CrossMark

electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449
Editors: **E. N. Baker and Z. Dauter**

Structure of the oncoprotein Rcl bound to three nucleotide analogues

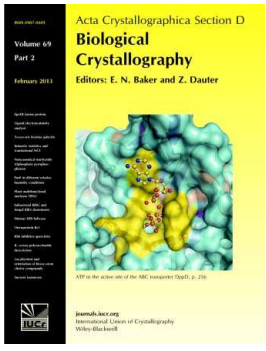
André Padilla, Claire Amiable, Sylvie Pochet, Pierre-Alexandre Kaminski and Gilles Labesse

Acta Cryst. (2013). D**69**, 247–255

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see http://journals.iucr.org/services/authorrights.html



Acta Crystallographica Section D: Biological Crystallography welcomes the submission of papers covering any aspect of structural biology, with a particular emphasis on the structures of biological macromolecules and the methods used to determine them. Reports on new protein structures are particularly encouraged, as are structure–function papers that could include crystallographic binding studies, or structural analysis of mutants or other modified forms of a known protein structure. The key criterion is that such papers should present new insights into biology, chemistry or structure. Papers on crystallographic methods should be oriented towards biological crystallography, and may include new approaches to any aspect of structure determination or analysis. Papers on the crystallization of biological molecules will be accepted providing that these focus on new methods or other features that are of general importance or applicability.

Crystallography Journals Online is available from **journals.iucr.org**

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the oncoprotein Rcl bound to three nucleotide analogues

André Padilla,[a,b] Claire Amiable,[c,d,e] Sylvie Pochet,[c,d] Pierre-Alexandre Kaminski[c,d] and Gilles Labesse[a,b]*

[a]CNRS, UMR5048, Université Montpellier 1 et 2, Centre de Biochimie Structurale, F-34090 Montpellier, France, [b]INSERM, U1054, 29 Rue de Navacelles, F-34090 Montpellier, France, [c]Unité de Chimie et Biocatalyse, Institut Pasteur, F-75015 Paris, France, [d]CNRS, UMR3523, F-75015 Paris, France, and [e]Université Paris Descartes Sorbonne Paris Cité, Institut Pasteur, F-75015 Paris, France

Correspondence e-mail: labesse@cbs.cnrs.fr

Received 25 July 2012
Accepted 30 October 2012

PDB References: Rcl, complex with phosphotriciribine, 4fyh; complex with 6-cyclopentyl-AMP, 4fyi; complex with adenosine 5′-phosphorothioate, 4fyk

Rcl is a novel N-glycoside hydrolase found in mammals that shows specificity for the hydrolysis of 5′-monophosphate nucleotides. Its role in nucleotide catabolism and the resulting production of 2-deoxyribose 5-phosphate has suggested that it might fuel cancer growth. Its expression is regulated by c-Myc, but its role as an oncoprotein remains to be clarified. In parallel, various nucleosides have been shown to acquire pro-apoptotic properties upon 5′-monophosphorylation in cells. These include triciribine, a tricyclic nucleoside analogue that is currently in clinical trials in combination with a farnesyltransferase inhibitor. Similarly, an N^6-alkyl-AMP has been shown to be cytotoxic. Interestingly, Rcl has been shown to be inhibited by such compounds *in vitro*. In order to gain better insight into the precise ligand-recognition determinants, the crystallization of Rcl with these nucleotide analogues was attempted. The first crystal structure of Rcl was solved by molecular replacement using its NMR structure in combination with distantly related crystal structures. The structures of Rcl bound to two other nucleotides were then solved by molecular replacement using the previous crystal structure as a template. The resulting structures, solved at high resolution, led to a clear characterization of the protein–ligand interactions that will guide further rational drug design.

1. Introduction

Rcl is a putative oncoprotein and as such it may constitute a new and attractive target for anticancer therapies. Its gene is a c-Myc target and it has been shown to be overexpressed in various cancers (Lewis *et al.*, 2000). Rcl is responsible for nucleotide hydrolysis and it produces free nucleobase and 5′-phosphodeoxyribose, which is a strong elicitor of angiogenesis (Ghiorghi *et al.*, 2007). This may fuel cancer growth while detoxifying the rapidly growing cells. However, a stronger and more direct demonstration of the role of Rcl in cancer development is still lacking. The design of potent inhibitors may help in validating this hypothesis while opening the road to mechanistic insights as well as potential future treatments.

Although the solution structure of Rcl is known (Yang *et al.*, 2009), a more detailed view of the interactions between the protein and its ligands is necessary for efficient drug design. Accordingly, we engaged in the determination of the crystal structure of Rcl bound to purine monophosphate derivatives (6-cyclopentyl-AMP, triciribine 5′-monophosphate and adenosine 5′-phosphorothioate).

The first crystal structure of rat Rcl was solved by molecular replacement using a template derived from the NMR structure of Rcl and distantly related proteins for which crystal structures are known but that show only low levels of sequence

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved

electronic reprint

identity (~20%). Solution of the first crystal structure of Rcl opened the road to the rapid determination of structures of this protein bound to two other ligands. Here, we describe the high-resolution structure of Rcl in complex with two chemical compounds of high biological relevance, 6-cyclopentyl-AMP (hereafter referred to as C6P) and phosphotriciribine (hereafter referred to as TCN-P), as well as adenosine 5′-phosphorothioate (hereafter referred to as AMPS). C6P has been shown to induce apoptosis through an unknown mechanism (Mlejnek & Dolezel, 2005), while TCN-P is the monophosphorylated form of TCN, a drug candidate currently in clinical trials in combination with a farnesyltransferase inhibitor (Balasis *et al.*, 2011). Recently, direct interactions of TCN-P with the PH domain of Akt have been demonstrated and have been shown to induce the relocalization of Akt, a protein kinase involved in cell survival (Berndt *et al.*, 2010). This may explain the impact of TCN-P on cancer growth, but our results suggest that Rcl could be another relevant target *in vivo*. Finally, another crystal structure was solved in the presence of the substrate mimic AMPS.

These structures highlight subtle changes in the active site that cannot be easily detected by NMR. Conversely, NMR and X-ray crystallography agree on the important flexibility of the long and central loop (residues 47–59). We discuss the impact of the high-resolution structure on the analysis of the proposed reaction mechanism as well as the design of better inhibitors.

2. Experimental procedures

Initial crystallization trials were performed using a Tecan Genesis crystallization robot equipped with eight independent needles and a liquid-detection system. For crystallization, 96-well plates were used with the Anions, pHClear and AmSO4 Suites from Qiagen. The wells were filled with 75 µl crystallization buffer, and 0.5 µl crystallization buffer was mixed with 0.5 µl of the protein sample (at 22 mg ml^{-1} in 50 m*M* phosphate buffer pH 7.2 and 4 m*M* GMP) prior to deposition at the drop location. The plate was then sealed with transparent plastic film (Greiner G-676070) and kept at 291 K. Several hits were obtained but with poor crystal quality; those corresponding to high concentrations of salts (1.6 *M* ammonium sulfate, HEPES pH 7 and Tris pH 8) were further refined manually using the hanging-drop method in 24-well plates. Firstly, we successfully tested replacement of the phosphate buffer by citrate buffer. The protein was dialyzed against 50 m*M* citrate buffer pH 7.3, concentrated in a Centricon and mixed at 12 mg ml^{-1} with 4 m*M* GMP. In the screen, the precipitant (ammonium sulfate) concentration varied from 1.0 to 2.0 *M* and the pH varied from 7.0 to 8.5. In this case, the wells were filled with 0.5 ml crystallization buffer, and 1 µl crystallization buffer was mixed with 1 µl protein sample prior to deposition at the drop location. Starting from selected conditions (those with the best crystal quality), further refinement was performed with 1.1, 1.2 and 1.3 *M* ammonium sulfate and 1.2, 1.3 and 1.4 *M* lithium sulfate in mild basic buffer (100 m*M* Tris pH 7.4–8.6) with 20 m*M* of one of the following additives: ethanolamine, magnesium sulfate, 1,6-hexanediol or 1,2,3-heptanetriol. These screening conditions were used for cocrystallization assays with various ligands including GMP, 6-cyclopentyl-AMP (C6P), triciribin phosphate (TCN-P) and adenosine 5′-phosphorothioate (AMPS). The crystals were cooled in liquid nitrogen and those obtained in the presence of C6P, TCN-P or AMPS diffracted to up to 1.6 Å resolution on the synchrotron beamlines at the ESRF, Grenoble.

Diffraction data were collected on beamlines ID14-4, ID14-2 and ID29 at the ESRF, integrated using *MOSFLM* (Leslie, 2006) and merged using *SCALA* from the *CCP*4 suite (Winn *et al.*, 2011). The reciprocal-space parameters and the systematic absences corresponded to space group $P2_12_12$. Molecular replacement was performed using *MOLREP* (Vagin & Teplyakov, 2010). Rigid-body and restrained refinement were performed using *REFMAC*5 (Murshudov *et al.*, 2011) with alternating manual rebuilding in *Coot* (Emsley & Cowtan, 2004).

3. Results

3.1. An engineered variant of Rcl

Previous NMR studies of rat Rcl in complex with GMP (Yang *et al.*, 2009) and in the apo form (Doddapaneni *et al.*, 2009) revealed a well folded core and several regions with poorly defined conformations. The flexible regions mainly include the N- and C-terminal tails and a 12-residue-long loop between β2 and α2 in the central portion of the protein (amino acids 47–59 in rat Rcl). Sequence comparison of close orthologues from mammals showed rapid divergence of both terminal tails, while the internal disordered region was rather well conserved. This is in agreement with the positions of these segments relative to the active site. Indeed, both protein ends are far away from the ligand-binding site, while the flexible loop 47–59 lies at the entrance to the active-site cavity. This suggested that we could remove the flexible tails but not the central loop in order to engineer a variant of a mammalian Rcl useful for crystallization.

At the same time, it has also been shown by NMR and SAXS that unbound Rcl is in an equilibrium between a monomeric and a dimeric form (Yang *et al.*, 2009). This behaviour is likely to hinder crystallization of Rcl. According to SAXS measurements, Rcl forms a stable dimer upon the addition of GMP. However, NMR studies showed slow hydrolysis (on an hour time scale) of GMP and dissociation of the protein dimer (Yang *et al.*, 2009). These results could explain the failure to grow useful crystals of Rcl in previous attempts.

This information prompted us to (i) use a shortened rat Rcl and (ii) design inactive mutants of this variant to attempt crystallization. The N-terminal sequence MAASGEQAPC was changed to MRR in order to remove the flexible segment and also to remove the cysteine to avoid the formation of unwanted interchain disulfide bridges over time. Similarly, the C-terminal sequence PQKTASSSHPSA was removed and replaced by the polyhistidine tag VEHHHHHH (see Fig. 1).

electronic reprint

Subsequently, the aspartate involved in catalysis (Asp69 in rat Rcl) was substituted by either an asparagine (a conservative substitution) or an alanine. Both mutants behaved like the full-length construct in solution but with increased stability of the dimeric form over several weeks in the presence of GMP (Yang *et al.*, 2009).

For the crystallization trials, we decided to use the most conservative mutant (shortened rat Rcl D69N) in order to better model possible interactions with the ligands. Optimal crystal-growth conditions were searched for using 96-well plates and a nanodispenser robot. Crystals were obtained in conditions with various salts and pH values in the presence of either GMP or C6P. We were able to improve the size and the diffraction quality of the crystals by manual refinement around the starting conditions in the case of cocrystallization with C6P. Crystals grown with GMP were mostly thin needles and did not lead to satisfactory diffraction patterns. In order to rapidly explore the potential crystallization of rat Rcl with other inhibitors, a focused 48-condition screen was set up. The main precipitant was either lithium sulfate (1.2–1.4 *M*) or ammonium sulfate (1.1–1.3 *M*). This focused screen yielded crystals of Rcl in the presence of TCN-P and AMPS. The crystals of the shortened and stabilized Rcl in the presence of C6P, TCN-P or AMPS diffracted to up to 1.6 Å resolution.

3.2. Solving the Rcl structure by molecular replacement

High-resolution diffraction data were collected on beamline ID14-4, processed using *MOSFLM* (Leslie, 2006) and merged using *SCALA* from the *CCP*4 suite (Winn *et al.*, 2011) from a crystal grown at high salt concentration in the presence of the ligand C6P. The unit-cell parameters, merging statistics and systematic absences were consistent with space group $P2_12_12$. The calculated Matthews coefficient (V_M = 2.3 Å^3 Da^{-1}) suggested that four monomers were present in the asymmetric unit, possibly forming two biologically relevant dimers (with a solvent content of 46%). Knowledge of the dimeric forms observed using NMR and SAXS was expected to help in molecular replacement. However, the detection of pseudo-translation suggested potential difficulties. Indeed, the liganded NMR structure (PDB entry 2klh; Yang *et al.*, 2009) alone did not yield a solution for molecular replacement using either *MOLREP* (Vagin & Teplyakov, 2010) or *Phaser* (McCoy *et al.*, 2007), which is in agreement with the general trend for medium-resolution NMR structures (Mao *et al.*, 2011; DiMaio *et al.*, 2011). In parallel, fold-recognition was performed to identify possible templates for comparative modelling using *@TOME*-2 (Pons & Labesse, 2009). Owing to the low level of sequence identity (∼20%) with known but distantly related crystal structures (PDB entries 1f8x, 2f62 and 3ehd; Armstrong *et al.*, 1996; Bosch *et al.*, 2006; Midwest Center for Structural Genomics, unpublished work), no molecular-replacement solution was detected using dimeric models derived from models of Rcl based on these structures.

The known NMR structure of Rcl was used to guide a realignment of its sequence with those of distantly related proteins for which crystal structures are known. The manual refinement was performed using the software *VITO* (Catherinot & Labesse, 2004) in order to connect sequence alignment and three-dimensional structure while surveying the common

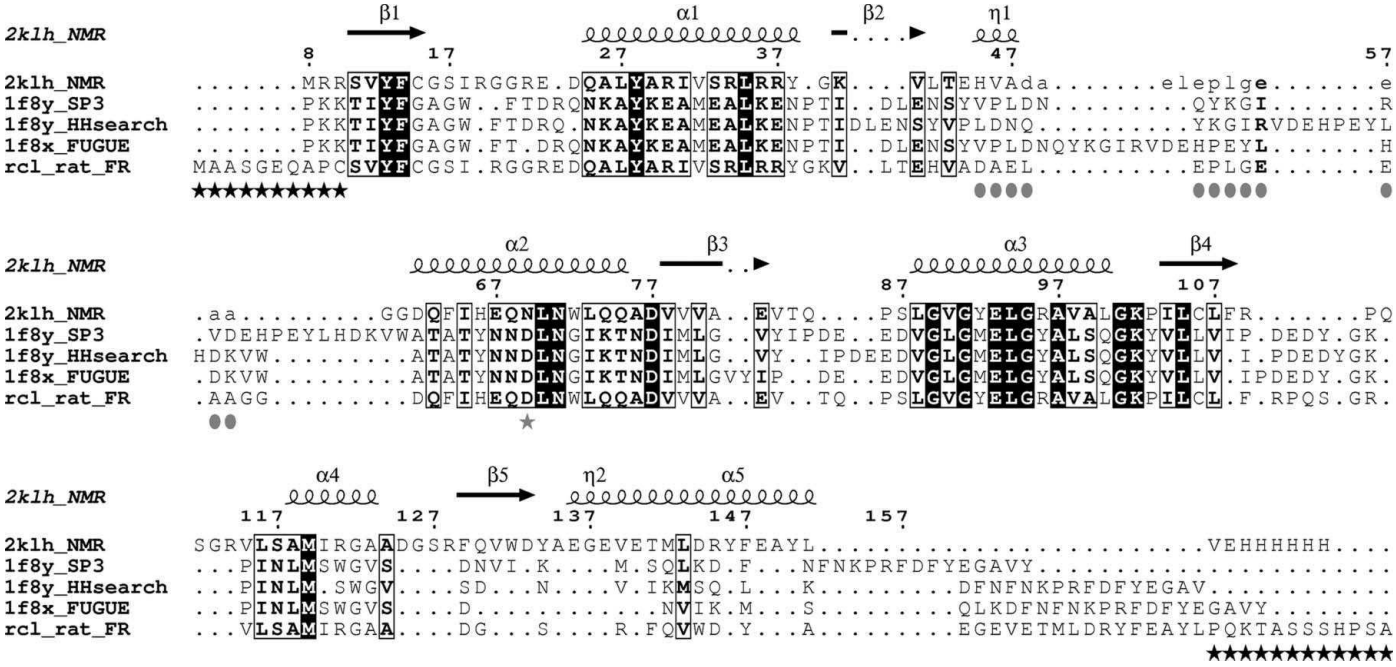


Figure 1
Sequence–structure alignment. The structural alignment of the NMR structure of Rcl and the sequence alignment deduced from fold-recognition using *@TOME*-2 (Pons & Labesse, 2009). Secondary structures in Rcl are shown above the sequence alignment. Well conserved residues are boxed and strictly conserved residues are shown in white on black. The black stars and grey circles indicate the positions of the residues removed in the recombinant variant of Rcl and those removed in the model for molecular replacement, respectively. This figure was generated by the *ESPript* server (Gouet *et al.*, 1999).

electronic reprint

Table 1
Data-collection, phasing and refinement statistics for the Rcl structures.

Values in parentheses are for the outer shell.

	Rcl–C6P	Rcl–TCN-P	Rcl–AMPS
Beamline	ID14-4	ID14-2	ID29
No. of crystals	1	1	1
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit-cell parameters			
a (Å)	91.6	92.7	86.4
b (Å)	119.0	118.7	128.8
c (Å)	57.5	57.7	57.5
No. of molecules in asymmetric unit	4	4	4
Wavelength (Å)	0.97950	0.93340	0.97625
Resolution (Å)	1.96 (2.01–1.96)	2.44 (2.50–2.44)	1.79 (1.84–1.79)
$R_{\rm merge}$† (%)	4.0 (22.9)	8.5 (37.6)	6.4 (13.3)
$\langle I/\sigma(I)\rangle$	19.4 (5.3)	8.0 (1.9)	10.3 (4.5)
Completeness (%)	97.4 (96.8)	100.0 (100.0)	92.7 (68.1)
Multiplicity	8.0 (7.8)	7.2 (7.3)	3.2 (2.4)
Wilson *B* factor (Å^2)	24.2	38.1	16.1
Refinement			
Resolution (Å)	42.5–1.96	38.9–2.60	35.9–1.79
No. of reflections	42198	23302	55536
$R_{\rm work}/R_{\rm free}$‡ (%)	18.1/21.7	20.9/24.6	19.6/21.9
No. of atoms			
Protein	4790	4153	4790
Ligand	4	4	4
Ions	—	5	3
Waters	404	123	626
B factors (Å^2)			
Protein	24.1	26.9	15.8
Ligand	20.2	17.6	9.9
Ions	—	58.9	44.1
Waters	35.5	28.8	35.7
R.m.s. deviations§			
Bond lengths (Å)	0.008	0.010	0.009
Bond angles (°)	1.3	1.1	1.1

† $\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl) \times 100$. ‡ $R_{\rm free}$ was calculated using a subset of reflections that were not used in the refinement (5%). § Deviations from ideal values.

dimeric interface. The main goal was to define a structural core common to both the NMR structures and the distantly related crystal structures in order to better define a useful template for molecular replacement. While minor changes were made to short segments mostly corresponding to small loops in the N-terminal and central parts of the structural alignment, some important changes were performed in the region corresponding to the last 50 residues (see Fig. 1). This part included the last β-strand (β5) and the very C-terminal α-helix (α5). This segment is highly variable in sequence, although it is partially constrained by a common dimeric interface observed in all of the distantly related structures (*e.g.* PDB entries 3ehd and 1f8x) as previously described by Yang *et al.* (2009). Indeed, fold-recognition provided us with a structural alignment that varies greatly in this region, with only *FUGUE* (Shi *et al.*, 2001) correctly aligning strand β5, while helix α5 was not correctly aligned. The revised sequence–structure alignment was used to derive new models of Rcl based on the conserved parts shared with the related crystal structures. The final models were built using *SCWRL*4.0 (Krivov *et al.*, 2009) for each monomer and were subsequently combined as dimers. Finally, the long and mobile loop (amino acids 47–59) as well as the C-terminal histidine tag were omitted from the model.

We then searched for two dimers using the program *MOLREP*. A weak solution was obtained (contrast 2.1; *R* factor 63.1). Following ten steps of rigid-body refinement in *REFMAC*5 ($R_{\rm work}$ = 0.543, $R_{\rm free}$ = 0.564), restrained refinement (with NCS restraints) led to smooth minimization and improvement of the working and free *R* factors (to 0.503 and 0.558, respectively). The electron density for the bound nucleotide was partially visible although it was omitted from the search model. Additional steps of restrained minimization were performed varying the high-resolution limit (from 2.96 to 1.96 Å) and subsequently lowering the NCS restraints (from tight to medium). This led to a significant decrease in the working and free *R* factors (to 0.443 and 0.517, respectively). Then, thanks to the recording of high-resolution data and the known NMR structure, complete rebuilding and straightforward refinement could be performed. Alternating manual rebuilding in *Coot* (Emsley & Cowtan, 2004) and restrained refinement with looser NCS restraints in *REFMAC*5 (Murshudov *et al.*, 2011) led to a satisfactory and almost complete model. The overall improvement can be witnessed in the C-terminal region (see Figs. 2*a* and 2*b*), in which manual alignment refinement was important for successful molecular replacement and easier model rebuilding. Finally, restrained refinement with *TLS* (Winn *et al.*, 2001) with no NCS restraints resulted in a high-quality structure ($R_{\rm work}$ = 0.188, $R_{\rm free}$ = 0.230; see Table 1).

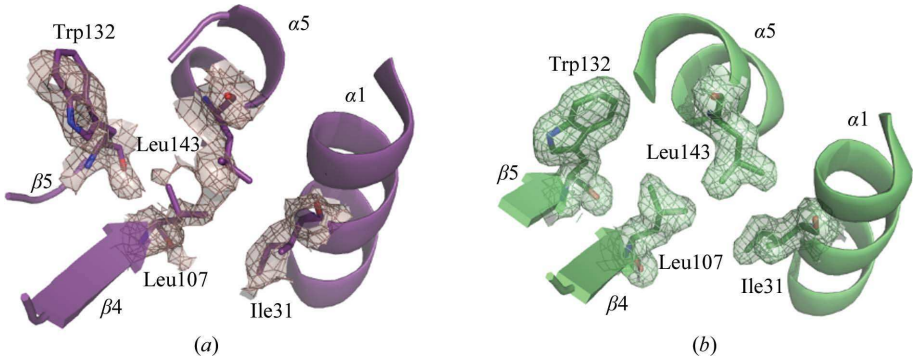


Figure 2
Comparison of the electron densities before and after refinement. (*a*) Local view of the partial model used for molecular replacement and the corresponding electron-density map drawn around the side chains of Ile32, Leu107, Leu143 and Trp132. This region includes segments that were correctly aligned automatically (*e.g.* helix α1) and segments that required manual refinement (*e.g.* helix α5). (*b*) The same region shown after final refinement with a much better match of the model to the electron density. The figures were generated using *PyMOL* (http://www.pymol.org).

3.3. Comparison of the crystal and the solution structures

Although the structures of Rcl solved by NMR (PDB entries 2klh and 2khz; Doddapaneni *et al.*, 2009) and X-ray diffraction are globally very similar (Fig. 3*a*), the overall r.m.s.d. was quite high

electronic reprint

(1.9 Å with PDB entry 2klh and 1.5 Å with PDB entry 2khz over 125 residues forming the common core of a monomer). The r.m.s.d. (computed in *Coot*) for the dimer increases to 2.3 and 1.7 Å with the solution structures (PDB entries 2klh and 2khz, respectively) and to 2.0 Å (over 208 C^{α} atoms) with the crystal structure of the template (PDB entry 1f8y; Fig. 3*b*; Armstrong *et al.*, 1996). The dimerization interface and central β-sheet showed little difference, while two α-helices (α1 and α5, both of which are solvent-exposed) showed significant tilts or shifts (a translation of up to 3 Å).

In agreement with the previously described structures of Rcl in solution, the long loop (47–59) appeared to be disordered, with only weak electron density visible in the vicinity of the bound ligand. However, the electron density was too weak to allow proper modelling of the missing residues. Similarly, the first two residues in the crystallized construct and the very C-terminus (residues 149–154 and the polyhistidine tag) appeared to be flexible and were not modelled.

Further analysis indicated that differences appear in many details. Firstly, the interactions between the protein and ligand are better defined (see below). Indeed, the precise orientations of several side chains in the active site are much better defined (*e.g.* Arg19), although one should keep in mind that the crystal lattice and the crystal-growth conditions (high salt concentration) might locally affect the protein structure. In addition, many water molecules are also clearly observed in the electron density, including in the vicinity of the bound nucleotide.

3.4. A conserved catalytic mechanism in NDTs and Rcl

A search for similar structures using *FATCAT* (Ye & Godzik, 2004) identified all of the NDTs solved to date, with an r.m.s.d. (for the monomer) of around 2.5 Å, and also the so-called receiver domain, with a higher r.m.s.d. (above 3.0 Å). The latter matches correspond to conservation of the Rossmann fold, but no similarities in the active site can be found and local rearrangements in the receiver domain close the ligand-binding pocket found in NDT and Rcl. In parallel, the global structural divergence of Rcl and NDT is correlated with high sequence variation, while the catalytic centre seems to be perfectly conserved. Nevertheless, significant rearrangements occurred, in agreement with the shift in substrate specificity (5′-monophosphorylated nucleotides for Rcl *versus* nucleosides for NDTs). This change in function relies mainly on the build-up of a phosphate-accommodating loop in Rcl. It is composed of loop β1–α1, which is longer in Rcl and adopts a distinct conformation compared with the equivalent loops in NDTs. In addition, several residue substitutions leave room for the bulky phosphate group and simultaneously remove a negative charge (see below).

In order to gain a better insight into the mechanism of catalysis by Rcl, we focused our structure comparison on the catalytic centre of Rcl and the NDTs. The catalytic residues of Rcl (Tyr13, Asp69 and Glu93 in rat Rcl) have previously been identified using directed mutagenesis and sequence comparison among mammalian Rcls and also with known NDTs (Ghiorghi *et al.*, 2007; Dupouy *et al.*, 2010). The catalytic triad in the crystal structure of Rcl adopts nearly the same orientation despite the mutation of the aspartate Asp69 to asparagine. In all three Rcl structures solved to date, these three residues occupied positions nearly identical to those observed for the catalytic residues in the NDT structures (⟨r.m.s.d.⟩ = 0.45 Å). For both NDTs (Armstrong *et al.*, 1996; Anand *et al.*, 2004) and Rcl (this study), the crystal structures show that

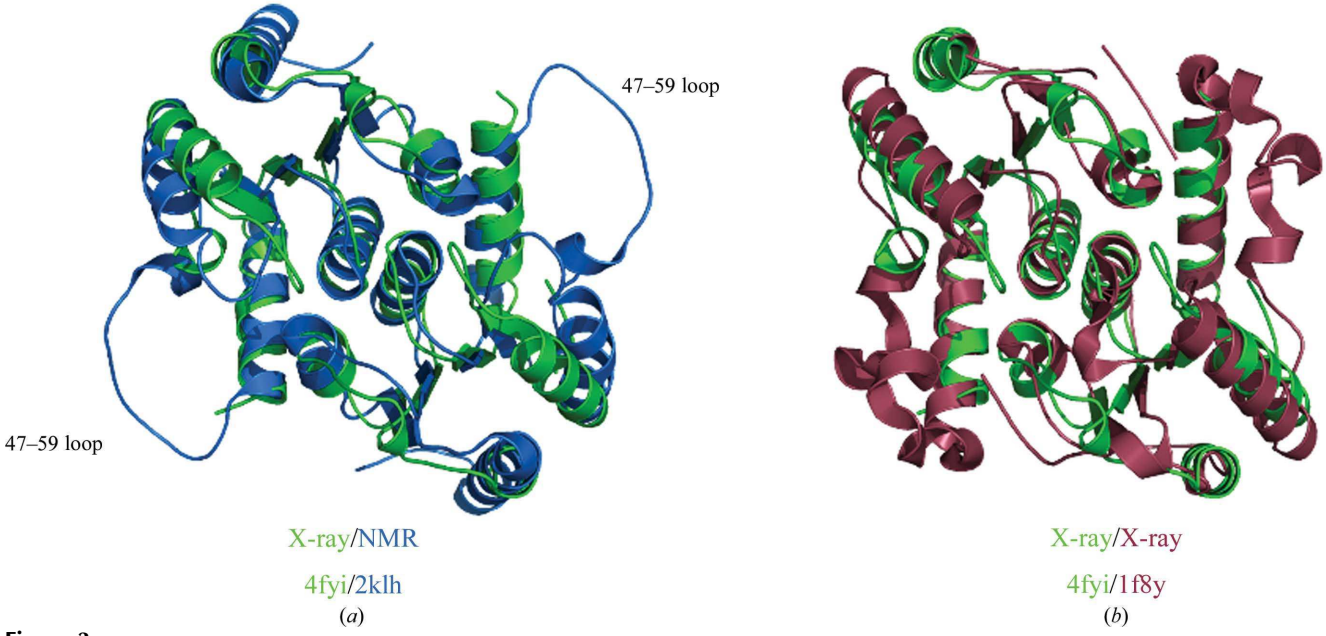


Figure 3
Comparison of solution and crystal structures. (*a*) The crystal structure (green ribbon) of rat Rcl in the dimer (this study) is shown superposed on its solution structure determined in the presence of GMP (PDB entry 2klh; blue). (*b*) Superimposition of rat Rcl (green ribbon) with a crystal structure of a distantly related nucleoside 2-deoxyribosyltransferase (PDB entry 1f8y; violet). In all cases, ligands were omitted for clarity. The figures were generated using *PyMOL* (http://www.pymol.org).

electronic reprint

these three residues form an intricate network of hydrogen bonds that also involves the bound ligand (if present). The perfect geometric conservation of this triad suggested that these enzymes share a similar mechanism of action. In NDT from *Lactobacillus leichmannii*, the catalytic glutamate (Glu98) has been shown to be the active-site nucleophile (Porter *et al.*, 1995) and a ribosylated enzyme intermediate involving an equivalent glutamate (Glu101) has been trapped using 2-deoxy-2-fluoro derivatives (Anand *et al.*, 2004). In NDTs and in Rcl the carboxylate group of the catalytic glutamate is at a close distance to the C1′ atom of the ribose moiety. This suggested that this residue can also attack the C1′ atom of the ribose to form a ribosylated enzyme. As in NDTs, this glutamate interacts with a tyrosine (Tyr13) and an aspartate (Asp69) in wild-type Rcl. These two residues are supposed to stabilize the different states of the glutamate residue in NDTs and, by similarity, in Rcl. In agreement with these hypotheses, conservative mutations (Y13F, D69N or E93Q) in Rcl abolish the activity or dramatically decrease the affinity for the substrate (Dupouy *et al.*, 2010).

Interestingly, Rcl is purely a hydrolase, while the distantly related NDTs are mainly transferases. In NDT, reaction proceeds by transfer of the deoxyribose moiety from the ribosylated enzyme intermediate to an acceptor nucleobase (Short *et al.*, 1996). In Rcl, in contrast, this transfer step involves a solvent molecule, as demonstrated by a methanolysis study (Doddapaneni *et al.*, 2011). A survey of the various NDT structures in unbound, substrate-bound and ribosylated forms shows that very few water molecules can occupy the ligand-binding sites. The very few water molecules found in the apo form of NDT from *L. leichmannii* (PDB entry 1f8x) occupy the positions of the 3′-hydroxyl group of the ribose or polar atoms of the nucleobase found in the bound state (PDB entry 1f8y). In the first case, the water molecule cannot enter into the ribosylated intermediate state. In the second case, the water molecules may be too distant to attack and free the ribosyl moiety. The absence of attacking water may correlate with the rather hydrophobic nature of the ligand-binding pocket. In NDTs, beside the catalytic triad, large aromatic residues constitute the active site and it is rather closed by various loops and the C-terminus. Conversely, the active site in Rcl appears to be largely open as the loop 47–59 is disordered. Furthermore, we observed two water molecules in the vicinity of the bound ribose, while the presence of the hydrophobic side chain of a proline (Pro42 and Pro44 in the NDTs from *L. leichmannii* and *L. helveticus*, respectively) prevents solvent entrance in NDTs (Fig. 4). In addition, the preference for solvent attack over nucleobase substitution may correlate with weaker interactions between the Rcl binding site and the nucleobase (see below).

3.5. Ligand recognition in the C6P-bound Rcl

In the crystal structure of Rcl in complex with C6P, the position of the inhibitor in the active site is clear (Fig. 5*a*). The adenine, ribose and phosphate moieties are very well defined in the electron density, while the 6-cyclopentyl moiety appears to be mobile or in an alternative conformation in the four independent monomers (data not shown). The new crystal structure clarified the position of the phosphate group and the orientations of the phosphate O atoms. Previously, in the NMR structure, the phosphate group was not directly positioned but was predicted to form hydrogen bonds according to low-field chemical shifts of the backbone N atoms of three residues (Ile18, Arg19 and Ala118). The orientation of the side chain of Arg19 is also better defined in the crystal structure compared with the NMR structures and showed a surprising feature. The positively charged guanidinium group did not point toward the negatively charged phosphate group. Rather, it points towards a nearby aspartate (Asp62) to form a salt bridge (distance of 3.9 Å), forming a lid over the phosphate group (Fig. 5*d*). It also appeared to stabilize the dimer and the ligand-binding site through a hydrogen bond to the carbonyl of Leu116′. The arginine also weakly interacts with the 6-cyclopentyl group of the ligand. In parallel, the crystal structure revealed an intricate network of seven hydrogen bonds formed between the phosphate group

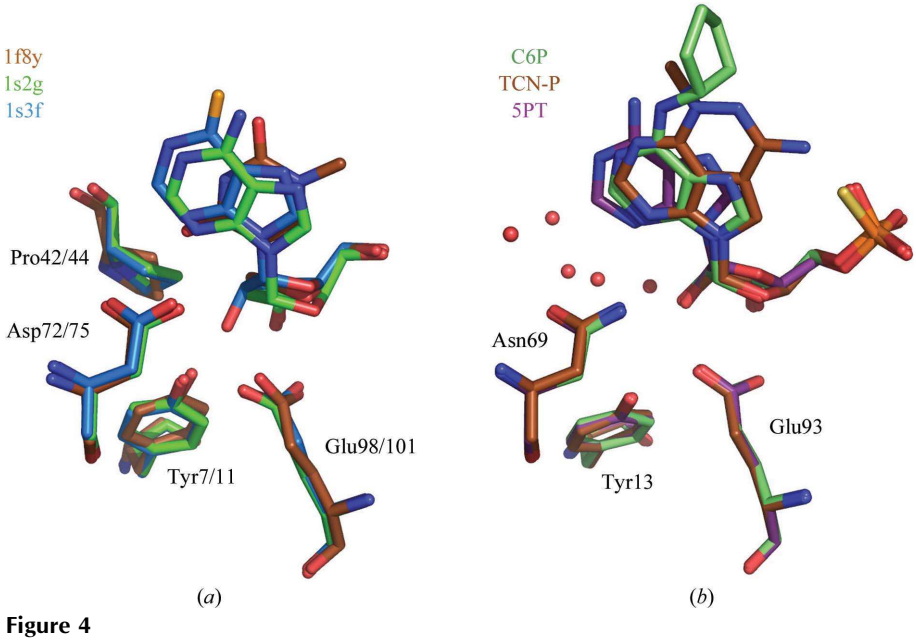


Figure 4
NDT and Rcl catalytic triads. (*a*) Superimposition of the catalytic triad from liganded NDT structures (PDB entries 1f8y, 1s2g and 1s3f; Anand *et al.*, 2004). The catalytic residues and the bound ligand are shown as a wireframe in CPK colours. In addition, a neighbouring proline that prevents the entrance of water molecules into the active site (Pro44 in *L. helveticus* and Pro442 in *L. leichmannii*) is shown. (*b*) Catalytic triad in Rcl structures. Red spheres show water molecules in the vicinity of the ribose from the three structures. The figures were generated using *PyMOL* (http://www.pymol.org).

electronic reprint

and the protein (Fig. 5*d*) that involves the side-chain hydroxyl groups from two serines (Ser87 and Ser117′), the backbone N atoms of Ile18, Arg19, Gly20, Gly89 and, indirectly, Ala118′ (through a water molecule). The same water molecule is hydrogen-bonded to the N7 atom of the adenine moiety. This intricate network of hydrogen bonds results in tight recognition of the phosphate group and this may explain the strong specificity of Rcl for monophosphated nucleotides compared with 5′-sulfated analogues, as described previously (Dupouy *et al.*, 2010). In parallel, the ribose interacts with Gly16, Gly89, Glu93 and Met119′. The latter protrudes from a second monomer and forms short van der Waals contacts with the O4′ and C1′ atoms of the ribose (C^{γ}—O4′ and C^{γ}—C1′ distances of 3.2 and 3.7 Å, respectively). The O4′ atom is also in van der Waals contact with the C^{α} atom of Gly89 (distance of 4.0 Å), while the N atom of Gly16 is hydrogen-bonded to the O3′ atom of the ribose. The two hydroxyls O2′ and O3′ are hydrogen-bonded to the carboxylate group of the catalytic Glu93. The O2′ hydroxyl ($O^{\varepsilon 1}$—O2′ distance of ∼2.6 Å) seems to attract the carboxylate group of Glu93. Accordingly, the presence of an additional hydroxyl group in a ribonucleotide may limit attack at C1′, and this would explain the observed inhibitory activity of nucleoside monophosphates towards Rcl. Tyr13 also forms a hydrogen bond to the $O^{\varepsilon 1}$ atom of the catalytic glutamate and is expected to stabilize its orientation. The neighbouring Asn69 (Asp69 in wild-type Rcl) shows a weak interaction with C1′ ($N^{\delta 2}$—C1′ distance of 3.5 Å). However, in the wild-type enzyme the negatively charged aspartate may activate the hydrolysis of the N-glycosidic bond. The ribose and phosphate moieties appear to be tightly bound in the crystal structure. Conversely, the nucleobase seems to be weakly bound, being sandwiched by only two hydrophobic isoleucines (Ile18 and Ile65), although the neighbouring Asn69 forms a weak hydrogen bond to the N3 atom ($N^{\delta 2}$—N3 distance of 3.1 Å). This situation implies little specificity in nucleobase recognition, which is in agreement with the hydrolytic activity of Rcl towards a large spectrum of 2′-deoxynucleotides.

In conclusion, the crystal structure of Rcl–C6P confirmed and more precisely revealed the organization of its nucleotide-binding site. It strongly suggests that large variations in the nucleobase may be used to derive new inhibitors of this therapeutic target. Furthermore, the binding of C6P, a pro-apoptotic compound, to Rcl will fuel interest in this putative oncoprotein. Our structure also revealed that Rcl could be the actual target of various nucleotide analogues formed of a phosphoribose moiety bearing a nucleobase-like moiety. Several nucleosides are already in clinical use or under

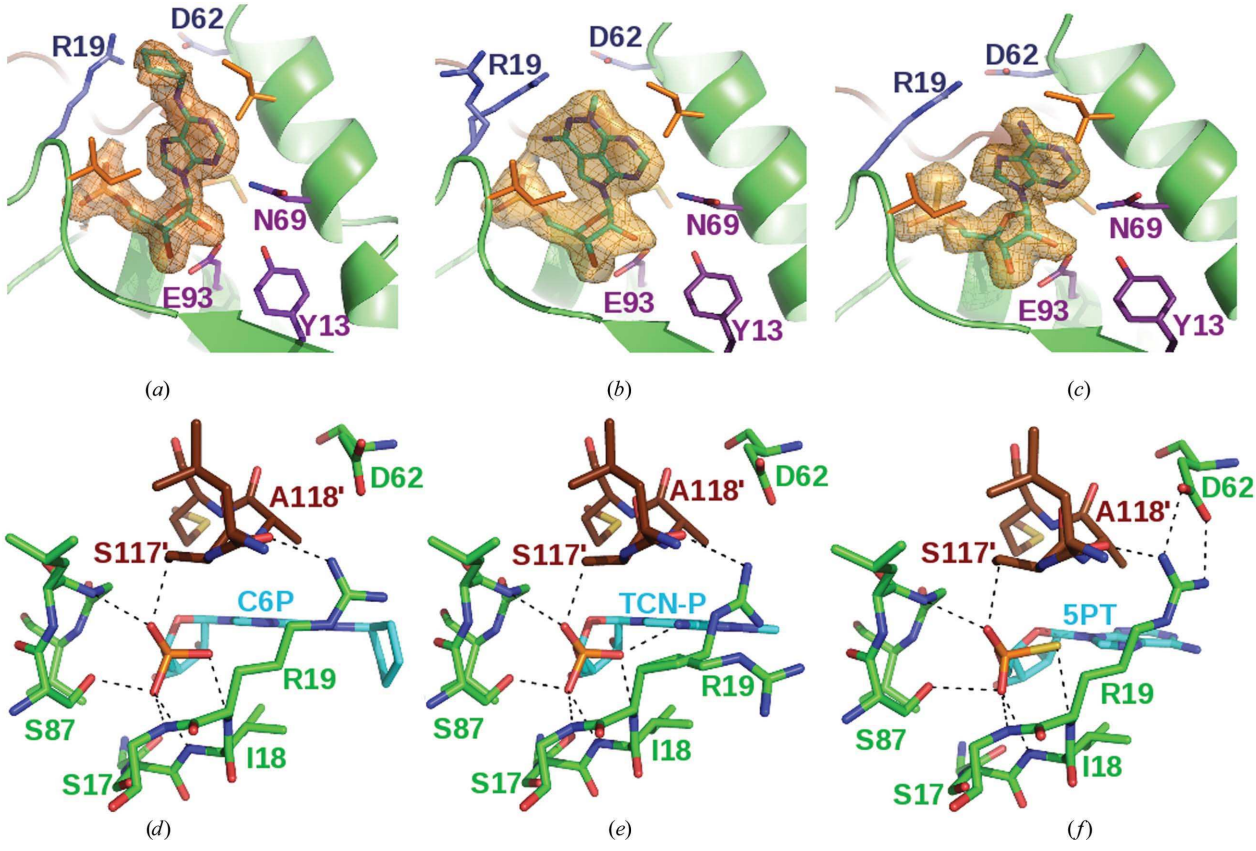


Figure 5
Crystal structures of complexes of Rcl with C6P (*a*, *d*), TCN-P (*b*, *e*) and AMPS (*c*, *f*). (*a*)–(*c*) Views of the $2F_o - F_c$ electron-density map in the active site with the ligand molecule omitted from the Fourier synthesis. The map is contoured at 0.9σ. (*d*)–(*f*) Alternative view of the binding site highlighting the hydrogen-bond networks involved in the tight recognition of the phosphate/phosphorothioate groups. The figures were generated using *PyMOL* (http://www.pymol.org).

electronic reprint

development and some may interact with Rcl after monophosphorylation, as illustrated by TCN-P (see below).

3.6. Crystal structure of Rcl in complex with phosphotriciribine

Among the nucleosides that are already in clinical use or that are promising drug candidates for cancer therapy, we selected triciribine as a potential inhibitor of Rcl in its monophosphorylated form. Triciribine is currently under clinical trials in combination with a farnesyltransferase inhibitor (Balasis *et al.*, 2011). However, the actual mechanism of action of triciribine has yet to be unravelled, although its binding to the PH domain of Akt represents an attractive mechanism, as this protein kinase is important for cell survival (Berndt *et al.*, 2010).

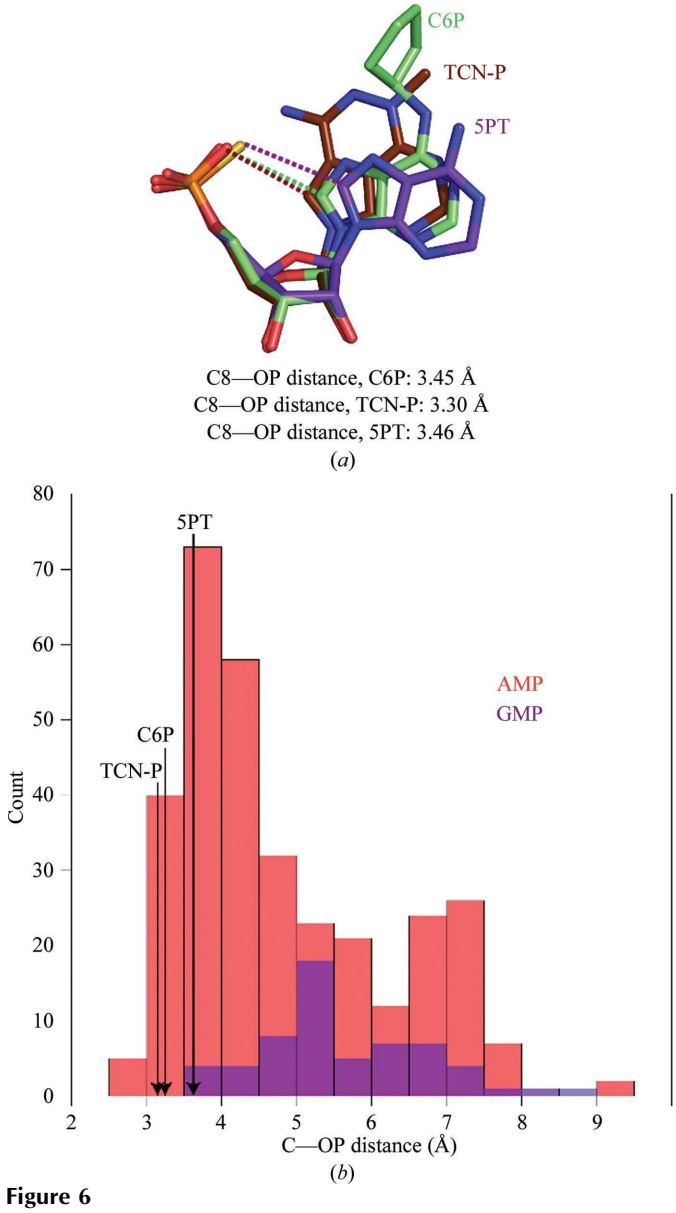


Figure 6
Comparison of the ribose conformations. (*a*) Protein structures were superposed in *Coot* and the ribose conformation or the position of the nucleobase was analyzed. The shortest distances between C^8 and an O (or S) atom of the phosphate group are shown by dotted lines. (*b*) Histogram of the distances observed between C^8 and the closest O atom of the phosphate group of either AMP or GMP found in the protein structure deposited in the PDB (Berman *et al.*, 2000).

Here, we present the crystal structure of Rcl in complex with triciribine monophosphate (TCN-P). Crystals were grown using the same method as used for the Rcl–C6P complex. Data were collected on beamline ID14-2, integrated and merged as for the Rcl–C6P structure (see Table 1). The overall structure closely resembles that of C6P-bound Rcl. The ribose was oriented similarly as in the Rcl–C6P complex. The planar tricycle mimics the adenine moiety apart from the extra ring linking positions N6 and C7 in triciribine (Fig. 5*b*). The main variation occurred in the vicinity of the phosphate group, where a nearby water molecule was replaced by the amino group protruding from the additional ring. The amino group is directly hydrogen bonded to an O atom of the phosphate group (distance of ∼2.65 Å). Concomitantly, the side chain of Arg19 is observed in two alternating conformations. In one conformation it points towards Asp62 to form a salt bridge as in the C6P complex, while in the other it points towards the bulk solvent and the salt bridge with Asp62 is broken (Fig. 5*e*). The intramolecular hydrogen bond and the additional van der Waals contacts between the tricycle and the sandwiching Ile18 and Ile65 (shortest distances of 4.1 and 3.6 Å, respectively) may explain the higher affinity of triciribine monophosphate compared with AMP (K_i of ∼14 *versus* 40 μ*M*). This structure confirmed the accommodation of larger nucleobases in the active site of Rcl. The long and disordered loop (47–59) remained highly flexible in the new structure.

In conclusion, this structure revealed a change in the water shell surrounding the phosphate group, suggesting the possibility of subtle variations in an attempt to stabilize the bound conformation of the nucleotide prior to binding and accordingly lower the entropy cost of binding. Our structure also confirmed that triciribine might target Rcl *in vivo* upon its phosphorylation.

3.7. Crystal structure of Rcl in complex with adenyl-phosphorothioate

Phosphorothioate is currently used as a surrogate phosphate group in RNA molecules for therapeutic applications and shows enhanced stability *in vivo* (owing to its resistance to phosphatases) as well as improved membrane permeability. We wondered whether this moiety could adequately mimic the phosphate group despite the strong specificity of Rcl described previously (Dupouy *et al.*, 2010). AMPS has been shown to be a micromolar inhibitor (14 μ*M*) with a slightly better efficiency than that of AMP (40 μ*M*), while 5′-sulfo-adenosine is a poor inhibitor.

Here, we present the crystal structure of Rcl in complex with phosphorothioate adenosine (AMPS). Crystals were grown using the same method as used for the Rcl–C6P complex. Data were collected on beamline ID29, integrated and merged as for the above structure (see Table 1). We solved the structure of Rcl bound to AMPS by molecular replace-

electronic reprint

ment using the structure of the Rcl–C6P complex as a starting template. The high resolution allowed clear identification of the position of the S atom in the electron density (Fig. 5*c*). Again, as in the Rcl–TCN-P complex, the water molecule lying between the phosphate group and the nucleobase is also displaced in the new crystal structure. This resembles the above structure solved with TCN-P, although in this case the variation was provided by a different chemical modification. As in the previous complex, the presence of the phosphorothioate also induced a change in the position of Arg19. The guanidinium group of its side chain is now in close contact with Asp62 as two hydrogen bonds are formed ($N^{\eta1}-O^{\delta1}$ and $N^{\eta2}-O^{\delta2}$ distances of 2.8 Å) and strengthen the salt bridge (Fig. 5*f*). The crystal structure also revealed a significant change in the conformation of the nucleotide, with a slightly more 'opened' ribose (Fig. 6*a*). This conformational rearrangement of the sugar moiety may be driven by the substitution of an O atom by a larger S atom. Indeed, in the two previous complexes a short distance is observed between an O atom of the phosphate group and the C8 position of C6P or TCN-P (3.45 and 3.29 Å, respectively), while the equivalent distance increases to 3.64 Å in the Rcl–AMPS complex. The observed variations match the conformational variability observed in the other AMP-bound protein complexes that we surveyed in the PDB (Fig. 6*b*).

In conclusion, this structure revealed a subtle change in the sugar conformation and the perfect accommodation of a phosphorothioate in the active site of Rcl. This suggested a slightly more flexible definition of the pharmacophoric feature in order to design new inhibitors with potentially improved affinity and/or better pharmacokinetic properties.

4. Discussion

In this study, we have combined information from an NMR structure of Rcl and fold-recognition models in order to crystallize and then solve the crystal structure of rat Rcl, a putative oncoprotein. The crystal structures of the complexes formed by Rcl with three distinct inhibitors were refined to atomic resolution. Among the three cocrystallized nucleotides, two are already of interest in cancer therapy (Mlejnek & Dolezel, 2005; Balasis *et al.*, 2011). The precise orientation of these ligands in the active site as well as the role of important side chains was precisely determined. Similarly, the positions of water molecules in the active site were defined with high accuracy. This structural information agrees well with functional data obtained by means of enzymology and directed mutagenesis (Ghiorghi *et al.*, 2007; Dupouy *et al.*, 2010). The new crystal structures will help to guide the design of improved inhibitors. Indeed, the tight recognition of the phosphate group in the active site of Rcl suggests that only a few substitutions will readily be accepted at the 5′ position of the nucleotide. Conversely, most positions on the nucleobase appear to be amenable to large variations and substitution in order to design better inhibitors. Chemical synthesis of new ligands is currently in progress.

We would like to thank Martin Cohen-Gonsaud and Stefano Trapani for helpful discussion. This project was funded by the CNRS and INSERM. The authors acknowledge financial support from the CNRS, INSERM and the Institut Pasteur.

References

Anand, R., Kaminski, P. A. & Ealick, S. E. (2004). *Biochemistry*, **43**, 2384–2393.
Armstrong, S. R., Cook, W. J., Short, S. A. & Ealick, S. E. (1996). *Structure*, **4**, 97–107.
Balasis, M. E., Forinash, K. D., Chen, Y. A., Fulp, W. J., Coppola, D., Hamilton, A. D., Cheng, J. Q. & Sebti, S. M. (2011). *Clin. Cancer Res.* **17**, 2852–2862.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Berndt, N., Yang, H., Trinczek, B., Betzi, S., Zhang, Z., Wu, B., Lawrence, N. J., Pellecchia, M., Schönbrunn, E., Cheng, J. Q. & Sebti, S. M. (2010). *Cell Death Differ.* **17**, 1795–1804.
Bosch, J. *et al.* (2006). *J. Med. Chem.* **49**, 5939–5946.
Catherinot, V. & Labesse, G. (2004). *Bioinformatics*, **20**, 3694–3696.
DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H. L., Das, D., Vorobiev, S. M., Iwaï, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature (London)*, **473**, 540–543.
Doddapaneni, K., Mahler, B., Pavlovicz, R., Haushalter, A., Yuan, C. & Wu, Z. (2009). *J. Mol. Biol.* **394**, 423–434.
Doddapaneni, K., Zahurancik, W., Haushalter, A., Yuan, C., Jackman, J. & Wu, Z. (2011). *Biochemistry*, **50**, 4712–4719.
Dupouy, C., Zhang, C., Padilla, A., Pochet, S. & Kaminski, P. A. (2010). *J. Biol. Chem.* **285**, 41806–41814.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Ghiorghi, Y. K., Zeller, K. I., Dang, C. V. & Kaminski, P. A. (2007). *J. Biol. Chem.* **282**, 8150–8156.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. (2009). *Proteins*, **77**, 778–795.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Lewis, B. C., Prescott, J. E., Campbell, S. E., Shim, H., Orlowski, R. Z. & Dang, C. V. (2000). *Cancer Res.* **60**, 6178–6183.
Mao, B., Guan, R. & Montelione, G. T. (2011). *Structure*, **19**, 757–766.
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.
Mlejnek, P. & Dolezel, P. (2005). *Toxicol. In Vitro*, **19**, 985–990.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Pons, J.-L. & Labesse, G. (2009). *Nucleic Acids Res.* **37**, W485–W491.
Porter, D. J., Merrill, B. M. & Short, S. A. (1995). *J. Biol. Chem.* **270**, 15551–15556.
Shi, J., Blundell, T. L. & Mizuguchi, K. (2001). *J. Mol. Biol.* **310**, 243–257.
Short, S. A., Armstrong, S. R., Ealick, S. E. & Porter, D. J. (1996). *J. Biol. Chem.* **271**, 4978–4987.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Yang, Y., Padilla, A., Zhang, C., Labesse, G. & Kaminski, P. A. (2009). *J. Mol. Biol.* **394**, 435–447.
Ye, Y. & Godzik, A. (2004). *Nucleic Acids Res.* **32**, W582–W585.

electronic reprint