Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A novel inhibitor-binding site on the HIV-1 capsid N-terminal domain leads to improved crystallization *via* compound-mediated dimerization

Christopher T. Lemke,[a]* Steve Titolo,[b]‡ Nathalie Goudreau,[a] Anne-Marie Faucher,[a]§ Stephen W. Mason[b]¶ and Pierre Bonneau[a]

[a]Department of Chemistry, Boehringer Ingelheim (Canada) Ltd, 2100 Cunard Street, Laval, Quebec H7S 2G5, Canada, and [b]Department of Biological Sciences, Boehringer Ingelheim (Canada) Ltd, 2100 Cunard Street, Laval, Quebec H7S 2G5, Canada

‡ Present address: AL-G Technologies, Lévis, Quebec, Canada.
§ Present address: Université de Montréal, Département de Chimie, Montréal, Quebec, Canada.
¶ Present address: Bristol-Myers Squibb, Virology, Wallingford, Connecticut, USA.

Correspondence e-mail: RESGeneral.LAV@boehringer-ingelheim.com

Received 28 January 2013
Accepted 5 March 2013

PDB References: HIV-1 capsid N-terminal domain, complex with compounds 1 and 2, 4e91; complex with compounds 1 and 3, 4e92

Despite truly impressive achievements in the global battle against HIV there remains a need for new drugs directed against novel targets, and the viral capsid protein (CA) may represent one such target. Intense structural characterization of CA over the last two decades has provided unprecedented insight into the structure and assembly of this key viral protein. Furthermore, several inhibitor-binding sites that elicit antiviral activity have been reported on CA, two of which are located on its N-terminal domain (CA_{NTD}). In this work, the binding of a novel capsid-assembly inhibitor that targets a unique inhibitory site on CA_{NTD} is reported. Moreover, whereas cocrystallization of CA_{NTD} in complex with ligands has proven to be challenging in the past, the use of this inhibitor as a tool compound is shown to vastly facilitate ternary cocrystallizations with CA_{NTD}. This improvement in crystallization is likely to be achieved through the formation of a compound-mediated homodimer, the intrinsic symmetry of which greatly increases the prospect of generating a crystal lattice. While protein engineering has been used in the literature to support a link between the inherent symmetry of a macromolecule and its propensity to crystallize, to our knowledge this work represents the first use of a synthetic ligand for this purpose.

1. Introduction

The World Health Organization has estimated that as of 2011 the population infected by HIV has reached 34 million individuals, with 1.7 million AIDS-related deaths per year (http://www.who.int/hiv/data/2012_epi_core_en.png). The current standard of care, known as highly active antiretroviral therapy (HAART), or more contemporarily as combination antiretroviral therapy (cART), combines multiple drugs from various mechanistic classes and has considerably improved the life expectancy of HIV-positive patients (Broder, 2010; McManus *et al.*, 2012). Many challenges remain, however, including the continued emergence of multi-drug-resistant isolates, which underscores the ongoing need to discover new classes of HIV inhibitors (Moreno *et al.*, 2010; Taiwo *et al.*, 2010).

The HIV-1 capsid protein (CA) plays key roles in both the early and the late phases of the HIV replication cycle and therefore has significant potential for antiviral intervention (for reviews, see Prevelige, 2011; Waheed & Freed, 2012). CA is expressed within the 55 kDa Gag polyprotein, which subsequently assembles and undergoes proteolytic cleavage during virus maturation (reviewed in Sundquist & Krausslich, 2012). During the maturation process, the viral RNA genome, along with several viral enzymes, becomes encased within a

lattice of multimerized CA to generate the mature viral core of infectious HIV.

Monomeric CA is comprised of two independently folded domains that are connected by a short flexible linker: the CA N-terminal domain (CA_{NTD}; residues 1–146) and the CA C-terminal domain (CA_{CTD}; residues 151–231). NMR and X-ray structures of the isolated domains, as well as dimeric, pentameric and hexameric assemblies, have been solved (Gamble *et al.*, 1996; Gitti *et al.*, 1996; Pornillos *et al.*, 2009, 2011; Wong *et al.*, 2008), thereby elucidating the detailed interactions that make up the capsid lattice (reviewed in Ganser-Pornillos *et al.*, 2012; Sundquist & Krausslich, 2012). The assembled mature CA lattice is now known to contain extensive intermolecular interfaces, including CA_{NTD}–CA_{NTD}, CA_{NTD}–CA_{CTD} and CA_{CTD}–CA_{CTD}. Since it has been shown that CA mutations that prevent assembly result in non-infectious virus particles (Ganser-Pornillos *et al.*, 2004; von Schwedler *et al.*, 2003), all of these inter-domain interactions may be considered as potential targets for therapeutic intervention. Moreover, it has been shown that the overall stability of the core is critical to proper coating and uncoating, with cores that are either too stable or too unstable also showing reduced infectivity (Forshey *et al.*, 2002). Therefore, molecules that simply stabilize or destabilize these lattice interactions may also yield a therapeutic effect.

Several peptidic and small-molecule inhibitors of capsid assembly have been reported (for a recent review, see Bocanegra *et al.*, 2012); two of these chemotypes stem from our own efforts and both target CA_{NTD}. CA_{NTD} has proven to be a very structurally malleable protein; significant variation has been noted among its several loops and even among its constituent helices. Two small-molecule inhibitory sites have been reported for the domain: the first is exemplified by the CAP-1 inhibitor (Kelly *et al.*, 2007; Tang *et al.*, 2003) and the second by PF-3450074 (Blair *et al.*, 2010). The majority of the hits from our screening efforts were found to target the CAP-1 site, as described in Goudreau *et al.* (2013). A defining feature of the CAP-1 site is a large variably inducible hydrophobic pocket located at the base of the CA_{NTD} helical bundle (Kelly *et al.*, 2007; Lemke *et al.*, 2012). The aforementioned inherent flexibility of CA_{NTD} coupled with varying induction of the CAP-1 pocket initially rendered X-ray support of CA_{NTD} hit-to-lead quite challenging; success rates varied widely, with crystallization conditions changing significantly between, and sometimes within, the various chemotypes.

In previously reported studies, the inherent flexibility of the capsid was somewhat mitigated through either the formation of CA complexes with cyclophilin A (Gamble *et al.*, 1996; Howard *et al.*, 2003) or Fab from anti-CA antibodies (Berthet-Colominas *et al.*, 1999; Momany *et al.*, 1996) or by major loop truncation (Blair *et al.*, 2010). Although potentially effective, such methods may also jeopardize correct compound binding, particularly in the case of unoptimized low-affinity ligands. In the present study, we describe the complex of a novel inhibitor of capsid assembly (Fig. 1; compound 1) which binds to a unique apical site on CA_{NTD}. The discovery and preliminary structure–activity relationship of this interesting new inhibitor series is simultaneously being reported elsewhere (Goudreau *et al.*, 2013). In addition to increasing the number of reported CA_{NTD} inhibitory sites to three, we here report the use of this compound as a broadly applicable tool for the synthetic dimerization of CA_{NTD}. This use of this compound as a crystallization tool led to a vast improvement in the success rate of CA_{NTD} cocrystallization with a minimal effect on compound binding elsewhere on the target. As examples, we describe the ternary cocrystallizations of two previously reported CAP-1 site inhibitors (Fig. 1; compounds 2 and 3), both of which had originally presented significant challenges in binary cocrystallization efforts. We propose that the observed highly improved propensity to crystallize stems principally from a fortuitous compound-mediated dimerization of CA_{NTD}.

2. Materials and methods

2.1. Plasmid construction, expression and purification

A plasmid containing CA_{NTD} (WISP-96-19; CA residues 1–146) was generously provided by the Sundquist laboratory. The CA_{NTD}-coding fragment was transferred to the pET-11a expression vector by PCR amplification using primers that introduced an *Nde*I site and a start codon at the 5′-end and a *Bam*HI site and a stop codon at the 3′-end. For expression, the construct was transformed into *E. coli* BL21 (DE3) cells (Novagen). LB medium was inoculated with overnight precultures, which were then grown at 310 K until mid-log phase (Abs_{600} of ∼0.6). Protein expression was then induced by the

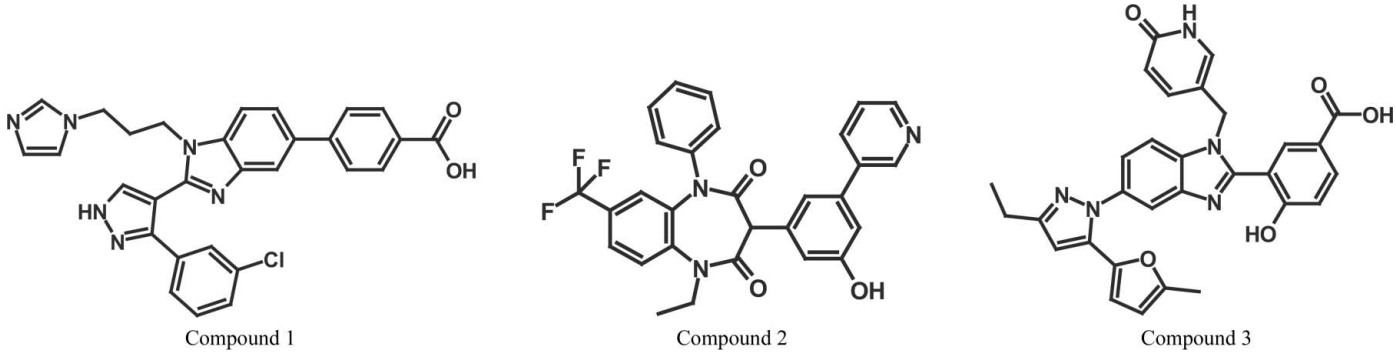

Figure 1
Compounds used in this study. Compounds 2 and 3 are referred to as BD3 and BM4, respectively, in Lemke *et al.* (2012).

addition of 0.5–1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) and was carried out for 4–6 h at 303 K. Cells were harvested by centrifugation and pellets were stored at 193 K until purification.

Purification of CA_{NTD} took place as follows: 5–10 g cell paste was lysed by sonication in 40 ml buffer *A* (20 m*M* MES pH 6.5, 10 m*M* β-mercaptoethanol) supplemented with 0.5 *M* NaCl and Complete EDTA-free protease-inhibitor tablets (Roche). Nucleic acids and cell debris were removed by adding 0.11 volumes of 0.2 *M* ammonium sulfate and an equivalent volume of 10% poly(ethyleneimine) pH 8.0 and stirring the sample for 20 min at 277 K followed by centrifugation at 30 000*g* for 20 min. CA_{NTD} protein was recovered from the supernatant by adding 0.6 volumes of saturated ammonium sulfate solution followed by centrifugation at 10 000*g* for 15 min. The pellet was dissolved in 10 ml buffer *A* and dialyzed overnight in the same buffer with a 10 kDa molecular-weight cutoff. The sample was clarified by centrifugation and sequentially passed through HiTrap TMSP HP and Q HP columns (GE Healthcare) pre-equilibrated in buffer *A*. CA_{NTD} was recovered in the flowthrough and wash fractions, concentrated and passed through a Superdex 75 gel-filtration column using 15 m*M* MES pH 6.5 for pre-equilibration and as the running buffer. Peak fractions were concentrated to 15 mg ml^{-1} as determined by Abs_{280} using a calculated molar extinction coefficient (ε = 25 320 M^{-1} cm^{-1}). Aliquots were then flash-frozen and stored at 193 K for later use.

2.2. Ternary cocrystallization screening of CA_{NTD} complexed with compound 1 and CAP-1 site binders

Frozen aliquots of CA_{NTD} at a concentration of 15 mg ml^{-1} were thawed in room-temperature (approximately 291 K) running water. A 150 m*M* stock of compound 1 in DMSO was diluted 100-fold with protein solution and the mixture was incubated on ice for approximately 15 min. A stock of 150 m*M* target compound in DMSO was then also diluted 100-fold with the protein–compound 1 mixture. The final protein–inhibitor solution (consisting of approximately 15 mg ml^{-1} CA_{NTD}, 15 m*M* MES pH 6.5, 1.5 m*M* compound 1, 1.5 m*M* target compound, 2% DMSO) was then used for crystallization screening *via* hanging-drop vapour diffusion (2–4 µl drops equilibrated over 500 µl wells) at 291 K using commercially available screens. Crystals typically appeared within two weeks.

After several rounds of crystallization screening using a variety of CAP-1 site binders, a six-solution mini-screen was developed based on two clusters of common hits. The ingredients of the focused screen were as follows: condition 1, 22.5%(*w*/*v*) PEG 8000, 0.1 *M* bicine pH 8.5, 0.1 *M* sodium acetate; condition 2, 5% PEG 20 000, 15% PEG 400, 0.1 *M* TAPS pH 9.0, 150 m*M* NaBr; condition 3, 15% PEG 20 000, 0.1 *M* TAPS pH 9.0, 0.1 *M* NaBr, 20% glycerol; condition 4, 0.5 *M* sodium/potassium tartrate, 0.2 *M* NaI; condition 5, 0.7 *M* sodium/potassium tartrate, 0.1 *M* bis-tris pH 7.0, 200 m*M* NaI, 15%(*v*/*v*) glycerol; condition 6, 0.85 *M* sodium/potassium tartrate, 0.1 *M* citrate pH 6.25, 175 m*M* ammonium sulfate.

New ternary CA_{NTD} complexes were first screened against this focused screen, thus often greatly reducing the reagents and the time required to obtain diffraction-quality crystals. Crystallization often occurred spontaneously, but in many cases was improved by streak-seeding using crystals of CA_{NTD} complexed with compound 1 and a second unrelated compound ('cross-seeding'). Use of the focused screen was particularly advantageous in cases where seeding was necessary; streak-seeding the six-solution screen was relatively quick and simple compared with seeding larger commercial screens. Further simplifying matters, it was later found that stably microseeded focused screen solutions could be generated by crushing crystals in the well solutions that were used to grow them (using Hampton Research Seed Beads as per the manufacturer's instructions) and then diluting this concentrated seed stock 10 000-fold in each of the solutions of the focused screen. Several versions of the focused screen were prepared over the course of the project with varying numbers and species of seed crystal forms. The relationship between the number and identities of seed crystal forms present in the seeded screens and crystallization success was not assessed. In many cases the crystallization screening hits were of sufficient quality for immediate X-ray diffraction and analysis, particularly those from the focused screen.

2.3. Crystallization of CA_{NTD} in complex with compounds 2 and 3

The ternary cocrystallization of CA_{NTD} in complex with compounds 1 and 2 was as described above except that the final protein–inhibitor mixture consisted of approximately 15 mg ml^{-1} CA_{NTD}, 15 m*M* MES pH 6.5, 2 m*M* compound 1, 3 m*M* compound 2, 2.8% DMSO. The optimized crystallization solution was 0.7 m*M* sodium/potassium tartrate, 0.1 *M* bis-tris pH 7.0, 0.2 *M* NaI, 16% glycerol. The crystals grew spontaneously at pH 6.25 and lower, but could be grown at higher pH *via* streak-seeding to yield larger and better defined crystals. The optimized crystals grew as hexagonal blocks that could exceed 400 µm in length and 100 µm in width (Fig. 3*g*). Individual crystals were mounted in cryoloops and were flash-cooled in liquid nitrogen without additional cryoprotectant.

The ternary cocrystallization of CA_{NTD} in complex with compounds 1 and 3 was as described above except that the final protein–inhibitor mixture consisted of approximately 15 mg ml^{-1} CA_{NTD}, 15 m*M* MES pH 6.5, 1 m*M* compound 1, 1 m*M* compound 3, 2% DMSO. The optimized crystallization solution was 0.5 m*M* sodium/potassium tartrate, 0.2 *M* NaI. The first crystals of this form required cross-seeding in order to nucleate crystal growth, as a multiply seeded focused screen yielded crystals while the equivalent unseeded screen did not. Thereafter, streak-seeding of this crystal form was used to generate optimized crystals. The crystals were bipyramidal and grew to approximately 100 µm across the base (Fig. 3*f*). Immediately prior to flash-cooling in liquid nitrogen, individual crystals were briefly transferred into a well solution

supplemented with 15% glycerol together with 1 m*M* each of compounds 1 and 3.

2.4. X-ray diffraction, structure solution and model building

Diffraction data were collected at 100 K on an FR-E X-ray generator equipped with Osmic HiRes^2 optics and a MAR345dtb image-plate detector. Data reduction and scaling were performed using *HKL*-2000 (Otwinowski & Minor, 1997). Preliminary models were obtained *via* rigid-body refinement in *PHENIX* (Adams *et al.*, 2010) using a model ultimately derived from PDB entry 1gwp (Tang *et al.*, 2002). Further iterations of refinement using *PHENIX* and manual model building using *Coot* (Emsley & Cowtan, 2004) yielded the final models. The data-processing statistics and model-refinement statistics are listed in Table 1. The stereochemical quality of both models was assessed using *MolProbity* (Chen *et al.*, 2010); there were no Ramachandran outliers, with 97–99% of residues lying in favoured regions.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outermost shell.

	Compounds 1 + 2 (PDB entry 4e91)	Compounds 1 + 3 (PDB entry 4e92)
Data collection		
Space group	$P3_121$	$P4_12_12$
Unit-cell parameters (Å)	$a = b = 65.78$, $c = 143.32$, $\alpha = \beta = 90.0$, $\gamma = 120.0$	$a = b = 81.71$, $c = 90.15$, $\alpha = \beta = 90.0$, $\gamma = 120.0$
Resolution range (Å)	57–1.7 (1.76–1.70)	34–1.8 (1.86–1.80)
Completeness (%)	97.7 (99.5)	97.3 (97.2)
Multiplicity	8.4 (5.9)	15.6 (7.6)
$\langle I/\sigma(I)\rangle$	14.7	14.8
$R_{\rm merge}$	0.067 (0.406)	0.050 (0.547)
Refinement		
Resolution range (Å)	37–1.7 (1.74–1.70)	34–1.8 (1.86–1.80)
No. of reflections, working set	39464 (2444)	28153 (2322)
No. of reflections, test set	1979 (128)	1427 (130)
Final $R_{\rm cryst}$	0.220 (0.319)	0.220 (0.411)
Final $R_{\rm free}$	0.243 (0.327)	0.264 (0.480)
No. of non-H atoms		
Protein	2101	2184
Ions	4	0
Ligands	114	156
Waters	275	165
Total	2494	2505
R.m.s. deviations		
Bond lengths (Å)	0.006	0.004
Bond angles (°)	1.02	0.806
Average *B* factors (Å^2)		
Protein	20.4	34.0
Ions	21.1	n/a
Ligands	19.0	28.0
Waters	25.0	35.3
Ramachandran plot, residues in (%)		
Favoured regions	98.4	96.7
Additionally allowed regions	1.2	2.9

3. Results and discussion

3.1. Initial compound 1 characterization

Compound 1 was tested in an *in vitro* immobilized capsid-assembly assay and an HIV-1 NL4-3 antiviral assay, as described in Lemke *et al.* (2012), yielding an IC_{50} of 1.2 µ*M* and an EC_{50} of 57 µ*M*. While the cellular toxicity level of compound 1 was not evaluated, other members of this chemotype displayed improved antiviral activity that was clearly distinct from activity arising from cytotoxicity. As an interesting hit-stage series, two-dimensional ^{1}H–^{15}N HSQC NMR spectra were collected for uniformly ^{15}N-labelled HIV-1 CA_{NTD} in the presence and the absence of closely related compounds, as reported elsewhere (Goudreau *et al.*, 2013). Unlike the vast majority of hits, which were found to perturb residues in and around the base of the helical bundle (*i.e.* the CAP-1 binding site), the compound 1 chemotype was found to perturb a distinct subset of residues at the apex of the helical bundle. This combination of interesting assay activities with a distinct NMR profile led to the prioritization of this compound for crystallographic studies. While binary cocrystallization of CA_{NTD} with compound 1 readily yielded well diffracting crystals and the structure of the complex, the use of compound 1 in ternary CA_{NTD} structure determinations proved to be much more interesting, as discussed below.

3.2. Models of the ternary complexes

The ternary complex of CA_{NTD} with compounds 1 and 2 (CA_{NTD}–C1–C2) crystallized in space group $P3_221$ with two CA_{NTD} molecules per asymmetric unit. Owing to protein flexibility, the final CA_{NTD}–C1–C2 model is missing residues 5–10 and 87–95 in monomer *A* and residues 86–95 in monomer *B*. Interestingly, although both monomers bind a molecule of compound 1, only monomer *A* has a molecule of compound 2 bound; the CAP-1 site of monomer *B* remains in an unliganded apo conformation. It is likely that a combination of compound insolubility and crystal-packing preferences lead to the asymmetry in this crystal form. The ternary complex of CA_{NTD} with compounds 1 and 3 (CA_{NTD}–C1–C3) crystallized in space group $P4_12_12$ with two CA_{NTD} molecules per asymmetric unit. Owing to protein flexibility, the final CA_{NTD}–C1–C3 model is missing residues 88–93 in both monomers *A* and *B*. Each monomer binds one molecule of compound 1 and one of compound 3. All ligand molecules are well defined by the experimental electron density. In this paper, we focus on the binding properties of compound 1. The binding of compounds 2 and 3, and an in-depth discussion of the biological effects thereof, has been described elsewhere (Lemke *et al.*, 2012).

The X-ray structures reported here confirmed the NMR chemical perturbation experiments that singled out compound 1 as binding to a novel inhibitory site on CA_{NTD}. The compound 1 'apical' binding site is well removed from the two previously reported inhibitory CA_{NTD} sites (Fig. 2*a*). The pocket is primarily comprised of residues from the base of the large loop that connects helices 4 and 5 (the 'cyclophilin A binding loop') and from the much shorter loop that connects helices 6 and 7. The shallow pocket extends approximately 20 Å along the surface and is relatively featureless, save for two small subpockets, one of which is occupied by the chlorophenyl moiety of compound 1 and the other by two ordered water molecules (Fig. 2*b*). Compound 1 appears to primarily

bind *via* hydrophobic interactions that anchor the molecule in the chlorophenyl pocket and extend along the length of the inhibitor. Residues with side chains that make significant hydrophobic contact with compound 1 are Trp80, His84, Met96, Trp117, His120, Pro122, Ile124, Pro125 and Ile129 (Fig. 2*c*). In addition to these hydrophobic interactions, the acid moiety of compound 1 forms a key bidentate salt bridge with the side chain of Arg132, thus providing a second anchor at the opposite end of the molecule. While no other direct intermolecular hydrogen bonds are observed in compound 1 binding, a well defined water molecule links the benzimidazole core of the inhibitor to the backbone carbonyl O atom of Pro123. Finally, the *N*-propylimidazole moiety of compound 1 is likely to form a favourable electrostatic interaction with the side chain of Glu98, although both moieties typically appear to be quite mobile. The compound 1 binding mode is essentially identical in the four models reported here (Fig. 2*d*), although in monomer *A* of the CA_{NTD}–C1–C3 structure compound 1 does appear to be slightly shifted in the pocket, possibly stemming from crystal-contact interference.

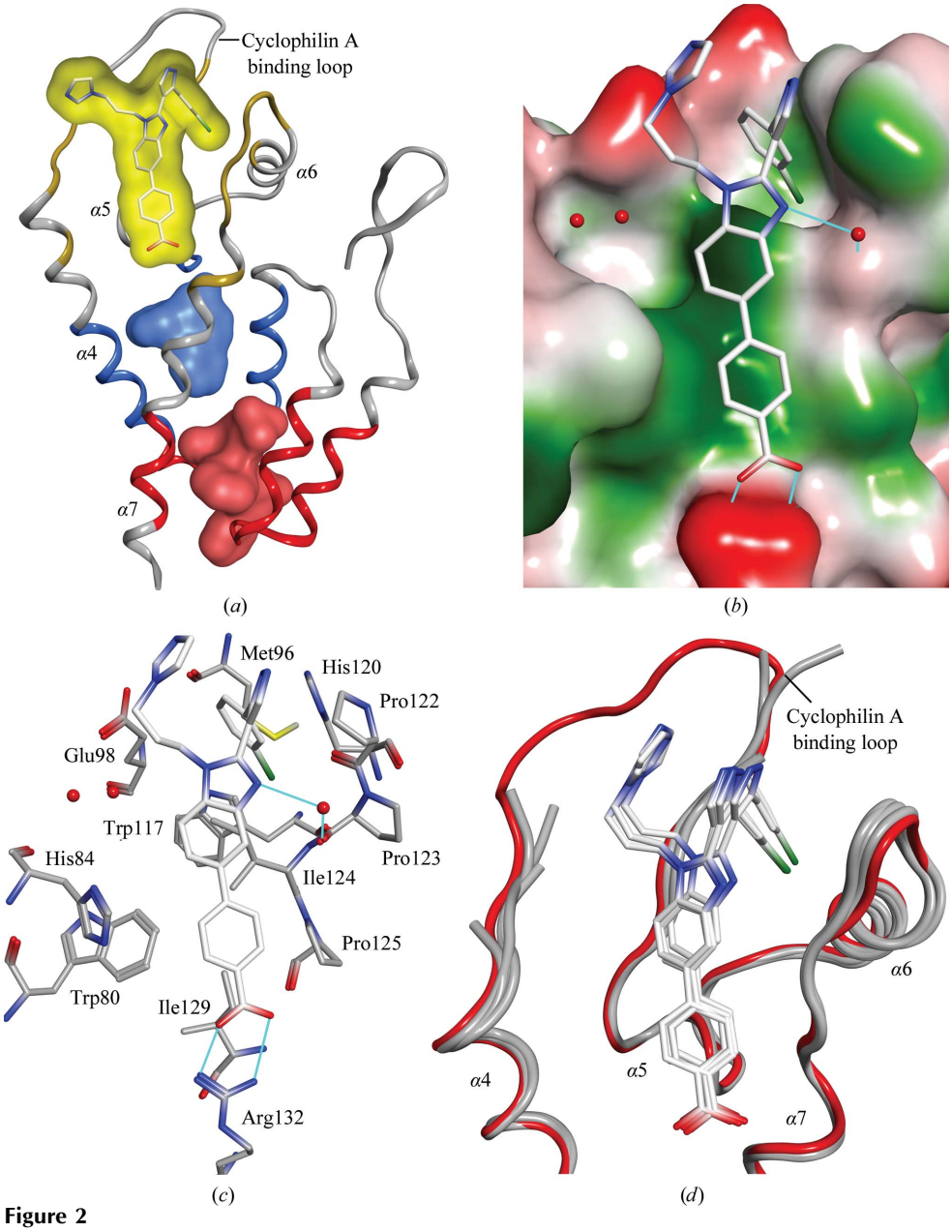

Figure 2
Binding of compound 1 to CA_{NTD}. (*a*) Relative positions of the three reported CA_{NTD} inhibitory sites. CA_{NTD} is represented in ribbon form, with the CAP-1, PF-3450074 and compound 1 sites indicated in red, blue and yellow, respectively. Surfaces of bound representatives are similarly colour-coded, with that of compound 1 made semi-transparent to reveal compound 1 in stick form. (*b*) Surface representation of the compound 1 binding site coloured by hydrophobicity: hydrophobic (green), neutral (white) and hydrophilic (red). Compound 1 is shown in stick form, while water molecules are indicated by red spheres and hydrogen bonds are indicated by blue lines. (*c*) Detailed interactions of compound 1. Compound 1 and interacting CA_{NTD} residues are illustrated as white and grey sticks, respectively. Water molecules are indicated by red spheres and hydrogen bonds are indicated by blue lines. (*d*) Effects of compound 1 binding. Superposition of the four bound compound 1 molecules reported here (grey) onto the 3mge structure (red).

Given that this series of compounds was identified by its ability to block capsid assembly in an *in vitro* assay, it was anticipated that they might demonstrate an observable morphological effect on CA_{NTD}. However, using a high-resolution apo CA_{NTD} structure for comparison (PDB entry 3mge; Pornillos *et al.*, 2010), the binding of compound 1 appears to induce little secondary-structure change in the pocket. The apo *versus* bound r.m.s.d.s calculated for the backbone atoms of selected residues within 9 Å of compound 1 (Ile73–Pro85, Met96–Ala105 and Glu113–Trp133) range between 0.45 and 0.57 Å. A close examination suggests that the two adjacent loops connecting helices 4 and 5 and helices 6 and 7 may slightly close in upon the compound (Fig. 2*d*). The most significant backbone differences appear within the cyclophilin A binding loop (residues 85–99); however, it is difficult to ascertain the relevance of these deviations owing to the high mobility of this region in the present structures and owing to its involvement in a significant crystal contact in the 3mge crystal form. Similarly, the relevance of the backbone shifts observed in the loop linking helices 6 and 7 is also unclear, as a range of positions are observed for this relatively mobile loop. At the level of side-chain conformation there also appears to be very

little perturbation observed on compound 1 binding. The chlorophenyl pocket is largely pre-formed, requiring only slight shifts of residues Met96, His120 and Ile124, as are the hydrophobic contacts along the length of the inhibitor. Even residue Arg132, which is highly solvent-exposed and free of crystal contacts in the 3mge structure, is largely pre-oriented for its intermolecular salt bridge with compound 1. Taken together, despite the fact that CA_{NTD} is often regarded as a highly flexible protein, the binding of compound 1 does not appear to result in any large structural deviations that would provide an obvious mechanism of action for the series.

3.3. Improved crystallization of CA_{NTD} using ternary complexes

Distinct from its potential antiviral role, the compound 1 series proved invaluable as a crystallization tool. The method of cocrystallizing ternary CA_{NTD} complexes described here was applied to a wide variety of CA_{NTD} ligands, several of which had previously failed to produce diffraction-quality crystals in binary complex with CA_{NTD}. These ternary CA_{NTD} cocrystallizations were extremely successful in the presence of compound 1, yielding more than 12 distinct well diffracting crystal forms (Fig. 3 and Table 2), from which dozens of complexed crystal structures were readily obtained. As might be expected given the distance between the compound 1 and CAP-1 binding sites, biochemical, NMR and X-ray experiments all indicated independent binding at the two sites (Goudreau *et al.*, 2013).

This unusually high number of distinct crystal forms resulting from the same protein construct is likely to be a reflection of the conformational flexibility of CA_{NTD}. For different compounds binding at the CAP-1 site we have observed variable shifts of helices 1 and 2 relative to the rest of the α-helical bundle at the core of CA_{NTD}, as well as a multitude of rearrangements of the loop connecting helices 3 and 4 (for examples, see Lemke *et al.*, 2012). Evidently, this continuum of possible conformations combined with the relatively small size of CA_{NTD} can give rise to a multitude of sufficiently different protein shapes to yield a large number of

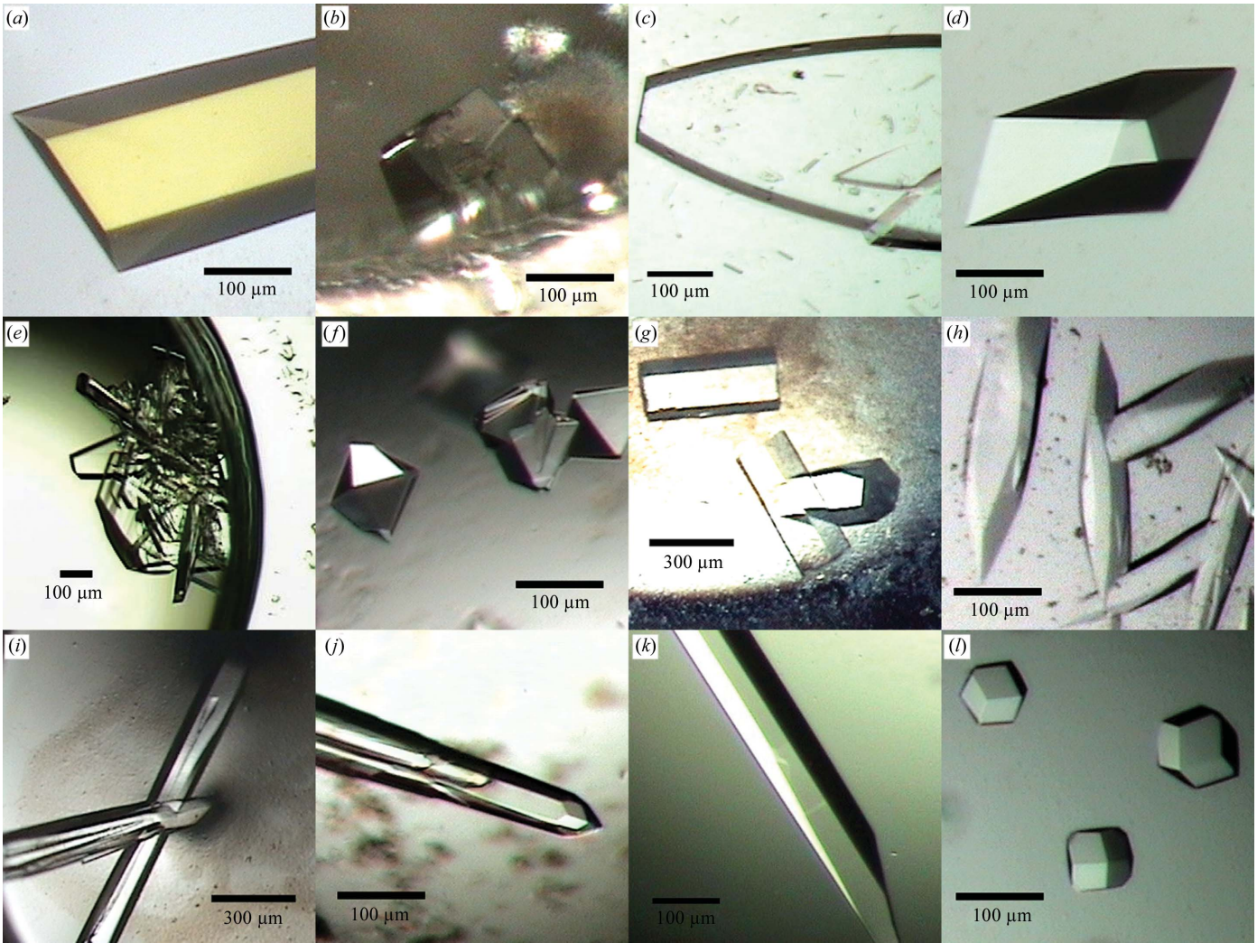

Figure 3
Twelve distinct CA_{NTD} crystal forms obtained by cocrystallization of CA_{NTD} in the presence of both compound 1 and a second CA_{NTD} ligand as described in §§2.2 and 2.3. Scale bars are indicated for reference.

different crystal-packing arrangements. Until the discovery of compound 1 as a crystallization tool for use in ternary cocrystallizations, this same conformational variability often confounded cocrystallization efforts by lowering the propensity for the complexes to crystallize and also by frequently rendering the hard-won cocrystallization conditions non-transferable, even to closely related compounds.

The wealth of tertiary CA_{NTD} structures resulting from this method provides significant insight as to how the binding of compound 1 improves the propensity of the ternary complexes to crystallize. Given the position of the compound 1 binding site, it is conceivable that the mobility of the large cyclophilin A binding loop is reduced by the binding of compound 1 (Fig. 2*a*). Reducing conformational variability is a well established strategy for improving crystallization, and indeed others have achieved impressive crystallization results using a construct in which the cyclophilin A binding loop was deleted (Blair *et al.*, 2010). However, an analysis of the numerous CA_{NTD} ternary structures does not strongly support this hypothesis. While our data cannot rule out the possibility that the cyclophilin A binding loop is less mobile in the presence of compound 1, the fact that the loop remains unresolved in the majority of our ternary CA_{NTD} structures indicates that this loop still remains quite mobile in the presence of bound compound 1.

A more convincing theory emerges from an analysis of the crystal contacts present among the 12 different crystal forms. While isolated CA_{NTD} is a nonsymmetric monomer in solution (Gitti *et al.*, 1996), analysis of the crystallographic data shows that compound 1 in complex with CA_{NTD} forms a synthetic compound 1-mediated dimer, as illustrated in Fig. 4(*a*), in 11 of the 12 different CA_{NTD} crystal forms shown in Table 2 and Fig. 3. The intermolecular contacts giving rise to this synthetic dimer are almost entirely mediated by compound 1 (Fig. 4*b*). The compound primarily makes contact with the N-terminus of helix 2, forming hydrophobic interactions with Pro34,

Table 2
Crystallographic parameters of the 12 crystal forms illustrated in Fig. 3.

Forms *f* and *g* were obtained as tertiary complexes with compounds 2 and 3, respectively. The number of compound 1-mediated dimers per crystallographic asymmetric unit (ASU) is indicated in the last column; crystal form *d* was the only form that did not contain such a dimer.

		Unit-cell parameters						
	Space group	*a* (Å)	*b* (Å)	*c* (Å)	*α* (°)	*β* (°)	*γ* (°)	Dimers per ASU
a	*P*1	53	83	89	117	106	93	4
b	$P2_1$	40	84	44	90	103	90	1
c	$P2_1$	49	74	50	90	112	90	1
d	$P2_12_12_1$	40	42	85	90	90	90	0
e	$P2_12_12_1$	71	126	129	90	90	90	3
f	$P4_12_12$	82	82	90	90	90	90	1
g	$P3_221$	66	66	143	90	90	120	1
h	$P3_221$	71	71	261	90	90	120	2
i	$P6_1$	81	81	91	90	90	120	1
j	$P6_1$	94	94	75	90	90	120	1
k	$P6_122$	73	73	125	90	90	120	$\frac{1}{2}$
l	*I*23	144	144	144	90	90	90	1

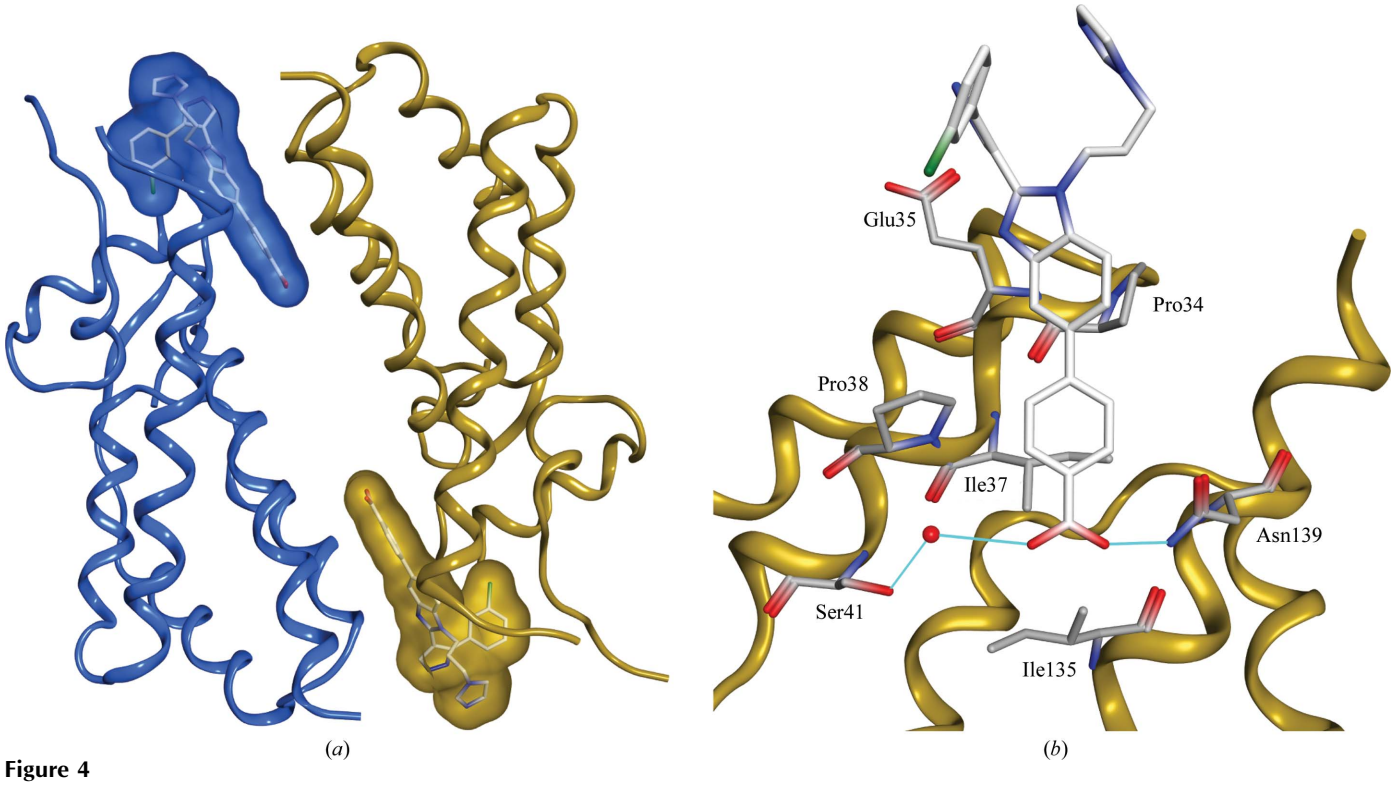

Figure 4
Compound 1-mediated dimerization of CA_{NTD}. (*a*) Ribbon diagram of two CA_{NTD} molecules dimerized by two molecules of compound 1 (rendered as semi-transparent surfaces showing compound 1 in stick form, with each unit of the dimer represented in either blue or yellow). (*b*) Intermolecular interactions between bound compound 1 and the second CA_{NTD} molecule of the dimer. Compound 1 (white sticks) from the blue monomer in Fig. 4(*a*) is shown interacting with residues (grey sticks) of the yellow monomer.

Glu35, Ile37 and Pro38, while its benzoic acid moiety forms a direct hydrogen bond to Asn139 and a water-mediated hydrogen bond to Ser41. In contrast, intermolecular protein–protein interactions stabilizing the dimer are limited to slight hydrophobic interactions made by the side chains of Arg132 and Ile135 and a potentially complementary electrostatic interaction between Arg143 and the negative dipole at the C-terminus of helix 4. Given the far greater proportion of compound-mediated intermolecular interactions, it is quite clear that this dimer would not exist in the absence of compound 1.

The 11 compound-1-mediated dimer-containing crystal forms are comprised of a total of 17 independent copies of the dimer. In most cases the crystallographic asymmetric unit was made up of integer multiples of this dimer; in only one crystal form did the symmetry axis of the dimer coincide with a crystallographic twofold (form *k* in Fig. 3 and Table 2). Nevertheless, in all cases the dimer was essentially symmetric (noncrystallographic symmetry was not employed in any of the refinements). An all-residue C^{α}-based superposition of the 17 dimers clearly illustrates the conservation of the compound-1-mediated dimer (Supplementary Fig. S1[1]), yielding pairwise r.m.s.d. values ranging from 0.38 to 1.63 Å and an overall average r.m.s.d. value of 0.80 Å (Supplementary Fig. S2). The conservation of this dimer across multiple crystal forms is a strong indication of its importance for crystallization. In support of this, while cross-seeding often proved important for crystallization, the resulting crystals did not always belong to the same space group as the seeds, again suggesting that the common element, the compound-mediated dimer, was important for nucleating crystallization.

It has been proposed that by virtue of their inherent rotational symmetry, homo-oligomeric macromolecules such as the compound-mediated CA_{NTD} dimer reported here can be viewed as existing in an advanced position along the path to a crystalline state and therefore can crystallize more readily than nonsymmetric molecules (Banatao *et al.*, 2006). More quantitatively, it has been shown that such inherently symmetric macromolecules need to participate in fewer distinct fortuitous crystal contacts to form a connected network in three dimensions (Wukovitz & Yeates, 1995). The ability to improve crystallization by the engineering of symmetrized homo-oligomers was first experimentally demonstrated using disulfide-stabilized homodimers of lysozyme (Banatao *et al.*, 2006; Forse *et al.*, 2011) and subsequently with leucine-zipper-stabilized homodimers of human pancreatic ribonuclease (Yamada *et al.*, 2007). In both studies the symmetrization approach led to new crystal forms that could not be obtained in the unmodified monomeric forms. We therefore suggest that with this work we have extended this principle to compound-mediated symmetrization, which has led to the greatly improved crystallization of CA_{NTD}.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: NJ5150). Services for accessing this material are described at the back of the journal.

4. Conclusions

HIV continues to be a global problem and innovative viral targets such as CA may well provide the next generation of therapeutically important HIV antivirals. The novel CA assembly-inhibiting site identified by compound 1 is the third distinct inhibitor-binding site reported for CA_{NTD} and at least the fifth for CA as a whole. This ever-increasing number of CA inhibitory sites again underscores the validity and potential utility of this target.

Distinct from their potential antiviral role, the compound 1 series of inhibitors proved to be extremely useful crystallization tools that greatly increase the crystallizability of CA_{NTD}; a very beneficial finding for this otherwise unruly crystallization target. We hypothesize that this increased propensity to crystallize stems from the formation of a compound-mediated intrinsically symmetrical homodimer that greatly improves the prospect of generating a crystal lattice. In this work, the symmetric homodimer was fortuitously created by nonhomotypic binding of an asymmetric ligand. It is intriguing to consider the potential of symmetric molecules specifically designed for use as generic additives during crystallization. In any case, while protein-engineering experiments have been used in the literature to help to establish the link between inherent macromolecular symmetry and crystallizability, to our knowledge this work represents the first use of a synthetic ligand for this purpose and in our opinion provides perhaps the most compelling experimental evidence to date linking the inherent symmetry of a molecule to its propensity to crystallize.

The authors would like to acknowledge Patrick Deroy, Lee Fader, Jean-Eric Lacoste and Chantal Grand-Maître for chemical syntheses, Jean-Francois Mercier and Elizabeth Wardrop for activity assays, Mireille Cartier for bioinformatics analyses and Axel Brilot for laboratory assistance. We would also like to thank Michael Bös, Paul Anderson, Richard Bethell and Michael Cordingley for leadership and guidance.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Banatao, D. R., Cascio, D., Crowley, C. S., Fleissner, M. R., Tienson, H. L. & Yeates, T. O. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 16230–16235.
Berthet-Colominas, C., Monaco, S., Novelli, A., Sibaï, G., Mallet, F. & Cusack, S. (1999). *EMBO J.* **18**, 1124–1136.
Blair, W. S. *et al.* (2010). *PLoS Pathog.* **6**, e1001220.
Bocanegra, R., Rodríguez-Huete, A., Fuertes, M. Á, Del Álamo, M. & Mateu, M. G. (2012). *Virus Res.* **169**, 388–410.
Broder, S. (2010). *Antiviral Res.* **85**, 1–18.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Forse, G. J., Ram, N., Banatao, D. R., Cascio, D., Sawaya, M. R., Klock, H. E., Lesley, S. A. & Yeates, T. O. (2011). *Protein Sci.* **20**, 168–178.
Forshey, B. M., von Schwedler, U., Sundquist, W. I. & Aiken, C. (2002). *J. Virol.* **76**, 5667–5677.
Gamble, T. R., Vajdos, F. F., Yoo, S., Worthylake, D. K., Houseweart, M., Sundquist, W. I. & Hill, C. P. (1996). *Cell*, **87**, 1285–1294.

Ganser-Pornillos, B. K., von Schwedler, U. K., Stray, K. M., Aiken, C. & Sundquist, W. I. (2004). *J. Virol.* **78**, 2545–2552.
Ganser-Pornillos, B. K., Yeager, M. & Pornillos, O. (2012). *Adv. Exp. Med. Biol.* **726**, 441–465.
Gitti, R. K., Lee, B. M., Walker, J., Summers, M. F., Yoo, S. & Sundquist, W. I. (1996). *Science*, **273**, 231–235.
Goudreau, N., Lemke, C. T., Faucher, A.-M., Grand-Maître, C., Goulet, A., Lacoste, J.-E., Rancourt, J., Malenfant, E., Mercier, J.-F., Titolo, S. & Mason, S. W. (2013). *ACS Chem. Biol.*, doi:10.1021/cb400075f.
Howard, B. R., Vajdos, F. F., Li, S., Sundquist, W. I. & Hill, C. P. (2003). *Nature Struct. Biol.* **10**, 475–481.
Kelly, B. N., Kyere, S., Kinde, I., Tang, C., Howard, B. R., Robinson, H., Sundquist, W. I., Summers, M. F. & Hill, C. P. (2007). *J. Mol. Biol.* **373**, 355–366.
Lemke, C. T. *et al.* (2012). *J. Virol.* **86**, 6643–6655.
McManus, H., O'Connor, C. C., Boyd, M., Broom, J., Russell, D., Watson, K., Roth, N., Read, P. J., Petoumenos, K. & Law, M. G. (2012). *PLoS One*, **7**, e48839.
Momany, C., Kovari, L. C., Prongay, A. J., Keller, W., Gitti, R. K., Lee, B. M., Gorbalenya, A. E., Tong, L., McClure, J., Ehrlich, L. S., Summers, M. F., Carter, C. & Rossmann, M. G. (1996). *Nature Struct. Biol.* **3**, 763–770.
Moreno, S., Lopez, A. J., Arribas, J. R., Domingo, P., Iribarren, J. A., Ribera, E., Rivero, A. & Pulido, F. (2010). *J. Antimicrob. Chemother.* **65**, 827–835.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pornillos, O., Ganser-Pornillos, B. K., Banumathi, S., Hua, Y. & Yeager, M. (2010). *J. Mol. Biol.* **401**, 985–995.
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.
Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. (2011). *Nature (London)*, **469**, 424–427.
Prevelige, P. E. Jr (2011). *J. Mol. Biol.* **410**, 634–640.
Schwedler, U. K. von, Stray, K. M., Garrus, J. E. & Sundquist, W. I. (2003). *J. Virol.* **77**, 5439–5450.
Sundquist, W. I. & Krausslich, H. G. (2012). *Cold Spring Harb. Perspect. Med.* **2**, a006924.
Taiwo, B., Hicks, C. & Eron, J. (2010). *J. Antimicrob. Chemother.* **65**, 1100–1107.
Tang, C., Loeliger, E., Kinde, I., Kyere, S., Mayo, K., Barklis, E., Sun, Y., Huang, M. & Summers, M. F. (2003). *J. Mol. Biol.* **327**, 1013–1020.
Tang, C., Ndassa, Y. & Summers, M. F. (2002). *Nature Struct. Biol.* **9**, 537–543.
Waheed, A. A. & Freed, E. O. (2012). *AIDS Res. Hum. Retroviruses*, **28**, 54–75.
Wong, H. C., Shin, R. & Krishna, N. R. (2008). *Biochemistry*, **47**, 2289–2297.
Wukovitz, S. W. & Yeates, T. O. (1995). *Nature Struct. Biol.* **2**, 1062–1067.
Yamada, H. *et al.* (2007). *Protein Sci.* **16**, 1389–1397.