doi:10.1016/j.jmb.2008.08.075 J. Mol. Biol. (2009) 386, 37–59

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

A Mechanistic Model of the Cysteine Synthase Complex

Anna Feldman-Salit[1,2]*, Markus Wirtz[3], Ruediger Hell[3] and Rebecca C. Wade[1]*

[1]*Molecular and Cell Modeling Group, EML Research, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany*

[2]*Interdisciplinary Center for Scientific Computing, Faculty of Biosciences, University of Heidelberg, 69120 Heidelberg, Germany*

[3]*Heidelberg Institute for Plant Science, University of Heidelberg, 69120 Heidelberg, Germany*

Received 4 June 2008; received in revised form 7 August 2008; accepted 20 August 2008
Available online 5 September 2008

Plants and bacteria assimilate and incorporate inorganic sulfur into organic compounds such as the amino acid cysteine. Cysteine biosynthesis involves a bienzyme complex, the cysteine synthase (CS) complex. The CS complex is composed of the enzymes serine acetyl transferase (SAT) and *O*-acetyl-serine-(thiol)-lyase (OAS-TL). Although it is experimentally known that formation of the CS complex influences cysteine production, the exact biological function of the CS complex, the mechanism of reciprocal regulation of the constituent enzymes and the structure of the complex are still poorly understood. Here, we used docking techniques to construct a model of the CS complex from mitochondrial *Arabidopsis thaliana*. The three-dimensional structures of the enzymes were modeled by comparative techniques. The C-termini of SAT, missing in the template structures but crucial for CS formation, were modeled *de novo*. Diffusional encounter complexes of SAT and OAS-TL were generated by rigid-body Brownian dynamics simulation. By incorporating experimental constraints during Brownian dynamics simulation, we identified complexes consistent with experiments. Selected encounter complexes were refined by molecular dynamics simulation to generate structures of bound complexes. We found that although a stoichiometric ratio of six OAS-TL dimers to one SAT hexamer in the CS complex is geometrically possible, binding energy calculations suggest that, consistent with experiments, a ratio of only two OAS-TL dimers to one SAT hexamer is more likely. Computational mutagenesis of residues in OAS-TL that are experimentally significant for CS formation hindered the association of the enzymes due to a less-favorable electrostatic binding free energy. Since the enzymes from *A. thaliana* were expressed in *Escherichia coli*, the cross-species binding of SAT and OAS-TL from *E. coli* and *A. thaliana* was explored. The results showed that reduced cysteine production might be due to a cross-binding of *A. thaliana* OAS-TL with *E. coli* SAT. The proposed models of the enzymes and their complexes provide mechanistic insights into CS complexation.

Keywords: cysteine biosynthesis; comparative modeling; protein–protein docking; cysteine synthase

Edited by D. Case

**Corresponding authors.* Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany. E-mail addresses: anna.feldman-salit@ eml-r.villa-bosch.de; rebecca.wade@eml-r.villa-bosch.de.

Abbreviations used: CS, cysteine synthase; SAT, serine acetyl transferase; OASS/OAS-TLC, *O*-acetyl-serine sulfhydrylase; CoA, coenzyme A; SAT3, serine acyl transferase; PLP, pyridoxal 5′-phosphate; C-SDA, constrained simulation diffusional association; F-SDA, free SDA.

Introduction

The investigation of the structure and dynamics of macromolecular interactions is of great importance for understanding how macromolecules function within the cell, how they recognize one another and how they contribute to cellular processes. Experimental techniques, such as the yeast two-hybrid system,[1,2] surface plasmon resonance,[3] atomic force microscopy[4] and isothermal titration calorimetry,[5] are widely used to identify possible binding partners and characterize protein complexes. X-ray crystallography,[6] NMR spectroscopy[6] and electron microscopy[7] can provide detailed information on the structures of proteins and their complexes. However, the experimental structure determination of protein complexes is challenging, and there are many protein complexes for which structural information is lacking. In this context, structure-based computational techniques provide a means to predict the interactions of proteins and their mode of recognition.

The aim of this work was to derive a model for the cysteine synthase (CS) complex by using computational techniques in conjunction with experimental data. The CS complex is a multiprotein complex that catalyzes cysteine biosynthesis in *Arabidopsis thaliana* mitochondria.

Plants and bacteria can assimilate and incorporate inorganic sulfur into essential organic compounds such as the amino acid cysteine. Cysteine constitutes a metabolic entrance of sulfur into cell metabolism, where it is required for the biosynthesis of essential components. Cysteine biosynthesis in plants and bacteria proceeds via a two-step pathway, shown schematically in Fig. 1a. The catalytic reaction includes acetylation of the amino acid serine and production of cysteine (Fig. 1b). The reactions are catalyzed by the enzymes serine acetyl transferase (SAT) in the first step and *O*-acetyl-serine-(thiol)-lyase (OAS-TL) in the second step.[9,10] In the literature, the second enzyme is also named *O*-acetyl-serine sulfhydrylase (OASS). Cysteine biosynthesis involves a bienzyme complex, termed the CS complex, which is composed of homohexameric SAT and homodimeric OAS-TL enzymes.[11–18] Crystallographic data for the individual structures of SAT and OAS-TL from bacteria and OAS-TL from *A. thaliana* cytosol are available, but no structure of a plant SAT or the CS complex is available.[19–29] A detailed study of protein–protein interactions revealed that the C-terminus of SAT plays a crucial role in the association of SAT and OAS-TL.[12,13,17,18] Crystal structures of OAS-TL with the last four C-terminal residues of SAT from *Haemophilus influenzae* and the last eight residues from *A. thaliana* cytosolic isoform have been obtained.[25,28] The existent experimental evidence on the formation of the CS complex has been triggered by its special properties, although its precise function remains unclear. Based on biochemical evidence, a regulatory model proposes a role in sulfide sensing to maintain sulfur homeostasis in plant cells.[30] Sulfide and OAS are known to stabilize and disisociate, respectively, the complex in plants and bacteria.[11,31,32] CS complex formation has a dissociation constant K_d of about 25 nM, is reversible and is prevented by OAS with a half-maximal dissociation rate at 50–80 μM OAS according to surface plasmon resonance studies.[15] The two enzymes show opposite activities inside and outside the complex: SAT is inactivated or less active when released from the complex, whereas OAS-TL is highly active as a free dimer but almost completely inactive when complexed with SAT.[11,17,31] Consequently, the intermediate OAS leaves the complex and is not channeled between the subunits as might be expected in a metabolic protein complex. Instead, OAS has been shown to activate transcription of *Sultr* genes and a whole spectrum of sulfate starvation response genes in feeding experiments.[33]

The cellular concentrations of OAS and sulfide have therefore been assumed to drive the equilibrium of CS complex association and dissociation.[34] Sulfate deficiency results in a decrease of sulfide levels and an accumulation of OAS. This leads to dissociation of the CS complex, inactivation of SAT and a halting of OAS formation from acetyl-coenzyme A (CoA). Accumulated OAS triggers transcription of high-affinity sulfate transporter genes to improve sulfate uptake. New cellular sulfate is reduced and consumed for cysteine synthesis by active OAS-TL dimers until OAS levels drop. Then, the CS complex reassociates and SAT resumes activity. This is a feature unique to the plant CS complex.[31] In contrast, free SAT in bacteria has the same enzymatic kinetic properties as when it is in the CS complex.[11,16] Thus, the plant CS complex serves the cellular homeostasis of cysteine by sensing levels of sulfide from reductive assimilation. This mechanism is unique as a sensing and regulatory system and obviously distinguishes plant cysteine synthesis from bacteria.

In this study, we applied established protein structure modeling methods[35] and a validated protein–protein docking technique[36] to model the CS complex from *A. thaliana* mitochondria. The three-dimensional structures of serine acyl transferase (SAT3) and O-acetyl-serine-(thiol)-lyase (OAS-TLC) from *A. thaliana* mitochondria were modeled by comparative techniques. SAT3 and OAS-TLC stand, respectively, for the isoforms of SAT (EC 2.3.1.30) and OAS-TL (EC 2.5.1.47) in *A. thaliana* mitochondria.[37,38] Diffusional encounter complexes of SAT3 and OAS-TLC were generated by Brownian dynamics (BD) simulation. Incorporation of experimental constraints during BD simulation allowed identification of complexes consistent with experiments. The C-terminus of SAT3, missing in the experimental homologous template structures, was modeled *de novo*. Selected encounter complexes were then refined with molecular dynamics (MD) simulation to generate tighter bound complexes. Since it is not clear at which stage the CS complex forms, we modeled complexes in the presence and in the absence of CoA. These models reflect different

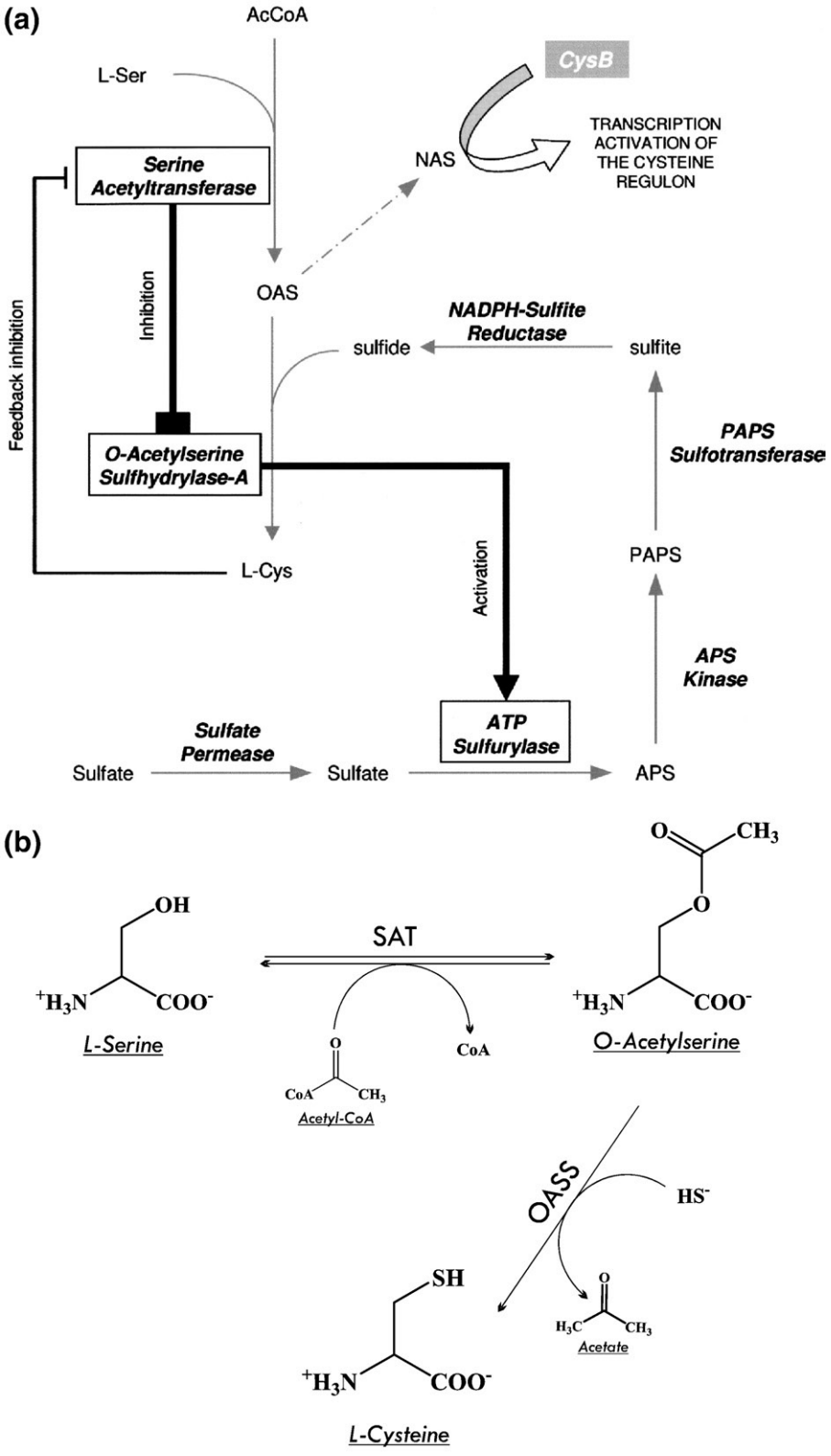

Fig. 1. Biosynthetic pathway for cysteine in bacteria and plants. (a) General overview adapted from Kredich.[8] Black thick lines connect proteins that can form bienzymatic complexes. OAS, O-acetyl-serine; NAS, *N*-acetyl-serine; APS, adenosine 5′-phosphosulfate; PAPS, 3′-phosphoadenosine 5′-phosphosulfate. CysB is a regulatory protein controlling the transcription of the cysteine regulon. (b) Reactions catalyzed by SAT and OASS to produce cysteine from serine.

possible mechanisms of complexation. We also address the question of the stoichiometric ratio of SAT3 and OAS-TLC in the CS complex, comparing the experimental results with our models. Experimentally, based on size-exclusion chromatography data, the CS complex is composed of one SAT

hexamer and two OAS-TL dimers.[11,15,16] Pye *at al.*[29] envisaged higher efficiency of the complex if six OAS-TL dimers would bind to one SAT hexamer. Here, we modeled a complex of six OAS-TLC dimers and one SAT3 hexamer in *A. thaliana* mitochondria. However, the binding free energy calculations for the modeled CS complexes suggest that a ratio of only two OAS-TLC dimers to one SAT3 hexamer is more likely, which is in agreement with experiments. The effect of mutations in the SAT-binding loop[27] of OAS-TLC was computationally tested by evaluating the change in electrostatic interaction energy during BD simulation. The mutation of residues that are experimentally important for the CS formation resulted in hindrance of the complexation of OAS-TLC and SAT3 due to a less favorable electrostatic binding free energy. The docking procedure for CS complex modeling was tested with a reference docking test for unfavorable binding of an isoenzyme of OASS (encoded by the gene *CysM*) that has been shown experimentally not to form a complex with a cognate SAT. We also explored the potential for cross-species binding of SAT and OAS-TL from *Escherichia coli* and *A. thaliana*, since the enzymes from *A. thaliana* were experimentally expressed in *E. coli*.

Results and Discussion

Comparative modeling

Template selection

To define the atomic structures of SAT3 and OAS-TLC from *A. thaliana* mitochondria, we applied the comparative modeling protocol (see Methods).

SAT3. SAT3 from *A. thaliana* mitochondria belongs to the family of SATs (EC 2.3.1.30) and has a sequence length of 391 amino acid residues. The pairwise sequence alignment demonstrated a high percentage of sequence similarity between five potential templates (Table 1) and SAT3. Because of the differences in sequence length between the bacterial and plant enzymes, the 123 N-terminal residues of SAT3 (31% of the sequence) were not aligned. Since there is no experimental evidence for the importance of this N-terminal region in CS formation, we did not attempt to model the structure of the N-terminus but did consider the possible effects of the N-terminus on CS formation. Using multiple-sequence alignment for bacterial and plant SATs, we found no significant sequence similarity in the primary structure of the C-terminal regions (8% of the SAT3 sequence). However, since, according to experiments, the C-terminal region plays a significant role in CS formation, we modeled its structure in this work. The remaining 61% of the SAT3 sequence (124–360) can be modeled by comparison with any of the potential templates (Table 1). A pairwise sequence alignment of SAT3 with 1s80 is shown in Fig. 2.

There are differences in the crystallographic features of the potential templates (Table 1). The structures 1s80, 1ssm and 1sst from *H. influenzae* are about 90% complete, with 10% from the C-terminal end missing. The structures 1ssq from *H. influenzae* and 1t3d from *E. coli* exhibit more structural information about the C-terminus. The pairwise sequence alignment of 1t3d and 1ssq to SAT3 reveals that 14 residues are relevant for more complete modeling of SAT3 in its C-terminal tail region. With a gap of 5 amino acid residues, these residues provide a possibility to model SAT3's residues, from 361 to 379 (GGKDNPKTHDKIPGLTMDQ). The pairwise alignment of the C-terminus of 1t3d with SAT3 is shown in Fig. 2.

For more precise comparative modeling, we decided that the apoenzyme with the highest sequence similarity to the target and the most complete quaternary structure would be a good choice for the template. Among the given potential templates, these requirements are fulfilled by combining two structures, 1s80 and 1t3d (or 1ssq). 1s80 is an apoenzyme that shows a high sequence identity with SAT3, while 1t3d (or 1ssq) provides a partially resolved C-terminus. In this work, the quaternary structure of SAT3 is reported with the structures 1s80 and 1t3d used to make a chimeric template.

OAS-TLC. OAS-TLC from *A. thaliana* mitochondria belongs to the family of *O*-acetyl-serine sulfhydrylases (EC 2.5.1.47) with a sequence length of 330 amino acid residues. Nine sequences of *O*-acetyl-serine sulfhydrylases (OAS-TLs or OASSs) with known three-dimensional structure were selected as potential templates for modeling a structure of OAS-TLC (see Table 1). The first six enzymes belong to the bacterial kingdom, while the last three (1z7w, 1z7y and 2isq) are from *A. thaliana* cytosol. The three-dimensional structure of 1z7y differs from that of the apoenzyme 1z7w ($AtSAT^{cyt}$) by the presence of the product analog in the active site and by the mutation of the active-site lysine to alanine (K46A).[27] The details of 1z7y's structure are intentionally missing in the table, since, according to Bonner *et al.*,[27] the backbones of structures 1z7w and unliganded 1z7y are nearly identical with an RMSD of 0.37 Å.

The results of pairwise sequence alignments show a high percentage of sequence similarity for all enzymes. The highest sequence identity (69%) is for 1z7w from *A. thaliana* cytosol. The alignment of 1z7w to OAS-TLC is presented in Fig. 2. Six percent of the missing residues in OAS-TLC are comprised by 12 residues of the N-terminus and 8 C-terminal residues. Although the missing termini could be significant for CS formation, no structural reference for their comparative modeling exists. Thus, considering the completeness of the crystal structures (Table 1) and the highest level of identity with the OAS-TLC primary structure, 1z7w was selected as the main template for modeling OAS-TLC.

Table 1. Potential templates and their characteristics for comparative modeling of SAT3 and OAS-TLC from *A. thaliana* mitochondria

Serine acyl transferases									
PDB ID	Organism	Resolution (Å)	Structure features	Seq. length[a]	Completeness of the structure[a]				Overall sequence identity to SAT3 (%)
					Resolved (residues / approx. %)		Missing (residue / approx. %)		
1s80[19]	*H. influenzae*	2.70		267	2–240	90	1, 241–267	10	51
1ssm[20]	*H. influenzae*	2.15	Truncated	242	1–240	99	241– 242	1	51
1ssq[20]	*H. influenzae*	1.85	Complex with Cys	267	Short chain: 1–241 Long chain: 1–257	90	242–267	10	51
1sst[20]	*H. influenzae*	2.00	Complex with CoA	267	1–181, 189–240	87	182–188, 241–267	13	51
1t3d[21]	*E. coli*	2.20	Complex with Cys	273	1–262	96	263–273	4	49

O-acetyl-serine-sulfhydrylases									
PDB ID	Organism	Resolution (Å)	Structure features[c]	Seq. length[a]	Completeness of the structure[a]				Overall sequence identity to OAS-TLC (%)
					Resolved (residues / approx. %)		Missing (residue / approx. %)		
1oas[22]	*S. typhimurium*	2.20		322	1–315	98	316–322	2	49
1fcj[23]	*S. typhimurium*	2.00		322	1–302	94	303–322	6	49
1ve1[d]	*Thermus thermophilus*	1.45		304	1–302	99	303–304	1	47
1o58[24]	*Thermotoga maritima*	1.80	Tetramer with phosphates	291	1–291	100	–	–	44
1y7l[25]	*H. influenzae*	1.55	Complex with terminal decapeptide of HiSAT[b]	316	1–311	98	312–316	2	47
2bht[26]	*E. coli*	2.00		303	1–294	97	295–303	3	35
1z7w[27]	*A. thaliana, cytosol*	2.20		322	3–322	99	1–2	1	69
2isq[28]	*A. thaliana, cytosol*	2.80	Complex with terminal decapeptide of AtSAT[b]	322	1–322	100	–	–	68

[a]The given properties are per monomer.
[b]Decapeptide here refers to the last 10 residues of SAT from *H. influenzae* (1y7l) and from *A. thaliana* (2isq).
[c]All crystal structures of OAS-TL contain a PLP cofactor, except for 1o58, where PO_4 is used to mimic PLP at its binding site. The shaded areas indicate the templates chosen for comparative modeling (e.g., 1s80 and 1t3d for SAT3 and 1z7w for OAS-TLC).
[d]The structure was solved by Goto in 2004, but no paper is available yet.

SAT3 vs. 1s80

```
SAT3C    124  DVWAKIREEAKSDIAKEPIVSAYYHASIVSQRSLEAALANTLSVKLSNLNLPSNTLFDLF  183
              DVW  IR+EAK     EP +++++H++I+  ++L  AL+  L+ KL+N   P+ +L ++
1s80     4    DVWQHIRQEAKELAENEPXLASFFHSTILKHQNLGGALSYLLANKLANPIXPAISLREII  63

SAT3C    184  SGVLQGNPDIVESVKLDLLAVKERDPACISYVHCFLHFKGFLACQAHRIAHELWTQDRKI  243
                  Q NP I++    D+ AV+ RDPA   +    L+ KGF A Q++RI H LW Q+RK
1s80     64   EEAYQSNPSIIDCAACDIQAVRHRDPAVELWSTPLLYLKGFHAIQSYRITHYLWNQNRKS  123

SAT3C    244  LALLIQNRVSEAFAVDFHPGAKIGTGILLDHATAIVIGETAVVGNNVSILHNVTLGGTGK  303
              LAL +QN++S AF VD HP AKIG GI  DHAT IV+GET+V+ N+VSIL  VTLGGTGK
1s80     124  LALYLQNQISVAFDVDIHPAAKIGHGIXFDHATGIVVGETSVIENDVSILQGVTLGGTGK  183

SAT3C    304  QCGDRHPKIGDGVLIGAGTCILGNITIGEGAKIGAGSVVLKDVPPRTTAVGNPARLL  360
              + GDRHPK+ +GV IGAG  ILGNI +G+ AKIGA SVVL  VP   TA G PAR++
1s80     184  ESGDRHPKVREGVXIGAGAKILGNIEVGKYAKIGANSVVLNPVPEYATAAGVPARIV  240

Score =  239 bits (611),  Expect = 2e-61
Identities = 122/237 (51%), Positives = 163/237 (68%), Gaps = 0/237 (0%)
```

SAT3 vs. 1t3d (C-terminal alignment)

```
SAT3C    361  GGKDNPKTHDKIPGLTMDQ----TSHISEWSD  388
              G  D+ K     P    DQ     +H  E+ D
1t3d     245  GKPDSDK-----PSXDXDQHFNGINHTFEYGD  271
```

OAS-TLC vs. 1z7w

```
OAS-TLC  13   IADNVSQLIGKTPMVYLNSIAKGCVANIAAKLEIMEPCCSVKDRIGYSMVTDAEQKGFIS  72
              IA +V++LIG TP+VYLN++A+GCV  +AAKLE+MEPC SVKDRIG+SM++DAE+KG I
1z7w     5    IAKDVTELIGNTPLVYLNNVAEGCVGRVAAKLEMMEPCSSVKDRIGFSMISDAEKKGLIK  64

OAS-TLC  73   PGKSVLVEPTSGNTGIGLAFIAS-RGYRLILTMPASMSMERRVLLKAFGAELVLTDPAKG  131
              PG+SVL+EPTSGNTG+GLAF A+ +GY+LI+TMPASMS ERR++L AFG ELVLTDPAKG
1z7w     65   PGESVLIEPTSGNTGVGLAFTAAAKGYKLIITMPASMSTERRIILLAFGVELVLTDPAKG  124

OAS-TLC  132  MTGAVQKAEEILKNTPDAYMLQQFDNPANPKIHYETTGPEIWDDTKGKVDIFVAGIGTGG  191
              M GA+ KAEEIL  TP+ YMLQQF+NPANPKIHYETTGPEIW  T GK+D FV+GIGTGG
1z7w     125  MKGAIAKAEEILAKTPNGYMLQQFENPANPKIHYETTGPEIWKGTGGKIDGFVSGIGTGG  184

OAS-TLC  192  TITGVGRFIKEKNPKTQVIGVRPTESDILSGGKPGPHKIQGIGAGFIPKNLDQKIMDEVI  251
              TITG G+++KE+N   ++ GV P ES ILSGGKPGPHKIQGIGAGFIP  L+  ++DEV+
1z7w     185  TITGAGKYLKEQNANVKLYGVEPVESAILSGGKPGPHKIQGIGAGFIPSVLNVDLIDEVV  244

OAS-TLC  252  AISISKLEAIETAKQLALKEGLMVGISSG-AAAAAIKVAKRPENAGKLIAVVFPSFGE-Y  309
               +S    E+I+ A+QLALKEGL+VGISSG AAAAAIK+A+RPENAGKL   +FPSFGE Y
1z7w     245  QVSSD--ESIDMARQLALKEGLLVGISSGAAAAAAIKLAQRPENAGKLFVAIFPSFGERY  302

OAS-TLC  310  LSTPLFQSIREE  321
              LST LF + R+E
1z7w     303  LSTVLFDATRKE  314

Score =  437 bits (1124),  Expect = 5e-121
Identities = 218/312 (69%), Positives = 265/312 (84%), Gaps = 5/312 (1%)
```

Fig. 2. Pairwise sequence alignments of SAT3 and OAS-TLC with their best templates from the BLAST2 server.[39] The blue line indicates the identity and similarity. The alignment of SAT3 with 1t3d is partially shown to demonstrate the alignment of the C-terminal part of the enzymes. Residues shown in green in 1t3d refer to the solved three-dimensional structure, which were later taken for comparative modeling. The residues squared in blue belong to the SAT-binding site of OAS-TLs. The residues squared in red represent the conserved Asn loop. The conserved residues of the active sites of OAS-TLs are shown in red. Residue X stands for selenomethionine.

Modeling quaternary structures

From SAT3 to AtSAT^{mit}/AtSAT·CoA^{mit}. The structures of 1s80 from *H. influenzae* (HiSAT^{1s80}) and 1t3d from *E. coli* (EcSAT^{1t3d}) are homohexamers made of a dimer of trimers with 3·2 symmetry.[19,29] The ribbon representations of HiSAT^{1s80} and EcSAT^{1t3d} are presented in Fig. 3a and b, respectively. The monomers of both enzymes are composed of two domains, an α-helical domain (SATase_N domain, PF06426) of eight helices and a left-handed parallel β-helix domain. The latter contains a series of hexapeptide repeats.[19,29] The serine catalytic sites are located in the clefts formed by pairs of adjacent threefold-related β-helix domains.

It has been suggested by Gorman and Shapiro[19] that the twofold interface between the trimers is mediated by the interaction of SATase_N domains. The threefold interface is assembled from the interaction along the β-helix domains and the SATase_N domains. The authors discuss the conservation of SATase_N domains within a subset of serine-acetyl transferases from plants and bacteria. The multiple-sequence alignment of SAT3 with 1s80 and 1t3d revealed that in SAT3 the region similar to the SATase_N domain corresponds to residues 124 to 191 and not to the N-terminal tail region. The sequence identity of these domains is about 35%, and the homology is about 61%. However, the conserved residues of the SATase_N domain[19] that lie at the twofold interface were not found in SAT3.

The chimeric template for comparative modeling of the quaternary structure of SAT3 was composed of the first 237 residues contributed by residues 4–240 from HiSAT^{1s80} and the following 14 residues contributed by residues 245–258 from EcSAT^{1t3d}. Five quaternary structures of SAT3 were modeled and constitute residues 124–382. Residues 368–372 in SAT3, which match the gap of 5 amino acids in 1t3d, were treated as a loop (Fig. 2).

The structure of 1s80 as a main template and the models generated were validated (see Methods). All five models exhibit a Ramachandran plot core percentage around 88%, with a *G*-factor of about −0.2 and approximately 60 bad contacts that were not present in the reference structure. No "poor" or "bad" rating was given to any of the models by WHATCHECK. Minimization of the models reduced the number of bad contacts to zero but caused dihedral angles to worsen for several residues. Some structural parameters also became rated as "poor" or "loose" by WHATCHECK. Since the aim was to model CS formation rather than accurate structures for the isolated SAT3 and OAS-TLC enzymes, we selected the model with the lowest RMSD (0.32 Å) to the 1s80 structure as the most suitable. This model is here designated as multimeric AtSAT^{mit}, and its structure is shown in Fig. 3c.

AtSAT·CoA^{mit}, another model with a shorter length of the sequence of SAT3 (residues 124–376), was derived from the modeled AtSAT^{mit} (see Methods).

The structures AtSAT^{mit} and AtSAT·CoA^{mit} were modeled in accordance with two crystal structures of SATs, previously introduced as 1ssq and 1sst from *H. influenzae*. The structure 1ssq reveals a fixed position of the C-terminal tail within the cavity between the two adjacent subunits. The structure 1sst differs by the presence of CoA between the subunits and thus a floppy C-terminal tail. Therefore, the model of the more complete structure, AtSAT^{mit}, relates to the initial potential mode of CS complex formation when no CoA is present. The second one, AtSAT·CoA^{mit}, is relevant to complexation when CoA is bound.

The N- and C-termini of SAT3. The multiple sequence alignment of more than 90 representatives of SATs from different organisms reveals that the N-terminal region as it appears in the full sequence of SAT3 is also present in other higher organisms but does not exist in bacterial SATs. Experimental data suggest that since the SAT3 is a mitochondrial protein, the mature SAT3 contains a shorter sequence, which starts from either L107 or T109.[40] However, since no clear secondary structure in the 15 (17) amino acid residues of the N-terminal tail of SAT3 was predicted by PredictProtein[41] and there is no experimental evidence for the importance of the N-terminal tail in CS complex formation, we did not model its structure in this work. Nevertheless, the possible steric impact of the SAT3 N-termini on OAS-TLC docking is discussed later.

Due to the importance of the C-terminal tails in CS complex formation, the modeling of its structure is presented separately below.

From OAS-TLC to AtOASS^{mit}. The three-dimensional structure of 1z7w from *A. thaliana* cytosol (AtOASS^{cyt}) was previously selected to serve as the main template for modeling the quaternary structure of OAS-TLC.

AtOASS^{cyt} (1z7w) is a symmetric dimer with monomers related by a 2-fold crystallographic axis. Each monomer is built of two α/β structural domains and is composed of 322 amino acid residues. The crystal structure 1z7w (3–322) was determined to 2.2 Å by molecular replacement, using the structure 1oas from *Salmonella typhimurium* (StOASS^{1oas}) as a search model.[27] The active site for the substrate OAS in AtOASS^{cyt} is defined by the location of the cofactor pyridoxal 5′-phosphate (PLP), which covalently binds to K46 (K46-PLP) and forms an internal Schiff base. The asparagine loop, which was previously suggested to be involved in substrate (OAS) binding in StOASS^{1oas}, is found to be highly conserved for both AtOASS^{cyt} and OAS-TLC (Fig. 2). A ribbon representation of AtOASS^{cyt} and its catalytic site with PLP cofactors is shown in Fig. 4a.

To model the quaternary structure of OAS-TLC, we considered only the homologous residues in 1z7w. Namely, the three-dimensional reference structure consisted of residues 5–314 from AtOASS^{cyt} (see Fig. 2). The models exhibit nearly the same results for the

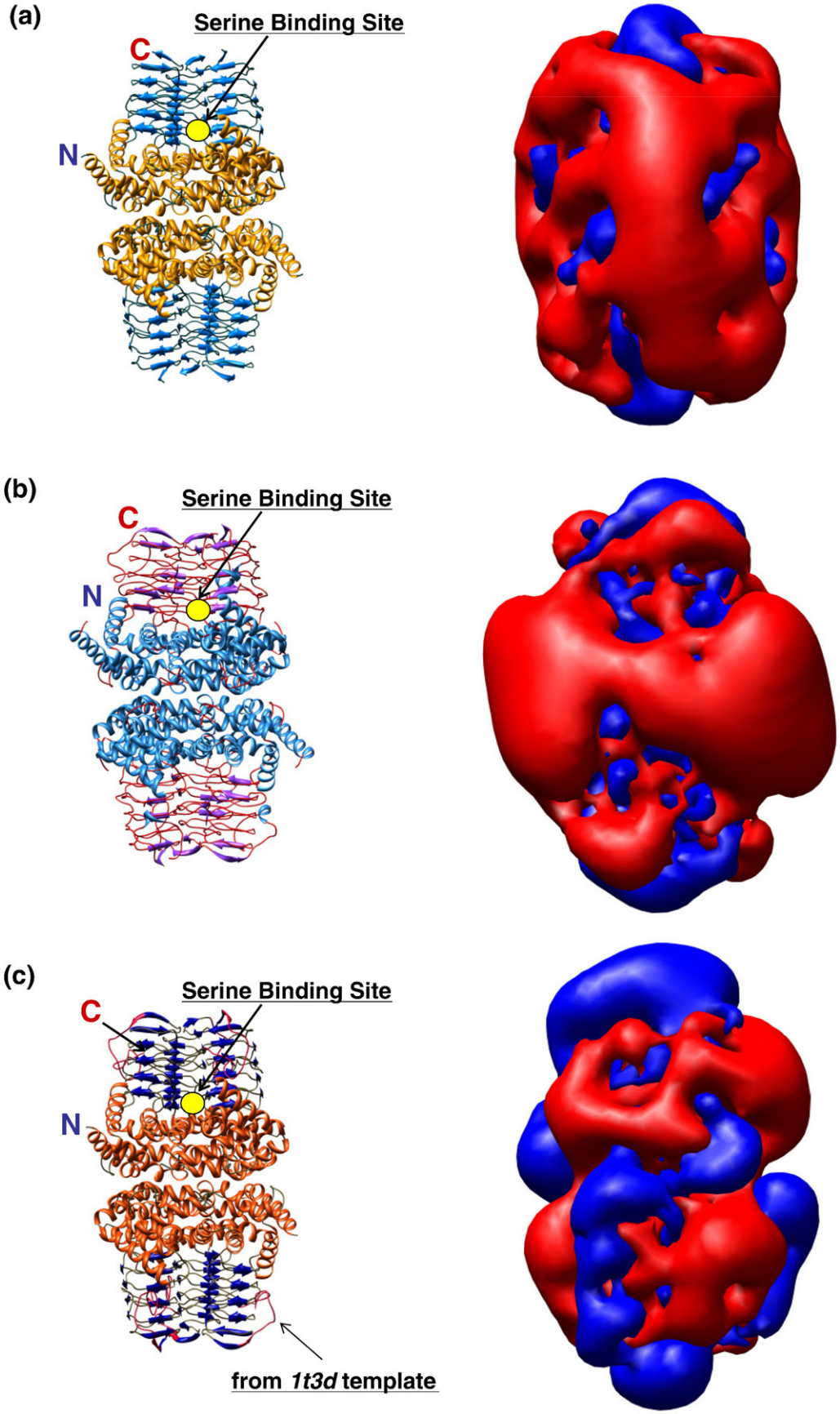

Fig. 3 (*legend on next page*)

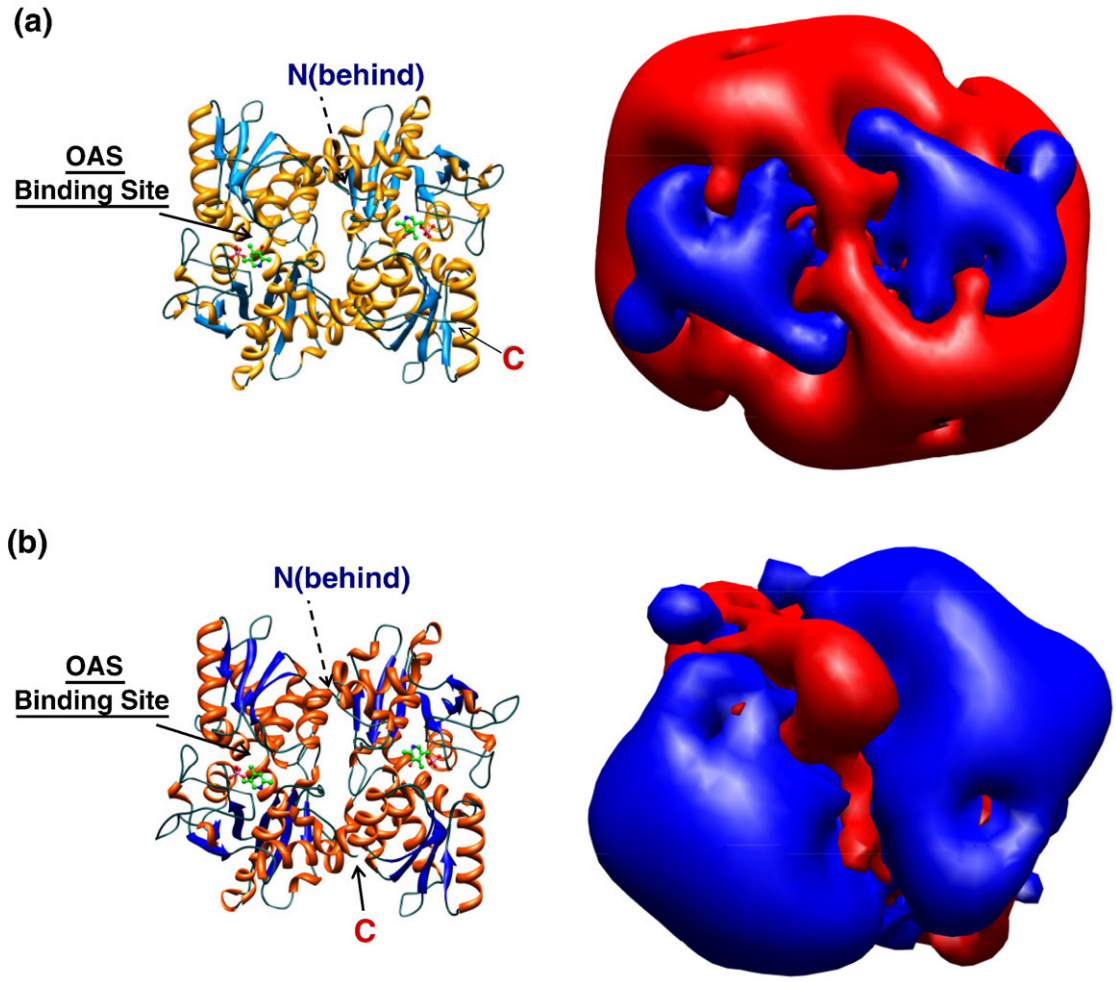

Fig. 4. Three-dimensional structures and the isopotential contours of AtOASS^{cyt} and AtOASS^{mit}. The three-dimensional structures of AtOASS^{cyt} and the modeled AtOASS^{mit} are shown in cartoon representation, and their corresponding electrostatic isopotential contours at $+0.2kT/e$ (blue) and $-0.2kT/e$ (red) are shown in (a) and (b), respectively.

PROCHECK Ramachandran plot core percentage (approximately 92%), overall *G*-factor (approximately −0.02) and number of bad contacts (approximately 10). None of the structures was scored as "bad" or "poor." Therefore, similarly to SAT3, for further simulations, we chose the model with the lowest RMSD (0.23 Å) to AtOASS^{cyt} as the most suitable. This model is designated as AtOASS^{mit} and corresponds to residues 13–321 of OAS-TLC. The complete quaternary structure of the modeled AtOASS^{mit} is presented in Fig. 4b.

Toward the CS complex

In this study, CS complex formation was modeled by mimicking the physical process of diffusional association of AtSAT^{mit}/AtSAT·CoA^{mit} and AtOASS^{mit}. Diffusional association is expected to be driven by long-range electrostatic forces, resulting in encounter complex formation,[42] and this was simulated using the BD technique. Short-range forces[42] are important for the transformation of encounter complexes into bound complexes, and MD simulations were therefore performed to refine the encounter complexes and obtain tighter complexation (see Methods).

Electrostatic properties of the copartners

To evaluate the contribution of the protein electrostatics to the diffusional association, we computed the net charges and the spatial charge distributions of the enzymes from the three organisms most relevant to our study. In Table 2, the net

Fig. 3. Three-dimensional structures and the isopotential contours of the templates and the modeled AtSAT^{mit}. The three-dimensional structures of HiSAT^{1s80}, EcSAT^{1t3d} and the modeled AtSAT^{mit} are shown in cartoon representation, and their corresponding electrostatic isopotential contours at $+0.2kT/e$ (blue) and $-0.2kT/e$ (red) are shown in (a), (b) and (c), respectively. The electrostatic potential was calculated for structures that are lacking C- and N-terminal tails for AtSAT^{mit}. The catalytic binding site for serine is shown by a yellow ball.

Table 2. Net charges of the enzymes from *H. influenzae*, *E. coli* and *A. thaliana* according to their primary and quaternary structures as well as to the missing fragments

Enzyme	Organism	Full sequence (e)	Three-dimensional structures (e)		Missing N-terminus[a] (sequence length/charge) (e)	Missing C-terminus[a] (sequence length/charge) (e)
SATs						
1s80	*H. influenzae*	−6	$HiSAT^{1s80}$	−3	—	27 aa/−3
1t3d	*E. coli*	−5	$EcSAT^{1t3d}$	−3	—	11 aa/−2
Mature SAT3	*A. thaliana* (mit)	−6	$AtSAT^{mit}$	0	15 aa or 17 aa/−4	12 aa/−2
			$AtSAT{\cdot}CoA^{mit}$	+6		18 aa/−3
			6CoA	−24		

Enzyme	Organism	Full sequence (without/with PLP) (e)	Three-dimensional structures (with PLP) (e)		Missing N-terminus[a] (sequence length/charge) (e)	Missing C-terminus[a] (sequence length/charge) (e)
OAS-TLs						
1y7l	*H. influenzae*	−2/−3	$HiOASS^{1y7l}$	−1	—	5 aa/−2
EcCysK	*E. coli*	−3/−4	$EcOASS^{CysK}$	−3	—	7 aa/−2
1z7w	*A. thaliana* (cyt)	−2/−3	$AtOASS^{cyt}$	−3	—	—
OAS-TLC	*A. thaliana* (mit)	0/−1	$AtOASS^{mit}$	+2	12 aa/0	9 aa/−3

All the results were computed for the monomer subunit with the ISREC SAPS server assuming standard protonation states at pH 7.[35]
[a] The net charges for the missing N- and C-termini were calculated without consideration of the charge of the terminal amino/carboxy groups.

charges at pH 7 for SATs and OASSs from *H. influenzae* (1s80 and 1y7l), *E. coli* (1t3d and *EcCysK*) and *A. thaliana* (1z7w and the models) are summarized. *EcCysK* refers to an OASS from *E. coli* that is encoded by the *CysK* gene and experimentally known to form a complex with a cognate EcSAT (discussed later). The structure was modeled and designated as $EcOASS^{CysK}$.

The results reveal that in *H. influenzae*, *E. coli* and *A. thaliana*, the association appears to be between two negatively charged enzymes and occurs most probably due to the topology of the distributed charges. If the CS complex forms when CoA is bound to SAT, then the net charge of the SAT hexamer is reduced by $(-4e)\times(6)$, making the apices of SAT even more negatively charged.

The electrostatic potential of each protein was calculated by solving the linearized finite-difference Poisson–Boltzmann equation (see Methods). The nonhomogeneous distribution of the electrostatic potentials is shown in Figs. 3 and 4 by the isopotential contours at $\pm 0.2kT/e$. $HiSAT^{1s80}$, $EcSAT^{1t3d}$ and $AtSAT^{mit}$ reveal a distribution of mixed negative and positive potentials along the 3-fold axes and clear positive patches at the C-terminal apices of the trimers. Despite the relatively high sequence similarity between $HiSAT^{1s80}$, $EcSAT^{1t3d}$ and $AtSAT^{mit}$, their electrostatic potentials exhibit different distributions. The isosurface of $AtSAT^{mit}$'s potential is, compared with that of $HiSAT^{1s80}$ for instance, "twisted" along the 3-fold axis. It looks even more different from the potential of $EcSAT^{1t3d}$, in particular, in the α-helical domain. The electrostatic isopotential surfaces of $AtOASS^{cyt}$ and $AtOASS^{mit}$ (Fig. 4) show positive patches at the catalytic sites. They acquire the form of "blue eyes" symmetrically positioned on both sides of the 2-fold axis. The isosurface of $AtOASS^{cyt}$ is for the complete three-dimensional structure, whereas for $AtOASS^{mit}$, the negatively charged C-termini are missing.

The enzymes are expected to exhibit electrostatic complementarity at their binding interfaces. Considering the full structures of the enzymes, the negatively charged C-termini of SATs in conjunction with CoA are complementary to the positively charged catalytic sites of OASSs. This is in agreement with experiments that demonstrate the formation of the complex via the C-termini (discussed below). The lack of structural information on certain protein fragments makes the electrostatic landscape incomplete and therefore might bias the modeling of complexation.

Steric interference of the N-termini of SAT3

Keeping in mind the typical structure of bacterial SAT, it is most probable that the N-terminal tail in $AtSAT^{mit}$ would point outward from the plane of the interface,[18] where two trimers of SAT monomers contact. Since no secondary structure was predicted for the N-termini, their floppy structure around $AtSAT^{mit}/AtSAT{\cdot}CoA^{mit}$ may interfere with the steering of $AtOASS^{mit}$ toward $AtSAT^{mit}/AtSAT{\cdot}CoA^{mit}$. On the other hand, the N-termini may also contribute to correct docking of $AtOASS^{mit}$ by navigating it toward the apices of the SAT where the C-termini are available for binding.

Rigid-body docking

Constrained simulation diffusional association docking. Several groups have experimentally studied CS formation in bacteria[13,16,18,25,43] and plants,[12,15,17,28] identifying the sequence fragments crucial for complexation. It was concluded that the last 10 amino acid residues at the C-terminus of SAT are largely responsible for the interaction with OASS. Comparing the C-terminal regions of *E. coli*, *S. typhimurium*, *A. thaliana* and *Chlorella vulgaris* SAT, Mino *et al.*[18] noticed that the C-termini contain

conserved Asp and Glu residues and a common last residue, Ile. It was thus suggested that these residues could play a key role in establishing the interactions between SAT and OASS. Another group of researchers, Huang *et al.*,[25] determined the crystal structure of HiOASS with the C-terminal decapeptide (C10) of HiSAT, which is required for bienzyme complex formation [Protein Data Bank (PDB) entry 1y7l]. The crystal structure at 1.55 Å resolution revealed clear electron density for the last 4 residues of the C-terminal region (~NLNI) in the active site of $HiOASS^{1y7l}$. The conformation of these residues suggests competition between the last C-terminal residue (Ile) and the substrate OAS for the active site of $HiOASS^{1y7l}$. Recently, a new structure of $AtOASS^{cyt}$ with the C10 peptide of $AtSAT^{cyt}$ was solved at 2.8 Å resolution with the last 8 residues (~TEWSDYVI) resolved (PDB entry 2isq).[28] These results confirm that the complexation of SAT and OASS occurs when the C-terminus of SAT occupies the same binding pocket as the substrate of OASS. Considering the sequence similarity of 1y7l, 2isq and OAS-TLC, the conservation of the active sites and the identity of the last amino acid residue (Ile) of HiSAT, cytosolic AtSAT and SAT3, we assumed that the mechanism of CS formation in *A. thaliana* mitochondria is similar to that in *H. influenzae* and *A. thaliana* cytosol.

The information that the C-terminal tail of SAT is located in the active site of OASS when SAT–OASS is formed was thus incorporated into the docking protocol for $AtSAT^{mit}$–$AtOASS^{mit}$ and $AtSAT{\cdot}CoA^{mit}$–$AtOASS^{mit}$. For the latter, the parameters of CoA were considered (Methods). In this stage, we assumed that the last five amino acids of SAT3 (~SDYVI) possess the same conformation in the active site as in the crystal structure 2isq. The structures of all eight residues of $AtSAT^{cyt}$ in 2isq were not considered since the preceding ~TEW~ is exposed, pointing out of the active site of $AtOASS^{cyt}$, and therefore may occupy a different conformation upon protein complexation. For constrained simulation diffusional association (C-SDA) docking, we defined a distance constraint of 25 Å between the last modeled residue of $AtSAT^{mit}$/$AtSAT{\cdot}CoA^{mit}$ and the "imaginary $Ile^{C\alpha}$" in the active site of $AtOASS^{mit}$ (see Fig. 5a Methods; Methods).

For the constrained $AtSAT^{mit}$–$AtOASS^{mit}$ docking [in contrast to free SDA (F-SDA), where $AtOASS^{mit}$ mostly docked to the central part of $AtSAT^{mit}$], the generated positions of $AtOASS^{mit}$ at the apices of $AtSAT^{mit}$ are energetically less favorable. There is a reduction of ~29% in electrostatic binding energies on average. However, the presence of CoA in the system, as tested in constrained $AtSAT{\cdot}CoA^{mit}$–$AtOASS^{mit}$ docking in comparison with $AtSAT^{mit}$–$AtOASS^{mit}$, increases the magnitude of the electrostatic binding energy by 13%. For both systems, we observed symmetry in the docked positions of $AtOASS^{mit}$ with respect to the 3-fold symmetry axis of SAT.

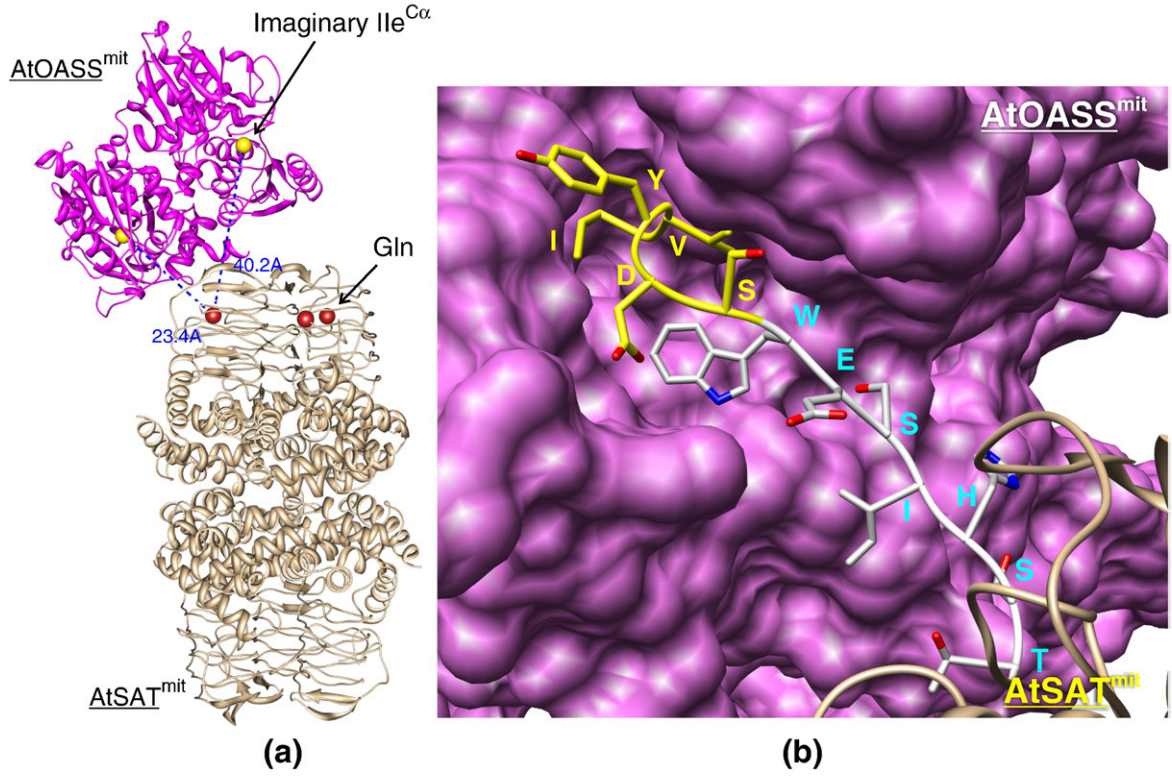

Fig. 5. C-SDA docking and C-terminus modeling. (a) Exemplification of the C-SDA results by one docked solution (encounter complex). The residues of the "imaginary $Ile^{C\alpha}$" are represented by yellow balls, whereas the last modeled Glu residues of $AtSAT^{mit}$ are represented by red balls. (b) Modeled C-terminal tail. The positions of the last five residues shown in yellow were taken from the structure of $AtOASS^{cyt}$ (2isq), while those shown in white were modeled with the MODELLER program.

C-terminal tail modeling. Although generating the encounter complexes of AtSAT^{mit}–AtOASS^{mit} (AtSAT·CoA^{mit}–AtOASS^{mit}) with the C-SDA protocol provides information on how AtOASS^{mit} might be oriented with respect to AtSAT^{mit}/AtSAT·CoA^{mit}, it does not specify how the C-terminus might be positioned at the entrance to the active site of AtOASS^{mit}. To address this question, we modeled the 12 amino acids of the C-terminal tail of SAT3 for each encounter complex AtSAT^{mit}–AtOASS^{mit}. For AtSAT·CoA^{mit}–AtOASS^{mit} complexation, we modeled 18 amino acid residues (see Methods).

The conformations of the modeled SAT3 C-termini revealed that in some cases the backbone of the modeled C-terminal tail pierces through neighboring parts of AtSAT^{mit}, such as loops. This was mostly observed when a longer C-terminus was modeled with CoA bound. Such complexes were discarded. The results of PROCHECK validation of the C-terminus for AtSAT^{mit}–AtOASS^{mit} complexation demonstrated a very high "core score" up to 90% and no residue with unacceptable dihedral angles according to the Ramachandran plot. The number of bad contacts did not exceed 1, and the overall *G*-factor was approximately −0.3. Although the validation results for the C-terminus of the AtSAT·CoA^{mit}–AtOASS^{mit} complexes were less good, we chose the best ones mostly according to the Ramachandran plot. The worst case accepted had 2 of 30 validated residues with dihedral angles in an "unacceptable" area of the Ramachandran plot. The 30 residues refer to the entire C-terminus of SAT3, which is floppy in the majority of SAT structures, and starts from the last residue of the main template (HiSAT^{1s80}) in our models. Structurally, these residues are found in the tightest helical loop region of the C-terminal tail; therefore, their conformation might not be well optimized by automated modeling (Fig. 5b). Moreover, the results were acceptable because the model was later improved by refinement with MD. A modeled C-terminus within the active site of AtOASS^{mit} is shown in Fig. 5b.

The structure of the CS complex

Six AtOASS^{mit} dimers to one AtSAT^{mit}/AtSAT·CoA^{mit} hexamer. There is experimental evidence on the stoichiometric ratio of the entire CS complex.[11,15,16] It is believed that one SAT hexamer binds two OASS dimers on average. However, SAT is a homohexamer and contains six C-terminal tails; each of them has equal probability of interacting with OASS. Pye *et al.*[29] envisaged that the high efficiency of cysteine synthesis could be facilitated if six OASS dimers would bind to the SAT hexamer on the exterior of the interface between the β-helical domains. Here, we model the CS complex in *A. thaliana* mitochondria while testing the complexation of three AtOASS^{mit} dimers with one (upper) trimer of the AtSAT^{mit}/AtSAT·CoA^{mit} hexamer (Fig. 6a). It is our expectation that similar binding behavior would be observed for the second trimer of AtSAT due to the 2-fold symmetry.

Enzymatically inactive CS complex. We simulated the sequential assembly of AtOASS^{mit} on AtSAT^{mit} with the C-SDA protocol. The procedure and the complex are shown in Figs. 6 and 7a, respectively. As mentioned by many experimental groups, upon CS complex formation, positive cooperativity of SAT is observed (i.e., the production of the substrate OAS is

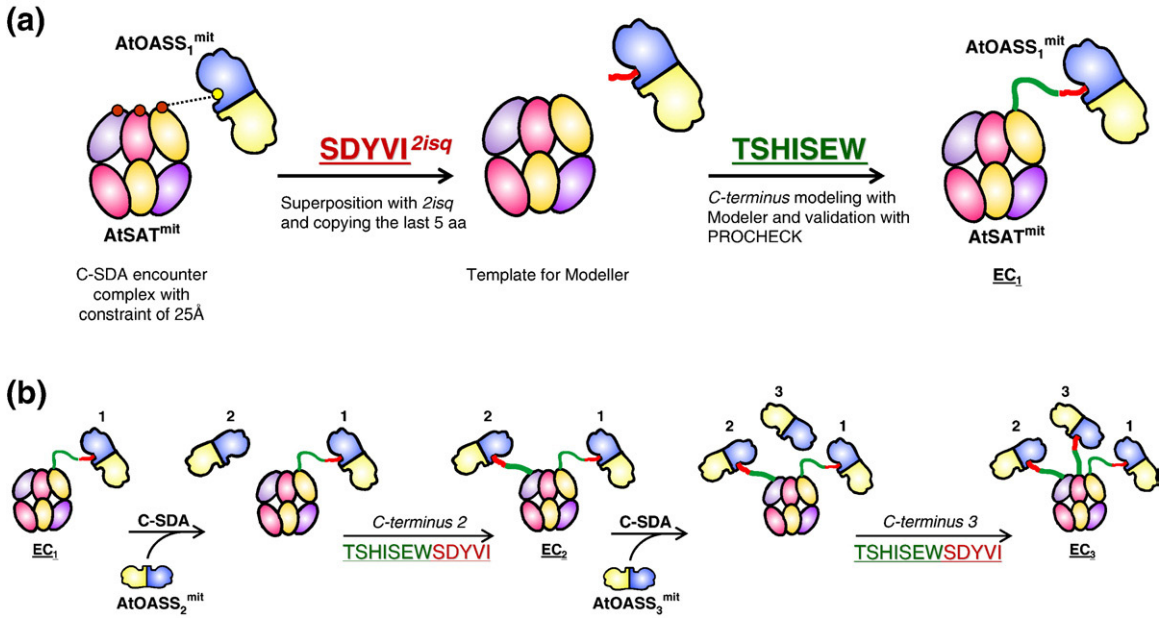

Fig. 6. Procedure of CS complex modeling by sequential docking of AtAOSS^{mit} on AtSAT^{mit} by C-SDA. (a) The detailed construction of the C-terminal tail to the first encounter complex (EC_1). (b) Subsequent docking of the second AtOASS^{mit} and the third AtOASS^{mit} and the construction of the C-terminal tails to obtain encounter complexes EC_2 and EC_3.

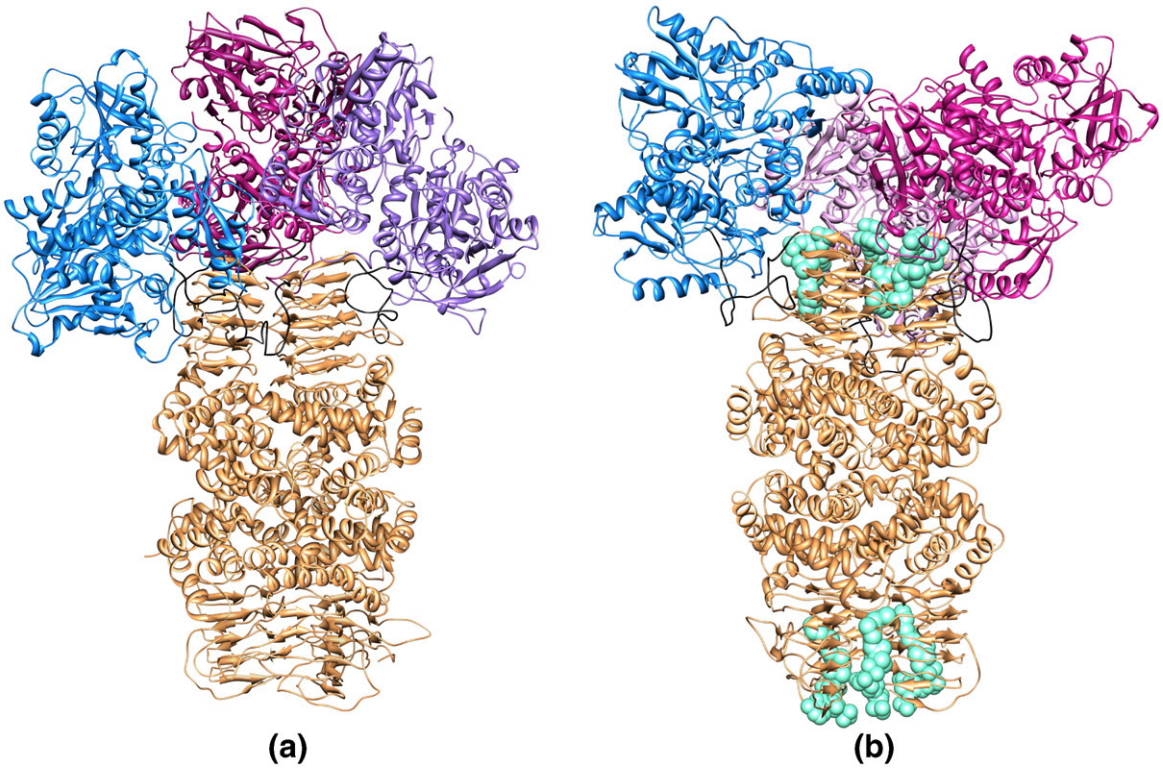

Fig. 7. Modeled CS complexes of six AtOASS^{mit} dimers to one AtSAT^{mit} hexamer. (a) Enzymatically inactive CS complex. The modeled C-terminal tails occupy acetyl-CoA-binding sites between the subunits of AtSAT^{mit} and are shown in black. (b) Enzymatically active CS complex. The CoAs are represented by spheres in light green. The modeled C-terminal tails are located outside the acetyl-CoA-binding sites and are shown in black.

enhanced). On the other hand, CS complex formation negatively affects the activity of OASS by reducing the production of cysteine. The CS complex modeled by sequential docking is an enzymatically inactive complex since the C-terminal tail occupies the binding site of acetyl-CoA, thereby hindering the production of OAS by SAT. Only when the C-terminal tail is not present in the acetyl-CoA-binding site does positive cooperativity occur. Nevertheless, one could speculate that this complex may form prior to binding of acetyl-CoA to AtSAT^{mit}. Regarding the activity of AtOASS^{mit} when the CS complex is formed, the active site of OAS is blocked by the AtSAT^{mit} C-terminus. According to the model, only one catalytic site of OAS is occupied when the CS complex forms, which is consistent with experiments (not published). This manner of interaction could lead to a reduction of cysteine production by AtOASS^{mit}.

Enzymatically active CS complex. We also simulated the complex of AtSAT·CoA^{mit}–AtOASS^{mit}. We used the results of the C-SDA docking that generated three main clusters of possible solutions for positioning AtOASS^{mit}. The best representatives of these clusters seem to be oriented in accordance with the 3-fold symmetry with respect to AtSAT·CoA^{mit}, showing no severe clash. After modeling and validating the C-terminal tails for each of these solutions, we manually constructed the CS complex by assembling three docked AtOASSs^{mit} (see Methods). Due to the presence of CoA, the conformations of the C-terminal tails are different from those in the previous complex. This model may thus represent the immediate result of the first step of cysteine biosynthesis, when acetyl-CoA has already reacted with serine and the by-product CoA is still in the system. This complex may be considered as an enzymatically active one and is presented in Fig. 7b.

Molecular dynamics refinement. The CS complexes constructed by rigid-body docking were refined with MD. The refinement was designed to generate plausible bound complexes.

We performed 400 ps of refinement. The backbone RMSD of the models after 400 ps exhibited fluctuations of ~2.9 and ~2.7 Å for the complexes AtSAT^{mit}–AtOASS^{mit} and AtSAT·CoA^{mit}–AtOASS^{mit}, respectively.

We analyzed the change in the size of interface between each docked AtOASS^{mit} and corresponding AtSAT^{mit}/AtSAT·CoA^{mit} in the complexes during the refinement. According to Lo Conte *et al.*,[44] protein–protein interfaces typically have a size of 1600 Å^2 (±400 Å^2), while large interfaces in the range of 2000–4660 Å^2 mostly occur between different components of the signal transduction system. The average size of the interfaces between each docked AtOASS^{mit} and AtSAT^{mit} in the complex modeled by sequential docking is 1570.5 Å^2 before refinement, whereas for the complex modeled by symmetry, AtOASS^{mit}–AtSAT·CoA^{mit}, it is 2057.5 Å^2.

Table 3. Binding energy calculations for formation of the CS complex performed with the APBS package

Complex	ΔE_{ele} (kcal/mol)	ΔE_{ap} (kcal/mol)	ΔE_{bind} (kcal/mol)
[$AtSAT^{mit}$]–$AtOASS_1^{mit}$	5.8	30.0	−24.2
[$AtSAT^{mit}$–$AtOASS_1^{mit}$]–$AtOASS_2^{mit}$	12.0	30.5	−18.5
[$AtSAT^{mit}$–$AtOASS_1^{mit}$–$AtOASS_2^{mit}$]–$AtOASS_3^{mit}$	14.6	52.8	−38.2

The binding energies were calculated on the CS complex, constructed by sequential docking. ΔE_{ele}, ΔE_{ap} and ΔE_{bind} are electrostatic, apolar and total binding energies, respectively. The binding energy is computed for systems consisting of the $AtSAT^{mit}$ hexamer and one, two or three $AtOASS^{mit}$ dimers bound to the same (upper) trimer of the SAT hexamer (see Figs. 6 and 7). The binding free energy was computed for the energy-minimized snapshot taken after 400 ps of flexible refinement (see Methods). An additional test calculation done for the structure extracted after 350 ps gave energies within 10% of those given in the table.

The average size of the interfaces after refinement increased to 2627 $Å^2$ for the first complex and to 2785.7 $Å^2$ for the second complex.

Two $AtOASS^{mit}$ dimers to one $AtSAT^{mit}$/$AtSAT{\cdot}CoA^{mit}$ hexamer

The stoichiometric ratio of OASS and SAT, which constitute a CS complex in bacteria and plants, was proposed from experiments to be two dimers of OASS to one hexamer of SAT. To address the question of whether all docked $AtOASSs^{mit}$ in the modeled CS complex bind with the same strength to $AtSAT^{mit}$, we computed the binding free energy of each $AtOASS^{mit}$ in the sequentially docked complex using APBS (see Table 3).

The results reveal that binding of the second $AtOASS^{mit}$ to the same $AtSAT^{mit}$ trimer in the enzymatically inactive CS complex is ~1.3 times less favorable after binding of the first $AtOASS^{mit}$. However, if binding of the second $AtOASS^{mit}$ occurs, the third $AtOASS^{mit}$ will also bind (to the same $AtSAT^{mit}$ trimer). A similar relation of the energies for binding of $AtOASSs^{mit}$ was observed with UHBD and the DFIRE scoring function. The same behavior of $AtOASS^{mit}$ binding could be seen from the electrostatic energies of the clusters generated during C-SDA docking. These results suggest that the most probable stoichiometric ratio in the enzymatically inactive CS complex is two dimers of $AtOASS^{mit}$ to one hexamer of $AtSAT^{mit}$, which is consistent with experiments.

To model the formation of the CS complex with a ratio of two dimers of $AtOASS^{mit}$ to one hexamer of $AtSAT{\cdot}CoA^{mit}$, we applied the sequential docking protocol to the binding of the second $AtOASS^{mit}$ to each of the trimers of the $AtSAT{\cdot}CoA^{mit}$ hexamer. According to the electrostatic binding energies from C-SDA, sequential binding of the second $AtOASS^{mit}$ to the second trimer is ~1.5 times more favorable than that to the first one (with one $AtOASS^{mit}$ dimer already bound). This implies that the most probable stoichiometric ratio in the enzymatically active CS complex also correlates with the experimental one. The modeled experimental enzymatically active CS complex is shown in Fig. 8.

The preference of the second OAS-TL to bind to the second trimer of SAT in both modeled complexes is mainly due to the change in topology of the electrostatic potential of SAT after binding of the first OAS-TL. The modified electrostatic potential of the trimer of SAT with OAS-TL attached seems to be less favorable for the second OAS-TL to bind. In contrast, the second trimer with no OAS-TL bound, and thus with no change in the electrostatic potential, is more preferable for the second OAS-TL to dock.

Binding of six dimers of $AtOASS^{mit}$ to one hexamer of $AtSAT^{mit}$/$AtSAT{\cdot}CoA^{mit}$ might nevertheless be plausible when the second enzyme ($AtOASS^{mit}$) is present in excess. The intermediate ratio of four dimers of $AtOASS^{mit}$ to one hexamer of $AtSAT^{mit}$/$AtSAT{\cdot}CoA^{mit}$ is not likely to persist.

A mechanistic overview of the models

The C-terminal tail influences the production of OAS in SAT by either blocking or exposing the active site to acetyl-CoA. For enzymatic activity in SAT, the C-terminus must be replaced by acetyl-CoA. This process occurs independently of CS complex formation.[20] However, the order of the steps in complexation is still not clear. Here, the structures of the models reveal two possible scenarios for CS complex formation. The first could occur when the C-terminus is located in the active site of acetyl-CoA. In this case, considering the most probable stoichiometric ratio of two OAS-TL dimers to one SAT hexamer, the remaining four sites might continue functioning independently of the presence of the OAS-TLs. The structure of this complex corresponds to the modeled enzymatically inactive CS complex. Another possible scenario is that complexation occurs in the presence of acetyl-CoA. In this case, the C-terminus is exposed to the exterior and the solvent. Although the four remaining sites could function in the same way as before, the activity of SAT would be enhanced by keeping the C-terminal tail away from the binding site of acetyl-CoA due to the presence of OAS-TL. This complex is the modeled enzymatically active CS complex. It is also possible that the two structures interchange as a result of the presence or absence of acetyl-CoA.

Hindrance of $AtSAT^{mit}$–$AtOASS^{mit}$ complexation in computational mutagenesis

Bonner *et al.*[27] discussed the evolutionary conservation of the loop of plant and bacterial OASS that forms

Fig. 8. Modeled CS complex of two $AtOASS^{mit}$ dimers to one $AtSAT{\cdot}CoA^{mit}$ hexamer. The positions of the docked $AtOASS^{mit}$'s were generated with a sequential docking C-SDA protocol and are found to be electrostatically the most favorable. The CoAs are represented by spheres in light green.

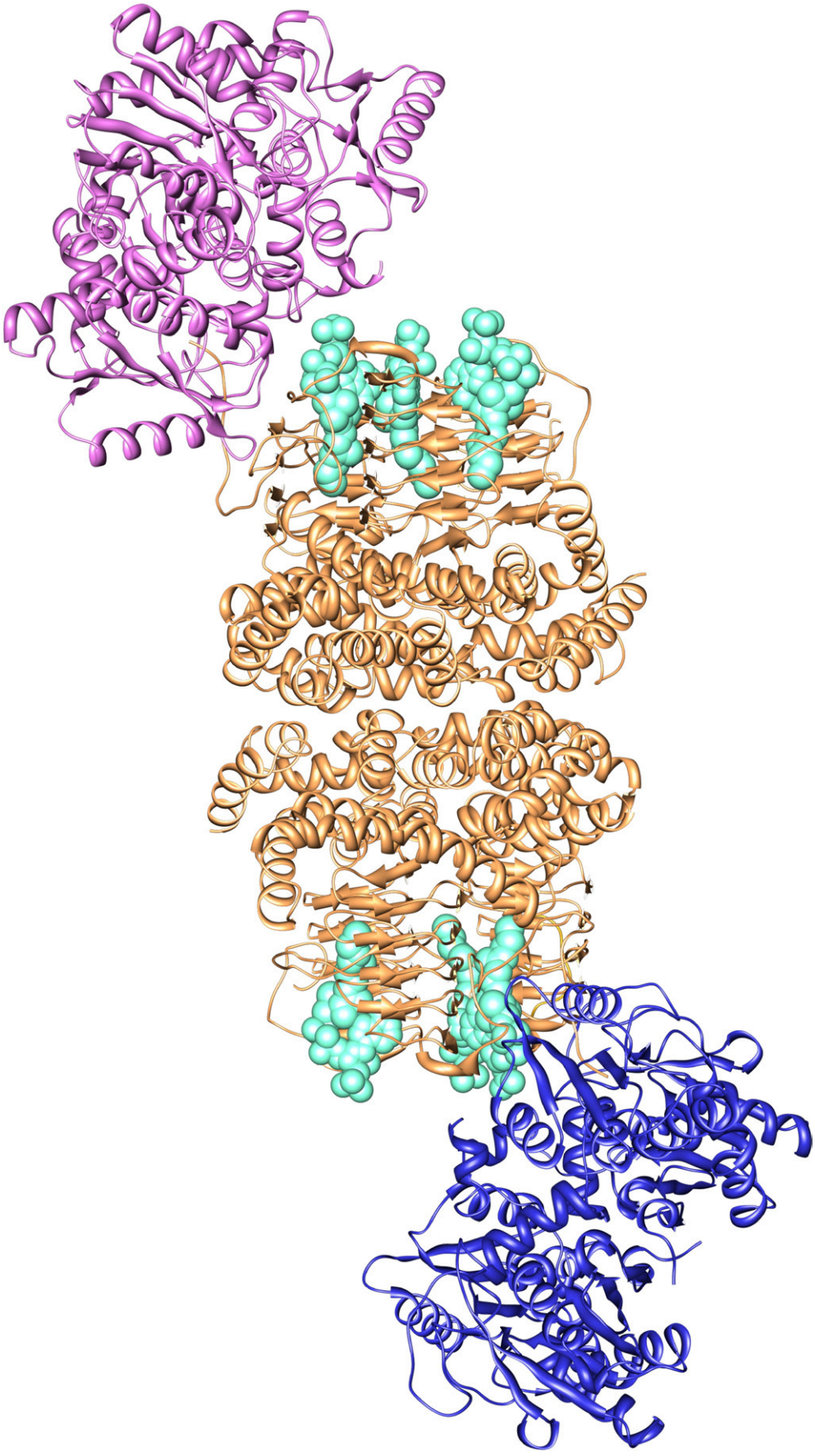

Fig. 8 (*legend on previous page*)

Table 4. The effect of mutations in the SAT-binding loop of $AtOASS^{mit}$ on $AtSAT^{mit}$–$AtOASS^{mit}$ complex formation during C-SDA docking

Mutation $AtOASS^{mit}$ ($AtOASS^{cyt}$)	Averaged E_{ele}[a] (kT)	Δ(mutant E_{ele} − wild-type E_{ele}) (%)
No mutation	−11.5	—
K212A (K217A)	−8.8	23
H216A (H221A)[b]	−11	4
K217A (K222A)	−10.0	13

[a] The electrostatic binding energy was averaged over the members of the docked clusters observed at the iterative step at which the first significant change in RMSDs between the clusters has occurred.
[b] The His is double protonated in the wild-type protein as a result of the strong hydrogen-bonding network to neighboring atoms.

part of the surface cleft ~15–20 Å^2 away from the active site. These residues are highlighted in Fig. 2. It was claimed that since the loop lies at the entrance to the active site, it may play an important role in CS complex formation. The loop contains two positively charged lysines and a histidine that is doubly protonated, donating hydrogen bonds to neighboring acceptor atoms. These residues were experimentally mutated to alanine, and their effect on formation of the CS complex was validated with a protein pull-down assay. The results revealed that mutagenesis of each of these residues prevents $AtSAT^{cyt}$ and $AtOASS^{cyt}$ from forming a complex. Later, this loop was designated the "SAT-binding site loop."

We performed a computational mutagenesis of these residues to alanine and redocked the modified $AtOASS^{mit}$ to $AtSAT^{mit}$. Both SDA protocols were applied, and the results of C-SDA are shown in Table 4. Each of the individual mutations increases the electrostatic binding energy, thereby making complexation less favorable. Mutation of the first lysine (K224 in OAS-TLC or K212 in the model) reduces the probability of binding by 23% during C-SDA docking. A similar tendency of lower binding is also observed during F-SDA. Such a behavior is consistent with the experimental results.

The importance of the SAT-binding loop was also experimentally tested for OASS-A from *E. coli*.[45] Site-directed mutagenesis of residues K221–P224 to K221A/P222E/G223E/P224E resulted in disablement of the interaction of OASS-A with a cognate SAT. The sequence "KPGP" is found to be conserved in OASS-TLC at residues 224–227. The computational mutagenesis corroborates the experimental results[45] and highlights residue K224 as the most important for the binding to SAT3 in *A. thaliana* mitochondria.

Unfavorable binding of EcCysM as a reference docking test

Two isoenzymes of OASS, A and B, have been detected in a variety of bacteria.[10,46,47] The representatives of isoenzyme OASS-A are encoded by the gene *CysK* and are known to be capable of forming a complex with the corresponding SATs. In contrast, isoenzyme OASS-B, which is encoded by the gene *CysM*, experimentally demonstrated no significant complexation with SAT.[48] The two isoenzymes exhibit different mechanisms of cysteine synthesis that complement each other when adapting to different growth conditions.[26] A solved structure of OASS-B from *E. coli* (PDB entries 2bht and 2bhs) differs in the active site compared with other structures of OASS-A. This could explain the loss in the ability of OASS-B to form a complex with SAT.

Binding of $EcOASS^{CysM}$ (2bht) to $EcSAT^{1t3d}$ was computationally tested by applying the C-SDA procedure. The results revealed that encounter complex formation of $EcSAT^{1t3d}$–$EcOASS^{CysM}$ is electrostatically unfavorable. The average electrostatic energy of $-2.3kT$ is 3.3 times less favorable than that for the complexation of $EcSAT^{1t3d}$ and $EcOASS^{CysK}$ (see Table 5). These results are in good agreement with experiments.

Mixed complexation during expression in *E. coli*

To evaluate the CS cross-complexation of enzymes from *A. thaliana* and *E. coli*, we analyzed their modes of binding and compared the electrostatic energies of the complexes to the nonmixed ones. Here, the binding of OASS from *E. coli* encoded by the *CysK* gene is discussed. Its three-dimensional structure ($EcOASS^{CysK}$) was generated by comparative modeling from the structure of OASS from *S. typhimurium* (PDB entry 1oas), which revealed 99% sequence similarity. The electrostatic binding energies of the $AtSAT^{mit}$–$AtOASS^{mit}$, $EcSAT^{1t3d}$–$EcOASS^{CysK}$, $EcSAT^{1t3d}$–$AtOASS^{mit}$ and $AtSAT^{mit}$–$EcOASS^{CysK}$ complexes resulting from C-SDA are summarized in Table 5.

To compare the electrostatic energies of the docked complexes, for simplicity, we averaged the energies of all structures over the clusters that were observed at a specified iterative step. The averaged electrostatic binding energies for C-SDA complexation of $AtSAT^{mit}$–$AtOASS^{mit}$ and $EcSAT^{1t3d}$–$AtOASS^{mit}$ are both approximately $-11.5kT$. This suggests that there is an equal chance for $AtOASS^{mit}$ to bind to the apices of either $EcSAT^{1t3d}$ or $AtSAT^{mit}$. It means that the inhibition of plant cysteine production might be enhanced by cross-binding of

Table 5. The averaged electrostatic binding energies of non-mixed and mixed encounter complexes resulting from C-SDA docking

Encounter complexes	Averaged E_{ele}[a] (kT)
$AtSAT^{mit}$–$AtOASS^{mit}$	−11.5
$EcSAT^{1t3d}$–$EcOASS^{CysK}$	−7.6
$AtSAT^{mit}$–$EcOASS^{CysK}$	−6.7
$EcSAT^{1t3d}$–$AtOASS^{mit}$	−11.5

[a] The electrostatic energy was averaged over the members of the docked clusters observed at the iterative step when the first significant difference in RMSDs between the clusters has occurred.

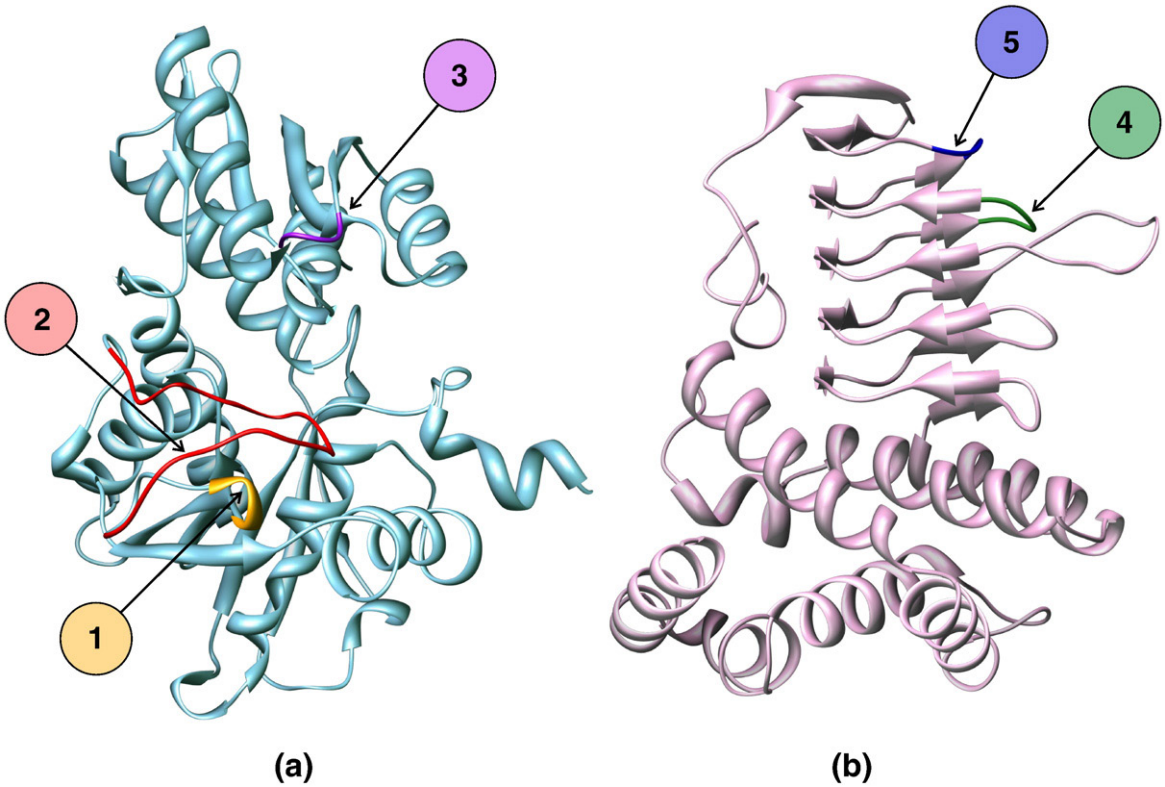

Fig. 9. Possible regions for mutation to affect CS formation. (a) Monomer of $AtOASS^{mit}$ with regions for mutation numbered 1, 2 and 3. (b) Monomer of $AtSAT^{mit}$ with regions for mutation numbered 4 and 5.

$AtOASS^{mit}$ to $EcSAT^{1t3d}$. $EcOASS^{CysK}$ prefers to bind to $EcSAT^{1t3d}$ rather than to $AtSAT^{mit}$.

In summary, during the expression of SAT3 and OAS-TLC in *E. coli*, there is a possibility of formation of the following CSs: $AtSAT^{mit}$–$AtOASS^{mit}$ and $EcSAT^{1t3d}$–$AtOASS^{mit}$ with a similar probability as well as $EcSAT^{1t3d}$–$EcOASS^{CysK}$ and $AtSAT^{mit}$–$EcOASS^{CysK}$ (in particular) at lower amounts.

Future work

The current models of the CS complex in *A. thaliana* mitochondria suggest further investigations of the interactions between SAT and OASS. More extensive mutation studies on both enzymes in conjunction with the flexibility of SAT and OASS upon binding might help determine more specifically the structure of the complex. Experimental mutagenesis could be performed on the regions that are found to be making tighter contacts in the refined CS complexes (see Fig. 9).

In addition to the "SAT-binding loop" (region 2), region 3 (D127–K130) also seems to be of high importance for CS complex formation. Its significance in *E. coli* was experimentally studied by Liszewska *et al.*[45] The deletion of residues E115–K118 led to hindrance of complex formation between OASS-A and SAT. In the refined modeled complexes, this region is found to closely interact with the C-terminal tail of SAT3. Site-directed mutagenesis might aid in identifying more specifically the residues that are involved in the binding of OAS-TLC to SAT3.

Conclusions

We investigated CS complex formation in *A. thaliana* mitochondria. As little structural and mechanistic information exists to date and the CS complex is a large multisubunit complex, modeling such a system is a very challenging task. Comparative modeling was used to construct the three-dimensional structures of SAT3 and OAS-TLC, based on structural information on the homologous proteins from bacteria and plants. The missing parts of the structure of SAT3, such as the C-termini that are crucial for CS complexation, were modeled *de novo*. The diffusional encounter complexes were generated with a modified version of the SDA program using BD simulation. To obtain tighter bound complexes, we performed refinement by MD simulation.

The question on the stoichiometric ratio of the OAS-TLC and SAT3 molecules in the CS complex was one of the key questions for this work. Although it was experimentally observed in bacteria and plants that only two OAS-TL dimers bind to one SAT hexamer to form a complex, the reason for the presence of four other equivalent binding sites on the SAT3 hexamer is puzzling. Therefore, we demonstrate here a model for the CS complex as an ensemble of six dimers of OAS-TLC and one SAT3 hexamer. However, the computed binding energies of each docked OAS-TLC in the modeled complex explain the reason for the experimentally observed ratio in the complex. The computational mutagenesis

results are found to be consistent with the experimental ones. Mutation of the charged residues on the SAT-binding loop of AtOASS^{mit} hinders CS complex formation. The buried surface area analysis showed that the interaction areas between SAT3 and OAS-TLC in the modeled CS complexes are within the range typical for protein–protein interactions.

The results presented in this work offer new insights into the mode of molecular recognition of SAT and OAS-TL enzymes in plants. The proposed structures of the enzymes and the models for CS complex suggest new scientific investigations. Experimental site-directed mutagenesis would assist in exploring the feasibility of the models. Charged residues found at the binding interface of the modeled CS complex could be tested experimentally. The suggested regions might include loops at the entrance to the active site of OAS (of OAS-TL) together with the upper loops of the β-sheet domain and C-terminal tail (of SAT). Further research and development for biotechnological and pharmaceutical purposes could be based on these findings.

Methods

Comparative modeling

Sequence alignment

To select the best candidates for templates for SAT3 and OAS-TLC structural modeling, we performed a pairwise sequence alignment with the enzymes with known crystal structures using BLAST2†.[39] Multiple-sequence alignment was carried out with the ClustalW program.[49] The entities for exploring the N-terminus in SATs were retrieved from the National Center for Biotechnology Information (NCBI)/Entrez Protein database‡.

SAT3 (EC 2.3.1.30). Five SATs with solved crystal structures (1s80, 1ssm, 1ssq, 1sst and 1t3d)[19,20,29] are available in the PDB§.[50] All enzymes belong to the bacterial kingdom. Their sequences were downloaded from the Swissprot/Trembl database‖ and aligned with the primary structure of SAT3. The main characteristics of the alignment are presented in Table 1.

Considering the importance of the C-terminus of SAT for CS complex formation, we selected the enzymes with more complete structures. Therefore, the crystal structures 1s80 from *H. influenzae*[19] with 51% sequence identity and 1t3d from *E. coli*[29] with a partially solved C-terminus were chosen to serve as the main templates for chimeric modeling of a structure of SAT3.

The N- and C-termini of SAT3. To explore the N- and C-terminal tails of SATs by multiple-sequence alignment, we retrieved more than 90 entities of SATs from the NCBI/Entrez Protein database‡. The secondary structure was predicted by PredictProtein¶.[41]

OAS-TLC (EC 2.5.1.47). Nine sequences of OAS-TLs with known crystal structures (1oas, 1fcj, 1ve1, 1o58, 1y7l, 2bht, 1z7w, 1z7y and 2isq)[21,23–28] were downloaded from the PDB and the Swissprot/Trembl database and aligned to the primary structure of OAS-TLC. Only the structures 1z7w, 1z7y and 2isq are from the *A. thaliana* cytosol; the rest are from bacterial organisms. The structural properties of the potential templates and their alignment to OAS-TLC are shown in Table 1.

The protein with the highest sequence similarity to the target with structure 1z7w was chosen as the template for comparative modeling.

Modeling quaternary structures

The modeling of SAT3 and OAS-TLC was performed using MODELLER 8v2 with an automated protocol for optimization and minimization of the models[a].[35]

The crystallographic coordinates of the templates were taken from the PDB,[50] and crystallographic water molecules were removed.

Verification of the templates and the models was performed using PROCHECK[b] [51] and WHATCHECK[c].[52] We retained models with reasonable PROCHECK Ramachandran plot core percentage score and WHATCHECK structural parameters. The models rated with WHATCHECK as "poor" or "bad" were rejected. Models with a high number of bad contacts were minimized with MD and validated again.

SAT3. Since the main template 1s80 demonstrates 51% sequence identity with the target, its 240 residues constitute a good structural reference for 61% of the residues of SAT3 (124–360). Since SAT3 is a mitochondrial enzyme, residues 1–106 (or 1–108) correspond to the N-terminal part, which is presumably truncated by proteases.[40] Residues 107–123 (or 109–123) constitute the N-terminal tail, which is not included in modeling because of lack of a structural reference. Information about the following 5%, which includes part of the C-terminus, could be derived from the second template, 1t3d or 1ssq. The structure of 1ssq was found to be very similar to that of 1t3d.

The chimeric template was constructed from 237 residues of 1s80 with no gap and 14 residues of 1t3d with a gap of 5 residues between 251 and 252, for

† http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi
‡ http://www.ncbi.nlm.nih.gov/entrez
§ http://www.rcsb.org
‖ http://www.expasy.org/sprot/
¶ http://www.predictprotein.org
[a] http://salilab.org/modeller/modeller.html
[b] http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
[c] http://swift.cmbi.kun.nl/WIWWWI/

each of the six monomers. The gap corresponds to residues 368–372 in SAT3 and is clearly observed in the pairwise alignment of SAT3 with 1t3d (Fig. 2). Based on this chimeric template, 256 of 391 residues of SAT3 were modeled, while MODELLER itself optimized the conformations of the 5 residues with no given reference structure. Five models were generated and structurally validated with PROCHECK and WHATCHECK. The model with the lowest RMSD calculated with PyMOL or MODELLER was selected for further calculations. The structure for SAT3 that constitutes 124–382 residues is denoted by AtSAT^{mit} in this article.

The additional model AtSAT·CoA^{mit} was constructed by chopping the last 7 C-terminal amino acid residues of AtSAT^{mit}. AtSAT·CoA^{mit} provides a structure for residues 124–376 of SAT3. This model was created to predict the CS complex when CoA is bound (discussed in Results and Discussion).

OAS-TLC. OAS-TLC was modeled against the 1z7w template, which has 69% sequence identity. For modeling, the cofactor PLP covalently bound to K46 (K46-PLP) in the template was simply substituted by K46 with no cofactor attached. K42 in the model of OAS-TLC corresponds to K54 in OAS-TLC's complete sequence. Later, K42 was replaced by K42-PLP after the template and modeled structures were superimposed. Ten models were generated and then validated with PROCHECK and WHATCHECK. The model with the lowest RMSD was explored further. This model is denoted as AtOASS^{mit}.

Electrostatics

Polar hydrogen atoms were added to AtSAT^{mit}/AtSAT·CoA^{mit} and AtOASS^{mit} without cofactor PLP with the WHATIF 5.0 program.[52] Protonation of K42-PLP and CoA was performed separately with the Sybyl program (Tripos Inc.). K42 in each monomer of AtOASS^{mit} was then manually replaced in the coordinate file by the protonated K42-PLP.

The electrostatic potentials were calculated with the UHBD program[53] by numerical solution of the linearized finite-difference Poisson–Boltzmann equation. Partial atomic charges and atomic radii were assigned from the OPLS parameter set.[54] The parameters for K42-PLP were taken from the CHARMM parameter set,[55] ensuring that the net charge of K42-PLP was (−1e). The parameters for CoA were derived with the Antechamber program of AMBER8 package. The net charge of CoA was required to be (−4e). Relative dielectric constants of 2 and 78 were assigned to the protein interior and solvent, respectively. Potentials were computed at an ionic strength of 50 mM. Cubic grids with 1.0 Å spacing and dimensions of 250×250×250 and 150×150×150 points were used to compute the electrostatic potentials of AtSAT^{mit}/AtSAT·CoA^{mit} and AtOASS^{mit}, respectively. To compute the electrostatic potential of the complexes, we used a grid dimension of 350×350×350 points with a spacing of 1.0 Å.

The ECM program[56] was used to calculate the effective charges for each protein by fitting so that they reproduce its external potential in a homogeneous dielectric environment corresponding to that of the solvent. The effective charges were placed on the carboxylate oxygen atoms of Asp and Glu residues and the C-terminus as well as on the amine nitrogen atoms of Lys and Arg residues and the N-terminus. The cofactor was not taken into account here, assuming that its effect on rigid-body docking would be negligible. Regarding CoA, we included the charges of the oxygen atoms of the phosphate groups.

The same protocol was also applied to calculate the electrostatic potentials for the template structures 1s80, 1t3d and 1y7w as well as for mutated AtOASS^{mit} structures.

Rigid-body docking protocol

The SDA program[57] was applied to simulate encounter complex formation for CS. The program was originally designed for computation of the bimolecular diffusional association rate constants for two proteins. It was modified for protein–protein docking applications to predict the structures of complexes. The modified SDA allows monitoring of the bimolecular complexes that form during the diffusional motion of the binding partners.[36] The proteins are treated as rigid bodies that diffuse with respect to one another under the influence of their electrostatic potentials. An encounter complex is considered to be formed when the two proteins come within a certain distance, called the "distance criterion."[57]

CS complexation was simulated for the modeled *A. thaliana* mitochondrial proteins AtSAT^{mit}/AtSAT·CoA^{mit} and AtOASS^{mit} and the *E. coli* EcSAT and EcOASS. The cofactor PLP and CoA were also included. The formation of the possible mixed CS complexes AtSAT^{mit}–EcOASS and EcSAT–AtOASS^{mit} was also tested.

The proteins were treated as rigid bodies. The translational and rotational motions were simulated for OASS relative to rotating SAT, which was positioned at the center of the simulation space. At the beginning of each trajectory, OASS was positioned with a randomly chosen orientation at a randomly chosen point on the surface of the sphere of radius b=300 Å centered on the geometric center of SAT. The b radius was chosen to exceed the sum of the maximum diameters of both proteins of ~134 Å (AtSAT^{mit}) and ~90 Å (AtOASS^{mit}). The simulations were performed until OASS diffused outside a sphere of radius c=450 Å centered on SAT.

For each pair of enzymes studied, the simulation of diffusional association was limited to either 99,999,999 trajectories or a total of 100 μs of simulated time.[36] Complex formation was monitored during each trajectory by recording when two proteins came within a certain distance. For F-SDA docking, the distance criterion was set to a center-to-center distance of 120 Å. This distance was chosen to exceed the sum of the maximum radii of AtSAT^{mit}

and AtOASS^{mit}. For C-SDA docking, the distance was 25 Å between the last modeled C-terminal residue of SAT and the "imaginary Ile$^{C^{\alpha}}$" in the binding site of OASS (as described below).

Constrained SDA docking: distance criterion

Based on the information from the crystal structure 2isq,[28] where the last eight residues of AtSAT cytosol bind in the active site of AtOASS, we assumed that CS complexation in *A. thaliana* mitochondria requires the two proteins to be within a distance equal to the length of SAT's C-terminal tail. We tested distances of different lengths, from 50 Å (which corresponds to the unfolded C-terminus) to 20 Å. To allow for the missing C-terminus to freely fold within the space between the two docked structures and avoid docking of AtOASS^{mit} close to the center part of AtSAT^{mit}/AtSAT·CoA^{mit}, we considered the distance of 25 Å as suitable. The distance was defined between the last modeled residue of each monomer of AtSAT^{mit}/AtSAT·CoA^{mit} and the C$^{\alpha}$ of the last residue (Ile) of cytosolic AtSAT in the structure 2isq, superimposed with AtOASS^{mit}. The C$^{\alpha}$ of Ile of the superimposed 2isq within the active site of AtOASS^{mit} was designated as "imaginary Ile$^{C\alpha}$".

To avoid a clash of the docked AtOASS^{mit} with the N-terminal tail for modeling the complex, we selected only those solutions that were found farther from the central region of SAT and closer to the apices of SAT. More specifically, the clashing was considered when the position of the docked AtOASS^{mit} was along the α-helical domain of the SAT, blocking its N-terminal endings.

Clustering and scoring

To cluster the resultant complexes, we applied an in-house program.[36] The configurations of the docked protein were clustered by a hierarchical clustering algorithm and an average linkage rule for the intercluster distance calculation. The distance between the clusters was defined as the average backbone RMSD between docked protein structures in one cluster relative to structures in another cluster. At the iterative cycle of clustering at which the clusters become distinguishable (larger than an empirical threshold of ~4%), the best representatives were monitored for each of observed clusters.[36]

Enzymatically inactive CS complex by sequential docking

C-SDA docking with one AtOASS^{mit}

The general protocol of rigid-body docking with constraints was first applied to simulate encounter complex formation of AtSAT^{mit} with one AtOASS^{mit} molecule.

The representative structures of the clusters were taken for further applications. These encounter complexes are designated here as EC$_i$.

C-terminal tail modeling

For each EC$_i$ complex, 20 C-terminal tails of SAT3 were modeled with the program MODELLER 8v2.

Due to sequential and structural conservation of the active sites of AtOASS^{mit} and AtOASS^{cyt} (2isq), the coordinates for the last five residues ~SDYVI for each encounter complex were derived from the superposition of the docked AtOASS^{mit} with the structure AtOASS^{cyt}. Once superimposed, the last residues ~SDYVI of AtSAT^{cyt} found in the AtOASS^{cyt} active site were positioned in the active site of AtOASS^{mit}. The remaining seven residues, TSHISEW, were treated as a loop and modeled with MODELLER. The encounter complex EC$_i$ and the five residues positioned within the active site of AtOASS^{mit} constituted the template (see Fig. 7). Twenty modeled C-terminal tails of SAT3 for each EC$_i$ were then validated by PROCHECK, and the best complexes E$_i^j$ were selected.

C-SDA docking with a second AtOASS^{mit}

The second AtOASS^{mit} was docked in a similar way to the first one but with the sampling space centered on E$_i^j$. Note that the electrostatic potential and effective charges were recalculated for each E$_i^j$. Moreover, the second AtOASS^{mit} was constrained to dock to any of the two remaining monomers. The best representatives were selected based on the sampling and scoring procedure as well as on the modeling of the C-terminal tail of AtSAT^{mit} in the second AtOASS^{mit} binding site. The modeled peptides were validated as previously with PROCHECK.

C-SDA docking with further AtOASSs^{mit}

This protocol of modeling was used to sequentially dock three AtOASS^{mit} molecules on three adjacent monomers of AtSAT^{mit}. The C-terminal tails of SAT3 were modeled into each of the docked AtOASS^{mit} molecules. Seven possible complexes with three docked AtOASS^{mit} enzymes were modeled. The same procedure might be applied to simulate complexation of the entire CS with a theoretically maximal number (i.e., 6) of the AtOASS^{mit} molecules.

Enzymatically active CS complex by symmetry

The C-SDA docking simulation of one AtOASS^{mit} on AtSAT·CoA^{mit} generated seven clusters. By visualizing the best representatives of the clusters, we omitted four clusters because of their being docked to the central part of AtSAT·CoA^{mit}. This step is to avoid a possible clash between the docked AtOASS^{mit} and the N-terminal tails that would be located at the central part of AtSAT·CoA^{mit}. The remaining three solutions demonstrated a quite symmetric orientation at the apex of AtSAT·CoA^{mit}, retaining the 3-fold symmetry of AtSAT·CoA^{mit}. The C-terminus of 13 amino acid residues was modeled in each of three solutions and validated with the

procedures described previously. The AtSAT·CoA^{mit}–AtOASS^{mit} complex was then manually constructed from the three solutions from the C-SDA simulation and the best modeled C-termini.

Refinement with MD

The complexes AtSAT^{mit}–AtOASS^{mit} and AtSAT·CoA^{mit}–AtOASS^{mit} were refined with MD simulation by using the AMBER8 package.[58,59] AMBER parameters for K-PLP were generated with Antechamber. AtSAT·CoA^{mit} contained CoA. Both CS complexes were immersed into a box with explicit water molecules of the TIP3 model. The systems were minimized for 200 steps and equilibrated for 40 ps, gradually increasing the temperature of the solute and the solvent to 300 K. The production steps were performed for 400 ps without restraints, with periodic boundaries and at constant pressure. Constant temperature with the weak-coupling algorithm (ntt = 1) and SHAKE constraints on bonds involving hydrogen (ntc = 2) were applied.

The energy changes during the MD simulation were analyzed. RMSD fluctuations of the backbone after equilibration were computed. The buried surface areas between each docked AtOASS^{mit} and AtSAT^{mit}/AtSAT·CoA^{mit} in the complexes before and after refinement were calculated with the NACCESS program[d].

Binding free energy calculations with APBS

The last snapshot of the "sequential docking" complex after refinement was minimized with AMBER8. The required PQR file was prepared from the RST file. The complex was then disassembled into the AtSAT^{mit} and each of the AtOASS$_i^{mit}$ molecules. The index i indicates the order of the sequentially docked AtOASS^{mit}. The binding energy was computed by using the APBS program for the following complexes: AtSAT^{mit} and AtOASS$_1^{mit}$; AtSAT^{mit}–AtOASS$_1^{mit}$ and AtOASS$_2^{mit}$; and AtSAT^{mit}–AtOASS$_1^{mit}$–AtOASS$_2^{mit}$ and AtOASS$_3^{mit}$.

The electrostatic term of the binding energy was calculated with the ELEC module by solving the nonlinearized Poisson–Boltzmann equation at an ionic strength of 50 mM. The grid dimension was taken to be 193 × 193 × 193 with a coarse grid length of 350 Å/side and a fine grid of 100 Å/side. The coarse and fine grids were centered on the binding interface of each of the complexes. The dielectric constants for the solute and for the solvent were assigned to 4 and 78, respectively. The solvent probe radius was chosen to be 1.4 Å for the electrostatic desolvation term.

To calculate the binding free energy governed by apolar solvation, we used the APOLAR module. The surface tension coefficient and the probe radius were assigned to 0.105 U and 1.4 Å, respectively.

[d] http://wolf.bms.umist.ac.uk/naccess/

Overall, the binding free energy ΔE_{bind} was computed as $\Delta E_{bind} = \Delta E_{ele} - \Delta E_{ap}$. The binding free energies were also calculated with UHBD[53] and scored with the DFIRE free energy method.[60] To check the reliability of the results, we also applied the protocol of computing binding free energy with APBS to the snapshot at 350 ps of the flexible refinement.

Acknowledgements

We gratefully acknowledge the support of the Klaus Tschira Foundation. We thank Martin Zacharias for discussions at an early stage of this work and Razif Gabdoulline and Domantas Motiejunas for assistance.

Supplementary information

Availability of the modeled structures of SAT3 and OAS-TLC, and the CS complexes addressing the reader to the Supplementary information for publication on the following webpage: http://projects.eml.org/mcm/data.

References

1. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R. *et al.* (2000). A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.
2. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA*, **98**, 4569–4574.
3. Cooper, M. A. (2003). Label-free screening of biomolecular interactions. *Anal. Bioanal. Chem.* **377**, 834–842.
4. Yang, Y., Wang, H. & Erie, D. A. (2003). Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. *Methods*, **29**, 175–187.
5. Velázquez Campoy, A. & Freire, E. (2005). ITC in the post-genomic era…? Priceless. *Biophys. Chem.* **115**, 115–124.
6. Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Pagé, N. *et al.* (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**, 2364–2368.
7. Baumeister, W., Grimm, R. & Walz, J. (1999). Electron tomography of molecules and cells. *Trends Cell Biol.* **9**, 81–85.
8. Kredich, N. M. (1992). The molecular basis for positive regulation of *cys* promoters in *Salmonella typhimurium* and *Escherichia coli*. *Mol. Microbiol.* **6**, 2747–2753.
9. Kredich, N. M. & Tomkins, G. M. (1966). The enzymic synthesis of L-cysteine in *Escherichia coli* and *Salmonella typhimurium*. *J. Biol. Chem.* **241**, 4955–4965.
10. Becker, M. A., Kredich, N. M. & Tomkins, G. M. (1969). The purification and characterization of *O*-acetylserine sulfhydrylase-A from *Salmonella typhimurium*. *J. Biol. Chem.* **244**, 2418–2427.
11. Kredich, N. M., Becker, M. A. & Tomkins, G. M. (1969). Purification and characterization of cysteine synthetase,

a bifunctional protein complex, from *Salmonella typhimurium*. *J. Biol. Chem.* **244**, 2428–2439.
12. Kumaran, S. & Jez, J. M. (2007). Thermodynamics of the interaction between *O*-acetylserine sulfhydrylase and the C-terminus of serine acetyltransferase. *Biochemistry*, **46**, 5586–5594.
13. Campanini, B., Speroni, F., Salsi, E., Cook, P. F., Roderick, S. L., Huang, B. *et al.* (2005). Interaction of serine acetyltransferase with *O*-acetylserine sulfhydrylase active site: evidence from fluorescence spectroscopy. *Protein Sci.* **14**, 2115–2124.
14. Johnson, C. M., Roderick, S. L. & Cook, P. F. (2005). The serine acetyltransferase reaction: acetyl transfer from an acylpantothenyl donor to an alcohol. *Arch. Biochem. Biophys.* **433**, 85–95.
15. Berkowitz, O., Wirtz, M., Wolf, A., Kuhlmann, J. & Hell, R. (2002). Use of biomolecular interaction analysis to elucidate the regulatory mechanism of the cysteine synthase complex from *Arabidopsis thaliana*. *J. Biol. Chem.* **277**, 30629–30634.
16. Mino, K., Imamura, K., Sakiyama, T., Eisaki, N., Matsuyama, A. & Nakanishi, K. (2001). Increase in the stability of serine acetyltransferase from *Escherichia coli* against cold inactivation and proteolysis by forming a bienzyme complex. *Biosci., Biotechnol., Biochem.* **65**, 865–874.
17. Wirtz, M., Berkowitz, O., Droux, M. & Hell, R. (2001). The cysteine synthase complex from plants. Mitochondrial serine acetyltransferase from *Arabidopsis thaliana* carries a bifunctional domain for catalysis and protein–protein interaction. *Eur. J. Biochem.* **268**, 686–693.
18. Mino, K., Hiraoka, K., Imamura, K., Sakiyama, T., Eisaki, N., Matsuyama, A. & Nakanishi, K. (2000). Characteristics of serine acetyltransferase from *Escherichia coli* deleting different lengths of amino acid residues from the C-terminus. *Biosci., Biotechnol., Biochem.* **64**, 1874–1880.
19. Gorman, J. & Shapiro, L. (2004). Structure of serine acetyltransferase from *Haemophilus influenzae* Rd. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 1600–1605.
20. Olsen, L. R., Huang, B., Vetting, M. W. & Roderick, S. L. (2004). Structure of serine acetyltransferase in complexes with CoA and its cysteine feedback inhibitor. *Biochemistry*, **43**, 6013–6019.
21. Burkhard, P., Rao, G. S., Hohenester, E., Schnackerz, K. D., Cook, P. F. & Jansonius, J. N. (1998). Three-dimensional structure of *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.* **283**, 121–133.
22. Burkhard, P., Tai, C.-H., Ristroph, C. M., Cook, P. F. & Jansonius, J. N. (1999). Ligand binding induces a large conformational change in *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.* **291**, 941–953.
23. Burkhard, P., Tai, C.-H., Jansonius, J. N. & Cook, P. F. (2000). Identification of an allosteric anion-binding site on *O*-acetylserine sulfhydrylase: structure of the enzyme with chloride bound. *J. Mol. Biol.* **303**, 279–286.
24. Heine, A., Canaves, J. M., von Delft, F., Brinen, L. S., Dai, X., Deacon, A. M. *et al.* (2004). Crystal structure of *O*-acetylserine sulfhydrylase (TM0665) from *Thermotoga maritima* at 1.8 Å resolution. *Proteins*, **56**, 387–391.
25. Huang, B., Vetting, M. W. & Roderick, S. L. (2005). The active site of *O*-acetylserine sulfhydrylase is the anchor point for bienzyme complex formation with serine acetyltransferase. *J. Bacteriol.* **187**, 3201–3205.
26. Claus, M. T., Zocher, G. E., Maier, T. H. & Schulz, G. E. (2005). Structure of the *O*-acetylserine sulfhydrylase isoenzyme CysM from *Escherichia coli*. *Biochemistry*, **44**, 8620–8626.
27. Bonner, E. R., Cahoon, R. E., Knapke, S. M. & Jez, J. M. (2005). Molecular basis of cysteine biosynthesis in plants: structural and functional analysis of *O*-acetylserine sulfhydrylase from *Arabidopsis thaliana*. *J. Biol. Chem.* **280**, 38803–38813.
28. Francois, J. A., Kumaran, S. & Jez, J. M. (2006). Structural basis for interaction of *O*-acetylserine sulfhydrylase and serine acetyltransferase in the *Arabidopsis* cysteine synthase complex. *Plant Cell*, **18**, 3647–3655.
29. Pye, V. E., Tingey, A. P., Robson, R. L. & Moody, P. C. (2004). The structure and mechanism of serine acetyltransferase from *Escherichia coli*. *J. Biol. Chem.* **279**, 40729–40736.
30. Wirtz, M. & Hell, R. (2006). Functional analysis of the cysteine synthase protein complex from plants: structural, biochemical and regulatory properties. *J. Plant Physiol.* **163**, 273–286.
31. Droux, M., Ruffet, M. L., Douce, R. & Job, D. (1998). Interactions between serine acetyltransferase and *O*-acetylserine (thiol) lyase in higher plants—structural and kinetic properties of the free and bound enzymes. *Eur. J. Biochem.* **255**, 235–245.
32. Wirtz, M., Droux, M. & Hell, R. (2004). *O*-acetylserine (thiol) lyase: an enigmatic enzyme of plant cysteine biosynthesis revisited in *Arabidopsis thaliana*. *J. Exp. Bot.* **55**, 1785–1798.
33. Hirai, M., Fujiwara, T., Awazuhara, M., Kimura, T., Noji, M. & Saito, K. (2003). Global expression profiling of sulfur-starved *Arabidopsis* by DNA macroarray reveals the role of *O*-acetyl-L-serine as a general regulator of gene expression in response to sulfur nutrition. *Plant J.* **33**, 651–663.
34. Hell, R. & Hillebrand, H. (2001). Plant concepts for mineral acquisition and allocation. *Curr. Opin. Biotechnol.* **12**, 161–168.
35. Sánchez, R. & Sali, A. (2000). Comparative protein structure modeling. Introduction and practical examples with Modeller. *Methods Mol. Biol.* **143**, 97–129.
36. Motiejunas, D., Gabdoulline, R., Wang, T., Feldman-Salit, A., Johann, T., Winn, P. J. & Wade, R. C. (2008). Protein–protein docking by simulating the process of association subject to biochemical constraints. *Proteins*, **71**, 1955–1969.
37. Jost, R., Berkowitz, O., Wirtz, M., Hopkins, L., Hawkesford, M. J. & Hell, R. (2000). Genomic and functional characterization of the oas gene family encoding *O*-acetylserine (thiol) lyases, enzymes catalyzing the final step in cysteine biosynthesis in *Arabidopsis thaliana*. *Gene*, **253**, 237–247.
38. Hell, R. & Bogdanova, N. (1995). Characterization of full-length cDNA encoding serine acetyl-transferase from *Arabidopsis thaliana*. *Plant Physiol.* **109**, 1498.
39. Tatusova, T. A. & Madden, T. L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**, 247–250.
40. Mentzel, T. (2006). *Regulation und Prozessierung der mitochondrialen Serin-Acetyltransferase aus Arabidopsis thaliana*. Diplomarbeit, Heidelberg, Germany.
41. Rost, B., Yachdav, G. & Liu, J. (2004). The PredictProtein server. *Nucleic Acids Res.* **32**, W321–W326.
42. Fersht, A. (1999). In *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Fersht, A., ed), 4th edit. W. H. Freeman and Co., New York, NY.

43. Mino, K., Yamanoue, T., Sakiyama, T., Eisaki, N., Matsuyama, A. & Nakanishi, K. (1999). Purification and characterization of serine acetyltransferase from *Escherichia coli* partially truncated at the C-terminal region. *Biosci., Biotechnol., Biochem.* **63**, 168–179.
44. Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.
45. Liszewska, F., Lewandowska, M., Płochocka, D. & Sirko, A. (2007). Mutational analysis of *O*-acetylserine (thiol) lyase conducted in yeast two-hybrid system. *Biochim. Biophys. Acta*, **1774**, 450–455.
46. Garvis, S. G., Tipton, S. L. & Konkel, M. E. (1997). Identification of a functional homolog of the *Escherichia coli* and *Salmonella typhimurium cysM* gene encoding *O*-acetylserine sulfhydrylase B in *Campylobacter jejuni*. *Gene*, **185**, 63–67.
47. Sirko, A. E., Zatyka, M. & Hulanicka, M. D. (1987). Identification of the *Escherichia coli cysM* gene encoding *O*-acetylserine sulfhydrylase B by cloning with mini-Mu-lac containing a plasmid replicon. *J. Gen. Microbiol.* **133**, 2719–2725.
48. Zhao, C., Kumada, Y., Imanaka, H., Imamura, K. & Nakanishi, K. (2006). Cloning, overexpression, purification, and characterization of *O*-acetylserine sulfhydrylase-B from *Escherichia coli*. *Protein Expression Purif.* **47**, 607–613.
49. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**, 3497–3500.
50. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
51. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
52. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics*, **8**, 52–56.
53. Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Luty, B. A., Ilin, A. *et al.* (1995). Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **91**, 57–95.
54. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236.
55. MacKerell, A. D., Jr, Brooks, B., Brooks, C. L., 3rd, Nilsson, L., Roux, B., Won, Y. & Karplus, M. (1998). CHARMM: the energy function and its parameterization with an overview of the program. In *The Encyclopedia of Computational Chemistry* (Schleyer, P. v. R., ed), vol. 1, pp. 271. John Wiley & Sons, Chichester, UK.
56. Gabdoulline, R. R. & Wade, R. C. (1996). Effective charges for macromolecules in solvent. *J. Phys. Chem.* **100**, 3868–3878.
57. Gabdoulline, R. R. & Wade, R. C. (1998). Brownian dynamics simulation of protein–protein diffusional encounter. *Methods*, **14**, 329–341.
58. Ponder, J. W. & Case, D. A. (2003). Force fields for protein simulations. *Adv. Protein Chem.* **66**, 27–85.
59. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M. *et al.* (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688.
60. Zhou, H. & Zhou, Y. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.* **11**, 2714–2726.