

**Andrey Feklistov\*** and **Seth A. Darst**Laboratory of Molecular Biophysics,  
The Rockefeller University, 1230 York Avenue,  
New York, NY 10065, USACorrespondence e-mail:  
afeklistov@rockefeller.edu

Received 11 June 2013

Accepted 23 July 2013

**PDB Reference:**  $\sigma_{2-3}^A$ -TGTACAATGGG, 4ki2

# Crystallographic analysis of an RNA polymerase $\sigma$ -subunit fragment complexed with $-10$ promoter element ssDNA: quadruplex formation as a possible tool for engineering crystal contacts in protein–ssDNA complexes

Structural studies of  $-10$  promoter element recognition by domain 2 of the RNA polymerase  $\sigma$  subunit [Feklistov & Darst (2011), *Cell*, **147**, 1257–1269] reveal an unusual crystal-packing arrangement dominated by G-quartets. The 3'-terminal GGG motif of the oligonucleotide used in crystallization participates in G-quadruplex formation with GGG motifs from symmetry-related complexes. Stacking between neighboring G-quadruplexes results in the formation of pseudo-continuous four-stranded columns running throughout the length of the crystal (G-columns). Here, a new crystal form is presented with a different arrangement of G-columns and it is proposed that the fortuitous finding of G-quartet packing could be useful in engineering crystal contacts in protein–ssDNA complexes.

## 1. Introduction

Bacterial promoter opening is initiated when the promoter-specificity  $\sigma$  subunit of the RNA polymerase (RNAP) recognizes the non-template DNA strand of the  $-10$  element (consensus sequence TATAAT, located 10 bp prior to the transcription start site). To gain insight into this pivotal event in bacterial transcription, we determined crystal structures of the complex between the  $\sigma_2^A$  domain of *Thermus aquaticus* RNAP and an oligonucleotide comprising the  $-10$  element sequence (Feklistov & Darst, 2011; Liu *et al.*, 2011). The structure revealed the oligonucleotide bound to the positively charged surface of  $\sigma_2^A$  with two bases flipped out of the single-stranded DNA base stack and into protein pockets.

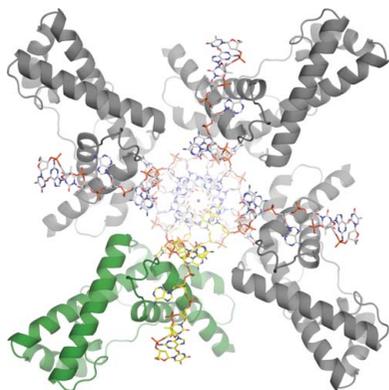
The unexpected finding, which was only briefly mentioned in the original publication, came from analysis of the packing arrangement in  $-10$  DNA– $\sigma_2^A$  crystals. The 3' part of the oligonucleotide used in cocrystallization trials, which was expected to interact with  $\sigma_2^A$ , did not form contacts with the protein but instead engaged in the formation of a G-quadruplex with symmetry-related protein–DNA complexes. Stacked on each other, the G-quadruplexes formed massive pseudo-continuous four-stranded columns running throughout the entire length of the crystal. Packed layers of the protein complexed with the single-stranded part of the oligonucleotide therefore appeared to be organized on a massive DNA scaffold in a manner reminiscent of the first attempts to engineer crystals, where proteins would be arranged on a nucleic acid scaffold formed by virtue of the self-assembling properties of DNA (Malo *et al.*, 2005; Seeman, 1982).

In the present manuscript, we analyze the observed crystal-packing arrangement in detail, present a new crystal form with a different symmetry and orientation of the G-columns relative to each other and discuss the structural aspects of G-quadruplexes with respect to their possible use for bottom-up crystal engineering.

## 2. Experimental procedures

### 2.1. Protein and nucleic acid preparation

Protein and nucleic acid preparation was performed as described in Feklistov & Darst (2011).

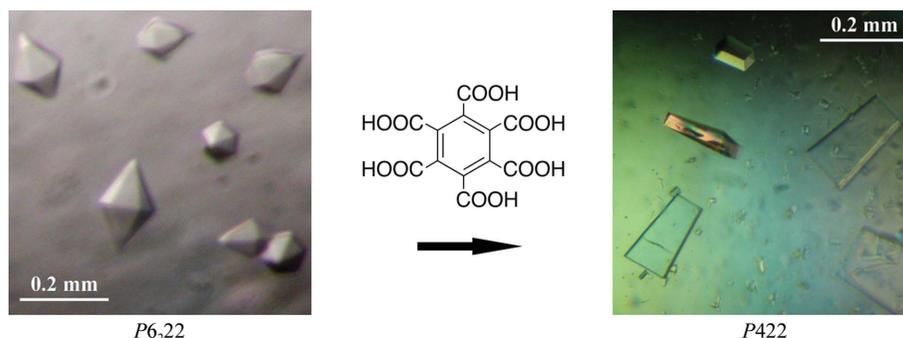


## 2.2. Crystallization, data collection and structure determination

The ssDNA- $\sigma_{2-3}^A$  complex was prepared on ice (molar ratio 2:1) at a final protein concentration of 10 mg ml<sup>-1</sup>; two different oligo sequences were used: TGTACAATGGG and TGTATAATGGG. The original hit was found in The JCSG Core II Suite screen (Qiagen) [condition No. 8; 0.1 M Tris pH 8.5, 5% (w/v) PEG 8000, 20% (w/v) PEG 300, 10% (v/v) glycerol]. Crystals were grown at 295 K using hanging-drop vapor diffusion by mixing equal volumes of the complex solution and the reservoir solution. Hexagonal bipyramidal-shaped crystals grew to maximum dimensions of  $\sim 0.1 \times 0.1 \times 0.2$  mm within 2–3 d. These initial crystals belonged to space group  $P6_22$ , contained two ssDNA- $\sigma_{2-3}^A$  complexes per asymmetric unit and diffracted to  $\sim 3.5$  Å resolution. A systematic search of additive screens led to the discovery that the addition of 0.1–0.2% mellitic acid (Silver Bullets, Hampton Research) yielded plate-like crystals that grew to maximum dimensions of  $0.05 \times 0.2 \times 0.2$  mm within 2–5 d. The new crystal form belonged to space group  $P422$ , contained one ssDNA- $\sigma_{2-3}^A$  complex per asymmetric unit and diffracted to beyond 2 Å resolution (Fig. 1). For data collection, crystals were flash-cooled in liquid nitrogen directly from the mother liquor.

Diffraction data were collected on NE-CAT beamline 24-ID-E at the Advanced Photon Source (Argonne National Laboratory) and on beamline X29 at the National Synchrotron Light Source (Brookhaven National Laboratory). The diffraction images were processed using the *HKL-2000* program suite (Otwinowski & Minor, 1997). For both crystal forms the structure was solved by molecular replacement using the structure of *T. aquaticus*  $\sigma_2^A$  (PDB entry 1ku2; Campbell *et al.*, 2002) as a search model with *Phaser* (McCoy *et al.*, 2007). The resulting electron-density maps were improved by density modification with *CNS* (Adams *et al.*, 1997). Iterative rounds of model building and refinement were carried out using *Coot* (Emsley & Cowtan, 2004) and *PHENIX* (Adams *et al.*, 2010), respectively. In the case of the  $P6_22$  crystal form, after rigid-body refinement, strict noncrystallographic symmetry (NCS) was applied for the first round of refinement. NCS restraints were not used in further rounds of refinement. Domain  $\sigma_3^A$  was disordered in the  $P422$  crystal form, but one of the two copies of  $\sigma_3^A$  was found in the  $P6_22$  crystal form. Molecular replacement failed to find the correct orientation for  $\sigma_3^A$ ; therefore, it was fitted manually in *Coot*.

Data collected from the  $P6_22$  crystals yielded a model that was refined at 3.5 Å resolution to  $R$  and  $R_{\text{free}}$  values of 0.219 and 0.268, respectively (oligo sequence TGTACAATGGG; Table 1). The  $P422$  crystal form gave the most detailed model, which was refined at 2.1 Å resolution to  $R$  and  $R_{\text{free}}$  values of 0.194 and 0.237, respectively (for oligo TGTACAATGGG), and a 2.7 Å resolution model with  $R$  and  $R_{\text{free}}$  values of 0.196 and 0.238, respectively (for oligo TGTATAATGGG) (for details, see Feklistov & Darst, 2011).



**Figure 1**

Crystals of the  $\sigma_{2-3}^A$ -DNA complex: the hexagonal crystal form (space group  $P6_22$ ) and the tetragonal crystal form ( $P422$ ) grown in the presence of mellitic acid (formula shown).

**Table 1**

Data-collection and refinement (molecular replacement) statistics.

Values in parentheses are for the highest resolution shell.

Crystal form	$P6_22$
Oligonucleotide	TGTACAATGGG
PDB code	4ki2
Data collection	
Unit-cell parameters (Å, °)	$a = b = 118.11$ , $c = 200.85$ , $\alpha = \beta = 90.00$ , $\gamma = 120.00$
Resolution (Å)	30.00–3.30 (3.42–3.30)
$R_{\text{merge}}$	0.1768 (1.55)
$R_{\text{meas}}^\dagger$	0.1836
$\langle I/\sigma(I) \rangle$	12.66 (1.92)
Completeness (%)	99.98 (99.76)
Multiplicity	14.4 (13.5)
Refinement	
Resolution (Å)	29.53–3.30
No. of reflections	13056
$R_{\text{work}}/R_{\text{free}}$	0.212/0.262
No. of atoms	
Total	3798
Protein	3224
DNA	569
K <sup>+</sup> ions	5
Water	0
B factors (Å <sup>2</sup> )	
Protein	110.26
DNA	86.34
K <sup>+</sup> ions	89.60
Water	—
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.38
Ramachandran plot	
Favored (%)	90
Outliers (%)	2.4

<sup>†</sup> Calculated according to Diederichs & Karplus (1997).

## 3. Results and discussion

### 3.1. Overall structure and analysis of crystal packing

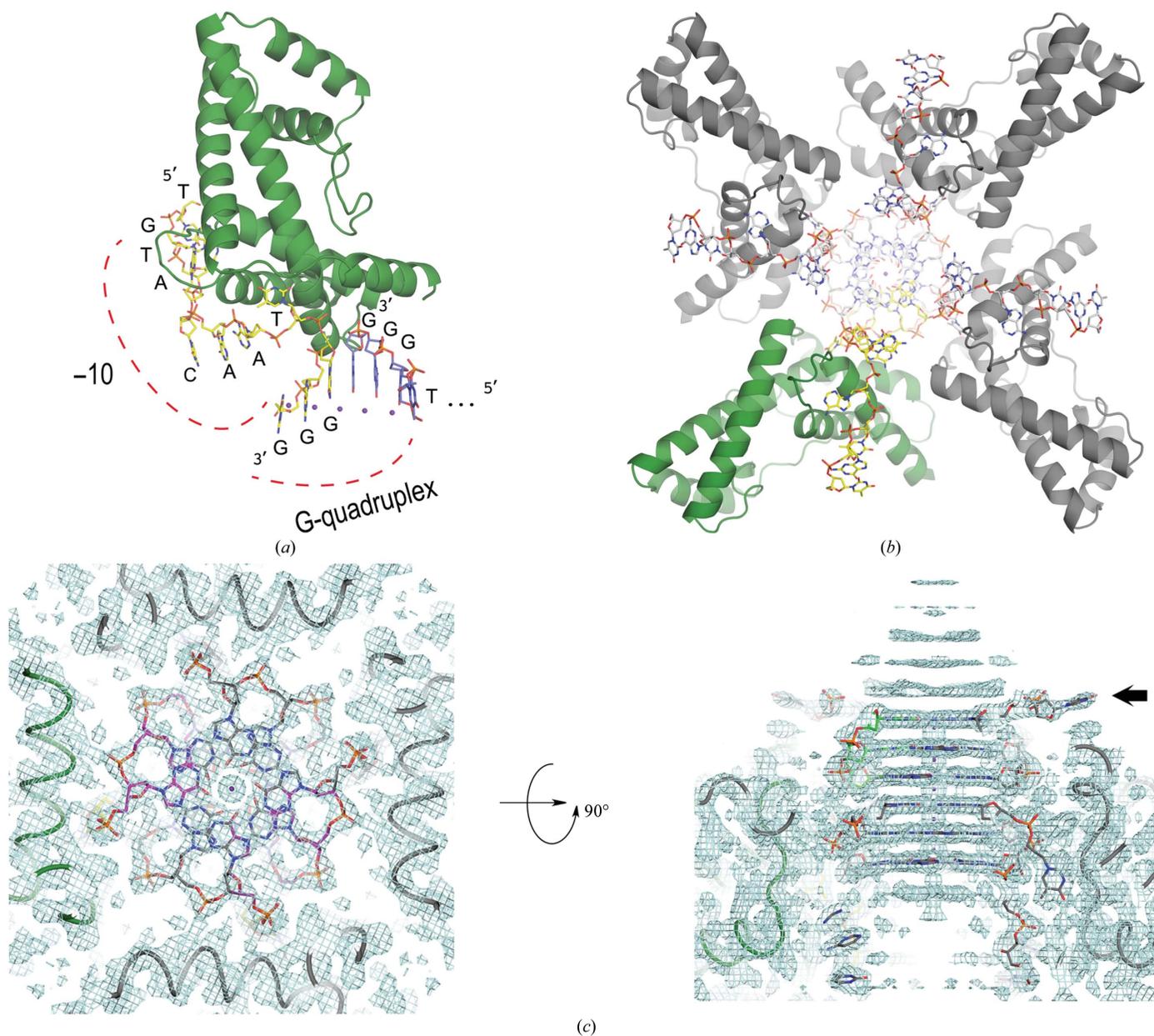
We solved structures of the complex of  $\sigma_{2-3}^A$  with –10 element ssDNA in two different crystal forms (in space groups  $P6_22$  and  $P422$ ). Parts of the structures comprising the  $\sigma_2^A$  domain bound to the –10 element ssDNA were essentially identical in both crystal forms (r.m.s.d. of 0.264 Å over all atoms). The 2.1 Å resolution structure of the  $\sigma_{2-3}^A$ -ssDNA complex from the  $P422$  crystal form is shown in Fig. 2(a). The –10 element is bound across the conserved positively charged surface of  $\sigma_2^A$  previously implicated in –10 element recognition from genetic and biochemical data (Hook-Barnard & Hinton, 2007). The two most highly conserved bases of the –10 element (A<sub>-11</sub> and T<sub>-7</sub>; Shultzaberger *et al.*, 2007) are flipped out of the single-stranded DNA base stack and buried deep in protein pockets. The oligonucleotide used in crystallization (5'-TGT**ACA**ATGGG-3', with the –10 element in bold) contained a discriminator element (the

GGG motif downstream of the  $-10$  element) that is expected to interact sequence-specifically with region 1.2 of  $\sigma_2^A$  (Feklistov *et al.*, 2006; Haugen *et al.*, 2006). Unexpectedly, in the solved structures the guanine bases of the discriminator element did not interact with  $\sigma$ , but peeled away from the protein and participated in crystal contacts with GGG motifs from symmetry-related complexes, forming a G-quadruplex structure that is unlikely to be relevant to  $\sigma$ -factor function but played a critical role in crystal packing (Figs. 2*a* and 2*b*).

On the central axis of the G-quadruplex, electron density for  $K^+$  ions is clearly observed (Fig. 2*c*): the cations are sandwiched between the G-quartet planes and are coordinated by the O6 atoms of the guanine bases. In addition to the DNA oligonucleotide specifically bound to  $\sigma_2^A$ , the unit cell of both crystal forms contained free oligonucleotides (not bound to the protein) that also participate in

extended G-quadruplex formation. Intermolecular G-quadruplexes from  $\sigma_2^A$ -bound DNA molecules pack end-to-end with G-quadruplexes formed by free DNA strands, together forming long continuous pseudo-infinite G-quadruplex columns that extend throughout the crystal, running along the  $c$  axis of the unit cell (Figs. 2*a* and 3*a*). The presence of the free DNA molecules required for the packing explains why crystals in drops with a twofold to fourfold molar excess of DNA over protein nucleated and grew significantly faster.

The crystal lattice of the  $P422$  crystals is therefore composed of layers of protein–DNA complexes bridged by four-stranded columns formed by free DNA oligos (Fig. 3). The connecting regions exhibited poor electron density and attempts to model them gave rise to an increased  $R_{\text{free}}$ , but the planes of the bases were clearly visible



**Figure 2** Formation of G-quartets in the tetragonal crystals of the  $\sigma_2^A$ -DNA complex ( $P422$ ). (a) The asymmetric unit of the crystals. The sequence of the oligo used in cocrystallization is shown: 5'-TGTACAATGGG-3'. The  $\sigma_2^A$  protein is shown as a cartoon and the ssDNA is shown as sticks with C atoms shown in yellow, N atoms in blue, O atoms in red and P atoms in orange.  $K^+$  ions are shown as purple spheres. (b) View down a fourfold crystallographic axis. The three symmetry-related complexes are shown in gray. (c) Left, the same view as in (b). The experimental electron-density map (blue mesh, contoured at  $1\sigma$ ) is superimposed. Right, a view along a twofold axis (rotated  $90^\circ$  as shown). The arrow points to a T nucleotide flipped out of the quadruplex DNA.

(Fig. 3*a*). One of the nucleotides ( $T_{-7}$ ) in the connecting regions is flipped out of the four-helix bundle (Figs. 2*c* and 3*b*), while the remaining four-plane stacks are apparently composed of A, T and C quartets (Fig. 3*b*). Such structures with planar quartets formed by bases other than G have been observed previously (Patel & Hosur, 1999; Patel *et al.*, 2000; Searle *et al.*, 2004). Disorder in the connecting regions of the columns may reflect multiple ways of arranging non-guanine bases in quartets.

The complex with a TGTATAATGGG oligo yielded  $P422$  crystals that diffracted to slightly lower resolution than crystals of the  $\sigma_{2-3}^A$ -TGTACAATGGG complex (2.7 *versus* 2.1 Å). Both structures (with TATAAT and TACAAT –10 element sequences) appeared virtually identical (r.m.s.d. of 0.165 Å over all atoms; Feklistov & Darst, 2011). The single-nucleotide difference between the oligo sequences may be responsible for the difference in diffraction properties of the crystals, since the central part of the oligo forms the regions of the quadruplex columns that connect layers of packed protein (Fig. 3). We also observed crystal growth for protein–ssDNA complexes containing oligos with variations at the 5'-end: one- or two-nucleotide extensions still allowed crystal formation, although changes of the GGG 3'-end were not tolerated (data not shown). Therefore, different sequences can be accommodated in the quadruplex columns in the observed crystal-packing arrangement.

Examination of the packing in the  $P6_222$  crystal form revealed an organization related to the  $P422$  crystal form, with four copies of the  $\sigma_2^A$ -ssDNA complex (two asymmetric units) arranged around a G-quadruplex running along the twofold crystallographic symmetry axes corresponding to the *a* and *b* axes of the unit cell (Fig. 4). The organization of the G-columns in  $P6_222$  is similar to that in the  $P422$  crystal form, although their relative positions to each other are different: the columns are organized in a parallel fashion within the planes perpendicular to the *c* axis of the unit cell and are related by a

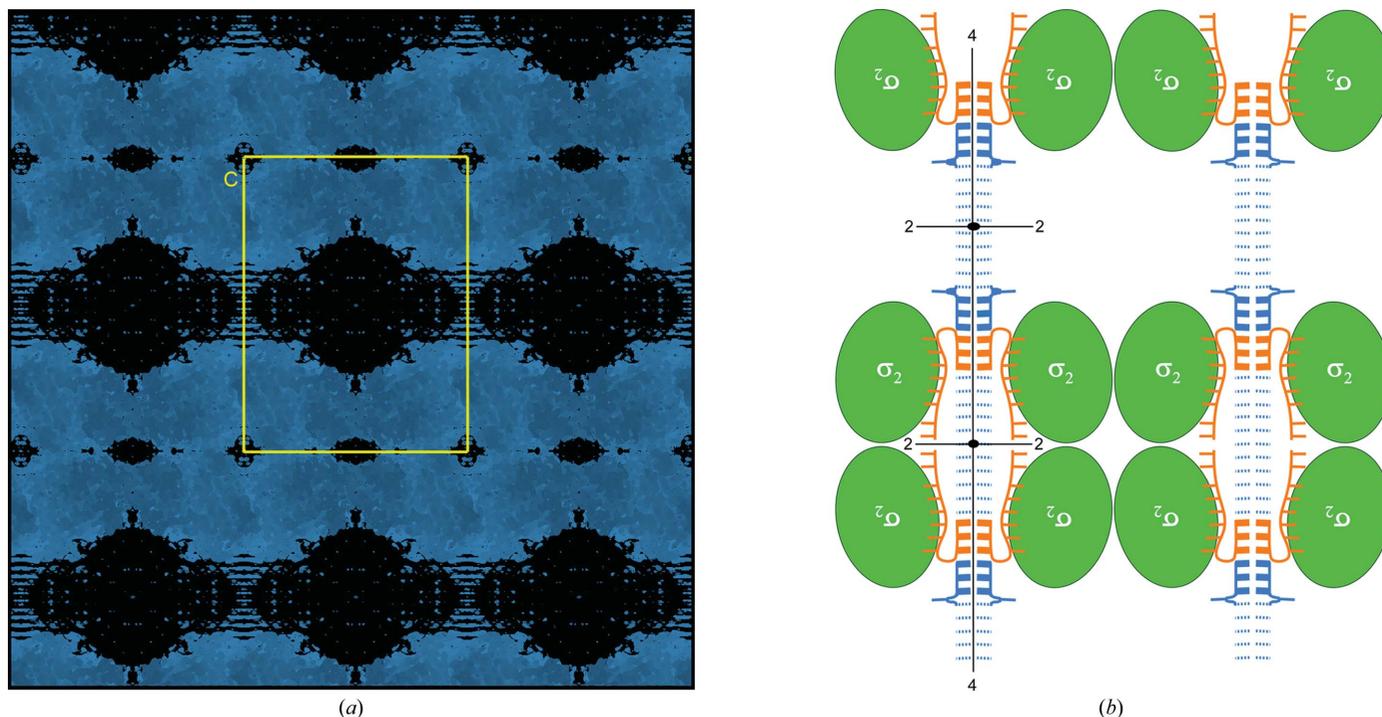
translation of 1/2 along either the *a* or *b* axis, whereas between the planes the columns run at a 120° angle (Fig. 4). Another important difference from  $P422$  crystal form, illustrated in Fig. 4(*a*), is that we were unable to observe even weak density for connecting quadruplexes (formed by oligos unbound to protein) except for the planes marked in blue in Fig. 4(*b*). The columns therefore appear to be discontinuous as opposed to the pseudo-continuous columns of the  $P422$  form.

It is unclear why the  $P422$  crystals diffracted better than the  $P6_222$  crystals; both crystal forms have very similar solvent contents (66.22 and 65.33%, respectively). We did not observe density for the mellitic acid (required for the  $P422$  crystal form), an additive from the Silver Bullets screen (Hampton Research), but this small molecule has been reported to be found in the intermolecular interface bridging packed protein molecules (Larson *et al.*, 2007).

Density for the  $\sigma_3^A$  domain (responsible for recognition of the –10 extended element  $T_{-15}G_{-14}$  (Murakami *et al.*, 2002) was absent from the  $P422$  crystals and it was presumed to be disordered. The asymmetric unit of the  $P6_222$  crystal form contains two  $\sigma_{2-3}^A$  molecules, but only one  $\sigma_3^A$  was ordered, although it did not form any biologically relevant interactions with the DNA (Fig. 4).

### 3.2. Possible applications for the crystallization of protein–ssDNA complexes

Packing in crystals of protein–DNA complexes is very often generated by interaction between ends of DNA molecules; therefore, variation of the terminal regions of DNA is a common approach in cocrystallization trials (Anderson *et al.*, 1984; Tan *et al.*, 2000). This is especially true for double-stranded DNA (dsDNA), but in ssDNA or ssRNA protein complexes terminal nucleotides also often participate in crystal packing (Werten & Moras, 2006; de Silva *et al.*, 2007; Phipps



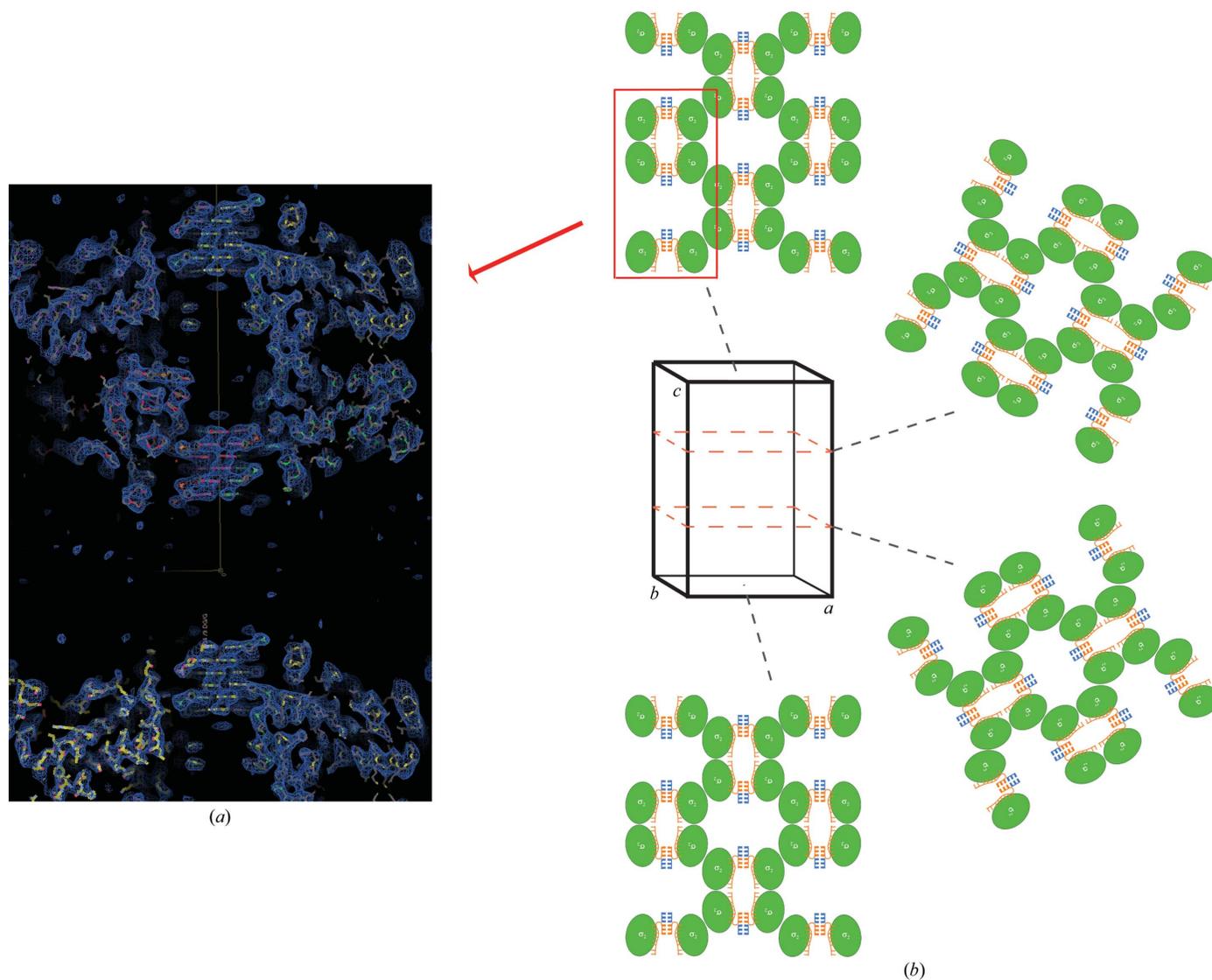
**Figure 3** Packing in the tetragonal crystals of the  $\sigma_{2-3}^A$ -DNA complex ( $P422$ ). (*a*)  $2F_o - F_c$  electron-density map showing the crystal packing (contoured at  $1\sigma$ ). Shown are layers of  $\sigma_2^A$  molecules connected by long DNA quadruplex columns. The unit cell is shown in yellow. (*b*) Schematic representation of the crystal packing shown in (*a*). Oligos complexed with  $\sigma_2^A$  (green) are shown in orange. Additional DNA strands not bound to protein and participating in DNA quadruplex formation are shown in blue. DNA present in the model is shown by bold lines and partially ordered DNA is shown by dashed thin lines representing stacked quartet planes

& Li, 2007; Warren *et al.*, 2009). A common feature of protein–dsDNA complexes is a pseudo-continuous helix formed by dsDNA molecules packed end to end. Variation of the overall length of the dsDNA fragment, as well as the nature of the terminal nucleotides, allows one to predict and control, to a certain extent, interactions between DNA ends as well as the helical repeat that positions DNA-bound protein relative to its neighbors in the crystal lattice (Tan *et al.*, 2000).

In the crystals discussed here, the pseudo-continuous G-quartet columns spanning the crystal lattices are formed by stacked G-quadruplexes. Although a number of G-quartet–protein complex structures have been reported, the crystal packing in each case is mediated only by protein–protein or protein–DNA contacts (PDB entries 1jb7, 4dih, 4dii, 1pa6 and 1ph1–1ph9; Horvath & Schultz, 2001; Russo Krauss *et al.*, 2012; Theobald & Schultz, 2003). The structures discussed here are, to our knowledge, the first example of a protein–DNA complex in which the crystal packing is determined by the formation of a pseudo-continuous G-quadruplex.

For dsDNA–protein complexes detailed strategic approaches to maximize the chances of success in crystallization trials have been formulated (Tan *et al.*, 2000), but no such rules have been suggested for ssDNA–protein complexes. Here, we report two different crystal forms of an ssDNA–protein complex formed by virtue of extended G-quadruplex packing. This type of packing can therefore be compatible with at least two types of symmetry arrangements (tetragonal and hexagonal) and different DNA sequences. Although this may be seen as a unique case, one could attempt to extend this approach to other protein–ssDNA crystals. We propose that systematically introducing short stretches of guanine nucleotides into terminal regions of ssDNA during cocrystallization trials may promote the formation of G-quartets mediating packing in crystals of protein–ssDNA or protein–ssRNA complexes.

A crystal lattice dominated by four-stranded helices leaves ample room for biologically relevant interactions in non-quadruplex parts, where specific protein–ssNA (single-stranded nucleic acid) interaction can reside. The periodic length of the G-column can be



**Figure 4** Packing in the hexagonal crystals of the  $\sigma_{2-3}$ -DNA complex (*P6<sub>22</sub>*). (a) Shown is a  $2F_o - F_c$  electron-density map (contoured at  $1\sigma$ ) corresponding to the fragment of the lattice delineated in (b) (red square). The view is along the *a* axis (yellow line). (b) The unit cell is schematically shown in the centre. Four parallel planes dissecting the unit cell demonstrate the arrangement of the protein–DNA complexes within each plane (viewed along the *c* axis).  $\sigma_2^A$  is represented by green ovals;  $\sigma_3^A$  domains were omitted for clarity.

adjusted by flipping bases out of the column (as observed in the tetragonal crystal form; Figs. 2c and 3b), expanding options for crystal packing.

#### 4. Concluding notes and future prospects

DNA self-assembly properties have been employed for building a great variety of ordered molecular patterns (Seeman, 2010; Carneiro *et al.*, 2013). One of the promising applications of DNA nanotechnology is in organizing other molecular species on nucleic acid scaffolds. The idea of the rational design of crystals with proteins which are otherwise recalcitrant to crystallization arranged on DNA arrays was first put forward by Nadrian Seeman in 1982 (Seeman, 1982). This method proved successful in building two-dimensional protein–DNA crystals using DNA Holliday junctions assembled from four oligonucleotides and RuvA protein, which naturally binds to Holliday junctions (Malo *et al.*, 2005). Alternative ways of attaching proteins to a DNA array include the use of aptamer sequences (Liu *et al.*, 2005; Chhabra *et al.*, 2007) and various affinity or chemical tagging methods (Saccà & Niemeyer, 2011). These examples mostly employed duplex DNA for building the structural scaffold, whereas the architectural properties of quadruplex DNA have not been systematically explored.

G-quadruplexes are four-stranded DNA or RNA structures formed by G-rich sequences (Burge *et al.*, 2006; Lane *et al.*, 2008). They are built from stacked G-quartets: planar structures of four guanine bases connected by Hoogsteen base pairing and coordinated by monovalent cations (most often K<sup>+</sup>) lying on the central axis of the G-quadruplex. The G-stacks can be connected by flexible single-stranded loops. Quadruplexes display a variety of topological folds: they can be built from one, two, three or four separate strands of a nucleic acid and the direction of the strands and loop sizes can also vary (Lane *et al.*, 2008; Zhou *et al.*, 2012). Therefore, quadruplexes demonstrate significant structural variability that can be sampled during the formation of a protein–ssNA crystal lattice. As a tool for crystal engineering, quadruplexes may present significant advantages over dsDNA owing to their rigidity and high stability.

G-quadruplexes have attracted interest for their potential use in molecular nanotechnology owing to their ability to self-assemble in continuous columns by stacking on one another (Davis & Spada, 2007; Aldaye *et al.*, 2008). Short DNA oligos have been shown to form structures reaching the micrometre scale (G-wires; Marsh *et al.*, 1995). The self-assembly of quadruplexes can be controlled by the nature of the solvent and the salt providing cations for chelation within the central cavity (González-Rodríguez *et al.*, 2009). Initial attempts to direct the assembly of G-quadruplexes by using proteins as anchor points have been reported (Borovok *et al.*, 2008).

The crystal-packing arrangements discussed here demonstrate the potential of G-quadruplexes for the building of highly ordered micrometre-scale crystals and should inspire the development of ‘bottom-up’ strategies employing G-quartets as tools for crystal engineering.

We thank M. Bick for help with data collection, E. A. Campbell and A. Weixlbaumer for advice with refinement procedures and K. R. Rajashankar and F. Murphy at APS NE-CAT beamline 24ID-E and W. Shi at NSLS beamline X29 for support with synchrotron data collection. AF was a Merck Postdoctoral Fellow at The Rockefeller University. This work was based, in part, on research conducted at the APS and the NSLS, supported by the US Department of Energy, Office of Basic Energy Sciences. The NE-CAT beamlines at the APS

are supported by award RR-15301 from the NCRR at the NIH. This work was supported by NIH RO1 GM053759 to SAD.

#### References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
- Aldaye, F. A., Palmer, A. L. & Sleiman, H. F. (2008). *Science*, **321**, 1795–1799.
- Anderson, J., Ptashne, M. & Harrison, S. C. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 1307–1311.
- Borovok, N., Iram, N., Zikich, D., Ghabboun, J., Livshits, G. I., Porath, D. & Kotlyar, A. B. (2008). *Nucleic Acids Res.* **36**, 5050–5060.
- Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. (2006). *Nucleic Acids Res.* **34**, 5402–5415.
- Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C., Weinman, O., Trester-Zedlitz, M. L. & Darst, S. A. (2002). *Mol. Cell*, **9**, 527–539.
- Carneiro, K. M. M., Avakyan, N. & Sleiman, H. F. (2013). *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **5**, 266–285.
- Chhabra, R., Sharma, J., Ke, Y., Liu, Y., Rinker, S., Lindsay, S. & Yan, H. (2007). *J. Am. Chem. Soc.* **129**, 10304–10305.
- Davis, J. T. & Spada, G. P. (2007). *Chem. Soc. Rev.* **36**, 296.
- Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Feklistov, A., Barinova, N., Sevostyanova, A., Heyduk, E., Bass, I., Vvedenskaya, I., Kuznedelov, K., Merkiene, E., Stavrovskaya, E., Klimauskas, S., Nikiforov, V., Heyduk, T., Severinov, K. & Kulbachinskiy, A. (2006). *Mol. Cell*, **23**, 97–107.
- Feklistov, A. & Darst, S. (2011). *Cell*, **147**, 1257–1269.
- González-Rodríguez, D., van Dongen, J. L., Lutz, M., Spek, A. L., Schenning, A. P. & Meijer, E. W. (2009). *Nature Chem.* **1**, 151–155.
- Haugen, S. P., Berkmen, M. B., Ross, W., Gaal, T., Ward, C. & Gourse, R. L. (2006). *Cell*, **125**, 1069–1082.
- Hook-Barnard, I. G. & Hinton, D. M. (2007). *Gene Regul. Syst. Biol.* **1**, 275–293.
- Horvath, M. P. & Schultz, S. C. (2001). *J. Mol. Biol.* **310**, 367–377.
- Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. (2008). *Nucleic Acids Res.* **36**, 5482–5515.
- Larson, S. B., Day, J. S., Cudney, R. & McPherson, A. (2007). *Acta Cryst.* **D63**, 310–318.
- Liu, X., Bushnell, D. A. & Kornberg, R. D. (2011). *Cell*, **147**, 1218–1219.
- Liu, Y., Lin, C., Li, H. & Yan, H. (2005). *Angew. Chem. Int. Ed. Engl.* **44**, 4333–4338.
- Malo, J., Mitchell, J. C., Vénien-Bryan, C., Harris, J. R., Wille, H., Sherratt, D. J. & Turberfield, A. J. (2005). *Angew. Chem. Int. Ed. Engl.* **44**, 3057–3061.
- Marsh, T. C., Vesenska, J. & Henderson, E. (1995). *Nucleic Acids Res.* **23**, 696–700.
- 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.
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. (2002). *Science*, **296**, 1285–1290.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Patel, P., Bhavesh, N. S. & Hosur, R. (2000). *Biochem. Biophys. Res. Commun.* **270**, 967–971.
- Patel, P. K. & Hosur, R. V. (1999). *Nucleic Acids Res.* **27**, 2457–2464.
- Phipps, K. R. & Li, H. (2007). *Proteins*, **67**, 121–127.
- Russo Krauss, I., Merlino, A., Randazzo, A., Novellino, E., Mazzarella, L. & Sica, F. (2012). *Nucleic Acids Res.* **40**, 8119–8128.
- Saccà, B. & Niemeyer, C. M. (2011). *Chem. Soc. Rev.* **40**, 5910–5821.
- Searle, M. S., Williams, H. E., Gallagher, C. T., Grant, R. J. & Stevens, M. F. (2004). *Org. Biomol. Chem.* **2**, 810–812.
- Seeman, N. C. (1982). *J. Theor. Biol.* **99**, 237–247.
- Seeman, N. C. (2010). *Annu. Rev. Biochem.* **79**, 65–87.
- Shultzaberger, R. K., Chen, Z., Lewis, K. A. & Schneider, T. D. (2006). *Nucleic Acids Res.* **35**, 771–788.
- Silva, U. de, Choudhury, S., Bailey, S. L., Harvey, S., Perrino, F. W. & Hollis, T. (2007). *J. Biol. Chem.* **282**, 10537–10543.
- Tan, S., Hunziker, Y., Pellegrini, L. & Richmond, T. J. (2000). *J. Mol. Biol.* **297**, 947–959.
- Theobald, D. L. & Schultz, S. C. (2003). *EMBO J.* **22**, 4314–4324.
- Warren, E. M., Huang, H., Fanning, E., Chazin, W. J. & Eichman, B. F. (2009). *J. Biol. Chem.* **284**, 24662–24672.
- Werten, S. & Moras, D. (2006). *Nature Struct. Mol. Biol.* **13**, 181–182.
- Zhou, J., Bourdoncle, A., Rosu, F., Gabelica, V. & Mergny, J.-L. (2012). *Angew. Chem. Int. Ed. Engl.* **51**, 11002–11005.