Journal of Theoretical Biology 407 (2016) 184–197

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Elucidating modes of activation and herbicide resistance by sequence assembly and molecular modelling of the Acetolactate synthase complex in sugarcane

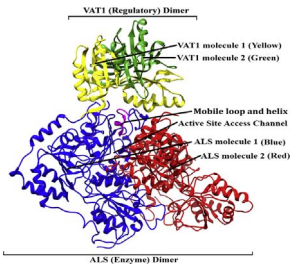
Dyfed Lloyd Evans [a,b,*], Shailesh Vinay Joshi [a,b]

[a] South African Sugarcane Research Institute, 170 Flanders Drive, Private Bag X02, Mount Edgecombe, Durban 4300, South Africa
[b] School of Life Sciences, College of Agriculture, Engineering and Science, University of Kwa-Zulu Natal, Private Bag X54001, Durban 4000, South Africa

HIGHLIGHTS

- Acetolactate synthase and its regulator VAT1 for sugarcane are assembled.
- Molecular models of sugarcane ALS and VAT1 were generated.
- ALS and VAT1 were docked using co-evolutionary studies to determine binding sites.
- The complexed model was used to predict the activation effect of VAT1 binding.
- The model was employed to predict the feedback inhibition mechanism of VAT1 on ALS.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 10 February 2016
Received in revised form
14 June 2016
Accepted 20 July 2016
Available online 22 July 2016

Keywords:
Acetolactate synthase
Molecular modelling
Sequence assembly
VAT1 activation
VAT1 inhibition

ABSTRACT

Acetolactate synthase (ALS) catalyzes the first portion of the biosynthetic pathway leading to the generation of branched-chain amino acids. As such it is essential for plant health and is a major target for herbicides. ALS is a very poorly characterized molecule in sugarcane. The enzyme is activated and inhibited by a regulatory subunit (known as VAT1 in plants) whose mode of action is entirely unknown. Using *Saccharum halepense* as a template we have assembled the ALS gene of sugarcane (*Saccharum* hybrid) and have modelled the structure of ALS based on an *Arabidopsis* template (the first ALS model for a monocot). We have also assembled the ALS regulatory proteins (VAT1 and VAT2) from sugarcane and show that VAT2 is specific to true grasses. Employing a bacterial model, we have generated a structural model for VAT1, which explains why the separate domains of the proteins bind to either leucine or valine but not both. Using co-evolution studies we have determined molecular contacts by which we modelled the docking of VAT1 to ALS. In conclusion, we demonstrate how the binding of VAT1 to ALS activates ALS and show how VAT1 can also confer feedback inhibition to ALS. We validate our ALS model against biochemical data and employ this model to explain the function of a novel herbicide binding mutant in sugarcane.

1. Introduction

Acetolactate synthase (ALS) EC 2.2.1.6 is an enzyme that catalyzes the first portion of the biosynthetic pathway leading to the generation of branched-chain amino acids (leucine, isoleucine and

* Corresponding author at: South African Sugarcane Research Institute, 170 Flanders Drive, Private Bag X02, Mount Edgecombe, Durban 4300, South Africa. Fax: +27 315087597.
E-mail address: dyfed.evans@sugar.org.za (D. Lloyd Evans).

http://dx.doi.org/10.1016/j.jtbi.2016.07.025
0022-5193/

valine) (Chipman et al., 1998; Duggleby and Pang, 2000). Specifically, it synthesizes the formation of acetolactate from pyruvate, a key step in the formation of valine. In addition, it can transfer the acetaldehyde from pyruvate to 2-oxobutanoate, forming 2-ethyl-2-hydroxy-3-oxobutanoate, also known as 2-aceto-2-hydroxybutanoate, a reaction in the biosynthesis of isoleucine (Duggleby and Pang, 2002).

Structurally, acetolactate synthase (ALS) is a highly ordered molecule, with the monomer comprising three domains: α, β and γ (Pang et al., 2002). The active form is a homodimer with the interface between the two subunits stabilized by the presence of flavine adenine dinucleotide (FAD) (McCourt et al., 2006). FAD also helps stabilize the structure of the active site, which is formed from the junctures of the α–subunit of the first molecule and the β– and γ–subunits of the second molecule. The active site itself is buried within the dimer interface and is reached via an access channel (McCourt et al., 2006; McCourt and Duggleby, 2006).

In weed control, ALS is the target of many herbicides (in particular sulfonylurea [SU], imidazolinone [IMI], triazolopyrimidine [TP], pyrimidinyl oxybenzoate [POBs], and sulfonylamino carbonyl triazolinone [SCTs]) (Zhou et al., 2007) which inhibit the function of ALS by binding to the enzyme's active site access channel thus preventing substrate from gaining entry (McCourt and Duggleby, 2006). This prevents the formation of branched chain amino acids, which is why herbicides that inhibit ALS function are lethal. Since their introduction in the 1980s, herbicides that target ALS figure amongst the most widely used weed control chemicals (Zhou et al., 2007).

Sugarcane is a major crop in the tropical and subtropical regions of the world; indeed, it ranks amongst the top 10 crop species world-wide. Sugarcane also provides between 60% and 70% of total world sugar output and is a major source of bioethanol (Reddy et al., 2008). ALS inhibitors are an important means of weed control in sugarcane production (Smeets et al., 2008). However, like most grass species, sugarcane itself is susceptible to ALS inhibiting herbicides (Monquero et al., 2011).

ALS is present in all kingdoms of life apart from animals (Gedi and Yoon, 2012); which is why animals must obtain branched chain amino acids via their diet. In fungi, ALS was derived from the mitochondrion (Avalos et al., 2013) whilst in plants ALS was originally derived from chloroplasts. In both kingdoms, ALS is expressed in the nuclear genome but is directed towards the plastome by means of a transit peptide that is cleaved within the organelle to form the active enzyme (Lee et al., 1999).

The catalytic subunit of ALS (CSU) is invariably associated with a regulatory subunit (RSU). In bacteria these regulatory subunits are small (~15 kDa), whilst in plants they are almost the same size as the catalytic subunit (50 kDa, as compared to 60 kDa for the CSU) (McCourt and Duggleby, 2006). The regulatory subunits bind to the catalytic subunit and enhance activity by between 7 and 11 fold (McCourt and Duggleby, 2006). The regulatory subunits also enable feedback inhibition by valine, leucine or isoleucine, though inhibition is never complete (between 11% and 67% activity always remains even at saturating concentrations of an inhibitory amino acid) (McCourt and Duggleby, 2006).

Globally, with the wide-scale use of ALS inhibiting herbicides, it is hardly surprising that many crop and weed species have developed resistance to these herbicides. The majority of these mutations affect the active site access channel, either preventing access of herbicides or reducing docking affinity (Guttieri et al., 1992; Bernasconi et al., 1995; Powles and Yu, 2010; Warwick et al., 2008).

Both the ALS catalytic (Chang and Duggleby, 1997) and regulatory subunits (Lee and Duggleby, 2001) have been cloned from *Arabidopsis thaliana*. Co-expression of the *A. thaliana* catalytic and regulatory subunits revealed that in the active complex, the CSUs and RSUs are present at a ratio of 1:1 and that the mature form of the enzyme contains four catalytic subunits (Lee and Duggleby, 2001). From crystallographic studies it is proposed that the mature form of *Arabidopsis* ALS is a dimer of homodimers (i.e. two catalytic homodimers bound together as a tetramer) (McCourt et al., 2006). This means that the mature form of the enzyme has four catalytic sites and that four regulatory subunits are bound to this structure.

The structure of the Arabidopsis RSU is not known, but structures from several bacterial orthologues have been solved (Petkowski et al., 2007; Kaplun et al., 2006) making molecular modelling feasible.

Despite the importance of ALS inhibiting herbicides in sugarcane weed management and sugarcane's own susceptibility to these herbicides, nothing is known about sugarcane ALS. The gene has not been sequenced in its entirety and there is only a single report of a sugarcane ALS mutant that confers herbicide resistance, but whose mechanism of action remains unknown, [GenBank: EU243999]. However, there is a high-depth short read whole genome archive for sugarcane [SRA:SRR427145] and the genomes of two close relatives, *Sorghum bicolor* (Paterson et al., 2009) and *Zea mays* (Schnable et al., 2009) have been assembled. Both the ALS CSU genes and the ALS RSU genes (known as VAT1) have been assembled in these genomes. This raises the intriguing possibility of using the maize and sorghum genes to bait out the cognate reads from the sugarcane dataset before *de novo* assembly of these regions in sugarcane.

As the structure of the Arabidopsis CSU and of bacterial LSUs are known it should be possible to model these subunits in sugarcane to yield structural models which would enable us to analyze the binding of herbicides to sugarcane ALS and also to determine the effect of all herbicide inhibiting mutants. The sequences of ALS and VAT1 from many plant species are known and these genes have cognates in both the bacterial and fungal kingdoms. This means that there is sufficient evolutionary sequence depth to examine the co-evolution of these two genes.

Recent advances in gene co-evolution studies to predict protein–protein binding sites reveal that these techniques are now the most accurate means of determining protein–protein interactions for docking studies. This means that, for the first time, it is possible to accurately predict how ALS and VAT1 bind together, leading to the activation of ALS.

Indeed, considerable work has demonstrated the accuracy of coevolution-based contact prediction for monomeric proteins using a global statistical model (Thomas et al., 2008) to distinguish between direct and indirect couplings (Marks et al., 2011; Morcos et al., 2011; Hopf et al., 2012; Nugent and Jones, 2012; Jones et al., 2012; Lapedes et al., 2012; Marks et al., 2012; Sułkowska et al., 2012; Kamisetty et al., 2013). More recent studies have shown that a pseudo-likelihood-based approach (Balakrishnan et al., 2011) results in more accurate predictions (Ekeberg et al., 2013; Kamisetty et al., 2012) for a range of alignment sizes and protein lengths. The latest advances have employed the determination of protein co-evolution using global statistical models as implemented in the papers of Tamir et al. (2014), Hopf et al. (2014) and Ovchinnikov et al. (2014). Such implementations are currently the most accurate methods for determining protein–protein contact sites for two molecules based on co-evolution studies.

In bacteria, ALS is expressed as multiple isozymes, each of which is an operon expressing a specific copy of the catalytic subunit with its cognate regulatory subunit (Vitreschak et al., 2006). In the IlvN form of *Escherichia coli* ALS (Karanth and Sarma, 2013) the regulatory subunit is formed from a dimer of ACT-like domains only. All other structures of bacterial ALS regulatory subunits show that the RSU is formed from the dimerization of two 180 bp proteins (Petkowski et al., 2007; Kaplun et al., 2006).