J Mol Model (2006) 12: 905–914
DOI 10.1007/s00894-005-0071-1

ORIGINAL PAPER

Christopher D. Dana · David R. Bevan · Brenda S. J. Winkel

Molecular modeling of the effects of mutant alleles on chalcone synthase protein structure

Received: 11 January 2005 / Accepted: 8 July 2005 / Published online: 31 March 2006

Abstract Chalcone synthase (CHS) catalyzes the first committed step in flavonoid biosynthesis, a major pathway of plant secondary metabolism. An allelic series for the *Arabidopsis* CHS locus, *tt4*, was previously characterized at the gene, protein, and end-product levels. In an effort to deduce the molecular basis for the observed phenotypes, homology models were generated for five of the tt4 proteins based on the crystal structure of CHS2 from *Medicago*. Molecular dynamics simulations provided insights into how even those substitutions that are not in close spatial proximity to key functional residues may still alter the architecture and dynamic movement of the enzyme, with dramatic effects on enzyme function. Simulations carried out at different temperatures pointed to optimized positioning of key residues in the active site or dimerization domain, rather than enhancement of overall structure, as underlying the higher activity of two temperature-sensitive variants at lower temperatures. Extending this type of analysis to account for protein–protein interactions may offer additional insights into the mechanisms by which single amino-acid substitutions can affect diverse aspects of protein function.

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00894-005-0071-1 and is accessible for authorized users

C. D. Dana · B. S. J. Winkel (✉)
Department of Biology and Fralin Biotechnology Center,
Virginia Tech,
Blacksburg, VA 24061, USA
e-mail: winkel@vt.edu
Tel.: +1-540-2313013
Fax: +1-540-2317126

D. R. Bevan
Department of Biochemistry, Virginia Tech,
Blacksburg, VA 24061, USA

Present address:
C. D. Dana
Department of Biochemistry and Molecular Biology,
Michigan State University,
East Lansing, MI 48824, USA

Keywords Homology modeling · Molecular dynamics simulation · Allelic series · Chalcone synthase · *transparent testa* · *Arabidopsis thaliana*

Abbreviations CHS: Chalcone synthase · *tt*: Transparent testa · RMSD: Root mean square deviation

Introduction

The type III polyketide synthases (PKSs) comprise a structurally related class of enzymes that are key to the production of an array of biologically active natural products in plants and some bacteria [1, 2]. These enzymes, also known as homodimeric iterative PKSs, include the extensively studied chalcone synthase (CHS), an enzyme that is ubiquitous among plants and that catalyzes the first committed step in the synthesis of flavonoids. In the CHS reaction, one molecule of *p*-coumaroyl CoA and three molecules of malonyl CoA are used to generate a tetraketide that is then cyclized into 4,2',4',6'-tetrahydroxychalcone. This molecule serves as a backbone for the production of a wide array of flavonoid compounds that function in such diverse roles as flower pigmentation, UV protection, signaling, male fertility, and defense against microbial pathogens [3].

CHS2 from *Medicago sativa* was the first type III PKS, and the first flavonoid enzyme, for which a crystal structure was solved [4]. Together with the subsequent determination of structures for mutagenized forms of this enzyme and for 2-pyrone synthase (2PS) from *Gerbera hybrida* [5, 6], significant insights have been gained into the molecular basis of the reaction series catalyzed by the type III PKSs. For example, it was known that, although each of the two active sites in CHS and the closely related enzyme, stilbene synthase (STS), function independently, dimerization is required for activity [7]. It is now clear that this is because a methionine in each monomer helps shape the active-site cavity of the adjoining subunit [4]. It has also been shown that each active site is buried within the enzyme and that substrates enter *via* a long (16 Å in *Medicago* CHS2) CoA-

binding tunnel [4]. The three key catalytic residues are located within an initiation/elongation cavity, one lobe of which forms a coumaroyl binding pocket, while the other accommodates and determines the length of the growing polyketide chain. Although CHS and 2PS use different substrates, produce different products, and exhibit only 74% amino-acid sequence identity, the overall architecture of the two enzymes is nearly identical, with a root mean square deviation (RMSD) value of 0.64 Å for the C_α atoms [8]. Remarkably, it is possible to engineer CHS for 2PS activity by altering just three residues that affect the overall dimensions of the initiation/elongation cavity [8].

Molecular dynamics simulations of protein structures are increasingly being used to explore the effects of amino-acid substitutions on protein function. For example, Ceruso et al. [9] were able to suggest a nucleotide-exchange mechanism for the Gα subunit of transducin by modeling a series of substitutions in residues that had been previously characterized biochemically. Such simulations have also been used to show that two gain-of-function mutations in the glucocorticoid receptor appear to act *via* different mechanisms [10] and to provide insights into the structural and dynamic effects of several mutations associated with type 2A von Willebrand disease [11]. Here we describe a similar approach to understanding the molecular basis of mutations in CHS from *Arabidopsis thaliana* identified in screens for plants with altered seed-coat pigmentation. A homology model of the wild-type enzyme based on the structure of *Medicago* CHS2 previously showed that most of these mutations affected residues located a significant distance from the catalytic center [12]. We have now extended this analysis, generating homology models for five of the variant enzymes and using molecular dynamics simulations to examine the defects in these proteins in greater detail, and providing insights into both the local and long-range effects of single amino acid substitutions on enzyme function.

Materials and methods

A multiple sequence alignment of 15 representative members of the plant type III PKS superfamily was generated with Clustal W [13] using the Megalign program in DNAStar (Madison, WI). Included in this analysis were CHS from *Arabidopsis* (accession number P13114), *Medicago* (P30074), *Zea mays* (SYZMCC), *Petroselinum crispum* (S42523), *Rubus idaeus* (AAK15174), *Antirrhinum majus* (P06515), and *Camellia sinensis* (P48387); acridone synthase 2 from *Ruta graveolens* (Q9FSC0); bibenzyl synthase from *Phalaenopsis sp.* (S71619); coumaroyl triacetic acid synthase from *Hydrangea macrophylla* (BAA 32733); dihydropinosylvin-forming stilbene synthase from *Pinus sylvestris* (Q02323); 2PS from *Gerbera hybrida* (P48391); trihydroxystilbene synthase sequences from *Vitis vinifera* (P51070) and *Arachis hypogaea* (S00334); and valerophenone synthase (O80400) from *Humulus lupulus*.

Homology models were generated for proteins encoded by wild type *Arabidopsis CHS* and five *Arabidopsis tt4* alleles [*tt4*(85), *tt4*(UV113), *tt4*(38G1R), *tt4*(UV01), and *tt4*(UV25)] based on the crystal structure of *Medicago* CHS2 [pdb id:1BI5, 4]. Five models were produced for each protein using Modeller6 [14] and merged by coordinate averaging. Each averaged structure was then minimized using the Sander module of Amber8 [15] for 200 steps using steepest descent. The structures were placed in a box of water molecules such that no solute atom was closer than 8 Å to the edge of the box. Sodium ions were added to give the system a net charge of zero. Molecular dynamics on the water and sodium ions was first performed for 100 ps. Minimization of the solvent and counterions was performed for 200 steps, followed by a minimization of the entire system. Molecular dynamics was then performed using the Sander module for a total of 500 ps with a time step of 2.0 fs and a structure collected at every 1 ps. Although these simulations are too short to detect large conformational changes, they do provide information about the local reorientation of residues. All simulations were performed using four to eight processors on Virginia Tech's Laboratory for Advanced Scientific Computing and Applications Linux cluster. The dynamics simulations were composed of an 80 ps heating phase to raise the system temperature from 0 K to either 285 K, 293 K, or 300 K, a 120 ps constant volume equilibration phase, and a 300 ps constant pressure phase. Final models were generated by coordinate averaging structures from the last 100 ps of the dynamics simulation and minimization. Assessments of the models were performed using Procheck on the Whatif server (http://swift.embl-heidelberg.de/servers/). For individual structures, the RMSD of the C_α atoms and phi-psi torsion angles were determined over the last 100 ps of the dynamics simulations using the Carnal module of Amber7 [15]. Local structural differences between the *Arabidopsis* proteins and *Medicago* CHS2 (pdb id 1BI5) and *Gerbera* 2PS (pdb id 1QLV) were assessed by using Swiss-Pdb Viewer, version 3.7 [16] to align the structures using the least squares fit function and then calculate a RMSD. All images of protein structures were generated using SwissPDB Viewer v. 3.7 [16] and rendered with POV-Ray, version 3.1a (http://www.povray.org).

Results

Seven alleles of the *Arabidopsis* CHS locus (At5g13930), induced using either ethyl methane sulfonate or gamma radiation, were previously characterized for changes in gene expression, protein stability, and the accumulation of flavonoid end products (Table 1) [12, 17, 18]. One of these alleles, *tt4*(UV118a), contains a complex rearrangement. The other six contain point mutations that either disrupt splicing to generate a severely truncated coding region [*tt4* (2YY6)], introduce a stop codon that shortens the enzyme by 15 residues [*tt4*(38G1R)], or alter a single amino acid [*tt4*(85), *tt4*(UV01), *tt4*(UV25), and *tt4*(UV113)]. Four of these mutations completely abolish the accumulation of flavonoids in affected plants, one is leaky, and two others display a temperature-sensitive phenotype.

Table 1 Molecular and biochemical characteristics of *Arabidopsis* CHS alleles

Allele	Mutation[a]	Allele type	CHS mRNA levels[b]	CHS protein levels[c]	Flavonoid glycoside levels[d]	Dimerization[e]
wild-type	N.A.[f]	Wild-type	++++	++++	+++	+++
tt4(85)	G268S	Leaky	++++	++++	+	+++
tt4(2YY6)	3' splice site (G to A)	Null for CHS protein and flavonoid accumulation	+	–	–	N.A.
tt4(UV118a)	complex rearrangement	Null for gene, CHS protein, flavonoid accumulation	–	–	–	N.A.
tt4(UV113)	T174P	Null for flavonoid accumulation	++++	++++	–	+++
tt4(38G1R)	P381S, G382H, L383C, T384stop	Null for flavonoid accumulation	+	+	–	+
tt4(UV01)	T49I	Temperature-sensitive	++	++	+	+++
tt4(UV25)	R334C	Temperature-sensitive	++	++	+	+

[a]Amino acid(s) altered in the predicted protein
[b]Determined by RNA blot analysis [12, 17, 18]
[c]Determined by immunoblot analysis [12, 17]
[d]Determined by HPLC analysis (data not shown) [12, 17]
[e]Determined using *in situ* crosslinking with disuccinimidyl glutarate and *N*-succinimidyl-[4-vinylsulfonyl] benzoate [12]
[f]N.A., not applicable

A multiple sequence alignment of 14 representative members of the plant type III PKS superfamily, including *Arabidopsis* CHS, shows that the five point mutations that introduce either a stop codon or an amino-acid substitution [*tt4*(85), *tt4*(UV01), *tt4*(UV25), *tt4*(UV113), and *tt4*(38G1R)] all affect residues that are highly, and often absolutely, conserved in this superfamily (Fig. 1). Most are also located in highly conserved stretches of sequence. However, none alter residues known to be directly involved in substrate binding or catalysis [5, 7, 19, 20]. A homology model of the *Arabidopsis* wild-type enzyme based on the structure of *Medicago* CHS2 previously showed that only *tt4*[UV113] and *tt4*[UV25] affect residues near important functional domains in the three-dimensional structure of CHS, suggesting that there may be other, indirect effects on various aspects of enzyme function [12]. To examine this possibility further, we generated homology models of the five CHS variants containing the C-terminal truncation or

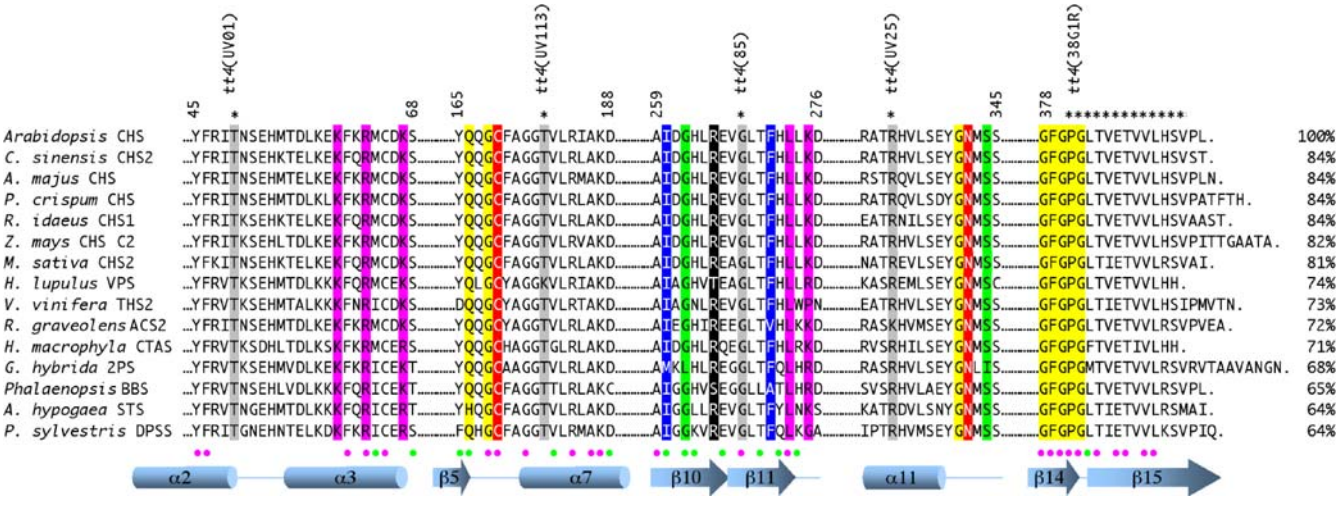

Fig. 1 Multiple sequence alignment of selected regions of a representative group of plant type III polyketide synthases. Highlighting indicates catalytic residues (red), CoA binding residues (purple), and active site residues that are structural (yellow), control polyketide length (green), or determine substrate specificity (blue), or are essential for activity of raspberry PKS1 [4, 5, 8, 22, 23, 27]. Residues that are altered in the five *Arabidopsis* enzymes are indicated with asterisks. Numbering is for the wild type *Arabidopsis* CHS protein sequence. Percentages to the right of the sequences refer to amino acid identity relative to *Arabidopsis* CHS. Pink dots below the sequences indicate residues that are absolutely conserved among type II PKSs; green dots indicate those conserved in CHS but not other type II PKSs. Elements of secondary structure in the crystal structure of *Arabidopsis* CHS are identified below the sequences. Abbreviations: acridone synthase (ACS), bibenzyl synthase (BBS), coumaroyl triacetic acid synthase (CTAS), dihydropinosylvin-forming stilbene synthase (DPS), 2-pyrone synthase (2PS), trihydroxystilbene synthase (THS), and valerophenone synthase (VPS)

amino-acid substitutions and subjected these to molecular dynamics simulations. These models were based on the structure of the CHS monomer and therefore did not allow predictions to be made regarding the position of the active site M142 residue, which is contributed by the adjacent subunit in the homodimer, or residues M1-E11, which remain largely unstructured in the monomer. However, these models did make it possible to analyze the effects of the mutations on specific residues known to be of central importance in the interaction of CHS with its substrate.

tt4(85)

The *tt4*(85) allele, the first *Arabidopsis* CHS mutation to be reported [21], encodes an enzyme in which glycine at position 268 is replaced with serine (G268S) [18]. Plants carrying this allele produce wild-type levels of CHS protein, but accumulate only trace amounts of flavonoids, even in tissues that normally produce these compounds at high levels (data not shown) [12]. Although G268 is located near the dimerization interface, the substitution in tt4(85) does not appear to affect the formation of homodimers [12]. This residue is absolutely conserved in all known type III PKSs and is located at one end of a β-sheet (β11) that helps form the cyclization pocket [4]. Interestingly, R265 in β10, an adjacent β-strand that is part of the cyclization pocket, has been shown to be essential for activity of raspberry CHS1 [22]. These residues are in linear and spatial proximity to F271, one of two "gatekeeper" residues that control entry of substrates into the active site, [4] as well as G262, a key determinant of substrate preference and polyketide chain length [23].

The RMSD of the homology model of the tt4(85) protein relative to the wild-type enzyme is 1.38 Å, indicating that the structures are quite similar overall (Table 2). This is reflected in the overlay of the C_α atoms of the two models (Fig. 2). The RMSD value calculated for the model over the last 100 ps of molecular dynamics was also comparable to that of the wild-type structure (Table 2). A number of substantial differences were found in the RMSD values for individual residues (values >2.750 Å are shown in red in Fig. 2). Several of these, particularly in β11 and surrounding residue I298 in the loop between α9 and β12, were also observed in the other tt4 protein, as well as in *Medicago* CHS2 and *Gerbera* 2PS (Fig. 2 and described below). These differences are therefore unlikely, by themselves, to account for the substantially reduced enzymatic activity resulting from the G268S substitution.

However, tt4(85) did exhibit uniquely high RMSD values relative to the wild type for residues T251 to G258 in β9, which is in close spatial proximity to the active site, for the active site C169 residue, and for M142, which contributes to the active site of the adjacent monomer [4] (Table 2). These differences were further reflected in the positioning of the side chains of the active site, gatekeeper, and cyclization-pocket residues relative to wild type (Fig. 3; Table 3). Atomic fluctuation analysis gave a value of 1.38 Å for the substituted residue vs. 0.70 Å in the wild-type protein and also indicated that this amino acid, as well as the residues of β9, the F271 gatekeeper, and several residues of the cyclization pocket, were substantially more dynamic than in the wild-type protein (Table 2 and supplementary material). These findings thus suggest that the G268S substitution in tt4(85) not only increases the mobility of functional residues to which it is in close spatial proximity, but also the mobility and positioning of critical amino-acid side chains that are located some distance away. This single amino-acid change thereby appears to impact both entry of the substrate into the active site and its subsequent conversion into product.

tt4(UV113)

The *tt4*(UV113) mutation was isolated as a second-site suppressor of the UV-hypersensitive phenotype of an *Arabidopsis* chalcone isomerase mutation, together with the three other alleles described below. This mutation results in substitution of proline for threonine at position 174 (T174P), a highly conserved amino acid located in helix α7 (Fig. 1; Fig. 3), five residues C-terminal to the active-site cysteine. As in the case of *tt4*(85), plants carrying this mutation produce wild-type levels of CHS protein that appears to dimerize normally [12]. However, no detectable flavonoids accumulate in these lines. T174 is absolutely conserved in all known CHS enzymes and most other type III PKSs, with the exception of valerophenone synthase from *Humulus lupulus*, where there is a lysine at this position (Fig. 1).

Analysis of the tt4(UV113) homology model identified only a few differences relative to the wild-type structure (Fig. 2), with an RMSD value of 1.11 for the C_α backbones (Table 2). The RMSD of the tt4(UV113) structure itself over the final 100 ps of simulations was also very close to that of the wild type (Table 2) and little difference was observed in atomic fluctuation analysis of the two structures, even in the region surrounding the substitution, which might be expected to have a local effect on the flexibility of the protein (see supplementary material). However, as in tt4(85), there was a marked effect on M142, which exhibited an RMSD value of 2.48 relative to the wild-type structure. There was also a notable change in the position of the sidechain for the F220 gatekeeper residue (Fig. 3; Table 3). Thus, it appears that the complete lack of detectable enzymatic activity for this protein is not due to local effects of the substitution on protein dynamics, but instead result from subtle changes in its structure that, as in tt4(85), do not affect the stability of the protein or its ability to dimerize, but do dramatically impact access to and activity of the catalytic center.

Table 2 Structural comparison by RMSD values of wild type and variant *Arabidopsis* CHS proteins

	27°C								20°C			12°C		
	***Arabidopsis* CHS**	***Medicago* CHS2**	***Gerbera* 2PS**	tt4 (85)	tt4 (UV113)	tt4 (38G1R)	tt4 (UV01)	tt4 (UV25)	**Wild-type**[e]	tt4 (UV01)	tt4 (UV25)	**Wild-type**[e]	tt4 (UV01)	tt4 (UV25)
RMSD vs AtCHS[a]	**N.A.**	**1.39**	**1.28**	1.38	1.11	1.35	1.29	1.11	**1.04**	1.49	1.47	**1.06**	1.66	1.31
RMSD over 100 ps[b]	**2.28 (0.09)**	**N.A.**	**N.A.**	1.94 (0.08)	2.45 (0.12)	3.58 (0.07)	2.89 (0.09)	2.11 (0.15)	**1.94 (0.05)**	2.35 (0.07)	2.06 (0.07)	**1.96 (0.06)**	2.36 (0.01)	1.92 (0.07)
Active site residues														
M142[c]	**N.A.**	**0.76**	**1.54**	*3.27*	*2.48*	*2.20*	*2.54*	*4.67*	**0.83**	1.52	*2.96*	**0.58**	*3.22*	*3.77*
	1.26	**N.A.**	**N.A.**	1.53	1.20	1.35	1.37	1.49	**1.56**	1.47	0.80	**1.14**	1.35	*2.54*
C169	**N.A.**	**0.86**	**1.36**	1.76	0.88	0.82	1.22	0.61	**1.13**	0.76	1.08	**0.31**	0.71	0.38
	0.55	**N.A.**	**N.A.**	0.96	0.65	0.94	0.89	0.85	**0.78**	0.89	0.72	**0.83**	1.06	0.70
H309	**N.A.**	**0.88**	**1.30**	1.31	0.79	0.36	0.39	0.52	**0.67**	0.67	1.33	**0.25**	0.61	0.73
	0.51	**N.A.**	**N.A.**	0.69	0.55	0.86	0.77	0.70	**0.69**	0.92	0.54	**0.79**	1.03	0.57
N342	**N.A.**	**0.97**	**0.86**	0.40	1.10	0.32	0.60	0.60	**0.57**	0.53	0.12	**0.43**	0.57	0.72
	0.57	**N.A.**	**N.A.**	0.75	0.62	0.88	0.83	0.68	**0.71**	0.95	0.55	**0.76**	1.15	0.66
Gatekeeper residues														
F220	**N.A.**	**1.24**	**0.70**	0.59	0.62	0.40	0.45	0.51	**0.55**	1.00	0.38	**0.51**	0.52	0.31
	0.62	**N.A.**	**N.A.**	0.88	0.85	0.93	0.92	0.79	**0.81**	0.89	0.69	**0.79**	1.13	1.11
F271	**N.A.**	*3.59*	*3.41*	*2.44*	*2.78*	*2.01*	1.86	*2.50*	**2.24**	0.51	1.16	***2.67***	0.80	0.96
	0.64	**N.A.**	**N.A.**	1.05	0.80	1.02	0.99	0.95	**0.84**	0.88	1.03	**0.99**	1.17	0.79
Cyclization pocket residues														
T137	**N.A.**	**0.35**	**1.01**	1.44	1.20	1.09	0.98	1.70	**1.36**	0.83	0.71	**0.61**	0.81	0.70
	0.58	**N.A.**	**N.A.**	0.94	0.72	0.89	0.76	0.90	**0.78**	0.75	0.55	**0.92**	1.09	0.69
I260	**N.A.**	**1.62**	**1.43**	1.03	1.50	0.54	1.34	1.24	**0.95**	1.14	1.82	**0.89**	*3.35*	1.18
	0.69	**N.A.**	**N.A.**	1.17	0.86	1.11	0.95	1.17	**0.78**	0.88	0.59	**0.99**	1.29	0.84
G262	**N.A.**	*2.75*	*2.60*	1.76	*2.65*	*2.25*	1.99	1.25	***2.37***	1.46	1.50	***2.92***	*2.10*	1.80
	0.59	**N.A.**	**N.A.**	1.33	0.88	0.97	0.93	0.95	**0.86**	0.81	0.77	**1.17**	1.12	1.09
P381[d]	**N.A.**	**1.29**	**1.39**	*2.31*	1.07	0.87	1.21	0.74	**1.09**	1.21	1.74	**0.86**	1.24	0.91
	0.59	**N.A.**	**N.A.**	1.20	0.76	1.10	0.90	1.18	**1.03**	0.89	0.56	**1.02**	1.12	0.80

[a]RMSD values for the C_α atoms in the averaged structures were determined using Swiss-PdbViewer v3.7. Values are for residues I12 through the C terminal L395 in the *Arabidopsis* CHS proteins (384 atoms), except for tt4(38G1R) where residues I12 through G380 were used (372 atoms), for *Medicago* CHS2 where residues I7 through I389 were used (383 atoms), and *Gerbera* 2PS where residues A18 through V394 were used (377 atoms). Comparisons were between the variant and wild type structures at the corresponding temperatures and for wild type at 20°C and 12°C relative to wild type at 27°C. All values are in angstroms. Bold characters indicate values for wild type enzymes; those in italics indicate values for individual residues of 2.00 Å or higher. N.A. = not applicable

[b]RMSD values for the C_α atoms of each structure were calculated using Carnal. Average values for the final 100 ps of molecular dynamics are given, with standard deviations in parentheses

[c]Values in the upper box for each residue are the RMSD determined as described in [a]; values in the lower box are the average atomic fluctuation over the final 100 ps of molecular dynamics simulations

[d] P381 is S in tt4(38G1R)

[e]Determined using *in situ* crosslinking with disuccinimidyl glutarate and *N*-succinimidyl-[4-vinylsulfonyl] benzoate [12]

tt4(38G1R)

The point mutation in *tt4*(38G1R) has a very different impact on CHS than is observed for *tt4*(85) and *tt4*(UV113) (Table 1). In this case the mutation shifts the reading frame of the CHS transcript starting at codon 381, introducing three altered residues and a premature stop codon so that 15 residues are lost from the C terminus. The protein accumulates at much lower levels than the wild-type CHS and appears to be catalytically inactive, apparently at least in part because it does not dimerize [12]. The first two residues altered by this mutation are part of the conserved GFGPG loop that provides a scaffold for the cyclization reaction in type III PKSs [1, 20].

The overall architecture of the C_α backbone of the modeled tt4(38G1R) structure is not dramatically different from that of the wild type protein, with an RMSD of 1.35 Å (Table 2; Fig. 3), despite the fact that the C-terminal β15 strand is entirely absent in this protein. However, there is a large local effect on the position of the adjacent β9 strand (S247 through D255). This results in a very high RMSD for the Cα atoms of these residues relative to the wild type (3.2 to 5.7 Å), as well as higher atomic fluctuation values (average 1.27 ± 0.11 Å versus

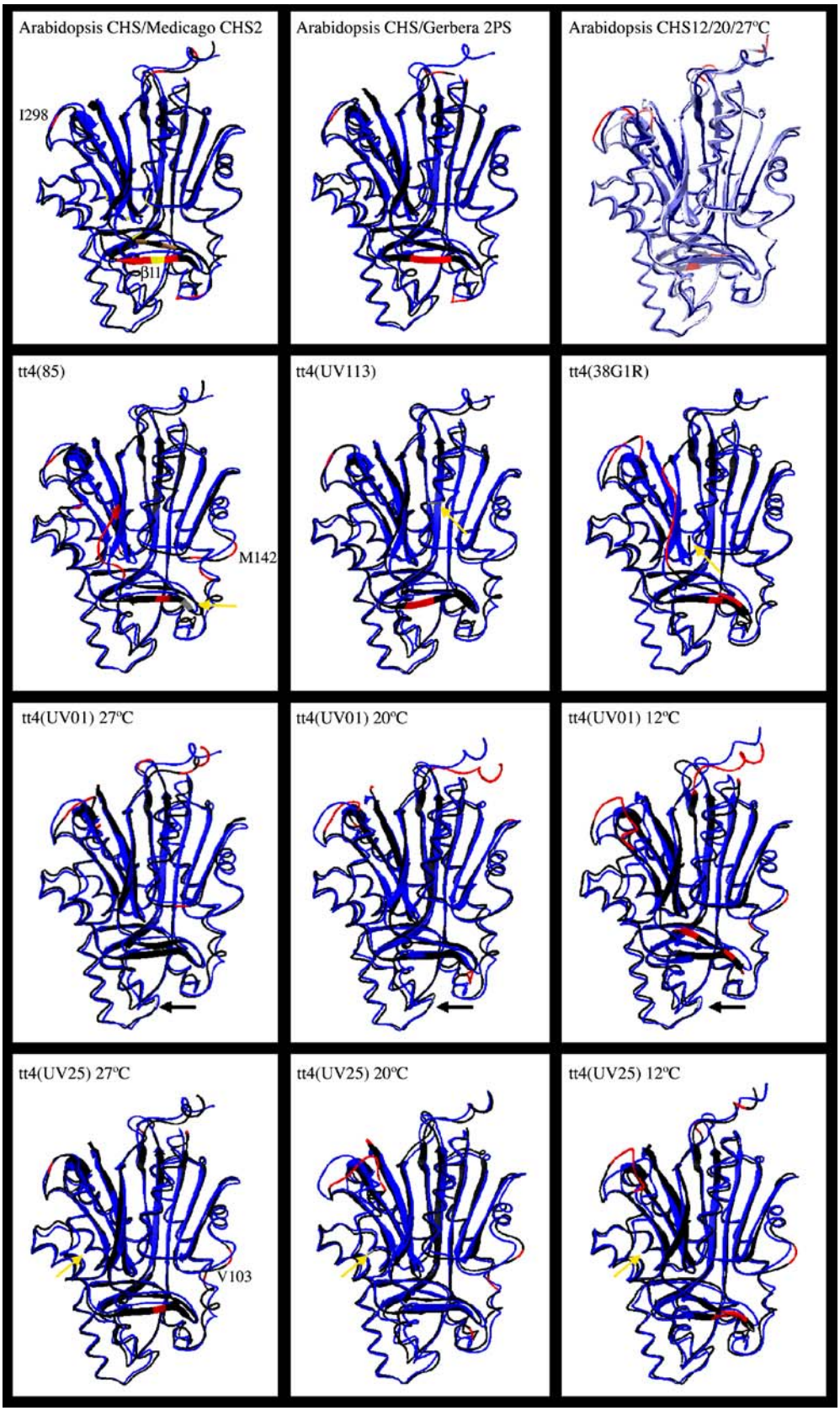

Fig. 2 Structural comparison of CHS wild-type and variant enzymes. Overlays of the C_α backbones of *Medicago* CHS2, *Gerbera* 2PS, and the *Arabidopsis* CHS variants (black) onto *Arabidopsis* wild type (blue) are shown, with regions that differ by a rms value of more than 2.75 Å shown in red. Active site and gatekeeper residues are shown in yellow and residues of the cyclization pocket are shown in brown in the *Medicago* CHS2 structure. Residues in gray, highlighted by the yellow arrows, are those that differ between the variants and wild type as a result of gene mutation. In the upper righthand panel the C_α backbone of *Arabidopsis* wild-type CHS modeled at 27°C is shown in dark blue, at 20°C in medium blue, and at 12°C in light blue

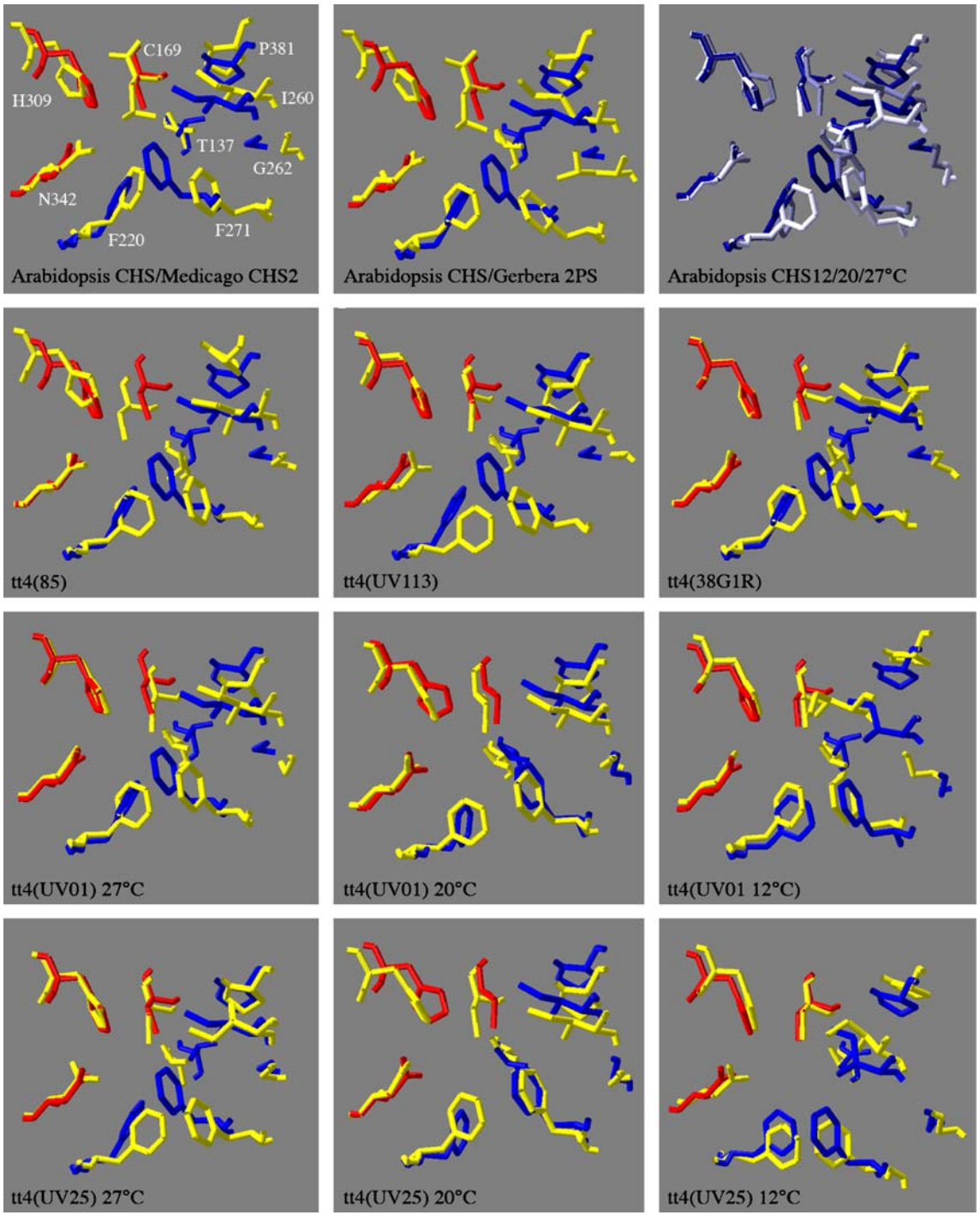

Fig. 3 Relative orientation of the sidechains of key functional residues. In the wild type, the sidechains of the catalytic triad (C169, H309, and N342) are shown in red, while those of the gatekeeper residues (F220 and F271) and other four residues of the cyclization pocket (Thr137, I260, G262, and P381) are in blue. The corresponding sidechains in *Medicago* CHS2, *Gerbera* 2PS, and the *Arabidopsis* CHS variants are shown in yellow. In the upper righthand panel residues in *Arabidopsis* wild-type CHS modeled at 27°C are shown in dark blue, at 20°C in medium blue, and at 12°C in light blue. The active site residue, M142, which is contributed by the other subunit in the homodimer, is not shown. The numbering scheme shown is for *Arabidopsis* CHS; the corresponding residues in *Medicago* CHS2 are T132, M137, C164, F215, I254, G256, F265, H303, N336, and P375 and in *Gerbera* 2PS are T137, M142, C169, F220, M259, L261, F270, H308, N341, and P380

0.71 ± 0.17 for wild type). Other critical residues exhibiting high C_α RMSD values are M142, F271, and G262, although the positions of the side chains of the remaining key functional residues do not appear to be substantially altered (Tables 2 and 3; Fig. 3). These predictions, together with the experimental evidence that

Table 3 Phi-psi torsion angles of active site and gatekeeper residues[a]

	27°C						20°C			12°C		
	Wild-type	tt4 (85)	tt4 (UV113)	tt4 (38G1R)	tt4 (UV01)	tt4 (UV25)	**Wild-type**	tt4 (UV01)	tt4 (UV25)	**Wild-type**	tt4 (UV01)	tt4 (UV25)
Active site residues												
C169 Φ	**–50.12**	–60.31	–53.06	–53.78	–50.70	–57.50	**–56.04**	–54.48	–60.65	**–57.29**	–52.32	–53.15
	(11.60)	(10.82)	(9.83)	(10.61)	(11.00)	(9.58)	**(8.43)**	(9.67)	10.43	**(10.48)**	(10.96)	9.77
C169	**–27.12**	–35.96	–35.09	–23.65	–33.80	–30.05	**–24.74**	–34.31	–21.42	**–22.64**	*–42.01*	*–38.14*
ψ	**(10.41)**	(12.90)	(10.82)	(9.55)	(11.30)	(10.74)	**(9.51)**	(12.22)	10.88	**(11.37)**	*(10.10)*	*12.04*
H309	**–72.57**	–61.36	–61.06	–71.16	–72.08	–67.08	**–71.70**	–66.77	–71.51	**–67.76**	–63.45	–65.01
Φ	**(9.96)**	(8.43)	(10.38)	(10.54)	(9.43)	(12.90)	**(12.22)**	(9.34)	11.16	**(10.84)**	(10.86)	12.24
H309	**116.30**	127.05	126.24	114.99	*152.42*	118.23	**107.40**	*149.10*	*149.35*	**113.38**	115.37	*144.73*
ψ	**(6.78)**	(6.49)	(7.67)	(7.07)	*(7.12)*	(8.16)	**(6.73)**	*(7.87)*	*8.54*	**(6.40)**	(6.65)	*11.09*
N342	**–88.90**	–78.57	–79.86	–80.16	*–153.68*	*–103.23*	**–105.41**	*–67.91*	*–65.74*	**–84.01**	–82.57	*57.39*
Φ	**(21.97)**	(8.14)	(8.41)	(15.48)	*(8.71)*	*(17.99)*	**(17.15)**	*(9.42)*	*9.72*	**(16.74)**	(8.29)	*12.70*
N342	**128.02**	*87.88*	*86.02*	*105.89*	135.46	94.73	**98.19**	*136.24*	*140.08*	**102.20**	90.25	114.93
ψ	**(16.96)**	*(7.04)*	*(8.60)*	*(11.62)*	(8.65)	(9.63)	**(13.40)**	*(14.90)*	*10.37*	**(15.07)**	(7.60)	9.59
Gatekeeper residues												
F220	**–79.21**	–75.90	*–62.59*	–77.26	*–61.70*	–57.60	**–68.23**	–61.86	–64.70	**–70.59**	–80.02	–65.60
Φ	**(10.38)**	(10.10)	*(9.25)*	(9.88)	*(9.03)*	(7.66)	**(8.95)**	(8.87)	9.79	**(10.29)**	(8.61)	11.76
F220 ψ	**148.17**	154.18	142.13	158.01	*114.63*	138.10	**141.34**	*118.89*	127.83	**151.18**	159.24	*118.34*
	(8.23)	(7.46)	(9.35)	(7.89)	*(7.40)*	(8.35)	**(8.56)**	*(7.12)*	7.19	**(7.86)**	(7.10)	*6.74*
F271 Φ	**–145.14**	*–62.07*	–150.95	*–60.03*	*–87.10*	*–73.93*	**–139.86**	*–81.48*	–89.51	**–140.35**	*–63.99*	*–80.83*
	(7.90)	*(10.74)*	(8.88)	*(8.87)*	*(11.20)*	*(19.01)*	**(11.63)**	*(8.95)*	13.77	**(8.91)**	*(9.11)*	*9.41*
F271 ψ	**119.20**	*137.79*	125.08	*142.15*	*95.34*	139.63	**155.42**	*86.60*	*98.61*	**148.25**	135.28	*87.93*
	(21.66)	*(8.81)*	(8.42)	*(7.66)*	*(8.81)*	(8.57)	**(7.39)**	*(9.39)*	*11.63*	**(8.64)**	(8.45)	*8.54*
Cyclization pocket residues												
T137 ψ	**–113.50**	–92.42	–91.72	*–80.95*	–77.55	*–116.28*	**–84.89**	–83.88	–81.22	**–99.03**	*–80.50*	–84.52
	12.39	(9.46)	(14.09)	*(12.43)*	(8.87)	*(11.15)*	**(11.87)**	(11.07)	9.94	**(15.37)**	*(10.47)*	12.28
T137 Φ	**–15.06**	*–42.03*	–29.87	–21.60	–29.05	12.03	**–28.36**	–27.43	–22.10	**–6.80**	*–25.71*	*–22.95*
	8.76	*(9.92)*	(8.79)	(9.74)	(9.40)	(9.04)	**(10.75)**	(8.71)	8.43	**(9.82)**	*(8.83)*	*8.01*
I260 ψ	**–115.92**	*–133.32*	–127.62	–107.23	*–133.16*	–129.00	**–116.32**	–122.01	*–134.84*	**–139.59**	*–79.93*	*–110.56*
	14.16	*(7.41)*	(9.66)	(16.29)	*(10.84)*	(19.36)	**(12.65)**	(11.24)	*10.11*	**(8.30)**	*(21.39)*	*17.03*
I260 Φ	**135.32**	*150.39*	*114.97*	137.30	131.11	*136.25*	**134.75**	132.90	131.33	**91.14**	83.15	*75.63*
	11.97	*(15.29)*	*(10.47)*	(12.78)	(10.09)	*(13.59)*	**(8.78)**	(11.74)	13.41	**(16.32)**	(10.14)	*20.44*
G262	**–145.09**	*–47.05*	–135.03	–152.60	*129.70*	–125.82	**–117.13**	*–76.18*	*22.92*	**134.10**	–126.17	*110.40*
ψ	**57.75**	*(161.57)*	(92.07)	(35.10)	*(36.40)*	(100.45)	**(107.79)**	*(148.43)*	*166.57*	**(16.17)**	(20.15)	*114.03*
G262	**24.93**	*46.80*	*59.34*	*65.55*	*142.38*	*23.77*	**136.11**	*55.27*	*44.83*	**70.68**	*118.12*	110.47
Φ	**13.76**	*(20.97)*	*(19.33)*	*(16.77)*	*(18.69)*	*(14.12)*	**(98.22)**	*(16.58)*	*23.16*	**(39.35)**	*(30.80)*	38.15
P381ψ[c]	**–48.11**	–66.23	–44.52	–51.73	–46.85	–48.50	**–43.25**	–44.49	–45.42	**–52.01**	–63.16	–53.61
	8.45	(9.03)	(9.45)	(8.58)	(10.23)	(9.88)	**(10.29)**	(8.26)	8.42	**(10.82)**	(9.78)	11.01
P381 Φ	**144.65**	*–22.89*	*121.98*	*112.44*	*121.66*	*145.02*	**138.53**	138.53	141.08	**126.33**	*154.07*	139.83
	9.20	*(14.14)*	*(10.95)*	*(9.10)*	*(11.90)*	*(9.37)*	**(10.52)**	(11.06)	10.09	**(40.55)**	*(9.33)*	10.21

[a]The phi angle (ϕ) indicates rotation around the N-C_α bond, while the psi angle (ψ), indicates rotation around the C_α -C bond. Positive and negative values indicate right-handed and left-handed rotation, respectively. Values calculated using Carnal were averaged for the final 100 ps of molecular dynamics and are given in degrees; standard deviations are shown in parentheses. Bold characters indicate values for the wild-type protein; those in italics indicate a substantial rotation (15° or more) relative to the wild-type

[b]P381 is S in tt4(38G1R)

the tt4(38G1R) protein is unstable and does not dimerize effectively, suggest that the major effect of the C-terminal truncation may be to interfere with proper folding of the protein and that the structure predicted by homology modeling may not occur in vivo.

tt4(UV01)

tt4(UV01) is a temperature-sensitive allele that results in the replacement of a threonine with an isoleucine at position 49 (T49I), a residue that appears to be invariant in all type III PKSs (Fig. 1). This amino-acid change does not

appear to affect the ability of the enzyme to dimerize, consistent with the fact that it is not located near the dimerization interface (Fig. 2), but does seem reduce protein stability at all temperatures [12] (Table 1). Plants that are homozygous for *tt4*(UV01) exhibit a temperature-dependent reduction in the accumulation of flavonoid end products, with almost no flavonoids present at 27°C, very low levels at 20°C (the optimum growth temperature for *Arabidopsis*), and moderate levels at 12°C [12].

The homology models of tt4(UV01) and the wild-type enzyme were subjected to molecular dynamics simulations at 27, 20, and 12°C. The wild-type protein appeared to be more different from the structure at 27°C when modeled at 12°C than at 20°C, as expected (Table 2; Fig. 2). Interestingly, the largest changes were near I298 and in β11, where the greatest differences with *Medicago* CHS2 and *Gerbera* 2PS also occur. However, the tt4(UV01) structure also exhibited an increasing RMSD with decreasing temperature, even relative to the wild-type protein under the corresponding conditions. This is in contrast to the expectation that the structures should converge as the temperature is reduced and may explain why protein stability is not enhanced at lower temperatures. This was also the case for the RMSD values of key functional residues, particularly M142 (Table 2). However, the torsion angles of several of these amino acids, particularly H309 and N342 of the catalytic triad, approached those of the wild-type structure at 12°C (Table 3; Fig. 3). This suggests that, although the tt4(UV01) protein may not recover full catalytic activity at lower temperatures, the increase in functionality observed in *tt4*(UV01) plants may be due to improved alignment of side chains in the active center of the enzyme.

tt4(UV25)

The *tt4*(UV25) allele is characterized by the replacement of arginine at position 334 with a cysteine (R334C) [12]. This mutation is located eight residues N-terminal to the active site asparagine (N342) and is highly conserved among type III PKSs. The known exceptions are two confirmed and two suspected isoforms of acridone synthase from rue [24] (accession numbers S60241, Q9FSC0, Q9FSC1, and Q9FSC2, respectively), which have conservative substitutions of lysine at this position (Fig. 1). Similar to tt4(UV01), the tt4(UV25) enzyme is present at reduced levels in mutant plants and exhibits a temperature-dependent reduction in flavonoid accumulation. However, it also appears to have lost the ability to form homodimers (Table 1).

Molecular dynamics simulations of the tt4(UV25) homology model were performed at 27, 20, and 12°C. As with tt4(UV01), there was not a convergence in overall structure with the wild type with decreasing temperature (Table 2). However, in this case there was an overall improvement in RMSD values for key functional residues at 12°C relative to 27°C (Table 2), but not in the torsion angles of these residues (Table 3; Fig. 3). Thus the outcome of this analysis is somewhat different from that observed for tt4(UV01), but still points to an improvement in the architecture of the key functional domains in this protein with decreasing temperature. Perhaps more important in this case is that the dimerization interface in tt4(UV25) appears to be more similar in structure to that of the wild-type protein at 12°C than at the higher temperatures (Fig. 2). In particular, the α5 helix exhibits a much closer fit at the lower temperature, with the RMSD for V103 going from 3.08 at 25°C, to 2.79 at 20°C, to 0.57 at 12°C. This suggests that a major factor in the recovery of activity at 12°C is due, at least in part, to improved dimerization of the protein.

Discussion

The current study indicates that homology modeling of variant proteins associated with defined physiological and biochemical phenotypes can provide new insights into structure–function relationships in enzymes such as CHS. This is the case even when the amino-acid changes caused by point mutations are not located at or near residues that are of known functional significance. Thus, it appears that the G268S substitution in tt4(85) alters the mobility and positioning of a number of key residues at diverse locations in the enzyme. In the case of tt4(UV113), contrary to expectations that insertion of a rigid proline residue adjacent to the active site would affect the dynamic character of the enzyme, the substitution instead appears to affect access to and orientation of key side chains in the active center. Molecular dynamics simulations of tt4(38G1R) underscored the importance of the C terminus as a structural component of this enzyme, while analysis of tt4(UV01) and tt4(UV25) pointed to effects on the positioning of key residues in the active site or dimerization domain, rather than global enzyme architecture, as being responsible for the temperature-sensitive phenotypes associated with these variant proteins.

The outcome of this study suggests that it will be useful to extend these analyses to account for the effects of substrate binding and interactions with other proteins. In the case of CHS, there is already good structural information for the homodimer and for the enzyme bound to substrate and to product analogs [25]; structural information is also starting to emerge for CHS complexed with other enzymes of flavonoid metabolism (Dana, Krueger, and Winkel, unpublished results). Models of the CHS variants that take this type of information into account could provide a more complete picture of the structure-function effects of amino-acid substitutions. An example of the utility of molecular modeling for assessing effects on protein–protein interactions is provided by a recent analysis of 60 human disease proteins, which showed that a disease mutation that disrupts dimerization of the adrenoleukodystrophy protein, ABCD1, affected a residue at the predicted homodimer interface [26]. As we understand more about the structural basis of protein–protein interactions, homology modeling should provide an increasingly powerful tool for characterizing protein function, including

efforts to dissect the molecular basis of both the beneficial and deleterious phenotypes resulting from genetic variation as well as rational approaches to engineering metabolism in plants and other organisms.

Acknowledgements The authors thank Jory Zmuda Ruscio for her assistance with the molecular dynamics simulations. This work was supported by the National Science Foundation (grants MCB-9808117 and MCB-0131010) and by grants from the Virginia Tech Graduate Research Development Project to C.D.D.

References

1. Austin MB, Noel JP (2003) Nat Prod Rep 20:79–110
2. Moore BS, Hertweck C, Hopke JN, Izumikawa M, Kalaitzis JA, Nilsen G, O'Hare T, Piel J, Shipley PR, Xiang L, Austin MB, Noel JP (2002) J Nat Prod 65:1956–1962
3. Winkel-Shirley B (2001) Plant Physiol 126:485–493
4. Ferrer J-L, Jez JM, Bowman ME, Dixon RA, Noel JP (1999) Nature Struc Biol 6:775–784
5. Jez JM, Ferrer J-L, Bowman ME, Dixon RA, Noel JP (2000) Biochem 39:890–902
6. Jez JM, Noel JP (2002) J Biol Chem 277:1361–1369
7. Tropf S, Kärcher B, Schröder G, Schröder J (1995) J Biol Chem 270:7922–7928
8. Jez JM, Bowman ME, Noel JP (2002) Proc Natl Acad Sci USA 99:5319–5324
9. Ceruso MA, Periole X, Weinstein H (2004) J Mol Biol 338:469–481
10. Stockner T, Sterk H, Kaptein R, Bonvin AM (2003) J Mol Biol 328:325–334
11. Sutherland JJ, O'Brien LA, Lillicrap D, Weaver DF (2004) J Mol Model 10:259–270
12. Saslowsky DE, Dana CD, Winkel-Shirley B (2000) Gene 255:127–138
13. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673–4680
14. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
15. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TA, DeBolt S, Ferguson DM, Seibel G, Kollman P (1995) Comput Phys Commun 91:1–41
16. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
17. Burbulis IE, Iacobucci M, Shirley BW (1996) Plant Cell 8:1013–1025
18. Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM (1995) Plant J 8:659–671
19. Lanz T, Tropf S, Marner FJ, Schröder J, Schröder G (1991) J Biol Chem 266:9971–9976
20. Suh D-Y, Fukuma K, Kagami J, Yamazaki Y, Shibuya M, Ebizuka Y, Sankawa U (2000) Biochem J 350:229–235
21. Koornneef M (1990) Arabid Inf Serv 28:1–4
22. Zheng D, Schröder G, Schröder J, Hrazdina G (2001) Plant Mol Biol 46:1–15
23. Jez JM, Bowman ME, Noel JP (2001) Biochem 40:14829–14838
24. Junghanns KT, Kneusel RE, Baumert A, Maier W, Groger D, Matern U (1995) Plant Mol Biol 27:681–692
25. Dana CD, Martins R, Krueger S, Winkel BSJ (2005) (submitted)
26. Zhou HX (2004) Curr Med Chem 11:539–549
27. Jez JM, Austin MB, Ferrer J-L, Bowman ME, Schröder J, Noel JP (2000) Chem Biol 7:919–930