Erschienen in: Acta Crystallographica Section D : Biological Crystallography ; 71 (2015), 10. - S. 2040-2053
https://dx.doi.org/10.1107/S1399004715014832

BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Crystallizing the 6S and 8S spliceosomal assembly intermediates: a complex project

Jann-Patrick Pelz,[a] Hermann Schindelin,[b] Katharina van Pee,[a] Jochen Kuper,[b] Caroline Kisker,[b] Kay Diederichs,[c] Utz Fischer[a] and Clemens Grimm[a]*

[a]Department of Biochemistry, Theodor Boveri Institute, University of Würzburg, Am Hubland, 97074 Würzburg, Germany, [b]Rudolf-Virchow-Zentrum, DFG Research Centre for Experimental Medicine, University of Würzburg, Josef-Schneider-Strasse 2/Haus D15, 97080 Würzburg, Germany, and [c]Protein Crystallography and Molecular Bioinformatics, University of Konstanz, 78457 Konstanz, Germany. *Correspondence e-mail: clemens.grimm@uni-wuerzburg.de

Received 20 March 2015
Accepted 7 August 2015

Edited by K. Miki, Kyoto University, Japan

Keywords: U snRNP assembly; spliceosomal assembly; contact engineering; surface-entropy reduction; SMN; Sm proteins; pICln.

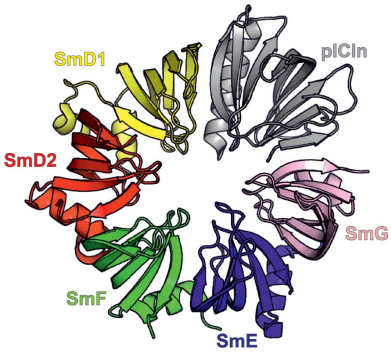
Supporting information: this article has supporting information at journals.iucr.org/d

The small nuclear ribonucleoproteins (snRNPs) U1, U2, U4/6 and U5 are major constituents of the pre-mRNA processing spliceosome. They contain a common RNP core that is formed by the ordered binding of Sm proteins onto the single-stranded Sm site of the snRNA. Although spontaneous *in vitro*, assembly of the Sm core requires assistance from the PRMT5 and SMN complexes *in vivo*. To gain insight into the key steps of the assembly process, the crystal structures of two assembly intermediates of U snRNPs termed the 6S and 8S complexes have recently been reported. These multimeric protein complexes could only be crystallized after the application of various rescue strategies. The developed strategy leading to the crystallization and solution of the 8S crystal structure was subsequently used to guide a combination of rational crystal-contact optimization with surface-entropy reduction of crystals of the related 6S complex. Conversely, the resulting high-resolution 6S crystal structure was used during the restrained refinement of the 8S crystal structure.

1. Introduction

U snRNPs are major constituents of the major and minor spliceosomes, which catalyze the splicing of cellular pre-mRNAs (Tarn & Steitz, 1997; Wahl *et al.*, 2009). The major spliceosome contains four individual snRNPs termed U1, U2, U4/6 and U5 and non-snRNP splice factors, which together perform decisive steps in the splicing reaction of canonical introns. All snRNP particles share a common core as a structural framework formed by the seven Sm proteins SmB/B′, SmD1, SmD2, SmD3, SmE, SmF and SmG that assemble on a specific single-stranded region of the snRNA, referred to as the Sm binding site (Branlant *et al.*, 1982; Liautard *et al.*, 1982; Lührmann *et al.*, 1990). The assembly of this so-called Sm core domain takes place in the cytosol and is dependent on *trans*-acting protein factors organized in the PRMT5 and SMN complexes (Yong *et al.*, 2004; Chari *et al.*, 2009). The Sm proteins are pre-organized on the PRMT5 complex into higher-order complexes in which the Sm proteins acquire the position that they occupy in the assembled Sm core domain. The most prominent assembly intermediate arising from this pathway is termed the 6S complex, which contains the five Sm proteins SmD1, SmD2, SmE, SmF and SmG and pICln. Remarkably, this complex does not bind to the snRNA and represents a kinetic trap in the assembly process (Pu *et al.*, 1999). The kinetically trapped Sm proteins can only be released by the action of the SMN complex, a multi-subunit entity consisting of SMN itself, Gemin2–Gemin8 and UNRIP (Neuenkirchen *et al.*, 2008).

In order to obtain structural insight into the assembly reaction, we recently solved two important structures of the



© 2015 International Union of Crystallography

assembly pathway: the 6S complex, which is the major product of the early assembly phase, and an additional complex representing the transition phase, termed the 8S complex. The latter represents a trapped intermediate consisting of the 6S complex bound to the SMN complex components Gemin2 and a fragment of SMN (Chari *et al.*, 2008; Grimm *et al.*, 2013). Owing to the initial lack of crystals with sufficient diffraction power, the two structures could only be solved after the application of several rescue strategies for crystallogenesis and crystal optimization. A frequently applied rescue strategy is the elimination of unstructured regions from the protein constructs to be crystallized. As the growth of protein crystals is mainly dependent on entropic changes (Gliko *et al.*, 2005; Paunov *et al.*, 2001; Yau *et al.*, 2000), regions with high conformational entropy are likely to hinder crystal formation. The systematic identification and subsequent stepwise deletion of such regions turned out to be decisive for the generation of well diffracting crystals of the 6S and 8S complexes. In addition, the amino-acid composition of the protein surface is often crucial for the formation of crystal contacts. While a high number of large polar side chains correlates negatively with the propensity of a protein to crystallize (Price *et al.*, 2009), small polar amino acids seem to be preferred in the vicinity of crystal contacts (Cieślik & Derewenda, 2009). This is the foundation for the established and successful surface-entropy reduction (SER) strategy that aims to optimize surface properties for crystallization (Longenecker *et al.*, 2001).

Here, we present the particular recalcitrant cases of the 8S and 6S assembly intermediates and describe how their crystal structures could eventually be solved by applying the aforementioned optimization strategies in combination with an approach to rationally foster crystal contacts. We describe how the low-resolution 8S crystal structure provided the basis for successful crystal-contact engineering of crystals of the related 6S complex. Conversely, the resulting high-resolution 6S structure was crucial for the final refinement of the 8S structure at low resolution by acting as a reference model.

2. Experimental

2.1. Protein preparation and complex reconstitution

SmD1/D2 and SmE/F/G were expressed and purified as reported previously (Kambach *et al.*, 1999; Chari *et al.*, 2008; Grimm *et al.*, 2013). *Drosophila melanogaster* pICln cDNA (GenBank entry AF216522) was cloned into the pET-M13 expression vector. All engineered protein constructs were then derived from this parental vector preparation. Modifications of pICln included the truncation of flexible regions; in addition, several surface amino acids were mutated. Point mutations for surface modification were introduced by the application of site-directed mutagenesis (Carter, 1986) and deletion mutants were produced by overlapping extension PCR (Bryksin & Matsumura, 2010). *Escherichia coli* strain Rosetta(DE3)pLysS was transformed with the resulting expression vectors and protein overexpression was induced at an OD_{600} of 0.8 by the addition of 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by 15 h incubation at 290 K. The bacterial cells were harvested by centrifugation, resuspended in IMAC-*A* buffer [300 m*M* NaCl, 20 m*M* HEPES pH 7.5, 10%(*w*/*v*) glycerol, 2 m*M* β-mercaptoethanol] and lysed by sonication. The lysate was then cleared by ultracentrifugation and the fusion protein was bound to nickel–NTA agarose (Qiagen, Hilden, Germany). After elution with IMAC-*B* buffer (150 m*M* NaCl, 20 m*M* HEPES pH 7.5, 250 m*M* imidazole, 2 m*M* β-mercaptoethanol), the protein was applied onto a Superdex 75 column (GE Healthcare, Munich, Germany) equilibrated with size-exclusion chromatography (SEC) buffer (150 m*M* NaCl, 10 m*M* HEPES pH 7.5, 2 m*M* TCEP). Fractions containing the target protein were pooled and concentrated to 30 mg ml^{-1}.

The 6S complex was reconstituted by mixing equimolar amounts of purified SmD1/D2, SmE/F/G and pICln proteins in reconstitution buffer (1 *M* NaCl, 20 m*M* HEPES pH 7.5, 5 m*M* DTT). Subsequently, the salt concentration was reduced by overnight dialysis- against SEC buffer. The dialysate was then passed over a Superdex 200 column (GE Healthcare, Munich, Germany) and fractions containing the complex were pooled and concentrated to 15 mg ml^{-1}.

The 8S complex represents a stalled assembly intermediate that was stabilized by the introduction of a C-terminal truncation of the SMN protein (Chari *et al.*, 2008). *D. melanogaster* Gemin2 and *D. melanogaster* SMNΔC (comprising residues 1–122) fused to an N-terminal His_6-GST tag were each individually expressed in *E. coli* strain Rosetta(DE3)pLysS. The cells were grown to an OD_{600} of 0.8 and protein expression was induced by the addition of 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by 15 h incubation at 290 K. After harvesting, the cells were mixed to produce a co-lysate. A cleared lysate was prepared following the procedure described above. This was followed by two affinity chromatography steps using nickel–NTA agarose (Qiagen, Hilden, Germany) and glutathione Sepharose (GE Healthcare, Munich, Germany). IMAC affinity chromatography was performed following the procedure described for pICln purification. After elution with IMAC-*B* buffer, glutathione Sepharose and 5 m*M* EDTA were added to the eluate. After elution with GSH-*B* buffer (150 m*M* NaCl, 20 m*M* HEPES pH 7.5, 2 m*M* β-mercaptoethanol) the fusion tag was proteolytically cleaved and removed by re-incubation with nickel–NTA agarose. The protein solution was passed over a Superdex 200 column and fractions containing stoichiometric amounts of Gemin2 and SMNΔC were pooled. Reconstitution of the 8S complex was carried out by mixing the preformed 6S complex with an equimolar amount of SMNΔC/Gemin2. The preparation was then passed over a Superdex 200 column and fractions containing the complex were pooled, concentrated to 15 mg ml^{-1} and shock-frozen in liquid nitrogen for storage.

2.2. Crystallization

High-throughput crystallization screening was carried out with a Cartesian Honeybee Nano Dispenser (Hamilton

Table 1
Crystallization.

Crystal form	8S	6S-*C*2	6S-*P*321	6S-T17A-1	6S-T17A-2
Protein components	SMN, Gemin2, SmD1, SmD2, SmE, SmF, SmG, pICln (Δ181–215)	SmD1, SmD2, SmE, SmF, SmG, pICln (Δ90–125, Δ160–215, H144A)	SmD1, SmD2, SmE, SmF, SmG, pICln (Δ90–120, Δ160–215)	SmD1, SmD2, SmE, SmF, SmG, pICln (Δ181–215, T17A)	SmD1, SmD2, SmE, SmF, SmG, pICln (Δ181–215, T17A)
Image	Fig. 2(*b*)	Fig. 2(*f*)	Fig. 2(*c*)	Fig. 2(*d*)	Fig. 2(*e*)
Method	Hanging-drop vapour diffusion	Hanging-drop vapour diffusion	Hanging-drop vapour diffusion	Hanging-drop vapour diffusion	Hanging-drop vapour diffusion
Temperature (K)	289	289	293	289	289
Protein concentration (g l^{-1})	15	15	12	12	12
Buffer composition of protein solution	150 m*M* NaCl, 10 m*M* HEPES pH 7.5, 5 m*M* TCEP	150 m*M* NaCl, 10 m*M* HEPES pH 7.5, 5 m*M* TCEP	150 m*M* NaCl, 10 m*M* HEPES pH 7.5, 5 m*M* TCEP	150 m*M* NaCl, 10 m*M* HEPES pH 7.5, 5 m*M* TCEP	150 m*M* NaCl, 10 m*M* HEPES pH 7.5, 5 m*M* TCEP
Composition of reservoir solution	24% PEG 4000, 10% ethanol, 150 m*M* NaCl pH 7.5	15% Jeffamine ED-2003 titrated to pH 7, 10% ethanol	2.7 *M* sodium formate, 0.1 *M* imidazole pH 8.0	10% PEG 20 000, 0.4 *M* malonate, 0.1 *M* glycylglycine pH 8.5	20% polyvinylpyrolidone K15, 0.1 *M* bis-tris pH 5.5
Composition of cryobuffer	27% PEG 4000, 10% ethanol, 150 m*M* NaCl pH 7.5	18% Jeffamine ED-2003 titrated to pH 7, 10% ethanol, 30% glycerol	3 M lithium formate, 0.1 *M* imidazole pH 8.0	Reservoir solution with 30% glycerol	Reservoir solution with 30% glycerol
Volume and ratio of drop	7 µl, 7:1 (protein: precipitant)	1 µl, 1:1	4 µl, 1:1	2 µl, 1:1	2 µl, 1:1
Volume of reservoir (µl)	300	300	300	300	300

Robotics, Martinsried, Germany) in sitting-drop vapour-diffusion polypropylene plates (Greiner Bio-One, Frickenhausen, Germany) by mixing 0.2 µl each of the protein solution and the reservoir solution. Identical crystallization plates were incubated at 293, 289 and 277 K. Crystal optimization was carried out manually and included the addition of additives (Hampton Research, Aliso Viejo, USA) and techniques such as gel crystallization and counterdiffusion. The final 8S crystals used for the collection of diffraction data were grown by the hanging-drop vapour-diffusion method in polystyrene plates (Jena Bioscience, Jena, Germany). Optimization involved the screening of varying drop sizes from 0.5 to 8 µl and protein:reservoir solution ratios ranging between 1:1 and 8:1. Crystals of 6S preparations containing the engineered pICln constructs were grown by the hanging-drop vapour-diffusion method in polystyrene plates (Jena Bioscience, Jena, Germany) by mixing 0.5–1 µl each of the protein and the reservoir solution while keeping the protein:reservoir solution ratio constant at 1:1. See Table 1 for a detailed overview of the individual crystallization conditions.

2.3. Data collection, phasing and model refinement

The 8S crystals were slowly transferred into cryobuffer (Table 1); the transfer was facilitated *via* five intermediate buffer steps that were prepared by mixing the cryobuffer with decreasing amounts of reservoir solution. The crystals were mounted in LithoLoops (Molecular Dimensions, Suffolk, England) and plunged into liquid nitrogen. Before measurement, the LithoLoops with the crystals were placed and centred within a nitrogen-gas stream at a temperature of 100 K. The final data set was collected on beamline ID14-4 at the ESRF, Grenoble, France. Data-collection statistics are given in Table 2. Auto-indexing and reduction were performed with *XDS* (Kabsch, 2010). Detection of weak reflections indicative of the 351 Å *b* axis was possible after setting the value for 'STRONG_PIXEL' to 2 and using all 360 frames in the autoindexing step (see Supplementary Table S1 for the list of autoindexing solutions from `IDXREF.LP` including weak reflections and Supplementary Table S2 for the autoindexing step with default settings). Useful data were detected to 3.1 Å resolution within the data set (Karplus & Diederichs, 2012); however, the crystal showed strongly anisotropic diffraction. Therefore, ellipsoidal resolution limit truncation of the data set followed by anisotropic scaling was performed with values of 3.1, 3.8 and 4.0 Å along a^*, b^* and c^*, respectively (Strong *et al.*, 2006). Native Patterson functions were calculated with different programs to check for the presence of noncrystallographic translational symmetry. Molecular replacement was performed with *MOLREP* (Vagin & Teplyakov, 2010) incorporating the highest NCS translation peak and using the coordinates of PDB entry 3s6n as the search model (Zhang *et al.*, 2011). After the initial placement of 12 copies of the complex, the partial model was refined and eight further copies could be placed in the resulting map. The resulting model was then subjected to two cycles of automated rigid-body, coordinate and ADP refinement with *PHENIX* (Adams *et al.*, 2010) and manual rebuilding. The pICln PH fold extracted from PDB entry 1zyi (Fürst *et al.*, 2005) was placed manually into the corresponding difference density at the appropriate sites in the 20 Sm rings. The completed model was then subjected to another round of automated refinement, and missing and deviating parts of the model were manually rebuilt with *Coot* (Emsley & Cowtan, 2004). After a total of ten rounds of manual rebuilding and automated refinement, the R and $R_{\rm free}$ factors converged. At this stage, the final high-resolution model of the 6S structure was available. We therefore performed further rounds of refinement within

Table 2
Data collection and processing.

Values in parentheses are for the outer shell.

Data set	8S†	6S-*C*2†	6S-*P*321
No. of molecules in asymmetric unit	20	2	18
Diffraction source	ESRF beamline ID14-4	ESRF beamline ID14-4	SLS beamline PXII
Wavelength (Å)	0.9762	0.9393	1.000
Temperature (K)	100	100	100
Detector	ADSC Quantum 315r	ADSC Quantum 315r	PILATUS 6M
Crystal-to-detector distance (mm)	478.0	262.3	262.3
Rotation range per image (°)	0.5	0.5	0.25
Total rotation range (°)	180	127.5	37.5
Exposure time per image (s)	0.5	0.8	0.25
Space group	$P2_1$	*C*2	*P*321
a, *b*, *c* (Å)	150.9, 356.8, 230.8	180.7, 65.3, 99.4	203.9, 203.9, 215.9
α, β, γ (°)	90.0, 97.3, 90.0	90.0, 92.5, 90.0	90.0, 90.0, 120.0
Mosaicity (°)	0.154	0.372	0.115
Resolution range (Å)	59.5–3.1/59.5–4.0 (3.29–3.10/4.23–4.00)	49.6–1.9 (2.01–1.90)	50.0–3.99 (4.23–3.99)
Total No. of reflections	1090980	237308	444855
No. of unique reflections	281563	88058	84895
Completeness (%)	98.9 (96.7)/64.6 (5.8)‡	96.2 (96.2)	99.4 (96.6)
Multiplicity	3.9 (3.9)	2.7 (2.7)	5.2 (5.0)
$\langle I/\sigma(I)\rangle$	5.8 (2.3)/4.6 (1.4)‡	14.3 (1.8)	12.5 (1.7)
R_{merge} (%)	13.9 (50.5)/20.0 (92.8)‡	4.4 (58.2)	9.2 (109.1)
PDB code	4v98	4f7u	Not submitted

† These data sets and the corresponding crystal structures were published in Grimm *et al.* (2013). ‡ The values given are for the resolution ranges 60–4.0 Å (highest resolution shell 4.23–4.00 Å in parentheses) and 60–3.1 Å without ellipsoidal truncation (highest resolution shell 3.29–3.10 Å in parentheses).

Table 3
Structure solution and refinement.

Values in parentheses are for the outer shell.

	8S†	6S-*C*2†	6S-*P*321
Resolution range (Å)	59.47–3.10 (3.18–3.10)	49.63–1.90 (1.92–1.90)	—
Completeness (%)	98.9 (96.7)/64.6 (5.8)‡	96.2 (96.2)	—
No. of reflections, working set	281563 (733)	88054 (2518)	—
No. of reflections, test set	14120 (44)	4426 (117)	—
Final R_{cryst}	0.232 (0.367)	0.181 (0.3128)	—
Final R_{free}	0.256 (0.299)	0.221 (0.3502)	—
No. of non-H atoms			
Protein	121940	8140	—
Ion	0	0	—
Ligand	50	31	—
Water	0	372	—
Total	121990	8543	—
R.m.s. deviations			
Bonds (Å)	0.010	0.012	—
Angles (°)	1.29	1.52	—
Average *B* factors (Å^2)			
Protein	97.3	44.7	—
Ligand	122.9	48.8	—
Ramachandran plot			
Most favoured (%)	92.4	98.3	—
Allowed (%)	5.2	1.7	—
Disallowed (%)	2.4	0.0	—

† These data sets and the corresponding crystal structures were published in Grimm *et al.* (2013). ‡ The values given are for the resolution ranges 60–4.0 Å (highest resolution shell 4.23–4.00 Å in parentheses) and 60–3.1 Å without ellipsoidal truncation (highest resolution shell 3.29–3.10 Å in parentheses).

BUSTER (Smart *et al.*, 2012) using the 6S structure as a reference model. During the refinement, the automatic detection of deviating model parts depending on gradient and coordinate differences was activated and roughly 25% of the reference-model restraints were automatically deactivated.

The 6S crystals, one of which yielded the high-resolution data set (6S-*C*2, Table 1), were transferred into a cryobuffer consisting of reservoir solution supplemented with 30% glycerol (Table 1). The crystals were mounted in LithoLoops (Molecular Dimensions, Suffolk, England) and plunged into liquid nitrogen. Prior to data collection, the LithoLoops holding the crystals were positioned in a nitrogen-gas stream at a temperature of 100 K. The data set was collected on beamline ID14-4 at the ESRF, Grenoble, France. Auto-indexing and reduction were performed with *XDS*. Molecular replacement was performed with *Phaser* (McCoy *et al.*, 2007) using a single molecule of the 8S crystal structure as a search model, from which the Gemin2 and SMN parts had been removed. Two molecules were found within the asymmetric unit and the model was refined with *Coot* and *PHENIX* including TLS refinement. The *R* and R_{free} factors converged within five additional rounds of manual and automated refinement including the placement of a PEG molecule and 372 water molecules. After modelling several alternative residue and peptide conformations, a further decrease of roughly 2% in R_{free} could be achieved (Table 3).

Crystals of the initial 6S crystal form (6S-*P*321, Table 1) were transferred into a cryobuffer that contained lithium formate instead of the original precipitant sodium formate (Rubinson *et al.*, 2000). The crystals were mounted in LithoLoops (Molecular Dimensions, Suffolk, England) and plunged into liquid nitrogen. Prior to data collection, the LithoLoops with the crystals were placed and centred within a nitrogen-gas stream at a temperature of 100 K. The data set was collected on beamline PXII at the SLS, Villigen, Switzerland. Auto-indexing and reduction were performed with *XDS*. Molecular replacement was performed with *Phaser* using a single molecule of the 8S crystal structure as a search model, from which the Gemin2 and SMN parts were removed. 18 molecules could be found within the asymmetric unit. Owing to the limited resolution (Table 2) and the availability of the high-resolution data set, refinement with *Coot* and *PHENIX* was abandoned at an early stage and this structure was not submitted to the PDB.

3. Results and discussion

3.1. Initial complex reconstitution and crystallization

The spliceosomal 6S (Fig. 1*a*) and 8S (Fig. 1*b*) assembly intermediates were reconstituted *in vitro* from bacterially

overexpressed subunit preparations. The reconstitution protocol primarily involved mixing the complex constituents under high-salt conditions and subsequent dialysis to physiological salt concentrations. For the two complexes, an extensive high-throughput crystallization screening with commercial sparse-matrix screens did not yield any initial hits. As, apart from salts, most commercial screens rely almost exclusively on various PEG species as the main precipitant, we set out to expand the crystallization space by adding conditions containing alternative polymeric precipitants (Grimm *et al.*, 2010). Within such a screen, a crystallization batch that contained the polyacrylate SOKALAN CP 42 yielded small crystals of the 8S complex (Fig. 2*a*). These crystals diffracted to 8 Å resolution at different synchrotron beamlines. Neither an optimization of the crystallization conditions nor screening of diverse additives could improve the size of these crystals significantly. After the screening of nearly 20 000 different conditions including different temperatures, additives and techniques such as sitting-drop vapour diffusion, gel crystallization and counterdiffusion without significant improvements of the initial crystal forms, the possibilities for conventional crystallization optimization or the discovery of further crystal forms seemed to be exhausted. We therefore decided to optimize the protein constructs.

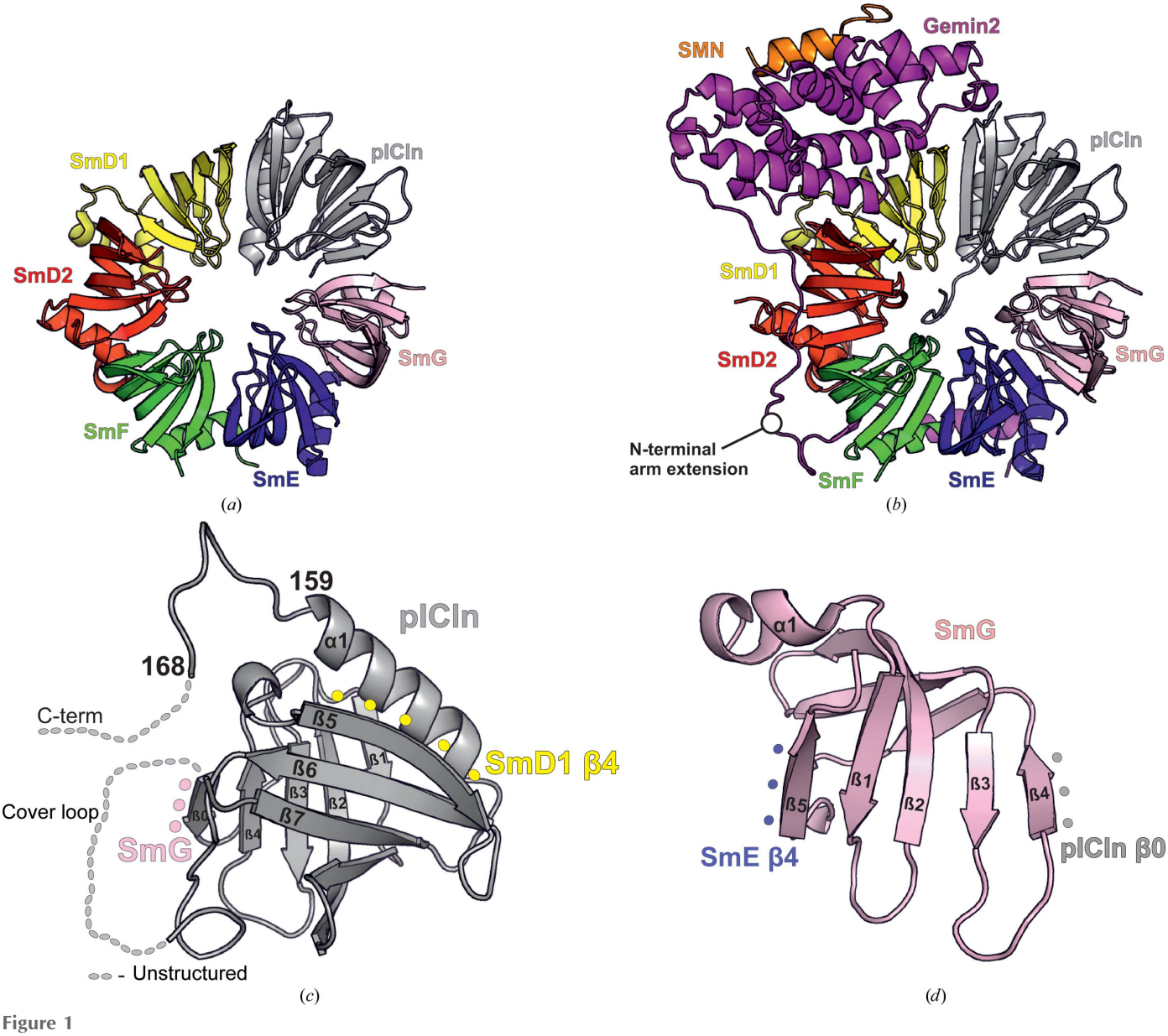


Figure 1
(*a*) Structure of the 6S complex from the 6S-*C*2 crystals in cartoon representation. The colour code for individual proteins (SmD1, yellow; SmD2, red; SmE, blue; SmF, green; SmG, pink; pICln, grey) is used throughout the manuscript. (*b*) Crystal structure of the 8S complex in cartoon representation. The colour code for the two additional proteins (SMN, orange; Gemin2, magenta) is used throughout the manuscript. The N-terminal 'arm' extension of Gemin2 is highlighted. (*c*) The crystal structure of pICln from the 8S crystals in cartoon representation. Unstructured regions are indicated by grey dots. Yellow dots indicate the position of SmD1 β-strand β4 and pink dots indicate the position of SmG β-strand β5. (*d*) Crystal structure of SmG from the 6S-*C*2 crystal in cartoon representation. The position of SmE β-strand β4 is indicated by blue dots and the position of pICln β0 is indicated by grey dots.

3.2. Protein-construct optimization

To optimize the 6S and 8S complex molecules for crystallization propensity, we decided to focus on a single complex subunit and selected the assembly chaperone pICln (Fig. 1*c*) as a target for the following reasons.

(i) pICln is a component of both the 6S and the 8S complexes.

(ii) An NMR structure of *Canis lupus familiaris* pICln was available at the time (Fürst *et al.*, 2005).

(iii) The NMR structure (and the derived homology model) revealed the presence of an extended flexible loop connecting strands β6 and β7 as well as an extended unstructured C-terminus.

The N-terminal residues (1–160) of *D. melanogaster* pICln form a pleckstrin-homology (PH) domain, whereas for the C-terminal residues (161–215) no secondary structure was predicted by *JPred* (Cole *et al.*, 2008) or *PSIPRED* (McGuffin *et al.*, 2000). An additional region of significantly increased flexibility was identified by NMR studies within an extended stretch of residues (residues 93–115 of *C. l. familiaris* pICln, corresponding to residues 78–132 of *D. melanogaster* pICln), which bridge β-strands β6 and β7 of the PH fold (Fürst *et al.*, 2005) and are referred to as the 'cover loop'. The incorporation of intrinsically unstructured protein regions into the crystal lattice increases the entropic cost of crystallization and the presence of extended termini or loop regions might therefore prevent crystallization (Derewenda, 2010). Hence, the elimination of flexible protein regions to increase the crystallization propensity of a given protein is a useful concept that has successfully been applied to many different targets (Chen *et al.*, 1998; Dale *et al.*, 1999; Martin *et al.*, 2000; Yeh *et al.*, 1996). In the case of pICln, the unstructured parts of the protein comprise three acidic regions (AR1–AR3; Schedlbauer *et al.*, 2011), which are the main determinants of its strong and evolutionarily conserved negative

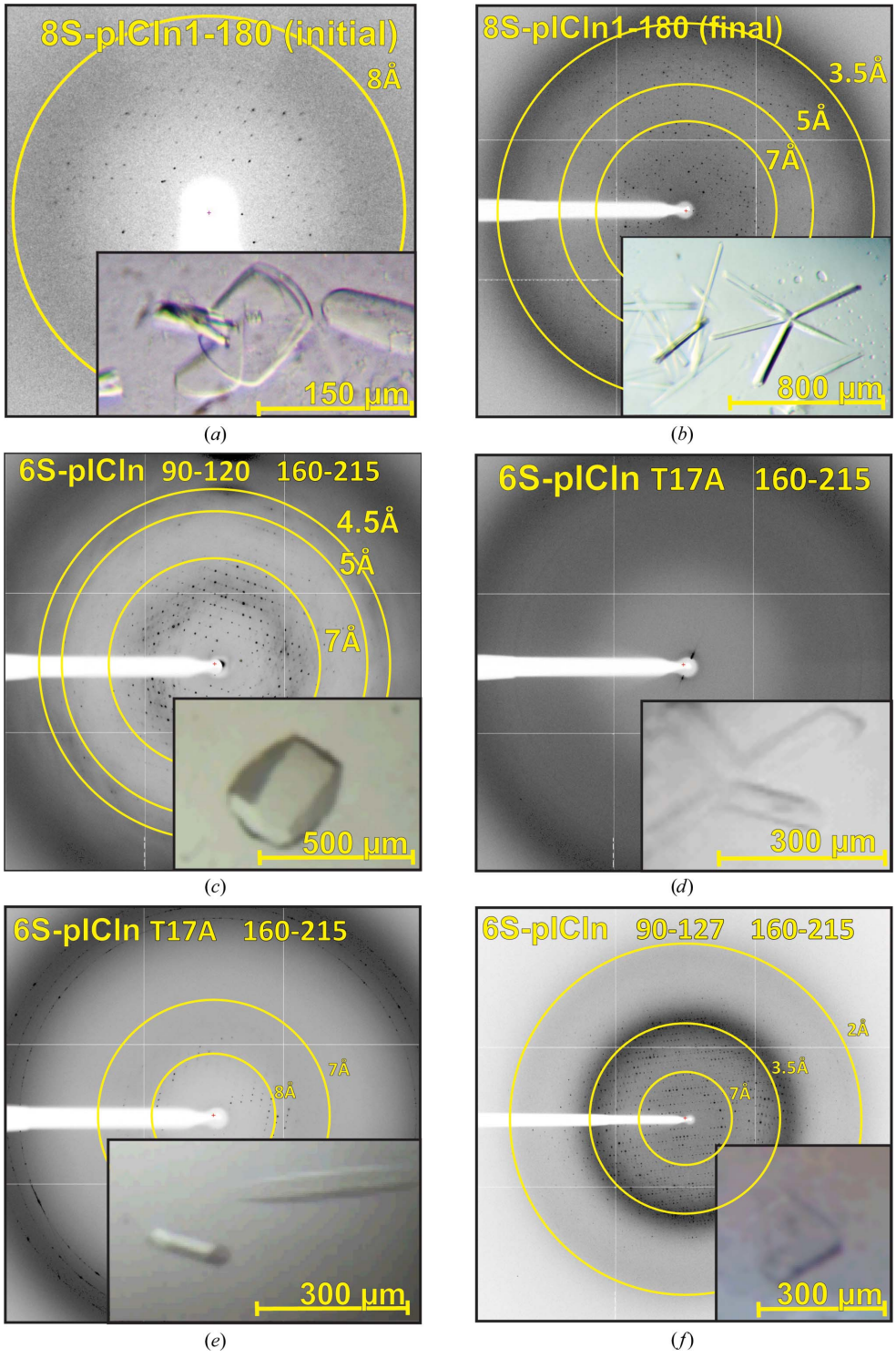


Figure 2
(*a*) Diffraction pattern and morphology of the initial 8S crystal form. In the crystal image, three plate-like crystals are visible. The crystal in the middle is twinned, while the left and right crystals appear to be single. (*b*) Diffraction pattern and morphology of the optimized 8S crystals obtained with 8S preparations containing pICln (Δ181–215). (*c*) Diffraction pattern and morphology of the 6S-*P*321 crystal form. (*d*) Diffraction pattern and morphology of the 6S-T17A-1 crystal form. (*e*) Diffraction pattern and morphology of the 6S-T17A-2 crystal form. (*f*) Diffraction pattern and morphology of the 6S-*C*2 crystal form.

overall charge (Grimm *et al.*, 2013). Hence, truncation of its unstructured regions is accompanied by a significant reduction in the magnitude of its strongly negative overall electrostatic charge. This is attributable to the fact that AR1 is situated in the cover loop and AR2 and AR3 in the unstructured C-terminus. By preparing different truncation and deletion constructs of pICln, we therefore expected pronounced changes in the physicochemical properties of the derived complex variants. Firstly, a series of C-terminal truncation variants was prepared in steps of five amino acids until the α-helix of the PH fold at residue 159 was reached (Fig. 1*c*). The respective constructs could be expressed in a soluble form and eluted from the SEC columns without exception within a single, monodisperse peak corresponding to a monomeric state. The resulting 6S and 8S complex variants were then subjected to a comprehensive crystallization screen. As a result, we observed a change in the crystallization behaviour of a particular 8S variant. An 8S complex reconstituted with *D. melanogaster* pICln (Δ181–215) yielded crystals (Fig. 2*b*) that grew significantly larger than the initial crystal form, although they belonged to the same space group with similar unit-cell parameters. Interestingly, this crystal form exhibited more reliable nucleation and appeared not only in conditions containing SOKALAN CP 42 but also in the presence of higher molecular-weight PEG and PEG MME polymers.

In the next step, C-terminal deletions of *D. melanogaster* pICln were combined with a deletion of the cover loop (Δ90–120). These double-deletion constructs likewise yielded soluble, monodisperse protein preparations and could be integrated into the 6S and 8S complexes. Notably, the shortest of these pICln constructs (Δ90–120, Δ160–215) did not contain any of the three flexible acidic regions. 6S complexes reconstituted with this protein yielded large single crystals with high reproducibility utilizing sodium formate as a precipitant (Table 1). However, the inherent diffraction of these crystals was still limited to roughly 4 Å resolution at synchrotron beamlines (Fig. 2*c*). We were able to solve the structure and to build an initial model (unpublished data). According to their space group, these crystals will be referred to as the 6S-*P*321 form in the following.

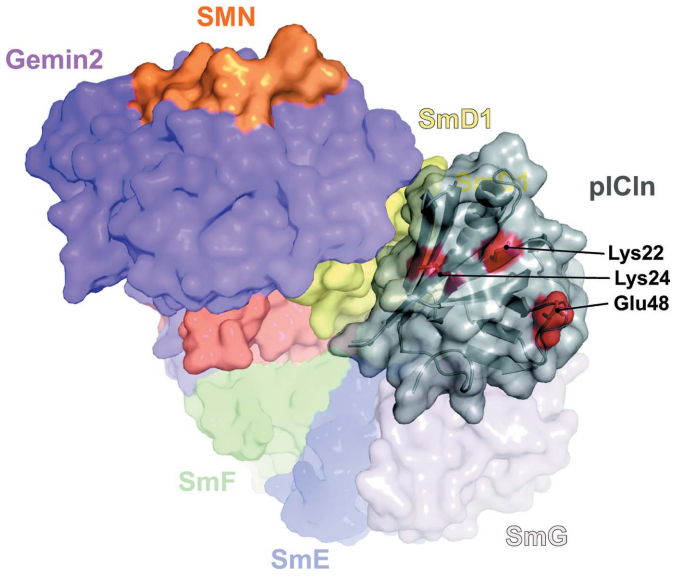


Figure 3
Crystal structure of the 8S complex shown as a Connolly surface. The positions of pICln residues Lys22, Lys24 and Glu48 (in red) on the surface of the protein are indicated by arrows.

3.3. Surface-entropy reduction

Encouraged by this initial success, we aimed at further modifications of the protein surface to improve the crystal quality and crystallization behaviour. An investigation of existing protein structures revealed that the occurrence of large polar side chains is negatively correlated with crystallization propensity (Price, 2009). Likewise, a preference for small hydrophobic amino acids in the vicinity of crystal contacts is observed (Cieślik & Derewenda, 2009). To optimize both complexes in this regard, we applied a strategy known as surface-entropy reduction (SER). The rationale behind SER is to minimize the local entropy loss during crystal formation by exchanging flexible surface residues for more rigid side chains such as alanine (Longenecker *et al.*, 2001). This has successfully been applied in the determination of several novel protein structures (Dale *et al.*, 1999; Prag *et al.*, 2003; Jiang *et al.*, 2005) as well as in the generation of improved crystals diffracting to higher resolution (Jin *et al.*, 2005; Pornillos *et al.*, 2009). *D. melanogaster* pICln surface residues were identified with the help of a homology model created from the *C. l. familiaris* NMR structure (Fürst *et al.*, 2005). We then selected highly solvent-exposed lysine and glutamate residues to convert them into residues with low conformational flexibility, including the point mutations K22A, K24A and E48S (Fig. 3). The three targeted residues are grouped together in a single, narrow surface patch, a factor that can reportedly enhance the effectiveness of the SER strategy (Cooper *et al.*, 2007; Derewenda, 2011). Different pICln variants carrying a single or up to three SER mutations in addition to truncations of the 'cover loop' and/or the C-terminus could all be overexpressed in *E. coli* in a soluble form. After purification, the SER-modified pICln preparations were used in reconstitution of the 6S and 8S complexes, which were then screened exhaustively for crystallization. The SER-modified 6S and 8S complexes yielded crystals similar to those observed for their SER-unmodified counterparts; however, no improvements in crystallization behaviour or X-ray diffraction were observed and no new crystal forms could be obtained.

3.4. Data collection and structure solution of the 8S crystal structure

Owing to the so far limited success of the construct-optimization strategy, we decided to abandon the search for further crystal forms of the 8S complex and to solve the structure using the existing crystals (Table 1). The inherently limited diffraction power of these crystals and the necessary trade-off between radiation damage and signal-to-noise ratio made it impossible to collect a complete data set from a single

crystal to an overall resolution that was better than 5 Å. To compensate for the weak diffraction, we focused on crystal size optimization. Therefore, we carefully adjusted the precipitant concentration of the crystallization batches until a very slow growth to a size of several hundreds of micrometres was achieved (Fig. 2*b*). To preserve their diffraction quality, it turned out to be crucial to transfer these large crystals very slowly over at least five steps of increasing cryoprotectant concentrations into their final cryocondition (Table 1). To utilize the large available crystal volume effectively during data collection, we pursued different protocols with stepwise shifting of the crystal along the φ axis as well as helical data collection at different synchrotron beamlines (Flot *et al.*, 2010). After testing more than 100 crystals with such data-collection protocols it became obvious that the amount of anisomorphism that existed over distant parts of the crystals prevented effective scaling of the different partial data sets. In contrast, several data sets collected conventionally on beamline ID14-4 of the ESRF, a beamline with a comparably wide and defocussed beam that is able to illuminate a comparably large crystal volume (McCarthy *et al.*, 2009), yielded reasonable data-collection statistics. Among roughly 100 data sets collected from different crystals on this beamline, a single data set finally stood out with an anisotropic diffraction limit of 3.1 Å resolution in the best direction. Remarkably, the initial data sets appeared to belong to space group *C*2, with unit-cell parameters $a = 256.8$, $b = 71.6$, $c = 149.9$ Å, $\alpha = 90$, $\beta = 117.6$, $\gamma = 90^\circ$ (see Supplementary Table S1). A close examination of diffraction images from the final, best data set, however, revealed a significant number of consistently weak reflections indicative of a larger unit cell. After fine-tuning of the auto-

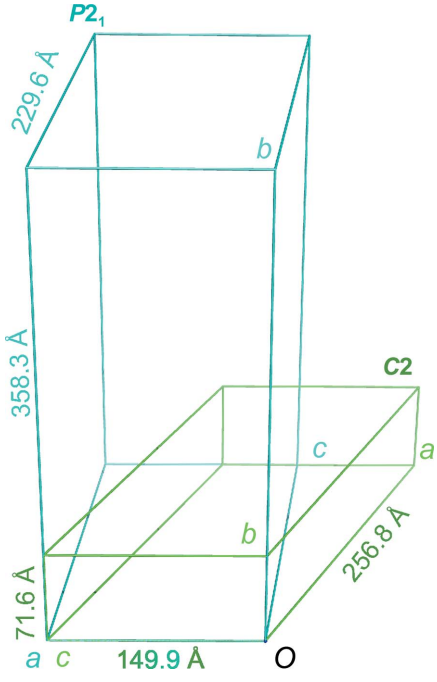


Figure 4
Crystal unit cells of the 8S complex. The unit cell of the originally indexed *C*2 cell is shown in green and the correct $P2_1$ cell is shown in blue. Unit-cell axes are labelled in the corresponding colours.

Table 4
NCS translation vectors as calculated with *MOLREP* using data in the resolution range 15–4 Å.

The *MOLREP* score for this run was 0.4.

Peak No.	Height (%)	*x*	*y*	*z*
1	84.5	0.500	0.301	0.500
2	62.3	0.000	0.398	0.000
3	40.7	0.500	0.102	0.500
4	27.9	0.000	0.199	0.000
5	22.0	0.500	0.500	0.500

indexing parameters to include these weak reflections, the raw data set could be processed in space group $P2_1$ with unit-cell parameters $a = 149.9$, $b = 358.3$, $c = 229.6$ Å, $\alpha = 90$, $\beta = 97.7$, $\gamma = 90^\circ$ with different data-processing suites (Table 2; see also Supplementary Table S2 and Fig. S1; Otwinowski & Minor, 1997; Kabsch, 2010). An inspection of the unit-cell parameters revealed obvious relationships between the *C*2 and $P2_1$ unit cells (Fig. 4). The length of the *c* axis of the *C*2 cell is identical to that of the *a* axis of the $P2_1$ cell, and the *b* axis of the $P2_1$ cell is five times longer than that of the *C*2 cell. The factor five is also reflected in the unit-cell volume, which increases from 2.46×10^6 to 12.34×10^6 Å^3. As a consequence, the total number of molecules increases from eight in the *C*-centred cell (two per asymmetric unit) to 40 in the primitive cell (20 per asymmetric unit). Native Patterson calculations in space group $P2_1$ with different programs revealed the existence of five significant NCS translation vectors with heights of greater than 20% of the origin peak (see Table 4). The observed significant translational NCS also influences statistical twinning tests (Supplementary Fig. S2). The translation vectors can be rationalized in light of the packing of the 20 8S complexes in the asymmetric unit, which are arranged into two groups of ten protomers. The two groups are related to each other *via* a translation along *x* and *z* of 0.5 (fractional coordinates). Within each group of ten 8S complexes the protomers are arranged in such a way that they can be interconverted *via* translations of roughly 2/5 and 1/5 along the *y* axis. The translation along 1/5 of the *y* axis again reflects the relationship in the *b* axis between the initially and incorrectly assigned space group *C*2 ($b = 71.6$ Å) and the true space group $P2_1$ ($b = 358.3$ Å).

Attempts to solve the structure by molecular replacement (MR) with the available coordinates of *C. l. familiaris* pICln, single Sm proteins (see Fig. 1*d* for the structure of the Sm fold, exemplified by SmG) or the SmD1/D2 or SmF/E/G hetero-oligomers failed. We therefore generated several hetero-hexameric search models with the PH fold of pICln integrated into the Sm ring in an Sm protein-like fashion. These modelling attempts were guided by the available structures of the homohexameric bacterial Sm protein homologue Hfq (Sauter *et al.*, 2003), the heteroheptameric Sm core (Pomeranz Krummel *et al.*, 2009; Weber *et al.*, 2010; Leung *et al.*, 2011) and the assumption that the pICln PH fold functionally mimics an Sm unit. However, MR with these manually constructed models likewise did not yield useful MR solutions. Presumably, this was owing to the large atomic differences of the

search model from the actual crystal structure and the unavailability of atomic data for Gemin2 and SMN. As only weak heavy-atom derivatives of the crystals could be prepared, phasing of the data set was not possible until the publication of a late snRNP assembly intermediate containing all of the components of the 8S complex except pICln (Zhang *et al.*, 2011). Using the coordinates of this complex, molecular replacement was carried out utilizing *MOLREP* incorporating the highest NCS translation vector ($x = 0.5$, $y = 0.301$, $z = 0.5$) into the search, resulting in the placement of 12 copies of the 8S complex. At this point it was not clear how many copies of the 8S complex should be expected, since values of between 12 and 24, corresponding to Matthews coefficients of 4.3 and 2.1 Å^3 Da^{-1}, respectively, were all compatible with commonly observed values. This partial model was refined with *PHENIX* and inspection of the resulting residual electron-density maps revealed the existence of additional copies of the 8S complex. These were located by an iterative procedure utilizing the known partial model as a fixed solution in *MOLREP* and performing a search for additional copies in the electron-density map. Via this procedure, eight additional copies of the 8S complex could be located. The final model thus contained 20 molecules of the 8S complex in the asymmetric unit that feature small but significant conformational differences, the implications of which have been described in detail elsewhere (Grimm *et al.*, 2013). These differences represent one of the reasons why the larger $P2_1$ space group represents the correct choice, as this conformational heterogeneity is not distributed evenly across the 8S complex but is localized primarily in the pICln subunit. In the case of an incorrect assignment of the larger $P2_1$ cell one would expect the changes to be distributed throughout all of the subunits, being most pronounced in the more poorly defined regions (loops and termini). Furthermore, the extra weak reflections (see above) were only accounted for with the larger $P2_1$ cell and should not exist in the smaller cell (see also Supplementary Fig. S1). Finally, the structure could not be satisfactorily refined to R factors below 0.39 for R_{free} and 0.35 for R in space group $C2$, reflecting a fundamental crystallographic problem. Taken together, these observations rule out space group $C2$ as the correct choice. Of note, the eight complex molecules that could not be located with the initial *Phaser* run did not stand apart in terms of r.m.s.d. or overall B factors. When the MR search was repeated with the latest version of *Phaser* (v.2.5.6), all 20 molecules were found at once.

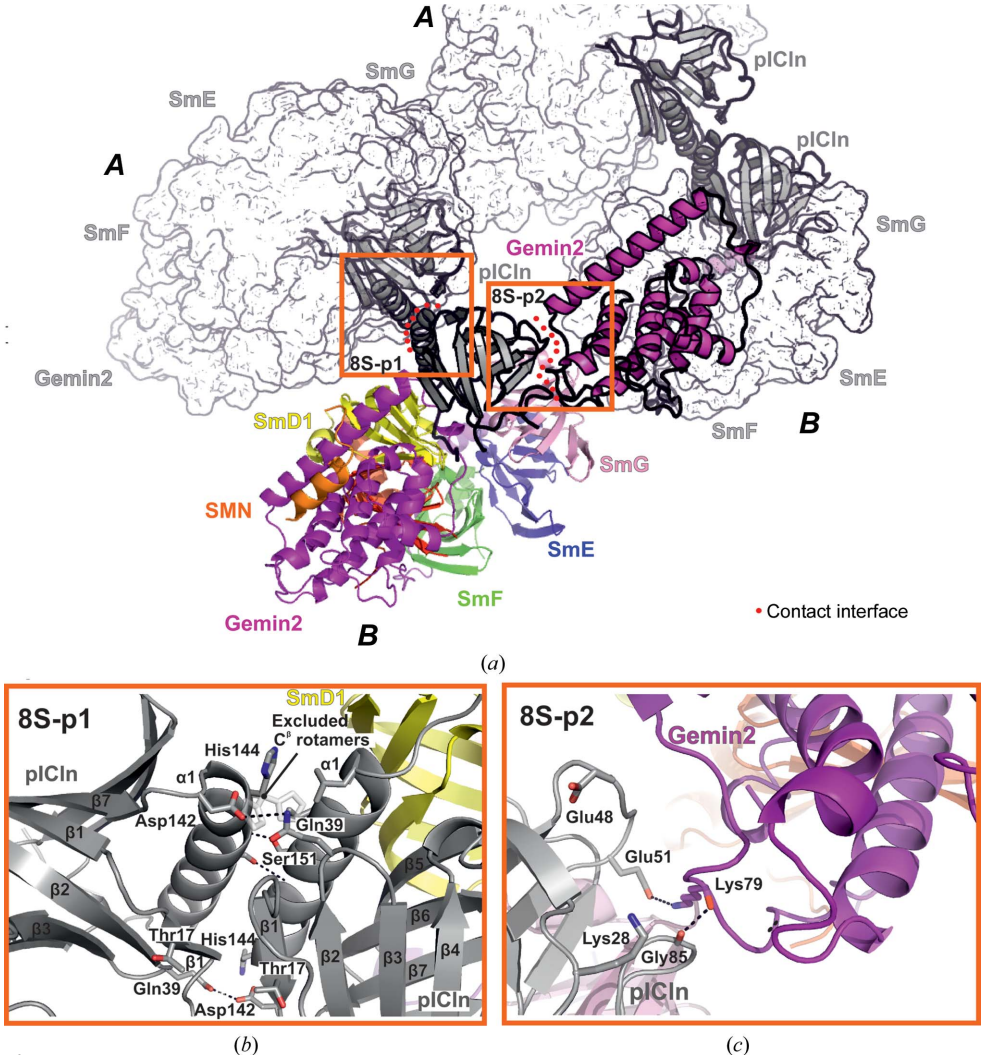
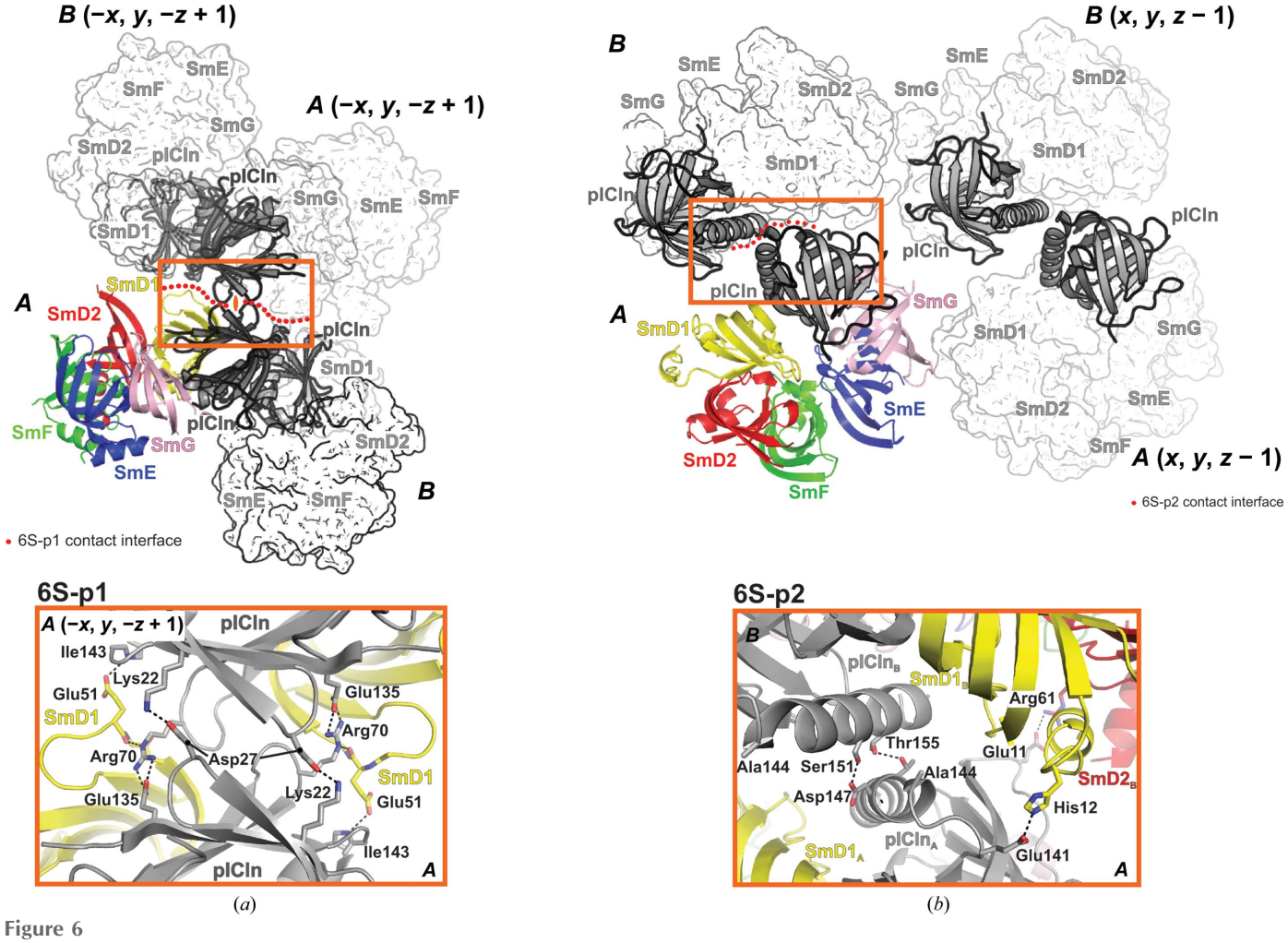


Figure 5
(*a*) Location of 8S-p1 and 8S-p2 within the 8S crystal structure. Four molecules of the 8S complex are shown. Both *A* molecules are represented as a Connolly surface and their pICln subunits are shown in cartoon representation. One *B* molecule is shown in cartoon representation and subunits are colour-coded as in Fig. 1. The second *B* molecule is shown as a Connolly surface and its pICln and Gemin2 subunits are shown in cartoon representation. The locations of the 8S-p1 and 8S-p2 contact patterns are indicated by orange boxes and the contact interface is indicated by red dots. (*b*) Detailed view of the 8S-p1 contact pattern. (*c*) Detailed view of the 8S-p2 contact pattern.

3.5. Rational engineering of 6S crystal contacts yields new crystal forms

To apply the insight gained from the 8S crystal structure to the optimization of the 6S complex constructs, we hypothesized that for the common set of subunits in the two complexes there might also be common, recurring contact patterns within the architecture of the 6S and 8S crystals. Therefore, we closely examined the crystallographic contacts within the 8S crystal

lattice. We recognized that the 20 molecules found within the asymmetric unit can be grouped into ten pairs of two 8S complex molecules (*A* and *B*) based on their interaction pattern (Fig. 5*a*). We could discern two distinct types of contact patterns that involved pICln. The first pattern, referred to as 8S-p1, is observed between the pICln subunits of adjacent *A* and *B* molecules (Figs. 5*a* and 5*b*). The second pattern, referred to as 8S-p2, is observed among molecules belonging to either the *A* or *B* group. It involves the pICln subunit of the first molecule and the Gemin2 subunit of the second molecule (Figs. 5*a* and 5*c*). Contact 8S-p1 is homotypic (two identical patches interact across the interface) and comprises the β1 strands (Leu15, Tyr16, Thr17 and Ala18) and the α1 helices (residues Asp142–Leu159) of the interacting pICln subunits (Fig. 5*b*). Each of the interfacing α1 helices of pICln harbours two equivalent histidine residues (His144) at their N-terminal ends. These are sterically highly restricted, as a rotation around their C^{β} atoms would result in clashes with adjacent residues. Hence, most rotamers of pICln His144 are sterically hindered in this context, resulting in a significant loss of conformational entropy upon the formation of this crystal contact (Fig. 5*b*). We therefore speculated that an H144A substitution at this position in pICln would boost the crystallization propensity of the respective 6S and 8S complex variants. Within the 8S-p1 contact, two equivalent pICln threonine residues (Thr17) are found to be located in close proximity to each other (Fig. 5*b*). This arrangement allowed the alteration of both contact sides with a single substitution. As all amino acids in the β1 strand of pICln had hydrophobic character (Leu15, Tyr16, Thr17 and Ala18), we introduced a T17A substitution to optimize this contact. Interface 8S-p2 is heterotypic (two dissimilar patches interact across the interface) and involves patches of pICln and Gemin2 (Figs. 5*a* and 5*c*). With an average of 386 Å^2 of buried surface area, the contact patterns of the 8S-p2 type are smaller than those of the 8S-p1 type, with an average buried surface of 652 Å^2. The



(*a*) (*b*)

Figure 6
(*a*) Location of the 6S-p1 contact pattern within the 6S-*C*2 crystal structure. The contact builds up the crystal lattice along the crystallographic *c* axis and shows twofold symmetry. Four molecules of the 6S complex are shown in the figure. Both *B* molecules and the *A* (−*x*, *y*, −*z* + 1) molecule are represented as Connolly surfaces. The *A* molecule is represented in cartoon style. All four pICln molecules are shown as grey cartoons. The orange box in the upper picture indicates the location of the contact. The interfacing region between the *A* (*x*, *y*, *z*) and *A* (−*x*, *y*, −*z* + 1) 6S complex molecules is indicated by red dots. The lower orange box contains a detailed view of the contact pattern. (*b*) Location of the 6S-p2 contact pattern within the 6S-*C*2 crystal structure. The contact in the 6S-*C*2 crystal lattice is indicated by the upper orange box and red dots indicate the contact interface. A detailed view of the contact is shown within the lower orange box.

residues of pICln that are buried in this interface are Lys28, residues 47–51 from the loop connecting β-strands $\beta2$ and $\beta3$ and residues 83–86 from the N-terminal part of the cover loop. These amino acids contact residues 76–88 of Gemin2, which are located in the N-terminal 'arm' extension (Fig. 1*b*), close to the C-terminal domain of the protein (Fig. 5*c*). Three pICln residues (Lys28, Glu48 and Glu51) with high conformational flexibility were located within the buried surface of this contact. Of these, Lys28 and Glu48 were chosen for substitution by alanine. Glu51 formed a salt bridge with Lys79 of Gemin2 and was therefore not chosen for substitution. The 6S and 8S complexes were reconstituted and screened for crystallization with four different *D. melanogaster* pICln (Δ160–215) constructs carrying single alanine substitutions replacing Thr17, Lys28, Glu48 or His144. The respective previous crystal forms of either the 6S or 8S complex were still obtained with these modified complex preparations, but unfortunately with no improvement in diffraction quality.

Two new crystal forms were obtained with 6S complex preparations containing the pICln T17A substitution (6S-T17A-1 and 6S-T17A-2; see Table 1). However, the crystals of the 6S-T17A-1 type showed no diffraction and those of the 6S-T17A-2 type only showed limited diffraction to roughly 7 Å resolution. No further optimization of the two crystal forms was attempted and only test diffraction images were collected for 6S-T17A-1 and 6S-T17-2 (Figs. 2*d* and 2*e*). No new crystal forms could be obtained for the corresponding 8S complex preparations.

3.6. A high-resolution crystal structure of the 6S complex

After the limited success of the optimization experiments for the crystallization of the 6S complex performed thus far, we combined the truncations of the pICln 'cover loop' (Δ90–125) and of its C-terminus (Δ160–215) with alanine substitutions of the surface residues Thr17, Lys28, Glu48 and His144. From a 6S complex preparation reconstituted with the pICln variant carrying the H144A substitution, a new, better diffracting crystal form (6S-*C*2; see Table 1 and Fig. 2*f*) was obtained, resulting in the collection of a data set to 1.9 Å resolution on beamline ID14-4 at the ESRF, Grenoble, France. The crystal structure was solved by molecular replacement using the coordinates of the hexameric Sm/pICln ring extracted from the previously solved 8S crystal structure. The crystals belonged to space group *C*2 and contained two molecules of the 6S complex in the asymmetric unit. The 6S-*C*2 crystals feature an unusually tight packing with a rather low solvent content of 39%. Within the rigid and dense crystal lattice, pICln is involved in two distinct contact types referred to as 6S-p1 (Fig. 6*a*) and 6S-p2 (Fig. 6*b*). Contact 6S-p1 builds up the crystal lattice along the crystallographic *c* axis and possesses twofold rotational symmetry. At 780 Å^2, it buries the largest surface area of all crystal contacts within the 6S crystal structure and covers more surface area than some of the biological intersubunit interfaces of the 6S complex. Contact 6S-p1 is formed between neighbouring *A*-type molecules (x, y, z and $-x$, y, $-z+1$) and buries pICln residues Lys22, Lys24, Asp27, Lys28 and Val29, which are located in the loops connecting β-strands $\beta2$ and $\beta3$, pICln residues 66–71 from the loop connecting β-strands $\beta5$ and $\beta6$, and pICln residues 135–143 that connect $\beta7$ to helix $\alpha1$. Several hydrogen bonds and salt bridges are formed between pICln residues Lys22, Asp27, Asn68, Pro69, Arg70, Glu135 and Ile143 (Fig. 6*a*, orange box). Of note, each of the pICln Lys22 residues from both sides of the contact forms a salt bridge with Asp27 of the respective neighbouring molecule. Lys22 had already been targeted during our initial SER experiments (Figs. 3 and 6*a*, orange box; compare with §3.3). Contact 6S-p2 is formed between the *A* and *B* copies of the 6S complex within the asymmetric unit (Fig. 6*b*). While 6S-p1 was not comparable to any of the contacts found in the 8S crystals, 6S-p2 resembles the 8S-p1 contact (compare Figs. 5*a* and 5*b* with Fig. 6*b*). 6S-p2 is likewise mainly based on the interaction between two adjacent $\alpha1$ helices of neighbouring pICln molecules. Accordingly, the substituted alanine residue (H144A) is also located in this crystal contact (compare Fig. 5*b* with Fig. 6*b*, orange box). However, the relative orientation of the complex molecules on both sides of the contact is not exactly preserved between the

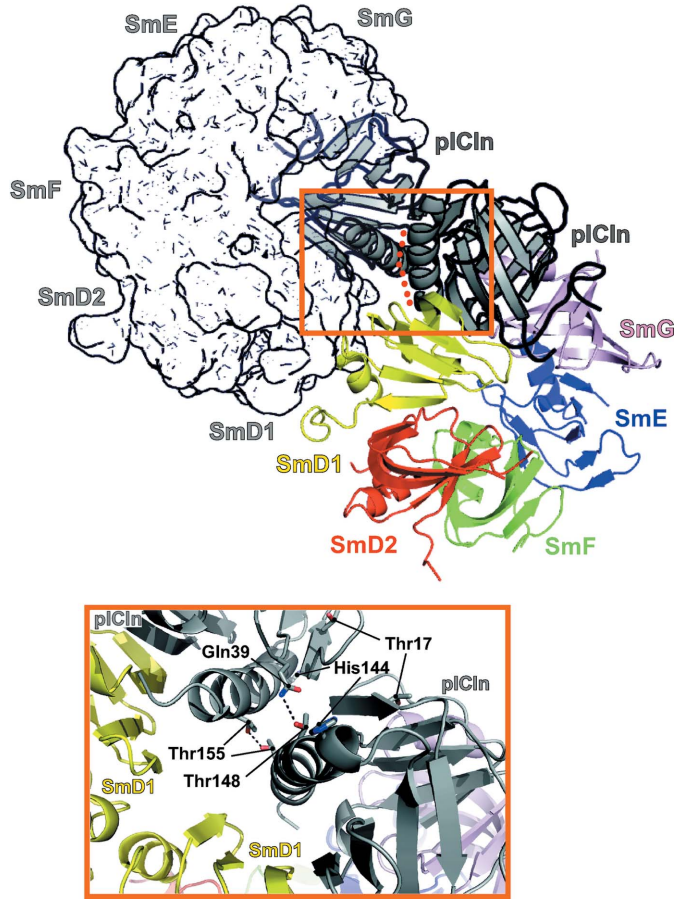


Figure 7
A contact pattern similar to the 8S-p1 (Figs. 5*a* and 5*b*) and 6S-p2 contacts is also found within the 6S-*P*321 crystal form. Two molecules of the 6S complex are shown in the figure. The colour-coded 6S molecule is orientated equivalently to the colour-coded 8S molecule in Fig. 5(*a*) and the colour-coded 6S molecule in Fig. 6(*b*). The orange box shows a detailed view of the crystal contact.

6S-*C*2 and 8S crystal forms. While both helices contact each other in an antiparallel fashion in the 8S-p1 region, these helices are tilted by only an angle of roughly 50° in the 6S-p2 contact (Fig. 6*b*, orange box). The latter arrangement leads to an asymmetric, heterotypic contact pattern in 6S-p2 as two different surface patches of both complexes approach each other. Still, the 8S-p1 and the 6S-p2 contact are both built upon a common set of residues. Of note, a similar crystal contact (Figs. 7*a* and 7*b*) is also observed in the 6S-*P*321 crystals (Tables 1 and 2). On the other hand, 8S-p2 has no equivalent in both the 6S-*C*2 and 6S-*P*321 crystals and the pICln surface patch that was involved in its formation points towards the interstitial space in the 6S-*C*2 crystals (Fig. 6*a*). Despite the comparably large area of contact 6S-p1 it seems to be unlikely that the biological structure of the complex is significantly influenced by the crystal environment, as the two molecules in the asymmetric unit do not exhibit significant differences. Furthermore, owing to its ring structure, the 6S complex is a rugged and mechanically stable particle.

3.7. An engineered alanine residue coordinates the formation of an extended water network

As the mechanically rigid and strongly diffracting 6S-*C*2 crystals grew exclusively from surface-engineered 6S preparations carrying the H144A point mutation, we wanted to elucidate the role of this single amino-acid substitution within the crystal lattice. We therefore carefully examined the environment of the *A* and *B* copies of residue Ala144 within crystals of the 6S-*C*2 type. Ala144 of the *B* molecule is located only in the periphery of the contact and hence a substitution by a histidine residue would not inflict any conformational restrictions. The corresponding histidine residue in molecule *A*, however, would be restricted conformationally and (similar to the situation in the 8S-p1 contact) only a single C^{β} rotamer would be sterically possible. After a closer inspection of this region of the crystal structure, we became aware of a remarkable, extended water network in the vicinity of Ala144. While a total of 372 ordered water molecules were found within the asymmetric unit, only 17 bridge neighbouring complex molecules by the formation of intermolecular hydrogen bonds. 13 of the 17 water molecules are located within the 6S-p2 contact interface formed between the NCS-related *A* and *B* complex molecules (Figs. 6*b* and 8*a*). These water molecules are arranged in a clathrate-like cage around Ala144 and form a continuous network of hydrogen bonds (Fig. 8*c*). Residues involved in water-mediated intermolecular hydrogen bonds include Tyr16, Asp142 and Thr148 of pICln from molecule *A* and Tyr150, Ser151, Thr155, Gln157 of pICln as well as His39 of SmD1 from molecule *B*. This interaction network is likely to contribute to the stabilization of the 6S-p2 crystal contact. A histidine at the position of Ala144, as occurs in the native pICln construct, would produce clashes with a majority of the intermolecular water molecules. This is a likely explanation for the fact that the 6S-*C*2 crystal form was exclusively obtained from 6S complex preparations harbouring the H144A substitution within the pICln subunit. Of note, an 8S complex preparation reconstituted from the 6S variant that yielded the 6S-*C*2 crystals formed crystals that were indiscernible from the previous 8S crystals in terms of morphology and diffraction behaviour.

3.8. Lessons learned for the rational optimization of crystal contacts

The successful outcome of a crystallographic project is highly dependent on the surface properties of the respective target protein (Dale *et al.*, 2003). Likewise, the intermolecular contacts between protein molecules of a crystal are highly specific for each individual protein and crystal form. To date, no reliable rational prediction of crystal contacts is possible. Hence, in cases where the original target protein did not crystallize, the utilization of orthologous proteins that might feature altered surface properties was suggested as early as 1972 (Campbell *et al.*, 1972). Today, genetic manipulations and the possibility of synthesizing entire coding sequences permit the efficient and targeted application of strategies such as SER. Nevertheless, as long as the SER-targeted surface

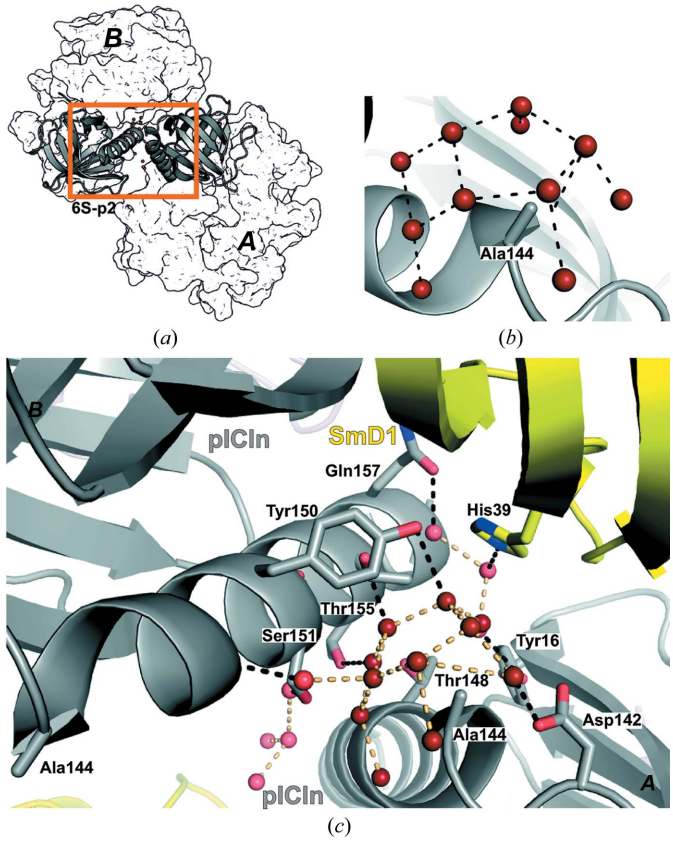


Figure 8
(*a*) Location of 13 water molecules within the 6S-p2 crystal contact between NCS-related 6S complex molecules of the 6S-*C*2 crystal form. pICln molecules are represented in cartoon style, while the rest of the complex subunit is shown as a Connolly surface. (*b*) 11 water molecules form a clathrate-like cage that is coordinated by pICln residue Ala144. (*c*) Detailed view of the hydrogen-bonded network formed in the 6S-p2 crystal contact. Hydrogen bonds between water molecules are indicated by dashed lines coloured beige. Hydrogen bonds between protein residues and water molecules are indicated by dashed lines coloured black.

patches are not involved in the formation of crystal contacts, they are unlikely to change the crystallization behaviour of the targeted protein constructs. In our attempts to crystallize the 6S and 8S complexes, we chose Lys22, Lys24 and Glu48 from a single pICln surface patch as targets for SER optimization. These modifications on pICln alone did not lead to new crystal forms of the resulting 6S and 8S complexes, and in two of the three crystal structures that were eventually solved the targeted surface patch was not buried within a crystallographic interface. However, our postulation that crystal contacts that had been observed in the newly solved 8S crystal structure might be important hotspots for contacts in potential crystals of the related 6S complex turned out to be fruitful for the identification of relevant SER sites. Conversely, the newly generated high-resolution crystal structure of the 6S complex was crucial for the accurate refinement of the low-resolution 8S crystal structure.

It has previously been reported (Smyth *et al.*, 2003) that crystal contacts can be facilitated by the addition of crystallization-optimized fusion proteins serving as 'crystallization chaperones'. While this strategy has mainly been successful for small monomeric proteins where the crystallization chaperone is the major component within the unit cell (Suzuki *et al.*, 2010; Chun *et al.*, 2012; Zou *et al.*, 2012), the identification and subsequent crystallization optimization of surface patches possessing a high propensity to form crystal contacts turned out to be the key for the generation of high-quality 6S complex crystals. The underlying principle of the crystallization-chaperone method and our strategy might be the same: the addition or generation of protein surfaces with a high propensity to form crystal contacts. Hence, in contrast to early macromolecular crystallographers, who were restricted to endogenous protein from species suitable for laboratory cultivation, the modern protein crystallographer seems to be less and less at the mercy of a purely stochastical crystallization process. In fact, the current bioinformatics, cloning and protein-production techniques in synergy with a solid understanding of the targeted protein system increasingly enables the application of rational strategies for the structure solution of recalcitrant complexes and proteins.

Acknowledgements

This work was financially supported by a grant to UF (Fi573/7-2) and the RVZ (FZ 82) from the DFG. We thank the team at beamlines 14.1, 14.2 and 14.3 of BESSY, the team at beamlines ID14-4, ID23-1, ID-23-2 and ID29 of the ESRF and the team at beamlines PXI and PXII of the SLS for their excellent support during data collection.

References

Adams, P. D. *et al.* (2010). *Acta Cryst. D***66**, 213–221.
Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Haendler, B. & Jacob, M. (1982). *EMBO J.* **1**, 1259–1265.
Bryksin, A. V. & Matsumura, I. (2010). *Biotechniques*, **48**, 463–465.
Campbell, J. W., Duée, E., Hodgson, G., Mercer, W. D., Stammers, D. K., Wendell, P. L., Muirhead, H. & Watson, H. C. (1972). *Cold Spring Harb. Symp. Quant. Biol.* **36**, 165–170.
Carter, P. (1986). *Biochem. J.* **237**, 1–7.
Chari, A., Golas, M. M., Klingenhäger, M., Neuenkirchen, N., Sander, B., Englbrecht, C., Sickmann, A., Stark, H. & Fischer, U. (2008). *Cell*, **135**, 497–509.
Chari, A., Paknia, E. & Fischer, U. (2009). *Curr. Opin. Cell Biol.* **21**, 387–393.
Chen, G.-Q., Sun, Y., Jin, R. & Gouaux, E. (1998). *Protein Sci.* **7**, 2623–2630.
Chun, E., Thompson, A. A., Liu, W., Roth, C. B., Griffith, M. T., Katritch, V., Kunken, J., Xu, F., Cherezov, V., Hanson, M. A. & Stevens, R. C. (2012). *Structure*, **20**, 967–976.
Cieślik, M. & Derewenda, Z. S. (2009). *Acta Cryst.* D**65**, 500–509.
Cole, C., Barber, J. D. & Barton, G. J. (2008). *Nucleic Acids Res.* **36**, W197–W201.
Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M. & Derewenda, Z. (2007). *Acta Cryst.* D**63**, 636–645.
Dale, G. E., Kostrewa, D., Gsell, B., Stieger, M. & D'Arcy, A. (1999). *Acta Cryst.* D**55**, 1626–1629.
Dale, G. E., Oefner, C. & D'Arcy, A. (2003). *J. Struct. Biol.* **142**, 88–97.
Derewenda, Z. S. (2010). *Acta Cryst.* D**66**, 604–615.
Derewenda, Z. S. (2011). *Acta Cryst.* D**67**, 243–248.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). *J. Synchrotron Rad.* **17**, 107–118.
Fürst, J. *et al.* (2005). *J. Biol. Chem.* **280**, 31276–31282.
Gliko, O., Neumaier, N., Pan, W., Haase, I., Fischer, M., Bacher, A., Weinkauf, S. & Vekilov, P. G. (2005). *J. Am. Chem. Soc.* **127**, 3433–3438.
Grimm, C., Chari, A., Pelz, J.-P., Kuper, J., Kisker, C., Diederichs, K., Stark, H., Schindelin, H. & Fischer, U. (2013). *Mol. Cell*, **49**, 692–703.
Grimm, C., Chari, A., Reuter, K. & Fischer, U. (2010). *Acta Cryst.* D**66**, 685–697.
Jiang, J., Prasad, K., Lafer, E. M. & Sousa, R. (2005). *Mol. Cell*, **20**, 513–524.
Jin, L., Pandey, P., Babine, R. E., Weaver, D. T., Abdel-Meguid, S. S. & Strickler, J. E. (2005). *Acta Cryst.* D**61**, 1418–1425.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Lührmann, R., Li, J. & Nagai, K. (1999). *Cell*, **96**, 375–387.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Leung, A. K. W., Nagai, K. & Li, J. (2011). *Nature (London)*, **473**, 536–539.
Liautard, J.-P., Sri-Widada, J., Brunel, C. & Jeanteur, P. (1982). *J. Mol. Biol.* **162**, 623–643.
Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). *Acta Cryst.* D**57**, 679–688.
Lührmann, R., Kastner, B. & Bach, M. (1990). *Biochim. Biophys. Acta*, **1087**, 265–292.
Martin, G., Keller, W. & Doublié, S. (2000). *EMBO J.* **19**, 4193–4203.
McCarthy, A. A., Brockhauser, S., Nurizzo, D., Theveneau, P., Mairs, T., Spruce, D., Guijarro, M., Lesourd, M., Ravelli, R. B. G. & McSweeney, S. (2009). *J. Synchrotron Rad.* **16**, 803–812.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). *Bioinformatics*, **16**, 404–405.
Neuenkirchen, N., Chari, A. & Fischer, U. (2008). *FEBS Lett.* **582**, 1997–2003.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Paunov, V. N., Kaler, E. W., Sandler, S. I. & Petsev, D. N. (2001). *J. Colloid Interface Sci.* **240**, 640–643.
Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. (2009). *Nature (London)*, **458**, 475–480.

Pornillos, O., Ganser-Pornillos, B. K., Kelly, B. N., Hua, Y., Whitby, F. G., Stout, C. D., Sundquist, W. I., Hill, C. P. & Yeager, M. (2009). *Cell*, **137**, 1282–1292.
Prag, G., Misra, S., Jones, E. A., Ghirlando, R., Davies, B. A., Horazdovsky, B. F. & Hurley, J. H. (2003). *Cell*, **113**, 609–620.
Price, W. N. *et al.* (2009). *Nature Biotechnol.* **27**, 51–57.
Pu, W. T., Krapivinsky, G. B., Krapivinsky, L. & Clapham, D. E. (1999). *Mol. Cell. Biol.* **19**, 4113–4120.
Rubinson, K. A., Ladner, J. E., Tordova, M. & Gilliland, G. L. (2000). *Acta Cryst.* D**56**, 996–1001.
Sauter, C., Basquin, J. & Suck, D. (2003). *Nucleic Acids Res.* **31**, 4091–4098.
Schedlbauer, A., Gandini, R., Kontaxis, G., Paulmichl, M., Furst, J. & Konrat, R. (2011). *Cell. Physiol. Biochem.* **28**, 1203–1210.
Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C. & Bricogne, G. (2012). *Acta Cryst.* D**68**, 368–380.
Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. (2003). *Protein Sci.* **12**, 1313–1322.
Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 8060–8065.
Suzuki, N., Hiraki, M., Yamada, Y., Matsugaki, N., Igarashi, N., Kato, R., Dikic, I., Drew, D., Iwata, S., Wakatsuki, S. & Kawasaki, M. (2010). *Acta Cryst.* D**66**, 1059–1066.
Tarn, W.-Y. & Steitz, J. A. (1997). *Trends Biochem. Sci.* **22**, 132–137.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wahl, M. C., Will, C. L. & Lührmann, R. (2009). *Cell*, **136**, 701–718.
Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R. & Wahl, M. C. (2010). *EMBO J.* **29**, 4172–4184.
Yau, S. T., Petsev, D. N., Thomas, B. R. & Vekilov, P. G. (2000). *J. Mol. Biol.* **303**, 667–678.
Yeh, J. I., Biemann, H.-P., Privé, G. G., Pandit, J., Koshland, D. E. Jr & Kim, S.-H. (1996). *J. Mol. Biol.* **262**, 186–201.
Yong, J., Wan, L. & Dreyfuss, G. (2004). *Trends Cell Biol.* **14**, 226–232.
Zhang, R., So, B. R., Li, P., Yong, J., Glisovic, T., Wan, L. & Dreyfuss, G. (2011). *Cell*, **146**, 384–395.
Zou, Y., Weis, W. I. & Kobilka, B. K. (2012). *PLoS One*, **7**, e46039.