Crystal structure of threonine synthase from *Arabidopsis thaliana*

KARINE THOMAZEAU,¹ GILLES CURIEN,² RENAUD DUMAS,² AND VALÉRIE BIOU¹

¹Institut de Biologie Structurale Jean-Pierre Ebel, Unité Mixte de Recherche 5075 Centre National de la Recherche Scientifique-Centre d'Études Nucléaires-Université Joseph Fourier, F-38027 Grenoble, France

²Unité Mixte de Recherche 1932 Centre National de la Recherche Scientifique-Institut National de la Recherche Agronomique-Aventis CropScience, F-69263 Lyon cedex 9, France

(RECEIVED October 18, 2000; FINAL REVISION December 20, 2000; ACCEPTED December 20, 2000)

Abstract

Threonine synthase (TS) is a PLP-dependent enzyme that catalyzes the last reaction in the synthesis of threonine from aspartate. In plants, the methionine pathway shares the same substrate, O-phospho-L-homoserine (OPH), and TS is activated by S-adenosyl-methionine (SAM), a downstream product of methionine synthesis. This positive allosteric effect triggered by the product of another pathway is specific to plants. The crystal structure of *Arabidopsis thaliana* apo threonine synthase was solved at 2.25 Å resolution from triclinic crystals using MAD data from the selenomethionated protein. The structure reveals a four-domain dimer with a two-stranded β -sheet arm protruding from one monomer onto the other. This domain swap could form a lever through which the allosteric effect is transmitted. The N-terminal domain (domain 1) has a unique fold and is partially disordered, whereas the structural core (domains 2 and 3) shares the functional domain of PLP enzymes of the same family. It also has similarities with SAM-dependent methyltransferases. Structure comparisons allowed us to propose potential sites for pyridoxal-phosphate and SAM binding on TS; they are close to regions that are disordered in the absence of these molecules.

Keywords: Allostery; MAD; pyridoxal phosphate; S-adenosyl-methionine; threonine synthesis

Aspartate is a precursor for the synthesis of several other amino acids, including lysine, threonine, isoleucine, and methionine. Threonine synthase (TS) (EC 4.2.99.2) is the last enzyme of the threonine synthesis pathway (Fig. 1A). It is a PLP-dependent enzyme and a 110-kD homodimer in solution. It catalyzes the conversion of O-phospho-L-homoserine (OPH) into threonine and inorganic phosphate (Fig. 1B). The bacterial reaction was studied by following the

absorbance of PLP and was shown to present at least seven steps (Laber et al. 1994).

In plants, OPH is a branch point intermediate between the methionine/S-adenosyl-methionine (SAM) and threonine/isoleucine pathways (Giovanelli et al. 1984). Thus TS competes for OPH with cystathionine- γ -synthase (CGS), the first committed enzyme in the methionine pathway (Fig. 1A). Flux regulation between the threonine and methionine pathways does not result from end-product inhibition but, rather, via a strong stimulation of TS by SAM (Giovanelli et al. 1984). Indeed, previous results showed that the allosteric binding of SAM leads to an eightfold increase in the rate of catalysis and to a 25-fold decrease in the K_M value for OPH (Curien et al. 1998). Saturation isotherms of TS by SAM disclosed positive cooperativity and indicated that at least two SAM molecules can bind on each dimeric enzyme (Curien et al. 1998). From these results, it was concluded that the dramatic modification in kinetic properties of plant TS originates from an allosteric and cooperative transition that is induced by SAM. Furthermore, this transition occurs at a

Reprint requests to: Valérie Biou, Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075 CNRS-CEA-Université Joseph Fourier, F-38027 Grenoble, France; e-mail: biou@ibs.fr; fax: 33-476-88-54-94 or Renaud Dumas, UMR 1932 CNRS-INRA-Aventis CropScience, 14-20 rue Pierre Baizet, F-69263 Lyon cedex 9, France.

Abbreviations: TS, threonine synthase; PLP, pyridoxal 5'-phosphate; OPH, O-phospho-L-homoserine; SAM, S-adenosyl-methionine; MAD, multiple wavelength anomalous dispersion; DTT, dithiothreitol; CGS, cystathionine- γ -synthase; TD, threonine deaminase; CBS, cystathionine- β -synthase; EC, enzyme classification.

Article and publication are at www.proteinscience.org/cgi/doi/10.1110/ps.44301.

Fig. 1. Reaction catalyzed by TS. (A) pathway from O-phosphohomoserine to threonine, methionine, and SAM. Enzyme names are written in italics, and compounds are written in bold. The thick arrow indicates the activation of TS by SAM. (B) reaction catalyzed by TS; below the reaction, activator SAM; above, cofactor pyridoxal phosphate.

much faster rate in the presence of the substrate OPH. In contrast with the plant enzyme, bacterial TS is not regulated by SAM, presumably because, in bacteria, homoserine (instead of OPH) is the branch-point metabolite between the threonine and methionine pathways. The specificity of plant TS activation by SAM is of great interest because, to our knowledge, only one other enzyme, the mammalian cystathionine-β-synthase (CBS) has been reported to be allosterically activated by SAM (Roper and Kraus 1992). In addition, TS is one of a few enzymes that are cross-activated by the product of another pathway (another example being chorismate mutase, which is activated by tryptophan (Schmidheini et al. 1990).

TS shows weak sequence homologies (9% identity) with threonine deaminase (TD) and tryptophan synthase, two other PLP proteins of known structure (Gallagher et al. 1998). These three enzymes belong to the fold type II family of PLP enzymes (Grishin et al. 1995; Schneider et al. 2000).

We describe in this article the first structure of a plant TS with the aim of determining the structural basis of the activation of plant TS by SAM.

Results and discussion

Oxidized and reduced MAD electron density maps

As explained in Materials and Methods, two MAD data sets were collected. Table 1 gives statistics about data collection. The fluorescence scan from nonreduced crystals showed a very strong white line for selenium (15 electrons for f'' max; data not shown). This indicated that the selenium atoms were probably oxidized in spite of the crystals being set up with 5 mM DTT and all solutions used being degassed. The fluorescence spectrum from reduced crystals was then more typical, with a white line showing 6.6 electrons for f'' max. Table 2 gives phasing statistics about the oxidized data set, which was used for structure refinement.

Table 1. MAD data collections

Unit cell	Oxidized MAD data set			Reduced MAD data set		
	57.8Å, 62.1Å, 76.6Å, 109.5°, 97.6°, 112.7°			57.8Å, 62.2Å, 76.8Å, 109.5°, 97.5°, 112.8°		
Wavelength (Å)	0.9788 (peak)	0.9789 (edge)	0.886 (remote)	0.9795 (peak)	0.9797 (edge)	0.9729 (remote)
Resolution (Å)	20 – 2.20	20 – 2.20	20 – 2.25	20 – 2.53	20 – 2.53	20 – 2.50
No. of reflections	172124 (24511)	170855 (23454)	124846 (17896)	105840 (6600)	106729 (6626)	108563 (9293)
No. of unique reflections	43625 (6233)	43254 (5954)	43719 (6265)	28359 (3129)	28358 (3223)	29154 (3953)
Completeness	97 (94.8)	96.6 (93.5)	97.2 (95.3)	95.7 (85.6)	96.3 (88.2)	95.8 (88.8)
Multiplicity	3.9 (3.9)	4.0 (3.9)	2.9 (2.9)	3.7 (2.1)	3.7 (2.1)	3.7 (2.4)
R-sym*	.042 (.256)	.036 (.184)	.033 (.183)	.053 (.151)	.050 (.100)	.046 (.113)
R-ano [#]	.074 (.192)	.050 (.129)	.042 (.139)	.073 (.133)	.044 (.109)	.047 (.112)
Anomalous completeness	96.6 (94.5)	96.4 (93.3)	95.7 (93.8)	90.5 (49.7)	93.5 (68.1)	92.6 (71.2)

Oxidized, data measured from nontreated crystal; Reduced, data measured from crystal soaked with DTT (see Materials and Methods). The numbers between parentheses indicate the value in the outer resolution shell. *R-sym = $\sum_j |I_j - \langle I \rangle| / \sum_j \langle I \rangle$ where I_j is the intensity of the j th observation and $\langle I \rangle$ is the average intensity over all observations.

$$^{\#}\text{R-ano} = \frac{|\bar{I}_+ - \bar{I}_-|}{\Sigma(\bar{I}_+ + \bar{I}_-)}.$$

Table 2. MAD phasing statistics using SHARP for the oxidized data set: *R* Cullis, *R* Kraut, phasing power

	ISO 11	ANO 11	ISO 12	ANO 12	ISO 13	ANO 13
<i>R</i> Cullis*		0.74	0.69	0.56	0.40	0.77
<i>R</i> Kraut [#]	0.019	0.042	0.033	0.063	0.024	0.047
Phasing power		2.1	3.6	2.9	4.5	2.4

11, remote wavelength used as reference; 12, peak wavelength; 13, inflexion point. ISO, isomorphous differences; ANO, anomalous differences. The overall figure of merit was 0.67 to 2.2 Å resolution. * R Cullis = $\sum |F_{PH} \pm F_P| - |F_{PH(calc)}| / \sum |F_{PH} - F_P|$ [#]*R* Kraut = $\sum |F_{PH} \pm F_P| - |F_{PH(calc)}| / \sum |F_{PH}|$.

A comparison of the raw MAD electron density maps of both data sets at the selenium positions allowed us to detect oxidized selenomethionines. Obvious differences were observed for residue 309 in both monomers, and smaller ones were observed for others (residues 322 and 391 in both monomers). Selenomethionines 309 from both monomers were interpreted as having an oxidized selenium, that is, with an oxygen atom double bonded in addition to the

methyl. The oxygen atom hydrogen binds to a water molecule, which shows that a polar atom is located at this position.

Structure of the monomer

Figure 2 describes the domain arrangement of TS. The monomer can be split into four structural regions: Domain 1 (residues 10–132) is mostly helical and contains disordered areas; domain 2 (residues 162–271) has a four-stranded parallel β -sheet with connecting helices. The strand topology is “3 2 1 4” (β 8– β 5– β 6– β 7). Residue 163 is the PLP-binding lysine. Domain 3 is not sequential; it is made up of residues 133–161 and 272–457. It is composed of a mixed β -sheet, the first two strands being antiparallel and the other four parallel with a topology “1 2 6 3 4 5” (β 3– β 4– β 12– β 9– β 10– β 11) and ends with two parallel helices connected by a β -strand. This domain is followed by a connecting helix, named “neck” in the threonine deaminase structure (Gallagher et al. 1998). Finally, the “Swap domain” is made up of

Fig. 2. Domain arrangement of TS. (A) Scheme for the secondary structure of TS. Program DSSP (Hooft et al. 1996) was used for secondary structure assignment. (B) Stereo view of the TS monomer. Domains are colored in the following way: Blue, domain 1; pink, domain 2; black, domain 3; green, Swap domain. Panel B was made with program Molscript (Kraulis 1991).

residues 458–486, which interact with the second monomer. It is mostly helical. The whole monomer has a globular core formed by domains 2 and 3, and two protruding ends with domain 1 and Swap domain (Fig. 2B).

Aromatic residues are particularly abundant in TS: 9 Trp, 14 Tyr, and 22 Phe out of 480 residues, making up 9% of all residues, whereas the average abundance of those residues in proteins is 6.2% (Schultz and Schirmer 1979). This is not the case for the structurally related enzymes TD and TS, which are poorer in tryptophans. Some aromatic–aromatic contacts stabilize long-distance interactions, whereas others are important for loop stabilization with respect to other secondary structures.

Structure of the dimer

The triclinic asymmetric unit contains two monomers. This dimer is most probably the functional dimer. It shows extensive interactions (4243 Å² out of a total accessible surface of 16582 Å² per monomer), and no other crystal contact is so extensive. Moreover, two other proteins of the same family have the same dimerization scheme (see PLP enzymes section). Figure 3 shows two views of the dimer. One view presents a deep canyon (Fig. 3A) that contains the PLP-binding lysine 163. The whole molecule has a compact ellipsoidal shape, $\sim 80 \times 60 \times 50$ Å³. The dimer interactions comprise 31 H bonds and important hydrophobic contacts that involve conserved residues. Aromatic residues play an important role in the dimer stabilization: Trp 108, Phe 127, and Tyr 103 are clustered together with Trp

101 that is perpendicular to Trp 135 from the other monomer. Other aromatic residues are at the dimer interface, such as Phe 400 or Tyr 478, which is stacked with Phe 233 from the other monomer. Phe 468 is stacked with Arg 249 from a symmetry-related molecule. Table 3 gives the root mean square (rms) deviation between atoms from the two monomers in the asymmetric unit.

Structure comparison with three PLP enzymes of the second fold type

Pyridoxal phosphate enzymes that act on amino acid substrates have been split into five fold types (Grishin et al. 1995; Schneider et al. 2000; Mehta and Christen 2000). TS is classified as part of the second type (fold II), with other enzymes that catalyze ligation reactions involving the C β carbon. TS is the only member of the family that catalyzes a β – γ replacement. The structures of three enzymes of this group have already been solved: Tryptophan synthase β subunit (e.g., 1UBS; Hyde et al. 1988), TD (1TDJ; Gallagher et al. 1998), and o-acetylserine sulfhydrylase, OAS (1OAS; Burkhard et al. 1998). The three structures were superimposed to TS monomer using program LSQMAN brute force alignment tool with a 2-Å distance cutoff (Kleywegt and Jones 1998). Table 4 gives the rms deviations and the number of aligned residues. The most similar structure to TS in terms of number of matched residues and identity among those residues is OAS. Figure 4A shows a superposition of the four structures as monomers and Figure 4B shows a sequence alignment based on the structure align-

Fig. 3. TS dimer. Two perpendicular views are shown. One monomer is colored in blue and the other in yellow. View A has the same orientation as Fig. 2B. The figure was made using Molscript and Raster 3D (Merritt and Bacon 1997).

Table 3. Refinement statistics and root mean square deviation between non crystallographic symmetry—related atoms

Resolution (Å)	30.–2.25	
R-factor*	22.3	
R-free [#]	24.3	
Total number of atoms	2789	
No. of protein atoms	6901	
No. of solvent atoms	388	
R.m.s. deviation		
Bond lengths (Å)	0.007	
Bond angles (°)	1.33	
Dihedral angles (°)	21.9	
Improper angles (°)	0.87	
Average B factor (Å ²)		
For all atoms	46.92	
For monomer A	46.18	
For monomer B	45.45	
R.m.s. ΔB for bonded atoms (Å ²)	1.28	
R.m.s.d. (Å)	All atoms	Side chains
Monomer	0.3 (0.2)	0.5
Domain 1	0.7 (0.5)	0.8
Domain 2	0.5 (0.06)	0.7
Domain 3	0.2 (0.1)	0.3
Domain Swap	0.07 (0.04)	0.08

The figures in parentheses indicate the values for Ca only. [#]The R-free was calculated on 5% of reflections left apart during the whole refinement.

$$*R\text{-factor} = \frac{\sum ||F_o| - k|F_c||}{\sum |F_o|},$$

where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes for a given reflection.

ment. All four proteins share domains 2 and 3. For the sake of simplicity, the oligomerization is not shown on Figure 5A. The dimer interface structure is rather similar between TS, Trp synthase, and OAS; however, TS has the largest hidden surface area and the highest proportion of nonpolar interactions (71%). The domain-swapping scheme is unique to TS. In fact, the first six residues of OAS interact with the other domain, whereas it is the C-terminal end of TS that is swapped. The most compact structure is OAS. The other three have additional regions. TD has a C-terminal regula-

Table 4. Statistics concerning structure alignment of TS with TD, tryptophan synthase beta subunit, and OAS

Second molecule (Mol2) ...	OAS	Trp synthase	TD
Nr of residues in TS	432	432	432
Nr of residues in Mol2	315	389	494
Nr of matched residues	135	129	114
Nr of identical residues	32	27	23
% identical of matched	23.7	20.9	20.2
% matched of TS	31.2	29.9	26.4
% identical of TS	7.4	6.2	5.3
% matched of Mol2	42.9	33.2	23.1
% identical of Mol2	10.2	6.9	4.6
Rms distance between matched Cα atoms (Å)	1.1	1.1	1.0

The superposition was done using program LSQMAN (see text).

Fig. 4. (Panels B and C appear on the following pages.) Structure superposition, sequence alignment between PLP enzymes (TS, TD, Trp synthase, and O-acetylserine sulfhydrylase) and TSs. (A) structure superposition. TS is drawn in magenta, TD in yellow, tryptophan synthase (subunits A and B) in cyan, and O-acetylserine sulfhydrylase in dark blue. Only one monomer was shown for each structure. Domains are indicated as well as the position of lysine 163, the PLP-binding lysine in TS. The figure was made using Molscript and Raster 3D. (B) sequence alignment based on the structure superposition. The TS secondary structure is shown on the first line; the second line shows TS residue numbering; the fifth line shows TD residue numbering and the last line shows the domain composition using the same coloring scheme as in Fig. 2. Residue conservation is coded in red with the following scheme: Red boxes indicate fully conserved residues, and red prints in black boxes indicate conservative modifications. Light blue boxes show residues that are involved in PLP binding (see text). (C) Sequence alignment between tTSs. *Arabidopsis thaliana* (arath), *Solanum tuberosum* (solen), *Escherichia coli* (ecoli), *Pseudomonas aeruginosa* (pseae), *Saccharomyces cerevisiae* (yeast), *Corynebacterium glutamicum* (corgl). Sequences were grouped into plant (first two lines) and micro-organism (last four lines). Red boxes show totally conserved residues, red letters show conservation inside groups, and yellow boxes show conservative replacements. Green boxes show regions with homology between plant TS and mammalian CBS, which also binds SAM (see text). The alignment figures were made using program ESPRIT (Gouet et al. 1999).

tory domain that is also involved in dimerization, tryptophan synthase has its α subunit (regulatory subunit) that points to a different direction, and TS has both its N- and C-terminal extensions (Fig. 4A). By analogy to the other enzymes, those extensions could be involved in regulation too. Very recently, Yao et al. (2000) published the structure of aminocyclopropane carboxylate deaminase, which has a similar fold to domains 2 and 3 of TS and is not allosteric.

PLP-binding site

PLP was shown to be bound on the enzyme with a stoichiometric ratio before crystallization as determined both by absorption measurements in alkaline conditions (Curien et

Fig. 4. (Continued.)

al. 1998) and by mass spectroscopy (data not shown). However, TS crystals are not yellow, which suggests that TS has lost its PLP during crystallization. In addition, we have never seen any PLP density in the active site pocket on crystals (both from native or selenomethionine protein) soaked with PLP solutions before data collection. An inten-

sive search for stabilizing conditions was performed using free radical scavengers, covalent binding using NaBH₄, soaks with substrate analog, data collections on different beamlines, and home diffractometer. As mentioned above, the electron density for the PLP-binding lysine 163 is truncated in the TS structure. We think X-ray radiation damage

C

```
1 THRC_ARATH ..ADGNHIKAPIETAVKPPHRTEDNIRDEARRNRNSHAVNPFSAIVYVFNAPGSTESYSLDEIVYRSRSG
THRC_SOLEN FTPIKATATNDIAISQQKRRPADENIREARRHCSSH..NFSAIVYVFNAPGSTESYSLDEIVYRSRSG
2 THRC_ECOLI ..LYNLKPDHN.EQVSFAQAVTQGLGKNQ
THRC_PSEAE ..YISTRGQA.PALNFEDVLLAGLASDG
THRC_YEAST ..NASQVYRSTRSSSPKTISFEAIIQGLATDG
THRC_CORGL ..YISTRDASRTFARFSDILLGGLAFDGG

1 THRC_ARATH T70 80 90 100 110 120 130
GLLDVQHDMEALKRFDDGAYWRDLFDSRVGKSTWPGYSGVSKKEWVLPFIDDDDIVSAFEG...NSNLF
THRC_SOLEN GLLDVQHDMDALKKFDGQYWRDLFDSRVGKTTWPGYSGVSKKEWVLPFIDSDDIVSAFEG...NSNLF
2 THRC_ECOLI GLFFHDLPEFSITRID.EMLKLDFTV...RSAKILSAFIDDEIPQEILEERVRA...AFA
THRC_PSEAE GLYVVEMLPRFTLEIA.SWVGLPYHE...LAFVRMFPFVAG.SIADADFKKILEET...YGVFA
THRC_YEAST GLFIFPTIPQVDQATLPNDWSKLSFDQ...LAPATMRPLYIAQRIIPDADLKLKRS.YSTFRSDE
THRC_CORGL GLYLFATYFQLDDAQLS.KWREVLANEG...YALAREVTSLEVD...DIPVEDIKAITARAY...TYPKF

1 THRC_ARATH 140 150 160 170 180 190
WAERFGKQFLGMNDLWVKHCGISHGSPKDLGNTVLVSQVNLRLKMKRP...VVGVGCASTGDT.
THRC_SOLEN WAERFGKQFLGMDTLWVKHCGISHGSPKDLGNTVLVSQVNLRLKMKRP...VVGVGCASTGDT.
2 THRC_ECOLI PP...APVANVESDVGCLELPHGPTLAFKDFGGRFMAQMLTHIAGDK...PVITLTATSGDTG
THRC_PSEAE DASGAAPVERRTMG.CVLELPHGPTLAFKDFALQLLGRLLDHVLAKRGE...RVVIMGATSGDTG
THRC_YEAST VTPLVQNVTCDEKMLHILELPHGPTLAFKDFALQFVGNLFLEYFLQRTNANLPEGEKKQITVVGATSGDTG
THRC_CORGL NSEDIVPVTELEDNIYILGHLSEGGCAAPKDNAMQLLGELFEYELRRNE...TINILGATSGDTG

1 THRC_ARATH 200 210 220 230 240 250
SAALSAYCASAGIPISIVFLANKISMAQLVQPIANG...AFVLSIDTDFDGCMLKIREITAEPL...
THRC_SOLEN SAALSAYCASAGIPISIVFLANKISMAQLVQPIANG...AFVLSIDTDFDGCMLKIREITAEPL...
2 THRC_ECOLI AAVAHAFYGLPHVVKVILYRGKISPLQEKLPCTLCG.NIETVAIDGDFDGCQALVKQAFDDEELKVALG
THRC_PSEAE SAAIAAAVADNVDFIIDHNVRVSEVRRQMTIHGDNIHNTAIEGNFDDQEMVKASPADGGFLKGTR
THRC_YEAST SAAIYGLRGKDVSVFIFYTGRISPIQEQMTVPDENVQTLSTVGTDFDGCQDIVKAIKPGDFEKNKHNN
THRC_CORGL SSAEYAMRGREGIRVPMLEAGRMTPPQQAQMPGLDDPNIFNTALDGVFDDCQDVVKAVSADAEFKKDNR

1 THRC_ARATH 260 270 280 290 300 310 320
IYLANSLNLSRLLEGQKTAIILQDFDW...VQPDWVIVEGGNLGNIYAFYKGFQMKCEILGLVDRIPLMV
THRC_SOLEN IYLANSLNLSRLLEGQKTAIILQDFDW...VQPDWVIVEGGNLGNIYAFYKGFQMKCEILGLVDRIPLMV
2 THRC_ECOLI LNSANSINISRLLAQICYFEEAVALPQET.RNQLVVSFVSGNFGDLTAGLLGLGLPVKRF.IAATNV
THRC_PSEAE LVAVNSINWASILMAQIVVYFHAALQLGAP..HRSVAPSVFTGNFGDIFAGYLANMGLPVSQQL.IVARNR
THRC_YEAST VGAVNSINWASILMAQIVVYFSPFQATNGKSKKKVSVESGNFGDILAGYFANKMGLPIRKL.AIATNR
THRC_CORGL IGAVNSINWASILMAQIVVYFVSWIRTTTSEN.DQKVSFSEATGNFGDICAGHIAKMGMLPIDRL.IVATNE

1 THRC_ARATH 330 340 350 360 370 380 390
CAQAANANPLYLHYKSGWVFKKMTASTTFASAIQIGPVSIDRAVVALKKCNQIVVEATEEELNDAMA.
THRC_SOLEN CAQAANANPLYLHYKSGWVFKKVKANTTFASAIQIGPVSIDRAVVALQCCNQIVVEATEEELNDAMA.
2 THRC_ECOLI NDTVPRFLHDG...QWSPKATQATLSNAMSVSPNNMWRVEELFRKRIWQLKELGYAAVDDE
THRC_PSEAE NDILHRP.MSG...NASTRHTLTPVSPSPMDIKVSNFPERLLFDLHGRNGKVAZELLDPAFKAS
THRC_YEAST NDILRFLKSG...LYERSDKVAATLSPAMDILISNFERLLWYLAREYLANGDGLKAGSIVN
THRC_CORGL NDVLEDFFRTGDY...RVRSSADTHESSPSMDISRAANPERFIFD.LGRDATRVNDLFGTQVRO

1 THRC_ARATH 400 410
QADSTG...MFICPSTGVALTALPKL
THRC_SOLEN QADSTG...MFICPSTGVALTALPKL
2 THRC_ECOLI TTQQTMRRLKELG...YTSECPAAVAYRALRDQ
THRC_PSEAE KGLSVEDQWRTEARKLFDSLAVSDEQTCETIAEVYRSSGE...LLDPHTAIGVRAAREK
THRC_YEAST NWQFLKLTNGKFOVDK.SIEGASKDPTSERVSNETSSTIKKIYESSVNPKHYILDPHTAIGVCAAREK
THRC_CORGL GGFSLADDANFERKAAAEYGFASGRSTHADRVATIADVHSRLD...VLIDPHTADGVHVARQW

1 THRC_ARATH 420 430 440 450 460 470
RNQGVIAFTDATVVVSTAHGLKFPQSKIDY...HSNAIPDMACRFSNPPVDVKADFGAVMDVLK
THRC_SOLEN RNSGVIAFTDATVVVSTAHGLKFPQSKIDY...HSKEIKDMCRFANPPVEVKADFGSVMDVLK
2 THRC_ECOLI LNPGE...YGLFLGTAHPAKFKESVEAI...LGETLDELKELAEADLPLLSHNL.ADFALALR
THRC_PSEAE RRSLS...VPMVTLGTAHPAKFPEAVEKAG...IGQAPALFAHLADLFEEREECTVLP.NELAKVQ
THRC_YEAST IAKDND.KSIQYISLS.TAHPAKFADAVNNALSGFSNYSFEKDVLPBELKKLS.TLKKKKLKFIBRADVELVK
THRC_CORGL RDEVN...TPIIVLE.TALPVKFACTIVEA...IGKAPQTPERFAAIMDAFFKVSVDLP.NDTDAVK

1 THRC_ARATH 480
SYLGSNTLTS...
THRC_SOLEN SYLLSQNSKL...
2 THRC_ECOLI KLMMKHQ...
THRC_PSEAE AFVSQHGNGPKPL
THRC_YEAST NAIBEELAKMKL
THRC_CORGL QYIVDAANTSTSVK
```

Fig. 4. (Continued.)

is responsible for this truncation, as shown on an electron density map calculated on partial (70% complete) data measured from a crystal that had been soaked in PLP. The electron density map showed a complete lysine 163 residue (but still no PLP), whereas this side chain became truncated when later data were added. It is easy to locate the PLP site by comparison with the other three structures. They are similar to each other in this region. Apart from the disordered area (loop 347–363), this region in TS has the same fold as in the other enzymes. This site is located at the interface between domains 2 and 3, and the free volume allows for the binding of a PLP molecule in each monomer, even though a few local rearrangements are necessary.

The electron density is very clear in this region and shows that the PLP absence has consequences on the orientation of several amino acids that are involved in interactions with the PLP in the other fold-II proteins. Asp 164 has a dual conformation and the side chain of Lys 163 points away from the PLP C4 atom with which it should make a Schiff base interaction. Asp 194 is in position to interact with the PLP ring atom O3. The residues hydrogen-binding PLP N1 nitrogen are Ser 272 in OAS aligned to His 404 in TS, whereas it is Ser 377 for Trp synthase and Ser 315 in TD (see Fig. 4B). Thr 432 in TS is in an equivalent position and is more likely to make contact with N1 atom. Moreover, the phosphate-binding loop (residues 296–299) has a conformation different from those of the structures with the PLP bound. It protrudes inside the PLP site so that this site is slightly too small to accommodate PLP. In addition, a loop between residues 347 and 363 is disordered. The corresponding loop is ordered in the other PLP-binding proteins of the same family (these are residues 215–234 in OAS, 225–247 in TD and 274–309 in Trp synthase), and its tip comes above the PLP ring. The importance of this loop ordering will be discussed further below. Therefore, we conclude that the PLP is absent from the crystals rather than disordered.

Conserved residues

The knowledge of the TS structure enabled us to make two types of sequence comparisons. The first one is the alignment of threonine synthase THRC gene products (TSs) and the second one is the alignment of the PLP proteins of the TS family, as mentioned above. The THRC alignment was performed almost automatically using program CLUSTALW, whereas the PLP family alignment relied on the structure superposition described above. In effect, the sequence alignments that used automatic programs gave results that were very different from structure superposition. This may explain our failure to solve the structure using molecular replacement. Figure 4C shows a sequence alignment of both known plant TS sequences (*Arabidopsis thaliana* and *Solanum tuberosum*) with sequences from mi-

croorganisms. Both plant sequences are very homologous, whereas microorganisms show more variations. However, totally conserved residues have obvious structural roles. For example, Lys 163 is H bonded to Thr 195 in the absence of PLP, whereas Asp 194 is H bonded to Lys 437 and is in position to make a hydrogen bond with PLP. In addition, Asn 264 interacts with Arg 267, which in turn interacts with Gln 271. Those residues are located on helix $\alpha 13$ opposite the phosphate loop (residues 294–297) and are therefore important in the active site.

Potential SAM-binding site

SAM is a widespread cofactor for methyltransferases and as many as 131 structures of methyltransferases were found in the protein databank. Among those, many were complexed either with SAM or with S-adenosyl-cysteine. Several structures were superimposed to the TS structure using program LSQMAN brute force option (Kleywegt and Jones 1998). Those that superposed best were cytosine-specific DNA methyltransferase HhaI (pdb code 2HMY; Cheng et al. 1993), adenine-M6 DNA methyltransferase M.TaqI (pdb code 2ADM; Labahn et al. 1994), and catechol methyltransferase (pdb code 1VID; Vidgren et al. 1994). The superposition involves TS domain 2 ($\beta 8$ – $\beta 5$ – $\alpha 8$ – $\beta 6$ – $\alpha 10$ – $\beta 7$ – $\alpha 11$), residues 162–261, which overlap with part of the methyltransferase SAM-binding site (Schluckebier et al. 1995). It is a mixed β sheet whose N-terminal part is mostly involved in SAM binding, whereas the C-terminal part is rather involved in DNA binding and catalysis. Further to this finding, we tried to determine whether this fold could be part of the SAM-binding site in TS. Figure 5A–D shows the structure comparison in this region between the methyltransferases and TS, with a SAM-binding mode proposed by comparison with the TaqI structure. The closeness in structure is visible, even though helix $\alpha 11$ is longer in TS than in the other structures. However, it does not interfere with SAM as bound in the methyltransferases.

Several arguments are in favor of the proposed site. First, the superposition of TS with the above-mentioned structures brings the SAM site in a region of TS where there is space for it to bind and in the vicinity of the PLP site as discussed above (but not overlapping, so that there is space for both ligands; see Fig. 5E). Second, a search for proteins containing this motif using program DALI (Holm and Sander 1999) yielded 305 protein databank entry codes, including the methyltransferases, all of which were nucleotide-binding proteins. Third, the motif is open on one side because of the disorder in two regions: The N-terminal region around residue 65 and the 344–364 loop that closes the active site in other PLP enzymes. All the other complexes of known structure with SAM involve proteins for which SAM is a methyl donor and is therefore involved in the reaction. In the TS case, SAM is not part of the reaction, which may

Fig. 5. Potential SAM-binding site. Structure comparison between the SAM-binding site in methyltransferases and a homologous site in TS. (A) TS. SAM is positioned in the TS figure (A) according to its position in TaqI. Part A shows residues 455–475 of the other TS monomer colored in orange. The β -strand is H bonded with β 7 strand. (B) HHAI; (C) Catechol methyl transferase; (D) TaqI. See text for details. Secondary structure elements coloring (as named in Fig. 2A): Red, α 7– β 5; blue, α 8– β 6; green, α 10– β 7; yellow, α 11– β 8. (E) Close up on the active site cleft of TS dimer oriented and colored as in Fig. 3A, with ligands positioned as described in the text. SAM is colored pink and PLP is in black. The figure was made using Molscript.

explain that the C-terminal part of the SAM-binding fold is missing and could be replaced in the allosteric conformational change by the disordered regions. Fourth, the other fold-II PLP enzymes usually have different amino acids and often have different sizes for the loops corresponding to those involved in SAM binding in the methyltransferases.

Finally, amino acid sequence comparison with mammalian CBS activated by SAM is also in favor of this argu-

ment. Interestingly, the sequence comparison between *Arabidopsis* TS and rat CBS points out region 341–346 (Fig. 4C, green boxes), at the edge of the disordered area. This sequence is conserved between both proteins and corresponds to an insertion with respect to TS sequences from microorganisms. Another such conserved sequence region is 41–46, that comes close to 341–346 in the structure and is also close to disordered residues 64–67. Again, this region has a different sequence in plants than in microorganisms TS. As mentioned above, loop 348–362 is well ordered in the structures with the PLP present in its binding site. Ours is the first structure of a PLP protein without its PLP bound and it is possible that the ordering of this loop is in relation with the presence of the PLP. Region 341–346 is well ordered and has an orientation that is different from that of the ordered loops. It is possible that SAM binding in the vicinity changes the orientation of region 341–346, thereby triggering the loop ordering and increasing the affinity for the substrate. Two other regions show distinct homologies with CBS: Residues 274–281 are located on helix α 14, which ends close to the active site, and residues 423–426 are part of the dimer interface opposite residues 16–17 from the other monomer.

Possible role for domain swapping

Figure 5A shows the position of swapped domain 458–486 in TS. Residues 462–466 form a slightly distorted β -like structure that prolongs the sheet by parallel hydrogen bonding to β 7 and might be part of the lever involved in the allosteric change. In addition, this domain is near the disordered N-terminal region that might become ordered on cofactor binding.

The crystallization conditions for the apo enzyme are different from those obtained by Laber et al. (1999) who crystallized it as a complex with SAM. This is not surprising, as soaking our crystals with SAM makes them crack. Therefore, the crystal contacts are probably disrupted on allosteric activation.

Conclusion

Plant TS has a unique feature among TSs: It shares its substrate OPH with the methionine synthesis pathway, and it is cross-activated by a downstream product of methionine. This allows for an increase in threonine synthesis to counterbalance the lower affinity for OPH of this enzyme with respect to CGS. In addition to its interest as one of the few enzymes that are cross-activated by the product of another pathway, it has potential applications as a target for herbicides. Indeed, the fact that the plant enzyme has unique features makes it a good target for specific inactivation. Sequence comparisons with other TSs show a unique N-

terminal extension, an insertion in the 340 area, and a unique C terminus.

We present here the first structure of a TS. The homodimer shows four main domains, which are termed domains 1, 2, 3, and Swap domains. Domain 1 is partially disordered, domain 2 is a mixed α - β domain with a nucleotide-binding fold, domain 3 interacts with the domain 2 at the PLP site, and the Swap domain interacts with the domain 2 β -sheet of the other monomer. Structure comparisons with PLP enzymes of the same family point toward the PLP-binding site that is partly disordered in our structure, in particular the loop between residues 347 and 363. This is next to the unique region 341–346. Structure comparisons with SAM enzymes point to a potential binding site close to the disordered areas and to the swapped C-terminal domain. This swapped domain as well as the N-terminal domain could be important for the allosteric movement. Indeed, a deletion mutant that lacks the first 36 amino acids was shown to behave as if it were activated by a single SAM molecule (Curien et al. 1996), and another mutant that lacks the first 77 amino acids had an activity that is comparable to that of the nonactivated enzyme but could not be activated by SAM (Laber et al. 1999). The allosteric movement induced by SAM binding could promote organization of the PLP loop and thus enhance substrate binding. Further work is needed to study the interaction with SAM and to characterize the allosteric movement.

Materials and methods

Selenomethionine protein cloning, purification, and crystallization

The native protein cloning and purification were described elsewhere (Curien et al. 1996). Numbering of the plant TS amino acids was performed from the first residue of the mature protein (i.e., position 39 in reference [Curien et al. 1996] becomes position 1).

The changes that are necessary for selenomethionine protein production and crystallization have been described elsewhere (K. Thomazeau et al., in prep.). The crystallization was performed using the hanging drop method, with reservoirs containing 1 M LiCl, 13% PEG 6000, 5 mM DTT, and MES-KOH pH 6.5 buffer. The drops were made by mixing 1.5 μ L of protein solution at 5 mg/mL and 1.5 μ L of reservoir solution. The crystals developed over a week period at 20°C to a size of ~0.1 mm in each dimension. The crystals were triclinic with a C2 pseudosymmetry. The unit cell was 57.8 Å, 62.1 Å, 76.6 Å, 109.5°, 97.6°, 112.7°.

Data collection

All data were measured on flash-cooled crystals at 100 K in a solution containing 10% glycerol in addition to the crystallization reservoir. 360° were measured for each wavelength. Data were collected at the European Synchrotron Radiation Facility (Grenoble, France). Two MAD data sets were collected on the selenomethionine enzyme. The first one (termed “oxidized” in the text) was collected from a nonsoaked crystal on beamline BM14 to

2.2 Å resolution using a MarResearch CCD detector. The fluorescence scan measured on the same crystal showed a strong white line (15 electrons f'' and absorption edge shift), which indicates that the selenium was oxidized. A second MAD data set (termed “reduced”) was later measured on beamline BM30 on a crystal that had been previously transferred into a solution containing 5 mM of freshly dissolved DTT. Here, the fluorescence scan gave a white line only 6.6 electrons high, showing that the selenium had been reduced. To speed up data collection, the small area of the MAR345 imaging plate detector was used (180 mm diameter). The detector edge was at 2.5 Å resolution.

Both data sets allowed structure solution. The reflections were integrated using program Mosflm (Leslie 1990) and scaled with program Scala (Evans 1997), using the remote wavelength as a reference data set and calculating scale factors on the complete data set before splitting the individual wavelengths.

Phasing and structure solution

The selenium sites were extracted one by one, starting from a first site at (0, 0, 0). Thus, 26 out of 28 sites were found. The missing two sites correspond to a disordered loop in both monomers. Sites were refined using program MLPHARE (Otwinowski 1991) and later using program SHARP (La Fortelle et al. 1997). Table 2 gives phasing statistics for the oxidized data set. After density modification using twofold noncrystallographic symmetry with program DM (Cowtan 1999), the initial electron density map allowed to trace 80% of residues (389 out of 486).

Final model

The final model contains residues 10–27, 34–64, 69–347, 363–486 for monomer A, residues 17–27, 36–62, 69–347, 363–479 for monomer B, and 388 water molecules. Missing residues were not built because the electron density was absent in both the 2Fo-Fc and the Fo-Fc maps. An electrophoresis gel run on dissolved crystals showed that the protein is not degraded; therefore, the missing density is attributed to disorder in the crystal. Selenomethionine residues 309 of both monomers were modeled with an additional oxygen atom. This addition resulted in a lower R-free and a reduction in the residual density in the Fo-Fc map. Lysine 163 that binds PLP has an incomplete side chain. In the map calculated from the MAD data set used to refine the structure, it is truncated after C ϵ , therefore the N ζ atom is missing.

The Ramachandran plot produced by PROCHECK shows 88.1% of all the residues in the most favored regions, 11.7% in additional allowed regions and 0.3% generously allowed regions.

Refinement

The oxidized remote wavelength data set was used for structure refinement because of its weaker anomalous signal and its higher resolution (2.25 Å). The structure refinement was performed using CNS version 0.9 (Brunger et al. 1998) and the maximum-likelihood “mlhl” target function (including experimental phases). During the whole process, 5% of the reflections were kept apart for R-free calculation. SIGMAA weighted electron density maps were calculated with CNS. Model modifications were performed using program O (Jones et al. 1991).

After initial model building, the first steps of crystallographic refinement consisted of a rigid body refinement at a resolution of 4 Å, with two rigid blocks per monomer. The R-factor was 35.3%

and the free R-factor was 35.2%. A correction of the geometry and a simulated annealing refinement followed at a resolution of 2.25 Å, with strict noncrystallographic constraints between the two monomers. Then atomic positions and grouped temperature factors were refined during several cycles to reach an R-factor of 30.3% and an R-free of 33.0%.

For the following refinement steps, strict noncrystallographic constraints were replaced by harmonic noncrystallographic restraints and refinement of individual temperature factors was introduced. Anisotropic global temperature factor and bulk solvent corrections were used. Water molecules were also added at each cycle of refinement, using the WATERPICK routine in CNS. The final steps of refinement consisted of progressively loosening the harmonic restraints from a weight of 300 kcal·mole⁻¹·Å⁻² to 100 kcal·mole⁻¹·Å⁻². The final R-factor is 22.5% and free R-factor 24.8%. Refinement statistics are summarized in Table 3.

The structure factors of the remote wavelength oxidized data set and the atomic coordinates have been deposited in the RSCB protein databank under accession numbers r1e5xsf and 1e5x, respectively.

Acknowledgments

We thank Andy Thompson and Vivian Stojanoff for their help in using beamline BM14 at the European Synchrotron Radiation Facility in Grenoble, Richard Kahn for his help in using beamline BM30, Laurent Chantalat for useful discussion. We acknowledge support for use of the ESRF. K.T. receives a thesis grant from the Ministère de l'Éducation Nationale et de la Recherche.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Brunker, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **54**: 905–921.
- Burkhard, P., Rao, G.S., Hohenester, E., Schnackerz, K.D., Cook, P.F., and Jansonius, J.N. 1998. Three-dimensional structure of O-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.* **283**: 121–133.
- Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., and Roberts, R.J. 1993. Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74**: 299–307.
- Cowan, K. 1999. Error estimation and bias correction in phase-improvement calculations. *Acta Crystallogr. D. Biol. Crystallogr.* **55**: 1555–1567.
- Curien, G., Dumas, R., Ravanel, S., and Douce, R. 1996. Characterization of an *Arabidopsis thaliana* cDNA encoding an S-adenosylmethionine-sensitive threonine synthase. Threonine synthase from higher plants. *FEBS Lett.* **390**: 85–90.
- Curien, G., Job, D., Douce, R., and Dumas, R. 1998. Allosteric activation of *Arabidopsis* threonine synthase by S-adenosylmethionine. *Biochemistry* **37**: 13212–13221.
- Evans, P.R. 1997. Scaling of MAD data. In *Proceedings of the CCP4 study weekend* (eds. K.S. Wilson, G. Davies, A.W. Ashton, and S. Bailey), pp. 97–102. Daresbury Laboratory, Warrington.
- Gallagher, D.T., Gilliland, G.L., Xiao, G., Zondlo, J., Fisher, K.E., Chinchilla, D., and Eisenstein, E. 1998. Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase. *Structure* **6**: 465–475.
- Giovanelli, J., Veluthambi, K., Thompson, G.A., Mudd, S.H., and Datko, A.H. 1984. *Plant Physiology* **76**: 285–292.
- Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. 1999. ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**: 305–308.
- Grishin, N.V., Phillips, M.A., and Goldsmith, E.J. 1995. Modeling of the spatial structure of eukaryotic ornithine decarboxylases. *Protein Sci.* **4**: 1291–1304.
- Holm, L. and Sander, C. 1999. Protein folds and families: Sequence and structure alignments. *Nucleic Acids Res.* **27**: 244–247.
- Hoof, R.W., Sander, C., Scharf, M., and Vriend, G. 1996. The PDBFINDER database: A summary of PDB, DSSP and HSSP information with added value. *Comput. Appl. Biosci.* **12**: 525–529.
- Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W., and Davies, D.R. 1988. Three-dimensional structure of the tryptophan synthase α 2 β 2 multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.* **263**: 17857–17871.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**: 110–119.
- Kleywegt, G.J. and Jones, T.A. 1998. Databases in protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **54**: 1119–1131.
- Kraulis, P.J. 1991. Molscript, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**: 946–950.
- La Fortelle, E. de, Irwin, J.J., and Bricogne, G. 1997. SHARP: A maximum-likelihood heavy-atom parameter refinement and phasing program for the MIR and MAD methods. In *Crystallographic computing* (eds. P. Bourne and K. Watenpaugh), pp. 250–262. IUCR, Oxford, UK.
- Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I., and Saenger, W. 1994. Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. *Proc. Natl. Acad. Sci.* **91**: 10957–10961.
- Laber, B., Gerbling, K.P., Harde, C., Neff, K.H., Nordhoff, E., and Pohlenz, H.D. 1994. Mechanisms of interaction of *Escherichia coli* threonine synthase with substrates and inhibitors. *Biochemistry* **33**: 3413–3423.
- Laber, B., Maurer, W., Hanke, C., Grafe, S., Ehlert, S., Messerschmidt, A., and Clausen, T. 1999. Characterization of recombinant *Arabidopsis thaliana* threonine synthase. *Eur. J. Biochem.* **263**: 212–221.
- Leslie, A.G.W. 1990. Molecular data processing. In *Crystallographic computing 5 From chemistry to biology*. (eds. D. Moras, A.D. Podjarny, and J.-C. Thierry), pp. 50–61. Oxford Science Publications, New York.
- Mehta, P.K. and Christen, P. 2000. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* **74**: 129–184.
- Merritt, E.A. and Bacon, D.J. 1997. Raster3D photorealistic molecular graphics. In *Methods in enzymology*. (eds. C. Carter and R.M. Sweet), pp. 505–524. Academic Press, Orlando, FL.
- Otwinski, Z. 1991. Maximum likelihood phasing. In *Proceedings of the CCP4 study weekends*, pp. 80–86 Daresbury Laboratory, Warrington, UK.
- Roper, M.D. and Kraus, J.P. 1992. Rat cystathionine β -synthase: Expression of four alternatively spliced isoforms in transfected cultured cells. *Arch. Biochem. Biophys.* **298**: 514–521.
- Schluckebier, G., O'Gara, M., Saenger, W., and Cheng, X. 1995. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**: 16–20.
- Schmidheini, T., Mosch, H.U., Evans, J.N., and Braus, G. 1990. Yeast allosteric chorismate mutase is locked in the activated state by a single amino acid substitution. *Biochemistry* **29**: 3660–3668.
- Schneider, G., Käck, H., and Lindqvist Y. 2000. The manifold of vitamin B6 dependent enzymes. *Structure* **8**: R1–R6.
- Schultz, G.E. and Schirmer, R.H. 1979. *Principles of protein structure*. Springer-Verlag, New York.
- Vidgren, J., Svensson, L.A., and Liljas, A. 1994. Crystal structure of catechol O-methyltransferase. *Nature* **368**: 354–358.
- Yao, M., Ose, T., Sugimoto, H., Horiuchi, A., Nakagawa, A., Wakatsuki, S., Yokoi, D., Murakami, T., Honma, M., and Tanaka, I. 2000. Crystal structure of 1-aminocyclopropane-1-carboxylate deaminase from *hansenula saturnus*. *J. Biol. Chem.* **275**: 34557–34565.