The effect of desolvation on the binding of inhibitors to HIV-1 protease and cyclin-dependent kinases: causes of resistance

Clifford W. Fong

Eigenenergy, Adelaide, South Australia, Australia

Email: cwfong@internode.on.net

Keywords

QSAR model, HIV-1 protease inhibitors, CDK inhibitors, kinase binding, quantum mechanics

Abstract

Studies of the cyclin-dependent kinase inhibitors and HIV-1 protease inhibitors have confirmed that ligand-protein binding is dependent on desolvation effects. It has been found that a four parameter linear model incorporating desolvation energy, lipophilicity, dipole moment and molecular volume of the ligands is a good model to describe the binding between ligands and kinases or proteases. The resistance shown by MDR proteases to the anti-viral drugs is multi-faceted involving varying changes in desolvation, lipophilicity and dipole moment interaction compared to the non-resistant protease. Desolvation has been shown to be the dominant factor influencing the effect of inhibitors against the cyclin-dependent kinases, but lipophilicity and dipole moment are also significant factors. The model can differentiate between the inhibitory activity of CDK2/cycE, CDK1/cycB and CDK4/cycD enzymes.

Abbreviations

HIV Human immunodeficiency virus, CDK cyclin-dependent kinases, MDR multi-drug resistant, $\Delta G_{\text{desolvation}}$ free energy of water desolvation, $\Delta G_{\text{lipophilicity}}$ free energy of lipophilicity or hydrophobicity, $\Delta G_{\text{desolv,CDS}}$ free energy of water desolvation of the cavitation dispersion solvent structure (CDS), $\Delta G_{\text{lipo,CDS}}$ free energy of lipophilicity or hydrophobicity for the CDS, DM dipole moment DM, EA electron affinity, IE ionization energy, ϵ dielectric constant, R^2 multiple correlation coefficient, F the F test of significance, SEE standard errors for the estimates, $SE(\Delta G_{\text{desolvation}})$ standard errors of $\Delta G_{\text{desolvation}}$, $SE(\Delta G_{\text{lipophilicity}})$, standard errors of $\Delta G_{\text{lipophilicity}}$, $SE(\text{Dipole Moment})$ standard errors for dipole moments, $SE(\text{Molecular Volume})$ standard errors for molecular volumes as calculated from “t” distribution statistics

Highlights

- Novel QSAR: desolvation energy+lipophilicity+dipole moment+molecular volume
- Binding cyclin-dependent kinase inhibitors: desolvation driven
- Binding HIV-1 protease inhibitors: desolvation, lipophilicity, dipole driven
- Resistance of HIV-1 proteases

Objectives

- To examine what impact the desolvation of ligand and protein has on ligand-protein binding and resistance to binding with HIV-1 proteases and cyclin-dependent kinases
- To examine the applicability of a novel four parameter QSAR method to ligand-enzyme binding of HIV-1 proteases and cyclin-dependent tyrosine kinases

This study will use a model (equation 1 or 1a) which has been successfully applied to various transport and drug-protein binding processes. A common model for both processes based on 4 molecular physico-chemical properties of the prospective drugs has been previously applied to passive and facilitated diffusion, and active organic anion transporter drug membrane transport, some competitive statin-CYP enzyme binding processes and some tyrosine kinase inhibitors.^{1,2,3} The model comprises four main properties: (a) desolvation energy in water (b) lipophilicity or hydrophobicity based on the solvation energy in a hydrophobic solvent such as n-octane or n-octanol; lipophilicity being a measure of how well a drug can interact with lipophilic cell membrane bilayers, and hydrophobicity being a measure of non-polar interaction between a drug and the hydrophobic sectors of a protein (c) dipole moment in water, as a measure of the polar attraction between the drug and its receptor target or cell membrane (d) the molecular volume of the drug in water as a measure of how well the prospective drug fits into the cavity of the target receptor protein or active protein transporter, or how well a drug can diffuse through a cell membrane. The basic model is shown in **equation 1**.

Transport or Binding

$$= \Delta G_{\text{desolvation}} + \Delta G_{\text{lipophilicity}} + \text{Dipole Moment} \\ + \text{Molecular Volume}$$

Or

Transport or Binding

$$= \Delta G_{\text{desol,CDS}} + \Delta G_{\text{lipophilicity}} + \text{Dipole Moment} \\ + \text{Molecular Volume}$$

It has been shown that equation 1 can be successfully applied to the competitive binding of a number of tyrosine kinase inhibitors and multi-kinase inhibitors to a range of kinase enzymes. Equation 1 also applies to the active competitive transport of these tyrosine kinase inhibitors by the hOCT3, OATP1A2 and OCT1 transporters. $\Delta G_{\text{desolvation}} = \Delta G_{\text{electrostatic}} + \Delta G_{\text{CDS}}$ where CDS is the cavity dispersion solvent structure. The CDS involves non-bulk solvent electrostatic contributions to the free energy of hydration. The SMD solvation model is based on $\Delta G_{\text{S}}^{\circ} = \Delta G_{\text{ENP}} + G_{\text{CDS}}$ where ENP is the electronic nuclear polarization: the change in the solute free energy due to electrostatic interactions between the solute and the bulk solvent and distortion of the solute's electronic structure in solution. The solvent is modelled as a dielectric continuum. The CDS represents first solvation shell effects. It involves atomic surface tension (geometry dependent proportionality constants). The CDS has been parameterized using extensive experimental data sets for optimization, and has the advantage of including a realistic experimentally based hydrogen bonding model. The CDS covers *shorter-range* polarization effects and shorter-range non-electrostatic effects such as cavitation, dispersion, and solvent structural effects (which includes both hydrogen bonding) and exchange repulsion effects.⁴ $\Delta G_{\text{desol,CDS}}$ can be substituted for $\Delta G_{\text{desolvation}}$ where this non-polar term may be more representative of the interaction processes being studied, as for

example in a ligand-protein binding study within the protein binding pocket where the dielectric constant for that environment is ca. $\epsilon=7-10$ compared to $\epsilon=78.3$ for the bulk water environment where the use of $\Delta G_{\text{desolvation}}$ is more appropriate.³

It has been shown that continuum based hydrophobic non-polar hydration solvation models can fairly accurately model particular protein-protein and ligand-protein complexes, although different protein systems may have unique attractive (van der Waals) properties not precisely transferable to other systems.⁵ The hydration hydrophobic effect is primarily due to a decrease in the number of ways that favourable hydrogen bonding can be achieved by solvent water molecules when they are near to a non-hydrogen bonding solute, thereby decreasing the solvent's entropy. The hydrophobic non-polar hydration free energy $\Delta G_{\text{non-polar}}$ can be considered to comprise of the repulsive ΔG_{cavity} and attractive ΔG_{vdw} terms. The ΔG_{cavity} term includes the repulsive electrostatic effect of forming a cavity in the solvent whereby the solute interacts with the water solvent, while the ΔG_{vdw} term includes *shorter-range* polarization effects and shorter-range non-electrostatic effects such as cavitation, dispersion, and solvent structural effects (which includes both hydrogen bonding) and exchange repulsion effects.

When a solvated ligand enters a protein binding cavity and begins to bind with the protein, the desolvation processes involving the protein and ligand may involve extensive rearrangement of water molecules, and possibly some expulsion of water molecules from the cavity. A significant change in solvation entropy should then occur on the side on the ligand and the protein receptor.

In this study it is postulated that the $\Delta G_{\text{desolv,CDS}}$ values may be a close approximation of how a solvated inhibitor reacts as it approaches and starts to enter the binding pocket (dielectric constant $\epsilon=20-30$) and leaves the bulk water environment ($\epsilon=78.3$) and binding starts to occur. This scenario has been modelled using more intensive molecular mechanics computations.⁶⁻⁸

It has also been shown² that equation 1 can be modified to model redox processes (eg metabolic oxidation of statins by CYP enzymes) particularly when the molecular volume changes are small. The electron affinity EA or ionization potential IE in water are measures of the ease that drugs can be reduced or oxidised by appropriate species such as enzymes.

Redox activity

$$= \Delta G_{\text{desolvation}} + \Delta G_{\text{lipophilicity}} + \text{Dipole Moment} \\ + \text{Electron Affinity}$$

HIV-1 Protease Inhibitors

Human immunodeficiency virus (HIV) is a lentivirus and HIV-1 causes the majority of the acquired immune deficiency syndrome (AIDS) epidemic. The development of selective approved inhibitors (**Figure 1**) for the HIV-1 protease within a short period is considered a notable success of molecularly designed drug development.⁹ Despite the success of the approved inhibitors, mutation induced resistance is a major problem facing the treatment of AIDS. The causes of resistance are not well understood. This study will focus on one study¹⁰

where it was suggested that desolvation of the inhibitors was an important factor in protease resistance.

Figure 1. Structures of approved HIV-1 Protease inhibitors

Cyclin-dependent Kinase Inhibitors

Cyclin-dependent kinases (CDKs) are among the most important of the many proteins involved in cell cycle control. CDKs are a family of multifunctional enzymes that can modify various protein substrates involved in cell cycle progression. Specifically, CDKs phosphorylate their substrates by transferring phosphate groups from ATP to specific stretches of amino acids in the substrates. Cyclins are a family of proteins that have no enzymatic activity of their own but activate CDKs by binding to them. CDKs must also be in a particular phosphorylation state. CDKs are potential targets for anti-cancer therapy. If it is possible to selectively interrupt the cell cycle regulation in cancer cells by interfering with CDK action, the cell will die. (**Figure 2**) Various CDK inhibitors have been approved or are being trialled for their anti-cancer potential.^{11,12}

Figure 2. Relationships amongst cyclin-dependent kinases CDK and cell cycle phases and effect of inhibitors

(4-Butoxy-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone (SQ-67563) has been shown to be a potent, selective inhibitor of CDK1 ($IC_{50}=0.15\ \mu\text{M}$) and CDK2 ($IC_{50}=0.11\ \mu\text{M}$) in vitro which binds at the ATP site.^{13,14} In cells SQ-67563 acts as a cytotoxic agent with the ability to block cell cycle progression and/or induce apoptosis. The X-ray structure shows SQ-67563 bound to CDK2 residues coincident with the ATP purine binding pocket and forms important H-bonding interactions with Leu83 on the protein backbone via the two N atoms of the pyrazolopyridine ring. The carbonyl group does not appear to be involved in any interaction with the CDK2 protein. Phenyl groups attached to the carbonyl group engage in π stacking with the Phe80 of the protein. Although the 4 position tolerated a wide number of large lipophilic substituents it was found that the introduction of smaller polar substituents at this location did not result in any activity against the CDK inhibitors. (**Figure 3**) The 4-alkoxy substituents extends into the space occupied by the ribose of ATP and does not appear to form any specific contacts with the protein. It was thought that the energetic penalty incurred during desolvation of the polar groups compromises the overall observed binding affinity.^{13,14}

1. Applying Equation 1 to HIV-1 Protease Inhibitors

The inhibitory effect of nine approved HIV inhibitors on a WT HIV-1 protease and four multi-drug resistant (MDR) HIV-1 proteases has been previously examined. The IC_{50} values for darunavir, atazanavir, lopinavir, tipranavir, nelfinavir, amprenavir, saquinavir, indinavir, ritonavir with the NL-43 WT HIV-1 protease, and the four clinical MDR HIV-1 protease variants, MDR 769, MDR 807, MDR 1385, and MDR 3761 have been determined. The four MDR protease variants were resistant to all of the FDA-approved HIV-1 protease inhibitors at various levels, compared to the WT HIV-1 protease.¹⁰

From a molecular mechanics study of the ligand binding to the protease MDR769, it was concluded that desolvation energy plays an important role in the ligand entry and protease-ligand complex formation. The MDR proteases required a higher energy to desolvate the substrate and the protease active site. The higher desolvation barrier for MDR proteases increases the difficulty of protease-ligand complex formation. From the crystal structure and direct measurement using pulsed double electron-electron resonance (DEER), it was found that there was a larger distance between the flaps of MDR 769 compared to the wild-type protease structure. The wide-open conformation of the MDR protease in solution exposed more active site regions to solvent which may further elevate the desolvation barrier. It was found that the desolvation energy and the non-polar component (analogous to CDS) solvation energy of the HIV-1 protease-p2/NC complex for the NL-43 WT HIV-1 protease (178 and 7.6 kcal/mol), were lower than the MDR HIV-1 protease variants, MDR 769 (188 and 7.7 kcal/mol), MDR 807 (210 and 7.9 kcal/mol), MDR 1385 (228 and 7.6 kcal/mol), and MDR 3761 (256 and 7.7 kcal/mol), suggesting that resistance had some dependency on the desolvation energy of the proteases.¹⁰

The high resolution X-ray and neutron structure of a HIV-1 protease – inhibitor complex has revealed the presence of water molecules as an integral part of the protease-inhibitor complex, particularly involved in 9 bridging (mediated) hydrogen bonds out of the 16 hydrogen bonds between the KNI-272 inhibitor and the protease HIV-1. This study and other related studies indicate that a number of water molecules are present in the inhibitor binding complexes formed with the HIV-1 proteases. HIV protease inhibitors usually contain a central core motif consisting of a hydroxyethylene scaffold. A critical feature is a hydroxyl

group on the core motif which forms a hydrogen bond with the carboxylic acid on the Asp-25 and Asp-25' residues in the binding site. Hydrogen bonds are formed between a water molecule, which is linked to Ile50 and Ile50', and carbonyl groups of the inhibitors are connected with the flap regions of the protease. Specific pockets in the binding site of the HIV protease, S1,S1',S2 and S2', recognize hydrophobic amino acids on natural substrates.¹⁷ The potency of inhibitors bearing hydrophobic groups complementing these areas is therefore increased.¹⁵

Mutations that are primarily in the active site of the HIV protease enzyme alter the conformational shape facilitate resistance to protease inhibitors. The cleavage sites have highly diverse sequences, so the protease recognizes its substrates by the conserved 3D shape (substrate envelope) the substrates share when bound at the active site.¹⁶

The x-ray results^{16,17} and the protease solvation studies¹⁰ suggest that there must be some pre-organization of the water within the protease binding pocket as well as changes to the solvation shells surrounding the incoming inhibitors as they leave the bulk cellular water environment and approach and then enter the protease binding pocket. When the protease binds to its substrate and buries nearly 1000 Å² of surface area, the symmetry of the protease is broken, yet most internal hydrogen bonds and waters are conserved. However, no substrate side chain hydrogen bond is conserved. Also the substrate-protease complexes show large numbers of water molecules, and several water molecules which act as bridging entities between the inhibitor and protease.^{16,17} The specificity of HIV-1 protease appears to be determined by an asymmetric shape rather than a particular amino acid sequence.

Consideration of how the hydration shell around the inhibitor changes as the inhibitor leaves the bulk water environment, enters the protease binding pocket, and then binds to the protease is critical to understanding the role of solvation during the binding process. The dielectric constant ϵ , for the bulk water environment is ca. 78, and then changes as the inhibitor approaches the entry of the protease to ϵ of ca. 20-30, then to a ϵ of ca. 6-10 within the protease environment.^{18,19} It is also clear that hydrophobic interactions between the S1,S1',S2 and S2' protease sites and the hydrophobic regions of inhibitors is also an important factor. In the analysis below, $\Delta G_{\text{solvation}}$ is considered to represent the inhibitor in the bulk water environment, the non-polar contribution to solvation, $\Delta G_{\text{desol,CDS}}$ is considered to represent the solvation shell around the inhibitor as it binds to the enzyme, $\Delta G_{\text{lipophilicity}}$ (the free energy of solvation in n-octane) represents the hydrophobic interaction between inhibitor and the enzyme, the dipole moment in water is taken to represent the polar interaction between the inhibitor and solvent and enzyme, and the molecular volume in water is taken to represent the molecular size of the solvated inhibitor.

The desolvation or dehydration, lipophilicity, dipole moment and molecular volumes of the various inhibitors are shown in the Table, and these values were used in deriving the following equations.

IC₅₀ values for darunavir, atazanavir, lopinavir, tipranavir, nelfinavir, amprenavir, saquinavir, indinavir, ritonavir with the NL-43 WT HIV-1 protease

Eq 1

$$IC_{50} = -0.07 \Delta G_{\text{desol,CDS}} - 0.12 \Delta G_{\text{lipophilicity}} - 0.085 \text{ Dipole Moment} - 0.005 \text{ Molecular Volume} + 1.77$$

Where $R^2 = 0.884$, $SEE = 0.209$, $SE(\Delta G_{\text{desolv,CDS}}) = 0.03$, $SE(\Delta G_{\text{lipophilicity}}) = 0.02$, $SE(\text{Dipole Moment}) = 0.04$, $SE(\text{Molecular Volume}) = 0.001$, $F=7.62$, Significance $F = 0.036$

IC_{50} values for darunavir, atazanavir, lopinavir, tipranavir, nelfinavir, amprenavir, saquinavir, indinavir, ritonavir with the NL-43 WT HIV-1 protease **Eq 1a**

$$IC_{50} = 0.02 \Delta G_{\text{desolvation}} - 0.10 \Delta G_{\text{lipophilicity}} - 0.02 \text{ Dipole Moment} - 0.004 \text{ Molecular Volume} + 13.420$$

Where $R^2 = 0.774$, $SEE = 0.29$, $SE(\Delta G_{\text{desolvation}}) = 0.02$, $SE(\Delta G_{\text{lipophilicity}}) = 0.035$, $SE(\text{Dipole Moment}) = 0.046$, $SE(\text{Molecular Volume}) = 0.003$, $F=3.41$, Significance $F = 0.130$

$\Delta G_{\text{binding}}$ from isothermal calorimetry for darunavir, atazanavir, lopinavir, tipranavir, nelfinavir, saquinavir, indinavir, ritonavir binding to HIV-1 protease (Freire 2008) **Eq 1b**

$$\Delta G_{\text{binding}} = -0.06 \Delta G_{\text{desol,CDS}} - 0.085 \Delta G_{\text{lipophilicity}} - 0.02 \text{ Dipole Moment} - 0.003 \text{ Molecular Volume} - 16.85$$

Where $R^2 = 0.852$, $SEE = 0.613$, $SE(\Delta G_{\text{desolv,CDS}}) = 0.10$, $SE(\Delta G_{\text{lipophilicity}}) = 0.03$, $SE(\text{Dipole Moment}) = 0.10$, $SE(\text{Molecular Volume}) = 0.003$, $F=4.30$, Significance $F = 0.13$

It can be seen that eq 1, 1a and 1b are very similar. Eq 1 and 1a are IC_{50} correlations from Wang¹⁰ where the inhibitors were the neutral species, and the $\Delta G_{\text{binding}}$ isothermal calorimetry values (kcal/mol) are from Freire^{10a}, where nelfinavir, saquinavir, indinavir, ritonavir were zwitterionic species which gave the best correlation. Clearly the HIV-1 proteases were different in each study, and experimental conditions were also different. The similarity of the three equations indicate that the general form of the four parameter equation eq 1 is substantially robust.

IC_{50} values for darunavir, atazanavir, lopinavir, tipranavir, nelfinavir, amprenavir, saquinavir, indinavir, ritonavir with the MDR-769 HIV-1 protease **Eq 2**

$$IC_{50} = 8.58 \Delta G_{\text{desol,CDS}} - 1.28 \Delta G_{\text{lipophilicity}} + 19.20 \text{ Dipole Moment} + 0.38 \text{ Molecular Volume} - 464.37$$

Where $R^2 = 0.911$, $SEE = 41.05$, $SE(\Delta G_{\text{desolv,CDS}}) = 6.88$, $SE(\Delta G_{\text{lipophilicity}}) = 2.24$, $SE(\text{Dipole Moment}) = 6.33$, $SE(\text{Molecular Volume}) = 0.22$, $F=10.22$, Significance $F = 0.022$

IC_{50} values for darunavir, atazanavir, lopinavir, tipranavir, nelfinavir, amprenavir, saquinavir, indinavir, ritonavir with the MDR-807 HIV-1 protease **Eq 3**

$$IC_{50} = 42.54 \Delta G_{\text{desol,CDS}} + 0.30 \Delta G_{\text{lipophilicity}} + 68.06 \text{ Dipole Moment} + 0.64 \text{ Molecular Volume} - 1410.89$$

Where $R^2 = 0.914$, $SEE = 117.0$, $SE(\Delta G_{\text{desolv,CDS}}) = 19.62$, $SE(\Delta G_{\text{lipophilicity}}) = 6.40$, $SE(\text{Dipole Moment}) = 18.04$, $SE(\text{Molecular Volume}) = 0.65$, $F=10.63$, Significance $F = 0.020$

IC_{50} values for darunavir, atazanavir, lopinavir, tipranavir, nelfinavir, amprenavir, saquinavir, indinavir, ritonavir with the MDR-1385 HIV-1 protease **Eq 4**

$$IC_{50} = -11.22 \Delta G_{\text{desol,CDS}} - 10.58 \Delta G_{\text{lipophilicity}} - 16.55 \text{ Dipole Moment} - 0.39 \text{ Molecular Volume} = 225.32$$

Where $R^2 = 0.826$, $SEE = 48.29$, $SE(\Delta G_{\text{desolv,CDS}}) = 8.09$, $SE(\Delta G_{\text{lipophilicity}}) = 2.64$, $SE(\text{Dipole Moment}) = 7.44$, $SE(\text{Molecular Volume}) = 0.27$, $F=4.75$, Significance $F = 0.08$

IC_{50} values for darunavir, atazanavir, lopinavir, tipranavir, nelfinavir, amprenavir, saquinavir, indinavir, ritonavir with the MDR-3761 HIV-1 protease **Eq 5**

$$IC_{50} = -14.18 G_{\text{desol,CDS}} - 17.543 \Delta G_{\text{lipophilicity}} - 23.23 \text{ Dipole Moment} - 0.58 \text{ Molecular Volume} + 253.91$$

Where $R^2 = 0.781$, $SEE = 91.60$, $SE(\Delta G_{\text{desolv,CDS}}) = 16.36$, $SE(\Delta G_{\text{lipophilicity}}) = 5.33$, $SE(\text{Dipole Moment}) = 15.05$, $SE(\text{Molecular Volume}) = 0.54$, $F=3.57$, Significance $F = 0.12$

It is apparent that the MDR proteases show very different behaviours than the WT protease, and MDR-769 and 807 are also quite different from MDR-1385 and 3761. The MDR proteases interact with the inhibitors with greater sensitivities to all four dependent variables than the wild type protease. This is consistent with Wang's observation that these latter two proteases have the higher solvation energies compared to MDR-769 and 807 (and the WT). The non-polar desolvation energies (CDS) for the proteases are 7.6 to 7.9 kcal/mol, which compares to a ligand range of 7.6 kcal/mol (ritonavir) to 17.7 kcal/mol (atazanavir). These data suggest that water solvent desolvation or solvent rearrangement energies are greater for the ligand than the receptor protein during binding, a key pointer regarding future drug development in these systems. It is noted that eqs 2-5 used the zwitterionic form for nelfinavir, saquinavir, indinavir, ritonavir since these species were large outliers from the eqs when included as neutral species, but gave consistently high correlations for all four MDR equations. Eqs 1 and 1a used the neutral form of the inhibitor species.

A common distinguishing feature is that all of the MDR proteases have much higher dependences on the dipole moment than the WT protease. Also molecular size does not seem to be a critical factor. The MDR-769 and 807 show a different class of interaction with the anti-viral drugs than MDR-1385 and 3761, particularly regarding their sensitivity to water desolvation, and lipophilicity. It would appear that the interactions with MDR-1385 and 3761 are driven by a more balanced non-polar, hydrophobic (or lipophilic) plus a hydrophilic interaction with the anti-virals than MDR-769 and 807 (and the WT protease even more so) which is more driven by hydrophilic interactions.

In summary, the resistance shown by MDR proteases to the anti-viral drugs is multi-faceted involving varying changes in desolvation, lipophilicity and dipole moment interaction compared to the non-resistant WT protease.

2. Applying Equation 1 to Cyclin-dependent Kinase Inhibitors

A study has been made of the inhibitory activity against CDK2/cycE by the series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone (Figure 3) where RO was the alkoxy group with R being n-propyl ($IC_{50}=0.18 \mu\text{M}$), n-butyl (0.11), 1-methylbutyl (0.24), 3-methylbutyl (0.18), n-pentyl (0.52), 2-hydroxyethyl (3.7), 2-dimethylaminoethyl (17), cyclohexyl (0.66), benzyl (0.28), phenyl (15), and (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone where R was n-butyl and phenyl-R was 2,6-difluorophenyl, 2,6-difluoro-4-methoxyphenyl, 2,4,6-trimethylphenyl, 2-fluorophenyl and 2-methylphenyl.^{13,14} [Misra 2003]

Figure 3. CDK inhibitors, (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone showing alkoxy (R = n-butyl) substituents, and (4-butoxy-1H-pyrazolo[3,4-b]pyridin-5-

yl)(phenyl-R)methanone illustrated with 2,6-difluoro-4-methyl substituents on the phenyl-R ring.

IC₅₀ values (μM) for CDK2/cycE of the inhibitor series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone where RO is the alkoxy group with R being n-propyl (0.18), n-butyl (0.11), 1-methylbutyl (0.24), 3-methylbutyl (0.18), n-pentyl (0.52), 2-hydroxyethyl (3.7), 2-dimethylaminoethyl (17), cyclohexyl (0.66), benzyl (0.28), phenyl (15), and (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone, where R was n-butyl and phenyl-R was R being n-butyl with phenyl-R being 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036) and 2-methylphenyl (0.073). R = phenyl omitted as outlier.

Eq 6a

$$IC_{50} = -3.87 \Delta G_{\text{desolv,CDS}} + 1.89 \Delta G_{\text{lipophilicity}} + 0.76 \text{ Dipole Moment} + 0.02 \text{ Molecular Volume} + 35.12$$

Where R² = 0.742, SEE = 2.42, SE(ΔG_{desolv,CDS}) = 0.67, SE(ΔG_{lipophilicity}) = 0.90, SE(Dipole Moment) = 0.54, SE(Molecular Volume) = 0.02, F = 8.61, Significance F = 0.0015, n=16

IC₅₀ values (μM) for CDK2/cycE of the inhibitor series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone where RO is the alkoxy group with R being n-propyl (0.18), n-butyl (0.11), 1-methylbutyl (0.24), 3-methylbutyl (0.18), n-pentyl (0.52), 2-hydroxyethyl (3.7), 2-dimethylaminoethyl (17), cyclohexyl (0.66), benzyl (0.28), phenyl (15), and (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone where R was n-butyl and phenyl-R was R being n-butyl with phenyl-R being 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036) and 2-methylphenyl (0.073). R = phenyl and 2-dimethylaminoethyl omitted as outliers.

Eq 6b

$$IC_{50} = -0.42 \Delta G_{\text{desolvation}} + 0.27 \Delta G_{\text{lipophilicity}} - 0.05 \text{ Dipole Moment} + 0.006 \text{ Molecular Volume} - 2.57$$

Where R² = 0.753, SEE = 0.52, SE(ΔG_{desolvation}) = 0.08, SE(ΔG_{lipophilicity}) = 0.19, SE(Dipole Moment) = 0.12, SE(Molecular Volume) = 0.006, F=8.38 Significance F = 0.002, n=16

Equations 6a and 6b are similar, with 6a being the better correlation. The differences between eq 6a and eq 6b reflects the stronger influence of the electrostatic bulk dehydration effect ΔG_{desolvation} in eq 6b compared to the non-polar CDS (as in the ΔG_{desolv,CDS} environment) in eq 6a on the inhibitory effect. As suggested in the introduction, the non-polar CDS parameter is a better representation of binding interaction with the kinase within the kinase binding pocket, where ε=6-10 compared to ε=78.3 for the bulk water environment, or ε=20-30 at the entry of the kinase binding pocket. Equations 6 and 6a show only small dependencies on dipole and molecular size, with water desolvation being the major influence on inhibitory binding, followed by a smaller dependency on lipophilicity. The phenoxy inhibitor is a strong outlier from both equations, and probably is attributable to a π stacking interaction or a steric clash with the kinase. The X-ray structure of the 4-butoxy substituent shows that the n-butyl group extends into the space occupied by the ribose of ATP and does not appear to form any specific contacts with the protein. The 2-dimethylaminoethyl groups is an outlier from eq 6a as well as the phenoxy group, which appears to indicate that some solvation around the –CH₂NMe₂ group is significant and causes some interaction with the protein residues compared to the non-polar n-butyl group which shows no protein interaction. It is possible

that the 2-dimethylaminoethyl inhibitor shows a larger deviation from the correlation in eq 6a, reflecting a stronger electrostatic contribution to hydration in eq 6b than the non-polar CDS hydration term in eq 6a (see eq 8 and 8a below).

IC₅₀ values (μM) for CDK2/cycE of the inhibitor series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone where RO is the alkoxy group with R being n-propyl (0.18), n-butyl (0.11), 1-methylbutyl (0.24), 3-methylbutyl (0.18), n-pentyl (0.52), 2-hydroxyethyl (3.7), 2-dimethylaminoethyl (17), cyclohexyl (0.66), benzyl (0.28), phenyl (15). R = phenyl omitted as outlier.

Eq 7

$$IC_{50} = -5.52 \Delta G_{\text{desolv,CDS}} - 0.96 \Delta G_{\text{lipophilicity}} + 0.53 \text{ Dipole Moment} - 0.02 \text{ Molecular Volume} + 15.62$$

Where R² = 0.905, SEE = 2.41, SE(ΔG_{desolv,CDS}) = 0.94, SE(ΔG_{lipophilicity}) = 2.13, SE(Dipole Moment) = 1.10, SE(Molecular Volume) = 0.016, F = 9.52, Significance F = 0.024

Eq 7 only includes inhibitors where R is varied, and does not include any substituted phenyl-R compounds. Again the R = phenyl is a large outlier. The strong influence of water desolvation as found in eq 6 is indicative of the rearrangement of water molecules in the inner solvent shell of the inhibitors primarily around the N atoms of the pyrazole ring (the main hydrogen and polar binding site with the CDK). The variation in R groups has only a minor effect as manifested by the small changes to the ΔG_{lipophilicity} term in equations 6 and 7. The small ΔG_{lipophilicity} terms suggest little hydrophobic interaction between the R groups and the CDK protein. This is consistent with the x-ray structure which shows little interaction between any of the R groups and the CDK. The x-ray structure of the (4-Butoxy-1H-pyrazolo[3,4-b]pyridin-5-yl)(2,6 difluorophenyl)methanone shows that the 4-butoxy substituent extends into the space occupied by the ribose of ATP and does not appear to form any specific contacts with the protein while the pendant difluorophenyl ring lies buried within the protein, π stacking with Phe80.^{13,14}

It is also clear that the pendant phenyl-R ring attached to the carbonyl group needs to be almost coplanar with the C=O group to allow in π stacking with the Phe80 of the protein. The minimal energy conformation of the 2,4,6-trimethylphenyl group attached to the carbonyl group is shown to be rotated almost orthogonally to the C=O plane, which prevents π stacking, and manifest itself in a very weak binding ability to the CDK. This is also consistent with the x-ray structure.^{13,14}

It has been shown that substitution of the 1-H on the N1 of the pyrazole ring by an ethyl group in the inhibitor (3-methylbutyloxy-1-ethyl-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone leads to a very large lowering of activity, IC₅₀ > 25 compared to the 1-H pyrazole parent.^{13,14} [Misra 2003] This change can obviously affect inhibitor hydration and lipophilicity simultaneously. Using eq 7 (which considers only the variation of activity of the 4-alkoxy inhibitors, and not inhibitor-CDK2 interaction via the pyrazolopyridine moiety) leads to a prediction that the IC₅₀ value is about 45. Since the hydrogen on N1 of the pyrazolopyridine ring serves as a H-bond donor to the carbonyl oxygen of Leu83 in the crystal structure of the complex of (4-n-butoxy-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone with CDK2,^{13,14} then the substitution of the 1-H on N1 by an 1-ethyl group prevents binding to the CDK2.

A more detailed thermochemical examination of the hydration of the CDK inhibitors has revealed that there is a linear relationship between ΔG_{desolv,CDS} and ΔG_{desolvation} and ΔH_{desolvation} for 13 inhibitors. Quantitative structural activity relationships (QSAR) based on linear free

energy changes in a series of active substrates can be observed if : (a) ΔH is linearly related to ΔS for the series, or (b) if ΔH is constant for the series, or (c) if ΔS is constant for the series. The non-polar hydration term $\Delta G_{\text{desolv,CDS}}$ can be considered to comprise of the repulsive ΔG_{cavity} and attractive ΔG_{vdw} terms, and in this series, the ΔG_{cavity} term should be proportional to changes in molecular volume, and ΔG_{vdw} proportional to changes in interaction with the kinase. All structures were optimized in water in this study to allow the close examination of hydration effects on these inhibitors.

Hydration study of the inhibitor series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone where RO is the alkoxy group with R being n-propyl (0.18), n-butyl (0.11), 1-methylbutyl (0.24), 3-methylbutyl (0.18), n-pentyl (0.52), 2-hydroxyethyl (3.7), 2-dimethylaminoethyl (17), cyclohexyl (0.66), benzyl (0.28), phenyl (15), and (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone where R was n-butyl and phenyl-R was and R being n-butyl with phenyl-R being 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methoxyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036) and 2-methylphenyl (0.73), R = 2-dimethylaminoethyl omitted as outlier.

Eq 8

$$\Delta G_{\text{desolv,CDS}} = 0.239 \Delta H_{\text{desolvation}} + 8.425$$

Where $R^2 = 0.315$, SEE = 0.686, $\text{SE}(\Delta G_{\text{solvation}}) = 0.088$, $F=7.36$, Significance F = 0.015

and

Eq 8a

$$\Delta G_{\text{desolvation}} = 0.205 \Delta H_{\text{desolvation}} + 7.906$$

Where $R^2 = 0.243$, SEE = 0.722, $\text{SE}(\Delta G_{\text{solvation}}) = 0.090$, $F=5.13$, Significance F = 0.037

Equations 8 and 8a show a linear relationship between the $\Delta G_{\text{desolv,CDS}}$ for 18 CDK inhibitors in water and $\Delta G_{\text{desolvation}}$ and $\Delta H_{\text{desolvation}}$ indicating that a proportionality exists between the polar electrostatic and non-polar solvent interactions for these inhibitors. Eq 8 and 8a are virtually identical which reflects the small contribution by T ΔS in this series of CDK inhibitors. The correlations are highly significant, despite low R^2 values which reflect the low slopes (as correlation coefficients are sensitive to the value of the slopes, so the SEE are better statistical indicators). However the (4-dimethylaminoethoxy-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone inhibitor was a clear outlier, which reflected a large T ΔS of -3.7 kcal/mol at 298K, which compares to an average |0.64| kcal/mol for the other 18 inhibitors. This observation may reflect a hydrogen bonding with the dimethylaminoethyl group, which does extend into the space occupied by the ribose of ATP when complexed with the CDK, unlike the other alkoxy substituents which have alkane-oxy groups.

Comparison of the same 15 CDK inhibitor activities against CDK2/cycE and CDK1/cycB, (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone where RO is the alkoxy group with R being n-propyl (0.18), n-butyl (0.11), 1-methylbutyl (0.24), 3-methylbutyl (0.18), n-pentyl (0.52), cyclohexyl (0.66), benzyl (0.28), and R being n-butyl with phenyl-R being 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methoxyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036) and 2-methylphenyl (0.73).

CDK2/cycE inhibitors

Eq 9

$$IC_{50} = -0.290 \Delta G_{\text{desolv,CDS}} + 0.035 \Delta G_{\text{lipophilicity}} + 0.002 \text{ Dipole Moment} + 0.008 \text{ Molecular Volume} + 0.42$$

Where $R^2 = 0.830$, SEE = 0.117, $\text{SE}(\Delta G_{\text{desolv,CDS}}) = 0.060$, $\text{SE}(\Delta G_{\text{lipophilicity}}) = 0.051$, $\text{SE}(\text{Dipole Moment}) = 0.029$, $\text{SE}(\text{Molecular Volume}) = 0.001$, $F = 12.20$, Significance F = 0.0007

CDK1/cycB inhibitors

Eq10

$$IC_{50} = -0.235 \Delta G_{\text{desolv,CDS}} + 0.010 \Delta G_{\text{lipophilicity}} + 0.010 \text{ Dipole Moment} + 0.004 \text{ Molecular Volume} + 0.52$$

Where $R^2 = 0.851$, $SEE = 0.083$, $SE(\Delta G_{\text{desolv,CDS}}) = 0.042$, $SE(\Delta G_{\text{lipophilicity}}) = 0.036$, $SE(\text{Dipole Moment}) = 0.026$, $SE(\text{Molecular Volume}) = 0.001$, $F = 14.28$, Significance $F = 0.0004$

Activity of 7 CDK inhibitors against CDK4/cycD and CDK1/cycB: (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone where R was n-butyl and phenyl-R was 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methoxyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036).

CDK4/cycD inhibitors

Eq11

$$IC_{50} = -25.35 \Delta G_{\text{desolv,CDS}} + 13.08 \Delta G_{\text{lipophilicity}} + 3.86 \text{ Dipole Moment} + 0.37 \text{ Molecular Volume} + 201.0$$

Where $R^2 = 0.857$, $SEE = 7.10$, $SE(\Delta G_{\text{desolv,CDS}}) = 7.83$, $SE(\Delta G_{\text{lipophilicity}}) = 8.62$, $SE(\text{Dipole Moment}) = 2.95$, $SE(\text{Molecular Volume}) = 0.36$, $F = 3.00$, Significance $F = 0.266$

CDK1/cycB inhibitors

Eq12

$$IC_{50} = -0.31 \Delta G_{\text{desolv,CDS}} + 0.12 \Delta G_{\text{lipophilicity}} + 0.02 \text{ Dipole Moment} + 0.007 \text{ Molecular Volume} + 1.65$$

Where $R^2 = 0.837$, $SEE = 0.090$, $SE(\Delta G_{\text{desolv,CDS}}) = 0.10$, $SE(\Delta G_{\text{lipophilicity}}) = 0.10$, $SE(\text{Dipole Moment}) = 0.035$, $SE(\text{Molecular Volume}) = 0.04$, $F = 2.56$, Significance $F = 0.300$

Comparison of the highly significant eq 9 and eq10 show very similar activity dependencies for CDK2/cycE and CDK1/cycB, suggesting that these kinases have very similar binding behaviours and properties. The data set for the CDK4/cycD inhibitors is restricted to those where substitution of the pendant phenyl-R ring of the inhibitors has been made. Eq 11 and eq 12, are much less statistically significant than eqs 9 and 10, but show that the CDK4/cycD has a lower sensitivity to the 7 inhibitors, but the relative proportionate sensitivities to $\Delta G_{\text{desolv,CDS}}$ and $\Delta G_{\text{lipophilicity}}$ are similar, but the sensitivities to dipole and molecular size of the inhibitors are lower for the CDK4/cycD enzyme.

Modifying general eq 1 to incorporate redox properties

IC_{50} values (μM) for **CDK2/cycE** of the inhibitor series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone where RO is the alkoxy group with R being n-propyl (0.18), n-butyl (0.11), 1-methylbutyl (0.24), 3-methylbutyl (0.18), n-pentyl (0.52), 2-hydroxyethyl (3.7), 2-dimethylaminoethyl (17), cyclohexyl (0.66), benzyl (0.28), phenyl (15), and (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone, where R was n-butyl and phenyl-R was R being n-butyl with phenyl-R being 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methoxyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036) and 2-methylphenyl (0.73). R = phenyl omitted as outlier.

Eq 13a

$$IC_{50} = -0.17 \Delta G_{\text{desolv,CDS}} - 0.00 \Delta G_{\text{lipophilicity}} + 0.01 \text{ Dipole Moment} - 0.73 \text{ Electron Affinity} + 2.58$$

Where $R^2 = 0.453$, $SEE = 0.22$, $SE(\Delta G_{\text{desolv,CDS}}) = 0.10$, $SE(\Delta G_{\text{lipophilicity}}) = 0.09$, $SE(\text{Dipole Moment}) = 0.05$, $SE(\text{Electron Affinity}) = 0.55$, $F = 2.07$, Significance $F = 0.15$, $no=15$

and

Eq 13b

$$IC_{50} = -0.17 \Delta G_{\text{desolv,CDS}} - 0.00 \Delta G_{\text{lipophilicity}} + 0.03 \text{ Dipole Moment} - 0.45 \text{ Ionization Energy} + 3.80$$

Where $R^2 = 0.388$, $SEE = 0.22$, $SE(\Delta G_{\text{desolv,CDS}}) = 0.11$, $SE(\Delta G_{\text{lipophilicity}}) = 0.10$ $SE(\text{Dipole Moment}) = 0.05$, $SE(\text{Ionization Energy}) = 0.68$, $F = 1.58$, $\text{Significance } F = 0.25$, $no=15$

Eqs 13a and 13b show correlations using the electron affinity (EA) and ionization energy (IE) instead of the molecular volume term used in eqs 6a and 6b (where the coefficients are very small, and can be ignored) to probe whether redox reactions may be significant in the binding interactions of CDK2/cycE with the inhibitors. The correlations in eq 13a and 13b are of moderate significance, but do show a relationship between IC_{50} and the redox properties, and the dominance of $\Delta G_{\text{desolv,CDS}}$. These observations suggest that redox reactions between the CDK2/cycE and inhibitors could involve non-polar water solvation effects.

IC_{50} values (μM) for **CDK1/cycB** of the inhibitor series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone where RO is the alkoxy group with R being n-propyl (0.18), n-butyl (0.11), 1-methylbutyl (0.24), 3-methylbutyl (0.18), n-pentyl (0.52), 2-hydroxyethyl (3.7), 2-dimethylaminoethyl (17), cyclohexyl (0.66), benzyl (0.28), phenyl (15), and (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone, where R was n-butyl and phenyl-R was R being n-butyl with phenyl-R being 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methoxyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036) and 2-methylphenyl (0.73). R = phenyl omitted as outlier.

Eq 14a

$$IC_{50} = -0.17 \Delta G_{\text{desolv,CDS}} - 0.01 \Delta G_{\text{lipophilicity}} + 0.01 \text{ Dipole Moment} - 0.49 \text{ Electron Affinity} + 1.96$$

Where $R^2 = 0.659$, $SEE = 0.125$, $SE(\Delta G_{\text{desolv,CDS}}) = 0.06$, $SE(\Delta G_{\text{lipophilicity}}) = 0.05$ $SE(\text{Dipole Moment}) = 0.03$, $SE(\text{Electron Affinity}) = 0.33$, $F = 4.83$, $\text{Significance } F = 0.019$, $no=15$

and

Eq 14b

$$IC_{50} = -0.16 \Delta G_{\text{desolv,CDS}} - 0.00 \Delta G_{\text{lipophilicity}} + 0.02 \text{ Dipole Moment} - 0.54 \text{ Ionization Energy} + 4.425$$

Where $R^2 = 0.652$, $SEE = 0.127$, $SE(\Delta G_{\text{desolv,CDS}}) = 0.06$, $SE(\Delta G_{\text{lipophilicity}}) = 0.058$ $SE(\text{Dipole Moment}) = 0.03$, $SE(\text{Ionization Energy}) = 0.38$, $F = 4.68$, $\text{Significance } F = 0.021$, $no=15$

It can be seen that eqs 13a and 13b are almost identical to eqs 14a and 14b, suggesting that the redox properties of the CDK2/cycE are very similar to the CDK1/cycB complexes.

IC_{50} values (μM) for **CDK4/cycD** of the inhibitor series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone, where R was n-butyl and phenyl-R was R being n-butyl with phenyl-R being 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methoxyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036)

Eq 15a

$$IC_{50} = -23.62 \Delta G_{\text{desolv,CDS}} + 22.79 \Delta G_{\text{lipophilicity}} + 13.47 \text{ Dipole Moment} + 97.82 \text{ Electron Affinity} + 113.22$$

Where $R^2 = 0.859$, $SEE = 7.05$, $SE(\Delta G_{\text{desolv,CDS}}) = 6.89$, $SE(\Delta G_{\text{lipophilicity}}) = 8.79$ $SE(\text{Dipole Moment}) = 7.71$, $SE(\text{Electron Affinity}) = 91.89$, $F = 3.034$ $\text{Significance } F = 0.26$, $no=7$

and

Eq 15b

$$IC_{50} = -23.70 \Delta G_{\text{desolv,CDS}} + 18.47 \Delta G_{\text{lipophilicity}} + 8.76 \text{ Dipole Moment} + 45.59 \text{ Ionization Energy} - 1.36$$

Where $R^2 = 0.818$, $SEE = 8.00$, $SE(\Delta G_{\text{desolv,CDS}}) = 8.61$, $SE(\Delta G_{\text{lipophilicity}}) = 8.89$ $SE(\text{Dipole Moment}) = 5.42$, $SE(\text{Ionization Energy}) = 74.45$, $F = 2.25$, $\text{Significance } F = 0.331$, $no=7$

IC_{50} values (μM) for CDK2/cycE of the inhibitor series (4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone, where R was n-butyl and phenyl-R was R being n-butyl with phenyl-R being 2,6-difluorophenyl (0.064), 2,4,6-trifluorophenyl (0.036), 2,6-difluoro-4-methoxyphenyl (0.009), 2,6-difluoro-4-methoxyphenyl (0.022), 2,6-difluoro-4-chlorophenyl (0.027), 2,6-difluoro-4-bromophenyl (0.020), 2-fluorophenyl (0.036) **Eq 16**

$$IC_{50} = -0.03 \Delta G_{\text{desolv,CDs}} + 0.045 \Delta G_{\text{lipophilicity}} + 0.03 \text{ Dipole Moment} + 0.35 \text{ Electron Affinity} - 0.22$$

Where $R^2 = 0.813$, $SEE = 0.013$, $SE(\Delta G_{\text{desolv,CDs}}) = 0.012$, $SE(\Delta G_{\text{lipophilicity}}) = 0.016$, $SE(\text{Dipole Moment}) = 0.014$, $SE(\text{Electron Affinity}) = 0.017$, $F = 2.18$ Significance $F = 0.33$, $no=7$

The correlations in eq 15a and 15b appear to show that the redox behaviour of inhibitor inhibition of the CDK4/cycD complex is quite different from the CDK2/cycE and CDK1/cycB complexes. However these eqs 15 have a limited number of data points compared to eq 13 and 14, but the main difference seems to be a greater sensitivity to lipophilicity and dipole moment in eqs 15 compared to eqs 13 and 14. Eq 16 provides a direct comparison with eq 15a for the same data points, and it can be seen that the coefficients in eq 15a are far more sensitive to non-polar water desolvation energy, lipophilicity and dipole moment relative to the sensitivity to EA. *It is known that redox potentials oscillate during the cell cycle, passing from a reductive to oxidative environment as the cycle progresses, and the redox potential increases in abnormal cell division such as apoptosis and necrosis, but decreases in normal cell division.*²⁰ (Figure 4) There is evidence that the CDK4/cycD has a strong reduction potential in the early G phase of the cell cycle, a weaker reduction potential later in the G phase or early S phase, but in the G2 to M phase of the cycle, CDK1/cycD has a strong oxidation potential. Eqs 15 and 13/14 are consistent with these known facets of cell cycle redox processes, and may be useful indicators of differences in cell cycle behaviour by allowing separation of desolvation, lipophilicity and dipole from electron transfer processes.

Figure 4. Schematic of two cell cycles showing parent and daughter generations illustrating the oscillating redox changes as a function of the cell cycle stages, and regulation of the cycle phases by cyclins and cyclin-dependent kinases. Rb is the retinoblastoma protein which releases the transcription factor E2F, which regulates transcription of the S-phase specific genes. The vertical bar graph illustrates the full range of the redox cycle in millivolts and shows some important biological redox couples such as the glutathione GSSG/GSH couple. The bottom graph illustrates typical redox potentials for the various cell cycles. (adapted from Sarsour 2009)

All correlations using eq 1 and its variants in this study were conducted using identical conditions.²¹ The Table lists the various drugs and their biophysical parameters that were used in all equations derived above.

Table

Anti-viral drugs	$\Delta G_{\text{desolvation}}$ kcal/mol (CDS) Water	$\Delta G_{\text{lipophilicity}}$ kcal/mol n-Octane	Dipole Moment D Water	Molecular Volume cm³/mol Water	Electron Affinity (Ionization Energy) eV
Darunavir	37.7 (-10.1)	-21.0	12.4	369	
Atazanavir	39.0 (-17.7)	-24.8	6.1	469	
Lopinavir	32.2 (-13.6)	-26.85	7.7	486	
Tipranavir	27.7 (-11.9)	-20.3	10.3	426	
Nelfinavir	22.0 (-11.3)	-30.4	8.4	410	
Amprenavir	33.7 (-11.1)	-20.0	8.0	365	
Saquinavir	41.3 (-12.0)	-23.9	6.1	492	
Indinavir	41.2 (-8.4)	-25.9	12.8	458	
Ritonavir	35.1 (-7.6)	-22.8	7.9	497	
4-RO-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone					
R= n-propyl	14.6 (-4.5)	-12.1	4.0	188	2.00 (6.16)
R=n-butyl	12.3 (-4.7)	-12.4	5.3	197	2.15 (6.33)
R=1-methylbutyl	12.1 (-5.1)	-12.3	5.5	238	2.14 (6.16)
R=3-methylbutyl	12.45 (-5.1)	-12.45	5.7	202	2.20 (6.34)
R=n-pentyl	14.2 (-4.9)	-13.0	3.1	212	2.00 (6.19)
R=2-hydroxyethyl	19.0 (-4.2)	-12.2	3.1	188	2.00 (6.13)
R=2-dimethylaminoethyl	16.5 (-2.0)	-12.2	5.0	209	2.16 (5.87)
R=cyclohexyl	12.5 (-4.1)	-12.7	4.9	238	2.11 (6.27)
R=benzyl	13.5 (-5.1)	-14.1	5.3	252	2.07 (6.27)
R=phenyl	14.3 (-4.7)	-13.4	4.0	240	2.07 (6.23)
4-n-Butoxy-1H-pyrazolo[3,4-b]pyridin-5-yl)(phenyl-R)methanone					
phenyl-R=2,6-difluorophenyl	11.7 (-5.8)	-11.35	5.5	228	2.22 (6.42)
phenyl-R=2,6-difluoro-4-methyphenyl	10.1 (-6.2)	-11.7	7.0	223	2.06 (6.27)
phenyl-R=2,6-difluoro-4-chlorophenyl	11.34 (-5.9)	-11.65	3.8	222	2.31 (6.43)
phenyl-R=2,6-difluoro-4-bromophenyl	11.5 (-5.15)	-12.25	3.9	204	2.30 (6.43)

phenyl-R=2,6-difluoro-4-methoxyphenyl	12.55 (-7.2)	-11.6	8.45	269	1.97 (6.24)
phenyl-R=2,4,6-trifluorophenyl	10.9 (-6.45)	-10.7	4.3	239	2.22 (6.42)
phenyl-R=2-methylphenyl	11.8 (-5.0)	-12.1	4.8	272	1.95 (6.22)
phenyl-R=2-fluorophenyl	11.2 (-5.1)	-11.6	5.5	226	2.11 (6.24)
phenyl-R=2,4,6-trimethylphenyl	11.4 (-5.7)	-12.65	4.4	231	
4-3-Methylbutyloxy-1-ethyl-pyrazolo[3,4-b]pyridin-5-yl)(phenyl)methanone	9.65 (-6.1)	-13.55	5.65	265	

In summary, eq 6a, 6b and 7 are consistent with desolvation being the dominant factor influencing the inhibitory effect of these drugs against the CDK2/cycE. With one exception, the desolvation required during complexation between the inhibitors and CDK2/cycB is enthalpy driven, and not entropy driven. Comparison of inhibitory activity for the CDK2/cycE and CDK1/cycB show very similar behaviours, indicating that CDK2/cycB and CDK1/cycD have similar binding behaviours and activity. However CDK4/cycD seems to be less sensitive to the same set of inhibitors than CDK1/cycB, which appears to be due to lower sensitivity to dipole and molecular size of the inhibitors.

Equations 13, 14 and 15 show that inhibitor activity can be successfully correlated with desolvation, lipophilicity, dipole moment and electron affinity (or ionization energy) where the influence of molecular size is negligible. These equations allow insights into redox behaviour of the cyclin-dependent kinase interactions during the various phases of the cell cycle.

Previous studies of the cyclin-dependent kinase inhibitors and HIV-1 protease inhibitors have sought to confirm whether ligand-protein binding is dependent on desolvation effects. It has been found that a four parameter linear model incorporating desolvation energy, lipophilicity, dipole moment and molecular volume of the ligands is a good model to describe the binding between ligands and kinases or proteases. The resistance shown by MDR proteases to the anti-viral drugs is multi-faceted involving varying changes in desolvation, lipophilicity and dipole moment interaction compared to the non-resistant protease. Desolvation has been shown to be the dominant factor influencing the effect of inhibitors against the cyclin-dependent kinases, but lipophilicity and dipole moment are also significant factors. The model can differentiate between the inhibitory activity of CDK2/cycE, CDK1/cycB and CDK/cycD enzymes.

References and notes

1. Fong, C.W.; Permeability of the Blood–Brain Barrier: Molecular Mechanism of Transport of Drugs and Physiologically Important Compounds, *J. Membr. Biol.* 2015, 248, 651.
2. Fong, C.W.; Statins in therapy: Cellular transport, side effects, drug-drug interactions and cytotoxicity - the unrecognized role of lactones, *HAL Archives*, hal-01185910v1, 2014.
3. Fong, C.W.; Drug discovery model using molecular orbital computations: tyrosine kinase inhibitors, manuscript submitted.
4. Marenich, A.V.; Cramer, C.J.; Truhlar, D.J. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem B*, 2009, 113, 6378.
5. Levy, R.M.; Zhang, L.Y.; Gallicchio, E.; Felts, A.K. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute-Solvent Interaction Energy, *J. Am. Chem. Soc.* 2003, 125, 9523.
6. Setny, P.; Baron, R.; McCammon, J.A. How Can Hydrophobic Association Be Enthalpy Driven? *J. Chem. Theory Comput.*, 2010, 6, 2866.
7. Mondal, J.; Friesner, R.A.; Berne, B.J. Role of Desolvation in Thermodynamics and Kinetics of Ligand Binding to a Kinase, *J. Chem. Theory Comput.* 2014, 10, 5696.
8. Shan, Y.; Kim, E.T.; Eastwood, M.P.; Dror, R.O.; Seeliger, M.A.; Shaw, D.E. How Does a Drug Molecule Find its Target Binding Site? *J. Am. Chem. Soc.* 2011, 133, 9181.
9. Turk, B. Targeting proteases: successes, failures and future prospects, *Nature Reviews Drug Discovery*, 2006, 5, 785.
10. Wang, Y.; Dewdney T.G.; Liu, Z.; Reiter, S.J.; Brunzelle, J.S.; Kovari, I.A.; Kovari, L.C. Higher Desolvation Energy Reduces Molecular Recognition in Multi-Drug Resistant HIV-1 Protease, *Biology* 2012, 1, 81-93. 10(a) Freire, E.; Do enthalpy and entropy distinguish best in class from first in class? *Drug Discov. Today*, 2008, 13, 869.
11. Singh, S.K.; Tripathil, S.K.; Dessalew, N.; Singh, P. Cyclin Dependent Kinase as Significant Target for Cancer Treatment, *Current Cancer Therapy Reviews*, 2012, 8, 225.
12. Peyressatre, M.; Prével, C.; Pellerano, M.; Morris, M.C. Targeting Cyclin-Dependent Kinases in Human Cancers: From Small Molecules to Peptide Inhibitors, *Cancers* 2015, 7, 179-237.
13. Misra, R.N.; Rawlins, D.B.; Xiao, H.; Shan, W.; et al, 1*H*-Pyrazolo[3,4-*b*]pyridine Inhibitors of Cyclin-Dependent Kinases, *Bioorganic & Medicinal Chemistry Letters*, 2003, 13, 1133.
14. Misra, R.N.; Rawlins, D.B.; Xiao, H.; Shan, W.; et al, 1*H*-Pyrazolo[3,4-*b*]pyridine Inhibitors of Cyclin-Dependent Kinases: Highly Potent 2,6-Difluorophenacyl Analogues, *Bioorganic & Medicinal Chemistry Letters* 13 (2003) 2405.
15. Blum, A.; Bottcher, J., Sammet, B., Luksch, T., Heine, A., Klebe, G., and Diederich, W.E. Achiral oligoamines as versatile tool for the development of aspartic protease inhibitors. *Bioorganic & Medicinal Chemistry*, 2008, 16, 8574.
16. Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C.A. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes, *Structure*, 2002, 10, 369.
17. Adachi, M.; Ohharaa, T.; Kurihara, K. et al. Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography, *PNAS*, 2009, 106, 4641.
18. Baldwin, R.L.; Chapter 6. Weak Interactions in Protein Folding: Hydrophobic Free Energy, van der Waals Interactions, Peptide Hydrogen Bonds, and Peptide Solvation,

- ed. J Buchner, T Kiefhaber, Protein Folding Handbook, Published Online: 31 Jan 2008. Wiley-VCH Verlag GmbH & Co. Weinheim. 2005: 127.
19. Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi, J. Chem. Theory Comput. 2013, 9, 212.
 20. Sarsour, E.H.; Kumar, M.G.; Chaudhuri, L.; Kalen, A.L.; Goswami, P.C.; Redox control of the cell cycle in health and disease, Antioxid. Redox Signal. 2009, 11, 2985.
 21. All calculations were carried out using the Gaussian 09 package at the B3LYP/6-31⁺G**(6d, 7f) level of theory with optimised geometries in water, as this level has been shown to give accurate electrostatic atomic charges, and was used to optimize the IEFPCM/SMD solvent model. With the 6-31G* basis set, the SMD model achieves mean unsigned errors of 0.6 - 1.0 kcal/mol in the solvation free energies of tested neutrals and mean unsigned errors of 4 kcal/mol on average for ions.⁴ The 6-31G** basis set has been used to calculate absolute free energies of solvation and compare these data with experimental results for more than 500 neutral and charged compounds. The calculated values were in good agreement with experimental results across a wide range of compounds.^{22,23} Adding diffuse functions to the 6-31G* basis set (ie 6-31⁺G**) had no significant effect on the solvation energies with a difference of less than 1% observed in solvents, which is within the literature error range for the IEFPCM/SMD solvent model. It is noted that high computational accuracy for each species in different environments is not the focus of this study, but comparative differences between various species is the aim of the study. The use of various literature values for K_m, IC₅₀ to develop the multiple regression equations have much higher uncertainties than the calculated molecular properties. The statistical analyses include the multiple correlation coefficient R², the F test of significance, standards errors for the estimates (SEE) and each of the variables SE($\Delta G_{\text{desolvation}}$), SE($\Delta G_{\text{lipophilicity}}$), SE(Dipole Moment), SE (Molecular Volume), SE(Molecular Volume), as calculated from “t” distribution statistics. Residual analysis was used to identify outliers.
 22. Rayne, S.; Forest, K. Accuracy of computational solvation free energies for neutral and ionic compounds: Dependence on level of theory and solvent model, Nature Proceedings, 2010, <http://dx.doi.org/10.1038/npre.2010.4864.1>.
 23. Rizzo, R.C.; Aynechi, T.; Case, D.A.; Kuntz, I.D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions, J. Chem. Theory Comput. 2006, 2, 128.