J Mol Model (2012) 18:4249–4262
DOI 10.1007/s00894-012-1419-y

ORIGINAL PAPER

In silico and *in vivo* studies of an *Arabidopsis thaliana* gene, *ACR2*, putatively involved in arsenic accumulation in plants

Noor Nahar · Aminur Rahman · Maria Moś · Tomasz Warzecha · Maria Algerin · Sibdas Ghosh · Sheila Johnson-Brousseau · Abul Mandal

Received: 21 February 2012 / Accepted: 26 March 2012 / Published online: 6 May 2012

Abstract Previously, our *in silico* analyses identified four candidate genes that might be involved in uptake and/or accumulation of arsenics in plants: arsenate reductase 2 (*ACR2*), phytochelatin synthase 1 (*PCS1*) and two multi-drug resistant proteins (*MRP1* and *MRP2*) [Lund et al. (2010) J Biol Syst 18:223–224]. We also postulated that one of these four genes, *ACR2*, seems to play a central role in this process. To investigate further, we have constructed a 3D structure of the *Arabidopsis thaliana* ACR2 protein using the iterative implementation of the threading assembly refinement (I-TASSER) server. These analyses revealed that, for catalytic metabolism of arsenate, the arsenate binding-loop (AB-loop) and residues Phe-53, Phe-54, Cys-134, Cys-136, Cys-141, Cys-145, and Lys-135 are essential for reducing arsenate to arsenic intermediates (arsenylated enzyme-substrate intermediates) and arsenite in plants. Thus, functional predictions suggest that the ACR2 protein is involved in the conversion of arsenate to arsenite in plant cells. To validate the *in silico* results, we exposed a transfer-DNA (T-DNA)-tagged mutant of *A. thaliana* (mutation in the *ACR2* gene) to various amounts of arsenic. Reverse transcriptase PCR revealed that the mutant exhibits significantly reduced expression of the *ACR2* gene. Spectrophotometric analyses revealed that the amount of accumulated arsenic compounds in this mutant was approximately six times higher than that observed in control plants. The results obtained from *in silico* analyses are in complete agreement with those obtained in laboratory experiments.

Keywords Arsenate reductase · *In silico* studies · *In vivo* analyses · ICP-DRC-MS · *Arabidopsis thaliana*

N. Nahar · A. Rahman · M. Algerin · A. Mandal (✉)
School of Life Sciences, University of Skövde,
PO Box 408, 541 28 Skövde, Sweden
e-mail: abul.mandal@his.se

M. Moś · T. Warzecha
Department of Plant Breeding and Seed Science,
University of Agriculture in Krakow,
Krakow, Poland

S. Ghosh · S. Johnson-Brousseau
Department of Natural Sciences and Mathematics,
Dominican University of California,
San Rafael, CA 94901, USA

Introduction

Millions of people worldwide are exposed, directly or indirectly, to arsenic (As) poisoning from drinking of contaminated water and/or consuming crops cultivated by irrigation with As-contaminated groundwater. Long-term exposure to As leads to several kinds of skin lesions, such as hyper- and hypo-pigmentation, hyperkeratosis and melanosis, as well as gangrene, skin cancer, lung cancer and bladder cancer [2, 3]. One possible strategy to avoid food-based As contamination in humans is to use genetically engineered plants for phytoremediation [4, 5], or to develop safe crop varieties that can be grown in contaminated environments [1]. Both these scenarios require understanding of the mechanisms of As-metabolism in plants in As-stressed conditions [6]. This requires specific information about the various components involved in catalysis, the mechanism of catalysis and overall function of the As reductase system. A structural study of this system would be a very useful step towards achieving this goal.

During the past decade, considerable progress has been made towards identification of the molecular mechanisms of the uptake, metabolism and translocation of As in plants [7]. After entering the roots via high affinity

phosphate transporters, arsenate is reduced readily to arsenite. The chemical similarity of arsenate and phosphate enables arsenate ions to enter the cell through the phosphate transport system [8]. Alternatively, under reducing conditions, arsenite is taken up by membrane proteins belonging to the aquaporin family [9]. Thus, it is clear that arsenic accumulation, translocation and sequestration are linked closely to the forms of arsenite that exist in plant tissues. Trivalent arsenite is most easily trapped/stored in the roots through vacuolar sequestration of thiol conjugates. On the other hand, under aerobic conditions, much of the As coming into the cells is in the form of pentavalent arsenate. However, high concentrations of arsenate are toxic to cells. Therefore, the reduction of arsenate to arsenite, which can be sequestered in the vacuoles of the plant cells, is a key step in As metabolism. To date, several means of arsenate reduction have been described. In *Escherichia coli* uses the glutaredoxin (Grx)-pathway [10, 11], which involves redox cycling of glutathione with glutaredoxin and glutathione reductase. In this pathway, arsenate ions ($H_2AsO_4^-$) are reduced to arsenite ions (AsO_2^-), which can bind glutathione for enzyme activity. Another pathway has been reported in *Staphylococcus aureus* [12] that employs coenzyme-A disulfide reductase with thioredoxoin for arsenate reductase activity. However, in the eukaryotic yeast *Saccharomyces cerevisiae*, the arsenic detoxification system a uses Grx-coupled arsenate reductase with the help of a thiol cysteine cascade to reduce arsenate to arsenite [13].

Few studies on ACR2-type arsenate reductases have been reported so far in eukaryotic systems. The arsenic reductases, *Saccharomyces cerevisiae* arsenic reductase2 protein (*ScAc2p*) [14] and *Leismania major arsenic reductase 2* (*LmACR2*) [15] are two examples. In the model plant *Arabidopsis thaliana*, arsenate reductase belongs to the ACR family and shares homology with the catalytic domain of the cell division cycle 25 (CDC25) protein that has been shown to reduce arsenate to arsenite [4]. This enzyme, also called *A. thaliana* arsenate reductase 2 (AtACR2), *A. thaliana* arsenate reductase (AtASR), or *A. thaliana* dual specificity phosphatase CDC25 (AtCDC25), has been identified and characterized as an arsenate reductase both in micro-organisms and in plants [4, 10, 11]. Disruption of the arsenate reductase gene leads to arsenate hypersensitivity in *E. coli*, *S. cerevisiae* and *A. thaliana* [4, 10].

Genome sequencing projects are producing linear amino acid sequences, and recent years have seen the development of a variety of techniques for predicting protein function based on three dimensional (3D) structures [16–18]. Experimental structure determination methods are providing high-resolution structural information about a subset of the proteins, whereas computational structure prediction methods provide valuable information for the large fraction of sequences whose structures cannot be determined experimentally [19]. The first class of protein structure prediction methods includes threading and comparative modeling, and is based on detectable similarity covering most of the modeled sequence and at least one sequence with known structure. The second class of methods includes de novo [20, 21] or ab initio methods [22, 23] that can predict a structure from a sequence alone, without relying on similarity, at the fold level, between the modeled sequence and any of the known structures. In addition, bioinformatics tools have enormous analytical and predictive potential in many areas, particularly in the fields of biochemistry, biomedicine and molecular biology. Significantly, with the help of these tools, one can predict and establish correlations between the structure and the function of the query molecule within a very short time. In our investigations, we believe that bioinformatics can provide deeper insight into the structural and functional aspects of arsenate reductase and thys help our understanding of the molecular mechanisms of As accumulation in plants.

To date, the structure of arsenate reductase and its catalytic site(s) in the model plant *A. thaliana* have been studied only poorly. Here, we report the results of *in silico* studies (validated by *in vivo* experiments) on the structure and catalytic activities of the ACR2 protein from *A. thaliana*. We propose a two-step catalytic pathway of the arsenate reductase enzyme, which is involved in metabolizing the pentavalent form of arsenic (arsenate, AsV) to trivalent arsenite (AsIII). Our work encompasses (1) *in silico* insights into the 3D structure, active site machinery and enzyme-substrate interactions of *A. thaliana* arsenate reductase enzyme; and (2) extensive *in vivo* laboratory experiments that validate the *in silico* results.

Materials and methods

Threading and structural association

The *A. thaliana* ACR2 amino acid sequence was retrieved from the NCBI protein sequence database (http://www.ncbi.nlm.nih.gov/protein?term=ACR2). The protein consists of 146 amino acids and its accession number is AED 90607.1.

To predict the secondary structure of the target protein ACR2, we used two programs available on the I-TASSER server [17]: position specific iterated prediction (PSI-PRED) [24, 25], and position specific iterated-BLAST (PSI-BLAST) [26]. In the first step, the query sequence was aligned against a non-redundant sequence database using PSI-BLAST, and the secondary structure was then predicted

using PSI-PRED based on a sequence profile generated by multiple alignments of the sequence homologs.

Another program on the I-TASSER server, local meta-threading-structure (LOMETS) [27], was used to thread the target protein. LOMETS is a locally installed meta-threading program combining ten state-of-art threading programs (COMA, FUGUE, HHSEARCH, MUSTER, PROSPECT2, PPA-1, HHSEARCH 1 SAM, SP3 and SPARKS) [28–35]. The predicted secondary structure of the query sequence was threaded through a representative library of PDB structures using LOMETS. The templates were ranked based on sequence-based profile–profile alignment, secondary structure match, structure-based profile–profile alignment, solvent accessibility match, backbone torsion angle match and general hydrophobicity scoring matrix. The top-ranked template hits obtained from each threading program were selected for further analysis.

Structure assembly and refinement

Assembly and refinement of the target protein structure was achieved using the programs Support Vector Machine (SVMSEQ) [36, 37] and SPICKER [38]. SPICKER is a simple and well-organized clustering program, in which clustering is performed in a one-step procedure using a contracted but representative set of decoy conformations; the pairwise root mean squared deviation (RMSD) cutoff is determined by self-adjusting iteration during computer simulations. Continuous fragments excised from the threading templates were used to assemble full-length models, with unaligned loop regions built by ab initio modeling in a lattice system [39]. The structure assembly simulations (for both the threading-aligned and the ab initio modeled regions) were guided by a unified knowledge-based force field, which included three components: (1) general knowledge-based statistics terms from the PDB C_α/side-chain correlations [40], H-bonds and hydrophobicity [41]; (2) spatial restraints from threading templates [42]; and (3) sequence-based contact predictions from SVMSEQ. The structural assembly process (Fig. 1) consisted of two sets of simulations. The first set was used to generate initial structures for threading templates. In the second set, the simulations started from the cluster centroids generated by SPICKER, which clustered all trajectories obtained in the first set of simulations. Spatial restraints collected from the PDB structure hits by TM-align [43] using the cluster centroids as query structures, were also incorporated into I-TASSER simulations. The purpose of the second stage was to refine the local geometry as well as the global topology of the SPICKER centroids.

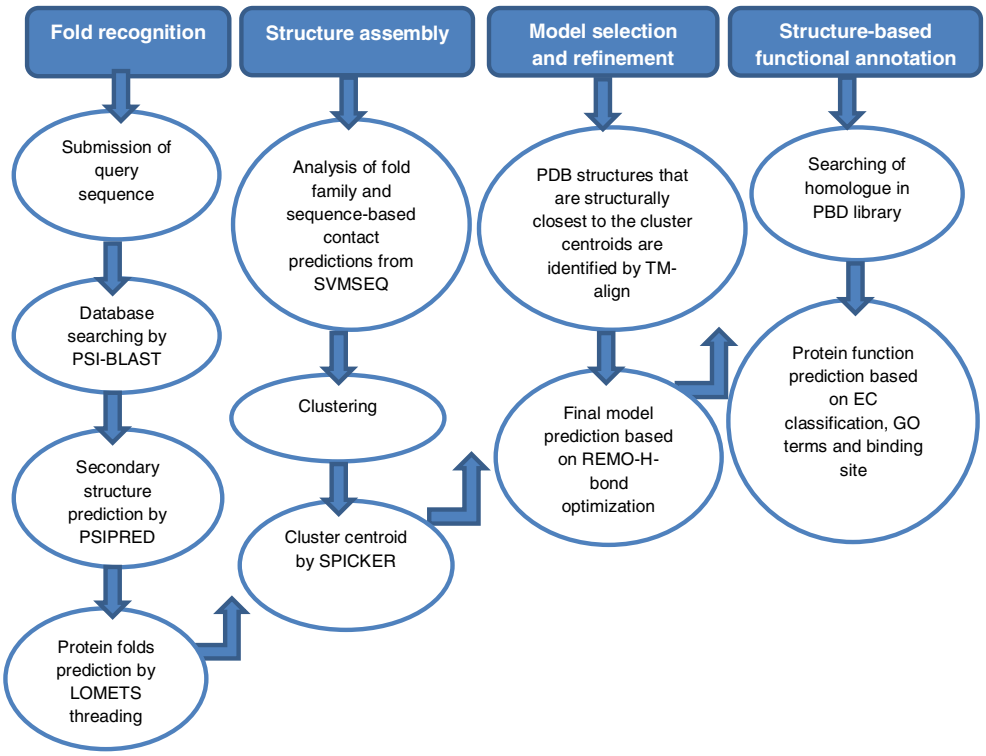

Fig. 1 Four major steps in the automated strategy for assigning structure and function of proteins using the iterative implementation of the threading assembly refinement (I-TASSER) protocol. *Step 1* Fold recognition, utilizes mainly three programs PSI-BLAST, PSIPRED and LOMETS. *Step 2* Structure assembly, involves two programs-SVMSEQ and SPICKER. *Step 3* Model selection and refinement, consists of two programs TM- align and REMO, respectively. *Step 4* Structure-based functional annotation, is based on enzyme commission (EC) classification, gene ontology (GO) terms and substrate binding sites

Modeling and ranking of 3D structure for prediction of protein function

The program REMO [44] was used to construct full-atomic models of the target protein. This program relies on constructing full-atomic models from C-alpha traces by optimizing H-bond networks. In our computer simulation analyses we used a secondary structure-specific backbone isomer. This isomer was constructed using the program (PSI- PRED). We used a reduced modeling simulation process to construct a full-atomic model of our protein (ACR2). In this process, the important intermediate step in protein structure prediction is to construct an appropriate full-atomic model from the reduced one. In reducing representation, models are generated based on the main simulated points, which are the Cα and side-chain centers. These have to be moved around so that the physical full atomic models with regular bond-length and bond-angle can be constructed. Thus, the better global topology of full atomic models depends on movement of the Cα atoms in the correct direction. In our present study, estimation of the movement of Cα atoms was verified by Cα-RMSD and TM-scores (results are presented in Table 2) relative to the native structure. The basic backbone fragments (C_α, C, N, and O) were matched from a secondary structure-specific backbone isomer library that consisted of a total of 68,206 non-redundant isomers from high-resolution PDB structures. The driving force in the REMO refinement protocol included H-bonding, clash/break-amendment, I-TASSER restraints and the CHARMM22 potential [45]. CHARMM 22 is used to identify the potential energy functions with the help of mathematical equations and corresponding force field parameters of biomolecules. On the basis of a test set of 230 non-homologous proteins, REMO had the ability to remove steric clashes while retaining a topology score (e.g., TM-score) similar to that of cluster centroids. The reduced models from I-TASSER were ranked based on the structure density in SPICKER clusters. The weights of the empirical score had been trained in benchmark tests [46]. The highest scoring 3D models were selected and compared with the proteins of known structure and function in the PDB. Three protein structures and/or function libraries were constructed independently for matching the predicted 3D models against known protein structures in PDB. The structural analogs of the query protein in the gene ontology (GO) library were matched based on the global topology using TM-align. A consensus was derived based on the frequency of occurrence of the GO terms. The structural analogs in the enzyme commission (EC) numbers and the binding site libraries were matched based on both global and local structural similarity.

Validation and evaluation of the model by laboratory experiments

To validate the *in silico* results, we performed laboratory experiments. These experiments were based on analysis of the model plant *A. thaliana*. Three types of this plant were used in these experiments: (1) wild type (Wt), ecotype Columbia-0; (2) mutant type (Mt) (T-DNA tagged SALK mutant) [47]; and (3) a vector transformed transgenic control line (VC). The SALK_ 005882C mutant was selected based on database information derived from the SIGnAL Arabidopsis T-DNA collection (http://signal.salk.edu/cgi-bin/tdnaexpress, verified 28 January 2011). It contains a T-DNA insertion in the promoter region of the *A. thaliana ACR2* gene (*At5g03455*). According to the information provided by The Arabidopsis Information Resource (TAIR), the genome of the SALK mutant (SALK_ 005882C) is homozygous with respect to T-DNA insertion (http://www.arabidopsis.org/servlets/TairObject?type=germplasm&id=1010320016). Wt and VC were used as controls. The results obtained in the SALK mutant were always compared with those from the control materials.

Preparation of plant materials

Seeds of *A. thaliana* Wt and VC (control plants) were obtained from our collection at the University of Skövde, Sweden, and seeds of SALK mutants (Mt) were collected from The European Arabidopsis Stock Centre, Nottingham (NASC, http://arabidopsis.info/). Prior to germination, all seeds were vernalized for 7 days at 4 °C in the dark. The seeds were germinated and grown in soil pots for 3 weeks and then some leaves/ whole plants were collected and submerged in liquid nitrogen and stored at −80 °C. These materials were used for isolation of DNA or RNA. To grow the plants in axenic conditions, the seeds were surface sterilized with 70 % ethanol for 5 min and, after five washes with sterile distilled water, the seeds were sown on Murashige-Skoog, MS basal salt medium (Sigma-Aldrich, St. Louis, MO; Cat. No. M5524). The medium was supplemented with 20 g/L sucrose and solidified with 8 g/L agarose (Sigma, Cat. No. A9539-500G). The pH of the medium was adjusted to 5.8. The soil pots and the Petri dishes were placed in separate plant growth chamber (growth cabinet Model- MLR-350, Sanyo, Osaka, Japan) maintaining a photoperiod of 16 h day, temperature 22 °C and a relative humidity of approximately 70 %. The SALK mutant and VC plants were grown on selective MS medium containing 50 mg/L kanamycin sulfate (Sigma-Aldrich, Cat. No. 60615). For exposure of plants to As, the MS medium was

supplemented with various concentrations of sodium arsenate dibasic heptahydrate/ $Na_2HAsO_4 \cdot 7H_2O$ (Sigma-Aldrich, Cat. No. A6756) and/ or sodium (meta) arsenite/ $NaAsO_2$ (Sigma-Aldrich, Cat. No. 71287) as a source of inorganic arsenics. Plants were exposed to different concentrations arsenate (25 μM or 100 μM) or arsenite (25 μM or 100 μM) for 3 weeks.

Verification of number of T-DNA copies and homozygocity in SALK mutants

To determine T-DNA copy number in the SALK mutant, we performed an inverse polymerase chain reaction (I-PCR). I-PCR is a special type of PCR based on amplifying self-ligated circular DNA by using only one known DNA sequence as primer. I-PCR is therefore very effective in identifying the number of T-DNA copies in the transgenic plants. For amplification of T-DNA flanking plant DNA sequences, primers can be designed only from the known sequences of the T-DNA but not from the unknown sequences of plant genomic DNA. Genomic DNA from Wt and Mt plants was isolated using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany, Cat. No. 69104) and digested with *Bam*HI (unique restriction site for T-DNA) from New England Biolabs (http://www.neb.com). Self-ligation of 200 ng DNA was performed by using 2U T_4 DNA ligase (New England Biolabs, Cat. No. M0202T) in a 50 μl reaction. T-DNA specific primers (Sigma-Aldrich) for SALK_ 005882C (forward primer 5′-AATGTATAATTGCGGGACTCTAA-3′; reverse primer 5′-TTTCCCCGTCAAGCTCTAAA-3′) were used to perform I-PCR. The PCR reactions were carried out as described by Svensson et al. [48] except the annealing temperature was adjusted to 60 °C for 1 min. The amplified I-PCR fragment was analyzed by agarose gel electrophoresis.

To verify if the SALK mutant was homozygous with respect to T-DNA insertion, we performed PCR using two T-DNA flanking plant genomic DNA sequences as primers. The forward primer (ACR2-F: 5′- ATCGAACAC GCGACCTTC-3′) was designed from the promoter regions, whereas the reverse primer (ACR2-R: 5′-ATGATGGC GATGTTGGGAC-3′) was constructed from the coding regions of the *ACR2* gene of *A. thaliana*. The PCR reaction was carried out in a total volume of 20 μl containing 10 ng plant DNA, 2.0 mM dNTPs, 1X phire plant PCR buffer, 0.5 μM primers and 1.0 U Hot-start DNA polymerase, (Finnzymes Cat. No. F-130; http://finnzymes.com). PCR was performed in a Piko Thermocycler, (Finnzymes) with an initial denaturation at 98 °C for 5 min followed by 34 cycles of 98 °C for 5 s, 59 °C for 5 sec and 72 °C for 2 min. The final elongation step was maintained at 72 °C for 10 min. The PCR products were analyzed by agarose gel electrophoresis.

Analysis of gene expression by RT-PCR

In order to verify whether the target gene *ACR2* (*At5g03455*) in the SALK_005882C plants was either inactivated (knock-out) or had reduced activity (knock-down), we performed reverse transcription-PCR (RT-PCR). Total RNA from 3-week old *Arabidopsis* (whole plants, shoots and roots) was isolated using an RNeasy Plant Mini Kit (Qiagen, Valencia, CA; Cat. No. 74904). The RNA was then treated with RNase free DNase (Qiagen Cat. No. 79254). The integrity and intensity of RNA were estimated by gel electrophoresis (1 μg RNA on a 1.2 % denaturing agarose gel). RT-PCR was performed using a Robust–TI RT-PCR Kit (Finnzymes, Espoo, Finland; Cat. No. F-580L). For each sample, the total volume of reaction mix was adjusted to 50 μl. The reaction mix contained 2,000 ng RNA, 1.5 mM $MgCl_2$, 1X PCR buffer, 200 mM each dNTP, 10 pmol of forward and reverse primers, 5U AMV-RT and 2U DyNAzyme EXT DNA polymerase. Tubes with reaction mixes were loaded in a Thermocycler (Flexcycler, Biostep, Jahnsdorf, Germany: Block assembly 96G). For synthesis of the first strand cDNA, the cycle setup was adjusted to 48 °C for 45 min. Second strand synthesis was carried out with an enzyme inactivation and initial denaturation at 94 °C for 2 min followed by 35 cycles of 94 °C for 30 sec, 63.1 °C for 30 sec and 72 °C for 1 min 20 sec. The final elongation step was maintained at 72 °C for 10 min. The following primers were used for RT-PCR: Art-Forward 5′-AAATGG CGATGGCGAGAA-3′ and Art-Reverse 5′-AATCGCCCTTGCAAGGAAC-3′. The primers were specific to the first and the third exons of *ACR2* gene. As a control, the transcription level of the *A. thaliana* actin gene was analyzed separately using two primers: 5′-TGCCAATCTACGAGGGTTTCTCTC-3′ and 5′-TGTACTTCCTTTCAGGTGGTGCAA-3′. RT-PCR was performed as described above except that the annealing temperature was adjusted to 61 °C for 30 sec. From each amplified PCR product, 10 μl was analyzed by agarose gel electrophoresis.

Analysis of arsenic accumulation in *A. thaliana* by ICP-DRC-MS

We performed inductively coupled plasma dynamic reaction cell mass spectrometry (ICP-DRC-MS) to measure the amount of arsenics accumulated in the plants. Seeds of Mt, Wt and VC plants were grown on MS medium without As. For exposure of plants to arsenics, 7-day old seedlings (15 seedlings for each treatment with eight replications) were then transferred to MS medium containing either arsenate (25 μM or 100 μM) or arsenite (25 μM or 100 μM). Plants were collected after 3 weeks and washed three to four times with deionized water to remove any surface contamination.

Plants were separated into two parts: shoots and roots. All plant materials were then oven-dried at 55 °C for 4 days. Dried tissues were ground using a mortar and pestle and stored at 4 °C until further use. For determination of the amount of arsenics in plants, all samples were prepared inside a trace metal free clean laminar flow hood. A known mass (approximately 0.120 g) of each sample was digested in a hot block digestion apparatus with aliquots of HNO_3 and H_2O_2 as described in EPA Method 3050B [49]. The resulting sample digests were analyzed for determination of total arsenic contents by ICP-DRC-MS. Prior to sample analysis, all calibration curves were verified using the second source of standards known as initial calibration verification standards (ICV).

Results

Protein homology modeling

Prior to prediction of the secondary structure of the protein sequence of ACR2, a profile-profile alignment was performed with 5,798 non-redundant protein structures in I-TASSER using PSI-BLAST. The secondary structure was then predicted using PSI-PRED. In the predicted structure, all α-helices, β-sheets and coiling were identified along with their confidence scores (Fig. 2). After prediction of secondary structure by PSIPRED, the query sequence was first threaded through the PDB library using LOMETS. In the initial step of LOMETS, to get the best performance from the ten state-of-art threading programs, the top ten models were taken from each program based on sequence identity coverage and combined energy Z-score values. These results are presented in Table 1. The fold 1t3kA, the top-ranked threading fold with 99 % sequence identity of the templates in the threading aligned region with the query sequence, was recognized by seven individual threading programs in LOMETS. All the threading folds were justified based on Z-scores. The higher the Z-score, the higher the significance level of the predicted structure. LOMETS also revealed three threading folds (ranked 8, 9 and 10 in Table 1) with Z-scores > 1 but, with low sequence identity ranging from 0.16 to 0.17, these were considered insignificant (Table 1). Template modeling align (TM-align) was performed to identify false positive matches and to highlight true positive ones. The output of this structural alignment was ranked based on the TM-score. These results are shown in Table 2. 1t3kA—a dual-specificity CDC25 phosphatase containing the signature motif (HCxxxxxR) with binding site—was ranked in first place (TM-score 0.7903; RMSD 2.05). For this reason, 1t3kA was selected as the first candidate for a possible native structure of the query sequence.

Prediction of 3D structure

Using the top ten threading folds as the native structures, five top-ranked 3D models were generated in the I-TASSER server (Fig. 3a). The accuracy of each model was estimated based on C-score, TM-score, RMSD and cluster density. According to Fig. 3a, the model 1 with the highest *C*-score (−0.19) reflects a model of better quality with structural similarity between the predicted model and the native structure (TM-score =0.69±0.12 and RMSD=5.1±3.3). For further prediction of the function of the query protein, the functional analogs and the confidence scores were estimated based on EC number, GO terms functional homology scores (Fh-score) and ligand binding sites.

Structure-based functional annotation

Prediction of EC number

The 3D structure of model 1, which has the highest C-score (−0.19, Fig. 3a) was further analyzed by large-

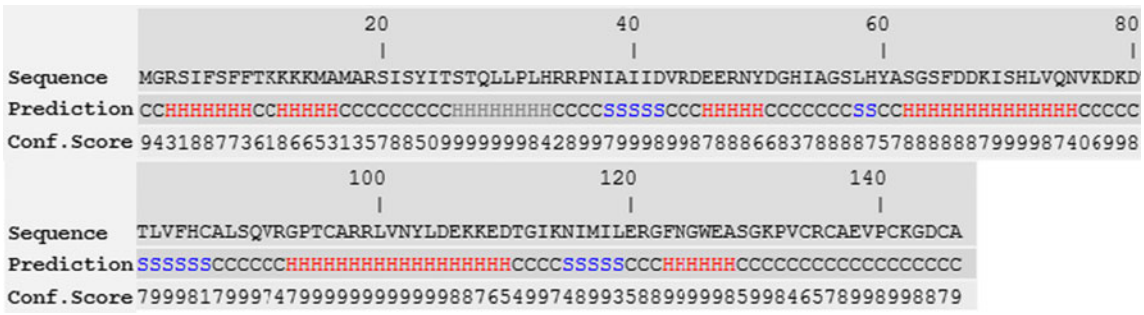

Fig. 2 Secondary structure of ACR2 protein predicted by PSI-PRED. *Red* Alpha helix (H), *blue* beta sheet (S), *black* coil (C). *Conf. score* Confidence score of each amino acid ranging from 1 to 9

Table 1 Accuracy of the predicted model-1 based on Z-score. The top ten threading folds were used by I-TASSER to construct the protein model. Alignment with a normalized Z-score >1 indicates a good alignment

Rank	Threading folds by PDB hit	Iden1[a]	Iden2[b]	Cov.[c]	Norm. Z-score[d]	Description and putative function of the identified folds
1	1t3kA	0.99	0.90	0.90	3.65	Dual-specificity CDC25 phosphatases—critical positive regulators of cyclin-dependent kinases (CDKs)
2	1t3kA	0.99	0.90	0.90	3.60	
3	1t3kA	0.99	0.90	0.90	3.38	
4	1t3kA	0.99	0.90	0.90	2.31	
5	1t3kA	0.99	0.90	0.90	3.13	
6	1t3kA	0.99	0.90	0.90	3.03	
7	1t3kA	0.99	0.90	0.90	2.38	
8	1qb0A	0.17	0.22	1.00	4.34	In Cdc25B, both sulfate and tungstate anions are shown to bind in the catalytic site containing the signature motif (HCxxxxxR) in a conformation similar to that of other protein tyrosine phosphatases and dual specificity phosphatases
9	1c25	0.16	0.24	1.00	3.96	Cdc25 phosphatases activate cell division kinases throughout the cell cycle. It is identical to rhodanese, a sulfur-transfer protein with only the active-site loop, containing the Cys-(X)5-Arg motif, shows similarity to the tyrosine phosphatases
10	1cwsA	0.16	0.22	0.98	2.47	Cdc25B, both sulfate and tungstate anions are shown to bind in the catalytic site containing the signature motif (HCxxxxxR) in a conformation similar to that of other protein tyrosine phosphatases and dual specificity phosphatases, with the exception of the Cdc25A

[a] Percentage of sequence identity of the templates in the threading aligned region with the query sequence

[b] Percentage of sequence identity of the whole template chains with query sequence

[c] Coverage of threading alignment = number of aligned residues divided by the length of query protein

[d] Normalized Z-score of the threading alignments

scale benchmarking tests for prediction of potential enzyme analogs. These analyses revealed that the I-TASSER-predicted model 1 matches with five enzyme analogs, and all five of these analogs exhibit almost identical EC nos.: either 3.1.3.16 or 3.1.3.48 (Table 3). However, the EC-scores of these analogs differed. The highest EC-score found in the PDB hit 1t3kA was 3.4296, whereas the lowest (0.9553) was observed in PDB hit 1cwsA (Table 3). According to the ExPaSy enzyme database, the name of enzyme analog 1t3kA is phosphotyrosine phosphatase (*PTPase*). PTPase belongs to a hydrolase family and acts on ester bond formation.

Table 2 Accuracy of structural analogs of the predicted model-1 based on TM-score

Rank	TM-score[a]	RMSD[b]	IDEN[c]	Cov.[d]	PDB Hit
1	0.7903	2.05	0.99	0.90	1t3kA
2	0.6515	3.82	0.14	0.92	1cwsA
3	0.6396	3.82	0.15	0.91	1c25A
4	0.6301	2.70	0.24	0.82	2j6pA
5	0.6256	2.94	0.23	0.81	3fs5A
6	0.6238	3.81	0.16	0.88	3op3A
7	0.5869	3.53	0.15	0.82	3hzuA
8	0.5854	4.16	0.14	0.87	3ippB
9	0.5719	3.14	0.13	0.76	3nhvA
10	0.5684	3.35	0.11	0.79	1orbA

[a] Measure of structural similarity between the predicted model and the native structure

[b] Root mean squared deviation (deviation between residues that are structurally aligned by TM-align)

[c] Percentage sequence identity in the structurally aligned region

[d] Coverage of the alignment by TM- align = number of structurally aligned residues divided by length of the model

Prediction of GO term and GO-score

To identify the functional similarity between the predicted structure and the detected enzyme analogs, we predicted the functional Fh-scores and associated GO terms (data not shown). To obtain a more reliable function of the enzyme analogs and to avoid unanimity of functionalities we performed a consensus prediction of the GO terms and determined their GO-scores. Table 4 shows the GO terms and GO-scores predicted separately for molecular function, biological process and cellular localization of the enzyme analogs. The higher the GO-score, the higher the confidence of the prediction. The

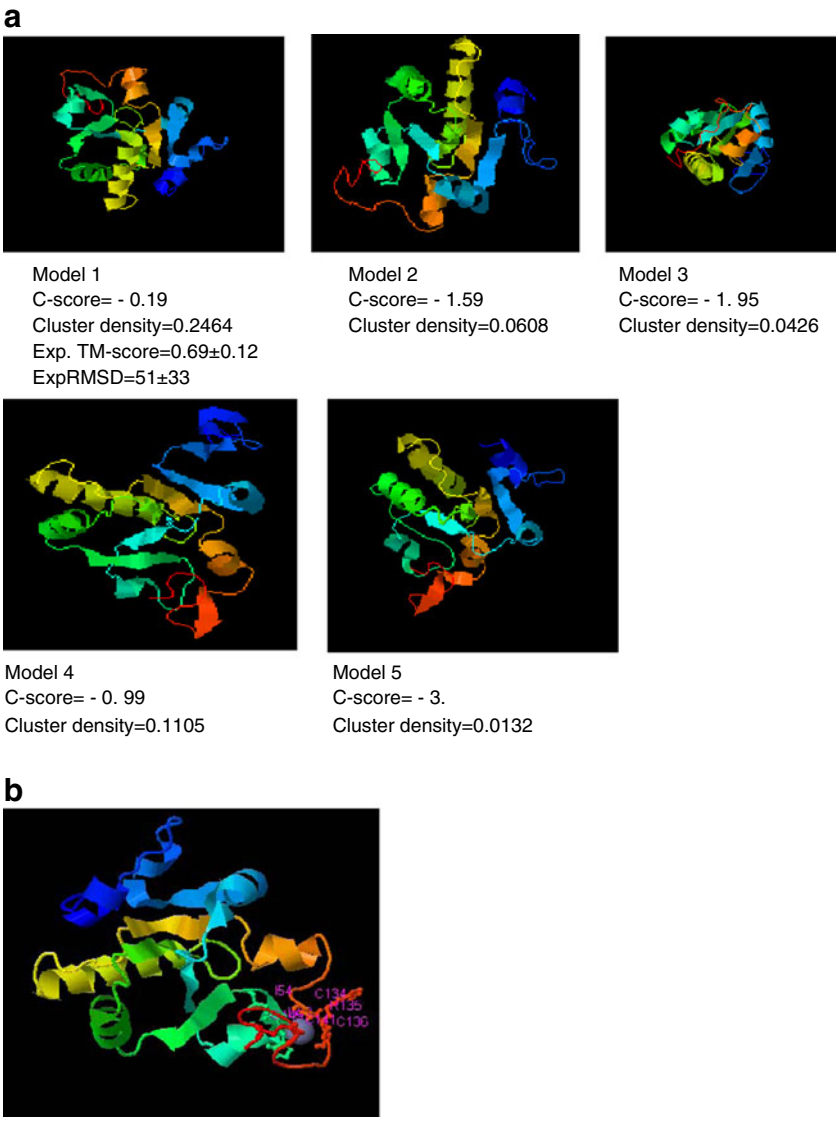

Fig. 3 **a** The top five 3D models of the ACR2 protein predicted by I-TASSER. The models were numbered according to their predicted ranking. The lowest number of the model (model 1) indicates the highest ranking by I-TASSER, whereas the highest number (model 5) represents the lowest ranking. Ranking was determined by the quality of the model which was estimated based on C-score, cluster density, TM-score and RMSD. The quality of a model increases with the increase of C-score and cluster density. Further refinement of the model is done by considering TM-score and RMSD. The C-score values of the models were estimated based on Z-score and the convergences of I-TASSER simulations as indicated in Table 1. **b** The final ribbon representation of ACR2 protein derived from model 1 as indicated in Fig. 3a. The *red ribbon* (AB loop) refers to the $HC(X)_5$ R motif and the substrate binding site residues are Phe-53, Phe-54, Cys-134, Cys-136, Cys-141, Cys-145, and Lys-135 (*magenta*)

highest GO-score (0.737) was found in GO term 0004725, which indicates that the target protein has phosphor-tyrosine activity. However, the lowest GO-score (0.368) was found in GO term 0016783, which is involved in sulfurtransferase activity. The highest GO-score (0.728) for biological process was obtained in GO term 0034960, which suggests that the target protein is involved in cellular macromolecule metabolism. Identical GO-scores (0.641) were found in GO terms 006796, 0043687 and 0043412, which are involved in phosphate metabolism, post-transcriptional amino acid modification and macromolecule metabolism, respectively. The highest

Table 3 Determination of the functional analogs of model 1 based on enzyme classification (EC) score. EC-Score is estimated based on the C-score and the similarity of model 1 with known enzyme structures determined using both global and local structural alignment programs. Ranking of the model was based on EC-score. The higher the EC-score, the higher the confidence. A prediction with an EC-score > 1.1 signifies a prediction with high confidence (up to 3 digit numbers of EC)

Rank	TM-score	RMSD	IDEN	Cov.	EC-Score	PDB Hit	EC No.
1	0.7903	2.05	0.99	0.90	3.4296	1t3kA	3.1.3.48
2	0.6067	3.03	0.25	0.78	1.1393	3f4aB	3.1.3.16
3	0.6396	3.82	0.15	0.91	0.9652	1c25A	3.1.3.48
4	0.6553	3.72	0.14	0.92	0.9606	2uzqF	3.1.3.48
5	0.6515	3.82	0.14	0.92	0.9553	1cwsA	3.1.3.48

Table 4 Determination of molecular function, biological process and cellular location of I-TASSER model1 based on GO-score. The GO-score associated with each prediction was defined as the average weight of the GO term, where the weights are assigned based on the Fh-score of the template from which the GO term was derived. A prediction with a GO-score > 0.5 signifies a prediction with high confidence

Molecular function		Biological process		Cellular location	
GO term	GO-score	GO term	GO-score	GO term	GO-score
GO:0004725	0.737	GO:0034960	0.728	GO:0005623	1.152
GO:0005488	0.709	GO:0000278	0.651	GO:0005622	1.152
GO:0004792	0.368	GO:0006796	0.641	GO:0005737	0.950
GO:0016740	0.368	GO:0043687	0.641	GO:0043229	0.873
GO:0016783	0.368	GO:0043412	0.641	GO:0005634	0.709

GO-score (1.152) for cellular location indicates that the target protein is located intracellularly.

Prediction of ligand binding site

For further verify protein function, we predicted the ligand binding sites of the query protein using I-TASSER. This prediction was based on determination of the binding site scores, which rely on the local sequence and structural similarity between the binding sites of the templates and the query structure. The highest confidence of the binding site score (1.9077) was observed when the template 1t3kA was compared with the query protein, whereas the lowest (0.6894) was found when the template 1okgA was compared with the query protein (Table 5). Based on the highest confidence of binding site score (1.9077), seven active binding residues of the query protein were identified. These residues are Phe-53, Phe-54, Cys-134, Cys-136, Cys-141, Cys-145, and Lys-135 (Table 5, Fig. 3b).

Estimation of gene function

Based on the results predicted by *in silico* analyses, we propose a possible pathway of arsenic reduction in *A. thaliana* (Fig. 4). This pathway is based on several redox reactions; we hypothesize that the first step (arsenylated enzyme–substrate complex formation, ES) involves activated Cys-136 and Cys-145 thiolates attacking the substrate, arsenate, to form an ES complex with the help of Phe-53 and Phe-54 as a general acid in an in-line associative manner. A water molecule leaves the pathway to facilitate the formation of the complex [10–14]. The second step (the arsenate reduction) involves four catalytically essential cysteines, Cys-134, Cys-136, Cys-141 and Cys145 (Fig. 3b), identified as two pairs of a conserved cysteine redox relay system to produce Cys-134–Cys-136 and Cys-141–Cys145 disulfide bonds and arsenite ions. The most important residue identified in this mechanism is the positively charged Lys-135, which plays a role in stabilizing the arsenate binding (AB) loop and binding the arsenate ion.

Analysis of SALK mutant

Determination of T-DNA copy number in the SALK mutant by I-PCR

To investigate the function of the *ACR2* gene in *A. thaliana*, we included the SALK_005882C mutant type (Mt) in our experiment. According to database information this mutant (SALK_005882C) harbors a T-DNA tag in the promoter region of the *ACR2* gene. To verify the number of T-DNA copies in this mutant, we performed I-PCR using T-DNA specific primers as described in "Materials and methods". The I-PCR experiments resulted in amplification of a single fragment of DNA of the expected size (3.5 kb, data not shown).

Table 5 Determination of substrate binding site residues of model 1 based on BS-score—a measure of local sequence and structure similarity between template's binding site and predicted binding site in the query structure. Based on large scale benchmarking analysis, binding site predictions of BS-score >1.1 signify predictions with high confidence

Rank	PDB hit	TM-score	RMSD	IDEN	Cov.	BS-Score	Lig Name	Predicted binding site residues in the modle.
1	1t3kA	0.7903	2.0500	0.9920	0.9041	1.9077	ZN	53,54,134,135,136,141,145
2	1yt8A	0.5656	3.2300	0.2120	0.7671	1.1642	SO3	
3	1gx7A	0.5080	3.9600	0.0750	0.7740	0.8963	SF4	85,86,87,88,89,90,91,92,93,95,96,118
4	3kt8C	0.5248	4.4400	0.0760	0.8699	0.7410	LTN	41,42,43,44,60,62,68,84,86,99,104
5	1qxnA	0.5280	3.5800	0.1540	0.7466	1.1648	PS5	86,87,88,89,90,121,122,123
6	1e08A	0.5211	4.0300	0.0440	0.7808	1.1213	POL	87,88,120,121,122,123,125
7	3kt0A	0.4923	4.5900	0.0780	0.8562	0.7904	EMN	41,42,43,44,65,68,84,99,100
8	1e08A	0.5211	4.0300	0.0440	0.7808	0.8762	SF4	43,49,54,56,57,59,85,86,123
9	2uzqC	0.6547	3.7100	0.1390	0.9178	1.1431	PO4	121,122,123,124,125
10	1okgA	0.5758	3.6300	0.1090	0.8014	0.6894	SO3	48,59,85,86

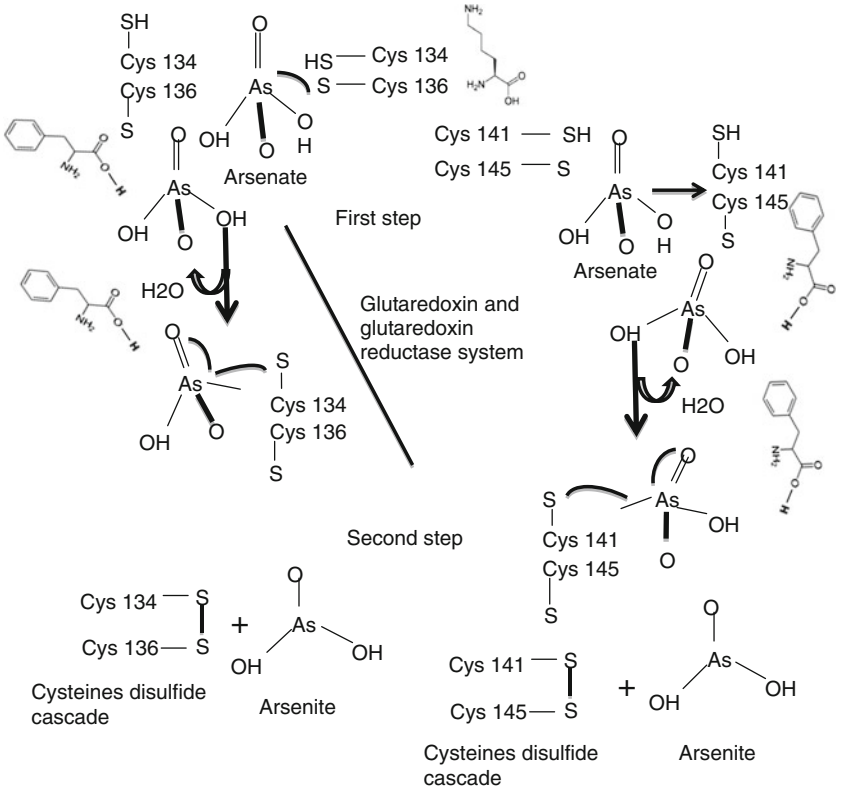

Fig. 4 The postulated pathway of arsenate reduction in *Arabidopsis thaliana*. In the first step, Cys-136 and Cys-145 are activated in the presence of arsenate. This activation leads to initiation of the reduction mechanism of arsenate in the plant cell. The nucleophilic attachment of thiol (SH) group of Cys-136 and Cys-145 initiates protonation of a hydroxyl (OH) group of each arsenate molecule. A stable cys136-S-$AsHO^{-}_{3}$ or cys145-S-$AsHO^{-}_{3}$ enzyme complex is formed by releasing water (H_2O) molecules. Phe53 and Phe54 act as a general acid in an in line reaction to stabilize the enzyme substrates. In second step, these enzyme substrates are attacked by the contiguous Cys-134 and Cys-141. The thiolate groups of these two cysteines are stabilized by the positively charged Lys-135. Then arsenate ions get reduced to arsenite and two disulfide bonds are formed combining Cys-136 with Cys-134 and Cys-145 with Cys-141 [15]. The disulfide bonds are then protonated by the glutaredoxin and glutaredoxin reductase system to cysteines for recycling of the pathway

Determination of homozygosity by PCR

To further verify if the mutant was homozygous with respect to T-DNA insertion we performed PCR as described in "Materials and methods". PCR of genomic DNA, isolated and pooled from several kanamycin-resistant plants of this mutant using primers from the promoter and coding regions of the *ACR2* gene resulted in amplification of a fragment of the expected size (4.9 kb), whereas a fragment of only 302 bp was obtained in Wt control plants (data not shown). These results indicate that plants of the mutant line SALK_005882C are homozygous with respect to T-DNA insertion.

RT-PCR analysis of expression of the ACR2 gene

To verify whether the T-DNA tag indeed resulted in inactivation or partially reduced expression of the *ACR2* gene, we performed RT-PCR (Fig. 5). The intensity of the amplified band obtained from the SALK mutant was approximately three-fold lower than that obtained in Wt or VC control plants (data not shown). These results indicate that the expression of *ACR2* mRNA in the SALK mutant was severely reduced compared to that in control plants (Wt and VC). An identical pattern of gene expression was observed in all three samples (whole plants, shoots and roots), suggesting that the *ACR2* gene in the SALK mutant is partially inactivated by T-DNA insertion.

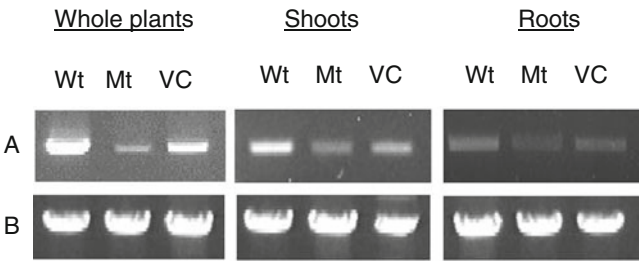

Fig. 5 Analysis of the expression level of the arsenic reductase (*ACR2*) gene in wild-type (Wt), SALK mutant (Mt) and vector transformed control (VC) plants by RT-PCR. Plants were grown for 3 weeks and total RNA was isolated from whole plants, shoots and roots. **A** Level of expression of *ACR2*, **B** level of expression of actin gene in *A. thaliana* (control). The sizes of the amplified *ACR2* and actin gene fragments were 439 bp and 521 bp, respectively

Exposure of plants to arsenics

To verify the role of the *ACR2* gene in tolerance/sensitivity of *A. thaliana* to arsenics, we exposed the SALK mutant to different concentrations of arsenics. When exposed to 100 μM arsenate, the SALK mutant exhibited a higher sensitivity as compared to control plants (Fig. 6). After 3 weeks of exposure, the leaves of these plants started to turn yellow (early senescence), whereas the control plants still remained green. In addition, after exposure to arsenate, the SALK mutant showed a decrease in fresh weight of about 1.5-fold compared to control plants (Fig. 7). However, when plants were treated with arsenite (25 μM or 100 μM), we observed no phenotypic difference between control and mutant plants. Growth of these plants was affected equally (Fig. 6). Figure 6 also indicates that, for *A. thaliana*, arsenite is more toxic than arsenate. Treatment of plants with 25 μM arsenite resulted in a higher level of stress (retarded growth and early decaying/senescence) in comparison with plants treated with 25 μM arsenate (Fig. 6).

Arsenic accumulation in plants

To verify if the target gene *ACR2* is indeed involved in arsenic accumulation in *A. thaliana* we have exposed the SALK mutant to various concentrations of arsenate and arsenite for 3 weeks. When exposed to 100 μM arsenate, the amount of arsenate accumulated in the shoots of the SALK mutant was approximately six-fold higher (1,240 μg/g dry wt) than that (234 μg/g dry wt) accumulated in the shoots of control plants (Fig. 8). On the other hand, when the SALK mutant was exposed to 100 μM arsenite, the amount of arsenite accumulated in the shoots was almost equal (106 μg/g dry wt) to that (131 μg/g dry wt) accumulated in the shoots of control plants. Similar results were obtained when the plants were exposed to 25 μM arsenite (Fig. 8).

Discussion

This paper describes a part of our attempt to generate new varieties of cultivated crops with eliminated or reduced arsenic accumulation in the edible parts of the plant. Previously, by employing *in silico* analyses, we identified four candidate genes (*PCS1*, *ACR2*, *MRP1* and *MRP2*) putatively involved in arsenic accumulation in plants [1]. Our simulation experiments, based on *Arabidopsis thaliana* as the model plant, have revealed that it is possible to significantly reduce arsenic accumulation in this plant (from 55 μg/g dry wt to 18 μg/g d rywt) by regulating the expression of the candidate genes [1]. One of the key genes is *ACR2*, the putative function in plants of which has been studied and reported by many researchers [4, 50]. This gene is involved in reduction of arsenate (AsV) to arsenite (As III), which can be sequestered in the vacuoles of the plant cell [9, 51]. However, to date, only a very few studies have reported the molecular mechanisms or biochemical pathway of this process in plants. Here, we report the results of our investigations, both *in silico* and *in vivo*, on the molecular mechanisms of the function of *ACR2* in the model plant *A. thaliana*.

For *in silico* studies, the ACR2 protein sequence of *A. thaliana* containing 146 amino acids retrieved from the NCBI database (accession number is AED 90607.1, http://www.ncbi.nlm.nih.gov/protein?term=ACR2) was analyzed mainly by using the I-TASSER server [17]. I-TASSER is a hierarchical protein structure modeling (Fig. 1) approach based on secondary structure-enhanced Profile-Profile threading Alignment (PPA). Using this method, we first predicted the secondary structure of the target protein and then constructed a 3D model indicating the substrate binding residues. Based on this 3D structure, the function of the target protein was predicted. Finally, the predicted function was validated by *in vivo* laboratory experiments.

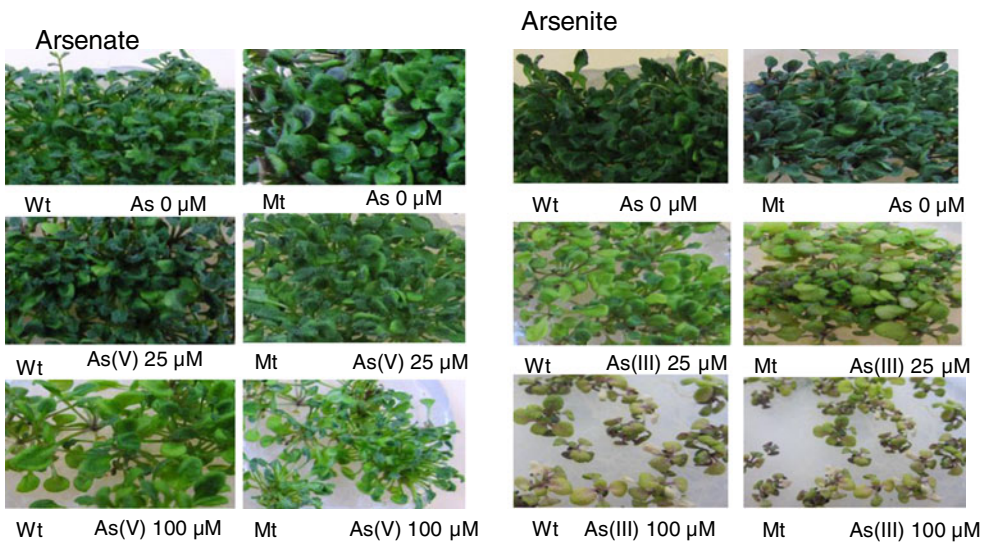

Fig. 6 Effect of arsenate (AsV) and arsenite (AsIII) on growth of *A. thaliana*. *Wt* Wild-type control, *Mt* SALK mutant plants. Seven-day old plants were exposed to arsenate (25 μM or 100 μM) and arsenite (25 μM or 100 μM). Photographs were taken after 3 weeks of exposure

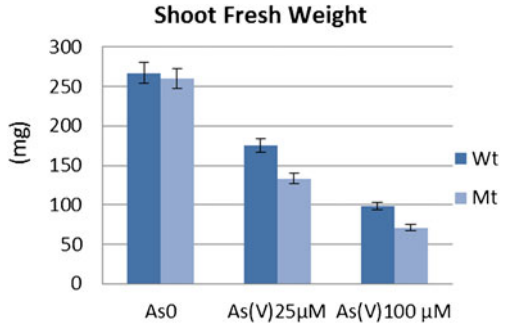

Fig. 7 Effect of arsenics on biomass accumulation in Wt and Mt *A. thaliana* plants. Fifteen plants were grown as described in Fig. 6. The experiments were performed in eight replicates. Plants were exposed to arsenate for 3 weeks and the fresh weight of shoots was measured

The bioinformatic analyses were based on several repeated benchmark tests [46] and, therefore, the results obtain in these experiments are reliable. Importantly, these methods are significantly faster, less laborious and inexpensive compared to laboratory experiments. A 3D model of a protein can also be constructed by using nuclear membrane resonance (NMR) and/or protein crystallography [52] but these are much more laborious, time consuming and expensive approaches. *In silico* prediction of plant gene function based on modeling of 3D structure of the protein and the substrate binding capacity of residues has been reported previously by us [48, 53] and many other researchers [54–56]. Previously, our research group had shown that the function of an unknown T-DNA tagged plant gene predicted by *in silico* analysis was indeed in complete agreement with that verified later by laboratory experiments [48, 53]. However, there are several drawbacks to *in silico* experiments. It often becomes very difficult to select the most suitable candidate when several parameters are considered. In these recent experiments, we selected the threading fold 1t3kA (Z-score 3.65) as a native structure for further analysis (Table 1) although the folds 1qb0A and 1c25 exhibited much

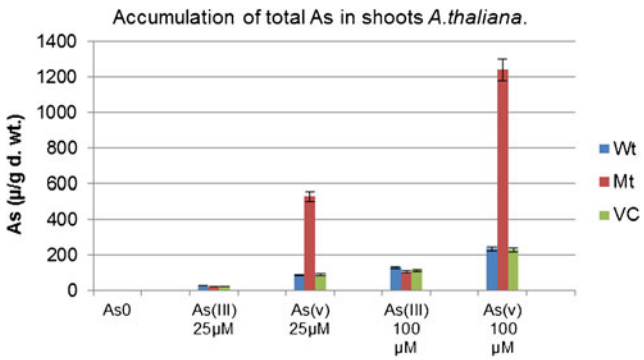

Fig. 8 Analysis of arsenic accumulation in shoots of *A. thaliana*. Seven day old seedlings were exposed to arsenate (25 μM or 100 μM) and arsenite (25 μM and 100 μM) for 3 weeks. Shoots were collected and dried, and the arsenic content was determined by ICP.DRC-MS (μg/g dry wt). *Blue bars* Wild-type control plants (Wt), *red bars* SALK mutant plants (Mt), *green bars* vector transformed control (VC) plants

higher Z-scores (4.34 and 3.96, respectively, Table 1). In this case, the decisive factors for ranking the top fold were the sequence identities (Ident1 and Ident2) and the putative function of the threading fold proposed by computer analysis. However, when all required parameters show the highest significant values, there is no problem with selection of the target structure or model. In our experiments, model 1 was top-ranked by I-TASSER and was based on the highest confidence score (C-score=−0.19; Fig. 3a). This model also indicated the highest number of cluster density (0.2464), TM-score (0.69±0.12) and RMSD (5.1±3.3). Therefore, we selected this model for further analyses.

Usually, the Fh–score is an indicator of functional similarity between a predicted structure and detected analogs. However, because protein function is multi-faceted, the unanimity of functionalities of the identified analogs usually yields many associated GO terms (data not shown). In our experiment, to select the most reliable associated GO terms, we performed another consensus prediction of GO terms based on GO-scores (Table 4). The highest GO-scores were then used for prediction of molecular, cellular and biological function of the target protein.

As mentioned in the Materials and methods, we performed extensive laboratory experiments to validate the results obtained with *in silico* analyses. To verify the postulated function of the *ACR2* gene, we studied a SALK mutant of *A. thaliana* [47]. As indicated in the database, the mutant harbors a T-DNA insertion tag in the upstream promoter region of the *ACR2* gene (*At5g03455*). Because of this insertion, the *ACR2* gene is knocked-down (reduced activity) but not knocked-out (inactive). For this reason, we observed partial mRNA expression of the *ACR2* gene in the SALK mutant (Fig. 5a,b). For our studies, the best option would be a mutant with a complete knocked-out *ACR2* gene exhibiting no transcriptional or translational activity. In the Arabidopsis Information Resource (http://arabidopsis.org/servlets/TairObject?type=locus&name=AT5G03455), seven SALK mutants were available. These mutants harbor a T-DNA insertion tag in the *ACR2* gene. In our experiments, we investigated all seven of these individual T-DNA tagged SALK mutants; unfortunately, none of them were knock-out mutants with an inactive *ACR2* gene (data not shown). Among the seven mutants, we selected one, SALK_ 005882C, for laboratory experiments to validate the *in silico* results. There are two reasons for this selection. In this mutant, expression of the *ACR2* transcript was lowest, and the level of arsenic accumulation highest in comparison with those observed in the remaining six SALK mutants tested (data not shown). When exposed to 100 μM arsenate, the accumulation of arsenics in the SALK_005882C was approximately six-fold higher than that observed in Wt or VC plants (Fig. 8). Several researchers have shown similar results while working with *ACR2* mutants. Dhankar and co-workers [4] exposed an

RNAi-derived *ACR2* mutant of *A. thaliana* to arsenics and reported that the mutant accumulated 10- to 16-fold higher amounts of arsenics compared to Wt-controls. The difference observed in arsenic accumulation between our SALK_005882C and the RNAi-induced mutant can possibly be explained by postulating that the level of expression of the *ACR2* gene in the RNAi mutant was lower than that in the SALK mutant. Physiological and growth conditions of the plants might play a significant role in differential accumulation of arsenics in these mutants. Further, it is not clear in the report of Dhankar and co-workers [4] whether the RNAi mutant had a Col-0 background as is the case for our SALK mutant. Thus, the differences in accumulation of arsenics observed in these two mutants may simply be self-explanatory.

In the *in silico* experiments, the final predicted model of our ACR2 protein consists of seven substrate binding site residues (Table 5, Fig. 3b), and all of these residues are highly specific for binding arsenate but not arsenite. We also verified this hypothesis by exposing the SALK mutants and control plants to arsenate and arsenite, separately. When exposed to 100 μM arsenite, we observe no difference in sensitivity between SALK mutants and control plants—they were equally sensitive to arsenite (Fig. 8). However, when exposed to 100 μM arsenate, the SALK mutant exhibited much higher sensitivity than control plants (Fig. 8). These results suggest that the *ACR2* gene is more responsive to arsenate than arsenite, and confirm the *in silico* predicted function of the ACR2 protein structure, which consists of substrate binding site residues specific for binding arsenate and not arsenite (Fig. 4, Table 5). To date, no other results on this topic have been reported in plant research and therefore we do not have the possibility to compare our results. However, similar studies on sensitivity of *ACR2* to arsenate have been performed in bacteria and yeast [10–12, 14]. In these latter studies, researchers have constructed models by either NMR or protein crystallography [51]. Our results, sensitivity of *ACR2* to arsenate, obtained in both *in silico* and *in vivo*, are almost identical to those observed in *E. coli* and yeast [10, 14]

By using *in silico* analyses, we have predicted the function of the *A. thaliana ACR2* gene. This gene is involved in arsenic accumulation in plants. Based on the *in silico* results, we propose a pathway through which arsenate is reduced to arsenite in plants. We also show that the predicted model of the ACR2 protein consists of seven substrate binding site residues with high affinity to bind arsenate. By performing extensive *in vivo* laboratory experiments we could show that the results obtained *in silico* are in complete agreement with those obtained *in vivo*. In addition, our study shows that, by using bioinformatic tools, it is possible to speed up molecular biology research, and verification of the function of target genes by laboratory experiments strengthens the reliability of the *in silico* analysis.

Acknowledgments This research was supported financially partly by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS, grant no. 229-2007-217) and partly by the Swedish International Development Cooperation Agency (SIDA, grant no. AKT-2010-018). We also acknowledge a grant from the Nilsson-Ehle Foundation (The Royal Physiographic Society in Lund). We would like to thank the European Arabidopsis Stock Center, Nottingham (NASC) for providing us with their SALK mutants and The I-TASSER Server Team for providing us with their programs and software for computational analyses.

References

1. Lund D, Larsson D, Nahar N, Mandal A (2010) J Biol Sys 18:223–224
2. Chen Y, Parvez F, Gamble M, Islam T, Ahmed A, Argos M, Graziano JH, Ahsan H (2009) Toxicol Appl Pharmacol 239:184–192
3. Zhao FJ, McGrath SP, Meharg A (2010) Annu Rev Plant Biol 61:1–25
4. Dhankher OP, Rosen BP, McKinney EC, Meagher RB (2006) Proc Natl Acad Sci USA 103:5413–5418
5. Zhu YG, Rosen BP (2009) Curr Opin Biotechnol 20:220–224
6. Srivastava S, Mishra S, Tripathi RD, Dwivedi S, Trivedi PK, Tandon PK (2007) Environ Sci Technol 41:2930–2936
7. Zhao FJ, Ma JF, Meharg AA, McGrath SP (2009) New Phytol 181:777–794
8. Lee DA, Chen A, Schroeder JI (2003) Plant J 35:637–646
9. Kamiya T, Tanaka M, Mitani N, Ma JF, Maeshima M, Fujiwara T (2009) J Biol Chem 284:2114–2120
10. DeMel S, Shi J, Martin P, Rosen BP, Edwards BF (2004) Protein Sci 13:2330–2340
11. Bennett SM, Guan Z, Laurberg M, Su DX (2001) Proc Natl Acad Sci USA 98:13577–13582
12. Zegers I, Martins JC, Willem R, Wyns L, Messens J (2001) Nat Struct Biol 8:843–847
13. Bobrowicz P, Wysocki R, Owsiananik G, Goffeau A, Ulaszewski S (1997) Yeast 13:819–828
14. Mukhopadhyay R, Shi J, Rosen BP (2000) J Biol Chem 275:21149–22115
15. Zhou Y, Messier N, Ouellette M, Rosen BP, Mukhopadhyay R (2004) J Biol Chem 279:37445–37451
16. Bazzoli A, Andrea G, Tettamanzi B, Zhang Y (2011) J Mol Biol 407:764–776
17. Roy A, Kucukural A, Zhang Y (2010) Nat Prot 5:725–738
18. Cozzetto D, Kryshtafovych A, Fidelis K, Moult J, Rost B, Tramontano A (2009) Proteins 77:18–28
19. Krivov GG, Shapovalov MV, Dunbrack RL (2009) Proteins 77:778–795
20. Floudas CA, Fung HK, McAllister SR, Monnigmann M, Rajgaria R (2006) Rev Chem Eng Sci 61:966–988
21. Fung HK, Welsh WJ, Floudas CA (2007) Ind Eng Chem Res 47:993–1001
22. Wu S, Skolnick J, Zhang Y (2007) BMC Biol 5:17
23. Bonneau R, Baker D (2001) Annu Rev Biophys Biomol Struct 30:173
24. McGuffin LJ, Bryson K, Jones DT (2000) Bioinformatics 16:404–405
25. Jones DT (1999) J Mol Biol 292:195–202
26. Margelevičium M, Venclovas C (2005) BMC Bioinf 6:185

27. Wu S, Zhang Y (2007) Nucleic Acids Res 35:3375–3382
28. Margelevicîium M, Laganeckas M, Venclovas C (2010) Bioinformatics 26:1905–1906
29. Shi J, Blundell TL, Mizuguchi K (2001) J Mol Biol 310:243–257
30. Sording J (2005) Bioinformatics 21:951–960
31. Wu S, Zhang Y (2008) Proteins Struct Funct Bioinf 72:547–556
32. Xu Y, Xu D (2000) Proteins 40:343–354
33. Karplus K, Karchin R, Draper J, Casper J, Mandel-Gutfreund Y, Diekhans M, Hughey R (2003) Proteins 53:491–496
34. Zhou H, Zhou Y (2004) Proteins 55:1005–1013
35. Zhou H, Zhou Y (2005) Proteins 58:321–328
36. Lee D, Redfern O, Orengo C (2007) Nat Rev Mol Cell Biol 8:995–1005
37. Wu S, Zhang Y (2008) Bioinformatics 24:924–931
38. Zhang Y, Skolnick J (2004) J Comput Chem 25:865–871
39. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
40. Holm L, Sander C (1991) J Biol 218:183–194
41. McDonald IK, Thornton JM (1994) J Mol Biol 238:777–793
42. Bennett MJ, Schlunegger MP, Eisenberg D (1995) Protein Sci 4:2455–2468
43. Zhang Y, Skolnick J (2005) Nucleic Acids Res 33:2302–2309
44. Li YQ, Zhang Y (2009) Proteins 76:665–676
45. Buck M, Bonnet BS, Pastor WR, MacKerell DA (2006) Biophys J 90:36–38
46. Lee SY, Skolnick J (2008) Biophys J 95:1956–1964
47. Alonso MJ, Stepanova NA, Leisse JT, Kim JC, Chen H, Shinn P, Stevenson KD, Zimmerman J, Barajas P, Cheuk R et al (2003) Science 301:653–657
48. Svensson M, Lund D, Ejdebäck M, Mandal A (2004) J Mol Model 10:130–138
49. Elke M, Lorentzen L, Kingston HM (1996) Anal Chem 68:4316–4320
50. Duan GL, Zhou Y, Tong YP, Mukhopadhyay R, Rosen BP, Zhu YG (2007) New Phytol 174:311–321
51. Blum R, Beck A, Korte A, Stengel A, Letzel T, Lendzian K, Grill E (2007) Plant J 49:740–749
52. Pittelkow M, Tschapek B, Smits SH, Schmitt L, Bremer E (2011) J Mol Biol 411:53–67
53. Sazzad K, Lund D, Holmström KO, Mandal A, Pirhonen M (2005) J Mol Model 11:226–236
54. Kumar R, Kumar S, Sangwan S, Yadav IS, Yadav R (2011) J Mol Graph Model 29:740–746
55. Biarnés X, Bongarzone S, Vargiu AV, Carloni P, Ruggerone P (2011) J Comput Aided Mol 25:395–402
56. Haqul Z, Khan W (2011) J Comput Aided Mol 25:81–101