ORIGINAL PAPER

In silico identification and characterization of 1-aminocyclopropane-1-carboxylate deaminase from *Phytophthora sojae*

Nidhi Singh · Sudhanshu Kashyap

Abstract As part of an effort to obtain a fungal 1-aminocyclopropane-1-carboxylate deaminase encoding gene from *Phytophthora sojae* expressed sequence Tag database, we identify and characterize the ACCD from *P. sojae* using bioinformatics data mining tools and techniques. Computed structural model of *P. sojae* ACCD was found to consist of mixed α/β motifs and probable loops. The predicted model resembles the structure of *Pseudomonas* ACCD (RMSD-0.44 Å). The main differences observed between them are the presence of partial length of domain one, and longer helix $\alpha4$. Ramachandran plot analysis revealed that portion of all residues falling into the most favorable regions was 95.0%. The substrate – and geometrical- docking of developed structure postulated functional capability of ACCD to carry out ACC cleavage reaction. The catalytic site in homo-tetrameric structure open to opposite directions separated by ~37.97 Å distance arranged around central axis. This study provides a comprehensive identification and characterization of the ACCD in *P. sojae* and it may be helpful in the transcriptional and expression based study of *P. sojae* pathogenesis.

Keywords Active site · Docking · ESTs · Homo-tetramer · Modeller · *Phytopthora sojae*

N. Singh · S. Kashyap (✉)
National Bureau of Agriculturally Important Microorganisms (ICAR),
Kusmaur, Kaithauli,
Mau Nath Bhanjan 275101 Uttar Pradesh, India
e-mail: sudhanshukshyp@gmail.com

Introduction

The role of compound 1-aminocyclopropane-1-carboxylate (ACC), a precursor of gaseous hormone ethylene in the regulation of plant physiology and development has been well established. Its role has also been implicated in biotic stress, both as a virulence factor of fungal and bacterial pathogens and as a signaling compound in disease resistance. Enhanced ethylene production is an early active response of plants in perception of pathogen attack and is associated with the induction of defense reactions [1]. An increased concentration of endogenous ethylene in plants or an enhanced sensitivity of certain plant species, mainly dicots, to ethylene can result in inhibition of seed germination, root growth, enhancement in flower and leaves senescence along with early fruits ripening [2, 3]. One of the mechanisms that a number of plant growth promoting soil and entophytic microbes uses to facilitate plant growth and development is by producing the enzyme 1-aminocyclopropane-1-carboxylate deaminase (ACCD), which sequester and hydolytically cleaves plant produced ACC and is quite effective in lowering the endogenous concentration of ACC (Fig. 1) and hence ethylene in plant tissues [3–6]. The ACCD has been studied in various plant growth promoting bacteria (PGPB) like *Enterobacter cloacae*, *Rhizobium* sp., *Variovorax* sp., *Alcaligenes* sp., *Bacillus* sp,. etc., and its structure is elucidated in *Pseudomonas* sp., [7–10] *Entrobacter cloaceae* [11], *Hansenula saturnus* [12] and *Penicillum citrinum* [13]. In spite of these, less or no information is available from phytopathogenic fungal group. In the present study we employed expressed sequence tag (EST) searching; multiple sequence comparison and other data mining techniques alongwith docking studies to identified and characterized coding sequence and tertiary and quaternary structure of putative ACCD of *Phytophthora sojae* (PsACCD).
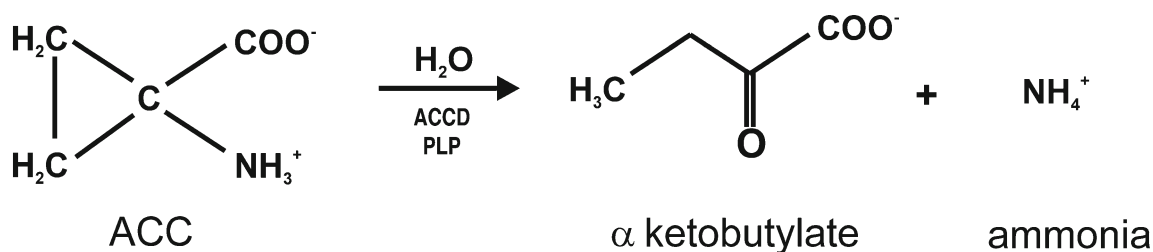
Fig. 1 Reaction scheme of ACCD

Material and methods

Data extraction of c-DNA sequence and protein prediction

The mRNA sequence (**XM_744146**) of *Aspergillus fumigatus* was used to find out similar EST of *P. sojae* (**CF842090.1**) by using Blastn against dbEST (containing about 28,357 EST of P. *sojae*). CpG islands revealing program (http://l25.itba.mi. cnr.it/cgi-bin/wwwcpg.pl) and CpG island searcher (http:// cpgislands.usc.edu/) were used to screen CpG islands using parameters (lower limit of %GC, ObsCpG/ExpCpG, length). The amino acid was deduced by using ExPASy translate tool (http://www.expasy.ch/tools/dna.html) and ORF Finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The homologous protein in other species is identified by BlastP (http://www.ncbi.nlm.nih.gov/BLAST/). Similarity analysis was performed by using blastx (http://www.ncbi.nlm.nih.gov/BLAST/) and SIB BLAST (http://expasy.org/tools/blast/). SOSUIsignal (http://bp.nuap.nagoya-u.ac.jp/sosui/sosuisignal/sosuisignal_submit.html) was used to predict transmembrane region in the deduced amino acid sequence. Signal peptide cleavage sites, of this amino acid sequence was predicted using SignalIP 3.0 server (http://www.cbs.dtu.dk/services/SignalP/). ProtPARAM (http://expasy.org/tools/protparam.html) and compute pI/Mw (http://expasy.org/tools/pi_tool.html), available at ExPASy were used to obtain physico-chemical parameters of a protein sequence by primary structure analysis. For secondary structure prediction GOR4 (http://www.compbio. dundee.ac.uk/~www-jpred/), program available at ExPASy was used. The two other programs used were PLOC (http:// www.genome.jp/SIT/plocdir/) to predict subcellular location of predicted protein, and ProSLP (http://proses.kisti.re.kr/) to know superfamily of coded protein sequence.

Phylogenetic analysis and structure modeling and rigid ligand docking of ACC deaminase

The phylogenetic tree of encoded amino acid sequence was constructed from the species with MEGA3.1 [14] by using the neighbor-joining method. The reliability of internal branches was assessed by using 1000 bootstrap replicated, and gaps were deleted in the analysis. The three dimensional structure was generated by aligning the deduced amino acid sequence

from *P. sojae* EST (**CF842090.1**) in HHpread Interactive server (http://protevo.eb.tuebingen.mpg.de/hhpred) under global alignment mode with PDB entries. Obtained result was manually narrowed down using the best ten structures alignment, HHpread was rerun on local alignment and zero settings. The so obtained end alignment was directly used in the MODELLER software [15] (version 9v8). The retrieved PDB was visualized using UCF Chimera (http://www.cgl. ucsf.edu/chimera). The best preliminary model was validated with PROCHECK [16] and WHATIF [17] by submitting the coordinates to the EMBL- (http://www.ebi.ac.uk) and ProSA-servers (https://prosa.services.came.sbg.ac.at.). Backbone confirmation was evaluated by the inspection of the Psi/Phi Ramachandran plot using the RAMPAGE web server (http:// mordred.bioc.cam.ac.uk/). Potential deviations were calculated with SUPERPOSE web server (http://wishart.biology.ual berta.ca/cgi-bin/) for root mean square deviation (RMSD) between *Pseudomonas* ACCD (PdACCD) structures (PDB ID: **1tyz**) [18] and developed model. The secondary structure visualization was made using PDBsum (http://www.ebi.ac.uk/ thronton-serv/PDBsum) and functional parameter was predicted using ProFunc (http://www.ebi.ac.uk/thronton-serv/ ProFunc). The coenzyme-substrate docking was performed with the PatchDOCK server [19] which works on local shape features matching and it reduces the complexity of the entire docking process by detecting only those molecular surface areas which have a high probability of being the binding site. The final docking pose results were obtained by submitting the ligand and receptor PDB files to the web interface. The clustering RMSD was kept at 4 Å while all other features were kept at default values. The ligand (PLP-ACC) was retrieved from and *Hansenula saturnus* ACCD (HsACCD) (PDB: 1j0d) molecule and developed PsACCD coordinates were used directly as a receptor. The selected single docking pose was submitted to SymmDOCK [19] for geometric base docking and generation of a quaternary structure. The SymmDOCK is a geometry – based algorithm for prediction of a cyclically symmetric complex, given the structure of its asymmetric unit. The final models were submitted to the Protein Model Data Base (PMDB) (www.caspur.it/ PMDB/) to obtain the PMDB identifiers **PM0077871** (monomer); **PM0077872** (homo-dimer); **PM0077873** (homo-tetramer).

Results

Nucleotide sequence analysis

Blast against dbEST showed three EST sequences of *P. sojae* of which **CF842090.1** was used for further analysis (Fig. 2). Similarity analysis with BlastP has suggested that the nucleotide sequence has high similarity with ACCD of bacterial species (Fig. 3). CpG islands reveling program displayed that nucleotide sequence contains %C + G=66; %CpG=11, Obs./Exp.=1.08. *P. sojae* ACCD mRNA was found to consist of 36 bp of 5' UTR, 606 bp of coding region, and 9 bp of a 3' UTR region. The PsACCD mRNA was predicted to encode a peptide of 201 amino acids.

Protein analysis

The BlastP analysis showed that PsACCD has various degrees of identity with corresponding proteins of *Phytopthora infestance*-T30-490(*D0N448*) 95.8%; putative *Zea mays* (*B6SNR6*) 85.84%; *Acidovorax avenae* aubsp *avenae* ATCC19860 (*B9MGI8*) 82.8%; *Variovorex parafdoxus*-S110 (*Q6J256*) 78.6%; *Pseudomonas putida* (*A0MNM8*) 79.7%; *Methlibium petroleiphilum*-pm1 (*A2SLW2*) 78.1%; *Pantoea* sp.At9b (*C8Q2T3*) 77.6%; *Burkholderia ambitaria*-cmc40-6 (*B1YYR8*) 76.6%; *Burkhoideria cepacia* (*B8R7S1*) 75.5% respectively (Fig. 3). Computed pI/Mw of deduced amino acid sequence has molecular weight of 21798.7 and theoretical pI of

6.90 which was also verified by ProtParam tool. Instability index classifies the protein as unstable. ProtParam displayed that predicted protein has 23 negatively charged residues (Asp + Glu) and 23 positively charged residues (Arg + Lys). Secondary structure prediction GOR4 predicted that deduced amino acid sequence has α helix, extended strand, and Random coil. SOSUIsignal result displayed that amino acid sequence is soluble protein and this sequence has no signal peptide. By using PLOC we found the subcellular location of *P. sojae* ACCD extracellular. ProSeS (http://proses.kisti.re.kr/help. php#ProSLP) predicted that the superfamily name of deduced amino acid sequence is ACCD (PIRSF006278). NCBI CDD displayed that deduced amino acid sequence shows high similarity with PRK12390 domain.
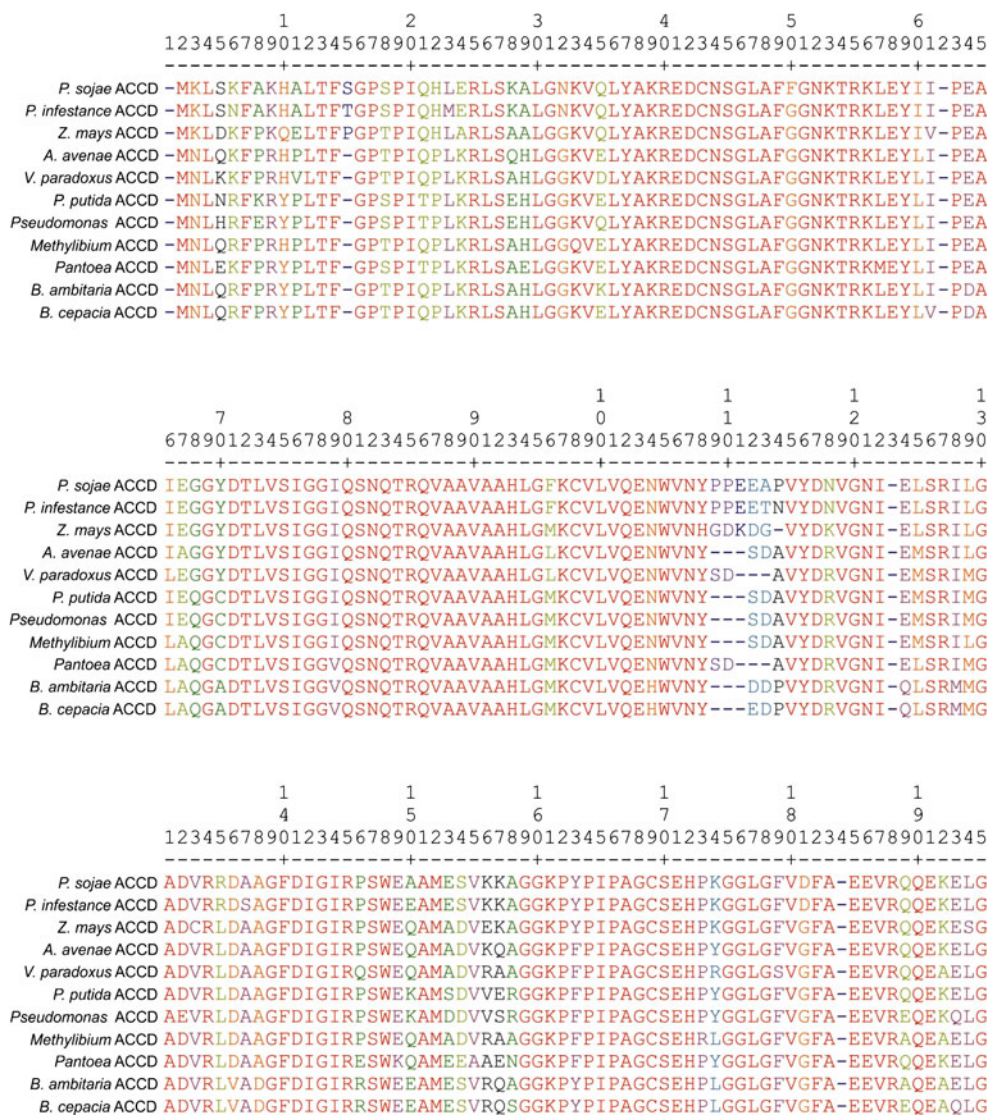
Phylogenetic analysis and structure modeling

The encoded amino acid sequence of *P. sojae* showed homology with bacterial species predicting that PsACCD is orthologous to reported bacterial ACCD (Fig. 4). The PdACCD *and* the PsACCD have a high degree of sequence similarity and the homologous tertiary structure consists of 198 amino acids. Reverse template analysis also confirmed it's similarity with ACCD structure's. According to the assigning of secondary structure, it consists of seven ά helices, six β strands. The final model is composed of seven helix coils and six beta sheets (Fig. 5), the mark differences observed between PdACCD and

Fig. 2 Nucleotide sequence alignment between mRNA sequence of *Aspergillus fumigates* (upper sequence line) and EST sequence from *P. sojae* (lower sequence line)

```
Aspergillus fumigates  37   ATGAAGCTCTCCAAGTTCGCCAAGCACGCGCTGACGTTCTCCGGTCCGTCGCCAATCCAG   96
                            ||||||||||||   || ||||||||||  || |||||| ||||||||||||  |||||
Phytopthora sojae       1   ATGAAGCTCTCCAATTTTGCCAAGCACGCTCTCACGTTCACTGGTCCGTCGCCGATCCAG   60

Aspergillus fumigates  97   CACCTGGAGCGTCTGTCCAAGGCGCTGGGCAACAAGGTGCAGCTCTACGCCAAGCGCGAG   156
                            ||| |||| || |||| ||||| ||| ||||||||||||||||||| ||| ||||||||||
Phytopthora sojae      61   CACATGGAACGGCTGTCGAAGGCTCTGGGCAACAAGGTGCAGCTGTATGCCAAGCGCGAG   120

Aspergillus fumigates 157   GACTGCAACTCGGGCCTGGCGTTT-TTCGGCAACAAGACGCGCAAGCTCGAGTACATCAT   215
                            |||||||||| ||| ||||| ||| |||||||||||||||||||||||||| |||||||||
Phytopthora sojae     121   GACTGCAACTCCGGTTTGGCCTTTGGT-GGCAACAAGACGCGCAAGCTCGAATACATCAT   179

Aspergillus fumigates 216   CCCGGAGGCCATCG-AGGGCGGCTACGACACGCTCGTGTCCATCGGCGGCATCCAGTCCA   274
                            ||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||
Phytopthora sojae     180   CCCGGAGGCCATCGAAGGG-GGCTACGACACGCTCGTGTCCATCGGCGGCATCCAGTCCA   238

Aspergillus fumigates 275   ACCAGACGCGCCAGGTAGCCGCCGTGGCCGCGCACCTCGGCTTCAAGTGCGTGCTGGTGC   334
                            |||||||||| ||||| |||| ||||||||||||| |||||||||||||||||| |||
Phytopthora sojae     239   ACCAGACGCGTCAGGTGGCCGCCGTAGCTGCTCACCTCGGGTTCAAGTGCGTGTTGGTCC   298

Aspergillus fumigates 335   AGGAGAACTGGGTCAACTACCCGCCCGAGGAGGCCC-CGGTGTACGACAACGTGGGCAAC   393
                            ||||||||||||| |||||||||||||  |||| || | ||||| || ||| |||||||
Phytopthora sojae     299   AGGAGAACTGGGTAAACTACCCGCCGGAAGAGACCAACG-TGTATGATAACGTAGGCAAC   357

Aspergillus fumigates 394   ATCGAGCTGTCCCGCATCCTGGGCGCCGACGTGCGTC-GCGACGCCGCGGGCTTCGACAT   452
                            || ||||||| || ||||| ||||||||||||||||| |  | ||| ||||| |||||
Phytopthora sojae     358   ATTGAGCTGTCGCGTATCCTCGGCGCCGACGTGCG-CAGGGACTCGGCCGGTTTCGATAT   416

Aspergillus fumigates 453   CGGCATCCGCCCGAGCTGGGAGGCCGCCATGGAGAGCGTCAAGAAGGCCGGCGGCAAGCC   512
                            ||||||||| |||||||||||||||| |||||||||||||||||||||||  ||||||||
Phytopthora sojae     417   CGGCATCCGACCAAGCTGGGAGGAGGCCATGGAGAGCGTCAAGAAGGCCGGTGGCAAGCC   476

Aspergillus fumigates 513   GTACCCCATTCCGGCGGGCTGCTCGGAGCACCCCAAGGGCGGCCTGGGCTTCGTGGACTT   572
                            ||| ||||||||||| |||| ||||||||||||||||| || ||||||||||||| ||||
Phytopthora sojae     477   GTATCCCATTCCGGCTGGCTGCTCGGAGCACCCCAAGGGTGGATTGGGCTTCGTCGACTT   536

Aspergillus fumigates 573   CGCCGAGGAGGTCCGTCAGCAGGAGAAGGAGCTCGGCTTCA-GTTCGACTACATCGTCGT   631
                            || ||||||||||||||||||||||||||||||| |||  || ||||||||||| ||||||
Phytopthora sojae     537   TGCTGAGGAGGTCCGTCAGCAGGAGAAGGAGCTGGGCTTTAAGTTCGACTACGTCGTCGT   596

Aspergillus fumigates 632   GTGCGCCGTGACGGGCAG   649
                            ||||||||||||| |||||
Phytopthora sojae     597   GTGCGCCGTGACCGGCAG   614
```

```
                      1         2         3         4         5         6
             1234567890123456789012345678901234567890123456789012345
             ---------+---------+---------+---------+---------+---------+-----
  P. sojae ACCD  -MKLSKFAKHALTFSGPSPIQHLERLSKALGNKVQLYAKREDCNSGLAFFGNKTRKLEYII-PEA
P. infestance ACCD  -MKLSNFAKHALTFTGPSPIQHMERLSKALGNKVQLYAKREDCNSGLAFGGNKTRKLEYII-PEA
  Z. mays ACCD  -MKLDKFPKQELTFPGPTPIQHLARLSAALGGKVQLYAKREDCNSGLAFGGNKTRKLEYIV-PEA
  A. avenae ACCD  -MNLQKFPRHPLTF-GPTPIQPLKRLSQHLGGKVELYAKREDCNSGLAFGGNKTRKLEYLI-PEA
V. paradoxus ACCD  -MNLKKFPRHVLTF-GPTPIQPLKRLSAHLGGKVDLYAKREDCNSGLAFGGNKTRKLEYLI-PEA
  P. putida ACCD  -MNLNRFKRYPLTF-GPSPITPLKRLSEHLGGKVELYAKREDCNSGLAFGGNKTRKLEYLI-PEA
Pseudomonas ACCD  -MNLHRFERYPLTF-GPSPITPLKRLSEHLGGKVQLYAKREDCNSGLAFGGNKTRKLEYLI-PEA
Methylibium ACCD  -MNLQRFPRHPLTF-GPTPIQPLKRLSAHLGGQVELYAKREDCNSGLAFGGNKTRKLEYLI-PEA
   Pantoea ACCD  -MNLEKFPRYPLTF-GPSPITPLKRLSAELGGKVELYAKREDCNSGLAFGGNKTRKMEYLI-PEA
B. ambitaria ACCD  -MNLQRFPRYPLTF-GPTPIQPLKRLSAHLGGKVKLYAKREDCNSGLAFGGNKTRKLEYLI-PDA
  B. cepacia ACCD  -MNLQRFPRYPLTF-GPTPIQPLKRLSAHLGGKVELYAKREDCNSGLAFGGNKTRKLEYLV-PDA


                              1         1         1         1         1
              7         8         9         0         1         2         3
             678901234567890123456789012345678901234567890123456789012345678901234567890
             ----+---------+---------+---------+---------+---------+---------+
  P. sojae ACCD  IEGGYDTLVSIGGIQSNQTRQVAAVAAHLGFKCVLVQENWVNYPPEEAPVYDNVGNI-ELSRILG
P. infestance ACCD  IEGGYDTLVSIGGIQSNQTRQVAAVAAHLGFKCVLVQENWVNYPPEETNVYDNVGNI-ELSRILG
  Z. mays ACCD  IEGGYDTLVSIGGIQSNQTRQVAAVAAHLGMKCVLVQENWVNHGDKDG-VYDKVGNI-ELSRILG
  A. avenae ACCD  IAGGYDTLVSIGGIQSNQTRQVAAVAAHLGLKCVLVQENWVNY---SDAVYDRVGNI-EMSRILG
V. paradoxus ACCD  LEGGYDTLVSIGGIQSNQTRQVAAVAAHLGLKCVLVQENWVNYSD---AVYDRVGNI-EMSRIMG
  P. putida ACCD  IEQGCDTLVSIGGIQSNQTRQVAAVAAHLGMKCVLVQENWVNY---SDAVYDRVGNI-EMSRIMG
Pseudomonas ACCD  IEQGCDTLVSIGGIQSNQTRQVAAVAAHLGMKCVLVQENWVNY---SDAVYDRVGNI-EMSRIMG
Methylibium ACCD  LAQGCDTLVSIGGIQSNQTRQVAAVAAHLGMKCVLVQENWVNY---SDAVYDRVGNI-EMSRILG
   Pantoea ACCD  LAQGCDTLVSIGGVQSNQTRQVAAVAAHLGMKCVLVQENWVNYSD---AVYDRVGNI-ELSRIMG
B. ambitaria ACCD  LAQGADTLVSIGGVQSNQTRQVAAVAAHLGMKCVLVQEHWVNY---DDPVYDRVGNI-QLSRMMG
  B. cepacia ACCD  LAQGADTLVSIGGVQSNQTRQVAAVAAHLGMKCVLVQEHWVNY---EDPVYDRVGNI-QLSRMMG


                     1         1         1         1         1         1
              1         1         1         1         1         1
              4         5         6         7         8         9
             1234567890123456789012345678901234567890123456789012345
             ---------+---------+---------+---------+---------+-----
  P. sojae ACCD  ADVRRDAAGFDIGIRPSWEAAMESVKKAGGKPYPIPAGCSEHPKGGLGFVDFA-EEVRQQEKELG
P. infestance ACCD  ADVRRDSAGFDIGIRPSWEEAMESVKKAGGKPYPIPAGCSEHPKGGLGFVDFA-EEVRQQEKELG
  Z. mays ACCD  ADCRLDAAGFDIGIRPSWEQAMADVEKAGGKPYPIPAGCSEHPKGGLGFVGFA-EEVRQQEKESG
  A. avenae ACCD  ADVRLDAAGFDIGIRPSWEQAMADVKQAGGKPFPIPAGCSEHPYGGLGFVGFA-EEVRQQEKELG
V. paradoxus ACCD  ADVRLDAAGFDIGIRQSWEQAMADVRAAGGKPFPIPAGCSEHPRGGLGSVGFA-EEVRQQEAELG
  P. putida ACCD  ADVRLDAAGFDIGIRPSWEKAMSDVVERGGKPFPIPAGCSEHPYGGLGFVGFA-EEVRQQEKELG
Pseudomonas ACCD  AEVRLDAAGFDIGIRPSWEKAMDDVVSRGGKPFPIPAGCSEHPYGGLGFVGFA-EEVREQEKQLG
Methylibium ACCD  ADVRLDAAGFDIGIRPSWEQAMADVRAAGGKPFPIPAGCSEHRLGGLGFVGFA-EEVRAQEAELG
   Pantoea ACCD  ADVRLDAAGFDIGIRESWKQAMEEAAENGGKPFPIPAGCSEHPYGGLGFVGFA-EEVRQQEKELG
B. ambitaria ACCD  ADVRLVADGFDIGIRRSWEEAMESVRQAGGKPYPIPAGCSEHPLGGLGFVGFA-EEVRAQEAELG
  B. cepacia ACCD  ADVRLVADGFDIGIRRSWEEAMESVRQSGGKPYPIPAGCSEHPLGGLGFVGFA-EEVREQEAQLG
```
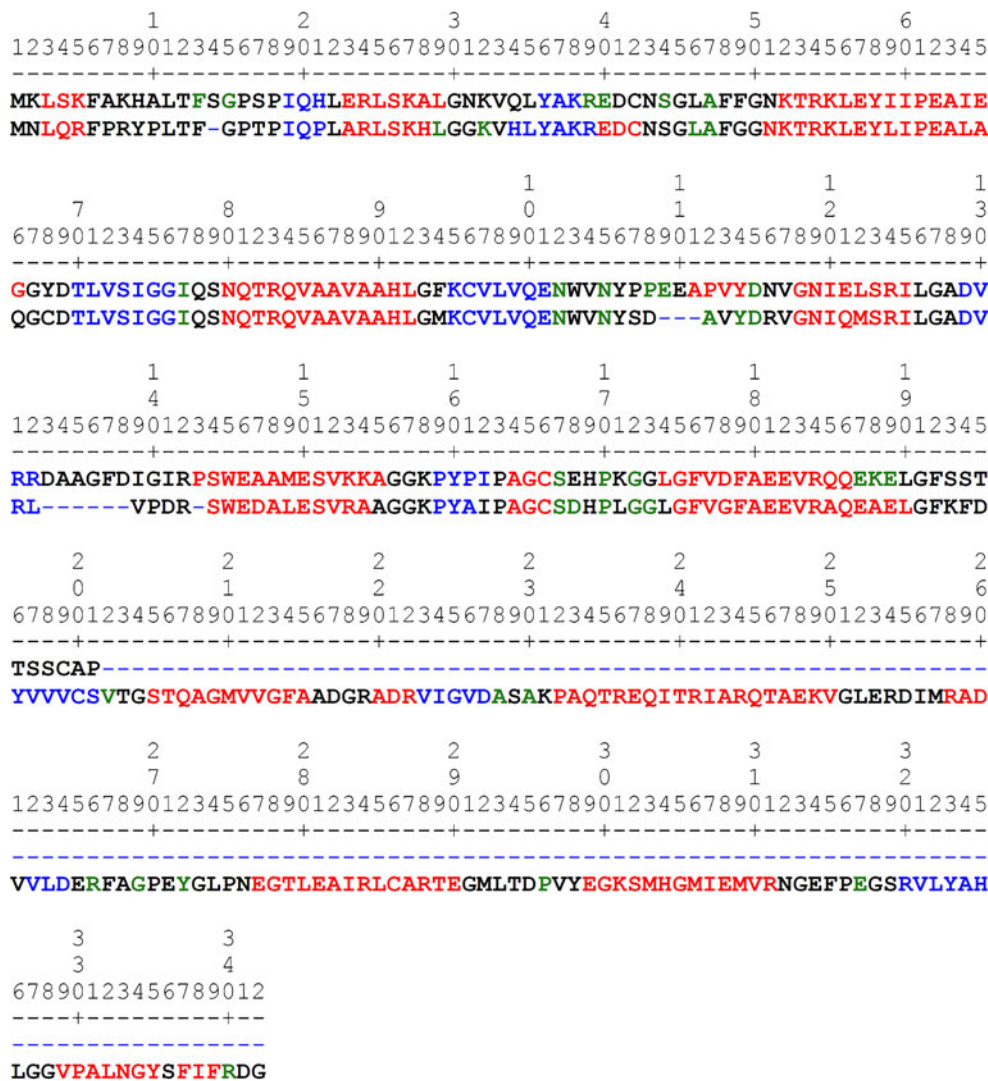
Fig. 3 Amino acid sequence alignment between *P. sojae* EST deduced sequence with putative and genuine ACCDs. The alignment and their identity percentages displays are from *Phytopthora sojae*; *Phytopthora infestance*-T30-490(*D0N448*) 95.8%; putative *Zea mays* (*B6SNR6*) 85.84%; *Acidovorax avenae* aubsp *avenae* ATCC19860 (*B9MGI8*) 82.8%; *Variovorex parafdoxus*-S110 (*Q6J256*) 78.6%; *Pseudomonas putida* (*A0MNM8*) 79.7%; *Methlibium petroleiphilum*-pm1 (*A2SLW2*) 78.1%; *Pantoea* sp.At9b (*C8Q2T3*) 77.6%; *Burkholderia ambitaria*-cmc40-6 (*B1YYR8*) 76.6%; *Burkhoideria cepacia* (*B8R7S1*) 75.5%. The residues highlighted in red represent complete conservation while those in blue are least (**Low 1 2 3 4 5 6 7 8 9 High**)

PsACCD are the presence of only seven helix (H1 to H7); longer loop length between aa 100 to 120 aa; longer size of strand β5(129 to 132 aa); relatively longer loop size between Glu[101] and Asn[119] than PdACCD (Glu[100] and Asn[115]); shorter C-terminal structure (domain one); length of loop between β strand 5 and helix 6 (133 to 142 aa-marked in figures as loop) (Fig. 5). The recognition of errors evaluated with PROCHECK showed all residues to be within the 0.07 to 0.75 values, while ProSA graphic assessment of Z score was found to be -4.4 (Fig. 6a). The Ramachandren plot analysis indicated that most (95%) residues have Φ and Ψ angles are in the core and favored regions, 97.5% or residues are in allowed regions and only 5 were in outlier regions (Fig. 6b). The superimposed backbone traces displayed 0.44 Å RMSD for Cα, 0.52 Å RMSD for back bone and 0.92 Å for all atoms calculated locally for developed model. The overall RMSD between the monomer of PsACCD and PdACCD (all chain) is 0.894 Å (Cα); 0.908 Å (back bone); 1.102 Å (heavy atoms); 1.102 Å (all atoms) while with HsACCD (all chain) it was 0.488 Å (Cα); 0.508 Å (back bone); 0.676 Å (heavy atoms); 0.676 Å (all atoms).

Fig. 4 Amino acid sequence alignment between PsACCD (upper line) and *Pseudomonas putida* ACCD (lower line). The number above the sequence represents the number of amino acids. The Smith –Waterman score - 874; %identity - 73.4%; amino acid overlap -192; Z score -829.7; E value -4.3e-39. Residues in red represents helix; those in blue represents strands; in green represents turn; those in black represents coil and are generated using SAS software (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/sas)

```
          1         2         3         4         5         6
 1234567890123456789012345678901234567890123456789012345
 ---------+---------+---------+---------+---------+-----
MKLSKFAKHALTFSGPSPIQHLERLSKALGNKVQLYAKREDCNSGLAFFGNKTRKLEYIIPEAIE
MNLQRFPRYPLTF-GPTPIQPLARLSKHLGGKVHLYAKREDCNSGLAFGGNKTRKLEYLIPEALA

                    1         1         1         1         1
 67890123456789012345678901234567890123456789012345678901234567890
 ----+---------+---------+---------+---------+---------+---------+
GGYDTLVSIGGIQSNQTRQVAAVAAHLGFKCVLVQENWVNYPPEEAPVYDNVGNIELSRILGADV
QGCDTLVSIGGIQSNQTRQVAAVAAHLGMKCVLVQENWVNYSD---AVYDRVGNIQMSRILGADV

          1         1         1         1         1         1
          4         5         6         7         8         9
 12345678901234567890123456789012345678901234567890123456789012345
 ---------+---------+---------+---------+---------+---------+-----
RRDAAGFDIGIRPSWEAAMESVKKAGGKPYPIPAGCSEHPKGGLGFVDFAEEVRQQEKELGFSST
RL------VPDR-SWEDALESVRAAGGKPYAIPAGCSDHPLGGLGFVGFAEEVRAQEAELGFKFD

   2         2         2         2         2         2         2
   0         1         2         3         4         5         6
 67890123456789012345678901234567890123456789012345678901234567890
 ----+---------+---------+---------+---------+---------+---------+
TSSCAP-----------------------------------------------------------
YVVVCSVTGSTQAGMVVGFAADGRADRVIGVDASAKPAQTREQITRIARQTAEKVGLERDIMRAD

          2         2         2         3         3         3
          7         8         9         0         1         2
 12345678901234567890123456789012345678901234567890123456789012345
 ---------+---------+---------+---------+---------+---------+-----
-----------------------------------------------------------------
VVLDERFAGPEYGLPNEGTLEAIRLCARTEGMLTDPVYEGKSMHGMIEMVRNGEFPEGSRVLYAH

   3         3
   3         4
 67890123456789012
 ----+---------+--
-----------------
LGGVPALNGYSFIFRDG
```

Ligand docking in ACCD modeled structure

The obtained ten best scoring poses, with least binding energy parameter were examined and compared; the pose having more energetically favorable ligand docking than others was selected and is presented in Fig. 7. The amino-acids lining the active sites were found to be Ser[17], Ala[37], Lys[38], Arg[39], Glu[40], Asp[41], Ala[47], Phe[49], Asn[51], Phe[137], Phe[179]. In comparisons HsACCD (1j0d) site is lined by Lys[54], Gly[75], Ser[78], Asn[79], Gly[79], Gln[80], Trp[102], Ala[163], Gly[164], Ser[166], Cys[200], Val[201], Thr[202], Gly[203], Ser[204], The[205], Tyr[295], Glu[295], Leu[323] and in PdACCD (1tz2) Asn[79], Gly[200], Lys[51], Lys[54], Ser[78], Thr[199], Thr[202], Val[198], Asn[50], Cys[196], Gln[80], Gly[161], Gly[295], Gly[323], Gly[324], Leu[322], Ser[197], Tyr[294] constituted the site. The comparison among geometrical docking of PsACCD-PLP-ACC homo-tetramer pose with those of PdACCD-PLP-ACC (1tz2) and HsACCD-PLP-ACC (1j0d) along with catalytic site

interactions detailing two dimensional diagram is presented in Fig. 8. In each case there are four monomer asymmetric units in the crystals. Further comparision of tetramer structure showed that the two active sites of the PsACCD open to opposite directions and are separated by ~37.97 Å distance. In the case of PdACCD this distance is ~23 Å and for HsACCD is ~36.592 Å. Each enzyme has different geometry. Comparison showed that HsACCD has a surface volume(SV) of 171.2e3 and surface area (SA) of 46.37e3 while in the case of PdACCD SV-165.2e3; SA-44.03e3 and for PsACCD it came about SV-93.08e3; SA-33.27e3.

Discussion

The plant growth promoting microorganisms stimulate plant growth through the activity of the enzyme ACCD which causes a lowering of plant ethylene levels. ACCD has been

Fig. 5 Linear depiction of PsACCD secondary structure. The structure ⌇ are helix labeled as H1, H2; and ⬭ as strands labeled by their sheets A, B while motifs β are beta turn and γ are gamma turn while ▭ is a beta hairpin. The important catalytic residues are marked with ▰ ▰. Short length region (135 to 145 aa) is labeled as loop
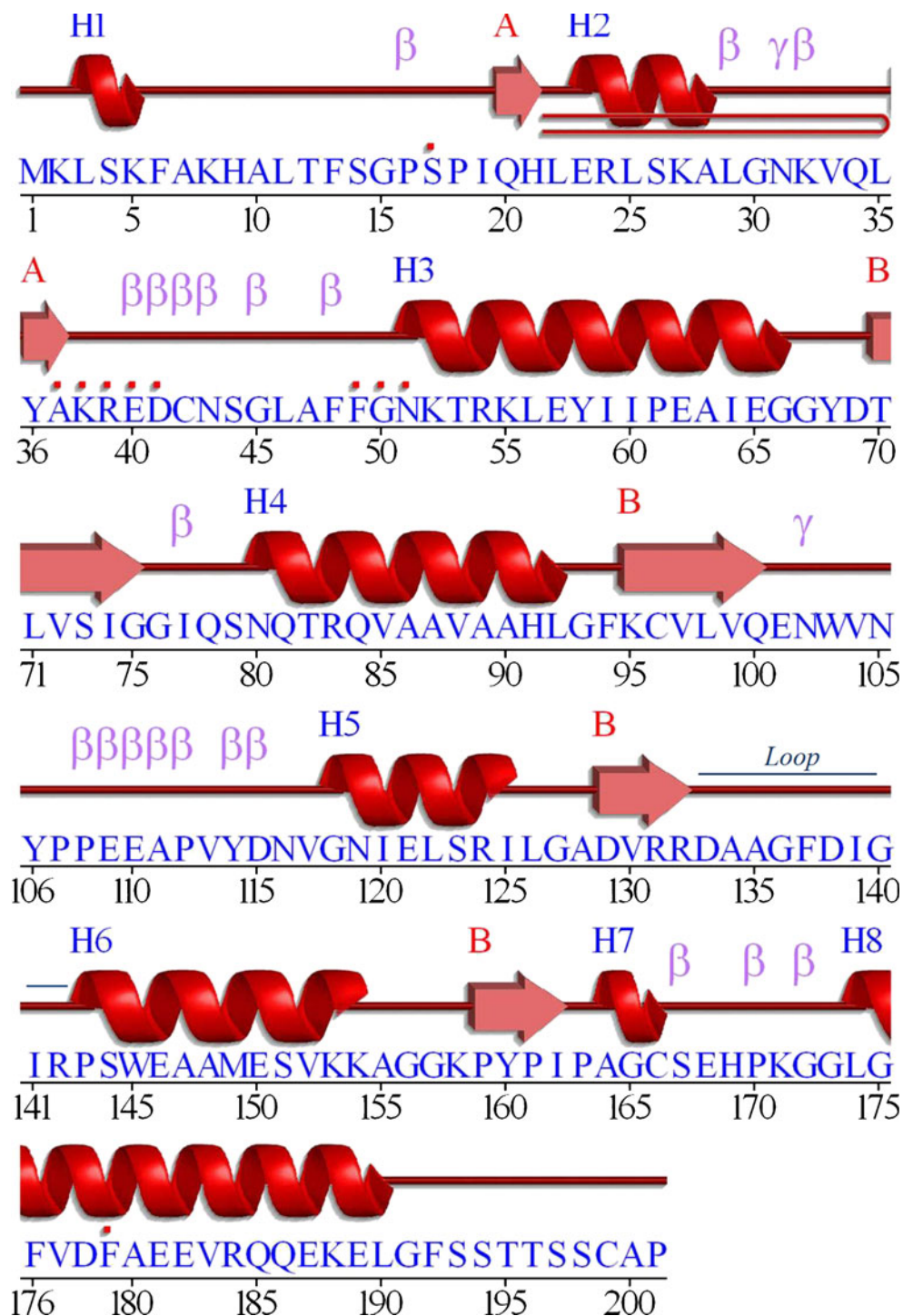
found in various bacteria, yeasts, and fungi, and it can convert ACC into α-ketobutyrate and ammonia [4] and thereby lower the level of ethylene in developing or stressed plants. Although ACCD has been identified and characterized in many organisms, there was no report from ACCD in *P. sojae*. Qutob et al. [20] and Trudy et al.[21] have reported unique ESTs of *P. sojae* spanning gene families encoding pathogenicity genes, elicitins, CRN "crinkler" proteins and endopolygalcturonases resulting from expansion and

diversification of the genes in response to selection pressure from Oomycete specific responses and the defense systems of host plants. Few functionally unidentified ESTs also

Fig. 6 a Evaluation of PsACCD model using ProSA web server. The plot indicating nearness of constructed model with the native structures. The Z-score of evaluated model was -4.4, shown as large black dot. **b** Ramachandren plot analysis showing placement of residues in deduced model (95.0% in favorable orientations). The structure residues orientation is separately considered for angle and torsions
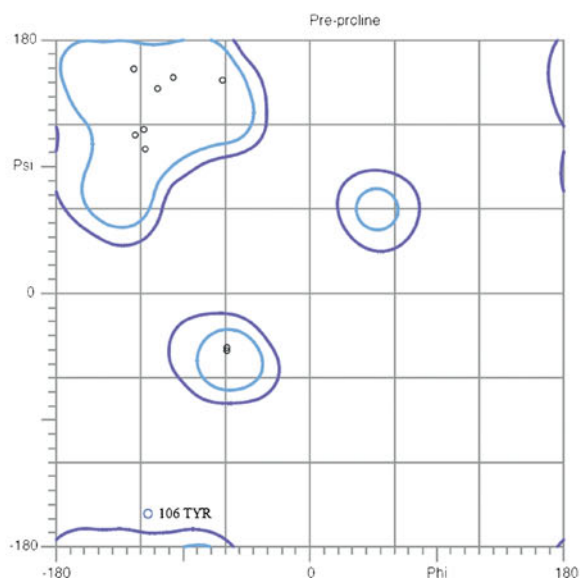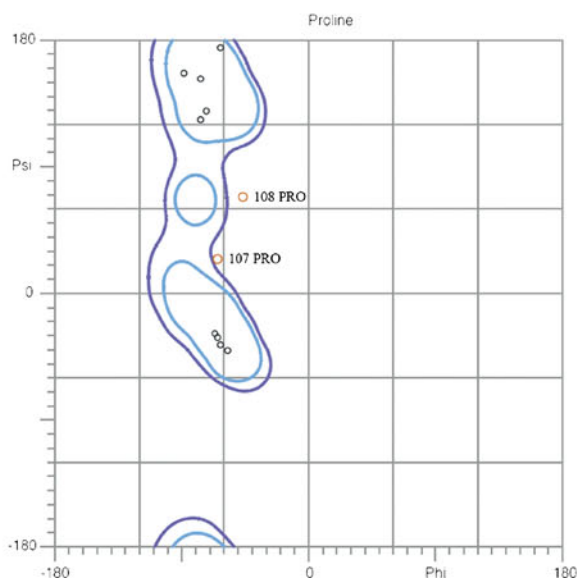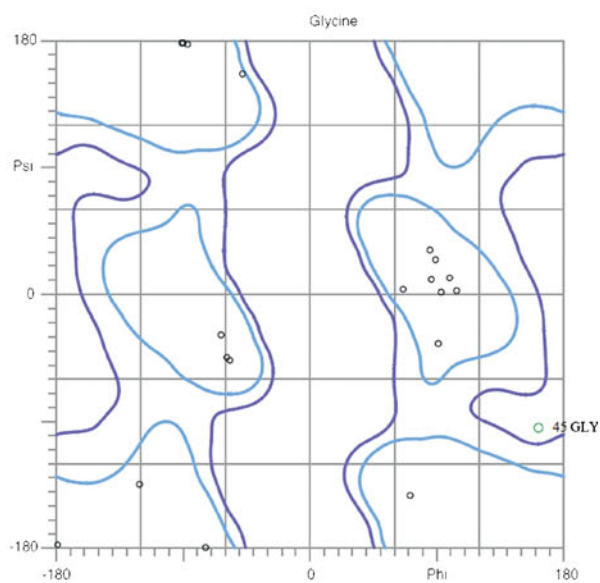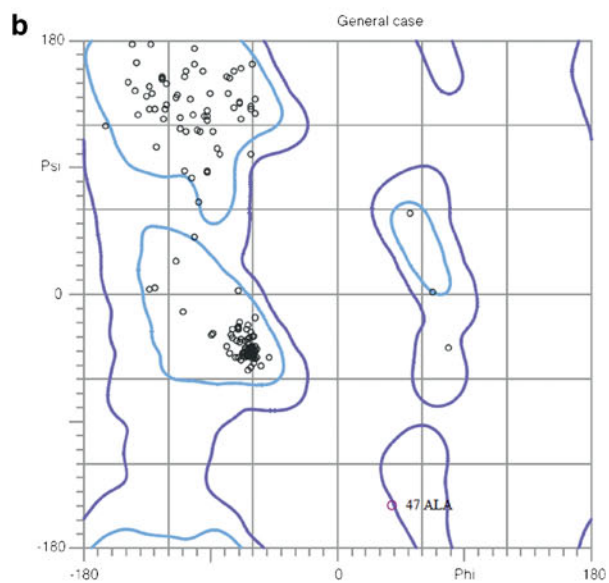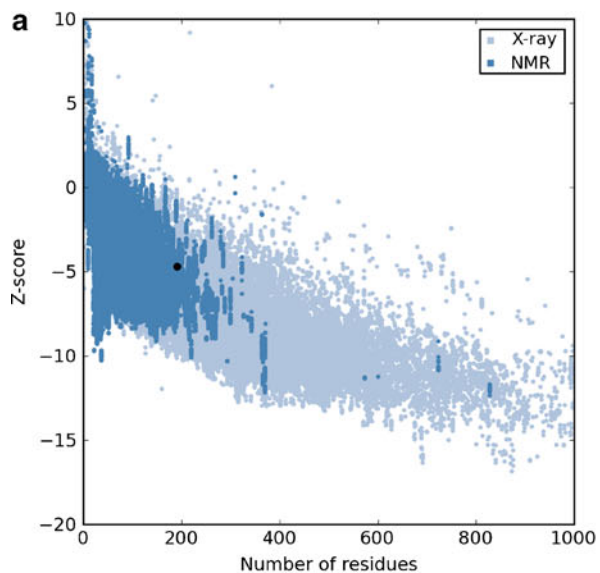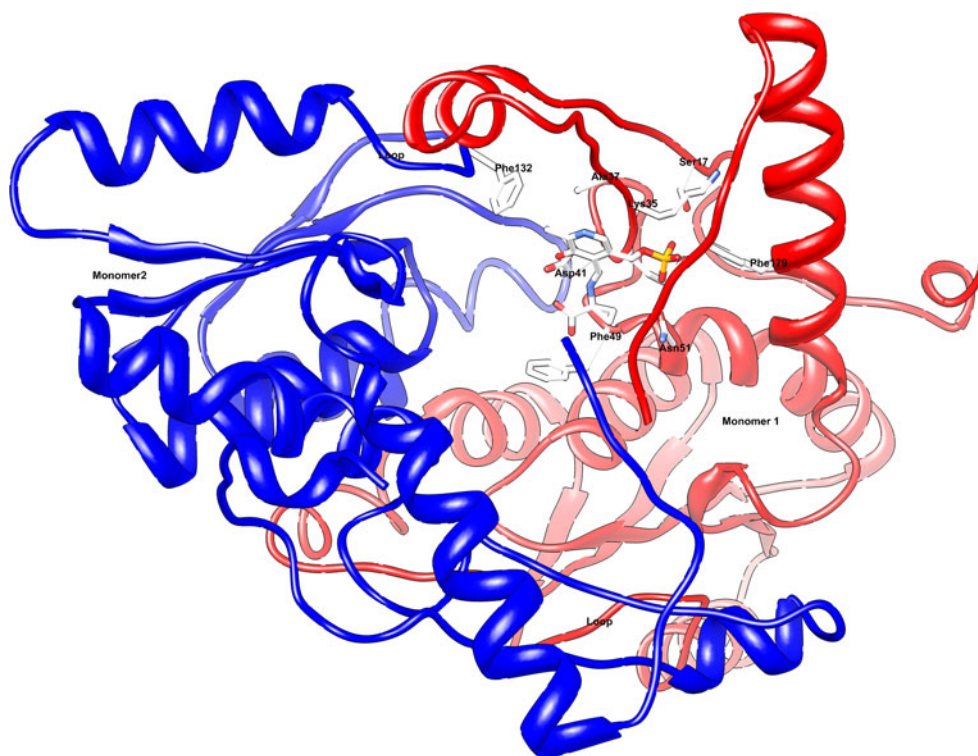
contribute to the *P. sojae* EST database. In the present study such functionally unidentified unique sequences representing the pathogen are analyzed using the ACCD mRNA sequence of *Aspergillus fumigates*. Out of few putative ESTs of *P.sojae* most similar EST was subjected to further analysis.

Clusters of CpG dinucleotides in GC rich regions of the genome called "CpG islands" frequently occur in the 5' ends of genes. This fact, that a better and increased frequency for the association of CpG islands with genes is found, indicates the usefulness of CpG island extraction for gene prediction [22]. The translated protein of the *P. sojae* EST showed its nearness with other ACCD of microbial origin. The Pro-Function analysis also showed its homology with DNA binding structure and this property, a new function attributed to it, further can only be speculated.

The similarity search within the PDBe entry invariably showed similarity with ACCD group of structures. In case of most bacterial phytopathogens it is shown that they manipulate their hosts by secreting effector proteins which may help to overcome host defenses and aid in the proliferation of the pathogen into the plant. Alternatively, it may induce defense responses in the plant and causes host resistance [23–25]. Secretion serves as an important mechanism for delivery of pathogenicity factors into host tissue. There is growing evidence that phyto-pathogenic oomycetes produce secreted effector proteins [26, 27]. In our case PLOC analysis showed that the protein has a tendency to be sub-cellular localized and may not be involved in secretion.

It was ascertained that the three dimensional structure of the identified protein was not available in the protein databank hence we further tried to ascertain the three dimensional protein structure of the identified EST. The structure folding of the PsACCD showed high similarity with the PdACCD. The view was further supported by the NCBI CDD prediction that the domain of PsACCD may catalyze a cyclopropane ring-opening reaction, the irreversible conversion of ACC to ammonia and alpha-ketobutyrate and may allow its growth on ACC as an alternative nitrogen source [28]. The protein sequence alignment highlighted the degree of sequence conservation and high sequence similarity predicted the orthologous nature of PsACCD along with other bacterial species. The phylogenetic analysis also indicated the above facts. The alignment shows that most of the important residues are conserved which are important for the ACCD reaction [28]. The PsACCD protomer consists of a large and a small domain (supplementary figure). The small domain is folded as a twisted ά/β sheet structure. The N-terminal helix (helix1) exists apart from the core region. The large domain (residues 1-53 and 161-325) is composed of nine ά helices and six β stands, which are folded as a twisted ά/β sheet somewhat resembling the small domains. The overall fold comparision among PsACCD, PdACCD and HsACCD indicated that it belongs to the PLP dependent proteins [29, 30]. These are also related to *O*-acetylserine sulfhydrylase [31], threonine deaminase [32], and tryptophan synthase *â* subunit [33, 34].

The low RMSD value of superimposition indicates fairly high similarity among the target and the template structure.

Fig. 8 The comparative figure showing tertramer structure ACCD enzyme substrate complexes of *Hansenula saturnus* (PDB: 1J0D), *Pseudomonas sp.* (PDB: 1TZ2) and *Phytopthora sojae* (PsACCD). The active site detail interactions of each structure were shown as 2D ligand plot reveling interacting amino acids (numbered as in sequences) with substrate
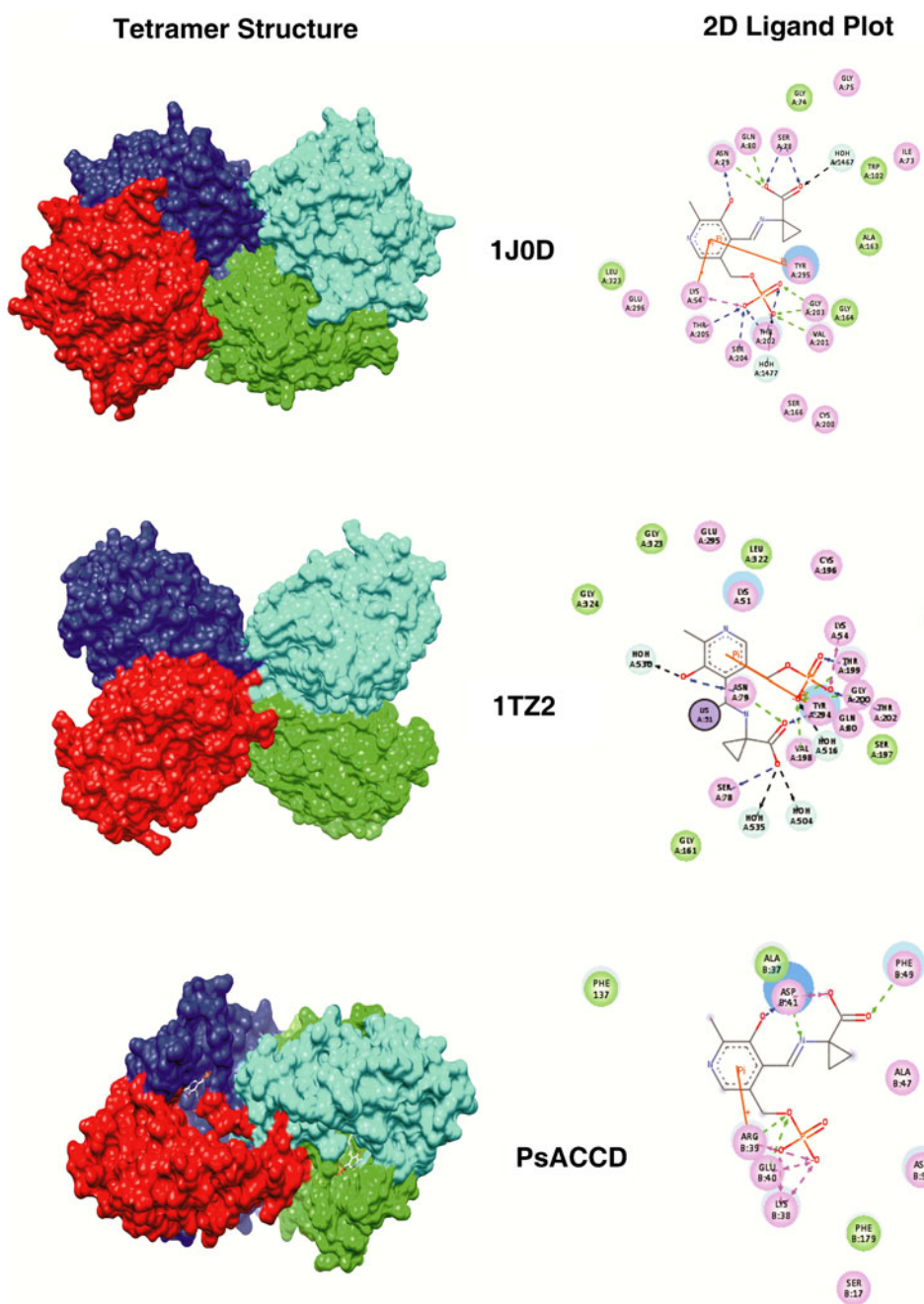
The peptide reduction of domain 2 might be the result of genome reduction phenomenon. This lack in domain 2 (Fig. 5; Suppl. Fig.) created an open enzymatic binding site in PsACCD monomer. Since, PdACCD likely forms dimers in solution and PsACCD monomer superimposes well with the PdACCD enzyme (all chain) with an RMSD of 0.894 Å, analogously indicating that the PsACCD should also have a dimeric structure. It also hypothesizes for more favorable binding and a functional PsACCD due to possible participation of residues from second monomer (Fig. 7). To ascertain the functionality of the enzymatic site, rigid ligand docking study was carried out using the PatchDOCK server

which uses local shape featured matching and detects high probability molecular surface areas as binding site. Moreover, this program has flexibility to process more than one substrate simultaneously, as in our case.

Observing the docking pose it was clearly evident that active site is not similar to other ACCD available structure. It is located between deep cervices of dimeric structure. While in the case of homotetramer structure its dual non interacting enzymatic sites open to opposite directions, separated by ~37.97 Å distance arranged around an axis. The two dimensional ligand interactions plot showed that Phe[49], Ala[37] and Asp[41] are simultaneously involved in the biding

with cofactor PLP via Schiff base and reorganizing the ACC. These interactions are considered to be important for ACC cleavage reaction [28], while Arg^{39}, Glu^{40} and Lys^{38} help in securing and retaining the cofactor within active site and may directly participate in the nucleophylic cleavage reaction of ACC molecule (Fig. 8). This confirmation might be near the actual active state of PsACCD. Based on the above facts it can be concluded that, the EST encoded portion matches with bacterial ACCD and the encoded protein is capable of enzymatically convert ACC into α-ketobutyrate and ammonia. This conformation may be near to the natural state in which PsACCD might naturally exist. This study provides identification and characterization of the PsACCD and it may be helpful in the transcriptional and expression based study of *P. sojae* pathogenesis.

References

1. Boller T (1991) Ethylene in pathogenasis and disease resistance. In: Mattoo AK, Suttle JC (ed) The plant hormone ethylene, CRC Press, pp 293-314
2. McKeon T, Yang SF (1987) Biosynthesis and metabolism of ethylene. In: Davies PJ (ed) Plant hormones and their role in plant growth and development. Martinus Nijhoff, Boston, pp 94–112
3. Madhaiyan M, Poonguzhali S, Sa T (2007) Characterization of 1-aminocyclopropane-1-carboxylate deaminase containing *Methylobacterium* spp. Isolated from rhizosphere soils of field-grown rice and regulation of ethylene levels in canola. Planta 226:867–876. doi:10.1007/s00425-007-0532-0
4. Glick BR, Karaturovic DM, Newell PC (1995) A novel procedure for rapid isolation of plant growth promoting Pseudomonads. Can J Microbiol 41:533–536. doi:10.1139/m95-070
5. Penrose DM, Glick BR (1997) Enzymes that regulate ethylene levels–1-aminocyclopropane-1-carboxylic acid (ACC) deaminase, ACC synthase and ACC oxidase. Indian J Exp Biol 35:1–17
6. Glick BR, Penrose DM, Li J (1998) A model for the lowering of plant ethylene concentration by the plant growth promoting bacteria. J Theor Biol 190:63–68. doi:10.1006/jtbi.1997.0532
7. Sheehy RE, Honma M, Yamada M, Sasaki T, Martineau B, Hiatt WR (1991) Isolation, sequence, and expression in *Escherichia coli* of the *Pseudomonas* sp. Strain ACP gene encoding 1-aminocyclopropane-1-carboxylate deaminase. J Bacteriol 173:5260–5265
8. Klee HJ, Hayford MB, Kretzmer KA, Barry GF, Kishore GM (1991) Control of ethylene synthesis by expression of a bacterial enzyme in transgenic tomato plant. Plant Cell 3:1187–1193. doi:10.1105/tpc.3.11.1187
9. Jacobson CB, Pasternak JJ, Glick BR (1994) Partial purification and characterization of ACC deaminase from the plant growth promoting rhizobacterium *Pseudomonas putida* GR12-2. Can J Microbiol 40:1019–1025. doi:10.1139/m94-162
10. Campbell BG, Thomson JA (1996) 1-Aminocyclopropane-1-carboxylate deaminase gene from *Pseudomonas* strains. FEMS Microbiol Lett 138:207–210. doi:10.1016/0378-1097(96)00108-5
11. Shah S, Li J, Moffatt BA, Glick BR (1998) Isolation and characterization of ACC deaminase gene from two different plant growth promoting rhizobacteria. Can J Microbiol 44:833–843. doi::10.1139/w98-074
12. Minami R, Uchiyama K, Murakami T, Kawai J, Mikami K, Yamada T, Yokoi D, Ito H, Matsui H, Honma M (1998) Properties, sequence, and synthesis in *Escherichia coli* of 1-aminocyclopropane-1-carboxylate deaminase from *Hensula saturnus*. J Biochem 123:1112–1118
13. Jia YJ, Kakuta Y, Sugawara M, Igarashi T, Oki N, Kisaki M, Shoji T, Kanetuna Y, Horita T, Matsui H, Honma M (1999) Synthesis and degradation of 1-aminocyclopropane-1-carboxylic acid by *Penicillium citrinum*. Biosci Biotechanol Biochem 63:542–549. doi:10.1271/bbb.63.542
14. Kumar S, Dudley J, Nei M, Tamura K (2008) MEGA: biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9:299–306
15. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Evaluation of comparative protein modeling by MODELLER. Protein 23:318–326. doi:10.1002/prot.340230306
16. Laskowski RA, MacArtur MW, Moss DS, Thrnton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291. doi:10.1107/S0021889892009944
17. Vriend G (1990) WHAT IF: a molecular modeling and drug designing program. J Mol Graph 8:52–56
18. Karthikeyan S, Zhou Q, Zhao Z, Kao CL, Tao Z, Robinson H, Liu HW, Zhang H (2004) Structural analysis of Pseudomonas 1-aminocyclopropane-1-carboxylate deaminase complexes: insight into the mechanism of a unique pyridoxal-5'-phosphate dependent cyclopropane ring-opening reaction. Biochemistry 43:13328–13339. doi:10.1021/bi048878g
19. Duhovny DS, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acid Res 33:W363–W367. doi:10.1093/nar/gki481
20. Qutob D, Hraber PT, Sobral BWS, Gijzen M (2000) Comparative analysis of expressed sequences in *Phytophthora sojae*. Plant Physiol 123:243–253. doi:10.1104/pp.104.055624
21. Trudy A, Tripathy S, Smith BM, Arredondo FD, Zhou L, Li H, Chibucos MC, Qutob D, Gijzen M, Mao C, Sobral BWS, Waugh ME, Mitchell TK, Dean RA, Tyler BM (2007) Expressed sequence tags from *Phytophthora sojae* reveal genes specific to development and infection. Mol Plant Microb Interact 20:781–793. doi:10.1094/MPMI-20-7-0781
22. Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. PNAS 99:3740–3745. doi:10.1073/pnas.052410099
23. Knogge W (1996) Fungal infections of plants. Plant Cell 8:1711–1722. doi:10.1105/tpc.8.10.1711
24. Collmer A, Badel JL, Charkowski AO, Deng WL, Fouts DE, Ramos AR, Rehm AH, Anderson DM, Schneewind O, van Dijk K, Alfano JR (2000) *Pseudomonas syringae* Hrp type III secretion system and effector proteins. PNAS 97:8770–8777
25. Staskawicz BJ, Mudgett MB, Dangl JL, Galan JE (2001) Common and contrasting themes of plant and animal diseases. Science 292:2285–2289. doi:10.1126/science.1062013
26. Huitema E, Bos JIB, Tian M, Win J, Waugh ME, Kamoun S (2004) Linking sequence to phenotype in *Phytophthora*–plant interactions. Trends Microbiol 12:193–200. doi:10.1016/j.tim.2004.02.008
27. Kamoun S (2006) A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol 44:2.1–2.20. doi:10.1146/annurev.phyto.44.070505.143436

28. Ose T, Fujino A, Yao M, Natanabe N, Honma M, Tanaka I (2003) Reaction intermediate structure of 1-aminocyclopropane-1-carboxylate deaminase. J Biol Chem 278:41069–41076. doi:10.1074/jbc.M305865200

29. Mehta PK, Christen P (1998) In: Purich DL (ed) Advances in enzymology and related areas of molecular biology. Wiley, New York, pp 129–184

30. Grishin NV, Phillips MA, Goldsmith EJ (1995) Modeling of the spatial structure of eukaryotic ornithine decarboxylases. Protein Sci 4:1291–1304. doi:10.1002/pro.5560040705

31. Burkhard P, Rao GS, Hohenester E, Schnackerz KD, Cook PF, Jansonius JN (1998) Three-dimensional structure of O-acetylserine sulfhydrylase from *Salmonella typhimurium*. J Mol Biol 283:121–133

32. Gallagher DT, Gilliland GL, Xiao G, Zondlo J, Fisher KE, Chinchilla D, Eisenstein E (1998) Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase. Structure 6:465–475. doi:10.1016/S0969-2126(98)00048-3

33. Rhee S, Miles EW, Davies DR (1998) Cryocrystallography of a true substrate, indole-3-glycerol phosphate, bound to a mutant (RD60N) tryptophan synthase R2â2 complex reveals the correct orientation of active site RGlu49. J Biol Chem 273:8553–8555

34. Weyand M, Schlichting I (1999) Crystal structure of wildtype tryptophan synthase complexed with the natural substrate indole-3-glycerol phosphate. Biochemistry 38:16469–16480. doi:10.1021/bi9920533