ChemPhysChem

Accepted Article

Title: Binding Energy Partition of Promising IRAK-4 Inhibitor (Zimlovisertib) for the Treatment of COVID-19 Pneumonia

Authors: César Arturo Zapata-Acevedo, José Manuel Guevara-Vela, Paul L. A. Popelier, and Tomás Rocha Rinza

WILEY-VCH

Binding Energy Partition of Promising IRAK-4 Inhibitor (Zimlovisertib) for the Treatment of COVID-19 Pneumonia

Dr. César A. Zapata-Acevedo[1,2,3], Prof. Dr. José Manuel Guevara-Vela[4], Prof. Dr. Paul L. A. Popelier[3], Prof. Dr. Tomás Rocha-Rinza[1*]

[1] Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Alcaldía Coyoacán, 04510 Ciudad de México, México.
[2] Tecnológico de Monterrey, Campus Santa Fe. Av. Carlos Lazo 100, Santa Fe, La Loma, Alcaldía Álvaro Obregón, 01389, Ciudad de México, México.
[3] Department of Chemistry, University of Manchester, Oxford Road, Manchester, M13 9PL, Great Britain.
[4] IMDEA Materials Institute, C/Eric Kandel 2, 28906, Getafe, Madrid, Spain.

* Corresponding author: Tomás Rocha-Rinza: trocha@iquimica.unam.mx

(https://iquimica.unam.mx/dr-tomas-rocha-rinza)

Abstract

The technique of Fragment-Based Drug Design (FBDD) considers the interactions of different moieties of molecules with biological targets for the rational construction of potential drugs. One basic assumption of FBDD is that the different functional groups of a ligand interact with a biological target in an approximately additive, that is, independent manner. We investigated the interactions of different fragments of ligands and Interleukin-1 Receptor-Associated Kinase 4 (IRAK-4) throughout the FBDD design of Zimlovisertib, a promising anti-inflammatory, currently in trials to be used for the treatment of COVID-19 pneumonia. We utilised state-of-the-art methods of wave function analyses mainly the Interacting Quantum Atoms (IQA) energy partition for this purpose. By means of IQA, we assessed the suitability of every change to the ligand in the five stages of FBDD which led to Zimlovisertib on a quantitative basis. We determined the energetics of the interaction of different functional groups in the ligands with the IRAK-4 protein target and thereby demonstrated the adequacy (or lack thereof) of the changes made across the design of this drug. This analysis permits to verify whether a given alteration of a prospective drug leads to the intended tuning of non-covalent interactions with its protein objective. Overall, we expect that the methods exploited in this paper will prove valuable in the understanding and control of chemical modifications across FBDD processes.

Introduction

The cost for research and development for a single new drug has fluctuated in the last 20 years, from $800 million in 2003 [1] to higher costs that have decreased to $1.3 billion in 2018 [2,3]. Recently, we have seen a growing relevance of computers in medicinal chemistry. Sophisticated algorithms capture the physicochemical principles that underlie the activity of drugs[4-6]. Lead compounds that are incorporated into drug candidates have to fulfil structural specifications (Lipinski's rules) and metabolic requirements of Absorption, Distribution,

Metabolism, Excretion and Toxicity (ADMET[7]) in order to be considered as promising candidates[7]. The use of hard thresholds for the discrimination of leads in ligand screenings may compromise further consideration of eventually successful leads because of inconsistencies in the quality and reliability of data[8]. The last 20 years have seen the rise of the Fragment-Based Drug Design (FBDD) method. In this approach, the most relevant fragment hits that bind to different parts of a protein pocket are linked together to increase the specificity towards a target[9]. Two drugs discovered with this method have been approved by FDA and are currently on the market (vemurafenib[10] and venetoclax[11]), while forty more have advanced to clinical trials[12]. The binding of fragments to the pocket are generally enthalpy-driven[13], a condition that validates the use of computational techniques that describe the internal energy of a chemical system to investigate the union of these ligands with their biological targets.

IRAK-4 (Interleukin-1 Receptor-Associated Kinase 4) is a serine/threonine protein kinase that is critical in the innate immune signalling by membrane-bound Toll-like receptors in humans and mice. Since IRAK-4 is central to all MyD88-dependent signalling pathways, it is an attractive therapeutic target to suppress uncharacteristic and prolonged inflammatory responses associated with diseases[14]. This circumstance is evidenced by the rise in the number of published patent applications for IRAK-4 inhibitors over the periods 2008-2014[15] and 2016-2018[16]. Several inhibitors are currently in clinical trials, such as ND-2158, ND-2110[17] as well as the most promising and advanced one, PF-06650833, which is now also known as Zimlovisertib. In fact, because the anti-inflammatory activity of PF-06650833 has shown such promising results in treating rheumatoid arthritis in trial NCT02996500, the drug candidate is now being tested in cases of COVID-19 pneumonia and of severe inflammation. At the time of writing, this trial, NCT04933799, is still recruiting participants.

Energy Decomposition Analyses (EDA)[18, 19], have proved to be valuable tools for the quantitative interpretation of chemical bonding in terms of quasiclassical electrostatic, repulsive Pauli exchange and covalent interactions[20]. Several factors, including the lack of a unified definition of many chemical concepts, such as chemical bonding itself, have marked the establishment of different EDA methods[21]. Nevertheless, the usefulness of these approaches resides in the fact that they can yield valuable insights that are not obtainable by experiment[22]. Indeed, drug design can benefit from EDA due to several reasons. EDA tools bridge the gap between Quantum Mechanics (QM) and the heuristic bonding models in chemistry[20]. More specifically, the methods of EDA use a QM representation of molecular systems, and thus they are more accurate than classical approaches[23]. EDA methods account for every energetic term in electronic systems and thereby, they provide a high detail in the description of the interactions within molecules and molecular clusters. This thorough characterisation would lead to useful insights about the contributions to protein-ligand binding from all the moieties in a potential drug and its target protein. For example, an EDA examination could pinpoint the strongest interacting zones in favour of and against protein-ligand complexation. The system size needed to apply EDA approaches in a proper drug design setting is dependent on the case, and in many instances, it represents the bottleneck for the analysis [24, 25]. There are current efforts to apply EDA to protein-ligand binding by considering either direct calculations[26] or carefully selected fragments [27]. The current outlook can benefit from the application of EDA techniques to case studies as diverse as possible. This work aims to contribute in the direction of this emerging trend.

The Interacting Quantum Atoms energy partition[28] (IQA) is an orbital invariant EDA technique[29], free of any parameters or reference states[30], based on the field of theoretical chemistry denominated as Quantum Chemical Topology (QCT)[31, 32]. QCT comprises rigorous methods to perform wave function analyses while building on the examination of quantum-mechanical observables to divide a molecular volume into atomic subspaces. In particular, IQA has proved successful in explaining the nature of different types of supramolecular interactions from a wide variety of chemical systems[33-41]. Such endeavour is undertaken by considering different

contributions to these interactions. These contributions have sound physical meanings, such as covalency, ionicity and steric effects. Covalency in IQA is defined via the exchange-correlation component of the pair density whereas ionicity and polarity are established through electrostatics. Steric effects are determined by way of the intra-atomic *self* energies[37, 42]. The identity of a molecule, and of an atom in a molecule, is always well-defined in IQA, even when these species are in close contact with other systems. IQA successfully dissects the electronic energy in molecules and molecular clusters under many different circumstances. This condition has led to clear distinctions between quantum and electrostatic effects in a wide variety of electronic systems without the need for arbitrary reference systems[28]. The interest to apply QTAIM and IQA to drug design is a possibility that has been considered for the last 15 years[43, 44], in which some efforts have produced only semi-quantitative results[45]. Back in 2012, a workflow of how future QTAIM protein-ligand studies could be performed was envisioned[46]. The increase in computational power and improvements in software packages that allow to perform IQA analyses has started to make these studies computationally attainable. Additionally, there is an effort to develop a machine learning-based force field using IQA interaction and self-energies as well as other QTAIM-related metrics [47-49]. This effort is aimed to extend the system sizes that can be studied using IQA with a potential application to drug design. Ultimately, such a force field needs data from regular IQA calculations. The increase in computational power has started to enable studies in small prototype systems in drug design, as the current work on molecular clusters of about 200 atoms demonstrates. This investigation also shows how IQA aids in (i) the determination of the effect of new interactions on the binding on an improved drug candidate and in (ii) the comparison of these contacts with those in previous prospects. In other words, this report includes an assessment of the core premise of FBDD, namely, that the interactions of conserved moieties in the ligand will remain essentially unaltered after the addition of new fragments to form a new drug candidate.

More precisely, we exploited the IQA energy partition throughout the different stages of fragment growth that resulted in the drug candidate PF-06650833 (Zimlovisertib). The different steps in this (and any other) drug design process are liable to examination by a multiple-step analysis with IQA. We considered all these steps to reveal how the pocket-ligand interactions result in improvements all over the series of drug candidates. Overall, this paper mainly aims to contribute to the improvement of the drug design paradigm by applying the rationale derived from a highly detailed, fragment-based EDA such as IQA.

Theoretical framework

QCT is an area of theoretical chemistry that employs the mathematical language of dynamical systems (for example: critical points, basins, attractors and gradient paths) to partition a quantum mechanical density function in real space to characterise the chemical bonding in an electronic system[50]. The origins of QCT are founded on the Quantum Theory of Atoms in Molecules (QTAIM) originally formulated by Bader and coworkers[31]. The QTAIM uses the gradient of the electron density $\nabla \rho(\mathbf{r})$ to define molecular structure and *topological atoms*. A topological atom in QCT is a three-dimensional region with a well-defined kinetic energy and a finite volume[51], enclosed by a surface with a null flux of $\nabla \rho(\mathbf{r})$. The constituents of topological atoms are a nuclear attractor and the associated stable manifold of the dynamical system defined by the gradient of the electron density denominated as basin. QTAIM basins are non-overlapping, space-filling regions, i.e., they completely encompass the three-dimensional space with no penetration among them. Such basins are malleable: they adapt to their environment, by means of changes of shape, volume, electron population and energy.

The Interacting Quantum Atoms energy partition considers the QTAIM topological atoms to carry out a rigorous EDA examination. This approach divides the total electronic energy E_{total} of a molecule or molecular cluster in

3

(i) steric (intra-atomic energy) components along with (ii) the interaction energies between basins, divided in quantum mechanical and classical electrostatic contributions. Herein, we will only mention the essence of the IQA formulation. A more thorough description of this method can be found elsewhere[28, 39]. We start by considering E_{total} under the Born-Oppenheimer approximation,

$$E_{total} = \frac{1}{2}\sum_{A \neq B}\frac{Z_A Z_B}{r_{AB}} + \int \hat{h}\rho_1(\boldsymbol{r}_1;\boldsymbol{r}_1')d\boldsymbol{r}_1 + \frac{1}{2}\int\int\frac{\rho_2(\boldsymbol{r}_1,\boldsymbol{r}_2)}{r_{12}}d\boldsymbol{r}_1 d\boldsymbol{r}_2. \qquad (1)$$

The first term on the right-hand side of Eq. (1), in which Z_X is the atomic number of nuclei X and r_{XY} is the distance between nucleus X and Y, represents the nuclear repulsion in the system. The one-electron energy, obtained via the integral $\int \hat{h}\rho(\boldsymbol{r}_1;\boldsymbol{r}_1')d\boldsymbol{r}_1$ wherein $\hat{h} = -\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{r_{A1}}$, is separated into a kinetic and a nuclear-electron attraction term. On the other hand, the integral of the pair density, $\rho_2(\boldsymbol{r}_1,\boldsymbol{r}_2)$, covers the contribution of the electron-electron repulsion to the electronic energy. The division of the three-dimensional space according to QTAIM, replaces the integrals over the whole three-dimensional space, into sums of integrals carried over the topological atoms A, B,

$$\int \hat{h}\rho(\boldsymbol{r}_1;\boldsymbol{r}_1')d\boldsymbol{r}_1 = \sum_A \int_A \hat{h}\rho(\boldsymbol{r}_1;\boldsymbol{r}_1')d\boldsymbol{r}_1 \quad \text{and} \quad \int\int\frac{\rho_2(\boldsymbol{r}_1,\boldsymbol{r}_2)}{r_{12}}d\boldsymbol{r}_1 d\boldsymbol{r}_2 = \sum_{AB}\int_A\int_B\frac{\rho_2(\boldsymbol{r}_1,\boldsymbol{r}_2)}{r_{12}}d\boldsymbol{r}_1 d\boldsymbol{r}_2. \qquad (2)$$

Hence, the grouping of the one-atom (i.e., intra-atomic) terms on one side and the two-atom (i.e., interatomic) terms on the other, allows to define respectively the *self*-steric energies, E_{self}^A, and interaction energies, E_{inter}^{AB}, between basins

$$E_{self}^A = T^A + V_{ne}^{AA} + V_{ee}^{AA} \qquad (3)$$

$$E_{inter}^{AB} = V_{nn}^{AB} + V_{ne}^{AB} + V_{ne}^{BA} + V_{ee}^{AB} \qquad (4)$$

in which T^X is the kinetic energy of atom X, while $V_{\sigma\tau}^{XY}$ is the potential energy between σ and τ in basins X and Y, respectively, where σ and τ can be either the nucleus (denoted n) or the electrons (denoted e). Note that n or e refer to the superscript (A or B) exactly above it. Since the second-order density matrix can be partitioned into a Coulombic component ("cl" for classic electrostatic energy) together with an exchange and correlation contribution ("xc"), the interaction energy can be separated in corresponding terms as well,

$$E_{inter}^{AB} = E_{cl}^{AB} + E_{xc}^{AB} \qquad (5)$$

Although $\rho_2(\boldsymbol{r}_1,\boldsymbol{r}_2)$ is not defined in DFT, one can perform the IQA division of the electronic energy (equations (3)-(4)) as put forward in references[35, 52] considering DFT electronic densities.

The IQA approach can be adapted to work with *superbasins* or molecules within molecular clusters[36, 53].

$$E_{total} = \sum_{\mathcal{G}}E_{add}^{\mathcal{G}} = \sum_{\mathcal{G}}\left(E_{self}^{\mathcal{G}} + \frac{1}{2}\sum_{\mathcal{H}\neq\mathcal{G}}E_{inter}^{\mathcal{G}\mathcal{H}}\right), \qquad (6)$$

in which the atoms of the system are gathered in the sets $\mathcal{G}, \mathcal{H}$ ... while $E_{add}^{\mathcal{G}}$ is the additive energy of basin $\mathcal{G}$. The self-energy of the set $\mathcal{G}$, $E_{self}^{\mathcal{G}}$, is given by the sum of the self-energy of every atom inside the cluster together with the pairwise interaction energies of all the atoms in the same group,

4

$$E_{self}^{\mathcal{G}} = \sum_{A\in\mathcal{G}} E_{self}^A + \frac{1}{2} \sum_{A,B\in\mathcal{G}} \sum_{B\neq A} E_{inter}^{AB} \qquad (7)$$

On the other hand, the interaction energy between groups $\mathcal{G}$ and $\mathcal{H}$ corresponds to the sum of the interaction energies of all pairs of elements of these sets,

$$E_{inter}^{\mathcal{G}\mathcal{H}} = \sum_{A\in\mathcal{G}} \sum_{B\in\mathcal{H}} E_{inter}^{AB} \qquad (8)$$

Equations (6)-(8) enables the use of the IQA approach to examine the formation of molecular clusters among drugs and their biological targets. Such extension of IQA allows to decompose the formation energy, ΔE, for the complexation of superbasins $\mathcal{G}, \mathcal{H}$ ... , to form the molecular cluster $\mathcal{G}\cdots\mathcal{H}\cdots$

$$\mathcal{G} + \mathcal{H} + \cdots \rightleftharpoons \mathcal{G}\cdots\mathcal{H}\;... \qquad \Delta E \qquad (9)$$
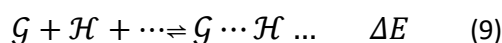
Indeed, ΔE can be written as the sum of (i) the corresponding deformation energies, $E_{def}^{\mathcal{G}}$, caused by the complexation

$$E_{def}^{\mathcal{G}} = E_{self}^{\mathcal{G}} - E_{iso}^{\mathcal{G}}, \qquad (10)$$

in which $E_{iso}^{\mathcal{G}}$ *is the energy of $\mathcal{G}$ when it is isolated in its equilibrium configuration,* along with (ii) the interaction energies among the superbasins,

$$\Delta E = \sum_{\mathcal{G}} E_{def}^{\mathcal{G}} + \sum_{\mathcal{G}>\mathcal{H}} E_{inter}^{\mathcal{G}\mathcal{H}} \qquad (11)$$

For a system of nearly $n\approx 200$ atoms, the number of unique IQA two-atom terms (E_{inter}^{AB}) is $\frac{n(n-1)}{2} \approx 20{,}000$ terms. Equation (5) shows that the separation into V_{xc} and V_{cl} introduces twice as many two-atom energy terms. Adding to these the *n* self-energy terms results in $2\frac{n(n-1)}{2} + n = n^2$ terms in total. The analysis of these numerous components of the electronic energy is only feasible by grouping the atoms of the systems as described above. The initially defined groups were *pocket* and *ligand*, and then we further divided these fragments into *amino acid residues of the pocket* and *moieties of the ligand*. The pocket includes the gatekeeper Tyr262, the catalytic Lys213 together with the surrounding residues Val200, Val246, Met265, Ala315, Leu318, and Ser328 with 160 atoms in total. Furthermore, this grouping allows one to focus on the amino acids and moieties that contribute most importantly to the binding between protein and ligand. We stress that chemical insight emerges after the determination of the parts of the total system whose energies determine the energy profile of the total system[54]. The establishment of the parts of the ligand or pocket that are mostly responsible for the behaviour of the overall energy profile can give directives for the improvement of a potential drug candidate.

Computational details

We performed B3LYP/6-31+G(d,p) and M062X/6-31+G(d,p) electronic structure calculations for the different phases of the development of Zimlovisertib via FBDD. The basis set 6-31+G(d,p) has been confirmed to be reliable for its small size, as a double zeta basis set enjoying a favourable trade-off between computational cost and reliability[55]. The B3LYP approximation has proved to yield adequate partitions of the IQA electronic energy

related to non-covalent contacts in respect of other methods in a recent assessment[56]. On the other hand, the M06-2X functional is an accurate approximation for the description of non-covalent interactions[57]. As implied above, the investigated ligands were taken from the 5 FBDD stages of drug candidate PF-06650833 reported by Pfizer[58]. The ligand and pocket atoms were taken from crystals 5UIQ (lead 51, stage 1), 5UIS (compound 14, stage 3), 5UIT (compound 20, stage 4) and 5UIU (drug candidate PF-06650833, stage 5) from the Protein Data Bank (PDB)[58]. No crystal structure was uploaded for the complex formed by IRAK-4 and compound 10, the second stage of the FBDD process for the creation of PF-06650833, in the PDB. Nevertheless, the 3D structure of this molecule was created after cutting and capping a segment from the very similar PDB structure 5UIR, obtained under the same conditions and uploaded to the PDB by the Pfizer researchers[58].

As mentioned above, we considered a protein pocket model for the ATP binding site in the kinase domain of IRAK-4, as shown in Figure 1(a). The amino acids within this pocket have the strongest interactions with the ligand as revealed by chemical intuition and the Ligand-Interactions tool[59] included in MOE[60] and LigPlot+[61]. Some backbone atoms from other residues were kept throughout the analysis by turning α carbons within the peptide chain into methyl caps in order to top the peptide skeleton without affecting the spin multiplicity of the backbone.

The cofactor-free (the apo-form) structure of IRAK-4 was obtained from crystal 2OIB[62] in the Protein Data Bank (PDB). The overall architecture of the IRAK-4 kinase domain is conserved regardless of the bonding of the ligand to the protein[62]. The pocket models show minimal rearrangement in the kinase site upon ligand binding, averaging a root mean square deviation of 0.31 Å for the pocket complexed in the 5 stages of the design of PF-06650833. We examined the non-covalent interactions within all the systems with the NCI-index[63] and the aid of the NCIPLOT software[64]. All the IQA calculations were performed with the AIMAll package[65], where the average of the integration error per atom is smaller than 0.02 kcal/mol in all systems. This error is far smaller than the formation energies of the investigated molecular clusters as reported below.
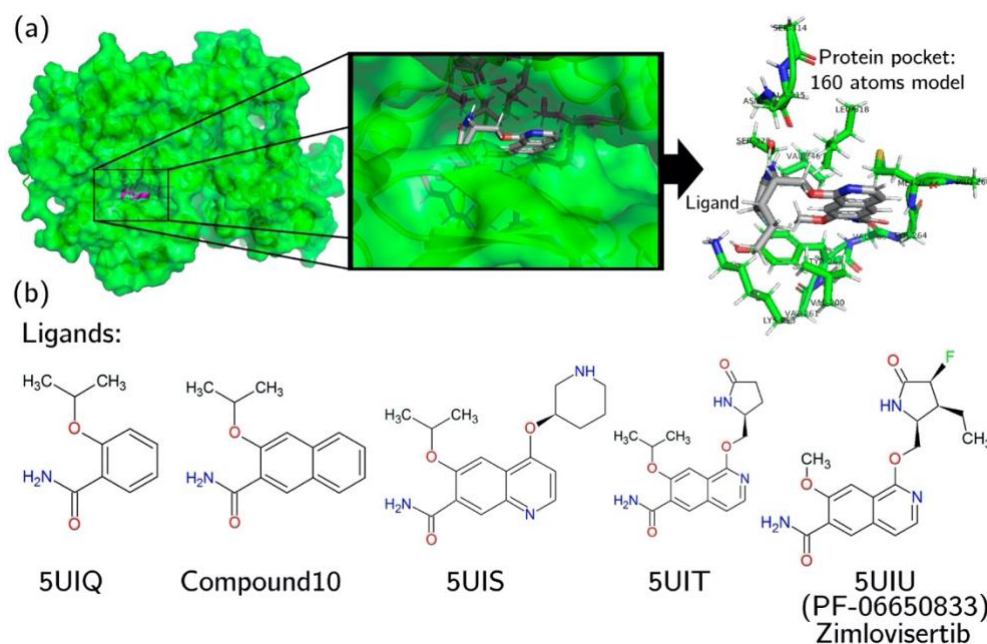
Figure 1. (a) Left. Location of the ATP binding site pocket in the IRAK-4 protein. Right: Close-up of the protein pocket (green backbones) interacting with the drug candidates considered in this study (grey backbone). (b) Ligands with an increasing affinity to IRAK-4. The four-letter codes come from the PDB. The lead/compound numbers have as their source the original FBDD work[58]. Based on Figure 13 from the study by Lee, *et al.*[58].

Results and Discussion

The NCI-index assessment of the protein-ligand complexes investigated (Figure 2), confirmed the occurrence of many weak interactions together with hydrogen bonds that had already been detected with the aid of LigPlot+. The numerous non-covalent interactions found warrant a further quantitative study using the IQA energy partition as reported below.
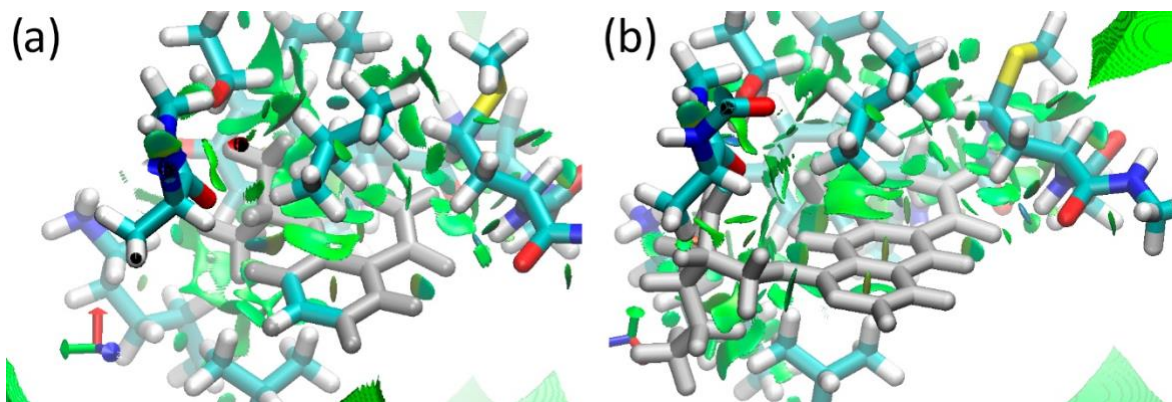
Figure 2. NCIPLOT calculations revealing a large amount of weak interactions (isosurfaces of the reduced gradient of the electron density, *s*=0.5, displayed in green) between the IRAK-4 protein pocket (carbon atoms shown in cyan) with (a) the ligand 5UIQ, i.e., the lead that represented the initial step for the drug design process and (b) the drug candidate PF-06650833, the last step of the drug design process. In both cases, the ligand is displayed in grey.

Analysis of binding contributions per moiety of the drug candidates

Our initial IQA energy partitions were performed on densities computed with the B3LYP functional and the corresponding results are reported in the Supporting Information. Apart from stage 3, the results concerning self-energies and interaction energies are similar to those calculated with the M06-2X exchange correlation functional (Table S1). Nevertheless, the B3LYP approximation fails to reproduce the progressive nature of the binding energy of the ligands in one out of the five FBDD steps for the study that led to PF-06650833 as reported in Figure S1. Indeed, the description offered by B3LYP of the complex formed between 5UIS and the protein is considerably deficient: even the formation energy of the complex, ΔE in equation (9), is non-stabilising. This observation indicates the relevance of dispersion in this complex which, as opposed to M06-2X, B3LYP is not able to describe this contribution to the interaction energy properly[66-70]. This defect can be so serious to the extent that B3LYP might even predict repulsive interactions for molecular clusters that mainly bound by dispersion[70] as it is the case for the molecular complex between 5UIS and the protein pocket. The relevance of dispersion forces in this step of the FBDD process can also be assessed by HF calculations as shown in Figures S1 (a). The HF calculations predict a repulsive character for the complex formed by the 5UI and the protein pocket, and they also fail to describe the progressive binding energy in step 3 of the investigated FBDD process in a similar fashion to the B3LYP functional. These observations support the important role of dispersion in this stage of the FBDD design.
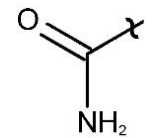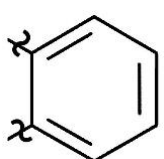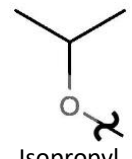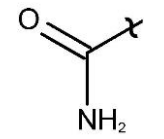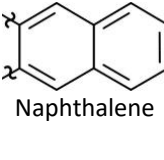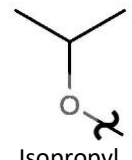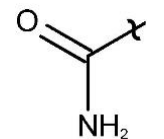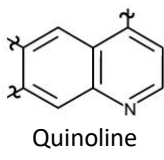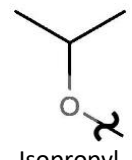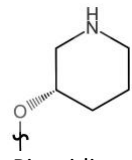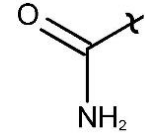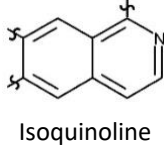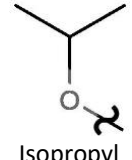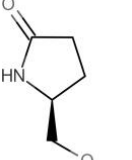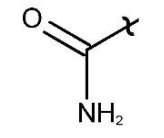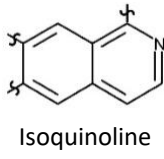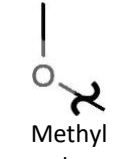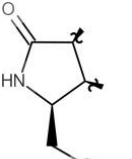
Therefore, only the M06-2X results are examined here. The interaction energy between two groups of atoms, denoted $\mathcal{G}$ and $\mathcal{H}$, within the IQA energy partition consists of the sum of all terms E_{inter}^{XY} in which $X \in \mathcal{G}$ and $Y \in \mathcal{H}$ (Eq. (7)). We start by considering $\mathcal{G}$ and $\mathcal{H}$ as the model pocket and the ligand respectively. The basic premise for the FBDD approach is that once a new fragment has been added to a compound, the interactions

7

of the previous functional groups of the ligand with the biological target will be unchanged. In other words, the energetics of these preceding contacts are independent of the new interactions by the newly added moieties of the compound in the next stages. In the original design of PF-06650833, the improvements in binding across the FBDD process were determined experimentally[58]. We aim to exploit the very high level of detail provided by IQA about interaction energies, in the examination of the energetics throughout the FBDD process of Figure 1(b).

Table 1 shows the relative contribution to binding from the added moieties compared to the initial screening hit, "lead 51", that is, 5UIQ. The central group in the ligands is a cyclic or bicyclic aromatic core. Benzene, naphthalene, quinoline and isoquinoline are used throughout the design process as scaffolds around which the fragments are linked. The contribution of this core to the ligand-pocket IQA interaction energy is small at best (9.5 % and 0.4% for naphthalene and benzene as reported in Table 1), and slightly repulsive, i.e., against the binding (-12.3 for quinoline as we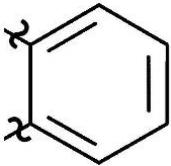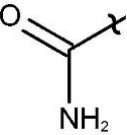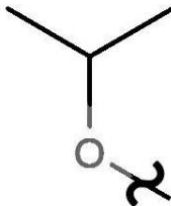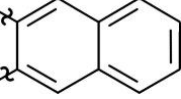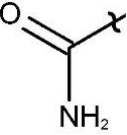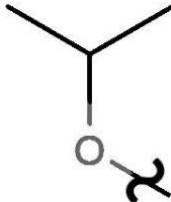ll as -12.2 and -10.8 % as be seen in Table 1) in the last three design stages. Although one might conjecture that these aromatic cores might have relevant π-π stacking with surrounding amino acids, our results indicate that these moieties serve merely as a junction point among the different functional groups that are added in the fragment growing strategy. Their contribution lies more in the realm of membrane permeability[58], a factor which is outside the scope of this work. As shown in Table 1, these central groups yield the smallest contribution to the ligand-pocket IQA interactions energies, which can be even slightly repulsive, as indicated by a negative sign in Table 1. Thus, they are the moieties that contribute the least to the formation of the protein-ligand contact. Naphthalene (and benzene, very slightly) is the only aromatic core with a favourable IQA contribution to the ligand-pocket binding. The FBDD approach can be fully analysed when looking at the moieties bonded to the above-mentioned aromatic cores. Not every stage of design has the same number of moieties, as the very nature of FBDD is to improve upon the previous stage by either substituting, removing or adding new functional groups. We observe that the IQA interaction energies become larger as we move forward in the stages of FBDD. With respect to reference compound 5UIQ, the relative percentages of E_{inter} are 115.7 %, 148.6 %, 157.0 % and 156.3 % for compound 10, 5UIS, 5UIT and 5UIU respectively. There is indeed a slight decrease in E_{inter} in the change from 5UIT to 5UIU in the last step of FBDD (last two rows of Table 1). In this case, the deformation energies for the latter compound are smaller than those for the former (Eq. (10)) to the extent that the binding of 5UIU to the protein pocket is more favourable than it is for 5UIT. As reported in Table 2, the deformation energies of 5UIT and 5UIU are 108.9 and 96.7 kcal/mol

8
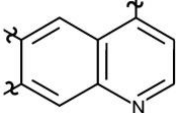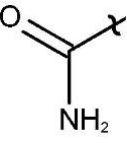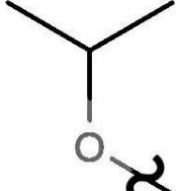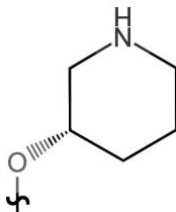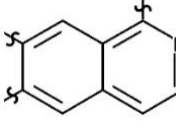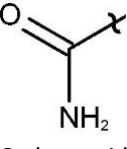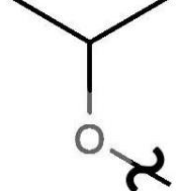
Table 1. IQA interaction energies, $E_{inter}^{\mathcal{GH}}$ in equation (8), between the different fragments of the ligand and the protein pocket. These energies are reported as a relative percentage compared to the interaction of the initial Lead 51 (5UIQ). The squiggles indicate contact points where different fragments are bolted together.
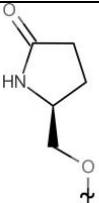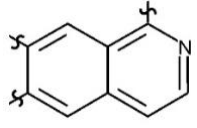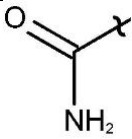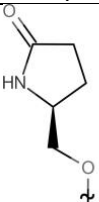
Ligand	Contribution of functional groups to the interaction						Total E_{inter} [%]
5UIQ (stage 1) (lead 51)	Carboxamide	Benzene	Isopropyl ether				
	43.4	0.4	56.1				100.0
Compound 10 (stage 2)	Carboxamide	Naphthalene	Isopropyl ether				
	50.3	9.5	55.9				115.7
5UIS (stage 3)	Carboxamide	Quinoline	Isopropyl ether	Piperidine ether			
	48.6	-12.3	59.4	52.9			148.6
5UIT (stage 4)	Carboxamide	Isoquinoline	Isopropyl ether	Gamma lactam ether			
	46.6	-12.2	57.5	65.1			157.0
5UIU (PF-06650833, stage 5)	Carboxamide	Isoquinoline	Methyl ether		Ethyl	Fluorine	

				Gamma lactam ether			
	50.7	-10.8	29.9	39.3	6.7	40.5	156.3

Table 2. IQA deformation and interaction energies related with the contact of the ligand and pocket (P) for different moieties (G) of the ligand throughout the FBDD stages examined herein. The contributions V_{cl} and V_{xc} are included as well. The data are reported in kcal/mol.

Moiety G	$E_{def\ G}$	$E_{inter\ G-P}$	$V_{xc\ G,P}$	$V_{cl\ G,P}$	$E_{def\ G} + E_{inter\ G-P}$
benzene	12.3	-0.7	-24.8	24.2	11.6
carboxamide	25.5	-64.1	-49.0	-15.1	-38.6
isopropyl ether	27.4	-82.8	-59.3	-23.5	-55.4
Inter-fragment interactions within the ligand $\sum_{G_1<G_2}\Delta E_{inter\ G_1-G_2}$	--	--	--	--	0.9
5UIQ (Stage 1)	**66.1**	**-147.5**	**-133.1**	**-14.4**	**-81.5**
Naphthalene	17.9	-14.1	-37.1	23.1	3.8
Carboxamide	27.5	-74.2	-54.3	-19.9	-46.7
	27.0	-82.4	-58.7	-23.7	-55.4

isopropyl ether					
Inter-fragment interactions within the ligand $\sum_{G_1 < G_2} \Delta E_{inter\ G_1 - G_2}$	--	--	--	--	0.4
Compound 10 (Stage 2)	**72.8**	**-170.7**	**-150.2**	**-20.5**	**-97.9**
Quinoline	12.6	18.1	-28.1	46.2	30.7
Carboxamide	27.5	-71.7	-52.2	-19.5	-44.2
isopropyl ether	31.9	-87.6	-64.9	-22.7	-55.7
piperidine ether	27.5	-78.0	-45.7	-32.3	-50.5
Inter-fragment interactions within the ligand $\sum_{G_1 < G_2} \Delta E_{inter\ G_1 - G_2}$	--	--	--	--	1.6
5UIS (Stage 3)	**101.2**	**-219.2**	**-190.9**	**-28.3**	**-118.1**
Isoquinoline	12.7	18.1	-25.7	43.7	30.8
Carboxamide	27.8	-68.8	-51.5	-17.3	-41.0
isopropyl ether	22.6	-84.8	-61.2	-23.7	-62.2

11

gamma lactam ether	22.9	-96.1	-51.3	-44.8	-73.2
Inter-fragment interactions within the ligand $\sum_{G_1<G_2}\Delta E_{inter\ G_1-G_2}$	--	--	--	--	23.0
5UIT (Stage 4)	**108.9**	**-231.6**	**-189.6**	**-42.0**	**-122.7**
Isoquinoline	13.5	15.9	-30.3	46.2	29.5
Carboxamide	29.8	-74.8	-57.4	-17.4	-45.0
Methyl ether	8.3	-44.2	-20.9	-23.3	-35.9
gamma lactam ether	26.1	-57.9	-52.7	-5.2	-31.8
Ethyl	7.2	-9.8	-14.3	4.5	-2.6
Fluorine	6.8	-59.8	-8.8	-51.0	-53.0
Inter-fragment interactions within the ligand $\sum_{G_1<G_2}\Delta E_{inter\ G_1-G_2}$	--	--	--	--	4.9
5UIU (Stage 5)	**96.7**	**-230.7**	**-184.4**	**-46.3**	**-133.9**

Regarding ligand 5UIQ or lead 51 as it was called in the original design work, we found that the ligand-pocket interaction was dominated in both exchange-correlation, $V_{xc}^{5UIQ,P}$, (in which P = pocket) and classical electrostatics ($V_{cl}^{5UIQ,P}$) by the isopropyl ether moiety. The contributions V_{cl} for the interaction between 5UIQ and the protein are -23.5, -15.1 and +24.2 kcal/mol for the isopropyl ether, the carboximade and benzene, respectively, as indicated in Table 2. The corresponding values for V_{xc} are -59.3, -49.0 and -24.8 kcal/mol respectively. This result is surprising because the carboxamide moiety forms two hydrogen bonds with residues Val263 and Met265. Still, the contributions of the isopropyl ether and the carboxamide groups to binding are

56.1 % and 43.4 % respectively (Table 1). The IQA interaction energy for the benzene moiety is slightly attractive to the protein pocket with a total 0.4% contribution to E_{inter}. Despite the two hydrogen bonds formed by the carboxamide group, the electrostatic classical component from the isopropyl ether moiety ($V_{cl}^{iPr,P}$) is larger in magnitude (-23.5 kcal/mol compared to -15.1 kcal/mol in the part corresponding to 5UIQ in Table 2). We also found evidence for an arene-H interaction, which involves the gatekeeper residue Tyr262. This interaction with Tyr262 is a crucial feature of the lead as this residue blocks access to an internal hydrophobic pocket and it effectively makes the binding site much smaller. The presence of the gatekeeper Tyr262 is exclusive to the IRAK family. The interaction between 5UIQ and Tyr262 is mostly covalent, roughly 90 % (Table S2).

The change from 5UIQ to compound 10, as it was called in the original design work, entailed the replacement of the benzene moiety with naphthalene. The naphthalene ring shows the best binding contributions from all the cyclic cores used in the design work (third column of Table 1). In other words, the IQA analysis indicates that the substitution of benzene by naphthalene is a good choice in this step of the FBDD study. The binding of naphthalene to the pocket was based solely on the exchange-correlation component $V_{xc}^{C_{10}H_6,P}$. As indicated in Table 2, $V_{xc}^{C_{10}H_6,P}$ = -37.1 kcal/mol. The corresponding classical electrostatic $V_{cl}^{C_{10}H_6,P}$ = 23.1 kcal/mol (Table 2) counterpart is repulsive, but the attraction from $V_{xc}^{C_{10}H_6,P}$ overcomes it. Still, the naphthalene moiety yields the smallest fragment IQA interaction energy in the contact of compound 10 with the protein pocket. The isopropyl ether and carboxamide moieties are virtually unaffected in both their deformation energies. The values of E_{def} for the isopropyl moiety in the 5UIQ and "Compound 10" stages are 27.4 kcal/mol and 27.0 kcal/mol, respectively (Table 2). Ditto for the carboxamide moiety with E_{def} = 25.5 kcal/mol and E_{def} = 27.5 kcal/mol for the 5UIQ and "Compound 10" ligands, respectively. As can be observed from Table 1, the contributions of these two moieties to ligand-pocket interactions remain practically unaltered in the change from stage 1 to stage 2 in the FBDD process. These groups have indeed the same behaviour as they do in stage 1, in accordance with the basic premise of FBDD. In other words, our IQA results prove that the basic premise of FBDD is justified in this case.

In the evolution from compound 10 to ligand 5UIS or compound 14 as it was originally called, a piperidine ether moiety was added and the central aromatic group was changed to quinoline. The carboxamide and isopropyl ether fragments remained unaffected in both deformation energies and IQA interaction energies concerning ligand-pocket binding as reported in Table 2. The values for E_{def}, for the isopropyl ether group are $E_{def, iso}$ = 27.0 kcal/mol and $E_{def, iso}$ = 31.9 kcal/mol for compound 10 and 5UIQ, respectively, while for the carboxamide moiety are $E_{def, carboxamide}$ = 27.5 kcal/mol for both stages 2 and 3 of the FBDD process. The IQA energies of these groups (i) $E_{inter}^{iso,P} = -82.4$ kcal/mol (for compound 10) and $E_{inter}^{iso,P} = -87.6$ kcal/mol (for 5UIUS) as well as (ii) $E_{inter}^{carboxamide,P} = -74.2$ kcal/mol (for compound 10) and $E_{inter}^{carboxamide,P} = -71.7$ kcal/mol (for 5UIUS) do not change substantially in the change from compound 10 to UIS. The new piperidine ether moiety is a strong contributor to the interaction between protein and ligand, even stronger than the carboxamide as reported in Table 1. The piperidine ether group forms a hydrogen bond with Ala315 and it accounted for nearly 90% of the contribution of the isopropyl ether (52.9% in comparison with 59.4% in the entries for 5UIS in Table 1). Nevertheless, the quinoline group was revealed to be a bad replacement for the naphthalene moiety in terms of the IQA energetics of the protein-ligand interaction. The quinoline moiety does not contribute favourably to binding in any of the last 3 FBDD steps in Table 1. There are no experimental results reported on how much the potency of the drug decreased when the naphthalene was switched to quinoline in stage 3. We can use the data in Table 1 to obtain an estimation for the impact of this replacement, using the fragmentation power of IQA to estimate this impact energetically. Let us consider a hypothetical compound that is identical to 5UIS but without the piperidine ether. The contribution of the three functional groups (carboxamide, quinolone and isopropyl ether) in this hypothetical ligand then amounts to 148.6% - 52.9% = 95.7% using the data in the

13

third row of Table 1. Thus, this hypothetical ligand would interact less strongly than compound 10 (115.7 %) by 20%. The relative contributions to the ligand-pocket interaction in Table 1 show that even benzene from the initial hit (lead 51 or 5UIQ) was a better option in terms of the energetics of the interaction than quinoline (and then isoquinoline in the two last stages). However, the bulkier double ring groups were preferred by the drug designers as a sacrifice serving better alternatives of fragment growing.

In the replacement of 5IUS (or compound 14) to 5UIT (compound 20) in the FBDD study, the designers decided to substitute the piperidine ether structure with a γ-lactam ether moiety. This fragment improved the energetics of the protein ligand interaction over its predecessor (157 % as compared with 148.6 % in Table 1) because the resulting ligand formed two hydrogen bonds with the peptide pocket. The hydrogen bond from the NH moiety to Ala315 was maintained and a new one from the γ-lactam carbonyl to Ser328 was established. The electrostatic component, $V_{cl}^{\gamma-lactam,P}$, accounted for 100 x -44.8 kcal·mol^{-1}/-96.1 kcal·mol^{-1} = 46.6 % of the interaction. The FBDD assumption was confirmed in this case, as the contributions to the interaction from all the previous moieties were maintained.

Finally, stage 5 of the FBDD study became the drug candidate PF-06650833, called compound 40 in the original study and nowadays called Zimlovisertib. This drug candidate has a PDB code 5UIU. As the clinical candidate, a substantial improvement over its precursor was expected. A couple of new moieties were added (an ethyl substituent to position β and a fluorine atom at position α of the γ-lactam). Additionally, the isopropyl ether moiety was changed to a methyl ether substituent. The relative contribution to the ligand-pocket interaction across all moieties was maintained as reported in the last two rows of Table 1, with the notable exceptions of the γ-lactam ether and the new methyl ether. We first conjectured that the differences between 5UIT and 5UIU were mostly caused by the α and β hydrogen atoms that are present in the γ-lactam ether in the former and not in the latter ligand because the substituents ethyl and fluorine replace those hydrogens. However, when we compared the equivalent remaining atoms in the γ-lactam, we found that their contributions to binding changed as revealed by the entries for the γ-lactam ether for 5UIT and 5UIU in Table 1. The thorough assessment of the corresponding interactions via IQA indicates that an important source for this difference is the α carbon atom which is bound to a fluorine atom in 5UIU (Table 3), while it forms part of an (unsubstituted) methylene group in stage 5UIT (Figure 1).

The drug designers saw an increase in potency of PF-06650833 compared to the de-fluoro analogue 5UIT. They hypothesised that the F atom increased the hydrogen bond donor capability of the γ-lactam. Our assessment of contributions per moiety does not support this hypothesis. Table 3 reports the IQA interaction energies computed with Equation (8) for the α and the carbonyl carbon in ligands 5UIT and 5UIU. Such IQA interaction energies constitute a critical part of the formation energy of the protein complex as indicated by Equation (11) and as discussed below. The only interactions that are substantially affected by the occurrence of the fluorine atom are those of the α carbon atom as can be seen in Table 3. Despite the substitution of -H by -F, the interaction energy between the carbonyl group of the γ-lactam and the protein pocket remains mainly unaltered. Interestingly, Table 3 shows that due to the bonding of such electronegative element, the contribution of the α-carbon atom to the ligand-pocket binding is highly repulsive, and this effect is solely caused by classical electrostatics.

Table 3. Difference between selected IQA interaction energies in systems 5UIT and 5UIU. The atoms considered in these values are, on one hand, the α and the carbonyl carbon atoms together with the atoms bonded to them in the γ-lactam ring, and on the other, the protein pocket. These differences are computed as $\Delta E_{inter}^{C-X,P} = E_{inter}^{C-X,P}(5UIU) - E_{inter}^{C-X,P}(5UIT)$ and ditto for the other quantities reported in the chart. Atom X represents H, F or O, depending on the atoms attached to the carbon atom under consideration. The data are reported in kcal/mol.

Carbon atoms in the γ–lactam	$\Delta E_{inter}^{C-X,P}$	$\Delta E_{inter}^{C,P}$	$\Delta V_{xc}^{C,P}$	$\Delta V_{cl}^{C,P}$	$\Delta E_{inter}^{X,P}$	$\Delta V_{xc}^{X,P}$	$\Delta V_{cl}^{X,P}$
C-H (carbon α)	−16.23	41.45	0.83	40.61	−57.68	−5.94	−51.74
C=O (carbonyl carbon)	2.62	5.20	−0.03	5.22	−2.58	−1.16	−1.41

The fluorine atom in Zimlovisertib shows an important attractive interaction with the pocket. It contributes more strongly to binding (40.5 %) that the rest of the atoms in the γ-lactam ether (39.26 %) as indicated in the last row of Table 1. The occurrence of F in the γ-lactam in 5UIU makes it roughly 15% stronger than the same moiety without the fluorine in 5UIT (40.5 % + 39.3% = 79.8 % compared to 65.1% in Table 1). The reduction of the contribution to binding of the α carbon atom after the fluorination in the last step of FBDD results in a severe decrease of the overall interaction of the equivalent atoms in the γ-lactam in 5UIU compared to 5UIT (65.1 % as opposed to 39.3 % in the rows corresponding to 5UIT and 5UIU in Table 1). Nevertheless, such a decrement is more than compensated by the contribution to binding of the fluorine atom as stated above.

We finally comment on the substitution of the isopropyl ether group for the methyl ether moiety in the design of 5UIU. This replacement reduces in half the interacting contribution compared to the equivalent moieties in previous FBDD stages: 29.9 % as compared to values which range from 55.9 % to 59. 4 % in previous stages of the FBDD process (Table 1). A Cheng-Prusoff analysis [71] shows a 0.4-fold decrease in potency when the isopropyl ether was substituted for the methyl ether in the last step of the examined FBDD process. We think that such a reduction is perhaps too high a price for the gained decrease in lipophilicity. Our results show that this ether moiety decreases the binding of the ligand as compared to the previous stage: at least a relative 25% (57.5 % as compared with 29.9 % in the last two rows of Table 1) of interaction energy is lost, as shown in Table 1. The internal energy assessment performed in IQA calculations only finds evidence against the substitution of the isopropyl ether moiety.

A large value of the IQA interaction energy is not the only factor to consider in IQA analyses. An atom can have strong interactions with other atoms or functional groups and yet the contact would not be stable because of the surge in steric energies due to strongly compressed basins against each other[72]. The energetics of any intermolecular contact between basins A and B result from the combined effect of the interaction and deformation energies (Equation (11)). In order to better understand this compromise, we calculated the deformation energies ($E_{def,G}$) for the different functional groups within the ligands in Table 1 along with their interactions with the pocket ($E_{inter\,G-P}$). In this way, not only can we determine which functional groups are most favourable to the ligand-pocket complexation, but we are also able to discern whether the stability brought by the interaction compensates the accompanying rise in the self-energy, i.e., deformation energy.

Table 2 shows the values for the IQA deformation energies of the different functional groups within the ligands considered in Table 1 as well as the interaction energies related to its non-covalent contacts with the pocket

15

for the five considered stages of the FBDD process. We note that for all the individual functional groups (and therefore for the ligands as well) the formation of ligand-pocket cluster is associated with a positive value of $E_{def,G}$. Likewise, the complexation of the ligand with the pocket is associated with a decrement of the interaction within the fragments in all the compounds considered throughout the FBDD process, i.e., $\sum_{G_1 < G_2} \Delta |E_{inter\, G_1 - G_2}| < 0$, wherein G_1 and G_2 are fragments within the ligand. As expected, $\sum_{G_1 < G_2} E_{inter\, G_1 - G_2} < 0$, and hence the fact that $\sum_{G_1 < G_2} \Delta E_{inter\, G_1 - G_2} > 0$ results in a reduction of the absolute value $|\sum_{G_1 < G_2} E_{int\, G_1 - G_2}|$. The value of $\sum_{G_1 < G_2} \Delta |E_{inter\, G_1 - G_2}|$ is relatively small for all stages apart from compound 5UIT in which $\sum_{G_1 < G_2} \Delta |E_{inter\, G_1 - G_2}| = 23.0$ kcal/mol. As a matter of fact, this contribution is critical to rationalise why the formation energy between the ligand and the IRAK-4 protein in stage 5UIU has a larger magnitude than it has in stage 5UIT. The larger deformation energy in 5UIT, which results from $\sum_{G_1 < G_2} \Delta |E_{int\, G_1 - G_2}|$ is one of the factors that determines that the formation energy of the ligand-protein complex in the last stage has a larger absolute value.

The consideration of the sum $E_{def,G} + E_{inter\, G-P}$ as reported in Table 2 and plotted in Figure 3, reflects stark differences among the different fragments of the ligand. For example, the aromatic cores (benzene, naphthalene, quinoline and isoquinoline) have small interaction energies with the pocket which do not compensate for the rise in their deformation energies. Indeed, $E_{def,G} + E_{inter\, G-P} > 0$ for all the aromatic cores in every stage of the investigated FBDD development as shown in the upper curve of Figure 3. The contribution of the inner fragments to the ligand-pocket contact, computed by considering both deformation and interaction energies, is relatively minor in the case of benzene and naphthalene as compared with other functional groups. Nevertheless, the situation drastically changes for quinoline and isoquinoline. The sum of deformation and interaction energies for these aromatic cores are similar in magnitude to other fragments but they oppose substantially to the contact. These results rule out the idea that the change of the central functional groups should barely affect the ligand-pocket recognition process because they are not in direct contact with the protein. Therefore, the alteration of these central moieties to facilitate the functionalisation of the drug under design can certainly affect the recognition of the ligand.

Concerning the rest of the moieties in the different stages through the investigated FBDD process, we found that the interaction energies are larger in absolute value than the corresponding deformation energies, and hence $E_{def,G} + E_{inter\, G-P} < 0$. Apart from the ethyl moiety in 5UIU, $E_{inter\, G-P}$ is considerably larger than $E_{def,G}$ for all the non-aromatic moieties of the ligands as can be observed in Table 2. Furthermore, the intervals for the interaction energy are much wider than those computed for the deformation energies. These observations suggest that $E_{inter\, G-P}$ might be a good indicator of the contribution to binding with a potential biological target for the functional groups that are in close contact with this protein. For example, the consideration of the deformation energy also supports the conclusion that the replacement of the isopropyl ether by a methyl ether does not contribute to the binding of the ligand to the protein. Furthermore, one can compute approximations to the classical and exchange-correlation components between fragments via Equations (12) and (13)[73, 74] :

$$E_{cl}^{AB} = \frac{Q^A Q^B}{R_{AB}}, \qquad (12)$$

$$E_{xc}^{AB} = -\frac{DI^{AB}}{2R_{AB}}, \qquad (13)$$

where Q^A is the QTAIM charge of atom A (likewise for Q^B) while R_{AB} and DI^{AB} are the distance between atoms A and B and its delocalisation index, respectively.

We would like to end this section by considering the basic premise of FBDD, i.e., that the contributions to binding of the different moieties of the ligand are not substantially altered throughout the FBDD design. Figure 3 shows that the most conserved moieties across the FBDD method, i.e., the carboxamide group (in stages 1-5) and the isopropyl ether (in stages 1-4) have remarkably similar values of $E_{def,G} + E_{inter\ G-P}$ across the total process. Concerning the atoms of the γ-lactam ether, we note that the inclusion of the fluorine atom alters strongly the sum of the interaction and the deformation energy of the ligand. The corresponding values for the γ-lactam ether are -73.2 kcal/mol and -31.8 kcal/mol in the 5UIT and 5UIU stages respectively as reported in the last column of Table 2. Therefore, we could expect that the fundamental assumption of FBDD is fulfilled throughout the conserved moieties of Zimlovisertib which were not attached to atoms that are powerful electron withdrawing groups such as the fluorine atom in the γ-lactam ether. We conjecture a similar situation for electron donating groups.
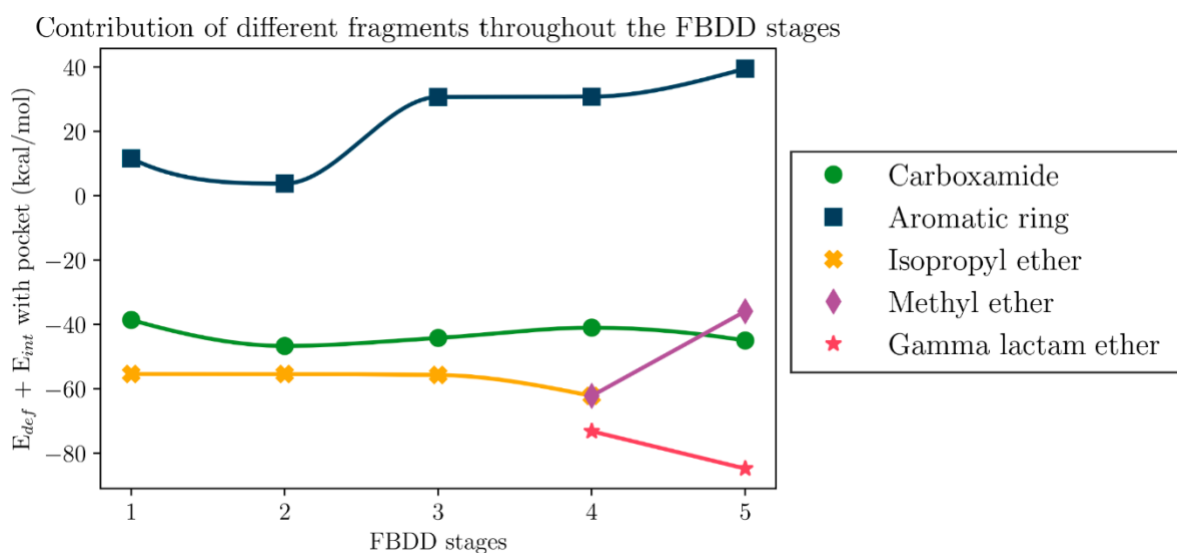
Figure 3. Values of $E_{def,G} + E_{inter,G-P}$ for different moieties of the ligands throughout the different stages of FBDD examined herein. The values are reported in kcal/mol.

Amino acid residue ranking

So far, we have discussed the relevance of the different fragments of the ligand to the energetics of the interaction of the considered drug candidates with the protein pocket. We now consider the importance of the surrounding amino acids in this energetics. The analysis of the contributions to the protein-ligand interaction from *specific* residues of the pocket permits finding the most significant amino acid residues for binding. We found that the top six residues (from a total of 14) contribute at least 75% of the total interaction towards the ligand for all studied FBDD stages (even up to 90% in the first two steps). The corresponding data for these amino acids can be found in Table S2. We analysed these amino acid residues that contribute the most to the binding with the ligand.

Hydrogen bonds play a crucial role in determining the specificity of ligand binding[75]. As expected, the contributions to the IQA interaction energy from the residues involved in hydrogen bonds are larger than those corresponding to other amino acids. Met265 is overall the residue with the strongest contribution to binding, as shown in Figure 4 and in Table S2. The backbone amide nitrogen in Met265 is a hydrogen bond donor in its

interaction with the carboxamide moiety of the investigated ligands. Table 4 shows the main features of this interaction, which is the strongest between the ligand and the pocket. The classical electrostatic component of the interaction of the hydrogen-bonded atom is preponderant (55-60%) throughout the whole process in the attractive contact of this QTAIM basin with the ligand (Table 4). On the other hand, the exchange-correlation part related to the nitrogen which acts as an H-bond donor is in favour of the ligand-pocket complexation while the classical electrostatic counterpart $V_{cl}^{N\cdots L}$ has the opposite behaviour as reported in Table 4. These results show that even among atoms involved in forming hydrogen bonds between the ligand and protein pocket, the different components that IQA describes can and should have different behaviours in favour of and against complexation.
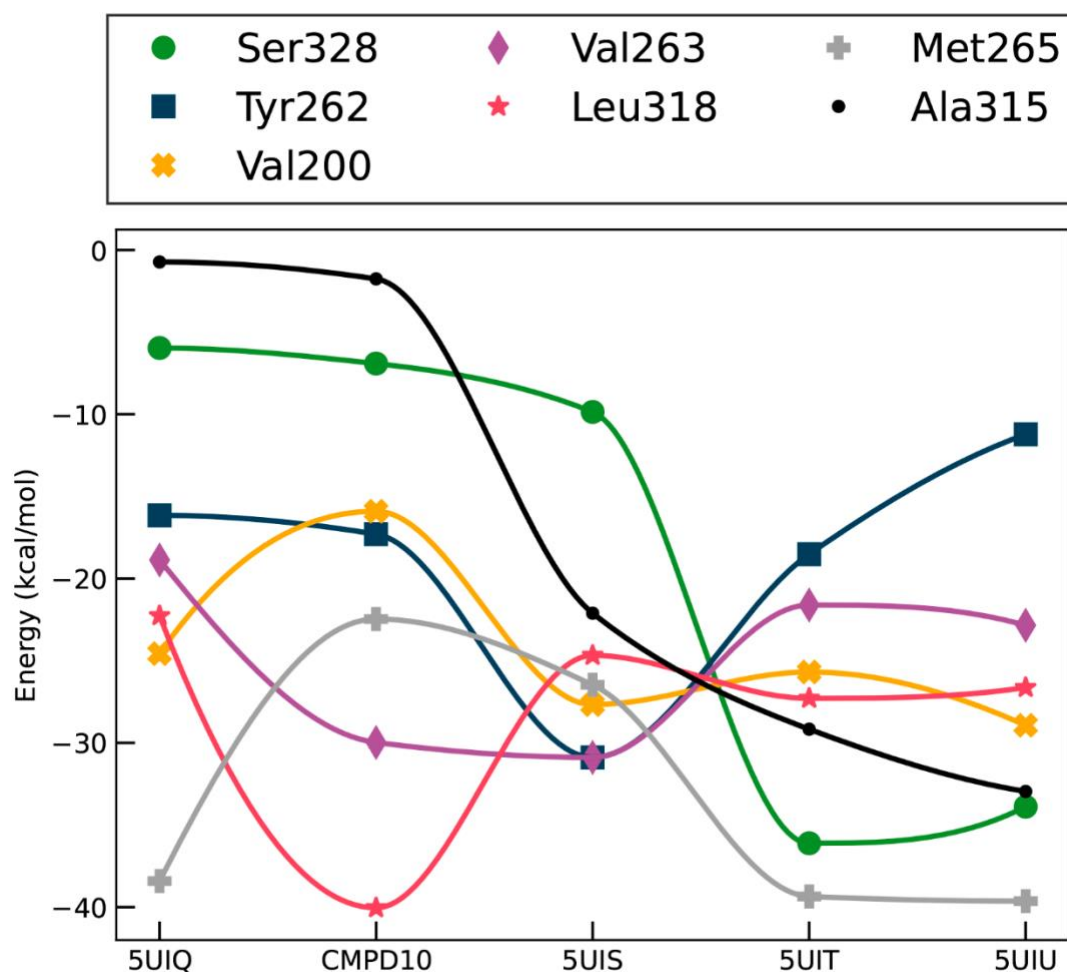
Figure 4. Contribution of the IQA interaction energy ($E_{inter}^{\mathcal{GH}}$ equation (8) wherein $\mathcal{G}$ is an amino acid and $\mathcal{H}$ is the ligand) to the formation of the protein-ligand cluster throughout the whole FBDD process examined herein. We show the results for only the seven amino residues with the largest values of IQA interaction energies ($E_{inter}^{\mathcal{GH}}$) with the potential pharmaceuticals throughout the whole FBDD process.

Table 4. IQA interaction energy (kcal/mol) as well as its classical and exchange-correlation components for the hydrogen bond donor and the hydrogen atom in the interaction between Met265 and the ligand. A negative/positive value in E_{inter} indicates an attractive/repulsive interaction.

Ligand	A	$E_{inter}^{A,ligand}$	$V_{xc}^{A,ligand}$	$V_{cl}^{A,ligand}$
5UIQ	N (Met 265)	12.5	-6.4	18.9
	H (Met 265)	-29.9	-13.5	-16.4
COMPOUND 10	N (Met 265)	15.2	-4.1	19.3
	H (Met 265)	-22.8	-9.2	-13.6
5UIS	N (Met 265)	17.5	-5.3	22.8
	H (Met 265)	-27.7	-10.4	-17.3
5UIT	N (Met 265)	13.5	-7.6	21.1
	H (Met 265)	-32.7	-14.5	-18.2
5UIU	N (Met 265)	16.3	-7.5	23.7
	H (Met 265)	-33.0	-13.7	-19.3

The data corresponding to the H-bonded atoms for Val263, Ala315 and Ser238 are reported in Tables S3-S5. The importance of hydrogen bonds suggests that residues (i) Met265 and Val263, as well as (ii) Ala315 and Ser328, which respectively form these interactions with the carboxamide and γ-lactam moieties, would be some of the most important amino acids in this regard. Since the γ-lactam moiety is only present in the ligand for stages 4 and 5, the average contribution ranking puts Ala315 and Ser328 as the sixth and seventh amino acids with the strongest contributions to binding (Table S2). However, when we consider only the last stage of the FBDD process, these residues are the second and third amino acids that contribute the most in the energetics of the ligand-pocket interaction, only behind Met265 as plotted in Figure 4. The presence of Leu318, Val200 and Tyr262 (in that order) as important contributors is surprising as there is no obvious hydrogen bond interaction involving these amino acids. This fact evidences the importance of other non-covalent interactions in the examined protein-ligand interactions.

Residue Leu318 interacts mainly via the two branches from its sidechain isopropyl group. This interaction occurs by way of CH$\cdots\pi$ contacts with the central cyclic moiety of the ligands. The corresponding IQA interaction energies are reported in Table S2. A noteworthy increase in stability occurs only in stage 2 because one of these branched methyl groups forms a C-H$\cdots$O hydrogen bond with the neighbouring oxygen from the isopropyl ether moiety. This hydrogen bond of Leu318 does not occur in any other step of the FBDD. Indeed, the contribution of this amino acid to the interaction is almost constant across the 5 stages (again, apart from step 2 in Table S2). Similarly, Val200 also interacts with the isopropyl group of its sidechain with the central cyclic aromatic moieties of the ligand, albeit from the other side of the ring. We note that the changes of the central group from benzene in stage 1 to bicyclic aromatic cores, have a very slight effect to the contribution from these residues. The interactions from the two residues Leu318 and Val200 are dominated by the exchange-correlation component (Table S2).

Tyr262, the gatekeeper residue and exclusive to the IRAK family of kinases, appears as the next ranked residue and the last one before the above-mentioned residues Ala315 and Ser328. The fact that Tyr262 is one of the top residues contributing to binding is consistent with the fact that PF-06650833 is so specific for proteins in the IRAK-4 family. We find that the carbon atoms in the aromatic ring of Tyr262 have favourable interactions towards the ligand. The corresponding IQA interaction energies are reported in Table S2. The original designers hypothesised that the isopropyl ether moiety from the ligand would interact with the aromatic core of this amino acid. Our results support this statement, because the contribution from Tyr262 to binding is affected by 10 kcal/mol when the isopropyl ether moiety is substituted by a methyl ether group in the change from 5UIT to 5UIU (Table S2). Finally, we would like to point out that a comparison between Figure 4 and Figure S2 reveals the importance of considering dispersion effects which are mostly absent in B3LYP, but they are recovered to some extent with the M06-2X functional.

Conclusions

We presented a partition of the formation energy of the complexes formed between (i) an amino acid pocket within the protein IRAK-4 and (ii) the different ligands throughout the FBDD process, which resulted in the conception of Zimlovisertib. This division of the formation energy of the investigated molecular clusters was performed with the IQA method, a physically sound approach based on state-of-the-art methods of wave function analysis such as the QTAIM.

The IQA dissection of the formation energy of the protein-ligand complex gives valuable insights with an atom-by-atom basis about the decisions made in the stages of an FBDD study. There are two ways of looking at the interaction: (i) per moiety (of the ligand) or (ii) contributions per residue of the protein pocket. The first approach allows to examine the suitability (or the lack of) the different alterations of the pharmaceutical throughout the FBDD process. For example, we can determine that the heterocyclic aromatic scaffolds, quinoline and isoquinoline, do not contribute to the attraction between the protein and the ligand. Ditto for the substitution of the isopropyl ether moiety with the methyl ether group at the end of the FBDD process. All these observations on the role of the fragments in the current drug candidate open the door for further, more targeted modifications in the future.

Regarding the investigated amino acids, we found that the residue Tyr262, which is exclusive of the IRAK family of kinases, is one of the largest contributors to binding. This observation rationalises the fulfilment of the objective to design a drug candidate with high specificity to the IRAK-4 family of kinases achieved by Pfizer. Altogether, and most importantly for the consequences of this investigation, the IQA approach provides a basis for the quantitative assessment of (i) the fulfilment of the basic premise of the FBDD design, i.e., how the interactions of different moieties of the ligand are affected throughout the whole FBDD process and (ii) the innocuousness of a given change of the ligand for its union with its biological target (e.g., the chemical modifications in a scaffold binding different functional groups). Furthermore, the IQA approach shown here can indicate a relevant interaction that could have been omitted in an FBDD design. Thus, we expect the IQA analysis to become an important tool to guide and to assess the decisions made throughout an FBDD approach.

- **Keywords:** COVID-19, Drug Design, Interacting Quantum Atoms, IRAK-4, Biological Chemistry and Chemical Biology.

References

[1] J. A. DiMasi, R. W. Hansen, H. G. Grabowski *J. Health Econ.* **2003**, *22*, 151-185.

[2] J. A. DiMasi, H. G. Grabowski, R. W. Hansen *J. Health Econ.* **2016**, *47*, 20-33.

[3] O. J. Wouters, M. McKee, J. Luyten *Jama*. **2020**, *323*, 844-853.

[4] F. J. Luque *Molecules*. **2018**, *23*, 2872.

[5] A. Mullard *Nature*. **2017**, *549*, 445-447.

[6] M. Bello, M. Martínez-Archundia, J. Correa-Basurto *Expert Opin Drug Discov*. **2013**, *8*, 821-834.

[7] M. Martínez-Archundia, M. Bello, J. Correa-Basurto *Methods Mol Biol*. **2018**, *1824*, 403-416.

[8] K. Martínez-Mayorga, A. Madariaga-Mazon, J. L. Medina-Franco, G. Maggiora *Expert Opin Drug Discov*. **2020**, *15*, 293-306.

[9] B. Lamoree, R. E. Hubbard *Essays Biochem*. **2017**, *61*, 453-464.

[10] G. Bollag, J. Tsai, J. Zhang, C. Zhang, P. Ibrahim, K. Nolop, P. Hirth *Nat. Rev. Drug Discov.* **2012**, *11*, 873-886.

[11] A. J. Souers, J. D. Leverson, E. R. Boghaert, S. L. Ackler, N. D. Catron, J. Chen, B. D. Dayton, H. Ding, S. H. Enschede, W. J. Fairbrother, D. C. S. Huang, S. G. Hymowitz, S. Jin, S. L. Khaw, P. J. Kovar, L. T. Lam, J. Lee, H. L. Maecker, K. C. Marsh, K. D. Mason, M. J. Mitten, P. M. Nimmer, A. Oleksijew, C. H. Park, C.-M. Park, D. C. Phillips, A. W. Roberts, D. Sampath, J. F. Seymour, M. L. Smith, G. M. Sullivan, S. K. Tahir, C. Tse, M. D. Wendt, Y. Xiao, J. C. Xue, H. Zhang, R. A. Humerickhouse, S. H. Rosenberg, S. W. Elmore *Nat. Med.* **2013**, *19*, 202-208.

[12] C. Jacquemard, E. Kellenberger *Expert Opin Drug Discov*. **2019**, *14*, 413-416.

[13] G. G. Ferenczy, G. M. Keserű *J. Chem. Inf. Model*. **2012**, *52*, 1039-1045.

[14] N. E. Genung, K. M. Guckian in *Chapter Four - Small Molecule Inhibition of Interleukin-1 Receptor-Associated Kinase 4 (IRAK4), Vol. 56* (Eds.: D. R. Witty, B. Cox), Elsevier, **2017**, pp.117-163.

[15] J. Hynes, S. K. Nair in *Chapter Nine - Advances in the Discovery of Small-Molecule IRAK4 Inhibitors, Vol. 49* (Ed. M. C. Desai), Academic Press, **2014**, pp.117-133.

[16] W. T. McElroy *Expert Opin Ther Pat*. **2019**, *29*, 243-259.

[17] P. N. Kelly, D. L. Romero, Y. Yang, A. L. Shaffer, 3rd, D. Chaudhary, S. Robinson, W. Miao, L. Rui, W. F. Westlin, R. Kapeller, L. M. Staudt *J. Exp. Med.* **2015**, *212*, 2189-2201.

[18] K. Kitaura, K. Morokuma *Int. J. Quantum Chem*. **1976**, *10*, 325-340.

[19] M. J. S. Phipps, T. Fox, C. S. Tautermann, C.-K. Skylaris *Chem. Soc. Rev.* **2015**, *44*, 3177-3211.

[20] L. Zhao, M. von Hopffgarten, D. M. Andrada, G. Frenking *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1345.

[21] J. Andrés, P. W. Ayers, R. A. Boto, R. Carbó-Dorca, H. Chermette, J. Cioslowski, J. Contreras-García, D. L. Cooper, G. Frenking, C. Gatti, F. Heidar-Zadeh, L. Joubert, Á. Martín Pendás, E. Matito, I. Mayer, A. J. Misquitta, Y. Mo, J. Pilmé, P. L. A. Popelier, M. Rahm, E. Ramos-Cordoba, P. Salvador, W. H. E. Schwarz, S. Shahbazian, B. Silvi, M. Solà, K. Szalewicz, V. Tognetti, F. Weinhold, É.-L. Zins *J. Comput. Chem.* **2019**, *40*, 2248-2283.

[22] I. Alkorta, J. C. R. Thacker, P. L. A. Popelier *J. Comput. Chem.* **2017**, *39*, 546-556.

[23] K. Raha, M. B. Peters, B. Wang, N. Yu, A. M. Wollacott, L. M. Westerhoff, K. M. Merz, Jr. *Drug Discov. Today*. **2007**, *12*, 725-731.

[24] J. C. Kromann, A. S. Christensen, Q. Cui, J. H. Jensen *PeerJ*. **2016**, *4*, e1994.

[25] F. Himo *J. Am. Chem. Soc.* **2017**, *139*, 6780-6786.

[26] N. Chaudhary, P. Aparoy *Heliyon*. **2020**, *6*, e04944.

[27] B. Thapa, K. Raghavachari *J. Chem. Inf. Model*. **2019**, *59*, 3474-3484.

[28] M. A. Blanco, A. Martín Pendás, E. Francisco *J.Chem.Theory Comput.* **2005**, *1*, 1096-1109.

[29] J. M. Guevara-Vela, E. Francisco, T. Rocha-Rinza, Á. Martín Pendás *Molecules*. **2020**, *25*, 4028.

[30] M. A. Vincent, A. F. Silva, P. L. A. Popelier *Z. Anorg. Allg. Chem.* **2020**, *646*, 1244-1251.

[31] R. W. F. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, Oxford, Great Britain, **1990**.

[32] P. L. A. Popelier, Atoms in Molecules. An Introduction., Pearson Education, London, Great Britain, **2000**.

[33] Z. Badri, C. Foroutan-Nejad, J. Kozelka, R. Marek *PCCP*. **2015**, *17*, 26183-26190.

[34] E. Romero-Montalvo, J. M. Guevara-Vela, W. E. Vallejo Narváez, A. Costales, Á. Martín Pendás, M. Hernández-Rodríguez, T. Rocha-Rinza *Chem. Commun.* **2017**, *53*, 3516-3519.

[35] P. Maxwell, Á. Martín Pendás, P. L. A. Popelier *PCCP*. **2016**, *18*, 20986-21000.

[36] J. M. Guevara-Vela, E. Romero-Montalvo, A. Costales, A. Martín Pendás, T. Rocha-Rinza *PCCP*. **2016**, *18*, 26383-26390.

[37] A. L. Wilson, P. L. A. Popelier *J.Phys.Chem.A*. **2016**, *120*, 9647-9659.

[38] K. Eskandari, M. Lesani *Chem.Eur.J.* **2015**, *21*, 4739-4746.

[39] F. J. Holguín-Gallego, R. Chávez-Calvillo, M. García-Revilla, E. Francisco, Á. Martín Pendás, T. Rocha-Rinza *J.Comput.Chem.* **2016**, *37*, 1753-1765.

[40] A. Fernández-Alarcón, J. L. Casals-Sainz, J. M. Guevara-Vela, A. Costales, E. Francisco, Á. Martín Pendás, T. Rocha-Rinza *PCCP*. **2019**, *21*, 13428-13439.

[41] O. A. Syzgantseva, V. Tognetti, L. Joubert *J. Phys. Chem. A*. **2013**, *117*, 8969-8980.

[42] B. C. B. Symons, D. J. Williamson, C. M. Brooks, A. L. Wilson, P. L. A. Popelier *ChemistryOpen*. **2019**, *8*, 560-570.

[43] N. Sukumar, C. M. Breneman *The Quantum Theory of Atoms in Molecules*. **2007**, 471-498.

[44] C. F. Matta *Foundations of Chemistry*. **2013**, *15*, 245-251.

[45] C. E. Whitehead, C. M. Breneman, N. Sukumar, M. D. Ryan *J. Comput. Chem.* **2003**, *24*, 512-529.

[46] P. L. A. Popelier *Current Topics in Med.Chem.* **2012**, *12*, 1924-1934

[47] P. Maxwell, The University of Manchester, **2017**.

[48] J. C. R. Thacker, A. L. Wilson, Z. E. Hughes, M. J. Burn, P. I. Maxwell, P. L. A. Popelier *Molecular Simulation*. **2018**, *44*, 881-890.

[49] Z. E. Hughes, E. Ren, J. C. R. Thacker, B. C. B. Symons, A. F. Silva, P. L. A. Popelier *J. Comput. Chem.* **2019**, *n/a*.

[50] P. L. A. Popelier in *On Quantum Chemical Topology, Vol.* (Eds.: R. Chauvin, C. Lepetit, E. Alikhani, B. Silvi), Springer, Switzerland, **2016**, pp.23-52.

[51] P. L. A. Popelier *AIP Conf.Proc.* **2012**, *1456*, 261-268.

[52] E. Francisco, J. L. Casals-Sainz, T. Rocha-Rinza, A. Martín Pendás *Theor. Chem. Acc.* **2016**, *135*, 170.

[53] A. Martín Pendás, M. A. Blanco, E. Francisco *J. Comput. Chem.* **2007**, *28*, 161-184.

[54] J. C. R. Thacker, P. L. A. Popelier *Theor. Chem. Acc.* **2017**, *136*, 86.

[55] Y. Zhao, J. Pu, B. J. Lynch, D. G. Truhlar *PhysChemChemPhys*. **2004**, *6*, 673-676.

[56] I. Cukrowski *PCCP*. **2019**, *21*, 10244-10260.

[57] Y. Zhao, D. G. Truhlar *Theor. Chem. Acc.* **2008**, *120*, 215-241.

[58] K. L. Lee, C. M. Ambler, D. R. Anderson, B. P. Boscoe, A. G. Bree, J. I. Brodfuehrer, J. S. Chang, C. Choi, S. Chung, K. J. Curran, J. E. Day, C. M. Dehnhardt, K. Dower, S. E. Drozda, R. K. Frisbie, L. K. Gavrin, J. A. Goldberg, S. Han, M. Hegen, D. Hepworth, H. R. Hope, S. Kamtekar, I. C. Kilty, A. Lee, L. L. Lin, F. E. Lovering, M. D. Lowe, J. P. Mathias, H. M. Morgan, E. A. Murphy, N. Papaioannou, A. Patny, B. S. Pierce, V. R. Rao, E. Saiah, I. J. Samardjiev, B. M. Samas, M. W. H. Shen, J. H. Shin, H. H. Soutter, J. W. Strohbach, P. T. Symanowicz, J. R. Thomason, J. D. Trzupek, R. Vargas, F. Vincent, J. Yan, C. W. Zapf, S. W. Wright *J. Med. Chem.* **2017**, *60*, 5521-5542.

[59] A. M. Clark, P. Labute *J. Chem. Inf. Model*. **2007**, *47*, 1933-1944.

[60] MOE in Molecular Operating Environment (MOE), Vol. (Ed.^Eds.: Editor), Chemical Computing Group Inc., City, **2017**.

[61] R. A. S. Laskowski, M.B. in LigPlot+: multiple ligand-protein interaction diagrams for drug discovery., Vol. (Ed.^Eds.: Editor), City, **2011**.

[62] A. Kuglstatter, A. G. Villasenor, D. Shaw, S. W. Lee, S. Tsing, L. Niu, K. W. Song, J. W. Barnett, M. F. Browner *J. Immunol.* **2007**, *178*, 2641-2645.

[63] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. Yang *J. Am. Chem. Soc.* **2010**, *132*, 6498-6506.

[64] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, W. Yang *J Chem Theory Comput*. **2011**, *7*, 625-632.

[65] T. A. Keith in AIMAll (Version 17.11.14), Vol. (Ed.^Eds.: Editor), City, **2017**.

[66] P. Hobza, J. šponer, T. Reschel *J. Comput. Chem.* **1995**, *16*, 1315-1325.

[67] S. Tsuzuki, H. P. Lüthi *J. Chem. Phys.* **2001**, *114*, 3949-3957.

[68] E. R. Johnson, R. A. Wolkow, G. A. DiLabio *Chem. Phys. Lett.* **2004**, *394*, 334-338.

[69] E. Torres, G. A. DiLabio *J. Phys. Chem. Lett.* **2012**, *3*, 1738-1744.

[70] E. R. Johnson, G. A. DiLabio *Chem. Phys. Lett.* **2006**, *419*, 333-339.

[71] Y. Cheng, W. H. Prusoff *Biochem. Pharmacol.* **1973**, *22*, 3099-3108.

[72] M. Gallegos, A. Costales, Á. Martín Pendás *ChemPhysChem*. **2021**, *22*, 775-787.

[73] M. Rafat, P. L. A. Popelier *The Quantum Theory of Atoms in Molecules*. **2007**, 121-140.

[74] A. Martín Pendás, E. Francisco *PCCP*. **2018**, *20*, 16231-16237.

[75] R. C. Wade, P. J. Goodford *Prog Clin Biol Res*. **1989**, *289*, 433-444.

Table of contents (ToC) entry: We divided the electronic energy throughout the Fragment-Based Drug Design (FBDD) process of the drug Zimlovisertib using the Interacting Quantum Atoms method. We could thereby establish the suitability of a chemical modification in each step of FBDD, providing an unprecedented and valuable test through an FBDD process.