Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the *Yersinia enterocolitica* molecular-chaperone protein SycE

Christine B. Trame and David B. McKay*

Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

Correspondence e-mail: dave.mckay@stanford.edu

Received 31 July 2002
Accepted 14 November 2002

PDB Reference: SycE, 1n5b, r1n5bsf.

The crystal structure of the *Yersinia enterocolitica* molecular-chaperone protein SycE, which specifically binds the YopE protein, has been solved to 2.0 Å resolution by molecular replacement. The crystal contains two SycE dimers per asymmetric unit; a novel feature of this crystal, when compared with closely related SycE structures, is a well ordered carboxy-terminal peptide in one protomer of each dimer. The peptide binds a hydrophobic patch of a neighboring molecule in a manner similar to that seen in a SycE–YopE chaperone–target complex, suggestive of low-affinity 'self-binding' through which the carboxy-terminal peptide might suppress counter-productive interactions with non-target proteins *in vivo*.

1. Introduction

Several bacterial virulence factors which are exported *via* type III secretion pathways are presented to the secretion apparatus as protein–chaperone complexes. Examples of such proteins include YopE, YopH and YopT of *Yersinia* ssp. and their cognate chaperones SycE/YerA, SycH and SycT (Hueck, 1998; Iriarte & Cornelis, 1998). The structures of the SycE proteins of *Y. pseudotuberculosis* (Birtalan & Ghosh, 2001) and *Y. pestis* (Evdokimov *et al.*, 2002) have been solved, revealing a dimeric molecule, the protomer of which has a unique fold when compared with the database of known structures. Correlation of aligned sequences of related chaperones with the SycE structure reveals the likelihood of a conserved structural core for this protein family corresponding to residues 1–120 of SycE, followed by carboxy-terminal extensions ranging from 0 to 30 residues for different representatives.

The structure of a complex between SycE and a chaperone-binding fragment of YopE ['YopE(Cb)'] has also been solved, showing a heterotrimeric complex, in which one SycE dimer binds a segment of YopE, residues ~22–78, in a relatively distended conformation (Birtalan *et al.*, 2002). The structure of a more distant chaperone–polypeptide complex, that of *Salmonella typhimurium* SicP and the chaperone-binding fragment of SptP, has also been solved; in this case, two fragments of SptP span two SicP dimers (Stebbins & Galan, 2001). Taken together, these structures of chaperones alone and chaperone–target complexes show (i) the novel fold of the protomer core of this family of chaperones, (ii) the dimeric structure of the chaperone and (iii) the specific interactions with cognate substrates, which span the chaperone dimers in distended conformations. In this context, information from additional structural work on this family of proteins can be viewed as elaborations upon this general framework.

As noted above, many of the chaperones have carboxy-terminal extensions beyond the protomer core that differ substantially in both length and sequence. In *Yersinia* SycE, this extension is approximately 10–12 residues in length; in one structure of SycE, the final two residues of the native protein plus two histidines from a carboxy-terminal His_6 tag are observed bound to a hydrophobic patch of SycE, predicated on 9–10 disordered residues (Evdokimov *et al.*, 2002); in the other SycE and SycE–YopE(Cb) complex structures the carboxy-terminal ~8–12 residues are all disordered (Birtalan & Ghosh, 2001; Birtalan *et al.*, 2002). Here, we present the structure of SycE of *Y. enterocolitica* in a crystal form in which the entire carboxy-terminal peptides of two out of four protomers in the asymmetric unit are well ordered and form specific intermolecular interactions in a manner suggestive of a self-binding activity that might suppress non-specific peptide-binding interactions of the chaperone.

2. Experimental

2.1. Protein expression, purification and crystallization

The coding sequence for the *sycE* gene of *Yersinia enterocolitica* was PCR-amplified from purified genomic DNA (a gift from the laboratory of Dr Stan Falkow, Stanford University) and ligated into a pTYB vector of the 'IMPACT' expression system (New England Biolabs) using *Xma*I and user-intro-

duced *Nsi*I restriction sites, yielding an in-frame fusion of the *sycE* gene with the coding sequence of the intein- and chitin-binding domains of the vector. Protein was expressed in *Escherichia coli* strain BL21(DE3) and cells were grown at 310 K to late exponential phase (A_{600} = 0.5); expression of the recombinant fusion protein was induced with 0.5 m*M* isopropyl-β-D-thiogalactoside (IPTG) at 293 K. Cells were harvested 6 h after induction, pelleted by centrifugation and resuspended in buffer *A*: 0.5 *M* NaCl, 50 m*M* Na HEPES pH 7.7, 1 m*M* EDTA, 1 m*M* tris(2-carboxyethyl)-phosphine hydrochloride (TCEP) plus one 'protease inhibitor cocktail tablet' (Roche) per 50 ml of cell suspension. The suspension was sonicated for ~10 min at 277 K and cell debris was removed by centrifugation. The soluble phase was filtered (0.22 µm pore size), loaded onto a chitin column pre-equilibrated in buffer *A*, washed with 60 column volumes of buffer *A* and cleaved from the intein with 50 m*M* dithiothreitol (DTT) in buffer *A*.

The eluted SycE protein was concentrated to ~10 mg ml^{-1} for crystallization. The final product consists of full-length *Y. enterocolitica* SycE protein plus an additional Pro-Gly dipeptide at the carboxy-terminus, which is introduced by the *Xma*I restriction site used in the cloning procedure.

Crystals of SycE were grown at 291 K in hanging drops containing equal volumes of protein stock (10 mg ml^{-1} in a buffer *A* plus 50 m*M* DTT) and a precipitant consisting of 1.75 *M* $(NH_4)_2SO_4$, 2% PEG 400 (polyethylene glycol of average molecular weight 400), 0.25 *M* NaCl, 100 m*M* Na HEPES pH 7.7, 1 m*M* TCEP. Microseeding was used to produce crystals of sufficient size for data collection.

2.2. Data collection and structure determination

Crystals were stabilized in 2.1 *M* $(NH_4)_2SO_4$, 2% PEG 400, 0.25 *M* NaCl, 100 m*M* Na HEPES pH 7.7, 1 m*M* TCEP and adapted to a cryoprotectant consisting of stabilization solution plus 20% glycerol with incremental increases of 5% glycerol, then flash-frozen in a nitrogen-gas stream at ~100 K. Data were collected from a single crystal using an ADSC CCD detector on beamline 11-1 of the Stanford Synchrotron Radiation Laboratory (SSRL). The crystal space group is $P2_12_12_1$, with unit-cell parameters a = 52.55, b = 92.67, c = 101.05 Å and four SycE protomers per asymmetric unit. All data were processed and scaled using the *DENZO*/*SCALEPACK* program package (Otwinowski & Minor, 1997). Data-collection statistics are summarized in Table 1.

The structure was determined by molecular replacement with the *CCP*4 program *MOLREP* using a single SycE protomer (PDB code 1k6z) as a search model. Four protomers that defined two dimers, similar to published dimeric SycE structures (Birtalan & Ghosh, 2001; Evdokimov *et al.*, 2002), were located in the asymmetric unit by iterative searches. The positions of the two dimers were consistent with an approximate non-crystallographic translation peak in the native Patterson at (0.5, ~0, 0.5). Rigid-body refinement of the search models yielded the statistics R_{cryst} = 0.402, R_{free} = 0.401. *B*-factor adjustment and simulated annealing with *CNS* (Brünger *et al.*, 1998) reduced R_{cryst} and R_{free} to 0.291 and 0.338, respectively. Manual model building with *O* (Jones, 1978; Jones *et al.*, 1991), including the addition of 12 carboxy-terminal residues in two protomers, further reduced R_{cryst} and R_{free} to 0.261 and 0.296, respectively. Further rebuilding and addition of solvent molecules yielded the statistics summarized in Table 1. The final model includes residues 3–130 of subunit *A*, 3–119 and 127–130 of subunit *B*, 3–118 and 126–129 of subunit *C* and 3–130 of subunit *D*, plus 89 water molecules. Each protomer has three regions which are difficult to trace owing to disorder: *A*20–26, *A*53–60 and *A*95–101; *B*19–24, *B*53–60 and *B*97–102; *C*20–22, *C*55–58 and *C*98–101; *D*17–22, *D*55–58 and *D*98–101. Additionally, the connections between the core of the protomer and the extreme carboxy-terminal residues of protomers *B* (residues 120–126) and *C* (residues 119–125) could not be traced unambiguously, although disconnected segments of electron density were present in these regions. The anisotropy of the data and disordered regions of the structure contribute to precluding reduction of the relatively high final values of R_{cryst} and R_{free} to below the values in Table 1; however, the accuracy of the overall structure is confirmed by its congruence with previously determined SycE structures (Birtalan & Ghosh, 2001; Evdokimov *et al.*, 2002).

Table 1
Crystallographic data-collection and refinement statistics.

Values in parentheses are for the last resolution shell.

Data collection	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 52.55, b = 92.67, c = 101.05
Wavelength (Å)	0.98
Resolution range (Å)	30–1.98 (2.01–1.98)
Observations (total/unique)	154025/34752
Completeness (%)	97.8 (89.2)
Average $I/\sigma(I)$	30.5 (5.0)
R_{sym}†	0.037 (0.235)
Refinement	
Resolution range (Å)	30.0–2.00 (2.03–2.00)
R_{cryst}‡	0.248 (0.308)
R_{free}	0.290 (0.321)
No. of reflections (working set)	29315
No. of reflections (test set)	3223
No. of protein atoms	3950
No. of solvent atoms	89
Average B value, all atoms (Å^2)	48.4
R.m.s.d. bond length (Å)	0.006
R.m.s.d. angles (°)	1.14

† $R_{sym} = \sum |I_{hkl} - \langle I_{hkl}\rangle| / \sum \langle I_{hkl}\rangle$, where I_{hkl} is a single value of measured intensity of the *hkl* reflection and $\langle I\rangle$ is the mean of all measured intensities of the *hkl* reflection. ‡ $R_{cryst} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$, where F_{obs} is the observed structure-factor amplitude and F_{calc} the structure factor calculated from the model. R_{free} is computed in the same manner as R_{cryst}, using the test set of reflections.

3. Results and discussion

The structure of the *Y. enterocolitica* SycE has been solved and refined to 2.0 Å resolution in a crystal form that contains two dimers per asymmetric unit. Not surprisingly, the tertiary structure of the protomer core of the molecule is similar to that of the *Y. pseudotuberculosis* and *Y. pestis* SycE proteins (Birtalan & Ghosh, 2001; Evdokimov *et al.*, 2002; root-mean-square deviations of C^α atoms of 0.70 and 0.62 Å, respectively), from which it differs in sequence by one and two amino acids, respectively. The novel feature of the structure in this crystal form, compared with previous structures, is the well defined carboxy-terminal peptide in one protomer of each dimer (protomers *A* and *D*), which is stabilized by specific interactions with neighboring protomers (Fig. 1*a*). Specifically, the backbone carbonyl O atom of Ser120 and N atom of Leu122 hydrogen bond to the side chains of Gln112 and Glu108, respectively, of one neighboring protomer, while both the N atom and carbonyl O atom of Arg127 hydrogen bond in a short antiparallel β configuration with Lys30 of another protomer, thereby positioning the side chain of Pro126 in a hydrophobic pocket (Fig. 1*b*). Additionally, the carboxy-terminal segments of the second protomer of each dimer (residues 127–130 of subunit *B* and 126–129 of subunit *C*) bind neighboring molecules in a similar manner, even though the connections between the extreme carboxy-terminal segments and the core domains were disordered.

The hydrophobic pocket in which Pro126 binds was noted in the original SycE structure (Birtalan & Ghosh, 2001). The interactions we observe with the carboxy-

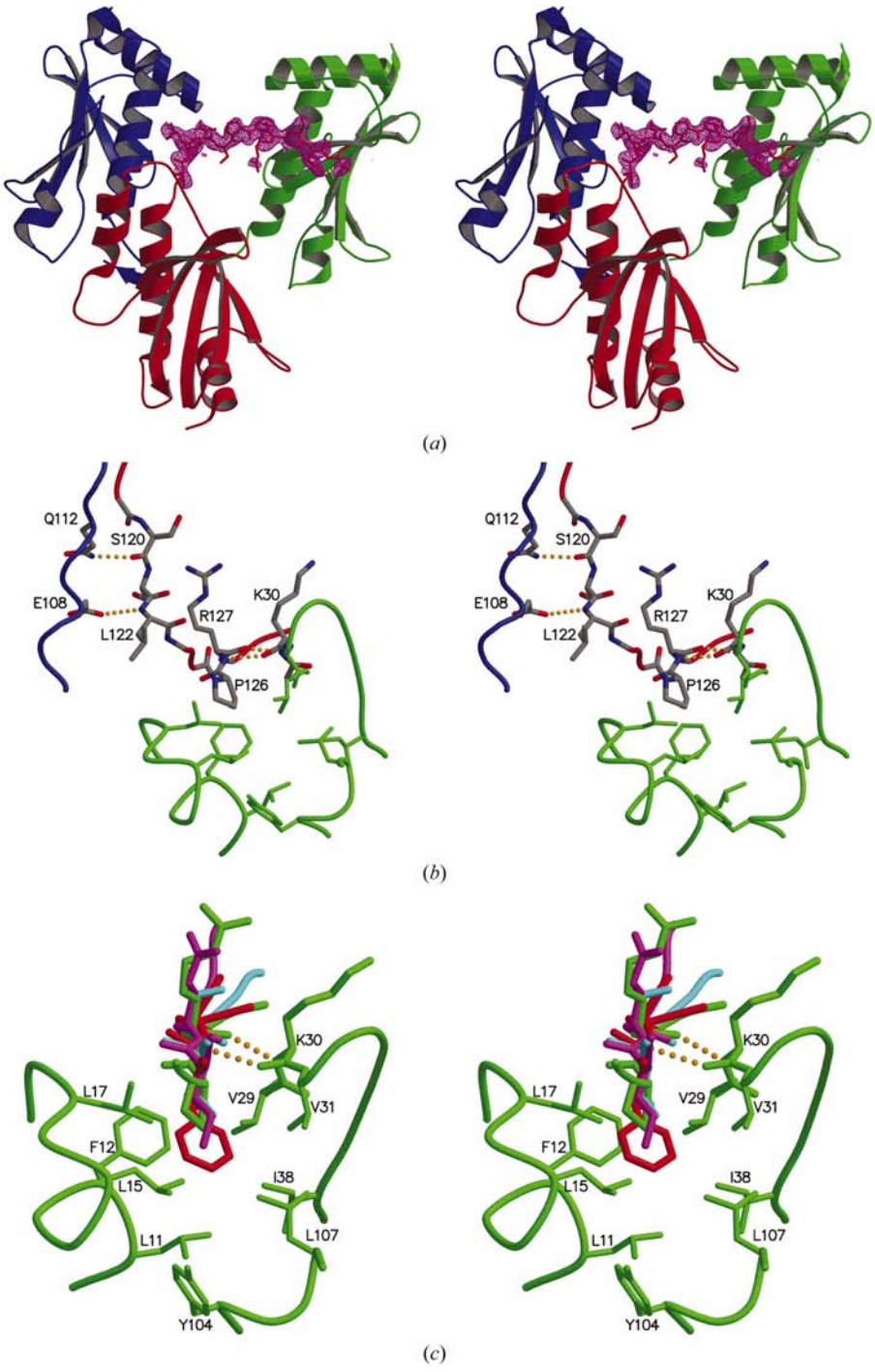

Figure 1
Interaction of carboxy-terminal peptide of SycE with neighboring protomers. (*a*) Simulated-annealing $F_o - F_c$ difference Fourier map, contoured at 3.2σ, in which the carboxy-terminal peptide of the *A* subunit (residues 120–130; shown as a stick model colored red) was omitted from the model used in refinement and phase calculation. A ribbon diagram of *A* subunit residues 3–119 (red) and two neighboring subunits (blue and green) with which the carboxy-terminal peptide interacts are shown. (*b*) Specific interactions of the carboxy-terminal peptide of the *A* subunit with neighboring subunits. The backbones of the subunits are shown in the same colors as in (*a*). Selected peptide-backbone segments and amino-acid side chains are shown as stick models. (*c*) Essential features of the SycE hydrophobic pocket, with superposition of the peptide segments which bind to it. C^{α} atoms of residues 10–19, 28–39 and 103–108 of each SycE protomer were superimposed on those of the *Y. enterocolitica* SycE (the backbone trace of which is shown in green) using *LSQMAN* (Kleywegt & Jones, 1997); the transformation was then applied to the segments that bind the hydrophobic pocket. Green, *Y. enterocolitica* SycE, Pro126 side chain in pocket (this work); red, SycE (PDB code 1k6z; Evdokimov *et al.*, 2002), Phe129 in pocket; cyan and magenta, SycE–YopE(Cb) complex (PDB code 1l2w; Birtalan *et al.*, 2002), Val31 and Ile60 in pocket, respectively. Figures were produced with *BOBSCRIPT* (Esnouf, 1997, 1999) and *MOLSCRIPT* (Kraulis, 1991) and rendered with *Raster3D* (Merritt & Bacon, 1997).

Table 2
Sequences of peptide segments showing binding to the hydrophobic pocket of SycE in crystals.

The amino-acid residues whose side chain is sequestered in the hydrophobic pocket is shown in bold and numbered. Six residues of each peptide are shown for comparison, but this does not imply that all residues interact with the hydrophobic pocket.

Peptide sequence	Molecule
SP**P$_{126}$**RSF	SycE (this work)
RS**F$_{129}$**SHH†	SycE with His$_6$ carboxy-terminal tag (Evdokimov *et al.*, 2002)
RS**V$_{31}$**SQQ	YopE(Cb) (Birtalan *et al.*, 2002)
RI**I$_{60}$**ERL	YopE(Cb) (Birtalan *et al.*, 2002)

† Two residues of a carboxy-terminal His$_6$ affinity tag are observed in the structure following the end of the native sequence at Ser130.

terminal peptide of *Y. enterocolitica* SycE are similar to those seen with the carboxy-terminal segment of the *Y. pestis* SycE-His$_6$ fusion (Evdokimov *et al.*, 2002), as well as with the YopE fragment in the YopE(Cb)–SycE complex (Birtalan *et al.*, 2002) (Fig. 1*c* and Table 2). In each case, the peptide backbone of the binding segment makes a short antiparallel β-strand interaction with Lys30, while a hydrophobic side chain interfaces with the hydrophobic pocket of SycE. Notably, the hydrophobic residue is different in each case: in the *Y. enterocolitica* structure described here it is Pro126, while in the *Y. pestis* SycE structure it is Phe129; in the YopE(Cb)–SycE complex, Val31 and Ile60 of the YopE fragment interface with the two different subunits of the SycE dimer. Additionally, in three out of four cases, an arginine two residues before the residue bound in the hydrophobic pocket makes a salt bridge to Glu115 of SycE.

A question which arises is whether the interaction we observe is a completely fortuitous crystal contact or whether it is related to the function of the chaperone. Specifically, it raises the question of whether the carboxy-terminal peptides of this family of chaperones, which show significant heterogeneity in both length and sequence, are responsible for a weak self-association of the chaperones that would suppress non-specific and unproductive interactions with other proteins in the cytoplasm. A precedent for such self-modulation of peptide-binding activity by a molecular chaperone is established for the 70 kDa heat-shock protein (Hsp70) family (McKay *et al.*, 1994). The first ~550 residues of Hsp70 proteins encode tandem ATPase and peptide-binding domains that are highly conserved in sequence. This is followed by 50–100 residues that are highly variable in sequence; this carboxy-terminal extension is responsible for reversible self-association of the

Hsp70 proteins, which is thought to attenuate counterproductive interactions with cellular proteins under conditions where Hsp70s are present in substantial excess over improperly folded polypeptide substrates. By analogy, the hydrophobic binding pocket of SycE to which the carboxy-terminal peptide binds appears to be adapted for relatively promiscuous (and presumably low-affinity) binding of short segments of unstructured hydrophobic polypeptides; the interaction with the SycE carboxy-terminal peptide is suggestive of a self-binding activity that may compete with unwanted interactions with non-target proteins in the cell in the absence of saturating levels of cognate target protein to sequester the chaperone.

This work was supported by grant NIH-GM39928 to DBM. We are grateful to members of the Falkow laboratory at Stanford for the gift of *Y. enterocolitica* genomic DNA. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences.

References

Birtalan, S. & Ghosh, P. (2001). *Nature Struct. Biol.* **8**, 974–978.
Birtalan, S. C., Phillips, R. M. & Ghosh, P. (2002). *Mol. Cell*, **9**, 971–980.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–134.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Evdokimov, A. G., Tropea, J. E., Routzahn, K. M. & Waugh, D. S. (2002). *Acta Cryst.* D**58**, 398–406.
Hueck, C. J. (1998). *Microbiol. Mol. Biol. Rev.* **62**, 379–433.
Iriarte, M. & Cornelis, G. R. (1998). *Mol. Microbiol.* **29**, 915–929.
Jones, A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 525–545.
Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.
McKay, D. B., Wilbanks, S. M., Flaherty, K. M., Ha, J.-H., O'Brien, M. C. & Shirvanee, L. L. (1994). *Heat Shock Proteins: Structure, Function and Regulation*, edited by R. Morimoto, A. Tissieres & C. Georgopoulos, pp. 153–177. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Stebbins, C. E. & Galan, J. E. (2001). *Nature (London)*, **414**, 77–81.