Classical MD and metadynamics simulations on back-pocket binders of CDK2 and VEGFR2: a guidepost to design novel small-molecule dual inhibitors

Andrés Felipe Vásquez and Andrés Fernando González Barrios

Grupo de Diseño de Productos y Procesos (GDPP), Department of Chemical Engineering, Universidad de los Andes, Bogotá, Colombia

Communicated by Ramaswamy H. Sarma

ABSTRACT

Cyclin-Dependent Kinase 2 (CDK2) and Vascular-Endothelial Growth Factor Receptor 2 (VEGFR2) are promising targets for the design of novel inhibitors in anticancer therapeutics. In a recent work, our group designed a set of potential dual inhibitors predicted to occupy an allosteric back pocket near the active site of both enzymes, but their dynamic and unbinding behavior was unclear. Here, we used molecular dynamics (MD) and metadynamics (meta-D) simulations to study two of these virtual candidates (herein called IQ2 and IQ3). Their binding mode was predicted to be similar to that observed in LQ5 and BAX, well-known back-pocket binders of CDK2 and VEGFR2, respectively, including H-bonding with critical residues such as Leu83/Cys113 and Asp145/Asp190 (but excepting H-bonding with Glu51/Glu111) in CDK2/VEGFR2, correspondingly. Likewise, while LQ5 and BAX unbound through the allosteric channel as expected for type-IIA inhibitors, IQ2 and IQ3 unbound via the ATP channel (except for CDK2-IQ2) as expected for type-I/IIA inhibitors. Interestingly, a C-C single/double bond difference between IQ2/IQ3, respectively, resulted associated with differences in the AS/T loop flexibility observed for CDK2. These insights will help developing scaffold modifications during an optimization stage, serving as a starting point to develop dual kinase inhibitors in challenging biological targets with a promising anticancer potential.

Introduction

Protein kinases (PKs) are a large superfamily of enzymes involved in the transfer of phosphate groups to other proteins in the cell and play a critical role in the regulation of numerous signaling pathways (Arslan et al., 2006; Roskoski, 2015). Many PKs are frequently implicated in cancer and, therefore, have been considered as attractive molecular targets for the design of novel small-molecule inhibitors (Cohen, 2002; Noble et al., 2004; Yarmoluk et al., 2013). Among them, Cyclin-dependent kinase 2 (CDK2) and Vascular endothelial growth factor receptor 2 (VEGFR2) have found to be of special interest in the last few years. CDK2 is one of the major regulators of cell growth and proliferation during G1 to S transition, the progression of S phase, and suppression of cellular senescence by controlling a complex signaling switch that involves the oncoproteins Myc and Ras (Campaner et al., 2010; Hydbring & Larsson, 2010; Sánchez-Martínez et al., 2015; Yuan et al., 2013). On the other hand, VEGFR2 is a key mediator of VEGF-induced effects, including migration, survival, permeability and, most important, angiogenesis and lymphangiogenesis (Olsson et al., 2006; Roskoski, 2007; Stacker & Achen, 2013). Interestingly, the simultaneous inhibition of both enzymes represents a potentially synergic strategy: it not only could hit various central cancer hallmarks previously described (Hanahan & Weinberg, 2011) but also the anti-angiogenic therapy, by forcing the blood vessels to shift from an irregular architecture to one more ordered and normalized, could improve the penetration and effect of other anticancer drugs (in this case inhibiting cell proliferation) and, hence, help halt critical steps in tumor spread that are necessary for invasion and metastasis (Hicklin & Ellis, 2005; Minchinton & Tannock, 2006; Waldner & Neurath, 2012). The hypothesis behind this specific synergistic combination was firstly posed by Bayer Pharma 15 years ago, when granted an European Patent Application (EP 1 568 368 A1) describing the use of single-target inhibitors of CDKs and VEGFRs (Gerhard Siemeister, 2005). Some years later, this same pharmaceutical company designed a third-generation dual compound (ZK 304709), which proved to be promising in certain human tumor xenografts and carcinoma models by highly reducing tumour microvessel density and primary tumor growth (up to 80% of the latter) (Romano, 2013; Scholz et al., 2009; Siemeister et al., 2006), but unfortunately could not reach the marker due in part to its poor ADMET profile (Scott et al., 2009). In the last decade, some well-known inhibitors have been developed for VEGFR2 and a few inhibitors for CDK2 are currently under investigation (Lapenna & Giordano, 2009), but until now there is no an approved kinase inhibitor able to bind and inhibit both of these enzymes. A dual compound design strategy could reduce side effects, resistance, and drug interactions, thus...
improving the efficacy and facilitating the eventual dosing and regulation process compared to a combination therapy (Stanković et al., 2019).

We currently know that the vast majority of kinase inhibitors interacts specifically with the ATP-binding (front) pocket in a so-called ‘active’ conformation (type I inhibitors), which depends on a particular configuration ‘in’ for both a conserved DFG motif and its αC-helix (Roskoski, 2015; Zhang et al., 2009). However, a most recent trend is intended to take advantage of a hydrophobic back pocket adjacent to the active site that is only accessible in an ‘inactive’ conformation when the DFG motif and the αC-helix (or at least one of them), adopt an ‘out’ configuration (Roskoski, 2016; Zuccotto et al., 2010) (Figure 1). This allosteric pocket may be harnessed to optimize molecular interactions and, more important, to achieve a better selectivity profile compared to the type I inhibitors, which have been considered in the last few years as potentially promiscuous compounds (Liu & Gray, 2006; Rabiller et al., 2010). Since a few protein kinases, including CDK2 (Alexander et al., 2015) and VEGFR2 (Roskoski, 2016), can adopt the inactive conformation, there is now a valuable opportunity to identify innovative and selective multi-target inhibitors able to extend into the back cleft of both kinases and to exert a potentially advantageous synergistic action in cancer treatment.

In this study, we developed a dynamic analysis of two virtual candidates for potential dual CDK2/VEGFR2 inhibitors recently found by combined protocol involving virtual screening (VS), fragment-based drug design (FBDD), and natural products (NPs) (Figure 2) (Vásquez et al., 2020). By performing molecular dynamics (MD) simulations we gained a major understanding of the role of key residues in protein-ligand interactions and the conformational dynamics of each complexed enzyme. Likewise, we carried out metadynamics (meta-D) simulations (Laio & Parrinello, 2002) to predict the exit channel of each compound during the unbinding process, comparing them to other back-pocket binders.

**Methods**

**Molecular dynamics simulations**

The 3D structures of both kinases under study, namely CDK2 and VEGFR2, co-crystallized along with a well-known back pocket-binder were retrieved from the Protein Data Bank (PDB) (Burley et al., 2019; Maesaka et al., 1978; Rose et al., 2015) (PDB entries: 5A14 (Alexander et al., 2015) and 4ASD (McTighe et al., 2012), respectively). Likewise, we included docked complexes that included a pair of virtual candidates (herein referred to as IQ2 and IQ3) designed also to occupy the back pocket of these enzymes (Figure 2). An apo (unbound) conformation was prepared for each enzyme by eliminating the co-crystallized ligand from the source structure in each case. The Protein Preparation Wizard and Superposition panels were employed from Maestro molecular modeling environment (Madhavi Sastry et al., 2013) (software Schrodinger v2018-10 (Schrödinger Release 2019-3: Maestro, Schrödinger, LLC, New York, NY, 2019, n.d.)) in order to construct each protein-ligand complex under study. Likewise, the module Prime (M. Jacobson et al., 2002; M. P. Jacobson et al., 2004) was used to build and add any missing side chains/loops on the structures and the module Epik (Greenwood et al., 2010; Shelley et al., 2007) was used to generate probable ionization and tautomeristic states in the pH range specified for all protein or ligand heteroatoms (i.e. any atom other than carbon or hydrogen). We used Amber ff14SB and GAFF force fields for the protein and the ligand, correspondingly (Maier et al., 2015; Wang et al., 2004), and then we calculated ligand charges at the Hartree-Fock (HF) level using a 6-31G* basis set as default parameters of Gaussian 09 package (Frisch et al., 2009). From this point on, we used GROMACS v2018 (Abraham et al., 2015; van Derveensop, 2014; Van Der Spoel et al., 2005) for the MD setup and the corresponding simulations. We used cubic boxes, solvation with explicit TIP3 water molecules (Jorgensen et al., 1983), and a concentration of 0.15 M for the neutralization step by Na+ and Cl- ions. Energy minimization was carried out on all protein-ligand systems with 5000 steps of steepest descent (Meza, 2010), followed by a heating process represented in a canonical (NVT) ensemble for 100 ps and then by equilibration in an isothermal-isobaric (NPT) ensemble for another 100 ps using a Berendsen barostat and a modified Berendsen thermostat (V-rescale) (Bussi et al., 2007; Fias et al., 2008). The NVT ensemble was used for a short time to get the velocity distribution reasonable, while the NPT ensemble was employed to relax the box volume and get the right density under constant (atmospheric) pressure (Braun et al., 2019; Lemkul, 2014; Mülders et al., 1998). In both cases, we used two different coupling groups (protein + ligand and solvent + ions) and backbone restraining. The temperature was kept to 300 K with a mesh spaced for 0.16 nm, and the pressure was kept at 1 atm. Bond lengths were constrained during the simulations using the LINCS algorithm (Hess et al., 1997). Finally, a 100 ns production simulation was carried out for all systems in a NPT ensemble (no backbone restraining) running on a Linux workstation using an Intel Xeon Gold Processor 5118 SP 2.3 GHz (3.2 GHz Turbo) with 24 Dual Cores and 64 GB of RAM, as well as two NVIDIA Titan Xp GPUs with 12GB GDDR5X 3840 CUDA cores. The van der Waals (vdW) interactions were cut off at 1.0 nm, and long range electrostatic interactions were calculated by the Particle Mesh Ewald (PME) (Darden et al., 1993) with a cut-off of 1.0 nm.

**Binding free energy calculations**

We evaluated the free binding energy of small ligands under study to CDK2 and VEGFR2 by the method MM-PBSA (Molecular Mechanics [MM] with Poisson-Boltzmann Surface Area), which allows implicit solvent calculations on snapshots from MD simulations (Genheden & Ryde, 2015; Kollman et al., 2000). For this purpose, we used the software G_mmpbsa (Kumari et al., 2014) in each evaluated enzyme-inhibitor complex. This program estimate the binding-free energy ΔG(binding) by solving the following formula:

$$\Delta G_{\text{binding}} = \Delta E_{\text{MM}} + \Delta G_{\text{solvation}}$$

where ΔE_{MM} includes the energy of both bonded and non-bonded interactions according to MM force-field typical parameters:
\[ E_{MM} = E_{\text{bonded}} + E_{\text{nonbonded}} = E_{\text{bonded}} + (E_{\text{vdW}} + E_{\text{elec}}) \]

\[ \Delta G_{\text{solv}} \] represents the sum of electrostatic and nonelectrostatic contributions to the solvation free energy, respectively:

\[ G_{\text{solvation}} = G_{\text{polar}} + G_{\text{nonpolar}} \]

\[ \Delta E_{\text{bonded}} \] is always taken as zero and no corrections will be applied for entropic changes (Kumari et al., 2014). MM-PBSA approach has already been used in workflows involving drug discovery campaigns with very good results, which proved highly similar to those obtained experimentally (Genheden & Ryde, 2015; Sgobba et al., 2012).

**Metadynamics simulations**

We developed well-tempered metadynamics (meta-D) simulations (Barducci et al., 2011; Ensing et al., 2006; Laio & Gervasio, 2008) with the two well-known co-crystallized inhibitors and our two virtual candidates IQ2 and IQ3 using PLUMED v2.5 (Bonomi et al., 2009). Based on the same MD setup, we run 20-ns well-tempered meta-D production simulations, respectively, using two collective variables (CVs) (Figure 3a and b): the distance between the center of mass (COM) of the ligand and either the COM of hinge region or the binding site of the protein. This last site was represented according to the alpha carbons (Cα) of the residues Val18, Ala31, Lys33, Glu51, Leu54, Leu58, Ile63, Val64, Leu78, Phe80, Phe82, His119, Val123, His125, Leu143, Ala144, Asp145 and Phe146 for CDK2, and the residues Val848, Ala866, Lys868, Glu885, Ile888, Leu889, Ile892, Val898, Val914, Val916, Phe918, Ala1020, Cys1024, His1026, Ile1044, Lys1045, Asp1046 and Phe1047 for VEGFR2, in line with the corresponding positions described by Casasnovas et al. (2017). Two extra CVs, previously described by Lovera et al. (2012) and Morando et al. (2016), were included to describe the DFG flip of the AS/T loop in both kinases (Figure 3c and d): the distance between the COMs of Lys33-Asp145 and Leu58-Phe146 in CDK2, and the distance between the COMs of Lys62-Asp190 and Ile86-Phe91 in VEGFR2. The bias factor was set to 14.0 and the Gaussian height was set to 1.4 kJ/mol with a deposition rate of 10 steps. The Gaussians’ width was 0.07 nm for all distances. The specific set of CVs was selected in terms of their ability to describe slow events relevant to the unbinding of the back-pocket binders under evaluation, this way surpassing energy barriers required to describe both the initial and final state (ligand bound and unbounded) as well as in-between metastable states, just as indicated by Laio and Gervasio (2008). Interestingly, CV1 and CV2 have not only been recently used in previous studies focused on ligands able to bind the back pocket of the same or similar kinases (Callegari et al., 2017; Capelli & Costantino, 2014), but also CV1 has been mentioned in other reports as absolutely critical to estimate the corresponding unbinding free energy profiles (FEPs). In line with this reasoning, the hinge motion must be explicitly considered for energy calculations not to lead to unrealistically high barriers (Morando et al., 2016).

**Results**

**Stability and flexibility of the systems**

The root-mean-square deviation (RMSD) of the backbone of CDK2 (Supplementary material Figure S1) and VEGFR2 (Supplementary material Figure S2) rapidly converged around 2 Å throughout most of the time course of the simulations, especially after the first 20 ns. The RMSD for IQ2 in complex with both enzymes was shown to be lower...
compared to all other systems (slightly lower than LQ5), whereas RMSD for IQ3 was the highest for CDK2 and VEGFR2 among the other compounds. Interestingly, all protein-ligand complexes proved to have RMSD values lesser than the unbound form of CDK2.

To understand the differences among the systems in more detail, we analyzed the per-residue root mean square fluctuation (RMSF) data and observed notable fluctuations in four loops and one helix of the CDK2 structure: the P-loop and the $\beta$3-$\alpha$C, $\beta$4-$\beta$5, and AS/T loops, as well as the $\alpha$D helix (Figure 4). For the first two loops, the fluctuation was observed to be similar (approximately 4 Å) for all the systems under evaluation. However, for the AS/T loop, the differences observed were certainly larger. Although this last region proved to be much more rigid when CDK2 was bound to LQ5 (and especially to IQ3) compared to free form, it gained more flexibility when bound to IQ2, reaching a peak between 6 and 7 Å. On the other hand, the P-loop and $\alpha$D helix in CDK2 were showed as more rigid when the enzyme was complexed with any compound of the study compared to the free form. On the other hand, per-residue RMSF calculation of VEGFR2 structure showed a large flexibility in two loops, $\alpha$E-$\alpha$F and AS/T (Figure 5), particularly for the complexes including either IQ2 or IQ3. For the $\alpha$E-$\alpha$F loop, the flexibility of the free form of the enzyme was shown as similar to the bound form complexed with LQ5, while the complex with IQ3 proved to be more rigid. Conversely, the complex with IQ2 was much more flexible, achieving a RMSF of almost 7 Å. For the AS/loop, which showed to reach a peak about 4 Å in the free form, proved to be much more rigid in the different protein-ligand complexes, within a range between 2 and 2.5 Å. Both, the potential energy and kinetic energy for CDK2 and VEGFR2 showed a rather constant value (Supplementary material Figure S5 and S6, respectively), which was in line with the the total energy fluctuation curves (Supplementary material Figure S7) for each protein system analyzed (both in bound and unbound form). Volume plots indicated a constant value for production phase of the simulations (Supplementary material Figure S8). On the other hand, it was clear that the temperature of the system, after quickly reaching the target value (300 K) during the heating phase, remained stable over the remainder of the equilibration and, particularly, the production phase (Supplementary material Figure S9). Finally, pressure and density values for each protein and protein-ligand system fluctuated only slightly over the course of the 100-ns production running, much as expected in this phase (Supplementary material Figure S10 and S11, respectively).

**Binding mode and energy in CDK2 and VEGFR2**

The binding mode of these virtual candidates was similar to that observed in active compounds LQ5 and BAX: H-bonding with critical residues such as Leu83 and Asp145 in CDK2 (Supplementary material Figure S3), which corresponds to Cys113 and Asp190, respectively in VEGFR2 (Supplementary material Figure S4), was maintained. The first residue for both kinases is located in their hinge region at the limits of the front pocket (FP), also known as the ATP binding site, while the second residue is nearer to the allosteric back pocket (BP). However, the H-bonding interactions with Glu51/Glu79 in the C-helix in both kinases were less frequent along the trajectories with IQ2 and IQ3 compared with LQ5.
and BAX. Remarkably, an H-bond observed for IQ2 and IQ3, but not for LQ5 and BAX, was present with the residue Glu81/Glu111 in CDK2 and VEGFR2, respectively, which is located in the hinge region of both kinases. The MM-PBSA results indicated favourable $\Delta G_{\text{binding}}$ values for compounds IQ2 and IQ3 compared to the reference compounds LQ5 and BAX in CDK2 and VEGFR2, respectively (Supplementary material Figure S12). Remarkably, in all protein-ligand complexes under study, the electrostatic contribution was 2–3 times lower than the van der Waals corresponding contribution. This finding was particularly true for CDK2 complexes, in which IQ2 and IQ3 showed much lower (favourable) $\Delta G_{\text{binding}}$ values compared to reference compound LQ5 (Table 1). The residues implicated in the binding energy contribution are in well agreement with the intermolecular H-bonds observed with each ligand under evaluation, and it is clear that 8 – 14 of these residues, depending of the particularly evaluated protein-ligand complex, accounted for more than 3 kcal/mol contribution in $\Delta G_{\text{binding}}$ (Supplementary material Figure S13).

Finally, the unbinding pathway of the well-known active compounds and our virtual hits was estimated. Two possible exit channels were possible for these molecules: the allosteric channel or the ATP binding catalytic channel (Figure 6). While the unbinding of LQ5 and BAX was shown to occur through the allosteric channel (also known as ‘deep’ pocket) as expected for type-IIa inhibitors (Roskoski, 2016), the unbinding of virtual candidates IQ2 and IQ3 occurred via the ATP channel (except for the complex CDK2-IQ2). It is important to mention also the conformational changes associated with the unbinding of the evaluated ligands in CDK2 and VEGFR2, especially those linked with type-II inhibition, showed an expected two-step binding process: a first, fast event and a second an slower conformational rearrangement. The first is mainly represented by the detach of the ligand from the hinge region of the kinase protein structure, while the second is dominated by the movement of DFG-motif and AS/T loop as the ligand slides through the so-

Figure 3. Graphical representation of the four collective variables (CVs) used for meta-D simulations involving a set of COMs of CDK2/VEGFR2 systems (CDK2-LQ5 complex used as reference). a) Distance between ligand and hinge region (CV1), b) distance between ligand and binding site (CV2), c) distance between Lys33 and Asp145 in CDK2 (Lys62 and Asp190 respectively in VEGFR2) (CV3), and d) distance between Leu58 and Phe146 in CDK2 (Ile86 and Phe191 respectively in VEGFR2) (CV4). The different COMs are represented by red spheres and the distance between them by discontinuous purple lines. The ligand is represented by cyan wires (CPK colored), the general protein backbone by gray ribbons (30% transparency), and the backbone of the hinge region, Lys33/Asp145, and Leu58/Phe146 by pink, coral, and gold ribbons, respectively.
called ‘deep pocket’ beneath the αC helix. The free energy profiles associated with the unbinding from CDK2 and VEGFR2 obtained by meta-D are shown in Figures 7 and 8, respectively. For CDK2, the profile for CV1 is characterized by a pronounced energy minimum, the deepest of all, which is followed a second basin as observed for the compounds LQ5 and IQ2, but much less prominent for IQ3 (Figure 7a). A similar result was obtained for CV2 in LQ5 and IQ2, in which the profile for three of the evaluated compounds is dominated by a large minimum that only for LQ5 and IQ2, is followed by a second less pronounced basin (Figure 7b). Likewise, we observed that the profiles for CV3 and CV4 were different among the evaluated compounds: while for LQ5 these profiles were characterized by a unique minimum, for IQ2 and IQ3 this minimum was much broader and shifted to either shorter (CV3, Figure 7c) or larger distances (CV4, Figure 7d).

For VEGFR2, the profile for CV1 is composed of a large and narrow minimum among all compounds, which is followed by a second minimum observed only for BAX (Figure 8a). The profile for CV2 was also characterized by two minima for all compounds, followed by a third minimum only observed for BAX (Figure 8b). Finally, the profiles for CV3 and CV4 were quite similar among the compounds, being dominated by a unique basin at very similar distances (Figure 8c and d).

**Discussion**

After running the first part of the MD simulations, the backbone RMSD profile data indicates a general structural stability of the enzymes CDK2 and VEGFR2 in both free and bound forms, which is not only in line with the H-bond occupancy estimated for each compound of interest, but also
with the conformational pose adopted by them inside both binding sites. Similarly, the MM-PBSA data confirms the binding free energy values derived from each protein-ligand complex, in which electrostatic contribution proved to be greater in all cases and supports, hence, the already mentioned H-bond analyses. Our previous studies of docking-

Figure 6. Superposition of CDK2 and VEGFR2 indicating the exit channels evaluated by meta-D simulations of protein-ligand complexes. Crystallized 3D structures of CDK2 (gray ribbons) and VEGFR2 (light green ribbons) are superimposed for depicting the two possible unbinding pathways (either ATP or allosteric channel) of complexed inhibitors (LQ5 and BAX colored gold and purple, respectively) (a). The surface of ATP and allosteric channel, 5 Å around the corresponding ligand, is illustrated in b) and c) for CDK2 and in d) and e) for VEGFR2, respectively. Each surface has a transparency of 30% and it is colored according to the Kyte-Doolittle hydrophobicity scale.

Figure 7. Free energy surface (FES) after running meta-D simulations for CDK2. FES for the protein-ligand complexes CDK2-LQ5, CDK2-IQ2 and CDK2-IQ3 according to the CVs1-4 (listed from a to d, respectively).
based virtual screening are in well agreement with these results (Vásquez et al., 2020). Furthermore, the RMSF analyses for CDK2 and VEGFR2 helped confirm a key role for some regions in both protein structures, considering that the rigidity observed for the P-loop and the αD-helix and the flexibility of AS/T loop in the bound form of both enzymes (in comparison with the free-form) has been previously reported in other kinases (Morando et al., 2016). We found worthy of note the much larger mobility observed for the AS/T loop in the complex CDK2-IQ3 compared to the compound LQ5, and especially IQ2, because the only divergent feature between the virtual compounds IQ2 and IQ3 is a double C-C bond located towards their middle region. This fact leads us to propose a potentially critical role of this chemical moiety in the interaction with this particular protein. Likewise, the large flexibility of the αD-αE loop observed for the VEGFR2-IQ2 complex is also intriguing because this region is distal from the active site. However, since the flexibility of this type of loops has previously been observed in other kinases (such as EGFR or c-Src) and associated with a 'cracking' phenomenon, in which the regions become more flexible because of partial unfolding required to satisfy entropic penalties associated with the ligand binding (Klein et al., 2015; Morando et al., 2016; Shan et al., 2013), we believe that our results reveal a similar role for the αD-αE loop in VEGFR2. The remarkable flexibility and rigidity in

### Table 1

| Complex    | ΔEelect | ΔEvdw   | ΔGpolar | ΔGnonpolar | ΔGbinding |
|------------|---------|---------|---------|------------|-----------|
| CDK2-LQ5   | -263.087 | -85.365 | 339.751 | -25.318 | -34.020   |
| CDK2-IQ2   | -222.641 | -52.057 | 164.637 | -22.397 | -132.458 |
| CDK2-IQ3   | -229.709 | -49.221 | 168.746 | -21.555 | -131.739 |
| VEGFR2-BAX | -249.618 | -80.076 | 173.852 | -23.456 | -179.297 |
| VEGFR2-IQ2 | -231.632 | -72.285 | 169.119 | -22.068 | -156.866 |
| VEGFR2-IQ3 | -226.830 | -78.401 | 178.396 | -21.769 | -148.604 |

Binding free energy values for the protein-ligand complexes CDK2-LQ5, CDK2-IQ2, CDK2-IQ3, VEGFR2-BAX, VEGFR2-IQ2, and VEGFR2-IQ3 using the 100-ns production MD simulation results in each case. A step jump of 10 was used for diminishing the number of MD-derived frames used during the calculations. ΔEelect=electrostatic interactions; ΔEvdw=van der Waals interactions; ΔGpolar=electrostatic contribution of solvation free energy; ΔGnonpolar=non-electrostatic contribution of solvation free energy; ΔGbinding=total binding energy. The solvent accessible surface area (SASA) model was used for the ΔGnonpolar calculations in each case. All values listed here are shown in kJ/mol.

**Figure 8.** Free energy surface (FES) after running meta-D simulations for VEGFR2. FES for the protein-ligand complexes VEGFR2-BAX, VEGFR2-IQ2, and VEGFR2-IQ3 according to the CVs1-4 (listed from a to d, respectively).
specific regions observed for CDK2 and VEGFR2 could be associated to a dominant ‘conformational selection’ mechanism discussed by Morando et al. (2016), in which ligand is able to recognize and bind to a DGF-out state of the proteins and, as a consequence, to their back pockets.

The most important observation from the unbinding studies is that not all the virtual candidates evaluated here use the same exit channel in the same or another enzyme (CDK2 or VEGFR2). From here, two different but interrelated facts are derived. On the one hand, a virtual candidate could be categorized as potential type II inhibitor for one specific enzyme, could also be a type I-1/2 for the other, as previously observed for other kinase inhibitors (Roskoski, 2016; Sun et al., 2015). The type I-1/2 inhibitors are known to bind to the adenine ring region similar to type-I inhibitors by establishing hydrogen bonds with the hinge region, but also extending into the back cavity and hence interacting with residues associated with the type-II pharmacophore (Zuccotto et al., 2010). On the other hand, and most important, specific structural features of the virtual candidates could also influence the capacity to access a particular exit channel. In this context, we believe that the single structural difference of IQ2 compared to IQ3 (i.e. the C-C single vs double bond, respectively) could be crucial not only to explain the recognition and interaction with the back pocket of CDK2, but also the large AS/T loop flexibility observed for this particular protein-ligand complex. It should be noted that the scaffold represented by this C-C bond has been previously reported as critical to interact with the so-called gate-keeper (GK) residue responsible to provide access to the back pocket (Gillis & McLeod, 2016; Rabiller et al., 2011; Roskoski, 2016) (especially in CDK2, where the GK residue is a bulky phenyalanine (Zuccotto et al., 2010)), as well as to maintain an orthogonal orientation with respect to the axis of the chemical moiety interacting with the hinge region (Backes et al., 2011). Therefore, the last point suggests that, in order to favor the type-II inhibition mode and harness its potential, especially in terms of their higher residence time in active site compared to other types of kinase inhibitors (Blanc et al., 2013; Callegari et al., 2017; Casasnovas et al., 2017; Iwata et al., 2011; Kokh et al., 2018), a single C-C bond (as observed in IQ2) should be maintained with the aim of improving rotatability required to access the deep pocket. In addition, we believe that the presence of a non-aromatic ring in this part of our virtual candidates, compared to LQ5 or BAX, could improve the developability profile of a potential dual CDK2/VEGFR2 compound: it would imply not only a lower MW, lipophilicity, and aromatic ring count (which are strongly associated with ADMET profile) (C. a. Lipinski, 2004; C. A. Lipinski, 2000; Ritchie & Macdonald, 2009), but also a better ligand efficiency (LE) (Hopkins et al., 2014) (or other similar metrics normalized by size) because of a lower number of non-hydrogen atoms. This is even more interesting considering that alterations in ring structures, including ring shortening, have been previously reported for scaffolds of kinase inhibitors with a multi-group capacity, it is, able to inhibit more than one kinase group (Hu & Bajorath, 2015).

From the above evidence, we can conclude that the applied MD and meta-D framework applied in this in silico study proved useful to gain a deeper understanding about the structural, binding and dynamic features of the back pocket-binders IQ2 and IQ3 as potential CDK2/VEGFR2 dual inhibitors. Based on the observations drawn from our MD simulations, we found a similar H-bond profile, binding pose and stability for these compounds compared to well-known inhibitors used as reference, and identified that the prominent flexibility degree of certain protein regions such as AS/T loop, among others, appears to play a key role in the kinases under study in the process of ligand recognition and interaction. Likewise, meta-D analysis allowed us to identify key differences between our back-pocket binders in the exit or escape channel (ATP or allosteric) compared to reference compounds, which indicates a strong dependency not only on the specific enzyme involved but also on very specific features (involving C-C bonds in the middle part of their chemical structure) associated with a potentially privileged access to the back pocket of the kinases. We consider these insights will be highly valuable as taking-decision criteria to propose potential scaffold alterations during eventual rounds of optimization, chemical synthesis, and biological tests of our virtual candidates, as well as to unravel structural complexities that help shed light on the intricacies behind the selectivity of multi-target inhibitors of promising therapeutic targets.

Acknowledgements

We acknowledge funds from the Colombian Ministry of Science, Technology and Innovation MINCIENCIAS Grant No. 727. Also, we thank funding sources from Universidad de los Andes under Educational Loan Program for Doctoral Students (Vice-Rectory of Research and Creation) and the Doctoral Studies Support Program of the Faculty of Engineering.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Colombian Ministry of Science, Technology and Innovation – MINCIENCIAS.

Data availability

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

References

Abraham, M. J., Murtola, T., Schulz, R., Päll, S., Smith, J. C., Hess, B., & Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1-2, 19–23. https://doi.org/10.1016/j.softx.2015.06.001
Alexander, L. T., Möbitz, H., Drueckes, P., Savitsky, P., Fedorov, O., Elkins, J. M., Deane, C. M., Cowan-Jacob, S. W., & Knapp, S. (2015). Type II inhibitors targeting CDK2. ACS Chemical Biology, 10(9), 2116–2125. https://doi.org/10.1021/acschembio.5b00398
