Inhibitor selectivity for CDK7 - a structural, thermodynamic and modelling study

Dr. Pascale Hazel¹, Dr. Sebastian H.B. Kroll², Dr. Alexander Bondke²,
Dr. Marion Barbazanges², Dr. Hetal Patel³, Dr. Matthew J. Fuchter², Prof. R Charles
Coombes³, Prof. Simak Ali³, Prof. Anthony G. M. Barrett², Prof. Paul S. Freemont^{1*}

¹Section of Structural Biology, Department of Medicine,

²Department of Chemistry,

Imperial College London

South Kensington Campus

London SW7 2AZ, UK

³Department of Surgery and Cancer,

Imperial College London,

Hammersmith Hospital Campus,

Du Cane Road, London W12 0NN

*correspondence p.freemont@imperial.ac.uk

Keywords

Cancer, CDK7, drug discovery, isothermal titration calorimetry, protein kinase

Table of contents entry

We have solved crystal structures of CDK7-selective inhibitors bound to CDK2. We have used a combination of structural, biophysical and modelling approaches to model the binding of these inhibitors to CDK7, and have used this information to explain their selectivity. We identify specific CDK7 residues that contribute to these inhibitors' specificity.

Abstract

Deregulation of the cell cycle by mechanisms that lead to elevated activities of cyclin-dependent kinases (CDK) is a feature of many human diseases, in particular cancer. We have identified small molecule inhibitors that selectively inhibit CDK7, the kinase that phosphorylates cell cycle CDKs to promote their activities. To investigate the selectivity of these inhibitors we have used a combination of structural, biophysical and modelling approaches. We have determined crystal structures of the CDK7 selective compounds ICEC0942 and ICEC0943 bound to CDK2, and used these to build models of inhibitor binding to CDK7. Molecular dynamics simulations of inhibitors bound to CDK2 and CDK7 have generated possible models of inhibitor binding. In order to experimentally

validate these models, we measured isothermal titration calorimetry (ITC) binding data for recombinant wild type and binding site mutants of CDK7 and CDK2. We have identified specific residues of CDK7, notably Asp155, that are involved in determining inhibitor selectivity. Our molecular dynamics simulations also show that the flexibility of the G-rich and activation loops of CDK7 is likely an important determinant of inhibitor specificity similar to CDK2.

Introduction

Cyclin-dependent kinases (CDKs) are serine/threonine kinases which control cell proliferation.^[1,2] Correct and timely progression through the cell cycle requires that CDK activity be tightly regulated. This is achieved by means of binding to an appropriate cyclin (levels of which vary depending on the cell cycle stage), the presence of endogenous CDK inhibitors (CDKIs, such as p21 and p27) and phosphorylation of specific residues. Deregulation of CDK activity is a component of many diseases, including cancer. This frequently occurs through the amplification of cell cycle regulators, such as cyclin D1, and through CDKI deletion/silencing, both of which can lead to the upregulation of cell cycle CDKs.^[3] Selective small molecule inhibitors of CDK1, CDK2 and CDK4/6 have been developed as cancer therapeutics with the aim of inhibiting cell cycle progression and several are in clinical evaluation.^[4-9] However, there is potential redundancy between CDKs, as highlighted by the finding that mice lacking CDK2 are viable, as are CDK4 or CDK6 knockout mice, although lack of both CDK4 and CDK6 is embryonic lethal.^[10-16] One option to tackle such redundancy is through the use of small molecule CDK inhibitors that have inhibitory activity against multiple CDKs; which may in turn provide a useful method for inhibiting cell cycle progression.^[17-19] An alternative is to selectively inhibit CDK7, due to its role in activating cell cycle CDKs.

Phosphorylation of the activation loop of cell cycle CDKs is required for full activation (Thr160 in CDK2 and corresponding residues in other CDKs).^[20] The CDK7-cyclin H and MAT-1 complex is responsible for phosphorylating cell cycle progression CDKs (CDK1, CDK2, CDK4 and CDK6) in metazoans. Inhibition of

CDK7 would therefore be expected to inhibit all the downstream cell cycle progression CDKs, and indeed the importance of CDK7 in cell cycle progression has been confirmed.^[21,22] As well as its role as a CDK activating kinase, CDK7 is a component of the transcription factor TFIIF and phosphorylates RNA polymerase II (PolII).^[23,24] CDK7-specific inhibitors therefore have the potential of inhibiting both transcription and cell cycle progression, making them attractive anti-cancer drugs.^[25-27]

There has been a large body of work looking at the selectivity of small molecule inhibitors for various CDKs, especially CDK4 and CDK9 over CDK2.^[28-34] Selectivity for CDK7 is much less well understood, and has generally been limited to the influence of a single amino acid substitution of a valine (Val100) for a bulky lysine (Lys89) in the ATP sugar binding region of CDK2.^[35] The role of CDK2 conformational flexibility in determining inhibitor affinity, especially the G-rich and activation loops, has already been reported.^[31,36] The G-rich loop forms the upper part of the ATP triphosphate binding site, while the activation loop (or T-loop) contains Thr160 which is phosphorylated upon activation of CDK2. This has also been observed in structures of CDK9 and is likely to be of importance in other CDKs.^[34] Crystal structures of CDK inhibitors bound to CDK2 are readily available, and several molecular dynamics studies have shown that these simulations can reliably reproduce experimental results.^[30,31,33,37-40]

We have previously reported the identification of several CDK inhibitory compounds, including a molecule selective for CDK7.^[25,37,41] Here, we use a combination of structural, biophysical and modelling techniques to develop an

understanding of the selectivity of representative inhibitors for CDK7 over CDK2. We have focused on three inhibitors from our compound library, two of which are CDK7 selective, and one which is CDK2 selective. In the absence of sufficiently high resolution diffracting crystals of CDK7-inhibitor complexes, we have used isothermal titration calorimetry (ITC) and Molecular Dynamics (MD) simulations to postulate binding modes of inhibitors to CDK7. Our MD simulations of inhibitors bound to CDK2 and CDK7 have identified key residues involved in binding, which we confirmed experimentally using ITC, similar to previous studies of CDK2.^[42] We introduced mutations in the CDK2 and CDK7 ATP binding sites and measured binding data, that enabled us to correlate the MD simulations with thermodynamic data.

Results and Discussion

Our CDK medicinal chemistry programme has led to the identification of potent CDK2- and CDK7-selective inhibitors, as previously reported.^[25,37] Here, we have focussed on three inhibitors from our library: BS-194, which is 140 times more active against CDK2 than CDK7,^[37] ICEC0942 which is 15 times more active against CDK7 than CDK2, and ICEC0943, an enantiomer of ICEC0942, which is a poor inhibitor of both CDK2 and CDK7 (Fig. 1).

Figure 1. Chemical structure of (A) BS-194, (B) ICEC0942 and (C) ICEC0943. IC₅₀ values for CDK7 and CDK2 are shown below each structure.

Binding of BS-194, ICEC0942 and ICEC0943 to monomeric and cyclin-bound CDKs

We measured binding affinities of the inhibitors to CDK2 and CDK7 using ITC. Our aim was to confirm that the inhibitory activities measured by the *in vitro* kinase assay were directly related to binding affinity of the compounds, and to investigate whether the inhibitors bound more favourably to monomeric or cyclin-bound CDKs. Table 1 shows the thermodynamic binding data collected for the three inhibitors. Interestingly, we find that all three bind more tightly to their cyclin-bound CDK targets (BS-194 to CDK2/cyclinA, and ICEC0942 and

ICEC0943 to CDK7/cyclinH) compared to monomeric CDKs. We also find that binding affinities to the monomeric CDKs correlate with the activities measured in kinase assays, in that ICEC0942 and ICEC0943 both bind more strongly to monomeric CDK7 than monomeric CDK2. However, BS-194, which is 140-fold more active against CDK2 than CDK7 in *in-vitro* kinase assays, binds with the same affinity to monomeric CDK7 as to CDK2.

Protein	BS-194				ICEC0942				ICEC0943			
	ΔG	ΔH	$-T\Delta S$	kD	ΔG	ΔH	$-T\Delta S$	kD	ΔG	ΔH	$-T\Delta S$	kD
CDK2	-9.2 ± 0.04	-11.0 ± 1.0	1.7 ± 1.0	173 ± 12	-7.7 ± 0.1	-8.1 ± 0.9	0.4 ± 0.9	2341 ± 400	ND	ND	ND	$> 2000^\dagger$
CDK2/A*	-10.5 ± 0.04	-12.2 ± 0.04	1.6	20 ± 2	ND	ND	ND	ND	ND	ND	ND	ND
CDK7	-9.3 ± 0.04	-14.7 ± 0.2	5.4 ± 0.2	152 ± 8	-10.1 ± 0.1	-11.2 ± 0.1	1.1 ± 0.2	40 ± 9	-8.7 ± 0.2	-8.7 ± 0.7	-0.03 ± 0.5	422 ± 100
CDK7/H	ND	ND	ND	ND	-11.1 ± 0.5	-7.7 ± 0.8	-3.4 ± 0.3	9 ± 7	-10.4 ± 0.3	-8.3 ± 0.7	-2.1 ± 0.4	27 ± 15

Table 1. Thermodynamic binding data for the three inhibitors studied binding to monomeric CDK2 and CDK7 and active CDK2/A and CDK7/H complexes (ΔG , ΔH and $T\Delta S$ in kcal/mol, kD in nM). Errors are standard deviations of three measurements. *CDK2/A measured once only, errors are from the Origin fitting program. † Binding too weak to be measured. ND: not determined.

The thermodynamic data also show that binding of all three inhibitors to monomeric CDK2 and CDK7 is enthalpically driven. In contrast, binding of ICEC0942 and ICEC0943 to CDK7/cyclinH is both enthalphy and entropically

driven. As CDK/cyclin complex formation is accompanied by large structural changes in the ATP binding side of CDKs, such differences in inhibitor binding thermodynamics is perhaps not unexpected.^[43]

Crystal structures of ICEC0942 and ICEC0943 bound to CDK2

Despite repeated attempts with CDK7 and CDK7/cyclinH complexes, we were unable to obtain suitable crystals for high-resolution structure determination. We did however solve the crystal structures of BS-194,^[37] ICEC0942 and ICEC0943 bound to monomeric inactive CDK2 by soaking each inhibitor into preformed CDK2 crystals.

We collected 2.0 Å data sets of CDK2-ICEC0942 and CDK2-ICEC0943 and after molecular replacement with our published structure of CDK2 in complex with BS-194 (PDB 3NS9), we observed clear electron density within the ATP binding site that allowed unambiguous fitting of both inhibitors. Both inhibitors bind in essentially the same orientation as BS-194 (Fig. 2, Table S1). All three inhibitors form hydrophobic stacking interactions with the side chains of Ile10, Phe80 and Leu134 and hydrogen bonds with the kinase hinge region (backbone N and O of Leu83). Assuming that ICEC0942 is protonated on N44 (Inhibitor numbering is shown in Fig. 1), it is able to form a hydrogen bond with Asn132 and electrostatic interactions with Asp145. This is in contrast to BS-194 binding to CDK2, where interactions between BS-194 and Asp145 and Gln131 (amongst others) were all water-mediated hydrogen bonds.^[37] ICEC0943 forms polar interactions with CDK2, however due to its opposite stereochemical configuration relative to ICEC0942, the orientation of hydrogen bonds is less optimal, and the distance of

N44 to Asp145 is short (2.3 Å). These interactions also affect the G-rich loop, which in the case of ICEC0943 exhibits higher B-factors. This then has a subsequent effect on the activation loop, which is much less well defined in the ICEC0943 complex than in the BS-194 or ICEC0942 complexes. This is shown in Fig. 3, where the G-rich and activation loops of CDK2 bound to BS-194, ICEC0942 and ICEC0943 are coloured according to B-factors. The network of water-mediated interactions formed by BS-194 enables a compact structure with the G-rich loop folded against the inhibitor, and the activation loop packed against the G-rich loop. This is also the case for ICEC0942, but ICEC0943 does not allow such a rigid packing of the G-rich loop, and the activation loop density is poorly visible in the crystal structure (no density was observed for residues Gly153 to His161). Higher binding affinity of CDK2 inhibitors has previously been shown to be correlated with decreased G-rich and activation loop flexibilities in monomeric CDK2.^[36] In the case of BS-194, this loss of loop flexibility is unlikely to reduce cyclin binding, as BS-194 also binds strongly to the activated CDK2.

Figure 2. Crystal structure of CDK2 in complex with (A) BS-194 (1.8 Å resolution), (B) ICEC0942 (2.0 Å resolution) and (C) ICEC0943 (2.0 Å resolution). Hydrogen bonds are shown as dotted lines and important binding site amino acids are labelled.

Figure 3. Crystal structure of (A) BS-194, (B) ICEC0942 and (C) ICEC0943 bound to CDK2, showing the G-rich (labelled G) and activation loops (labelled A) and coloured according to B-factors (from blue – low B-factor to red – high B-factor). In (C) the activation loop residues 153-161 were removed during refinement due to poor electron density.

Molecular dynamics simulations of ICEC0942 and ICEC0943 binding to CDK2 and CDK7

A structure-based sequence alignment of CDK2 and CDK7 shows that many of the binding site residues are conserved (Fig. 4). Inhibitor specificity is therefore likely to be determined by the different local structures of the binding sites, and influenced by the inherent flexibility of the proteins. CDK2 crystal structures have shown that flexibility of the protein contributes to the binding affinity of various inhibitors.^[36] This is likely the case for CDK7, and we therefore used molecular dynamics simulations in order to generate models of how ICEC0942

and ICEC0943 could bind to CDK7. To date, there is only one available crystal structure of CDK7 namely a monomeric complex with ATP, and we therefore used this structure for the MD simulations.^[44] We decided against building a CDK7/cyclin H model, as the large conformational changes upon cyclin binding would create too many uncertainties in our starting model. Our ITC measurements allowed us to verify that the inhibitors showed the same relative affinities towards monomeric CDKs as compared to cyclin-bound CDKs. We therefore used the monomeric CDK7 structure for the MD simulations.

Figure 4. Structure-based sequence alignment of CDK2 and CDK7 N-terminal lobe and ATP binding site. Blue boxes highlight CDK2 residues that are within 5 Å of ICEC0942 in the crystal structure. CDK2 numbering is above the sequence and CDK7 numbering is below. The CDK2 G-rich loop is highlighted in red and the activation loop in green.

The crystal structures of ICEC0942 and ICEC0943 bound to CDK2 were used to generate starting models for their binding to CDK7. The published CDK7 crystal structure was aligned with the CDK2 crystal structures using Modeller^[45] and the inhibitors positioned within the CDK7 ATP binding site. The resulting CDK7-ICEC0942 complex is shown in Figure 5. This model revealed that in the context of a static ATP-bound binding site, Asp155 is unavailable for binding to the inhibitors. Asp155 is located deep within the binding site, and access is sterically restricted by the presence of Lys41. This is in contrast to the CDK2 binding site where the equivalent Asp145 is sterically unhindered within the binding site and accessible to the inhibitors (Fig. 2B).

Figure 5. ICEC0942 positioned in the CDK7 ATP binding site after alignment of CDK7 with CDK2. CDK7 coordinates were obtained from the published crystal structure (PDB code: 1U2A).^[37]

MD simulations were carried for between 17 and 22ns using the starting models of CDK2 and CDK7 bound to BS-194, ICEC0942 and ICEC0943. We used the Amber force field, as this has been successfully applied to previous CDK MD simulations.^[30,31,38] Root mean square deviations (RMSD) with respect to the starting (crystal) structures show that all the inhibitor complexes are stable over the timescales used, although the CDK7 complexes show larger RMS deviations compared to CDK2 (Fig. S2). This is expected due to the lower resolution of the CDK7 crystal structure (3.0 Å; PDB 1UA2) used as a starting model for the simulations, compared to higher resolution CDK2 starting models (2.0 Å; PDB 5JQ5 and PDB 5JQ8). The stability of the trajectories shows that the force field and inhibitor parameterization were appropriate for the timescales studied. Interactions between the inhibitors and CDK2 which were not observed in the crystal structure, were observed during the simulations, and this has been observed previously for other small molecule-CDK2 interactions.^[46]

During the MD simulations of CDK7 with ICEC0942 and ICEC0943, the inhibitors remained bound in similar conformations as in CDK2. Hydrogen bonds between the inhibitors and the backbone of the hinge region (especially Met94) were maintained in all simulations. Different specific interactions were however observed, notably with the backbone of Gly21 (in the G-rich loop) and the side chains of Asp155 and Asp137. For analysis purposes, the trajectories were divided into clusters, based on the inhibitor RMSD. Representative snapshots of the most populated clusters during the simulations of CDK2-ICEC0942, CDK7-ICEC0942 and CDK7-ICEC0943 are shown in Fig. 6. Hydrogen bonds formed during the MD simulations of CDK2 and CDK7 are also shown in Table S3.

Figure 6. Representative snapshots of the most populated cluster during the MD simulations of (A) CDK2+ICEC0942, (B) CDK7+ICEC0942 and (C) CDK7+ICEC0943. Residues within hydrogen bonding distance of ICEC0942 (blue) and ICEC0943 (orange) are shown as sticks and hydrogen bonds are shown as dotted lines.

Starting models of ICEC0942 binding to the CDK7 crystal structure revealed that Asp155 was difficult to access for inhibitor binding, partly due to the position of the side chain deep in the ATP binding pocket, and partly due to the position of Lys41 (Fig. 5). However, motions of the protein during the MD simulations of CDK7 show that polar interactions between the inhibitors and Asp155 are possible, and were indeed observed for substantial periods of simulation time (~40% of the time for ICEC0942 and almost 90% for ICEC0943). This emphasizes the importance of considering dynamical flexibility within the target protein when investigating inhibitor binding, rather than relying solely on crystal structure information. ICEC0942 was also able to form long-lived hydrogen bonds with Asp137, although this was at the expense of one of the two hydrogen bonds formed with the hinge region (to Met94 N). In contrast, for the

CDK2-ICEC0942 simulation, Asp127 (the equivalent residue in CDK2) is never within hydrogen-bonding distance of ICEC0942 (Fig. 6A).

Figure 6B shows how ICEC0942 is able to bridge the CDK7 binding site between Asp155 and Gly21 in the G-rich loop. A hydrogen bond between ICEC0942 and Gly21 O was formed during ~ 70% of the simulation. ICEC0942 can also form a hydrogen bond with the backbone O of the equivalent CDK2 Glu12, however this is only formed for ~5% of the 20ns simulation. The ability of ICEC0942 to form a bridging set of polar contacts with the C-terminal lobe and the G-rich loop of CDK7 could help explain its selectivity for CDK7. As observed for CDK2 inhibitors^[36], a key determinant in inhibitor specificity is the effect on the flexibility of the G-rich loop, and the activation loop by extension (Fig. 3). Indeed, CDK7 displayed smaller G-rich and activation loop RMS deviations when bound to ICEC0942 compared to ICEC0943 (Fig. S4).

In contract to ICEC0942, ICEC0943 did not form any polar interactions with the G-rich loop during the timescales considered. Fig. 6C shows a representative snapshot of the most populated cluster, and is also representative of the full simulation, as the hydrogen bonds shown were mostly conserved over 20ns. Fig. 6C shows how ICEC0943, with opposite stereochemical configuration to ICEC0942, cannot interact with the G-rich loop while maintaining a hydrogen bond with Asp155 unlike ICEC0942 which can (Fig. 6B). Instead, the hydrogen bond between ICEC0943 and Asp155 directs the inhibitor O1 towards interactions with the backbone O of Asn141.

In summary, the MD simulations suggest an explanation for the selectivity of ICEC0942 for CDK7 compared to CDK2, and also for the higher affinity of the ICEC0942 enantiomer over ICEC0943. In CDK7, ICEC0942 is observed to form a bridge between the N and C lobes of the kinase, anchoring the G-rich loop with a hydrogen bond. On the other hand, ICEC0943 forms a set of non-polar contacts with the G-rich loop of CDK7, as does ICEC0942 when bound to CDK2. The MD simulation trajectories are dependent on the initial starting conformation of CDK2 and CDK7, however, they do provide information on the conformational space available for inhibitor binding. Due to the conformational flexibility of the protein, a broader range of residues are accessible to the inhibitors within the ATP binding site during the MD simulations,. ICEC0942 and ICEC0943 were able to form polar interactions with Asp155 and Asp137 (ICEC0942 only) in CDK7, despite their inaccessibility in the ATP-bound crystal structure (Note that Asp155 is deep within the ATP binding pocket, and Asp137 is ~ 5 Å away from ATP).

Binding of ICEC0942 and ICEC0943 to CDK2 and CDK7 mutants

The molecular dynamics simulations identified key residues which could be important for inhibitor binding and selectivity for CDK7. We therefore decided to use ITC to validate these results experimentally. We chose specific amino acid mutations for CDK7 and CDK2 according to two main criteria: (1) the mutation should not be expected to alter the overall structure of the binding site; (2) the mutated residues should still be expected to be in the same position in the cyclin-bound complex, so that conclusions drawn about binding can be applied to the activated structures.

Two key residues identified during the CDK7 simulations were Asp155 and Asp137. Upon inspection of the CDK2-ICEC0942 crystal structure, we decided not to mutate the side chain of Asp137 as the equivalent Asp127 in CDK2 is involved in a hydrogen bond with Lys129, the removal of which may affect the local protein structure. From alignments of CDK2 with CDK2/cyclinA, we know that Asp145 (Asp155 in CDK7) remains in a similar orientation upon cyclin binding, and that mutation to alanine is unlikely to alter the geometry of the binding site, given that a potential salt bridge with Lys33 is not observed in any of our CDK2 crystal structures. A salt bridge between CDK7 Asp155 and Lys41 is also only weakly observed in the ATP-bound CDK7 crystal structure. We therefore made recombinant CDK7 Asp155Ala and CDK2 Asp145Ala mutant proteins. The results of ITC binding experiments of ICEC0942 and ICEC0943 are shown in Table 2. For CDK2, Asp145Ala has a negligible overall effect on ICEC0942 binding to CDK2, as shown by $\Delta\Delta G$ (ΔG mutant - ΔG wild type = -0.1). This is despite observing a weak electrostatic interaction between ICEC0942 and Asp145 in the crystal structure. However, both ICEC0942 and ICEC0943 binding to CDK7 Asp155Ala show significant effects ($\Delta\Delta G = 0.7$ for ICEC0942 and $\Delta\Delta G = 0.9$ for ICEC0943). Binding is 3-4 times weaker for both inhibitors compared to wild type, suggesting that Asp155 is important in inhibitor binding to CDK7. Decomposition of $\Delta\Delta G$ into entropic and enthalpic contributions reveals that the CDK7 mutation is enthalpically favorable and entropically unfavorable compared to wild type enzyme. These results suggest a more complex binding mode in the mutant than the simple loss of a hydrogen bond between the inhibitor and Asp155, which would be enthalpically unfavorable. The latter is the case for

binding of ICEC0942 to the Asp145Ala mutant in CDK2. However, the enthalpic contribution is countered by an equal entropic contribution, leading to a negligible $\Delta\Delta G$. In summary, the ITC results support our hypothesis that Asp155 in CDK7 is a key residue in the binding of ICEC0942 and ICEC0943.

Protein	ICEC0942				ICEC0943			
	$\Delta\Delta G$	$\Delta\Delta H$	$-\Delta\Delta S$	kD	$\Delta\Delta G$	$\Delta\Delta H$	$-\Delta\Delta S$	kD
CDK2 Asp145Ala	-0.1	1.3	-1.4	2175 $\pm$ 900	ND	ND	ND	ND
CDK7 Asp155Ala	0.7	-3.0	3.8	147 $\pm$ 40	0.9	-2.5	3.4	1800 $\pm$ 300

Table 2. Thermodynamic binding data for ICEC0942 and ICEC0943 binding to CDK2 Asp145 and CDK7 Asp155 mutants. ($\Delta\Delta G$, $\Delta\Delta H$ and $\Delta\Delta S$ in kcal/mol, kD in nM). $\Delta\Delta G = \Delta G$ mutant - ΔG wild type, $\Delta\Delta H = \Delta H$ mutant - ΔH wild type, $\Delta\Delta S = \Delta S$ mutant - ΔS wild type. ND: not determined.

The crystal structures and simulations show that different inhibitors have different effects on the flexibility of the G-rich and activation loops of CDK2 and CDK7. Although we looked at specific inhibitor-kinase interactions, the size of the binding pocket could also play a role. To this effect, we investigated the role of Val100 in CDK7 inhibitor selectivity. In the absence of structural information of bound inhibitors, the importance of this residue has been suggested previously, as its equivalent Lys89 in CDK2 creates a smaller binding site.^[35] We made recombinant CDK7 Val100Lys and CDK2 Lys89Val mutant proteins and investigated ICEC0942 binding with ITC (Table 3). We indeed find that binding to CDK7 Val100Lys is less favourable than to wild type suggesting that the wider

binding pocket contributes to ICEC0942 binding affinity. In contrast, binding of BS-194 to CDK7 Val100Lys is unaffected compared to wild type. The corresponding mutant in CDK2 (Lys89Val) leads to a negligible $\Delta\Delta G$ for both BS-194 and ICEC0942 binding (Table 3). This is expected as Lys89 does not interact directly with either inhibitor in the crystal structures. In fact, electron density for the Lys89 side chain is poor, indicating that Lys89 is flexible even in the presence of bound inhibitor.

Protein	BS-194				ICEC0942			
	$\Delta\Delta G$	$\Delta\Delta H$	$-\Delta\Delta S$	kD	$\Delta\Delta G$	$\Delta\Delta H$	$-\Delta\Delta S$	kD
CDK2 Lys89Val	-0.1	2.4	-2.6	138 $\pm$ 35	-0.2	0.6	-0.8	1615 $\pm$ 270
CDK7 Val100Lys	-0.3	1.6	-1.9	87 $\pm$ 1	0.8	2.8	-2.0	152 $\pm$ 50

Table 3. Thermodynamic binding data for BS-194 and ICEC0942 binding to CDK2 Lys89 and CDK7 Val100 mutants. ($\Delta\Delta G$, $\Delta\Delta H$ and $\Delta\Delta S$ in kcal/mol, kD in nM). $\Delta\Delta G = \Delta G$ mutant - ΔG wild type, $\Delta\Delta H = \Delta H$ mutant - ΔH wild type, $\Delta\Delta S = \Delta S$ mutant - ΔS wild type.

Effect of G-rich loop flexibility on inhibitor binding

As mentioned above, we have observed that strong binding of BS-194 results in a reduced flexibility of the G-rich and activation loops of monomeric CDK2. This type of inhibitor effect on loop flexibility has been previously linked to inhibitors that bind more strongly to monomeric inactive CDK2^[36]. However, BS-194 is able to bind to monomeric CDK2 by constraining its G-rich loop conformation, but binds more strongly to the activated complex, where the activation loop is in a

different conformation. This shows that designing inhibitors to interact with, and anchor the G-rich loop down, will not necessarily reduce cyclin binding ability. Indeed, BS-194 is an example of an inhibitor that can reduce the G-rich loop flexibility, but also bind strongly to cyclin-bound CDK2.

Conclusions

Using a combination of structural and thermodynamic data, we have built models of how selective inhibitors can bind to CDK7, which lead us to the following conclusions:

- 1) The ATP binding site is flexible, especially the G-rich and activation loops and this flexibility is important for inhibitor binding. CDK2 –inhibitor complexes have less flexible G-rich and activation loops (Figure 3). Molecular dynamics simulations suggest a similar flexibility for CDK7, as CDK7-specific inhibitors are able to tether the G-rich loop (Figure 6).
- 2) We identify residues in CDK7 that could form polar contacts with inhibitors, in both the C-terminal lobe and the G-rich loop. Although these residues are conserved between CDK2 and CDK7 (for example Asp155 and Asp137 in CDK7), the different geometries and shapes of the binding sites seem to strongly impact the accessibility of these residues to the inhibitors studied. For example, mutating Asp155 to alanine in CDK7 had a much larger effect on inhibitor binding than the equivalent Asp145 in CDK2, suggesting that for CDK7, Asp155 has a more important role in inhibitor binding. In general, we have found a positively charged amine in this region of the binding site to be important for CDK7 selectivity, which supports our earlier observations.^[25]

- 3) The CDK7 ATP binding site is more open than that of CDK2, in part due to Val100 in CDK7 replacing Lys89 in CDK2. This steric effect is likely to be felt mostly for larger inhibitors, as Lys89 in CDK2 is flexible in our crystal structures. Although the presence of Val100 in CDK7 is beneficial for ICEC0942 binding, the presence of a valine or lysine in CDK2 does not affect ICEC0942's binding affinity for CDK2.

In summary, by using a combination of methods, we have been able build models to explain the observed selectivity of our inhibitors for CDK7 compared to CDK2. Although our study focuses on the binding of inhibitors to the inactive form of CDK7, due to a lack of structural data on the CDK7/cyclin-H complex, we have examined specific residues which we believe to be relevant in both monomeric and cyclin-bound CDK binding. Our study shows that despite the absence of high resolution crystal structures of inhibitor complexes, a combined use of careful biophysical measurements and MD simulation can provide insights in inhibitor selectivity and binding modes between highly related target proteins.

Materials and methods

Protein expression and purification:

Crystallization-grade CDK2 was expressed in E-coli fused with a SUMO tag, and purification, crystallization and structure solution by molecular replacement were as previously described.^[37] A GST-tagged CDK2 for ITC experiments was created by subcloning CDK2 into pGEX-6P-1 (GE Healthcare) and expressed in E-coli BL21 DE3 Gold. Protein expression was carried out overnight at 22°C after induction with 0.1 mM IPTG. The cell pellet was resuspended in lysis buffer (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM EDTA, 1 mM DTT) and the cells lysed by sonication. The clarified lysate was purified using a GST affinity column (GE Healthcare) and eluted in lysis buffer supplemented with 15 mM reduced glutathione. CyclinA was cloned into at pET21d vector and expressed as CDK2. After lysis and clarification in the above CDK2 lysis buffer, cyclinA and CDK2 lysates were mixed prior to GST column purification. The GST-CDK2/cyclinA complex was eluted as above. ITC experiments used the GST-CDK2 fusion protein or CDK2 and CDK2/A after cleavage with PreScission protease (GE Healthcare) and a further GST affinity column purification in the above buffers. The final proteins were dialysed overnight against ITC buffer (10 mM Hepes pH 7.5, 200 mM NaCl, 5% glycerol).

Wild type and mutant CDK7 fused to an MBP tag were expressed using baculovirus infection of sf21 insect cells. CDK7 was cloned into a pFastBac-his6-MBP vector with a PreScission cleavage site. CyclinH was cloned into a pFastBac

vector with no tag. The resulting pFastBac-his6-MBP-CDK7 and pFastBac-cyclinH plasmids were transformed into DH10Bac, following the manufacturer's instructions (Invitrogen). The resulting Bacmids were extracted and purified, and transfected into sf21 insect cells using Cellfectin or CellfectinII, following the manufacturer's instructions (Invitrogen). Virus (V0) was collected after 72 hours of incubation at 27°C, and used to infect 50 ml of sf21 at $0.4-0.8 \times 10^6$ cells/ml. After 72 hours, virus V1 was collected, titered by plaque assay and used to generate V2 virus for infection in 200ml sf21 at $0.4-0.8 \times 10^6$ cells/ml. V1 titers were usually between 1×10^7 and 1×10^8 pfu/ml. Protein expression was routinely carried out in 6 to 12 2L roller bottles, with 400ml of $1.4-1.8 \times 10^6$ cells/ml in each bottle. The CDK7/H complex was expressed by co-infecting cells with CDK7 and cyclinH bacmids. After 72 hours shaking at 27°C, the cells were pelleted at 630 xg, resuspended in lysis buffer (25 mM Hepes KOH pH 7.5, 10 mM NaCl, 5 mM imidazole) and sonicated briefly. The clarified cell lysate was brought up to 250 mM NaCl, loaded onto a nickel affinity column and eluted in lysis buffer with 500 mM imidazole. The eluted peak was loaded immediately onto an MBP affinity column (GE Healthcare) pre-equilibrated in binding buffer (25 mM Hepes pH 7.5, 1 mM EDTA, 1 mM DTT, 200 mM NaCl and 5% glycerol) and eluted in binding buffer with 15mM maltose. The MBP-CDK7 fusion protein was used in ITC experiments after an overnight dialysis against ITC buffer (10 mM Hepes pH 7.5, 200 mM NaCl, 5% glycerol). For MBP tag cleavage, the fusion protein was dialysed overnight against ITC buffer with 1mM DTT and PreScission protease, and further purified over an MBP affinity column.

To check that CDK2, CDK7 and the mutants were folded, an ATP affinity column was used (A6888, Sigma). The proteins were diluted into binding buffer (25 mM Hepes pH 7.5, 1 mM EDTA, 1 mM DTT, 10% glycerol, 25 mM NaCl for CDK2 and 100 mM NaCl for CDK7) and loaded onto the ATP column. Bound protein was eluted with a linear NaCl gradient.

All compounds were synthesised as previously reported.^[25,37] All compounds were >95% pure as determined by HPLC analysis. The analytical data for BS-194 has been reported previously.^[37] Analytical data for ICEC0942 and ICEC0943 can be found in the supporting information.

ITC experiments:

ITC experiments were carried out on a MicroCal VP-ITC system (GE Healthcare). All measurements were carried out at 25°C, in the ITC buffer above. The cell contained protein at concentrations between 15 and 40 μ M, and the syringe contained the inhibitor between 120 and 500 μ M. Inhibitor stocks in 100% DMSO were diluted to the appropriate concentration in ITC buffer, and the equivalent amount of DMSO was added to the protein sample in the cell. 25 or 30 injections of 12 or 10 μ L respectively were used depending on the expected binding affinity. Heats of dilution of the inhibitor into buffer were taken into account either by subtracting the heats at the end of the dilution when a plateau was observed, or by subtracting the heats of a blank run where the inhibitor was injected into buffer with DMSO alone. The Origin software was used for all data analysis. Unless otherwise indicated, each measurement was carried out in triplicate. As the fusion proteins were used without cleavage of the tag, controls

using wild type CDK2 and CDK7 after cleavage and GST were used. These experiments showed that the expression tags had no effect on the thermodynamic measurements.

In vitro kinase assays and IC50 determination:

The purified recombinant CDK2/cycE1 and CDK7/CycH/MAT1 were purchased from ProQinase GmbH. Kinase assays were performed according to manufacturer's protocols. Briefly, Rb-CTF (ProQinase GmbH) and RNA Polymerase II (Cambridge Research Biochemicals) peptide was used for CDK2 and CDK7 kinase assays, respectively. A luciferase assay (PKLight assay; Cambrex) was used to determine ATP remaining at the end of the kinase reaction, which provides a measure of kinase activity, according to the manufacturer's protocols. GraphPad Prism Software was used to generate the IC50s for each CDK.

Molecular dynamics simulations:

The inhibitors were parameterized using the Antechamber suite from AmberTools12^[47]. AM1-BCC charges were computed with sqm, and bond and angle parameters automatically assigned, then checked manually. The crystal structures of CDK2 with BS-194, ICEC0942 and ICEC0943 were used as starting structures for the molecular dynamics simulations. Crystallographic water molecules were stripped, apart for water molecules bridging interactions between BS-194 and CDK2. Missing loop residues (Ile35-Gly43) were built using Modeller⁴⁵. For CDK7, the crystal structure (PDB code 1UA2) was used after removal of ATP and water molecules. The inhibitors were placed within the ATP

binding site by alignment of CDK7 with the CDK2 crystal structures. Missing loop residues Lys44-Ile55 were built using Modeller. We used the ff99SB force field for the protein, and gaff for the inhibitors. The Leap program of AmberTools12 was used to create input files for the dynamics: ions (Na⁺ or Cl⁻) were added to neutralize the systems, and the system was solvated with TIP3P water molecules for at least 12.0 Å in each dimension. NAMD was used to run the molecular dynamics.^[48] Between 17 and 22 ns of MD were carried out for each system. The trajectories were visualized in VMD and analysed with the cpptraj module of AmberTools12.

Supporting Information Available: Analytical data for ICEC0942 and ICEC0943, x-ray data collection and refinement statistics, rms deviations and hydrogen bonding observed during MD simulations.

PDB ID Codes: X-ray structures were deposited with the Protein Data Bank, with codes 5JQ5 (CDK2-ICEC0942) and 5JQ8 (CDK2-ICEC0943).

References

- [1] M. Malumbres, E. Harlow, T. Hunt, T. Hunter, J. Lahti, G. Manning, D. O. Morgan, L-H. Tsai, D. J. Wolgemuth, *Nat. Cell Biol.* **2009**, *11*, 1275-6.
- [2] M. Malumbres, *Genome Biol.* **2014**, *15*, 122-31.
- [3] M. Malumbres, M. Barbacid, *Nat. Rev. Cancer* **2001**, *1*, 222-31.
- [4] M. Malumbres, P. Pevarello, M. Barbacid, J. Bischoff, *Trends Pharmacol. Sci.* **2008**, *29*, 16-21.
- [5] C. G. Murphy, M. N. Dickler, *Oncologist* **2015**, *20*, 483-90.

- [6] T. M. Pitts, S. L. Davis, S. G. Eckhardt, E. L. Bradshaw-Pierce, *Pharmacol. Ther.* **2014**, *142*, 258-69.
- [7] H. Galons, N. Oumata, O. Gloulou, L. Meijer, *Expert Opin. Ther. Pat.* **2013**, *23*, 945-63.
- [8] J. S. Blachly, J. C. Byrd, *Leuk. Lymphoma* **2013**, *54*, 2133-43.
- [9] T.A. Chohan, H. Qian, Y. Pan, J.Z. Chen, *Curr. Med. Chem.* **2015**, *22*, 237-63.
- [10] O. Tetsu, F. McCormick, *Cancer Cell* **2003**, *3*, 233-45.
- [11] D. Coudreuse, P. Nurse, *Nature* **2010**, *468*, 1074-9.
- [12] S. Ortega, I. Prieto, J. Odajima, A. Martín, P. Dubus, R. Sotillo, J. L. Barbero, M. Malumbres, M. Barbacid, *Nat. Genet.* **2003**, *35*, 25-31.
- [13] C. Berthet, E. Aleem, V. Coppola, L. Tessarollo, P. Kaldis, *Curr. Biol.* **2003**, *13*, 1775-85.
- [14] S. Rane, P. Dubus, R. Mettus, E. Galbreath, G. Boden, E. Reddy, M. Barbacid, *Nat. Genet.* **1999**, *22*, 44-52.
- [15] T. Tsutsui, B. Hesabi, D. Moons, P. Pandolfi, K. Hansel, A. Koff, H. Kiyokawa, *Mol. Cell. Biol.* **1999**, *19*, 7011-9.
- [16] M. Malumbres, R. Sotillo, D. Santamaría, J. Galán, A. Cerezo, S. Ortega, P. Dubus, M. Barbacid, *Cell* **2004**, *118*, 493-504.
- [17] A. Stone, R. Sutherland, E. Musgrove, *Crit. Rev. Oncogenesis* **2012**, *17*, 175-198.
- [18] J. Węsierska-Gądek, M. Kramer, *Expert Opin. Invest. Drugs* **2011**, *20*, 1611-28.
- [19] J. Węsierska-Gądek, M. Maurer, N. Zulehner, O. Komina, *J. Cell. Physiol.* **2011**, *226*, 341-9.
- [20] R. Fisher, D. A. Morgan, *Cell* **1994**, *78*, 713-24.
- [21] S. Larochelle, K. Merrick, M.-E. Terret, L. Wohlbold, N. Barboza, C. Zhang, K. M. Shokat, P. V. Jallepalli, R. P. Fisher, *Mol. Cell* **2007**, *25*, 839-50.
- [22] M. Ganuza, C. Sáiz-Ladera, M. Cañamero, G. Gómez, R. Schneider, M. Blasco, D. Pisano, J. M. Paramio, D. Santamaría, M. Barbacid, *EMBO J.* **2012**, *31*, 2498-510.
- [23] H. Lu, L. Zawel, L. Fisher, J. Egly, D. Reinberg, *Nature* **1992**, *358*, 641-5.
- [24] H. Serizawa, T. Mäkelä, J. Conaway, R. Conaway, R. Weinberg, R. Young, *Nature* **1995**, *374*, 280-2.

- [25] S. Ali, D. Heathcote, S. Kroll, A. Jogalekar, B. Scheiper, H. Patel, J. Brackow, A. Siwicka, M. J. Fuchter, M. Periyasamy, R. S. Tolhurst, S. K. Kanneganti, J. P. Snyder, D. C. Liotta, E. O. Aboagye, A. G. M. Barrett, R. C. Coombes, *Cancer Res.* **2009**, *69*, 6208-15.
- [26] N. Kwiatkowski, T. Zhang, P. B. Rahl, B. J. Abraham, J. Reddy, S. B. Ficarro, A. Dastur, A. Amzallag, S. Ramaswamy, B. Tesar, C. R. Jenkins, N. M. Hannett, D. McMillin, T. Sanda, T. Sim, N. D. Kim, T. Look, C. Mitsiades, A. P. Weng, J. R. Brown, C. H. Benes, J. A. Marto, R. A. Young, N. S. Gray, *Nature* **2014**, *511*, 616-20.
- [27] B.Y. Wang, Q.Y. Liu, J. Cao, J.W. Chen, Z.S. Liu, *Drug Des. Devel. Ther.* **2016**, *10*, 1181-9.
- [28] M. Ikuta, K. Kamata, K. Fukasawa, T. Honma, T. Machida, H. Hirai, I. Suzuki-Takahashi, T. Hayama, S. Nishimura, *J. Biol. Chem.* **2001**, *276*, 27548-54.
- [29] D. Pratt, J. Bentley, P. Jewsbury, F. Boyle, J. Endicott, M. Noble, *J. Med. Chem.* **2006**, *49*, 5470-7.
- [30] L. Aixiao, B. Florent, M. Francois, D. Michel, W. Baoshan, *J. Mol. Struct.: THEOCHEM* **2008**, *849*, 62-75.
- [31] H. Park, M. Yeom, S. Lee, *ChemBioChem* **2004**, *5*, 1662-72.
- [32] K. Rossi, J. Markwalder, S. Seitz, C.-H. Chang, S. Cox, M. Boisclair, L. Brizuela, S. L. Brenner, P. F. W. Stouten, *J. Comput.-Aided Mol. Des.* **2005**, *19*, 111-22.
- [33] N. Mascarenhas, D. Bhattacharyya, N. Ghoshal, *J. Mol. Graphics Modell.* **2010**, *28*, 695-706.
- [34] A. Hole, S. Baumli, H. Shao, S. Shi, S. Huang, C. Pepper, P. M. Fischer, S. Wang, J. A. Endicott, M. E. Noble, *J. Med. Chem.* **2013**, *56*, 660-70.
- [35] S. Wang, G. Griffiths, C. Midgley, A. Barnett, M. Cooper, J. Grabarek, L. Ingram, W. Jackson, G. Kontopidis, S. J. McClue, C. McInnes, J. McLachlan, C. Meades, M. Mezna, I. Stuart, M. P. Thomas, D. I. Zheleva, D. P. Lane, R. C. Jackson, D. M. Glover, D. G. Blake, P. M. Fischer, *Chem. Biol.* **2010**, *17*, 1111-21.
- [36] S. Wu, I. McNae, G. Kontopidis, S. J. McClue, C. McInnes, K. J. Stewart, S. Wang, D. I. Zheleva, H. Marriage, D. P. Lane, P. Taylor, P. M. Fischer, *Structure* **2003**, *11*, 399-410.
- [37] D. Heathcote, H. Patel, S. Kroll, P. Hazel, M. Periyasamy, M. Alikian, S. K. Kanneganti, A. S. Jogalekar, B. Scheiper, M. Barbazanges, A. Blum, J. Brackow, A. Siwicka, R. D. M. Pace, M. J. Fuchter, J. P. Snyder, D. C. Liotta, P. S. Freemont, E. O. Aboagye, R. C. Coombes, A. G. M. Barrett, S. Ali, *J. Med. Chem.* **2010**, *53*, 8508-22.
- [38] B. Zhang, V. Tan, K. Lim, T. Tay, *J. Chem. Inf. Model.* **2007**, *47*, 1877-85.

- [39] B. Zhang, V. Tan, K. Lim, T. Tay, S. Zhuang, *J. Mol. Model.* **2007**, *13*, 79-89.
- [40] S.K. Tripathi, R. Muttineni, S.K. Singh, *J. Theor. Biol.* **2013**, *334*, 87-100
- [41] M. Kaliszczak, H. Patel, S. H. Kroll, L. Carroll, G. Smith, S. Delaney, D. A. Heathcote, A. Bondke, M. J. Fuchter, R. C. Coombes, A. G. M. Barrett, S. Ali, E. O. Aboagye, *Br. J. Cancer* **2013**, *109*, 2356-67.
- [42] T. Mayhood, W. Windsor, *Anal. Biochem.* **2005**, *345*, 187-97.
- [43] P. Jeffrey, A. Russo, K. Polyak, E. Gibbs, J. Hurwitz, J. Massagué, N. P. Pavletich, *Nature* **1995**, *376*, 313-20.
- [44] G. Lolli, E. Lowe, N. Brown, L. Johnson, *Structure* **2004**, *12*, 2067-79.
- [45] A. Sali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779-815.
- [46] L. Heady, M. Fernandez-Serra, R. Mancera, S. Joyce, A. Venkitaraman, E. Artacho, C-K. Skylaris, L. Colombi Ciacchi, M. C. Payne, *J. Med. Chem.* **2006**, *49*, 5141-53.
- [47] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman, **2012**, AMBER 12, University of California, San Francisco.
- [48] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. *J. Comput. Chem.* **2005**, *26*, 1781-802.