

Construct optimization for studying protein complexes: obtaining diffraction-quality crystals of the pseudosymmetric PSPC1–NONO heterodimer

Mihwa Lee,^a Daniel M. Passon,^a
Sven Hennig,^b Archa H. Fox^b and
Charles S. Bond^{a*}

^aSchool of Biomedical, Biomolecular and
Chemical Sciences, University of Western
Australia, Crawley, WA 6009, Australia, and

^bCentre for Medical Research, University of
Western Australia, Crawley, WA 6009, Australia

Correspondence e-mail:
charles.bond@uwa.edu.au

Received 4 July 2011

Accepted 27 September 2011

The methodology of protein crystallography provides a number of potential bottlenecks. Here, an approach to successful structure solution of a difficult heterodimeric complex of two human proteins, paraspeckle component 1 (PSPC1) and non-POU domain-containing octamer-binding protein (NONO), that are involved in gene regulation and the structural integrity of nuclear bodies termed paraspeckles is described. With the aid of bioinformatic predictions and systematic screening of a panel of constructs, bottlenecks of protein solubility, crystallization, crystal quality and crystallographic pseudosymmetry were overcome in order to produce crystals that ultimately revealed the structure.

1. Introduction

The eukaryotic cell nucleus provides a challenging pool of targets for structural biology research. The various nuclear proteins and nucleic acids interact in numerous complexes, often transiently, and under regulation by post-transcriptional and post-translational modifications. In order to understand in detail how these interactions direct cell fate, the structural biologist is typically faced with the challenge of rationally designing protein fragments which are both experimentally tractable and biologically informative.

Paraspeckles are relatively newly identified subnuclear bodies that are specific to mammals (Fox *et al.*, 2002; Chen & Carmichael, 2009). Paraspeckles have been implicated in a novel post-transcriptional gene-regulation mechanism: RNA nuclear retention of hyperedited mRNAs (Prasanth *et al.*, 2005; Chen & Carmichael, 2009). This mechanism has been identified as crucial in the response of cells to the environment and in the differentiation of embryonic stem cells.

Paraspeckles are composed of a long non-protein-coding RNA (lncRNA), *NEATI* (also called *Mens1β*), and the core paraspeckle proteins PSPC1 (also known as PSP1), splicing factor, proline- and glutamine-rich protein (SFPO, also known as PSF) and NONO (p54nrb) (Bond & Fox, 2009; Fox & Lamond, 2010). *NEATI* forms the structural framework of paraspeckles by recruiting core proteins for their assembly. *NEATI* transcription is known to be essential to initiate the *de novo* assembly of paraspeckles and to regulate paraspeckle maintenance (Clemson *et al.*, 2009; Mao *et al.*, 2011). Although the three core paraspeckle proteins are not the initiating factors, they are nevertheless essential for paraspeckle assembly and integrity, as knock-down of the abundant SFPO or NONO leads to an absence of paraspeckles (Sasaki & Hirose, 2009).

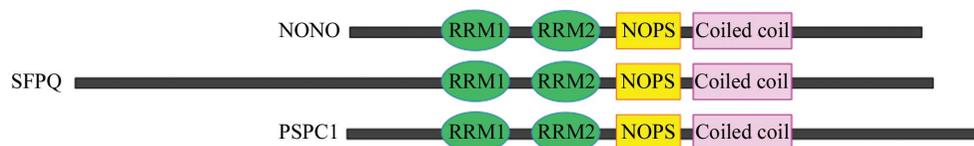


Figure 1

Domain structure of the human DBHS-family proteins (PSPC1, NONO and SFPQ). The N- and C-terminal regions flanking the core conserved region of each protein are predicted to be disordered.

PSPC1, NONO and SFPQ belong to the *Drosophila* behaviour and human splicing (DBHS) protein family. The putative DBHS domain comprises two tandemly arranged RNA-recognition motifs (RRMs) and a NONA/paraspeckle (NOPS) domain followed by a coiled-coil domain (Fig. 1). Proteins of the DBHS family have been implicated in a variety of functions in RNA production and processing, including transcription initiation and termination, splicing and paraspeckle-specific functions (Bond & Fox, 2009; Fox & Lamond, 2010). Sharing a high sequence identity (approximately 70%) within the DBHS domains, the three proteins are known to form dimers with each other *via* DHBS domain-mediated interactions (Myojin *et al.*, 2004; Fox *et al.*, 2005).

Understanding the intermolecular interactions required to sustain paraspeckle structure and function requires detailed structural knowledge of the component proteins and how they interact with each other, with *NEATI* and with nuclear-retained mRNA (Chen & Carmichael, 2009; Chen *et al.*, 2008; Zhang & Carmichael, 2001). As a key step towards this goal, we have applied the crystallographic approach to the core paraspeckle proteins PSPC1 and NONO. Here, we present the systematic process undertaken to optimize the properties of a truncated PSPC1–NONO heterodimer, despite a series of technical challenges, in order to produce crystals suitable for structure solution.

2. Materials and methods

2.1. Plasmid construction

Co-expression of human PSPC1 and NONO in *Escherichia coli* was achieved using the pETDuet-1 expression vector (Novagen). In all cases PSPC1 was cloned into the first multiple cloning site (MCS) and NONO into the second MCS. The cDNAs described by Fox *et al.* (2005) were used as templates for polymerase chain reaction (PCR) amplification to produce inserts containing *Bam*HI and *Eco*RI sites at the 5' and 3' ends of PSPC1 fragments and *Nde*I and *Xho*I sites for NONO, allowing the construction of the appropriate pETDuet-1 plasmids shown in Fig. 2. The sequences of the primers used for PCR reactions are listed in Supplementary Table 1[†]. All constructs were verified by DNA sequencing.

Initial attempts at expressing the heterodimer of full-length PSPC1–NONO (residues 1–523 and 1–471, respectively)

[†] Supplementary material has been deposited in the IUCr electronic archive (Reference: GX5190). Services for accessing this material are described at the back of the journal.

resulted in totally insoluble protein. The first attempt to improve expression (PN1 in Fig. 2) has been published previously (Fox *et al.*, 2005); construct PN1 encodes PSPC1 (residues 61–328, numbering relative to UniProt entry PSPC1_HUMAN) with a hexahistidine tag, FLAG tag and thrombin cleavage site at the

N-terminus and NONO (residues 75–313; NONO_HUMAN) with no tag. In the subsequent constructs PN2–PN6, PSPC1 was preceded by only a hexahistidine tag and a tobacco etch virus (TEV) protease cleavage site (sequence ENLYFQG) to allow tag removal at a later stage of protein purification. The resulting protein product contains a single glycine not derived from the gene sequence at the N-terminus.

2.2. Expression and purification

Competent *E. coli* Rosetta2 (DE3) cells (Novagen) were transformed with the appropriate plasmid. Transformants were selected from Luria–Bertani (LB) agar plates supplemented with the antibiotics ampicillin (100 µg ml⁻¹) and chloramphenicol (50 µg ml⁻¹). Expression trials were carried out with 0.5 l LB medium in 2 l conical flasks shaken at 310 K. When the optical density (600 nm) reached approximately 0.6, the cells were induced with IPTG (0.5 mM) and incubated with shaking for 16 h at 298 K. The cells were harvested by centrifugation (5000g), resuspended in lysis buffer [20 mM Tris pH 7.5, 150 mM NaCl, 5 mM imidazole with or without 10% (v/v) glycerol] and lysed with an Emulsiflex C5 high-pressure homogeniser (Avestin); the lysate was then clarified by centrifugation (18 000g). The supernatant was applied onto a 5 ml His-Trap column (GE Healthcare) and the PSPC1–NONO heterodimer was eluted with an imidazole gradient (50–500 mM over ten column volumes). Peak fractions were pooled and the concentration was estimated by absorption at 280 nm to facilitate tag cleavage by application of histidine-tagged recombinant TEV protease [1:50 (w:w)] at ambient temperature overnight. Subsequent filtration over a nickel-affinity column removed the tagged protease and residual uncleaved PSPC1–NONO heterodimer. The flowthrough was concentrated, loaded onto a HiLoad 16/60 Superdex 200 preparative-grade column (GE Healthcare) and developed with 20 mM Tris pH 7.5, 150 mM NaCl with or without 10% (v/v) glycerol. The resulting protein samples were concentrated to 4–10 mg ml⁻¹ as estimated from the absorption at 280 nm.

2.3. Crystallization

A standard crystallization strategy was employed: crystallization space was explored using a variety of commercially available crystallization screens set up in 96-well sitting-drop vapour-diffusion format using an Art Robbins Phoenix robot. Typically, 150 nl PSPC1–NONO heterodimer (4–10 mg ml⁻¹ in 20 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol) was mixed

with 150 nl reservoir solution and the drops were equilibrated over 80 μ l reservoir solution in CrystalQuick 96-well sitting-drop plates (Greiner). Optimization of the resulting hit conditions was performed using hanging-drop vapour diffusion by mixing 1–2 μ l protein solution and 1–2 μ l reservoir solution in a variety of ratios and equilibrating against 1 ml reservoir solution in 24-well plates (Linbro, ICN Biochemicals). All experiments were performed at 293 ± 0.5 K.

Small trigonal crystals of PN2 (4.2 mg ml⁻¹ in 20 mM Tris pH 7.5, 150 mM NaCl) grew within one week from NeXtal

PEG Suite (Qiagen) condition No. 23 [0.1 M Tris pH 8.5, 25% (w/v) PEG 1000]. Crystal size was optimized to $100 \times 30 \mu$ m using hanging drops composed of 2 μ l protein (10 mg ml⁻¹ in 50 mM Tris pH 7.5, 150 mM NaCl) and 2 μ l reservoir solution [0.1 M Tris pH 7.5, 26% (w/v) PEG 1000]. Cryoprotection was afforded by brief soaking in reservoir supplemented with 15% (v/v) glycerol immediately prior to freezing. The presence of both PSCP1 and NONO proteins in the crystals in equal quantities was confirmed by washing crystals in buffer and analysis by silver-stained SDS-PAGE.

The crystallization and preliminary analysis of crystals of PN5 has been described elsewhere (Passon *et al.*, 2011). Briefly, crystals of PN5 [9.5 mg ml⁻¹ in 20 mM Tris pH 7.5, 150 mM NaCl, 10% (v/v) glycerol] were obtained from Index (Hampton Research) condition No. 71 [0.1 M Bis-Tris pH 6.5, 0.2 M NaCl, 25% (w/v) PEG 3350] at 293 K.

2.4. Data collection and preliminary X-ray analysis

Data-collection and processing statistics are presented in Table 1. Synchrotron data to 4.0 Å resolution were recorded from a single trigonal crystal of PN2 cooled to 100 K on beamline MX1 of the Australian Synchrotron. The data were recorded on an ADSC 351r CCD detector. 100 successive frames with 1° oscillation were collected. The diffraction data were processed using XDS (Kabsch, 2010), POINTLESS (Evans, 2006) and SCALA (Evans, 2006).

Complete details of the data collection and X-ray analysis of native and selenomethionine-derivative PN5 crystals have been described elsewhere (Passon *et al.*, 2011). The structure (PDB entry 3sde) was determined by multi-wavelength anomalous dispersion phasing.

The structure of the PN2 crystals was solved by molecular replacement with Phaser (McCoy *et al.*, 2007; Winn *et al.*, 2011) using only chain A (representing PSCP1) of PDB entry 3sde as the search model.

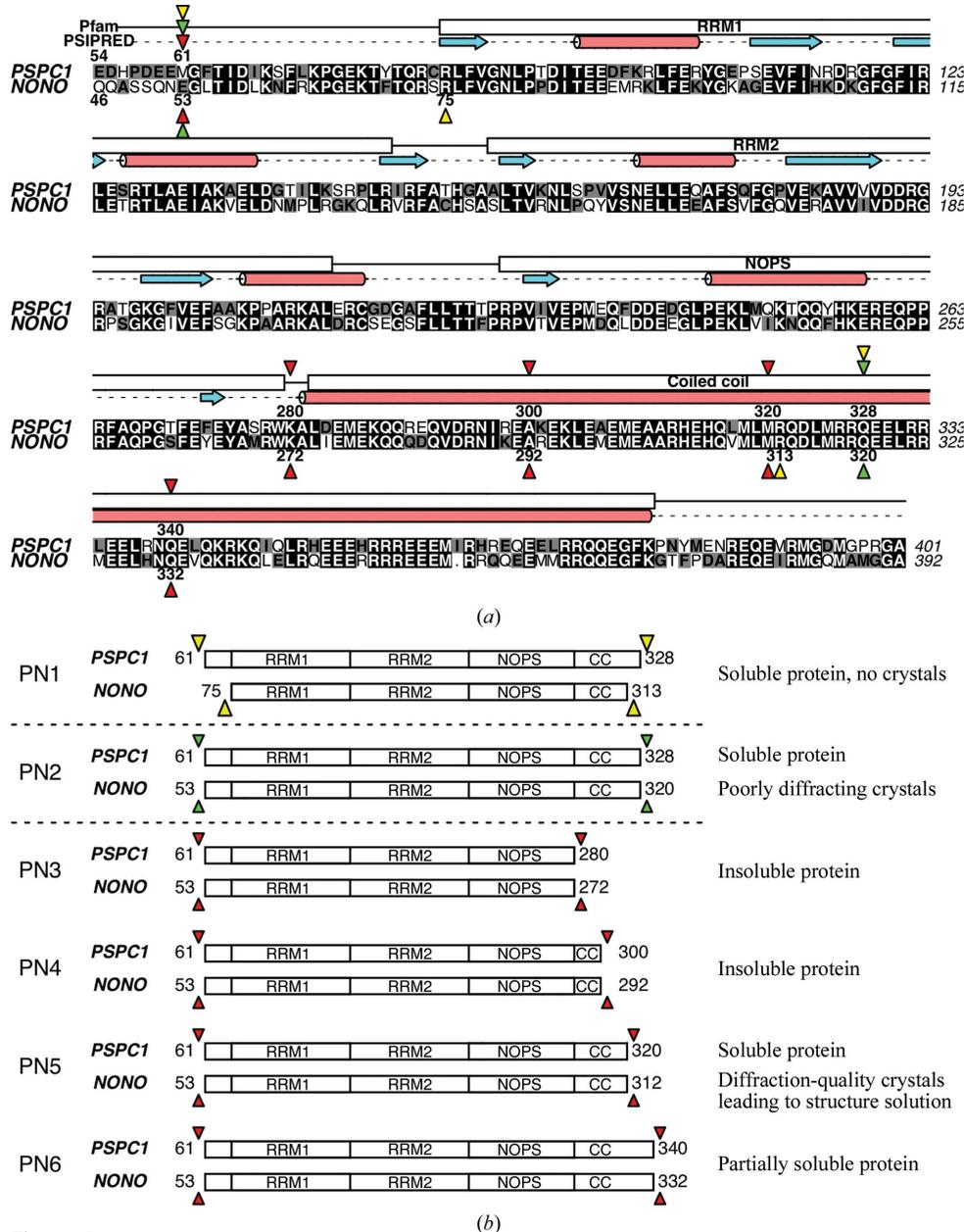


Figure 2 Construct design for the PSCP1–NONO complex. (a) Sequence alignment of human PSCP1–NONO within the predicted structured regions. PFAM domain definitions (white boxes) and secondary-structure predictions are overlaid on top of the sequences (arrows and cylinders represent β -sheets and α -helices, respectively). Triangles indicate the construct boundaries tested, with downward-pointing triangles for PSCP1 and upward-pointing triangles for NONO; yellow, construct PN1; green, PN2; red, PN3–PN6 as defined in (b). (b) Schematic diagram of each construct PN1–PN6 made during the course of optimization together with its properties.

3. Results and discussion

3.1. Domain analysis of paraspeckle proteins

Determining the three-dimensional structure of proteins by X-ray crystallography involves several key steps: target selection, production of pure protein, crystallization, diffraction data collection and determination of atomic positions. When studying multi-protein complexes, the production of diffraction-quality crystals is commonly the most challenging step of all these steps, even when pure protein samples are available. After the initial characterization of PSPC1 as the main core protein component of paraspeckle (Fox *et al.*, 2002) and a follow-up study on the heterodimerization between PSPC1 and NONO (Fox *et al.*, 2005), the PSPC1–NONO heterodimer was chosen as the principal target for structure determination. To increase the chance of obtaining substantial amounts of pure soluble protein, detailed analysis of the predicted domain structure, secondary structure and globularity of PSPC1 and NONO was performed.

PSPC1 (523 amino acids) and NONO (471 amino acids) share over 70% sequence identity within a 300 amino-acid region that is conserved amongst all DBHS proteins. This region is composed of two RRM, a NOPS domain and a coiled-coil domain (Figs. 1 and 2). The RRM is one of the most abundant protein motifs found in eukaryotes and is associated with many post-transcriptional events, including pre-mRNA processing, splicing, RNA editing and mRNA stability (Maris *et al.*, 2005). The tandem RRM are mostly likely to be involved in interactions with retained mRNAs in the nuclear RNA-retention mechanism as well as in interactions with structural lncRNA (*NEATI* in paraspeckle; Clemson *et al.*, 2009). The NOPS domain is defined in the PFAM database (Finn *et al.*, 2010) by sequence conservation and is definitive for the DBHS protein family. Together with the following coiled-coil domain, the NOPS domain has been reported as a protein–protein interaction domain in other DBHS-family proteins (Kiesler *et al.*, 2003). In the case of PSPC1 the second RRM, the NOPS domain and the coiled-coil domain are required to be functionally targeted to paraspeckles (Fox *et al.*, 2005). The coiled-coil domains in PSPC1 and NONO are predicted to stretch to around 100 amino acids in length by *COILS* (Lupas *et al.*, 1991) and *PairCoil* (Berger *et al.*, 1995). The flanking regions (residues 1–60 and 381–523 in PSPC1; residues 1–52 and 372–471 in NONO) outside the conserved core domains are predominantly composed of low-complexity sequences rich in proline and glycine residues and are predicted to be disordered by *IUPred* (Dosztányi *et al.*, 2005) and *PSIPRED* (Bryson *et al.*, 2005). Sequence analysis was performed with *ALINE* (Bond & Schüttelkopf, 2009).

3.2. Construct design with different C-termini

Owing to the low solubility and poor stability of the full-length heterodimer of PSPC1–NONO, combined with the presence of unstructured regions at the termini of the proteins (Figs. 1 and 2a), it was reasoned that truncation of both the N- and C-termini of both proteins to the well defined conserved region (RRM1, RRM2, NOPS and coiled coil) would increase

Table 1

Data scaling and merging statistics.

Values in parentheses are for the last shell.

	P6	P622
Resolution (Å)	68.04–4.00 (4.22–4.00)	68.04–4.00 (4.22–4.00)
No. of observations	28775 (4227)	28770 (4227)
No. of unique reflections	9174 (1340)	2737 (379)
Completeness (%)	98.3 (98.1)	98.9 (99.1)
Multiplicity	3.1 (3.2)	10.5 (11.2)
R_{merge} (%)	0.110 (1.461)	0.125 (1.656)
R_{meas} (%)	0.133 (1.762)	0.132 (1.734)
$R_{\text{p.i.m.}}$ (%)	0.074 (0.976)	0.041 (0.508)
Average $I/\sigma(I)$	6.8 (0.9)	12.9 (1.6)

the chance of crystallization while maintaining the maximum number of domains.

The choice of the N-terminal boundary, clearly delineated by the start of RRM1, was more obvious than that for the C-terminus: the complexity of the amino-acid sequence degrades throughout this coiled-coil region and there is variation in predicted coiled-coil length between members of the DBHS family, making it difficult to define a convincing C-terminal end.

Owing to the length (up to 100 amino acids) of the predicted coiled-coil domain, it was reasoned that a parallel heterodimeric coiled coil would form. Based on this assumption, several constructs differing in length at the C-terminus were designed.

PN1, which was designed to verify the heterodimeric interaction between PSPC1 and NONO, has significantly different termini for the two proteins to highlight the size difference after purification of the heterodimer (Fox *et al.*, 2005). It includes part of the conserved region, RRM1, RRM2, NOPS and part of the coiled coil (residues 61–328 in PSPC1 and residues 75–313 in NONO; Fig. 2). The production of soluble heterodimer from this construct was a major step forward; however, the protein could not be crystallized despite exhaustive attempts.

PN2 was designed to overcome crystallization problems based on the hypothesis that a sample in which the termini of the two proteins are the same would be more likely to crystallize. The resultant construct encodes the same partial conserved region of PSPC1 that was used in PN1 (residues 61–328) together with the equivalent domain in NONO (residues 53–320). This construct also produced soluble heterodimer and ultimately weakly diffracting crystals, the properties of which are discussed below.

In the third stage of optimization (constructs PN3–PN6) we explored the choice of C-terminus in order to screen the effect of coiled-coil length on solubility and crystallizability. It was expected that a construct with the full coiled-coil domain might result in elongated molecules which may hinder compact packing into any crystal lattice.

It has previously been reported that truncated PSPC1 constructs (residues 1–235, 81–286 or 231–358) transiently expressed in HeLa cells all failed to co-immunoprecipitate NONO (Fox *et al.*, 2005). The shortest construct in the report that successfully associated with NONO contains residues

125–358, encoding part of RRM1, RRM2, NOPS and the coiled coil, suggesting that although the coiled-coil domain is mostly likely to be involved in heterodimerization, the NOPS and coiled-coil domains are not sufficient for association with NONO. In contrast, interaction-domain mapping of the highly similar protein SFPQ with NONO (Peng *et al.*, 2002) indicated that deletion mutants of NONO lacking both RRMs (*e.g.* NONO Δ 17–220) were capable of binding SFPQ.

Based on these observations, we produced four different constructs (PN3–PN6) with different C-termini; the shortest construct PN3 included RRM1, RRM2 and NOPS but no coiled coil, while constructs PN4–PN6 included coiled coil in 20-amino-acid increments. To our surprise, the constructs with different C-termini resulted in wildly varying solubility. Constructs PN3 (no coiled coil) and PN4 (20 amino acids of coiled coil) were exclusively insoluble under the expression conditions tested. PN5 (40 amino acids of coiled coil) was soluble and formed a stable heterodimer, perhaps not surprisingly as it differed by being only eight amino acids shorter in both proteins relative to the soluble PN2 protein. Unexpectedly, PN6 (60 amino acids of coiled coil) exhibited very limited solubility, with only around 20% of the recombinant proteins expressed in a soluble form even at the lowest temperature tested (289 K). Even then, PN6 protein was poorly behaved in solution, showing aggregation at concentrations higher than 1.2 mg ml⁻¹. Ultimately, these results turned out to be entirely consistent with the solved structure of the PSPC1–NONO heterodimer determined from PN5 (discussed in §3.4).

3.3. Quest for diffraction-quality crystals

PN1 did not yield crystals despite extensive trials, in spite of the fact that it resulted in pure soluble PSPC1–NONO heterodimer. We reasoned that having different termini for two highly similar proteins might result in poor folding of the excess residues at both termini of PSPC1, thus preventing crystal formation.

This hypothesis was confirmed by the successful crystallization of PN2, in which both proteins have equivalent termini. These crystals, with a prolate hexagonal prism habit (Fig. 3*a*), typically diffracted to only 8 Å resolution on beamline MX1 of the Australian Synchrotron. Extensive attempts at optimization (re-screening, additive screening, dehydration, crystal annealing and room-temperature data collection) did very little to improve diffraction quality, although one single crystal was obtained that diffracted to 4.0 Å resolution with a trigonal unit cell of $a = b = 78.58$, $c = 159.80$ Å. The results of scaling (Table 1) indicated that the data collected from this crystal were consistent with point groups $P3$, $P6$ or $P622$. The presence of appropriate systematic absences along $(0, 0, l)$ clearly indicated the presence of a 6_1 or 6_5 screw axis. The equivalence of R_{meas} values for $P6$ and $P622$ (13.3 and 13.2%, respectively) would typically lead one to choose $P622$ as the space group. However, solvent-content analysis indicated that $P622$ was incompatible with the presence of a heterodimer in the asymmetric unit (predicted

solvent content 3%) and rather suggested the presence of a monomer (51% solvent).

We reasoned that we may have crystallized only one of the proteins (PSPC1 or NONO), as this would be compatible with the solvent content. To test this, crystals were washed and subjected to SDS–PAGE. Two distinct protein bands with the same intensity indicated that both PSPC1 and NONO are present in the crystals (Fig. 3*b*).

Based on the observation that we had made the two protein components as similar in length as possible, together with their high sequence similarity, led us to the conclusion that our crystals were pseudosymmetric, incorporating PSPC1 and NONO in a stochastic mixture of $A:B$ and $B:A$ orientations (Fig. 3*c*). This phenomenon is likely to be a cause of the pathologically poor-quality diffraction properties of this crystal form (see further discussion in §3.4) and this led us to design a panel of further constructs (PN3–PN6) to seek improved diffraction.

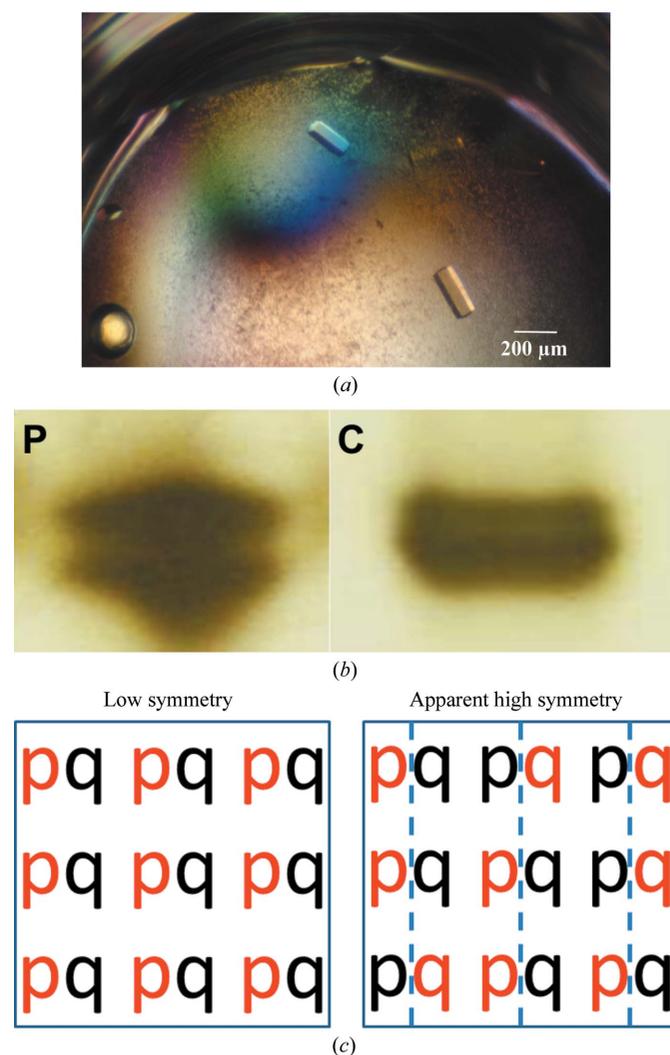


Figure 3 Pseudosymmetric crystals in space group $P6_22$. (*a*) Hexagonal crystals of PN2 viewed under crossed polarisers. (*b*) A silver-stained SDS–PAGE gel obtained from purified PN2 (lane P) and from washed and dissolved crystals of PN2 (lane C). (*c*) Accommodating a stochastic mixture of pseudosymmetric heterodimers in an apparently high-symmetry lattice

Among these four constructs (PN3–PN6), PN5, which is only eight amino acids shorter than PN2, shows the highest solubility and yielded diffraction-quality monoclinic crystals that led to eventual structure determination at 1.9 Å resolution. A full description of the structure, deposited as PDB entry 3sde, and its biological implications will be published elsewhere.

Molecular replacement of the hexagonal data (PN2) using PSPC1 extracted from the structure of the heterodimer was successful. The translation function clearly established $P6_522$ as the correct space group. The validity of this result was confirmed by the observation that the application of one of the twofold crystallographic transformations to the PSPC1 monomer generates a dimer with a very similar arrangement to the pseudosymmetric PSPC1–NONO heterodimer. We consider this as additional evidence that the hexagonal crystals are pathologically pseudosymmetric.

3.4. Assumptions versus reality

The most unexpected among the discrepancies between our predictions and the solved structure was the configuration of the 100-residue coiled-coil domain. The large predicted length of this coiled coil, its implication in DBHS protein dimerization and comparison with proteins of known structure led us to expect a parallel coiled coil. Based on this hypothesis, one might expect that any construct with sufficient coiled-coil content to promote dimerization should be soluble and that increased extension of the coiled coil should not impair this solubility, although it might impair crystallizability. Rather than observing the expected comparative solubility (0-coil \ll 20-coil = 40-coil = 60-coil), we observed 0-coil = 20-coil \ll 40-coil > 60-coil. The explanation for this behaviour was ultimately revealed by solution of the crystal structure obtained from PN5. This region of the protein formed an antiparallel coiled coil of length 42 amino acids. The N-terminal portion of the coiled coil intimately interacts with the NOPS and RRM domains; 42 amino acids are essential to form the core coiled coil and any residues beyond 42 amino acids would be predicted to protrude into space with no interaction partner. All of these observations post-rationalize our solubility results. They also, however, pose questions regarding the biological role of the 60 amino acids of predicted coiled coil that are not present in our structure, and studies are now under way to investigate the functional context of this region in paraspeckle biology.

4. Conclusions

In crystallographic studies, when the full-length target protein of interest cannot be expressed in substantial amounts or the full-length protein seems unlikely to be crystallized owing to embedded intrinsically disordered sections, the screening of truncated proteins offers a useful route to success. When the target protein is composed of several individual domains, it is the relative arrangement of these domains in three dimensions that specifies its function and therefore minimal truncation is

essential in order to best represent the structure of the functionally active full-length protein. Sometimes, however, drastic truncation may be unavoidable to produce diffraction-quality crystals. Domain prediction, secondary-structure prediction, globularity prediction and molecular cell biology or biophysical characterization can inform the choice of appropriate termini of truncated proteins. Nevertheless, even when the structures of individual domains have been solved and well characterized, choosing the correct boundaries for production of soluble protein and diffraction-quality crystals is not trivial. The journey taken to solve the structure of the PSPC1–NONO heterodimer is an illustrative example of hypothesis-based reasoning in which both valid and invalid assumptions were tested to ultimately produce the desired structure. Even with well developed analysis and prediction tools, in the end it is only when the structure of interest is solved that we can fully understand and explain the unexpected. Optimization of the C-terminus was based on the assumption of a parallel intermolecular coiled coil owing to its predicted length (~100 amino acids). The strategy worked, but the assumption was invalid as the structure ultimately indicated a 42-amino-acid antiparallel intermolecular coiled coil: the correct length at the C-terminus was thus essential to promote crystallization.

This work was supported by NHMRC Project Grant 588130 to CSB and AHF. ML is supported by an NHMRC Postdoctoral Training Fellowship (513935). Diffraction data were collected on beamline MX1 at the Australian Synchrotron, Victoria, Australia. We thank Tom Caradoc-Davies and Rachel Williamson for their assistance.

References

- Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M. & Kim, P. S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 8259–8263.
- Bond, C. S. & Fox, A. H. (2009). *J. Cell Biol.* **186**, 637–644.
- Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* **D65**, 510–512.
- Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S. & Jones, D. T. (2005). *Nucleic Acids Res.* **33**, W36–W38.
- Chen, L. L. & Carmichael, G. G. (2009). *Mol. Cell*, **35**, 467–478.
- Chen, L. L., DeCerbo, J. N. & Carmichael, G. G. (2008). *EMBO J.* **27**, 1694–1705.
- Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A. & Lawrence, J. B. (2009). *Mol. Cell*, **33**, 717–726.
- Dosztányi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). *Bioinformatics*, **21**, 3433–3434.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). *Nucleic Acids Res.* **38**, D211–D222.
- Fox, A. H., Bond, C. S. & Lamond, A. I. (2005). *Mol. Biol. Cell*, **16**, 5304–5315.
- Fox, A. H., Lam, Y. W., Leung, A. K., Lyon, C. E., Andersen, J., Mann, M. & Lamond, A. I. (2002). *Curr. Biol.* **12**, 13–25.
- Fox, A. H. & Lamond, A. I. (2010). *Cold Spring Harb. Perspect. Biol.* **2**, a000687.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kiesler, E., Miralles, F., Ostlund Farrants, A. K. & Visa, N. (2003). *J. Cell Sci.* **116**, 3949–3956.
- Lupas, A., Van Dyke, M. & Stock, J. (1991). *Science*, **252**, 1162–1164.

- Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. (2011). *Nature Cell Biol.* **13**, 95–101.
- Maris, C., Dominguez, C. & Allain, F. H. (2005). *FEBS J.* **272**, 2118–2131.
- 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.
- Myojin, R., Kuwahara, S., Yasaki, T., Matsunaga, T., Sakurai, T., Kimura, M., Uesugi, S. & Kurihara, Y. (2004). *Biol. Reprod.* **71**, 926–932.
- Passon, D. M., Lee, M., Fox, A. H. & Bond, C. S. (2011). *Acta Cryst.* **F67**, 1231–1234.
- Peng, R., Dye, B. T., Pérez, I., Barnard, D. C., Thompson, A. B. & Patton, J. G. (2002). *RNA*, **8**, 1334–1347.
- Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., Zhang, M. Q. & Spector, D. L. (2005). *Cell*, **123**, 249–263.
- Sasaki, Y. T. & Hirose, T. (2009). *Genome Biol.* **10**, 227.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Zhang, Z. & Carmichael, G. G. (2001). *Cell*, **106**, 465–475.