688    | Drosophila melanogaster | START     | 601       | 386-596     | -                      |
| NP_731907      | CG31319| RhoGAP88C  | Drosophila melanogaster | RhoGAP START | 1017  | 806-1007    | -                      |
| NP_492624      | 1K507| F52F1.72   | Caenorhabditis elegans | START     | 241       | 34-237      | -                      |
| NP_510230      | X0324| K02D3.2    | Caenorhabditis elegans | START     | 296       | 69-279      | -                      |
| NP_499460      | 3M432| T2BD6.7    | Caenorhabditis elegans | START     | 322       | 61-262      | -                      |
| NP_505830      | 5L607| C06H2.2    | Caenorhabditis elegans | START     | 397       | 121-337     | -                      |
| NP_498027      | 3F991| F26F4.4    | Caenorhabditis elegans | START     | 447       | 197-446     | 4 (o23-45|65-87o96-118i128-150o) |
| NP_492762      | IL133| F25H2.6    | Caenorhabditis elegans | PH Ser-Rich START | 573     | 338-567     | -                      |
| NP_497695*     | gei-1*| F45H7.2    | Caenorhabditis elegans | RhoGAP START | 722*   | 532-710     | -                      |
| NP_497694*     | gei-1*| F45H7.2    | Caenorhabditis elegans | RhoGAP START | 842*   | 652-830     | -                      |
| BAC01051       | BmCBP| CBP1       | Bombyx mori | START     | 297       | 60-294      | -                      |
| AAD37799       | ChtA | FbxA       | Dicyostelium discoideum | START F-Box WD40 | 1247   | 8-178       | -                      |

GenBank accession codes, locus, other names, and corresponding organism are given for each predicted protein. START domain-containing (StarD) nomenclature is given for the human proteins. The structure, total size in amino acids (aa), and position of the START domain are listed. Numbers of predicted transmembrane segments followed by the amino acid positions separated by ‘i’ if the loop is on the inside or ‘o’ if it is on the outside (in parentheses) are indicated. *There are two protein isoforms as the products of alternative splicing. †The internal loop in the Start1 START domain was not included in the analysis. All proteins are supported by cDNA clones.
predicted genes in rice (ssp. indica: 53,398 [21]) versus *Arabidopsis* (~28,000), the number of START domains per genome appears to be relatively constant: *Arabidopsis* and rice contain 35 and 29 STAR T genes, respectively. Thus, START-domain genes belong to the subset of *Arabidopsis* genes (estimated at two-thirds) that are present in rice [21]. However, one intriguing exception is the apparent absence of rice proteins orthologous to two unusual *Arabidopsis* START proteins (At4g26920 and At5g07260), which share sequence similarity to each other and to members of the HD-ZLZ START subfamily, but lack HD and zipper-loop-zipper (ZLZ) domains (Figure 2b; Tables 1,2). Their absence from rice makes them candidate dicot-specific START proteins.

Screening for expressed sequence tags (ESTs) by BLASTN was conducted to determine whether the types of START sequences from *Arabidopsis* and rice are also present in other plants (see Materials and methods). The screen detected 185 START domain-encoding sequences from a wide assortment of plants representing 25 different species. Consistent with our findings in *Arabidopsis* and rice (Tables 1,2), START domains were found in the plant-specific combinations (HD-START and PH-START) in both dicot and monocot members of the angiosperm division. ESTs for HD-START transcription factors were also identified from the gymnosperm *Picea* abies (AF328842 and AF172931), as well as from a representative of the most primitive extant seed plant, the cycad *Cycas* rumphii (CB093462). Furthermore, a HD-START sequence is expressed in the moss *Physcomitrella patens* (AB032182). Thus it appears that the HD-START plant-specific configuration evolved in the earliest plant ancestor, or alternatively has been retained in the complete plant lineage.

### Two different HD-associated leucine zippers are found in HD-START proteins

Sequence alignments and phylogenetic analysis revealed two distinct classes of HD-START proteins, which differ substantially in their leucine zippers and START domains (Figures 1,2,3). Both types of leucine zipper are unrelated in sequence
Figure 2 (see legend on next page)
to the homeobox-associated leucine zipper (HalZ), which is a plant-specific leucine zipper found in other HD proteins lacking START [22].

Most HD-START proteins (16/21 in Arabidopsis; 12/17 in rice) contain a leucine zipper with an internal loop (defined here as zipper-loop zipper, ZLZ; also termed ‘truncated leucine zipper motif’ [23]) immediately following a conserved HD domain (Figure 3a). The ZLZ motif appears to be less conserved than the classic basic region leucine zipper and seems to be plant specific. It was shown to be functionally equivalent to the HalZ leucine zipper domain for dimerization in an in vitro DNA binding assay [24].

The other HD-START proteins (5/21 in Arabidopsis; 5/17 in rice) contain a classic leucine zipper DNA-binding motif fused to the end of the HD, designated here as ZIP (Figure 3b). This leucine zipper shows strong sequence similarity to the basic region leucine zipper domains (bZIP and BRIZ) [25,26], which have overlapping consensus sequences and are found in all eukaryotic organisms.

Despite these differences, it is likely that both types of HD-START transcription factors originated from a common ancestral gene. They share a common structural organization in their amino-terminal HD, leucine zipper (ZLZ or ZIP) and START domains (Figure 2a). Moreover, the carboxy terminus of HD-ZLZ START proteins (approximately 250 amino acids) shares sequence similarity with the first 250 amino acids of the approximately 470 amino acids at the carboxy terminus of HD-ZIP START proteins. This is exemplified by a comparison between the carboxy-terminal sequences of ATML1 (HD-ZLZ START) and REV (HD-ZIP START), which are 20% identical and 39% similar.

**HD-START proteins are implicated in cell differentiation during plant development**

Several HD-ZLZ START genes correspond to striking mutant phenotypes in Arabidopsis, and for numerous HD-ZLZ START genes, functions in the development of the epidermis have been implicated. Proteins of the HD-ZLZ START subfamily share strong sequence similarity to each other along their entire lengths, including the carboxy-terminal sequence (approximately 250 amino acids) of unknown function that follows the START domain. The HD-ZLZ transcription factors ATML1 and PDF2 appear to be functionally redundant: double-mutant analysis shows that the corresponding genes are required for epidermal differentiation during embryogenesis [20]. The rice HD-ZLZ START protein RICE OUTER-MOST CELL-SPECIFIC GENE1 (ROC1) seems to have an analogous function to ATML1 in that its expression is restricted to the outermost epidermal layer from the earliest stages in embryogenesis [27]. Another HD-ZLZ gene from rice, Oryza sativa TRANSCRIPTION FACTOR 1 (OSTF1), appears to be developmentally regulated during early embryogenesis and is also expressed preferentially in the epidermis [23]. Mutations in Arabidopsis ANTHOCYANINLESS2 (ANL2) affect anthocyanin accumulation and the cellular organization in the root, indicating a role in subepidermal cell identity [28]. The GL2 gene is expressed in specialized epidermal cells and mutant analysis reveals its function in trichome and non-root hair cell fate determination [24,29]. GL2 functions as a negative regulator of the phospholipid signaling in the root [18], raising the possibility that the activity of GL2 itself is regulated through a feedback mechanism of phospholipid signaling through its START domain.

The HD-ZIP START genes characterized thus far are implicated in differentiation of the vasculature. Members of this subfamily are typically large proteins (more than 830 amino acids) that display strong sequence similarity to each other along their entire lengths, including the carboxy-terminal 470 or so residues of unknown function that follow the START domain. Mutations affecting PHABULOSA (PHB) and PHAVOLUTA (PHV), which have redundant functions, abolish radial patterning from the vasculature in the developing shoot, and perturb adaxial/abaxial (upper/lower) axis formation in the leaf [30]. Mutant analysis reveals that REV [31,32], isolated independently as IFL1 [17,33], is also involved in vascular differentiation. Although a mutant phenotype for A. thaliana HOMEobox-8 (ATHB-8) is not reported, its expression is restricted to provascular cells [34] and promotes differentiation in vascular meristems [35].

The presence of the START domain in HD transcription factors suggests the possibility of lipid/sterol regulation of gene transcription for HD-START proteins, as previously hypothesized [2]. One advantage of such a mechanism is that the metabolic state of the cell in terms of lipid/sterol synthesis could be linked to developmental events such as regulation of transcription during differentiation. Changes in the activity of a HD-START transcription factor could be controlled via a
lipid/sterol-binding induced conformational change. For instance, a protein-lipid/sterol interaction involving the START domain may regulate the activity of the transcription factor directly by affecting its DNA-binding affinity or interaction with accessory proteins at the promoter. Alternatively, or in addition, protein-lipid/sterol binding may positively or negatively affect transport or sequestration of the transcription factor to the nucleus.

**PH-START proteins differ in plants and animals**

A subset of animal and plant START proteins contain an amino-terminal PH domain, which is found in a wide variety of eukaryotic proteins implicated in signaling. PH domains are characterized by their ability to bind phosphoinositides, thereby influencing membrane and/or protein interactions [36]. In some cases, phosphoinositide interactions alone may not be sufficient for membrane association, but may require cooperation with other cis-acting anchoring motifs, such as the START domain, to drive membrane attachment.

Although both plant and animal genomes encode START domains in association with an amino-terminal PH domain, the sequences of the PH-START proteins are not conserved between kingdoms (Figure 1; data not shown). In plants, the START domain is adjacent to the PH domain, whereas in animals the PH and START domains are separated by two serine-rich domains [11]. The PH-START protein from humans, GPBP, has serine/threonine kinase activity and Goodpasture (GP) antigen binding affinity, two functions that involve the serine-rich domains. In contrast, the plant PH-START proteins contain a plant-specific carboxy-terminal domain (of around 230 amino acids) of unknown function, DUF1336 (Protein families database (Pfam)) [37]. In addition, amino-terminal sequence analysis (TargetP; see Materials and methods) predicts that three PH-START proteins from Arabidopsis (At3g54800, At4g19040 and At5g43560) and two PH-START proteins from rice (BAD07818 and BAC22213) localize to mitochondria. This suggests a common lipid/sterol-regulated function of these proteins that is related to their subcellular localization.

**Membrane localization of START-domain proteins**

Transmembrane segments may act to tether START-domain proteins to intracellular membranes. One START protein,
MLN64 from humans, is anchored to late endosomes via four putative transmembrane segments that are required for its localization [38,39]. The carboxy-terminal START domain of MLN64 faces the outside of this compartment, consistent with a role in cholesterol trafficking to a cytoplasmic acceptor [38]. Using sequence scanning and alignments with MLN64, putative transmembrane segments in the MLN64-related proteins from D. melanogaster (CG3522/Start1) [40] and C. elegans (3F991) are predicted (Table 3). No other START proteins from animals are predicted to contain transmembrane segments. A single putative amino-terminal transmembrane segment is predicted for one bacterial START protein (Xanthomonas axonopodis XAC0537), while one START protein from the unicellular protist Cryptosporidium parvum (1MX.08) contains seven putative transmembrane segments amino-terminal to the START domain (Table 4). Transmembrane segments are also predicted for several Arabidopsis and rice START proteins (Tables 1,2). Among Arabidopsis START-domain proteins, four (At1g55960, At5g04720, At3g13062, and At3g23080) are predicted to encode a single transmembrane segment at the amino-terminal end of the protein. Two START proteins of similar sequence (At4g14500 and At5g54170) are predicted to encode a putative transmembrane segment at the amino-termini, and the two proteins are predicted to have two putative transmembrane segments. One of the amino-terminal residues of the protein and the second very close to the carboxy terminus. This feature is also found in one rice protein (ID215312), and appears to be plant specific. Possible functions of these putative transmembrane proteins include lipid/sterol transport to and from membrane-bound organelles and/or maintenance of membrane integrity.

Putative ligands for START domains in plants

Cholesterol, phosphatidylcholine, carotenoid and ceramide are examples of lipids that are known to bind START domains from animals [3-6]. We explored the potential for predicting lipid/sterol ligands for plant-specific START domains by aligning START domains with related animal START domains having defined ligands. However, the mammalian MLN64 and STAR START domains, which have been shown to bind cholesterol, do not share convincing sequence conservation with any particular subfamily of Arabidopsis or rice START domains. Similarly, the START domains of CBP1, which binds the carotenoid lutein, and CERT, which binds ceramide, both show insufficient amino-acid conservation when compared in alignment to the most closely related Arabidopsis proteins. Knowledge of the precise ligand contacts within the cavity of these START domains is required to make more accurate predictions about their specific binding properties.

Although most START domains appear to have diverged in the evolution of plants and animals, one notable exception is the preservation of PCTP-like sequences (Figures 1,4a). The major function of mammalian PCTP, a START-domain minimal protein, is to replenish the plasma membrane with phosphatidylcholine and to allow its efflux. Mammalian PCTP binds phosphatidylcholine and its crystal structure in the ligand-bound form has recently been solved [4]. Phosphatidylcholine is the most abundant phospholipid in animals and plants. It is a key building block of membrane bilayers, comprising a high proportion of the outer leaflet of the plasma membrane.

To identify plant START domains that are most similar to mammalian PCTP, we constructed a phylogeny with PCTP against the complete set of 64 START domains from Arabidopsis and rice. Figure 4a shows the branch from this phylogeny grouping mammalian PCTP with START domains from 10 plant proteins. Unlike PCTP, a protein of 214 amino acids, the plant PCTP-like proteins are larger (around 400 amino acids) and contain divergent amino-terminal and carboxy-terminal sequences in addition to the START domain. Most (8/10) of these proteins are predicted to encode an amino-terminal transmembrane segment while a subset (3/8) also encode a putative carboxy-terminal transmembrane segment (Tables 1,2). A sequence alignment of PCTP against the START domains from these 10 PCTP-like proteins is illustrated in Figure 4b. Among 27 amino-acid residues in PCTP that contact the phosphatidylcholine ligand [4] (Figure 4b), 63% (17/27) are similar to residues in one or more of the plant proteins.

Homology modeling was used to generate three-dimensional structures of the 10 PCTP-like START domains from Arabidopsis and rice, incorporating templates from PCTP and two other mammalian START domain structures (MLN64 [3] and StarD4 [7]). The models proved most accurate for the At3g13062 and ID215312 START domains from Arabidopsis and rice, respectively. Figure 4c-f depicts the model of the 212-amino-acid START domain from At3g13062. Overall, the PCTP and At3g13062 START domain sequences are 24% identical and 45% similar.

The backbone of the At3g13062 START domain overlaps with the crystal structure of PCTP START [4]. The model exhibits α helices at the amino and carboxy termini separated by nine β strands and two shorter α helices (Figure 4c). The curved β sheets form a deep pocket with the carboxy-terminal α helix (α4) acting as a lid, resulting in an internal hydrophobic cavity (Figures 4e-g). The amino-terminal α helix (α1), which is in a region of the START domain that is least conserved, is peripheral to the hydrophobic cavity, and is not predicted to contact the ligand. Lipid entry or egress would require a major conformational change, most likely opening or unfolding of the carboxy-terminal α helix lid. In the case of PCTP, the α4 helix has been implicated in membrane binding and phosphatidylcholine extraction [41]. The overall conformation is distinct from other hydrophobic cavity lipid-binding proteins, such as the intracellular sterol carrier protein 2 (SCP2) [42] from animals, and the orthologous non-specific lipid-transfer protein from plants [43]. It is not clear how specific START is for its particular ligand, as the START domain
Figure 4

PTCP-like START domains are conserved in Arabidopsis and rice. (a) Branch of a neighbor-joining phylogenetic tree that was constructed assuming the Poisson correction model and a complete deletion algorithm (bootstrapped 2,000 replicates). Six START domains from Arabidopsis (green lettering), and four from rice (white lettering in green boxes) are similar to human PCTP (white lettering in red box). (b) Alignment of the 10 PCTP-like START domains from Arabidopsis and rice against the START domain of human PCTP (highlighted in red). Yellow highlighting indicates PCTP residues contacting the α1, β1, β2, β3, β4, β5, and β6 on one end of the hydrophobic cavity while α2 and α3 form an interior hydrophobic chamber that can accommodate a single ligand molecule. (d) Top of the hydrophobic tunnel with a clear view of α4, the carboxy-terminal α helix comprising the lid. α4 contains a conserved kink at glycine 280. (e) Lateral view showing the basket structure formed by the antiparallel β sheets [β4, β5], and β6 on one end of the hydrophobic cavity while β1, β2, β3, β7, β8, and β9 form the other side. (f) View through the empty START domain showing that the amino-terminal α1 helix is not predicted to contact the ligand within the hydrophobic interior.
of human StAR has been shown in vitro to have affinity for both cholesterol and sitosterol [44], the latter being the major plant sterol. However, modeling of highly similar START domains based on the crystal structure of PCTP implicates a similar molecular ligand, and suggests that these proteins have retained a common function in evolution.

**START domain-containing genes in unicellular organisms**

Of about 100 bacterial genomes sequenced, only eight species encode proteins with similarity to START, whereas one-half (4/8) of unicellular protist genomes thus far sequenced appear to contain a START domain (Figure 1, Table 4). Expression data regarding the putative START-encoding genes from bacteria and unicellular protists is lacking to date. Most of the bacterial sequences represent START-domain minimal proteins of less than 245 amino acids. A START gene from the opportunistic pathogen *Pseudomonas aeruginosa* (PA1579) was previously hypothesized to have arisen by horizontal gene transfer [7,8]. Here we show that similar START-domain genes also occur in one other pathogenic *Pseudomonas* species as well as in two related nonpathogens. The citrus-tree pathogen *X. axonopodis* and human pathogen *Vibrio vulnificus* also contain START genes. Nonpathogenic species that contain START genes include the anaerobic green sulfur bacterium *Chlorobium tepidum* and the anaerobic dehalogenating bacterium *Desulfitobacterium hafniense*. Among unicellular protists, the mammalian pathogens *C. parvum*, *Giardia lamblia*, *Plasmodium falciparum*, and *Plasmodium yoelii yoelii* contain START genes. By contrast, the genome of the unicellular green alga *Chlamydomonas reinhardtii* lacks START domains (Shinhan Shiu, personal communication).

START domains appear to be absent from the Archaea. Moreover the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, as well as another fungal species, *Neurospora crassa*, lack START domains. Perhaps START-domain genes are rarely found in prokaryotes and protists because their function is not required for normal unicellular proliferation. Their presence in a number of pathogenic species is consistent with the hypothesis of horizontal acquisition. This raises the possibility that in pathogens the START-domain proteins have a role in the interaction with host cells. For example, pathogens may recognize, transport and utilize lipids/sterols from the host organism via their START domains.

**Conclusions**

In this study we have explored putative START domains from plants in comparison to animal, protist and bacterial START domains. The function of START in binding lipid/sterol ligands has been demonstrated in only a few cases in the context of mammalian cells. One START domain configuration appears to be conserved from animals to plants: mammalian PCTP, which specifically binds phosphatidylycholine and exhibits sequence similarity to six *Arabidopsis* and four rice START domains. Using three-dimensional molecular modeling we predict that the plant proteins will bind orthologous lipids/sterols. Thus our finding provides a rationale for experimentally testing potential lipid/sterol interactions with PCTP-like START proteins. The majority of START domains appear to have evolved divergently in animals and plants. This is exemplified by RhoGAP START proteins, which are found in animals but are absent from plants. Strikingly, plant genomes contain two subfamilies of HD-START proteins that are not found in animals. Our results illustrate that these plant-specific START domains are amplified and conserved in dicot and monocot plant genomes, and that they originated in an ancient ancestor of the plant lineage. The modular coupling detected raises the possibility of regulation of signal transduction by lipids/sterols in plants. Although HD-START proteins are unique to plants, the presence of the lipid/sterol-binding domain in a transcription factor has a parallel in the animal kingdom (Figure 5). In animals, steroid-binding transcription factors of the nuclear receptor superfamily control numerous metabolic and developmental processes. The structure of a classic nuclear receptor consists of amino-terminal transactivation and zinc-finger DNA-binding domains and a carboxy-terminal steroid hormone receptor domain of around 240 amino acids [45]. Steroid nuclear receptors are complexed with the proteins Hsp90, Hsp70, immunophilins or cyclophilins in the cytoplasm and upon steroid hormone binding move into the nucleus to bind specific DNA response elements for transcriptional control of target genes. The
presence of HD-START transcription factors in plants leads us to postulate that these have an analogous role by binding regulatory lipids/sterols. The next hurdle in elucidating the role of the START domain in plant signal transduction will be the identification of in vivo binding partners.

**Materials and methods**

**Searching databases for putative START proteins and transcribed sequences**

Putative START proteins were identified by BLASTP against databases at the National Center for Biotechnology Information (NCBI) [46], The Institute for Genomic Research (TIGR) [47], and the Munich Information Center for Protein Sequences (MIPS) [48]. Putative START proteins from rice *Oryza sativa* ssp. *indica* were identified by BLASTP of sequences predicted on whole-genome shotgun deposited in MIPS *Oryza sativa* database (M0sDB) [49] using START-domain proteins from *Arabidopsis* (GL2, ATML1), rice (ROC1, AAP54082) and humans (PCTP) as query. START-domain proteins were verified using InterPro [50] and NCBI Conserved Domain (CD)-Search [51,52], using cutoff E-values set to 1e-05. *Arabidopsis* proteins were checked against *Arabidopsis* EST databases using WU-BLAST2 [53] to identify corresponding transcribed sequences. Supporting mRNA expression data was obtained from the microarray elements search tool [54] available through The *Arabidopsis* Information Resource (TAIR). Rice ESTs and cDNAs were identified using BLASTN at E-values of < 1e-60 (with most sequences showing an E-value = 0) and other plant ESTs were analyzed using TBLASTN at an E-value threshold of < 1e-80. Putative START protein-encoding plant cDNAs were identified from Sputnik EST database [55] and NCBI using *Arabidopsis* and rice START proteins as query. START domains were identified from the following plant species: *Arabidopsis thaliana*, *Beta vulgaris*, *Capsicum annuum*, *Cycas rumphii*, *Glycine max*, *Gossypium arboreum*, *Gossypium hirsutum*, *Helianthus annuus*, *Hordeum vulgare*, *Lactuca sativa*, *Lotus japonicus*, *Lycopersicon esculentum*, *Malus domestica*, *Medicago sativa*, *Medicago truncatula*, *Oryza sativa*, *Phalaenopsis sp.*, *Physcomitrella patens*, *Picea abies*, *Prunus persica*, *Solanum tuberosum*, *Sorghum bicolor*, *Triticum aestivum*, *Zea mays* and *Zinnia elegans*.

Signal peptide and transmembrane analysis of the predicted START proteins was accomplished with TargetP [56] and TMHMM [57] using default settings. For the human MLN64 protein, and the MLN64-like proteins from *D. melanogaster* (CG3522/Start1) and *C. elegans* (3F991), sequence alignments were utilized to confirm the number and coordinates of the transmembrane segments. Experimental data for MLN64 [38] was incorporated to predict the orientation of the transmembrane segments.

**Phylogenetic analysis and sequence alignments**

Alignments were made using ClustalW [58] with BLOSUM62 settings. Alignments of START domains were converted into the Molecular Evolutionary Genetics Analysis (MEGA) format in MEGA2.1 [59,60]. For sequences originating from the same species, alignments were made using PileUp from Genetics Computer Group (GCG), screening for similarities and identities manually. After eliminating redundancies, datasets were realigned with ClustalW and a phylogenetic tree was derived. Phylogenies containing sequences from different organisms were constructed by the neighbor-joining method, assuming that sequences from different lineages evolve at different rates, followed by bootstrapping with 2,000 replicates [60,61]. The Poisson correction was used to account for multiple substitutions at the same site. Gaps were handled using complete deletion, assuming that different amino-acid sequences evolve differently. Alignments were constructed by GCG's PileUp to examine trends seen in the phylogenies. For analysis of proteins from different organisms, pairwise deletion was used instead of complete deletion. Domain alignments and consensus sequences shown in Figures 3 and 4 were constructed using GCG’s Pretty [62].

**Tertiary structure analysis of START domains**

Homology modeling of *Arabidopsis* and rice PCTP-like proteins was accomplished using the SWISS-MODEL web server [63,64] and DeepView (Swiss-Pdb Viewer) [65]. PDB templates of structures used for homology modeling were PCTP (PDB ID 1LN1A), MLN64 (1EM2A), and StarD4 (1JSSA). The WhatCheck summary reports for the highest likelihood models for *Arabidopsis* and rice PCTP-like proteins, At3g13062 and ID215312, respectively, were as follows: structure Z-scores: 1st generation packing quality: -1.779, -2.132; 2nd generation packing quality: -4.818, -4.425; Ramachandran plot appearance: -2.674, -2.393; chi-1/chi-2 rotamer normality: 0.138, 1.450; backbone conformation: -1.861, -1.446; and root mean square (RMS) Z-scores: bond lengths: 0.785, 0.859; bond angles: 1.389, 1.243; omega angle restraints: 1.007, 0.960; side chain planarity: 2.374, 1.346; improper dihedral distribution: 1.877, 1.544; inside/outside distribution: 1.324, 1.269. Coordinates and complete WhatCheck Reports are available on request. Cartoon images of the models were produced in RasMol Version 2.7.2.1.

**Additional data files**

The following additional data are available with the online version of this article: a pdb file (Additional data file 1), the WhatCheck Report (Additional data file 2) and the log (Additional data file 3) for the 3D model of the At3g13062 START domain; a pdb file (Additional data file 4), the WhatCheck Report (Additional data file 5) and the log (Additional data file 6) for the 3D model of the ID215312 START domain; a text file (Additional data file 7) and a MEGA file (Additional data file 8) of the alignment corresponding to the tree in Figure 1; a text file (Additional data file 9) and the MEGA file...
Acknowledgements

The authors are grateful to Animesh Ray for the generous use of his facilities at the Keck Graduate Institute. We thank David Wild for help with protein modeling and Shinnan Shiou of the Department of Ecology and Evolution at the University of Chicago for searching the Chiromonomas reinhardtii genome database at the Department of Energy Joint Genome Institute [http://www.jgi.doe.gov for START domains]. We thank Elliot Meyerowitz, Animesh Ray, and David Wild for valuable discussions and critical reading of the manuscript, and Raymond Soccio for many helpful suggestions. This work was supported by the National Science Foundation (Animesh Ray: NSF EIA-0205061 and NSF EIA-0130059), a grant from the Deutsche Forschungsgemeinschaft (SCHR 744/1–2), and a research fellowship to K.S. (SCHR 744/1–1–3).

References

1. Stocco DM: StAR protein and the regulation of steroid hormone biosynthesis. Annu Rev Physiol 2001, 63:193-213.
2. Ponting CP, Aravind L: A START: a lipid-binding domain in STAMPs, HD-ZIP and signaling proteins. Trends Biochem Sci 1999, 24:130-132.
3. Tsujishita Y, Hurley JH: Structure and lipid transport mechanism of a START-related domain. Nat Struct Biol 2000, 7:408-414.
4. Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, Rajashekkaran S, Krunic J, Cohen DE: Structure of the human phosphatidylcholine transfer protein in complex with its ligand. Nat Struct Biol 2002, 9:507-511.
5. Tabunoki H, Yagi Y, Tanaka Y, Fuji H, Banno Y, Joumi ZE, Koba-yashi M, Sato R, Masahara H, Taichida K, Isolation, characterization, and cdna sequence of a cartoendin binding protein from the silk gland of Bombyx mori larvae. J Biol Chem 2002, 277:32133-32140.
6. Hanada K, Kumagai K, Yasuda S, Miura Y, Kawano M, Fukasawa M, Nishijima M: Molecular machinery for non-vesicular trafficking of ceramide. Nature 2003, 426:803-809.
7. Romanowski MJ, Soccio RE, Breslow JL, Burley SK: Crystal structure of the Mus musculus cholesterol-regulated START protein 4 (Stard4) containing a START-related lipid transfer domain. Proc Natl Acad Sci USA 2002, 99:6949-6954.
8. Iyer L, Koonin EV, Aravind L: Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 2001, 43:141-144.
9. The RCSB Protein Data Bank [http://www.rcsb.org/pdb]
10. SCOP: structural classification of proteins [http://scop.mrc-lmb.cam.ac.uk/scop]
11. Soccio RE, Breslow JL: START-related lipid transfer (START) proteins: mediators of intracellular lipid metabolism. J Biol Chem 2003, 278:22183-22186.
12. Ennis HL, Doo DN, Bukatzki SU, Kessin RH: Dictyostelium amoebae lacking an F-box protein form spores rather than stalk in chimera with wild-type. Proc Natl Acad Sci USA 2000, 97:3292-3297.
13. Nelson MK, Clark A, Abe T, Nomura A, Yadasa N, Funair CJ, Jermyn KA, Mohanty S, Firtel RA, Williams JG: An F-box/WD-40 repeat-containing protein important for Dictyostelium cell-type proportioning, slug behaviour, and culmination. Dev Biol 2000, 224:42-59.
14. Tekluay T, Ennis HL, Wu MY, Nelson M, Kessin RH, Ratner DI: Genetic interactions of the E3 ubiquitin ligase component FbxA with cyclic AMP metabolism and a histidine kinase signaling pathway during Dictyostelium discoideum development. Eukaryot Cell 2002, 1:618-626.
15. Riechmann JL: Transcriptional regulation: a genomic overview. In The Arabidopsis Book 2002 [http://www.aspb.org/downloads/arabidopsis/riechn.pdf]. Rockville, MD: American Society of Plant Biologists.
Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al.: The Pfam contribution to the annual NAR database issue. *Nucleic Acids Res* 2004, 32 Database issue:D138-D141.

38. Aly F, Stoeckel ME, Dierich A, Escola JM, Wendling C, Chenard MP, Vanier MT, Gruenberg J, Tomasetto C, Rio MC: The steriodogenic acute regulatory protein homolog MLN64, a late endosomal cholesterol-binding protein. *J Biol Chem* 2001, 276:4261-4269.

39. Rho G-Lutz C, Tomasetto C, Regnier CH, Wendling C, Lutz Y, Muller D, Chenard MP, Basset P, Rio MC: MLN64 exhibits homology with the steriodogenic acute regulatory protein (StAR) and is over-expressed in human breast carcinomas. *Int J Cancer* 1997, 71:183-191.

40. Roth GE, Gierl MS, Vollborn L, Meise M, Lintermann R, Korge G: The *Drosophila* gene Start1: a putative cholesterol transporter and key regulator of ecystosteroid synthesis. *Proc Natl Acad Sci USA* 2004, 101:1601-1606.

41. Feng L, Chan VW, Roderick SL, Cohen DE: High-level expression and mutagenesis of recombinant human phosphatidyl transfer protein using a synthetic gene: evidence for a C-terminal membrane binding domain. *Biochemistry* 2000, 39:15399-15409.

42. Choinowski T, Hauser H, Piontek K: Structure of sterol carrier protein 2 at 1.8 Å resolution reveals a hydrophobic tunnel suitable for lipid binding. *Biochemistry* 2000, 39:1897-1902.

43. Lee JY, Min K, Cha H, Shin DH, Hwang KY, Suh SW: Rice non-specific lipid transfer protein: The 1.6 Å crystal structure in the unliganded state reveals a small hydrophobic cavity. *J Mol Biol* 1998, 276:437-448.

44. Kallen C, Billheimer JT, Summers SA, Stayrook SE, Lewis M, Strauss JF 3rd: Steroidogenic acute regulatory protein (StAR) is a sterol transfer protein. *J Biol Chem* 1998, 273:26285-26288.

45. Laudet V, Gronemeyer H: The Nuclear Receptors FactsBook. London: Academic Press; 2002.

46. NCBI BLAST [http://www.ncbi.nlm.nih.gov/blast]

47. Arabidopsis thaliana sequence BLAST search [http://tigrblast.tigr.org/euk-blast]

48. MATDB BLAST query [http://mips.gsf.de/proj/thal/db/search/blast_arabi.html]

49. Karlowski WM, Schoof H, Janakiraman V, Stuempflen V, Mayer KF: MOsDB: an integrated information resource for rice genomics. *Nucleic Acids Res* 2003, 31:190-192.

50. InterPro: home [http://www.ebi.ac.uk/interpro]

51. NCBI CD-search [http://www.ncbi.nih.gov/Structure/cdd/wrpsb.cgi]

52. Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, et al.: CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* 2003, 31:383-387.

53. TAIR WU-BLAST 2.0 form [http://www.arabidopsis.org/wublast/index2.jsp]

54. TAIR: microarray elements search and download [http://www.arabidopsis.org/tools/bulk/microarray/index.html]

55. The Sputnik resource for the annotation of clustered plant ESTs [http://mips.gsf.de/proj/sputnik]

56. TargetP server v1.01 [http://www.cbs.dtu.dk/services/TargetP]

57. TMHMM server v.2.0 [http://www.cbs.dtu.dk/services/TMHMM]

58. ClustalW [http://www.ebi.ac.uk/clustalw/index.html]

59. Molecular Evolutionary Genetics Analysis MEGA [http://www.megasoftware.net]

60. Kumar S, Tamura K, Jakobsen IB, Nei M: MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 2001, 17:1244-1245.

61. Nei M, Kumar S: *Molecular Evolution and Phylogenetics* Oxford: Oxford University Press; 2000.

62. Devereux J, Haeblerl P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 1984, 12:387-395.

63. SWISS-MODEL [http://swissmodel.expasy.org]

64. Schwede T, Kopp J, Guex N, Peitsch MC: SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 2003, 31:3381-3385.

65. Guex N, Peitsch MC: SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* 1997, 18:2714-2723.