

**Hannes Simader, Michael  
Hothorn and Dietrich Suck\***Structural and Computational Biology Unit,  
European Molecular Biology Laboratory (EMBL),  
Meyerhofstrasse 1, D-69117 Heidelberg,  
GermanyCorrespondence e-mail: [suck@embl.de](mailto:suck@embl.de)

# Structures of the interacting domains from yeast glutamyl-tRNA synthetase and tRNA-aminoacylation and nuclear-export cofactor Arc1p reveal a novel function for an old fold

Eukaryotic aminoacyl-tRNA synthetases (aaRS) frequently contain additional appended domains that are absent from their prokaryotic counterparts which mediate complex formation between eukaryotic aaRS and cofactors of aminoacylation and translation. However, the structural basis of such interactions has remained elusive. The heteromerization domain of yeast glutamyl-tRNA synthetase (GluRS) has been cloned, expressed, purified and crystallized in space group  $C222_1$ , with unit-cell parameters  $a = 52$ ,  $b = 107$ ,  $c = 168$  Å. Phase information was obtained from multiple-wavelength anomalous dispersion with selenomethionine to 2.5 Å resolution and the structure, comprising two monomers per asymmetric unit, was determined and refined to 1.9 Å resolution. The structure of the interacting domain of its accessory protein Arc1p was determined and refined to 1.9 Å resolution in a crystal form containing 20 monomers organized in five tetramers per asymmetric unit (space group  $C2$ , unit-cell parameters  $a = 222$ ,  $b = 89$ ,  $c = 127$  Å,  $\beta = 99.4^\circ$ ). Both domains adopt a GST-like fold, demonstrating a novel role for this fold as a protein–protein interaction module.

Received 27 July 2006

Accepted 28 September 2006

**PDB References:** yeast GluRS heteromerization domain, 2hra, r2hrasf; yeast Arc1p heteromerization domain, 2hqt, r2hqtsf.

## 1. Introduction

Aminoacyl-tRNA synthetases (aaRS) catalyze the highly specific charging of their cognate tRNAs with the corresponding amino acid. This reaction establishes the link between the anticodon, *i.e.* the nucleotide code, and the amino acid attached to the 3' end of the tRNA, *i.e.* the protein code. Therefore, aaRS may be considered to be the real translators of the genetic code. In contrast to their prokaryotic counterparts, eukaryotic aaRS are integrated into an efficient tRNA nuclear-export and shuttling machinery and are interconnected closely with the translational apparatus. While prokaryotic aaRS typically function as monomers or dimers, eukaryotic enzymes frequently function as part of higher order complexes. This functional requirement is reflected by the presence of additional protein-binding and RNA-binding domains and a correspondingly higher degree of complex formation among eukaryotic aaRS (Mirande, 1991; Kisselev & Wolfson, 1994; Yang, 1996). The basic enzymatic function of prokaryotic and eukaryotic aaRS is structurally very well understood (Martinis *et al.*, 1999; Woese *et al.*, 2000) and a wealth of structural information is available for these enzymes. However, despite many reported interactions between eukaryotic aaRS and protein cofactors (Negrutskii *et al.*, 1999; Quevillon *et al.*, 1999; Kim *et al.*, 2000; Deinert *et al.*, 2001), a structural model of a protein–protein interaction involving a eukaryotic aaRS has not been reported.

In higher eukaryotes, nine aminoacyl-tRNA synthetases, including MetRS and GluProRS, associate to form a supra-

molecular multi-enzyme complex (Mirande *et al.*, 1982; Kerjan *et al.*, 1994; Lee *et al.*, 2004). In addition to the enzymes, this complex contains three associated proteins, p18, p38 and p43 (Quevillon & Mirande, 1996; Quevillon *et al.*, 1997; Kim *et al.*, 2002). In the yeast *Saccharomyces cerevisiae*, an evolutionary intermediate aaRS complex is formed by MetRS, GluRS and the tRNA-aminoacylation and nuclear-export cofactor Arc1p, which shares sequence homology with all three accessory components of the multi-synthetase complex of higher eukaryotes (Simos *et al.*, 1996). The yeast aaRS complex is a biochemically well characterized model system. Both synthetases and both tRNA<sup>Glu</sup> and tRNA<sup>Met</sup> can be efficiently copurified with protein A-tagged Arc1p from cellular yeast lysates (Simos *et al.*, 1996). The N-terminal appended domains of the three proteins are necessary and sufficient for formation of the MetRS–Arc1p–GluRS ternary complex, which appears to be stoichiometric (Simos *et al.*, 1998; Deinert *et al.*, 2001; Galani *et al.*, 2001). The interaction domain of Arc1p contains binding sites for MetRS and GluRS which enable it to bind both synthetases simultaneously and independently (Simos *et al.*, 1998). The N-terminal appended interaction domains of the synthetases are non-catalytic and are absent in the prokaryotic counterparts. MetRS cannot interact with GluRS in the absence of Arc1p, demonstrating that the association is mediated by Arc1p (Simos *et al.*, 1998; Galani *et al.*, 2001).

We have chosen the N-terminally appended heteromerization domains of yeast glutamyl-tRNA synthetase (subsequently referred to as GluRS-N) and the tRNA-aminoacylation and nuclear-export cofactor Arc1p (subsequently referred to as Arc1p-N) as a well established model system to structurally characterize the two interacting domains.

## 2. Materials and methods

### 2.1. Cloning and expression of GluRS-N

Residues 1–197, 17–207 and 1–207 of *S. cerevisiae* cytoplasmic glutamyl-tRNA synthetase (EC 6.1.1.17; GeneBank accession P46655) were PCR-amplified from plasmid pEMBLyex4-His8-GluRS (Deinert *et al.*, 2001) using primers Fw1 (TATAGAATTCCCGCTCCATATGACGAAACTATTTC AAAGGTTAAGG), Fw17 (TATAGAATTCCCGCTCCATATGCCATCTACCTTGACTATTAATGG), Rv197 (CGCTCTAGATTACTTACCCACATTAGCACTTTTCTT) and Rv207 (CGCTCTAGATTATTCAAAGTTAGCCTTGTA-GTTTC). PCR products were restricted with *FauI* and *XbaI* (New England Biolabs) and ligated into an *NdeI*- and *XbaI*-restricted modified pETm (Novagen) vector. The expression construct allows the removal of the N-terminal six-His tag through TEV protease cleavage, leaving a Gly-His dipeptide N-terminal of the natural starting methionine. The sequence of the expression constructs was confirmed by sequencing with T7 primer. The constructs were transformed into *Escherichia coli* strain BL21(DE3) Star cells, which were grown at 310 K in 6 l TB medium containing 100 µg ml<sup>-1</sup> carbenicillin (Sigma) to an OD<sub>600</sub> of 0.6. Upon induction of recombinant protein expression with 0.1 mM IPTG, the incubation temperature

was lowered to 291 K and the cells were grown for another 12 h.

### 2.2. Purification of GluRS-N and interaction analysis

Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris, 150 mM NaCl, 20 mM imidazole, 5 mM MgCl<sub>2</sub>, 10% glycerol, 10 mM β-mercaptoethanol, pH 7.5 with HCl) at 10 ml buffer per gram of wet cell paste and lysed with an Emulsiflex-C5 (Avestin, Canada). The lysate was cleared by centrifugation at 30 000g for 20 min and the supernatant was loaded onto nickel–nitrilotriacetic acid Superflow resin (Qiagen). The resin was washed with lysis buffer containing 1 M NaCl, followed by elution of the proteins with a linear gradient of 20–400 mM imidazole in lysis buffer. Protein peak fractions (typically >95% pure) were pooled, the buffer was exchanged to TEV protease buffer (50 mM Tris, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, 5 mM β-mercaptoethanol, pH 8.0 with HCl) using a Sephadex G-25 column and the sample was incubated with TEV protease in a 50:1 molar ratio at room temperature overnight for removal of the six-His tag. TEV protease, impurities and any remaining uncleaved six-His-tagged protein was removed by re-adsorption to nickel–nitrilotriacetic acid resin. The flowthrough was concentrated to 20 mg ml<sup>-1</sup> using centrifugal filter devices (Centricon) and further purified on a 16/60 Superdex-75 column (Pharmacia) equilibrated and run in gel-filtration buffer (20 mM HEPES, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, pH 7.2 with NaOH). Pure peak fractions from the gel filtration were pooled, concentrated to 20 mg ml<sup>-1</sup> and stored at 193 K for crystallization. Selenomethionine-substituted protein was expressed in *E. coli* B834 (DE3) and purified following the same basic protocol. Full-length GluRS is known to be monomeric in solution under physiological conditions (Deinert *et al.*, 2001). The monomeric state of all three N-terminal fragments of GluRS was confirmed by dynamic light scattering and gel filtration. The ability of all three fragments to form a stable complex with Arc1p was confirmed by copurification with Arc1p in gel filtration on a 16/60 Superdex-200 column (results not shown). Arc1p was purified for this interaction analysis as described in Deinert *et al.* (2001).

### 2.3. Crystallization of GluRS-N

Crystallization trials were set up at 293 K in 96-well format using a Mosquito nanolitre pipetting robot (TTP Labtech). Sitting drops of 600 nl (1:1 mixture of 20 mg ml<sup>-1</sup> protein in gel-filtration buffer and reservoir solution) were equilibrated by vapour diffusion against reservoirs containing 75 µl screen solution in Crystal Quick plates (Greiner Bio-One). 15 screens of 96 conditions each were tested with each GluRS fragment. While GluRS<sub>1–197</sub> and GluRS<sub>17–207</sub> did not yield crystalline material in any of the 1440 conditions tested, GluRS<sub>1–207</sub> crystallized under several closely related conditions from a commercial ammonium sulfate screen (Nextal). These crystals were optimized using microseeding in 24-well plates by vapour diffusion from hanging drops composed of equal volumes (typically 2 + 2 µl) of protein and reservoir solution equi-

**Table 1**

Data-collection statistics for GluRS<sub>1</sub>–207 selenomethionine multiple-wavelength anomalous dispersion data.

Values in parentheses are for the highest resolution shell.

Data statistics	Peak	Edge	Remote
Space group	C222 <sub>1</sub>		
Unit-cell parameters (Å)	$a = 52.27, b = 107.25, c = 168.85$		
Wavelength (Å)	0.97912	0.97934	0.95372
Resolution range (Å)	50–2.5 (2.7–2.5)		
Observed reflections	223191	166396	171682
Unique reflections	31000	30699	31028
Redundancy	7.2 (5.1)	5.4 (3.9)	5.5 (4.4)
Completeness (%)	97.8 (89.4)	96.8 (84.7)	97.9 (89.7)
$R_{\text{sym}}^{\dagger}$ (%)	7.2 (11.4)	7.1 (13.1)	7.3 (14.9)
Mosaicity (°)	0.103	0.104	0.105
Average $I/\sigma(I)$	20.77 (12.07)	17.13 (8.63)	16.95 (8.54)
Anomalous signal (asn) <sup>‡</sup>	1.09	1.02	1.02

<sup>†</sup>  $R_{\text{sym}} = \sum |I(h, i) - I(h, i)| / \sum I(h, i)$ . <sup>‡</sup> Anomalous signal-to-noise ratio as the mean  $\sigma(I)$  of acentric reflections assuming Friedel's law to be true divided by the mean  $\sigma(I)$  of acentric reflections assuming Friedel's law to be false.

**Table 2**

Data-collection statistics for Arc1p<sub>1</sub>–122 and GluRS<sub>1</sub>–207 native data.

Values in parentheses refer to the highest resolution shell.

Data statistics	Arc1p-N native	GluRS-N native
Space group	C2	C222 <sub>1</sub>
Unit-cell parameters (Å, °)	$a = 222.3, b = 89.5,$ $c = 126.8, \beta = 99.4$	$a = 52.0, b = 107.1,$ $c = 167.9$
Wavelength (Å)	0.95372	0.97925
Resolution range	50–1.90 (2.00–1.90)	50–1.90 (2.00–1.90)
Observed reflections	811593	176684
Unique reflections	187178	37390
Redundancy	4.3 (3.9)	4.4 (4.4)
Completeness (%)	97.0 (96.5)	99.5 (97.4)
$R_{\text{sym}}^{\dagger}$ (%)	4.1 (44.1)	7.3 (35.8)
Mosaicity (°)	0.218	0.174
Average $I/\sigma(I)$	18.5 (3.8)	13.2 (4.2)

<sup>†</sup>  $R_{\text{sym}} = \sum |I(h, i) - I(h, i)| / \sum I(h, i)$ .

brated against 1 ml reservoir solution [1.7–1.8 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 200 mM KSCN for selenomethionine-substituted crystals and 1.8–1.9 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 200 mM NaI for native crystals]. The optimized condition yielded crystals suitable for X-ray analysis (see supplementary material<sup>1</sup>).

#### 2.4. X-ray data collection

Crystals of GluRS-N were cryoprotected by sequential transfer to cryosolution [1.9 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 200 mM NaI or KSCN, 150 mM NaCl, 20 mM HEPES, 5 mM MgCl<sub>2</sub>, 1 mM DTT] with stepwise addition of glycerol to 15% prior to freezing in liquid nitrogen. Multiple-wavelength anomalous dispersion data were collected at 100 K from one selenomethionine-substituted crystal at the peak and inflection-point wavelength of selenium and at a high remote wavelength (Table 1) at the Swiss Light Source beamline PX1. Native data from a GluRS-N crystal frozen at 100 K were collected at the

<sup>1</sup> Supplementary material has been deposited in the IUCr electronic archive (Reference: HV5066). Services for accessing this material are described at the back of the journal.

**Table 3**

Refinement statistics for Arc1p<sub>1</sub>–122 and GluRS<sub>1</sub>–207.

Values in parentheses refer to the highest resolution shell.

Refinement statistics	Arc1p-N	GluRS-N
Resolution range (Å)	50–1.90 (1.949–1.900)	50–1.90 (1.949–1.900)
No. of reflections used	177795 (13032)	35496 (2580)
$R_{\text{cryst}}^{\dagger}$ (%)	20.9 (25.7)	21.4 (27.2)
$R_{\text{free}}^{\dagger}$ (5% test set) (%)	26.2 (30.7)	26.2 (31.3)
Asymmetric unit content		
Chains	20	2
Residues	2480	418
$V_M$ (Å <sup>3</sup> Da <sup>-1</sup> )	2.2	2.5
Solvent content (%)	44.1	51.3
No. of atoms		
Protein	18582	2731
Water	1397	302
Ligand	50	46
R.m.s. deviations		
Bond length (Å)	0.015	0.014
Angles (°)	1.48	1.47
Mean $B$ factor by atom type (Å <sup>2</sup> )		
Protein	46.1	31.4
Water	50.7	49.0
Ligand	66.4	50.2
Ramachandran plot <sup>‡</sup>		
Most favoured (%)	93.8	89.3
Allowed (%)	5.5	9.1
Generously allowed (%)	0.5	0.6
Disallowed (%)	0.2	1.0

<sup>†</sup> As defined in *REFMAC5* (Murshudov *et al.*, 1997). <sup>‡</sup> As calculated by *PROCHECK* (Laskowski *et al.*, 1993).

European Synchrotron Radiation Facility beamline ID-23. Native data from an Arc1p-N crystal frozen at 100 K were collected at the Swiss Light Source beamline PX1. Data reduction and scaling were performed with *XDS*. Data-collection statistics are given in Tables 1 and 2.

#### 2.5. Structure solution of GluRS-N

In order to solve the phase problem for this structure, we decided to attempt experimental phasing by the multiple-wavelength anomalous dispersion (MAD) method using selenomethionine-substituted crystals (Table 1) since molecular replacement was clearly not an option owing to the lack of any structures that were significantly related to GluRS-N by sequence homology. *SHELXC* (Sheldrick, 1990) was used to analyse and prepare the MAD data set for a heavy-atom site search using *SHELXD*. The heavy-atom site search with *SHELXD* (Sheldrick, 1990) on peak, inflection and high remote data sets in the resolution range 20–3.5 Å revealed six of the eight possible Se atoms in the asymmetric unit. Refinement of heavy-atom sites and calculation of the initial phases were performed with *SOLVE* (Terwilliger & Berendzen, 1999) and density modification and phase extension to the 2.5 Å high-resolution limit of the MAD data set were completed with *RESOLVE* (Terwilliger, 2000). The figure of merit for phasing was 0.76 for initial phases to 3.5 Å, 0.86 after density modification at 3.5 Å and 0.64 after extending the phases to 2.5 Å.

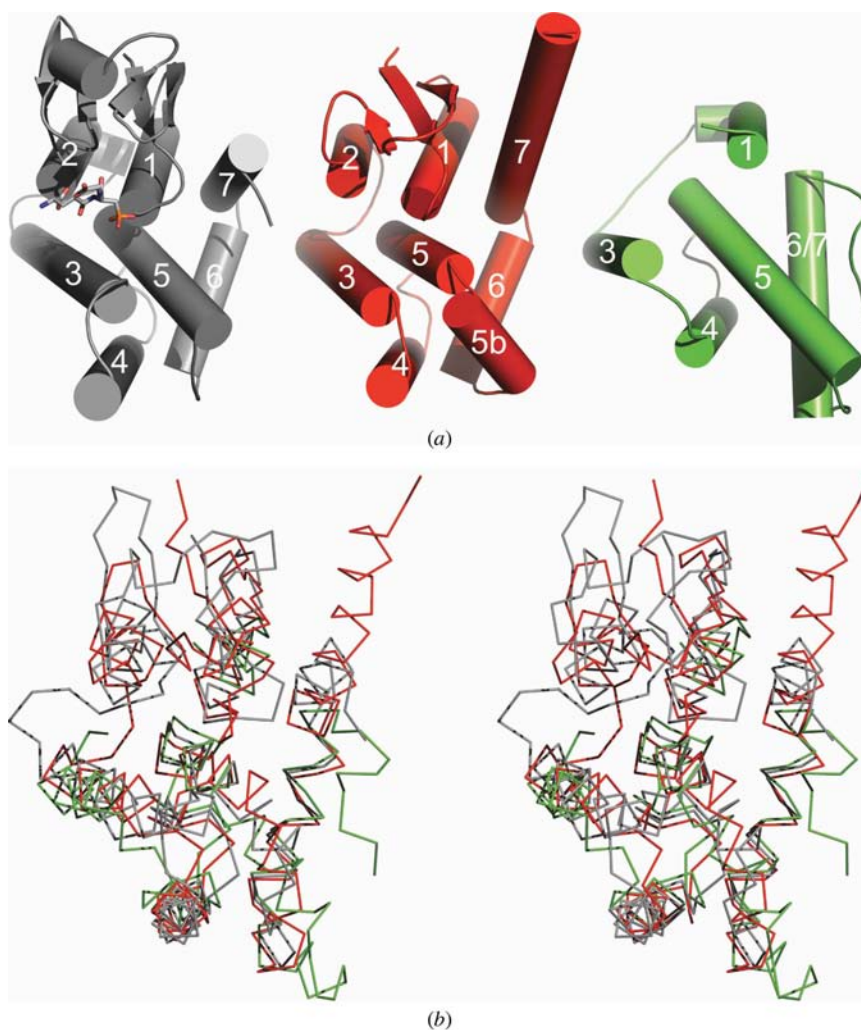
## 2.6. Model building and refinement of GluRS-N

The structure was built manually in *Coot* (Emsley & Cowtan, 2004). During refinement against the 1.9 Å native data set, 46 iodines were added to the structure based on difference-map analysis as NaI was absent from the crystallization solution of the selenomethionine-substituted crystals but was present at 200 mM in the crystallization solution of the native crystals. Refinement against the 1.9 Å native data set reached convergence (Table 3) after a few alternating cycles of model correction in *Coot* and restrained refinement allowing the anisotropic motion of rigid bodies, described as TLS parameters (Winn *et al.*, 2001), as implemented in *REFMAC5* (Murshudov *et al.*, 1997). Water molecules were automatically added in *Coot* and the occupancy of the iodines was manually adjusted based on difference-map analysis. Inspection of the refined model with *PROCHECK* (Laskowski *et al.*, 1993) revealed good stereochemistry (Table 3). The final refined model includes 189 amino-acid residues in both of the monomers in the asymmetric unit. Residues 1–16 and 197–207 of chain *A* and 1–13 and 194–207 of chain *B* were unidentified in the electron-density map and therefore molecular models for these segments are not present in the final model. Statistics for phasing and refinement are summarized in Tables 2 and 3.

## 2.7. Phasing, model building and refinement of Arc1p-N

Expression, purification, crystallization, X-ray diffraction analysis and preliminary multiple-wavelength anomalous dispersion phasing of the Arc1p N-terminal heteromerization domain has been described in Simader & Suck (2006). While the initial map was not interpretable for model building, fivefold NCS averaging over the five tetramers present in the asymmetric unit as implemented in *RESOLVE* (Terwilliger, 2000) subsequently produced a traceable map. The figure of merit for phasing by *RESOLVE* was 0.53 for initial phases to 3.5 Å, 0.65 after density modification at 3.5 Å and 0.59 after extending the phases to the 2.8 Å high-resolution limit of the selenomethionine MAD data set. The structure was built using the program *O* (Jones *et al.*, 1991) by manually building one tetramer and then using *MOLREP* (Vagin & Teplyakov, 1997) to automatically place the remaining four tetramers of the asymmetric unit into the phased electron-density map from *RESOLVE*. Strict NCS restraints between the five tetramers of the asymmetric unit as implemented in *REFMAC* were maintained during the initial stages of refinement and

rebuilding in *O* (Jones *et al.*, 1991). Ten sulfate molecules were located and identified by manual inspection of the difference electron-density maps and added manually. NCS restraints were progressively relaxed and replaced by TLS restrained refinement as implemented in *REFMAC* during the later stages of refinement. *Coot* was used for manual model correction and analysis during the later stages of refinement and to automatically add water molecules. Inspection of the refined model with *PROCHECK* revealed excellent stereochemistry (Table 3). The 20 monomers of the final refined model differ somewhat in respect to the amino-acid residues that could be built in each chain. On average, 115 of 124 residues were built per monomer, with two to four residues missing from each terminus. Residues 14–18 from the first chain of each tetramer (corresponding to chains *A*, *E*, *I*, *M* and *Q*) as well as 16–22 from chain *C*, 22–31 from chain *H* and 19–21 from chain *K* could not be identified in the electron-density maps and these segments are therefore missing in the final model. Statistics for refinement are summarized in Table 3.



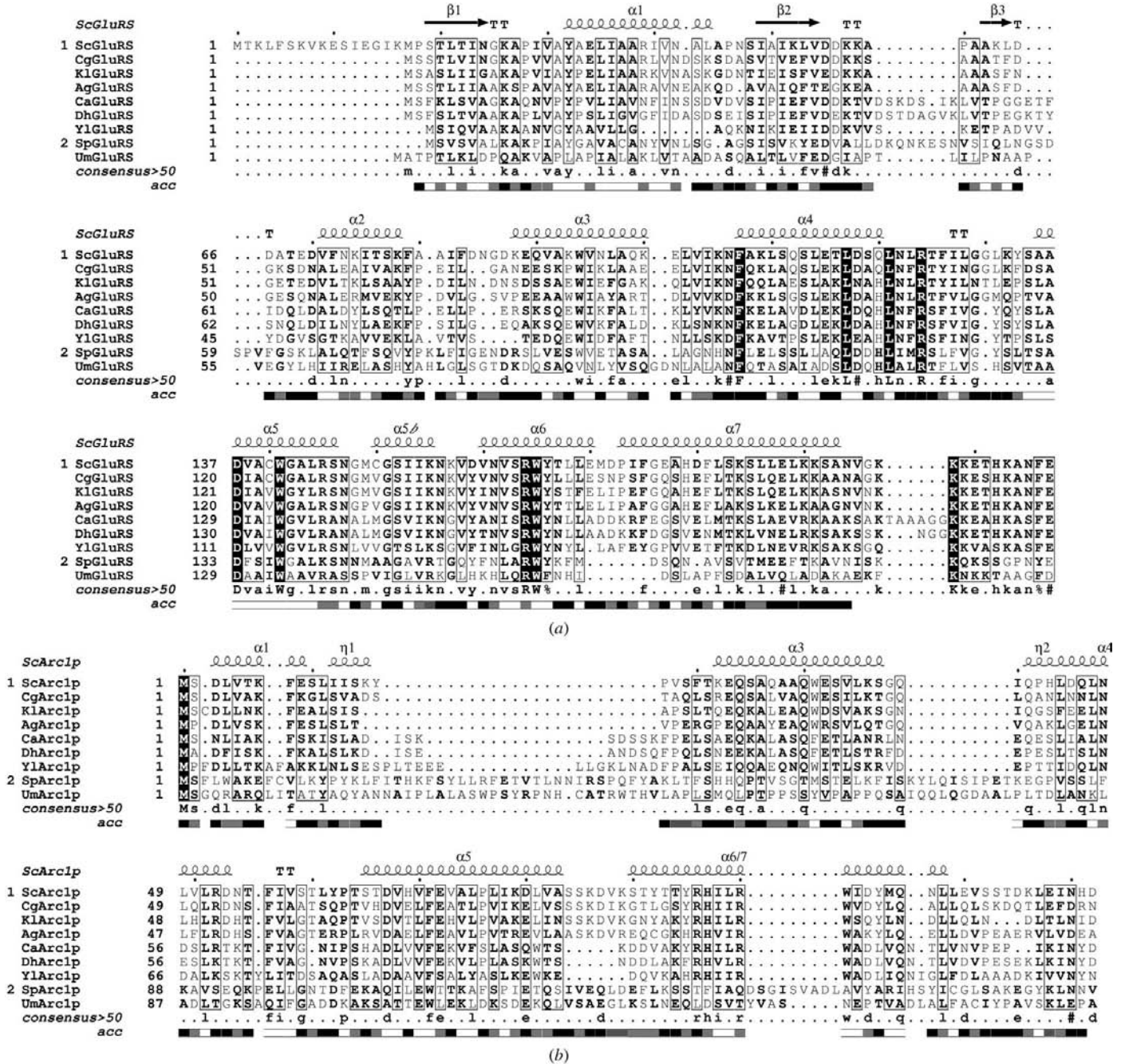
**Figure 1**

GluRS-N and Arc1p-N both adopt a GST-like fold. Shown are *E. coli* GST (grey; PDB code 1a0f), GluRS-N (red) and Arc1p-N (green). (a) Cartoon representations of *E. coli* GST, GluRS-N and Arc1p-N viewed in identical orientations, with the corresponding  $\alpha$ -helices of the GST domain labelled 1–7. (b) Stereo representation of the superimposed C $\alpha$  traces.

2.8. Miscellaneous

*PyMOL* (DeLano, 2006) was used to render structural models for figure preparation. *ClustalW* (Thompson *et al.*, 1994; Chenna *et al.*, 2003) was used to calculate multiple sequence alignments and for phylogenetic distance analysis. Multiple sequence-alignment figures were prepared with *ESPrpt* (Gouet *et al.*, 1999) applying the Risler scoring matrix

(Risler *et al.*, 1988) with a global similarity-score cutoff of 0.6, using consensus criteria from *MULTALIN* (Corpet, 1988) and using secondary-structure and accessibility information calculated using *DSSP* (Kabsch & Sander, 1983). The *DALI* server (Holm & Sander, 1993, 1998) was used to identify proteins that were structurally related to GluRS-N and Arc1p-N. We used automated error-scaled difference distance-



**Figure 2** Multiple sequence alignment of *S. cerevisiae* GluRS-N (a) and Arc1p-N (b) with orthologous sequences from other yeasts. Shown are the sequences from *Yarrowia lipolytica* (Yl), *Debaryomyces hansenii* (Dh), *Candida albicans* (Ca), *Kluyveromyces lactis* (Kl), *Ashbya gossypii* (Ag), *Saccharomyces cerevisiae* (Sc) (all group 1), *Ustilago maydis* (Um) and *Schizosaccharomyces pombe* (Sp) (group 2). Identical residues are shown as white characters in a black box, residues conserved among the sequences of a group are printed in bold and residues conserved among all sequences are shown in a frame. Secondary-structure elements observed in the reported structures of *S. cerevisiae* Arc1p-N and GluRS-N are represented graphically above the alignment, the consensus sequence is shown below the sequences and the accessibility of residues in the structures is represented schematically by the bar labelled 'acc' below the alignment, with accessible residues shown in black, buried residues in white and intermediate residues in grey.

matrix analysis of the 20 individual monomers contained in the asymmetric unit of the Arc1p-N structure as implemented in *ESCKET* (Schneider, 2000, 2002, 2004) in order to cluster the monomers by structural similarity at different  $\sigma$  levels and to distinguish between conformationally invariant and flexible regions of Arc1p-N. Clusters of highly similar conformers (with matrix elements larger than 0.98 at the  $2\sigma$  level in *ESCKET*) were reduced to a single representative with the minimal experimental error (mean e.s.d. in *ESCKET*). The remaining five conformers were kept for analysis of invariant and flexible regions in *ESCKET*. Protein interfaces and assemblies were analyzed using the protein interfaces, surfaces and assemblies service PISA at the European Bioinformatics Institute (Krissinel & Henrick, 2005).

### 3. Results and discussion

#### 3.1. Structure of GluRS-N

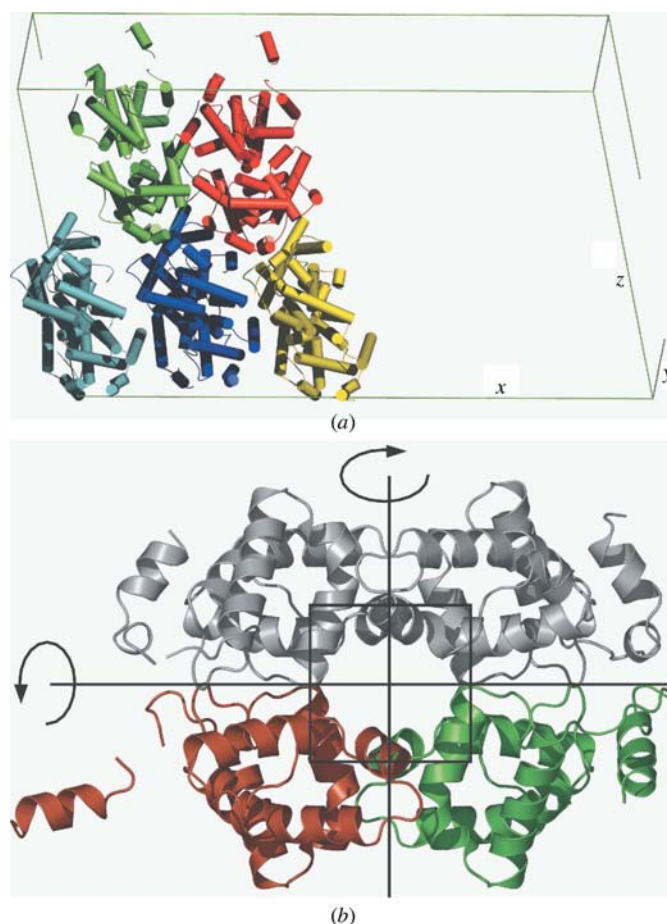
The N-termini of Arc1p and GluRS were originally described as domains with little or no homology to known proteins (Simos *et al.*, 1996, 1998). A relationship to glutathione *S*-transferase (GST) like proteins was subsequently predicted on the basis of conserved sequence motifs (Galani *et al.*, 2001). Indeed, the crystal structure of GluRS-N confirms that it adopts a GST-like fold (Fig. 1). The two monomers of GluRS-N contained in the asymmetric unit are conformationally almost identical, as suggested by analysis with *ESCKET* (Schneider, 2000, 2002, 2004; results not shown). The r.m.s.d. for superposition of GluRS-N chain *B* on chain *A* of *E. coli* GST (PDB code 1a0f; Nishida *et al.*, 1998) is 2.04 Å for 145 C $\alpha$  atoms with 17% sequence identity. However, the cleft formed between the N-terminal thioredoxin-like  $\alpha/\beta$  motif (GST\_N in PFAM; Bateman *et al.*, 2004) and the C-terminal  $\alpha$ -helical part of the classical GST fold (GST\_C in PFAM), which is normally the substrate-binding cleft of enzymatically active GSTs, lacks residues suitable for substrate binding and does not bind to glutathione Sepharose (results not shown).

The open reading frame for *S. cerevisiae* GluRS encodes 16 N-terminal amino acids followed by a second methionine which are not contained in any homologue (Fig. 2). This has led to disagreement about which methionine represents the real start of translation; consequently, two alternative versions are found in the databases, one including the first 16 codons (*e.g.* SwissProt accession P46655) and one starting with the second methionine (*e.g.* the SGD entry for GluRS, GUS1). There does not appear to be any experimental evidence available to support one or the other. We have therefore tried to express, purify and crystallize different versions of GluRS-N with and without the nonconserved first 16 amino acids. While all three fragments could be expressed in soluble form, the levels of solubly expressed residues 17–207 were much lower than those of residues 1–197 and 1–207, suggesting that the first 16 amino acids may play a role in the folding or solubility of GluRS. However, neither residues 1–197 nor 17–207 could be crystallized, while crystals of residues 1–207 were readily obtained under several conditions from the initial

screen, suggesting involvement of the first 16 and last ten amino acids in crystal lattice contacts. Surprisingly, the structural model of residues 1–207 reveals that the first 16 and the last 12 amino acids are not ordered in both monomers of the asymmetric unit of this crystal form: there is hardly any interpretable electron density N-terminal of the second Met17 or C-terminal of Asn194 in either monomer. While it still remains unclear why the shorter constructs could not be crystallized, the structure demonstrates that the basic GST fold of the GluRS N-terminus is complete without the first 16 residues.

#### 3.2. Unusual noncrystallographic symmetry in the structure of Arc1p-N

The structure of Arc1p-N comprises 20 monomers organized into five tetramers per asymmetric unit (Fig. 3). Given that Arc1p-N behaves as a monomer in solution (Deinert *et al.*, 2001 and data not shown), the structure of Arc1p-N is



**Figure 3**  
The structure of Arc1p-N contains 20 monomers organized in five tetramers per asymmetric unit. (a) The arrangement of the five tetramers (No. 1 in blue, 2 in green, 3 in red, 4 in cyan and 5 in yellow) in the unit cell. (b) The monomers within each tetramer are related by pseudotwofold rotational symmetry and fall into three conformers (No. 1 in brown, 2 in green and 3 in grey) that differ by the orientation of their N-terminal  $\alpha$ -helix. The central frame in (b) indicates the section shown in detail in the supplementary material.

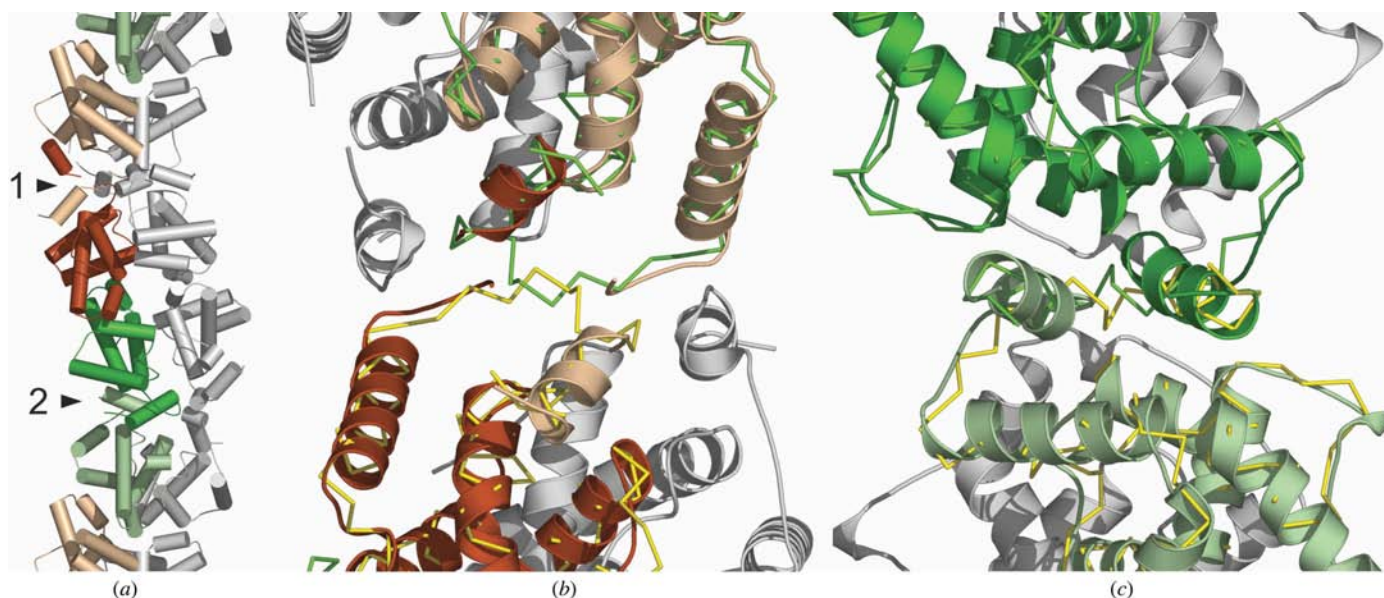
outstanding in respect to its very high noncrystallographic symmetry. The five tetramers (Fig. 3) are almost identical to each other (the average r.m.s.d. for pairwise superpositions is 0.60 Å), but their arrangement within the asymmetric unit is slightly irregular. They all differ in their relative orientation, with angular differences of between 4 and 14° in the *xy* plane of the crystal lattice (not shown), explaining why the data could not be indexed and processed in a crystal system with a smaller asymmetric unit.

Within each tetramer, the individual monomers are related by pseudo-twofold rotational symmetry axes as indicated in Fig. 3(b). Two sulfates are found at a distance of about 18.5 Å in the central part of each tetramer. They are coordinated by four arginines (Arg54 from each of two monomers and Arg98 from each of the two other monomers of the tetramer) and two lysine side chains (Lys91 from the same two monomers that contribute Arg98; see supplementary material). This arrangement suggests that the sulfates may play an important role in the stabilization of the tetramers by partially neutralizing the concentrated positive charges of the coordinating residues, which would otherwise create strong electrostatic repulsive forces that might prohibit tetramer formation. Interestingly, the arginines involved in this interaction are well conserved within most of the orthologous sequences (group 1 in Fig. 2b), suggesting a biological function for these residues. However, analysis of protein interfaces within each tetramer by PISA (not shown) also did not reveal any specific interactions within a tetramer that could result in the formation of stable quaternary structures. This is in accordance with the finding that Arc1p-N behaves as a monomer in solution (Deinert *et al.*, 2001 and results not shown). Thus, the arrangement of Arc1p-N monomers into tetramers as

observed in our structure is most likely to be an artefact of crystallization.

### 3.3. Alternative conformations of the N-terminal 22 residues in Arc1p-N are determined by crystal-packing constraints

Clustering of the Arc1p-N monomers by structural variance at the 3 $\sigma$  significance level in *ESCKET* (Schneider, 2000, 2002, 2004) reveals three major conformational variants corresponding to the three alternative conformations of the N-terminal 22 residues as suggested by manual analysis. The different conformations are adopted by corresponding monomers within each tetramer: conformation 1 (brown monomer in Figs. 3b and 4) is adopted by the first chain of each tetramer (chains *A, E, I, M, Q*), conformation 2 (green monomer in Figs. 3b and 4) is adopted by the second chain (chains *B, F, J, N, R*) and conformation 3 (grey monomers in Figs. 3b and 4) is adopted by the third and fourth chain of each tetramer (chains *C, D, G, H, K, L, O, P, S, T*). Analysis of the crystal-packing environment of this N-terminal segment reveals why it adopts three different conformations and suggests that conformer 3 represents the default conformation which is adopted in the absence of packing constraints. In conformers 1 and 2, this segment is deeply involved in crystal contacts between neighbouring tetramers stacking on top of each other along the *z* axis (Fig. 4a). Both conformations 1 and 2 in fact represent a 'mini' domain swap of the N-terminal segment with that of the corresponding monomer from the next tetramer (Figs. 4b and 4c). In the case of conformer 1 (Fig. 4b), this domain swap is not clearly defined in our structure as the residues connecting the domain-swapped N-terminal segment to the remainder of the monomer are not



**Figure 4**

Alternative conformations of the N-terminal segment of Arc1p-N are determined by crystal-packing constraints. Tetramers stack on top of each other along the *z* axis (a). Conformation 1 (brown) (b) and 2 (green) (c) represent two different ways of domain swapping of the N-terminal  $\alpha$ -helix between conformationally corresponding monomers from neighbouring tetramers. Conformation 3 (grey) represents the 'default' conformation adopted in the absence of packing constraints. The 'default' conformation 3 is shown as a yellow and green C $\alpha$  trace superimposed on the conformers 1 in (b) and 2 in (c). See text for explanation.

ordered and could not be built for any of the five tetramers in the asymmetric unit. However, the  $F_o - F_c$  and  $2F_o - F_c$  maps suggest this connectivity rather than that corresponding to conformer 3 and superposition of the putative 'default' conformer 3 (green and yellow  $C^\alpha$  traces in Figs. 4*b* and 4*c*) on the two domain-swapped conformer 1 monomers (brown and light brown in Fig. 4*b*) reveals that the crystal packing is sterically incompatible with conformation 3 in this region (clashes between the yellow and green  $C^\alpha$  traces in Fig. 4*b*). In the case of conformer 2 (Fig. 4*c*), this domain swap is clearly defined in the structure by continuous electron density at the  $1.7\sigma$  level of the  $2F_o - F_c$  map for all connecting residues in all five monomers corresponding to this conformer. As for conformer 1, superposition of the putative 'default' conformer 3 (green and yellow  $C^\alpha$  traces in Figs. 4*b* and 4*c*) on the two domain-swapped conformer 2 monomers (green and light green in Fig. 4*c*) reveals that the crystal packing is sterically incompatible with conformation 3 in place of conformation 2 (clashes between the yellow and green  $C^\alpha$  traces in Fig. 4*b*). Thus, the alternative conformations observed for the N-terminal 22 amino acids correspond to two different ways of domain-swapping of the first  $\alpha$ -helix of two Arc1p-N monomers and are fully explained by crystal-packing constraints.

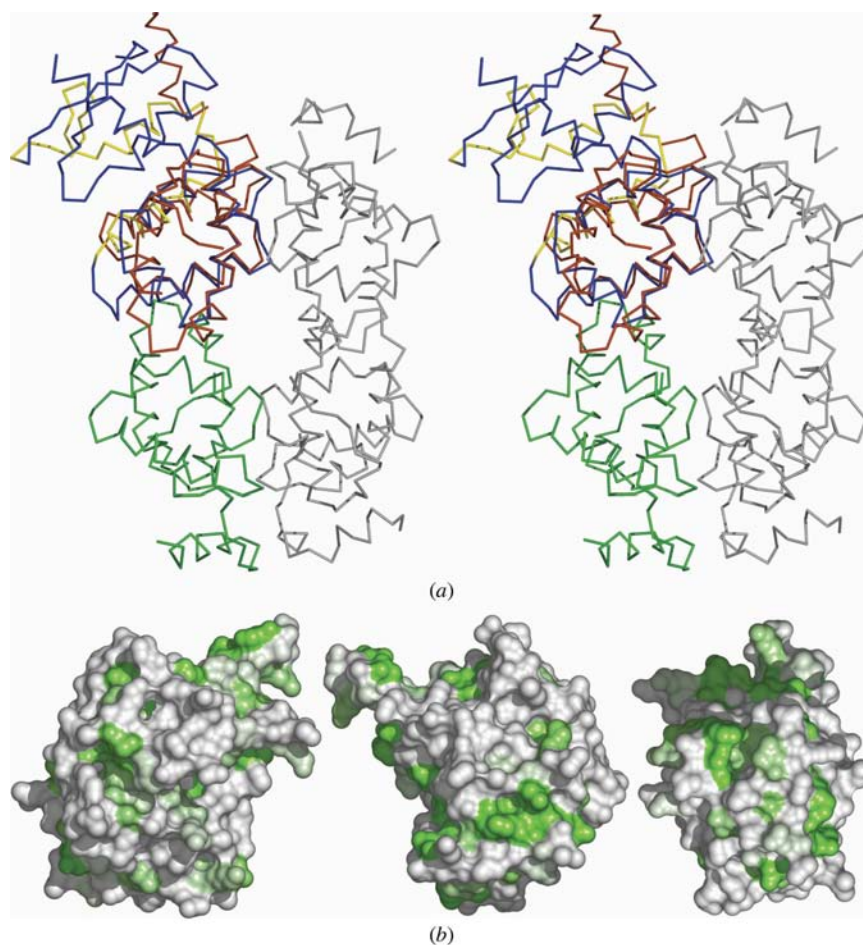
### 3.4. Structure of the Arc1p-N monomer

As for GluRS-N, the crystal structure of Arc1p-N confirms that the isolated monomer adopts a fold similar to the GST fold (Fig. 1). Weak similarity to GSTs is reflected by the r.m.s.d. for superposition of Arc1p1-N chain *D* on chain *A* of *E. coli* GST (PDB code 1a0f; Nishida *et al.*, 1998), which is 2.71 Å for 92  $C^\alpha$  atoms at 18.5% sequence identity. In contrast to GluRS-N, the Arc1p N-terminal domain corresponds to only the C-terminal  $\alpha$ -helical part of the classical GST fold (GST\_C in PFAM; Bateman *et al.*, 2004) and thus lacks the N-terminal  $\alpha/\beta$  motif, which contributes most of the substrate-binding residues in classical GSTs. As a consequence, the substrate-binding cleft of classical GSTs is missing and Arc1p cannot be enzymatically active as a GST. The multiple sequence alignment of Arc1p-N with orthologous sequences from other yeasts (Fig. 2*b*) suggests that the N-terminal  $\alpha/\beta$  motif is still contained in the sequences from the phylogenetic group 2 (*Ustilago maydis* and *Schizosaccharomyces pombe*). This is supported by secondary-structure prediction, fold-recognition and threading approaches (results not shown), suggesting that the truncation of the Arc1p-N fold

occurred in the time between the phylogenetic separation of groups 1 and 2.

### 3.5. Possible interactions between GluRS-N and Arc1p-N

Arc1p-N and GluRS-N are known to be necessary and sufficient for formation of the yeast aaRS complex *in vivo* and *in vitro* (Simos *et al.*, 1998; Deinert *et al.*, 2001; Galani *et al.*, 2001) and the ability of our Arc1p-N and GluRS-N proteins to form a stable complex with each other under near-physiological conditions was confirmed by copurification in gel filtration (results not shown). On the basis of the three-dimensional structures of the two isolated domains, we can now speculate about how the two domains might associate to form a binary complex.



**Figure 5**

Three suggested ways of interaction between GluRS-N and Arc1p-N. (a) Stereo representation of GluRS-N (blue/yellow) superimposed on conformer 1 (brown) of an Arc1p-N tetramer (color coding as in Fig. 3*b* and 4). Segments of GluRS-N corresponding to the interfacing elements of the *E. coli* GST homodimer are coloured yellow. (b) Analysis of GluRS-N surface conservation does not yield clues as to which area is likely to be involved in the GluRS-N–Arc1p-N interaction. Surface representations of GluRS-N are coloured in a white to green gradient corresponding to 10–90% sequence conservation (scoring by the Risler matrix; Risler *et al.*, 1988). The three views represent the three possible modes of interaction and are oriented and centred on the respective putative interface area: GST homodimer-like interaction (left), interaction corresponding to that indicated in (a) between GluRS-N and the grey Arc1p-N monomer (middle) and interaction corresponding to that indicated in (a) between GluRS-N and the green Arc1p-N monomer (right).



Both domains turned out to adopt a GST-like fold. Classical GSTs form homodimers and interact in a characteristic way that relates the monomers by twofold rotational symmetry. Therefore, Arc1p-N and GluRS-N may be expected to interact in a corresponding mode (Fig. 5*a*). However, this interaction has so far been described as a homomeric interaction only. Alternatively, GluRS-N and Arc1p-N may interact in a manner suggested by the crystal-packing contacts within a tetramer of the Arc1p-N structure (Fig. 5*a*). Superposition of GluRS-N onto a monomer of Arc1p-N within a tetramer reveals that the most closely matching segments are those involved in contacts between monomers of Arc1p within a tetramer. This is particularly true for the horizontal interaction shown in Fig. 5*a*), suggesting that GluRS and Arc1p might dock onto each other *via* this interface. The three possible interactions of GluRS-N with Arc1p-N, representing (i) the classical GST homodimerization and (ii) and (iii) the two different intermonomer contacts observed in each tetramer of Arc1p-N, would involve three distinct non-overlapping surface regions of GluRS-N.

N-terminal appendices homologous to Arc1p-N and GluRS-N are conserved in orthologous GluRS and Arc1p proteins from other yeast species (Fig. 2) and the only known biological function of this domains is to mediate complex formation between the proteins. The interfacing residues would therefore be expected to be the most conserved feature of the orthologous sequences. We therefore analysed the GluRS-N and Arc1p-N surface for patches of conserved residues in order to test the agreement between putative conserved surface patches and candidate interface regions. However, mapping of the conservation scores derived from the multiple sequence alignment of GluRS-N (Fig. 2) onto the surface of GluRS-N did not reveal a significant difference in surface conservation between the three putative interfacing surface regions (Fig. 5*b*). The result for Arc1p-N is similar (not shown), with the additional complication that Arc1p-N is known to harbour two independent interfaces for GluRS and MetRS (Simos *et al.*, 1998; Deinert *et al.*, 2001; Galani *et al.*, 2001). We have analysed the electrostatic surface potential of both proteins in a similar way (not shown), but could not identify any strong complementary electrostatic patches for the candidate interface surfaces. Whatever the real interface is, the contacting residues appear on average to be no more conserved than other surface residues not involved in the interaction. This unexpected finding raises the question whether GluRS-N and Arc1p-N may have additional previously unrecognized functions resulting in surface conservation in addition to the interfacing surface and suggests a high plasticity of the interface.

The authors are grateful to J. Basquin and A. Scholz (EMBL Heidelberg) for crystallization screen setup and would like to thank the beamline staff at the European Synchrotron Radiation Facility (Grenoble, France) ID-23 and Swiss Light Source (Villigen, Switzerland) PX-1 for their assistance during data collection. MH gratefully acknowledges

financial support from the Peter and Traudl Engelhorn Foundation, Penzberg, Germany. This work is supported by the EU grant 3D Repertoire, contract No. LSHG-CT-2005-512028.

## References

- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). *Nucleic Acids Res.* **32**, D138–D141.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). *Nucleic Acids Res.* **31**, 3497–3500.
- Corpet, F. (1988). *Nucleic Acids Res.* **16**, 10881–10890.
- Deinert, K., Fasiolo, F., Hurt, E. C. & Simos, G. (2001). *J. Biol. Chem.* **276**, 6000–6008.
- DeLano, W. L. (2006). *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Galani, K., Grosshans, H., Deinert, K., Hurt, E. C. & Simos, G. (2001). *EMBO J.* **20**, 6889–6898.
- Gouet, P., Courcelle, E., Stuart, D. I. & Metz, F. (1999). *Bioinformatics*, **15**, 305–308.
- Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
- Holm, L. & Sander, C. (1998). *Nucleic Acids Res.* **26**, 316–319.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Kerjan, P., Cerini, C., Sémériva, M. & Mirande, M. (1994). *Biochim. Biophys. Acta*, **1199**, 293–297.
- Kim, J. E., Kim, K. H., Lee, S. W., Seol, W., Shiba, K. & Kim, S. (2000). *Nucleic Acids Res.* **28**, 2866–2872.
- Kim, J. Y., Kang, Y.-S., Lee, J.-W., Kim, H. J., Ahn, Y. H., Park, H., Ko, Y.-G. & Kim, S. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 7912–7916.
- Kisselev, L. L. & Wolfson, A. D. (1994). *Prog. Nucleic Acid Res. Mol. Biol.* **48**, 83–142.
- Krissinel, E. & Henrick, K. (2005). *Computational Life Sciences*, edited by M. R. Berthold, R. Glen, K. Diederichs, O. Kohlbacher & I. Fischer, pp. 163–174. Berlin: Springer-Verlag.
- Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.
- Lee, S. W., Cho, B. H., Park, S. G. & Kim, S. (2004). *J. Cell Sci.* **117**, 3725–3734.
- Martinis, S. A., Plateau, P., Cavarelli, J. & Florentz, C. (1999). *EMBO J.* **18**, 4591–4596.
- Mirande, M. (1991). *FEBS Lett.* **283**, 1–3.
- Mirande, M., Kellermann, O. & Waller, J. P. (1982). *J. Biol. Chem.* **257**, 11049–11055.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Negrutskii, B. S., Shalak, V. F., Kerjan, P., El'skaya, A. V. & Mirande, M. (1999). *J. Biol. Chem.* **274**, 4545–4550.
- Nishida, M., Harada, S., Noguchi, S., Satow, Y., Inoue, H. & Takahashi, K. (1998). *J. Mol. Biol.* **281**, 135–147.
- Quevillon, S., Agou, F., Robinson, J. C. & Mirande, M. (1997). *J. Biol. Chem.* **272**, 32573–32579.
- Quevillon, S. & Mirande, M. (1996). *FEBS Lett.* **395**, 63–67.
- Quevillon, S., Robinson, J. C., Berthonneau, E., Siatecka, M. & Mirande, M. (1999). *J. Mol. Biol.* **285**, 183–195.
- Risler, J. L., Delorme, M. O., Delacroix, H. & Henaut, A. (1988). *J. Mol. Biol.* **204**, 1019–1029.
- Schneider, T. R. (2000). *Acta Cryst.* **D56**, 714–721.
- Schneider, T. R. (2002). *Acta Cryst.* **D58**, 195–208.
- Schneider, T. R. (2004). *Acta Cryst.* **D60**, 2269–2275.

- Sheldrick, G. M. (1990). *Acta Cryst.* **A46**, 467–473.
- Simader, H. & Suck, D. (2006). *Acta Cryst.* **F62**, 346–349.
- Simos, G., Sauer, A., Fasiolo, F. & Hurt, E. C. (1998). *Mol. Cell*, **1**, 235–242.
- Simos, G., Segref, A., Fasiolo, F., Hellmuth, K., Shevchenko, A., Mann, M. & Hurt, E. C. (1996). *EMBO J.* **15**, 5437–5448.
- Terwilliger, T. C. (2000). *Acta Cryst.* **D56**, 965–972.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 849–861.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* **D57**, 122–133.
- Woese, C. R., Olsen, G. J., Ibba, M. & Söll, D. (2000). *Microbiol. Mol. Biol. Rev.* **64**, 202–236.
- Yang, D. C. (1996). *Curr. Top. Cell. Regul.* **34**, 101–136.