Single-digit nanomolar inhibitors lock the aromatase active site via a dualsteric targeting strategy

(Article begins on next page)

18 November 2024

Single-digit nanomolar inhibitors lock the aromatase active site via a dualsteric targeting strategy

Jessica Caciolla,[a,#] Silvia Martini,[b,#] Angelo Spinello,[c] Federica Belluti,[a] Alessandra Bisi,[a] Nadia Zaffaroni,[b] Alessandra Magistrato,[d,*] Silvia Gobbi[a,*]

[a]Department of Pharmacy and Biotechnology, Alma Mater Studiorum-University of Bologna, via Belmeloro 6, 40126, Bologna, Italy

[b]Molecular Pharmacology Unit, Fondazione IRCSS Istituto Nazionale dei Tumori, via Amadeo 42, 20113, Milano, Italy

[c]Department of Biological, Chemical and Pharmaceutical Sciences and Technologies, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy

[d]National Research Council of Italy Institute of Materials (CNR-IOM) c/o SISSA, via Bonomea 265, 34136, Trieste, Italy

[#] These authors equally contributed to this manuscript
*Corresponding authors. E-mail addresses: silvia.gobbi@unibo.it (S. Gobbi), alessandra.magistrato@sissa.it (A. Magistrato)

Abbreviations: BC, breast cancer; ER+, estrogen receptor positive; SERM, selective estrogen receptor modulators; SERD, selective estrogen receptor degraders; AR, aromatase; AIs, aromatase inhibitors; LTZ, letrozole; MD, molecular dynamics; QM/MM, quantum mechanics/molecular mechanics; ΔG_b, binding free energy.

Abstract

The most frequently diagnosed breast cancer (BC) type in women expresses estrogen receptor (ER) and depends on estrogens for its growth, being classified as ER positive (ER+). The gold standard therapy for the treatment of this tumor relies on the inhibition of the aromatase enzyme, which catalyzes estrogen biosynthesis. Despite the clinical success of current aromatase inhibitors (AIs), after prolonged therapeutic regimens, BC ER+ patients experience acquired resistance and disease relapse. This points up the urgent need for a newer generation of AIs able to overcome resistance issues, while mitigating toxicity and side effects of current therapies.

Here we performed the synthesis, biological evaluation, and extensive structural characterization by advanced molecular simulation methods of a new generation of dualsteric non-steroidal AIs, which simultaneously target the enzyme's active and allosteric sites. Notably, **3d**, the most active AI of the series, exhibits a single-digit nM potency (IC$_{50}$ 2 nM). A detailed inspection of its binding mode reveals that the ancillary alkoxy chain predatorily takes advantage of the small hydrophobic cavities lining the allosteric site, triggering a remodeling of its residues and completely sealing the active site access-channel. As a result, the inhibitor is effectively locked in. This study sets a conceptual basis to develop a new generation of AIs exploiting a dualsteric targeting strategy.

Keywords: breast cancer, aromatase inhibitors, allosteric inhibition, molecular dynamics, QM/MM.

1. Introduction

Breast cancer (BC) represents the most common cancer among women, with almost 300K new estimated cases in the US in 2022 [1]. Among the different types of BC, the estrogen receptor-positive (ER+) subtype is the most frequent. In this type of BC the binding of estrogens to the estrogen receptor α (ERα) stimulates cell proliferation. As such, the tumor requires estrogens for its growth and progression, thus current mainstay therapies aim to suppress estrogen production/activity [2]. Indeed, the two currently available first-line therapies for the treatment of ER+ BC act either by blocking estrogen prooncogenic activity, *via* selective estrogen receptor modulators (SERMs, such as the prototype drug tamoxifen) [3] or degraders (SERDs, such as fulvestrant) [4] or by preventing estrogen production, through the inhibition of cytochrome P450 aromatase (AR), which catalyzes estrogen biosynthesis. [5, 6]. AR is expressed in distinct healthy tissues, such as ovaries and testes, placenta, adipose tissue, bone (chondrocytes and osteoblasts), vascular smooth muscle and numerous sites in the brain (hypothalamus, limbic system, cerebral cortex), but also in ER+ BC tumor cells [7]. Third generation AR inhibitors (AIs), such as letrozole, anastrozole and exemestane, have high selectivity for AR and do not interfere with the biosynthesis of other steroid hormones. Therefore, they have been FDA-approved as first-line agents for the treatment of post-menopausal women with ER+ BC. Anastrozole and letrozole are non-steroidal compounds containing azole moieties and, although their crystal structure in complex with the AR enzyme has not been solved to date, they are believed to coordinate the heme iron atom with their nitrogens [8]. Exemestane is a steroidal irreversible blocker that acts by mimicking the natural substrate of the enzyme [9]. Despite their remarkable therapeutic efficacy, all these marketed drugs are characterized by serious side effects and, alarmingly, their prolonged use triggers resistance onset and disease relapse in metastatic patients. Resistance onset owes to the identified somatic mutations of the ERα, which confer resistance to most currently available therapeutic strategies [10, 11]. To the best of our knowledge no resistant mutations of the AR enzymes have been hitherto identified. Side effects and resistance issues thus limit AIs therapeutic application and success [12]. For these reasons, the achievement of an effective and well-tolerated treatment for ER+ BC still represents a much needed, attractive, and challenging research field.

The need to overcome the different issues related to the use of AIs has led to investigating other potential mechanisms of inhibition of estrogens production, and recently allosteric inhibition of AR has emerged as an appealing alternative strategy [13]. Based on the non-competitive inhibition observed for tamoxifen metabolites [14], the mechanism of AR non-competitive inhibition was

extensively studied and, using different computational methods, three putative binding sites in different regions of AR structure were identified [15]. Among these, one was located in proximity of the most favorable access channel for the catalytic site, while the second was located close to the heme proximal cavity, involved in the electron transfer from cytochrome P450 reductase (CPR) to AR [16]. The binding of a ligand at these sites could interfere with AR activity, leading to enzymatic modulation. These findings paved the way for the design of potential allosteric modulators of the enzyme [17] or novel dual-mode inhibitors that could simultaneously interact with both the active site and one of the putative allosteric sites on AR [18].

1.1 Design strategy

As part of a long-lasting project aimed at the development of novel AIs, we recently made a first attempt to develop dual-mode inhibitors [19]. Briefly, a rigid alkoxy chain was inserted on some imidazolylmethylxanthones [20] that had previously shown competitive inhibitory activity against AR. In particular, a pentynyloxy chain was selected because it was reported to be the most suitable group to be inserted on the structure of exemestane [18]. This ancillary chain was shown to establish favorable interaction with the hydrophobic residues of the putative allosteric site placed at the AR access channel in the crystal structure. Among the new derivatives, **1** (Figure 1) showed the highest activity against AR (IC$_{50}$ value of 0.77 µM) [19]. However, the addition of a rigid side chain on the stiff xanthone core generally led to a drop in the inhibitory activity, since the inhibitor could not adapt to the rigid AR active site. In this work, to establish the most favorable structural features to reach and interact with the residues lining the enzyme access channel, we designed new derivatives with improved flexibility of the central core, also carrying tail-chains of different length and rigidity. Initially, the rigid xanthone core was replaced by a 4-imidazolylmethylbenzophenone scaffold, already reported by us in the structure of compounds endowed with moderate activity as AIs (**2e-h**, Figure 1) [21]. Based on this scaffold, the new derivatives **2a-d** (Figure 1) were designed. Moreover, since some of our 3-imidazolylmethylbenzophenones (**3g-h**, Figure 1) [22] had also proven to be potent AIs (IC$_{50}$ values of 5.3 and 7.3 nM, respectively), compounds **3a-f** (Figure 1) were conceived, enabling a more complete SAR investigation. The introduction of alkoxy tail-chains characterized by different structural features could provide precious information about the chemical space and diversity of ligands, able to establish optimal interactions with the AR allosteric site placed at the edge of the AR access channel and exert a potent inhibition.
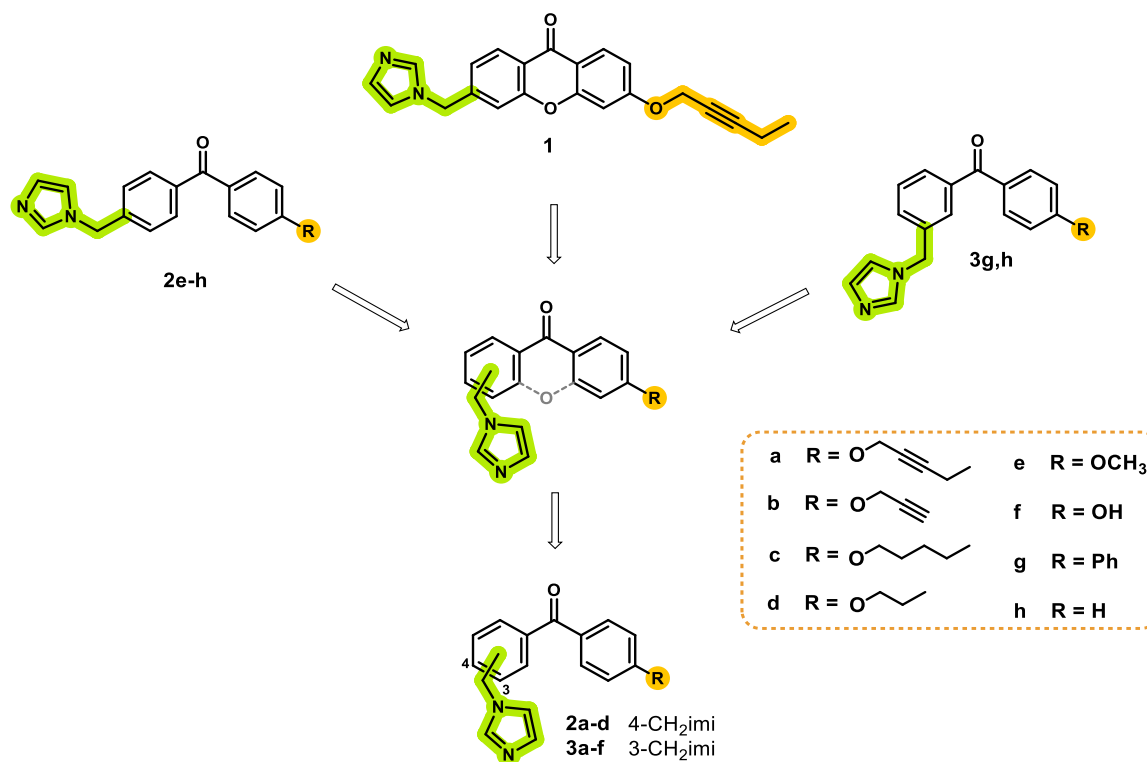
Figure 1. Structures of our previously reported compounds (**1**, **2e-h**, **3g,h**) and design strategy of new alkoxy imidazolylmethylbenzophenones (**2a-d** and **3a-f**).

2. Chemistry

For the synthesis of compounds **2a-d** (Scheme 1), the previously reported **2f** [21] was alkylated by reaction with the suitable alkyl bromide, in the presence of K_2CO_3 as base. While, for compounds **3a-f** bromination of the methyl group of (4-methoxyphenyl)(m-tolyl)methanone [23] with NBS followed by reaction with imidazole led to compound **3e**, which was then demethylated with 48 % HBr. The obtained hydroxyl derivative **3f** was then alkylated, in the same condition used for **2f**, to give **3a-d**.
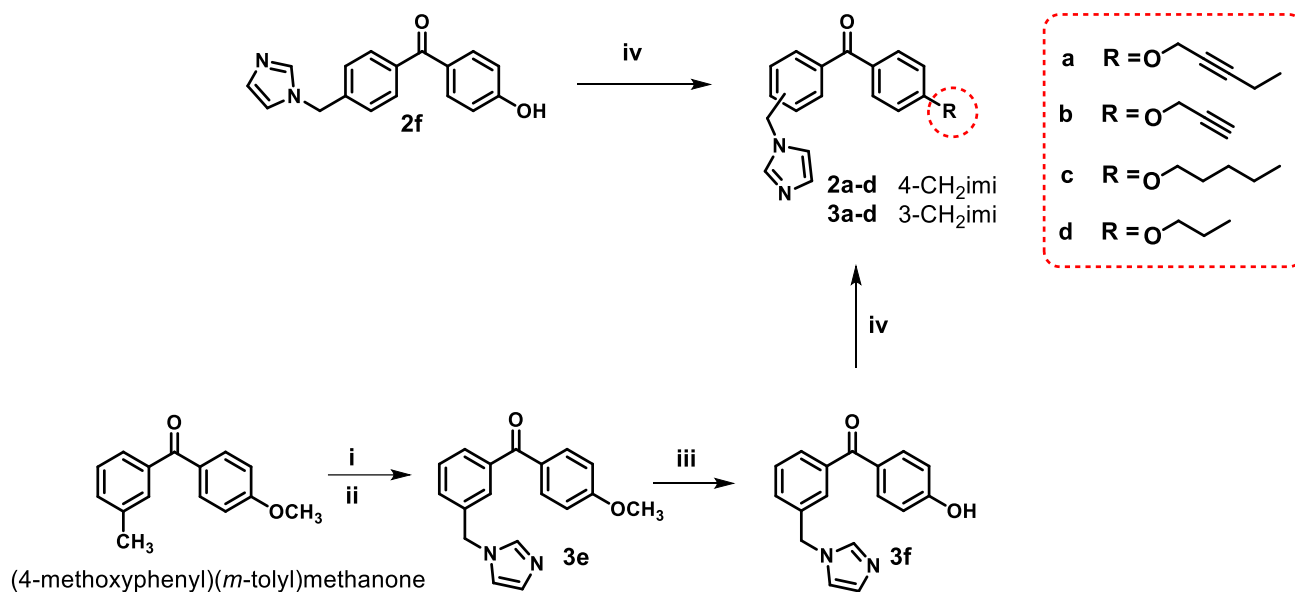
Scheme 1. Synthesis of compounds **2a-d** and **3a-f**. Reagents and conditions: i) NBS, BPO, CCl$_4$, hv, reflux 6-7 h; ii) imidazole, CH$_3$CN, N$_2$, reflux, 6 h; iii) 48 % HBr, reflux, 9 h; iv) selected alkyl bromide, K$_2$CO$_3$, acetone, reflux, 24 h.

3. Results

3.1 In vitro evaluation of AR inhibitory activity

The ability of the new compounds (**2a-d** and **3a-f**) to inhibit AR activity was tested at different concentrations, by monitoring the conversion of a fluorogenic substrate of the enzyme into a highly fluorescent metabolite, following an established protocol [19]. The results, expressed as IC$_{50}$ values, are reported in Table 1. AR inhibition data for previously described xanthone **1**, benzophenones **2e-h**, **3g-h**, and letrozole (LTZ) as reference were also reported to allow for a complete picture of the inhibition profile for the two series of structurally related compounds.

The replacement of the rigid xanthone nucleus of previously reported pentynyloxy compound **1** with a more flexible benzophenone core, maintaining the side chain in position 4' and the methylimidazole moiety in 4 (**2a**), caused a slight decrease in the activity (IC$_{50}$s of 770 and 1425 nM, respectively). Conversely, when the methylimidazole moiety was moved to position 3 in compound **3a**, a striking increase in potency was seen (IC$_{50}$ value of 9 nM) as compared to both **1** and **2a**. Briefly, except for compound **3b**, exhibiting a slightly higher IC$_{50}$ value than **2b** (650 nM vs 210 nM, respectively), 3-imidazolylmethylbenzophenones demonstrated stronger inhibitory potency with respect to the corresponding 4-imidazolylmethyl derivatives.

Considering the side chain in the 4-imidazolylmethyl series **2a-h**, the reduction of the length of the alkoxy tail from five carbon atoms (**2a** and **2c**, IC_{50} of 1425 and 1300 nM, respectively) to three (**2b** and **2d**, IC_{50} values of 210 and 374 nM, respectively) led to an improvement in activity, while its rigidity did not play a key role. An additional increase in activity was observed by further shortening the tail to one carbon atom (**2e**, IC_{50} value of 237.8 nM) or by removing it (**2h**, IC_{50} value of 252.4 nM). Contrariwise, the replacement of the alkoxy chain with a hydroxyl (**2f**), or a phenyl group (**2g**) induced a decrease in the activity (IC_{50} values of 1100 and 2531 nM, respectively).

Table 1. AR inhibition of studied imidazolylmethylbenzophenones **2a-h** and **3a-h**.
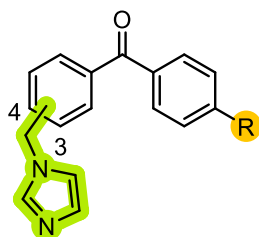
Compound	CH2-imi position	R	AR inhibition IC_{50} nM[a]
1[b]	-	-	765 ± 276[b]
2a	4		1425 ± 530
2b	4		210 ± 160
2c	4		1300 ± 100
2d	4		374 ± 230
2e[c]	4	OCH_3	237.8[c]
2f[c]	4	OH	1100[c]
2g[c]	4	Ph	2531[c]
2h[c]	4	H	252.4[c]
3a	3		9 ± 1
3b	3		650 ± 212
3c	3		42 ± 6
3d	3		2 ± 2

3e	3	OCH$_3$	13 ± 11
3f	3	OH	13 ± 10
3g[d]	3	Ph	5.3[d]
3h[d]	3	H	7.3[d]
LTZ[e]	-	-	10[e]

[a]Compound concentrations inhibiting by 50% AR activity. Data represent the mean ± SD of at least three independent experiments performed in triplicate. [b]ref [19]; [c]ref [21]; [d]ref [22]; [e]ref [24]

The role of the side chain in the 3-imidazolylmethyl series **3a-h** was also evaluated. Considering the compounds carrying a rigid unsaturated alkoxy chain (**3a** and **3b**), the nanomolar activity of pentynyloxy derivative **3a** (IC$_{50}$ of 9 nM) was significantly reduced by shortening the tail to three carbon atoms (**3b**, IC$_{50}$ of 650 nM). Conversely, in derivatives with a saturated and flexible tail a 20-fold increase of inhibitory activity was observed going from five carbon atoms (**3c**, IC$_{50}$ of 42 nM) to three (**3d**, IC$_{50}$ of 2 nM). Moreover, activity was maintained in the low nanomolar range when the chain was further shortened to one carbon atom (**3e**, IC$_{50}$ of 13 nM)), substituted with a hydroxyl (**3f**, IC$_{50}$ of 13 nM) or a phenyl group (**3g**, IC$_{50}$ of 5.3 nM) or even removed (**3h**, IC$_{50}$ of 7.3 nM). Considering the obtained results, analogue **3d**, carrying the methylimidazole moiety in position 3 and a saturated three carbon atoms alkoxy chain, emerged as the best compound of the whole series, with an IC$_{50}$ value of 2 nM.

The effects of the studied compounds on cell growth inhibition were also investigated. To this aim, two breast cancer cell lines, namely MCF-7 (ER$^+$) and MDA-MB-231 (ER$^-$), were treated with different concentrations of the compounds for 72 h in the absence or in the presence of testosterone. In order to evaluate the potential adverse effects, the compounds were also screened against the MCF-10A healthy breast cell line. The activity was then detected by cell counting assay. Despite the strong inhibition of AR in cell-free tests, none of the compounds showed a direct effect on growth inhibition in the examined cancer cell lines up to a concentration of 50 µM (data not shown). The lack of antiproliferative effect observed in the MCF-7 cell line could be explained, at least in part, by the low level of AR activity, despite the presence of *AR* gene amplification, in these cells [25, 26] which also makes them not susceptible to testosterone-mediated growth stimulatory effect. Notably, the compounds did also not display any antiproliferative activity on healthy breast cells, thus excluding they could induce off-target effects.

3.2 All-atom simulations

To provide a detailed structural analysis of the interactions established by the newly designed inhibitors and the AR enzyme we performed classical molecular dynamics (MD) and mixed quantum-classical (QM/MM) MD simulations. In particular, we focused on 3-imidazolylmethylbenzophenones (**3a-3h**, Table 1) as the most potent series of inhibitors.

First, we have performed docking calculations of the investigated inhibitors restraining the coordination bond between a nitrogen atom of the imidazole ring and the iron of the heme moiety as performed in our previous study [19]. The introduction of this metal-ligand restraint was necessary to obtain binding poses in which the nitrogen atom was placed at coordination distance from the iron atom of the heme moiety, which is not properly described in docking simulations. Due to the high computational cost of the QM/MM MD simulations, we have investigated in more detail only the 3-imidazolylmethylbenzophenones (**3a-3h**), which have shown lower $IC_{50}s$ compared to the corresponding 4-substituted derivatives. Each AR/inhibitor complex was initially relaxed by a 100 ns-long classical MD simulation (RMSD is shown in Figure S1), followed by 10 ps-long QM/MM MD simulations, to accurately describe the metal-ligand coordination bond, while explicitly considering the surrounding biomolecular environment. Indeed, classical MD simulations are inaccurate in describing metal-ligand interactions, and require a more accurate, yet more computationally expensive QM-based description [27, 28] Moreover, we also computed compound **2b**, which gave the most potent inhibitory effect among the 4-imidazolylmethylbenzophenones, to find a rationale between the position of the side chain and the measured $IC_{50}s$.

In line with previous studies [19, 29], all-atom simulations revealed that all the investigated inhibitors can form stable coordination bonds with the iron atom of the heme (Figure 2), and a detailed structural analysis disclosed that their coordination geometry (i.e. coordination bond lengths and angles) was not strictly linked to their measured potency (Table 2). Notably, the lack of hydrogen bonds with the AR active site residues was the largest difference with the reference inhibitor LTZ.

Table 2. Key structural parameters for the binding of the inhibitors **3a-3h** and **2b** within AR active site. The coordination bonds (Angstrom) and angles (deg) of the imidazole moiety of the ligand to the heme iron atom are reported. The clinically used inhibitor letrozole (LTZ) is also reported for comparison.

	IC$_{50}$ (nM)	Distance (Fe -N) Å	Angle (planes) deg
LTZ	10[a]	2.33 ± 0.15[b]	91.8 ± 2.7[b]
3a	9 ± 1	2.17 ± 0.09	88.0 ± 4.2
3b	650 ± 212	2.21 ± 0.13	90.1 ± 3.9
3c	42 ± 6	2.18 ± 0.10	86.5 ± 3.6
3d	2 ± 2	2.22 ± 0.12	85.5 ± 3.3
3e	13 ± 11	2.22 ± 0.10	87.5 ± 4.2
3f	13 ± 10	2.24 ± 0.13	88.4 ± 3.5
3g	5.3[c]	2.23 ± 0.11	91.4 ± 3.2
3h	7.3[c]	2.18 ± 0.08	93.2 ± 3.9
2b	210 ± 160	2.20 ± 0.10	90.9 ± 2.7

[a]ref [24], [b]ref [19], [c]ref [22]
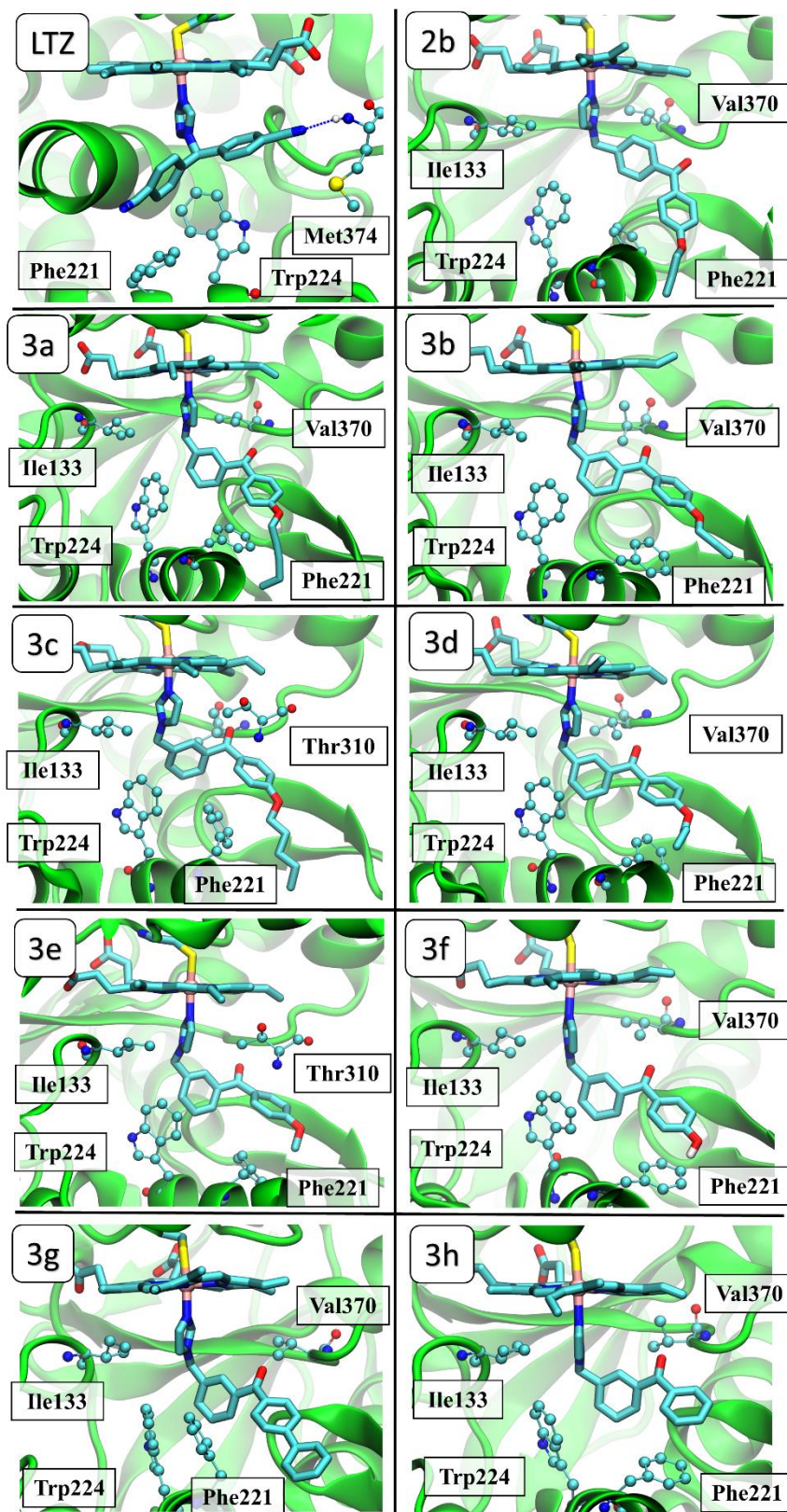
Figure 2. Representative structures as obtained from QM/MM molecular dynamics trajectories of the coordination complex between letrozole (LTZ) and selected inhibitors **2b** and **3a-3h** and the aromatase (AR)

active site. The AR structure is displayed in green new cartons, the heme moiety, Cys437 and the inhibitors are shown in licorice, the residues mostly contributing to the binding are shown in balls and sticks. Atoms are colored by atom name (O red, S yellow and N blue, C cyan).

Next, in order to dissect the most relevant intermolecular interactions of the inhibitors within the AR catalytic site, we have performed a per-residue decomposition analysis of the binding free energy (ΔG_b) with the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method [30] This type of analysis, being performed at force field level on selected equally spaced frames extracted from QM/MM trajectories, neglects the contribution of the inhibitor-iron coordination bond. Nevertheless, it provides important insights on the relevant non-bonded interactions (hydrophobic or electrostatic interactions) between the inhibitors and AR enzyme. This analysis revealed that all the inhibitors preferentially engage hydrophobic interactions with active site residues, which indeed display a highly hydrophobic character. Of particular relevance are i) the π-π stacking interactions between the aromatic rings of the inhibitors and Phe221 and Trp224; ii) the hydrophobic interactions with Thr310, a residue that plays a key role in the AR catalytic mechanism [31] (Table 3).

Table 3. Decomposition of the binding free energy (ΔG_b, kcal/mol) obtained with Molecular Mechanics Generalized Surface Area (MM-GBSA) method, along with their standard deviations, for the binding of compounds **3a-3h** and **2b** to the aromatase (AR) enzyme as obtained from QM/MM molecular dynamics trajectories. Residues involved in stabilizing the inhibitor binding pose are marked in yellow, in light green and in dark green when their contribution to ΔG_b is smaller than -1.0 kcal mol, when it ranges between -1.1 and -2.0 kcal/mol and when it larger than -2.1 kcal/mol, respectively.

	3a	3b	3c	3d	3e	3f	3g	3h	2b
IC$_{50}$	9 ± 1	650 ± 212	42 ± 6	2 ± 2	13 ± 11	13 ± 10	5.3	7.3	210 ± 160
Ile133	-1.0 ± 0.2	-0.6 ± 0.1	-1.4 ± 0.3	-1.4 ± 0.1	-1.3 ± 0.2	-1.5 ± 0.2	-1.3 ± 0.2	-1.4 ± 0.2	-1.2 ± 0.2
Phe221	-2.5 ± 0.5	-2.4 ± 0.5	-2.5 ± 0.5	-2.3 ± 0.4	-1.7 ± 0.4	-2.0 ± 1.7	-2.1 ± 0.4	-1.3 ± 0.3	-2.2 ± 0.4
Trp224	-1.3 ± 0.3	-1.0 ± 0.3	-1.2 ± 0.3	-1.7 ± 0.3	-1.6 ± 0.3	-1.7 ± 0.3	-1.4 ± 0.3	-1.7 ± 0.3	-1.6 ± 0.3
Thr310	-1.1 ± 0.4	-1.8 ± 0.4	-1.9 ± 0.4	-1.8 ± 0.3	-2.1 ± 0.3	-2.0 ± 0.3	-1.7 ± 0.3	-1.8 ± 0.3	-1.7 ± 0.2
Val313	-1.1 ± 0.4	-1.2 ± 0.3	-1.0 ± 0.3	-1.0 ± 0.2	-0.9 ± 0.3	-0.8 ± 0.2	-1.1 ± 0.3	-0.7 ± 0.2	-1.2 ± 0.3
Val369	-1.0 ± 0.4	-0.9 ± 0.3	-0.8 ± 0.4	-0.5 ± 0.2	-0.7 ± 0.2	-0.6 ± 0.2	-0.6 ± 0.2	-0.6 ± 0.3	-0.8 ± 0.3
Val370	-1.1 ± 0.3	-1.0 ± 0.3	-1.3 ± 0.4	-1.2 ± 0.2	-0.7 ± 0.2	-1.2 ± 0.3	-1.0 ± 0.2	-1.0 ± 0.2	-0.8 ± 0.2
Leu477	-1.0 ± 0.3	-0.4 ± 0.1	-0.6 ± 0.4	-1.0 ± 0.2	-1.2 ± 0.3	-1.1 ± 0.3	-1.3 ± 0.3	-1.0 ± 0.2	-1.0 ± 0.2

These combined computational approaches, using levels of theory of different and increasing accuracy, allowed to rationalize the IC_{50}s values observed experimentally. As an example, in the case of ligand **3b**, which showed the worst IC_{50} value among the 3-methylimidazole series, the rigid propargyloxy tail is not properly oriented along the aromatase access channel. Indeed, it protrudes towards the channel side and clashes on Pro481 and Leu188. This leads to a decrease in the ΔG_b values of some residues composing the binding pocket (e.g. Ile133, Tpr224, Table 3) thus disfavoring the binding of the inhibitor. Although the functionalization of the imidazole moiety in position 4 (series **2a-2h**) generally showed higher IC_{50}s values, in compound **2b** the propargyloxy tail is better oriented along the access channel as compared to the related inhibitor **3b** (Figure S2), leading to a more potent inhibition. This also explains why in the 3-methylimidazole series of compounds the inhibitory potency increases by making the alkoxy tail more flexible (**3c**) or even better by reducing its length (**3d-f**). At the opposite **3g**, (IC_{50} of 5.3 nM), in which the tail is replaced by a phenyl group, completely fills the allosteric cavity and creates a rigid hydrophobic gate to the entrance of water molecules in the active site (Figure 3A) and to the exit of the inhibitor. Surprisingly, a remarkable inhibitory activity remains even when the tail is removed in **3h** because, although the inhibitor does not reach the allosteric cavity, the non-functionalized aromatic ring of the benzophenone triggers a remodeling of the hydrophobic residues Val313, Phe221 and His480, which create a barrier to inhibit the entrance of incoming water molecule and the dissociation of the inhibitor. Strikingly, the functionalization with a propoxy chain in **3d** achieves the best inhibitory potency. Here, the alkoxy group settles within a hydrophobic pocket formed by Phe221, Ile217 and the π-electrons of Gln191. This triggers a remodeling of Gln191, Arg192, Glu483 and His480, which by establishing an extended H-bonds network effectively lock up the active site (Figure 3B).
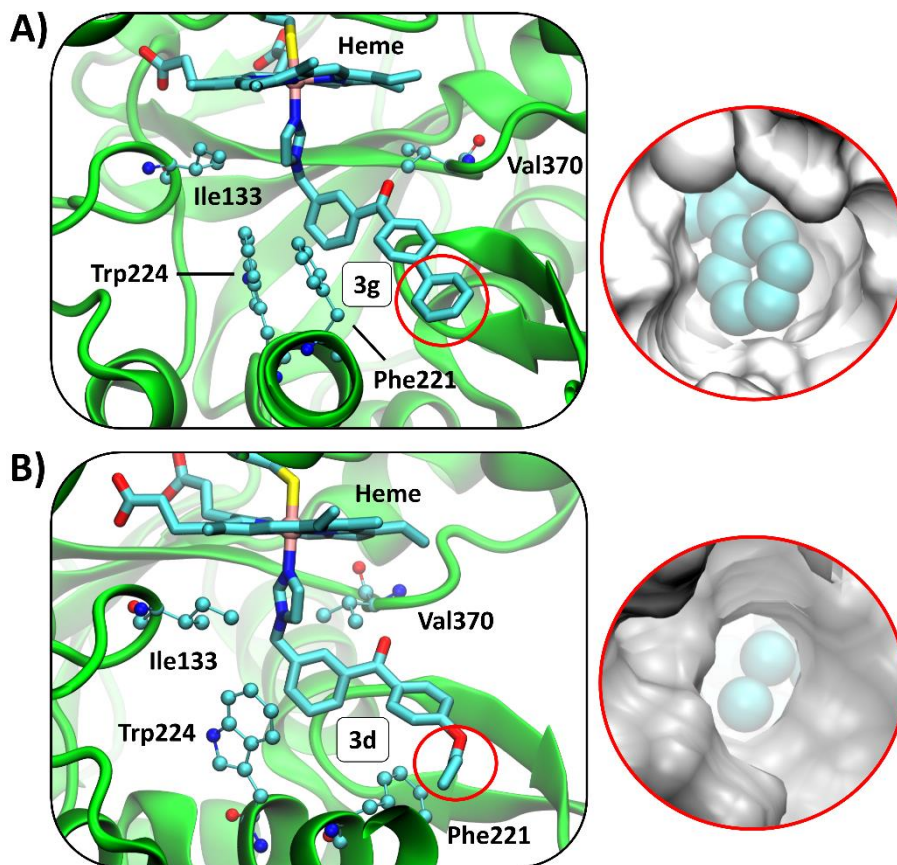
Figure 3. Representative structures extracted from QM/MM molecular dynamics simulations of A) compound **3g** and B) **3d** in complex with the aromatase (AR) active site. The insets show the occlusion of the aromatase access channel induced by the presence of the inhibitors. The AR structure is displayed in green new cartons, the heme moiety, Cys437 and the inhibitors are shown in licorice, the residues mostly contributing to the binding are shown in balls and sticks. In the insets we provide a view of the entrance to the active from the access-channel perspective. The protein is shown as a silver surface, while the compounds are displayed as van der Waals spheres. Atoms are colored by atom name (O red, S yellow and N blue, C cyan).

These findings suggest that the functionalization in position 3 of the benzophenone ring is in general required to achieve an optimal coordination geometry to the heme iron atom. Moreover, the allosteric site can be filled by a bulky substituent, such as a phenyl group (e.g. compound **3g**), which by occupying the AR access tunnel creates a rigid gate preventing the entrance of water/substrate molecules into the active site. Conversely, for less bulky substituents, the optimal activity is achieved by the flexible and medium-length hydrophobic tail, which exploits a small hydrophobic cavity on the access channel, inducing the remodeling of the flanking residues and the formation of an extended H-bond network which completely seals the active site. As such, the length and flexibility of the tail enable to establish the optimal set of non-bonded interactions which drive the active site closure.

In order to exclude the possibility that the compounds could exert a non-active site directed inhibition mechanism, docking studies were also performed on the previously identified allosteric sites laying in the AR access channel [16] and in the cytochrome P450 reductase binding interface [17]. The docking scores of compounds **3a-3h** and **2b**, when docked in the allosteric sites, were smaller than that of the allosteric inhibitor identified in our previous study [17], active in the μM range. This suggests that their binding at those sites is unlikely, and most likely it is not responsible for their observed inhibitory potencies (Table S1).

3.3 Discussion and conclusions

In this study, a structural simplification approach was attempted, with the aim to get insight into the dualsteric potential and improve the potency of our previously described imidazolylmethylxanthone **1**. This was accomplished by decreasing the rigidity of the central core through the removal of the ethereal oxygen bridge between the aryl rings. Notably, for the corresponding benzophenone **2a** no increase in activity was observed, proving an inappropriate fitting of this more flexible molecule in the active site of the enzyme following the primary interaction of the imidazole with the heme iron. Nevertheless, the modification of the pentynyloxy side chain of **2a** protruding towards the allosteric site placed at the edge of the substrate access channel, to give **2b-2h**, was seen to impact on the potency of this set of derivatives. Briefly, shorter alkoxy chains partially restored the activity, while the polar hydroxy and the bulky phenyl derivatives proved to be the least active of this subset.

Remarkably, the shift of the imidazolylmethyl moiety from position 4 (compound **2a**) to position 3 (**3a**) on the benzophenone core alone proved to strongly affect the potency of the AI. This result validated our design hypothesis postulating a boost in potency through an increase in flexibility of the molecule, further proving key role of the correct positioning of this crucial moiety for the appropriate interaction with the enzyme. Indeed, **3a** proved to be two orders of magnitude more potent than both **2a** and the xanthone **1**. Modification of the pentynyloxy side chain to give **3b-3h** did not seem to play a major role in the activity of this subset of compounds that, with the exception of **3b**, all showed nanomolar potency. Still, inhibitory potency was increased by more flexible or shorter alkoxy chains, which proved to be able to plastically adapt to the allosteric site by interacting with residues in the access channel area. Notably, the most potent derivatives **3d**, **3g** and **3h** were all seen

to induce a significant remodeling of crucial AR residues that locked the orthosteric site, effectively inhibiting enzymatic activity.

In conclusion, our findings unequivocally disclose that the dualsteric targeting strategy, which is exemplified by inhibitors able to simultaneously target the orthosteric and a flaking allosteric site, appears to be effective for the AR enzyme. Yet, an optimal inhibitory activity is achieved by a combination of different structural features: the ability to optimally establish a coordination bond to the active site (attained with a properly oriented imidazole moiety), the flexibility of the inhibitor core (obtained by shifting to the benzophenone) and the flexibility, length and orientation of its substituent. Here, the goal of an optimal AR inhibitory potency is accomplished with a molecule which, thanks to its ancillary substituent, predatorily exploits the small cavities present in the allosteric site. As a result, the inhibitor locks up and completely seals the active site by remodeling the residues of the allosteric site. The excellent activity of the herein newly developed inhibitors contributes to expand the currently limited portfolio of dual targeting AIs. Furthermore, by rationalizing the critical structural elements underlying their activity, this study sets a conceptual basis for future drug-discovery campaigns aiming to develop a new effective generation of dualsteric inhibitors, able to resolve or ameliorate the selectivity and resistance issues of current therapies.

4. Experimental Section

4.1 Chemistry

4.1.1 General Methods

All chemicals were purchased from Aldrich Chemistry, Milan (Italy), or Alfa Aesar, Milan (Italy), and were of the highest purity grade. Melting points were determined in open glass capillaries using a Büchi apparatus and are uncorrected. ^{1}H NMR and ^{13}C NMR spectra were recorded in CDCl$_3$, unless otherwise indicated, on a Varian VXR Gemini spectrometer working at 400 MHz and 101 MHz, respectively. Chemical shifts are reported in parts per million (ppm) relative to tetramethylsilane (TMS) as internal standard and spin multiplicities are given as s (singlet), d (doublet), t (triplet), m (multiplet) or br (broad). Chromatographic separations were performed by flash column chromatography on silica gel columns (Kieselgel 40, 0.040-0.063 mm; Merck). Organic

solutions were dried over anhydrous sodium sulfate. UHPLC−MS analyses were run on a Waters ACQUITY ARC UHPLC/MS system consisting of a QDA mass spectrometer equipped with an electrospray ionization interface and a 2489 UV/Vis detector. The detected wavelengths (λ) were 254 nm and 365 nm. The analyses were performed on an XBridge BEH C18 column (10 × 2.1 mm i.d., particle size 2.5 μm) with a XBridge BEH C18 VanGuard Cartridge precolumn (5 mm × 2.1 mm i.d., particle size 1.8 μm). The mobile phases were H_2O (0.1% formic acid) (A) and MeCN (0.1% formic acid) (B). Linear gradient: 0−0.78 min, 20% B; 0.78-2.87 min, 20−95% B; 2.87-3.54 min, 95% B; 3.54-3.65 min, 95-20% B; 3.65-5-73, 20% B. Flow rate: 0.8 mL/min. Electrospray ionization in positive and negative mode was applied in the mass scan range 50−1200 Da. All tested compounds were found to have > 95% purity. Compounds were named relying on the naming algorithm developed by CambridgeSoft Corporation and used in ChemDraw Professional 20.0.

4.1.2 (3-((1*H*-imidazol-1-yl)methyl)phenyl)(4-methoxyphenyl)methanone (3e). A mixture of (4-methoxyphenyl)(m-tolyl)methanone (1.09 g, 4.80 mmol), *N*-bromosuccinimide (NBS, 0.85 g, 4.80 mmol) and a catalytic amount of benzoyl peroxide (BPO) in CCl_4 (20 mL) was refluxed for 6 h. The mixture was hot filtered and the solvent was evaporated under reduced pressure. The obtained residue, without further purification, was dissolved in acetonitrile (25 mL) and imidazole (0.98 g, 14.40 mmol) was added. The mixture was refluxed for 6 h under N_2 atmosphere, the solvent was evaporated under reduced pressure and the residue was purified by flash column chromatography (gradient elution starting from toluene, then toluene/acetone 1:1, then 1:4) to give **3e** as an oil (0.41 g, yield 38 %). ^{1}H NMR: δ 3.89 (s, 3H, OCH_3), 5.18 (s, 2H, CH_2imi), 6.95-6.97 (m, 3H, arom), 7.10 (s, 1H, imi), 7.32 (d, *J* = 7.7 Hz, 1H, arom), 7.44-7.48 (m, 1H, arom), 7.57 (s, 1H, imi), 7.61 (s, 1H, imi), 7.68 (d, *J* = 8.0 Hz, 1H, arom) 7.77-7.80 (m, 2H, arom). ^{13}C NMR: δ 50.6, 55.6, 113.8 (2C), 119.3, 128.5, 129.0, 129.7, 129.8, 130.2, 130.6, 132.6 (2C), 136.7, 137.5, 139.2, 163.6, 194.9. MS (ES) *m/z*: 293.22 (M + H)

4.1.3 (3-((1*H*-imidazol-1-yl)methyl)phenyl)(4-hydroxyphenyl)methanone (3f). A solution of **3e** (1.00 g, 3.40 mmol) in 48 % HBr (20 mL) was heated to reflux for 9 h. The reaction mixture was basified with 6N NaOH solution and washed with dichloromethane (3 x 30 mL). The aqueous phase was acidified dropwise with 6N HCl until a precipitate was formed (pH ≈ 7). The solid obtained was filtered and dried to give **3f** as white solid (500 mg, yield 52 %), mp 199-201 °C. ^{1}H NMR (methanol-

d_4): δ 5.34 (s, 2H, CH$_2$imi), 6.87 (d, J = 2.0 Hz, 2H, arom), 7.04 (s, 1H, imi), 7.19 (s, 1H, imi), 7.49-7.57 (m, 3H, arom), 7.64-7.69 (m, 3H, arom), 7.85 (s, 1H, imi). ^{13}C NMR (methanol-d_4): δ 51.2, 116.2 (2C), 120.9, 129.4, 129.5, 129.5, 129.9, 130.2, 131.9, 134.0 (2C), 138.7, 138.8, 140.2, 163.9, 197.0. MS (ES) m/z: 279.22 (M + H)

4.1.4 <u>General method for the synthesis of alkoxylated compounds, 2a-d and 3a-d.</u>

A mixture of **2f** or **3f** (1.0 eq), K$_2$CO$_3$ (1.0 eq) and the suitable alkyl bromide (1.0 eq) in acetone (20 mL) was refluxed for 18-26 h (monitored by TLC). The mixture was hot filtered and evaporated to dryness to obtain a residue that was purified by flash column chromatography.

4.1.4.1 (4-((1*H*-imidazol-1-yl)methyl)phenyl)(4-(pent-2-yn-1-yloxy)phenyl) methanone (2a).
Starting from **2f** [21] (0.20 g, 0.72 mmol) and 1-bromopent-2-yne (0.08 mL, 0.72 mmol) a crude compound was obtained that was purified by flash column chromatography (ethyl acetate/methanol 4.75:0.25) to give compound **2a** as grey solid (0.04 g, yield 17 %), mp 92-96 °C. ^{1}H NMR: δ 1.14 (t, J = 7.2 Hz, 3H, CH$_3$), 2.21-2.27 (m, 2H, CH$_2$), 4.75 (t, J = 1.8 Hz, 2H, CH$_2$), 5.22 (s, 2H, CH$_2$imi), 6.94 (s, 1H, imi), 7.04 (d, J = 8.0 Hz, 2H, arom), 7.14 (s, 1H, imi), 7.24 (d, J = 8.1 Hz 2H, arom), 7.60 (s, 1H, imi), 7.75 (d, J = 8.0 Hz, 2H, arom), 7.80 (d, J = 8.4 Hz, 2H, arom). ^{13}C NMR: δ 12.6, 13.6, 50.6, 56.8, 73.5, 90.4, 114.7 (2C), 119.4, 127.0 (2C), 130.2, 130.3, 130.5 (2C), 132.5 (2C), 137.6, 138.3, 140.2, 161.7, 194.8. MS (ES) m/z: 345.29 (M + H)

4.1.4.2 (4-((1*H*-imidazol-1-yl)methyl)phenyl)(4-(prop-2-yn-1-yloxy)phenyl) methanone (2b).
Starting from **2f** [21] (0.20 g, 0.72 mmol) and 3-bromoprop-1-yne (80 % in toluene, 0.11 mL, 0.72 mmol) a crude compound was obtained that was purified by flash column chromatography (dichloromethane/acetone 2:3) to give **2b** as white solid (0.03 g, yield 13 %), mp 140-143 °C. ^{1}H NMR: δ 2.57 (t, J = 2.4 Hz, 1H, CH), 4.78 (d, J = 2.4 Hz, 2H, OCH$_2$), 5.22 (s, 2H, CH$_2$imi), 6.94 (s, 1H, imi), 7.05 (d, J = 8.8 Hz, 2H arom), 7.14 (s, 1H, imi), 7.24 (d, J = 8.4 Hz, 2H, arom), 7.59 (s, 1H, imi), 7.75 (d, J = 8.4 Hz, 2H, arom), 7.81 (d, J = 8.8 Hz, 2H, arom). ^{13}C NMR: δ 50.6, 56.0, 76.3, 77.8, 114.6 (2C), 119.4, 127.0 (2C), 130.2, 130.5 (2C), 130.7, 132.5 (2C), 137.5, 138.2, 140.3, 161.3, 192.8. MS (ES) m/z: 317.23 (M + H)

4.1.4.3 (4-((1*H*-imidazol-1-yl)methyl)phenyl)(4-(pentyloxy)phenyl)methanone (2c).
Starting from **2f** [21] (0.20 g, 0.72 mmol) and 1-bromopentane (0.09 mL, 0.72 mmol) a crude compound was

obtained that was purified by flash column chromatography (ethyl acetate/methanol 4.75:0.25) to give **2c** as a yellow oil (0.15 g, yield 60 %). ^{1}H NMR: δ 0.93 (t, *J* = 7 Hz, 3H, CH$_3$), 1.36-147 (m, 4H, 2 x CH$_2$), 1.78-1-83 (m, 2H, CH$_2$), 4.03 (t, *J* = 6.4 Hz, 2H, CH$_2$), 5.20 (s, 2H, CH$_2$), 6.92-6.95 (m, 3H arom), 7.13 (s, 1H, imi), 7.22-7.24 (m, 2H, arom), 7.59 (s, 1H, imi), 7.72-7.79 (m, 4H, arom). ^{13}C NMR: δ 14.1, 22.5, 28.2, 28.9, 50.6, 68.4, 114.2 (2C), 119.5, 127.0 (2C), 129.6, 130.1, 130.5 (2C), 132.6 (2C), 137.6, 138.4, 140.0, 163.1, 194.8. MS (ES) *m/z*: 349.29 (M + H)

4.1.4.4 (4-((1*H*-imidazol-1-yl)methyl)phenyl)(4-propoxyphenyl)methanone (2d). Starting from **2f** [21] (0.18 g, 0.65 mmol) and 1-bromopropane (0.07 mL, 0.65 mmol) a crude compound was obtained that was purified by flash column chromatography (dichoromethane/acetone 3:2) to give **2d** as an oil (0.15 g, yield 65 %). ^{1}H NMR: δ 1.06 (t, *J* = 7.4 Hz, 3H, CH$_3$), 1.84-1.86 (m, 2H, CH$_2$), 4.00 (t, *J* = 6.4 Hz, 2H OCH$_2$), 5.21 (s, 2H, CH$_2$imi), 6.93-6.96 (m, 3H, arom), 7.12 (s, 1H, imi), 7.23 (d, J = 8.0 Hz, 2H, arom), 7.59 (s, 1H, imi), 7.73 (d, J = 8.1 Hz, 2H, arom), 7.78 (d, J = 8.8 Hz, 2H, arom). ^{13}C NMR: δ 10.6, 22.5, 50.5, 69.9, 114.2 (2C), 119.4, 127.0 (2C), 129.6, 130.1, 130.5 (2C), 132.6 (2C), 137.6, 138.4, 140.0, 163.1, 194.8. MS (ES) *m/z*: 321.29 (M + H)

4.1.4.5 (3-((1*H*-imidazol-1-yl)methyl)phenyl)(4-(pent-2-yn-1-yloxy)phenyl) methanone (3a). Starting from **3f** [23] (0.20 g, 0.72 mmol) and 1-bromopent-2-yne (0.08 mL, 0.72 mmol) a crude compound was obtained that was purified by flash column chromatography (ethyl acetate/methanol 4.75:0.25) to give **3a** as beige solid (0.06 g, yield 24 %), mp 58-60 °C. ^{1}H NMR: δ 1.14 (t, *J* = 7.2 Hz, 3H, CH$_3$), 2.24 (tq, *J* = 7.5 Hz, *J* = 2.1 Hz, 2H, CH$_2$), 4.75 (t, *J* = 2.1 Hz, 2H, OCH$_2$), 5.19 (s, 2H, CH$_2$imi), 6.93 (s, 1H, aro), 7.02-7.05 (m, 2H, aro), 7.11 (s, 1H, imi), 7.31 (d, *J* = 7.7 Hz 1H, arom), 7.44-7.48 (m, 1H, arom), 7.57 (s, 1H, imi), 7.62 (s, 1H, imi), 7.69 (d, *J* = 7.7 Hz, 1H, arom), 7.79 (m, 2H, arom). ^{13}C NMR: δ 12.6, 13.6, 50.6, 56.8, 73.5, 90.4, 114.7 (2C), 119.3, 128.5, 129.0, 129.8, 130.1, 130.2, 130.6, 132.5 (2C), 136.7, 137.5, 139.1, 161.8, 194.9. MS (ES) *m/z*: 345.33 (M + H)

4.1.4.6 (3-((1*H*-imidazol-1-yl)methyl)phenyl)(4-(prop-2-yn-1-yloxy)phenyl) methanone (3b). Starting from **3f** [23] (0.20 g, 0.72 mmol) and 80 % 3-bromoprop-1-yne (0.11 mL, 0.72 mmol) a crude compound was obtained that was purified by flash column chromatography (dichloromethane/methanol 9.5:0.5) to give **3b** as with solid (0.04 g, 18 %), mp 119-121 °C. ^{1}H NMR: δ 2.57 (t, *J* = 2.4 Hz, 1H, CH), 4.79 (d, *J* = 2.4 Hz, 2H, OCH$_2$), 5.19 (s, 2H, CH$_2$-imi), 6.93 (s, 1H, arom), 7.07-7.02 (m, 2H, arom), 7.11 (s, 1H, imi), 7.33 (d, *J* = 7.7 Hz, 1H, arom), 7.47-7.44 (m, 1H, arom), 7.58 (s, 1H, imi), 7.62 (s, 1H, imi), 7.68 (d, *J* = 7.7 Hz, 1H, arom), 7.79-7.81 (m, 2H, arom).

^{13}C NMR: δ 50.6, 56.0, 76.3, 77.8, 114.6 (2C), 119.3, 128.5, 129.0, 129.7, 130.1, 130.5, 130.7, 132.5 (2C), 136.7, 137.5, 138.9, 161.3, 194.8. MS (ES) *m/z*: 317.23 (M + H)

4.1.4.7 (3-((1*H*-imidazol-1-yl)methyl)phenyl)(4-(pentyloxy)phenyl)methanone (3c). Starting from **3f** [23] (0.20 g, 0.72 mmol) and 1-bromopentane (0.09 mL, 0.72 mmol) a crude compound was obtained that was purified by flash column chromatography (ethyl acetate/methanol 4.75:0.25) to give **3c** as an oil (0.11 g, yield 44 %). ^{1}H NMR: δ 0.95 (t, *J* = 5.6 Hz, 3H, CH$_3$), 1.39-1.48 (m, 4H, 2CH$_2$), 1.73-1.84 (m, 2H, CH$_2$), 4.04 (t, *J* = 6.6 Hz, 2H, OCH$_2$), 5.18 (s, 2H, CH$_2$imi), 6.93-6.95 (m, 3H, arom), 7.11 (s, 1H, imi), 7.31 (d, *J* = 7.6 Hz, 1H, arom), 7.44-7.46 (m, 1H, arom), 7.57 (s, 1H, imi), 7.61 (s, 1H, imi), 7.68 (d, *J* = 8.0 Hz, 1H, arom), 7.80-7.74 (m, 2H, arom). ^{13}C NMR: δ 14.1, 22.5, 28.2, 28.9, 50.7, 68.4, 114.2 (2C), 119.3, 128.5, 129.0, 129.5, 129.7, 130.1, 130.5, 132.6 (2C), 136.6, 137.5, 139.3, 163.2, 194.9. MS (ES) *m/z*: 349.29 (M + H)

4.1.4.8 (3-((1*H*-imidazol-1-yl)methyl)phenyl)(4-propoxyphenyl)methanone (3d). Starting from **3f** [23] (0.18 g, 0.65 mmol) and 1-bromopropane (0.07 mL, 0.65 mmol) a crude compound was obtained that was purified by flash column chromatography (dichloromethane/acetone 3:2) to give **3d** as an oil (0.09 g, yield 43 %). ^{1}H NMR: δ 1.06 (t, *J* = 7.6 Hz, 3H, CH$_3$), 1.82-1.88 (m, 2H, CH$_2$), 4.00 (t, *J* = 6.6 Hz, 2H, OCH$_2$), 5.18 (s, 2H, CH$_2$imi), 6.94-6.96 (m, 3H, arom), 7.11 (s, 1H, imi), 7.30 (d, *J* = 7.6 Hz, 1H, arom), 7.44-7.46 (m, 1H arom), 7.58 (s, 1H, imi), 7.61 (s, 1H, imi), 7.68 (d, *J* = 7.7 Hz, 1H, arom), 7.76-7.78 (m, 2H, arom). ^{13}C NMR: δ 10.6, 22.5, 50.6, 69.9, 114.2 (2C), 119.3, 128.5, 129.0, 129.5, 129.7, 130.2, 130.5, 132.6 (2C), 136.6, 137.5, 139.2, 163.2, 194.9. MS (ES) *m/z*: 321.29 (M + H)

4.2 Biological evaluation

4.2.1 Aromatase inhibition assay

Inhibition of aromatase was quantified by the Aromatase Inhibitor Screening Kit (BioVision Inc., San Francisco, USA), using a fluorogenic substrate that is converted by aromatase activity into a highly fluorescent metabolite. Briefly, after the reconstitution of the reagents, a standard curve was generated by diluting the fluorescence standard. Test compounds were dissolved in DMSO at a final concentration of ≤ 0,25% (v/v), after having verified that such concentration of solvent does not

significantly affect AR activity. Each test was diluted in aromatase assay buffer to obtain a range of concentrations for generating a multi-point dose-response curve. The concentrations tested were 1000, 100, 10, 1 and 0.1 nM. The reaction was prepared by adding Aromatase mix containing Recombinant Human AR (2X), Aromatase assay buffer and NADPH-generating system to test compounds, no inhibitor control, background control and positive inhibition control (1 μM of letrozole). The reaction mixture was preincubated at 37°C for 10 min to allow test compounds to interact with AR, then, reaction initiated after the addition of 30 μl of Aromatase Substrate/NADP+ mixture containing buffer, AR substrate and β-NADP+ stock (100X). Assays were conducted in 96-well microtiter plates (Corning Incorporated, Corning, ME, USA) in a final reaction volume of 100 μl/well. Sample fluorescence was measured using a Tecan Spark Plate Reader (Tecan Trading AG, Switzerland) at dual wavelengths of 488/527 nm for 60 min. Results were expressed as relative fluorescence units (RFUs). Experiments were performed in triplicate and the average values were used to construct the dose-response curves.

4.2.2 Cell lines and culture conditions

MCF-7, MDA-MB-231 and MCF-10A cell lines were obtained from American Type Culture Collection (ATCC). MCF-7 cells were grown in RPMI-1640 medium without phenol red, containing 10% charcoal dextran-stripped FBS (Thermo Fisher Scientific), MDA-MB-231 in DMEM-F12 + 5% FBS while MCF-10A in MEBM medium (Lonza) supplemented with 10 ng/mL EGF, 10 ng/mL FGF, 600 U/L eparin and B-27 Supplement (Thermo Fisher Scientific).

Cells were tested for the absence of Mycoplasma fortnightly and maintained in logarithmic growth phase as a monolayer in a humidified 5% CO_2 atmosphere at 37 °C.

4.2.3 Cell proliferation assays

Cells were seeded in duplicate in 12-well plates (50.000 cells/well, 50.000 cells/well and 90.000 cells/well for MCF-7 cells, MDA-MB-231 cells and MCF-10A cells, respectively) and 24 h later exposed for 72 h to increasing concentrations of all compounds (50-0.1 μM) in the absence or in the presence of 1 μM of testosterone. The medium and the treatments were replaced after 48 h. Culture medium was then removed and adherent cells were harvested using trypsin and counted with a cell counter (Beckman Coulter,S.p.A., Milan, Italy). All experiments were performed three times.

4.3 Computational Methods

4.3.1 Docking calculations

Initial docking calculations of compounds **3a-3h** and **2b** were performed on the aromatase (AR) enzyme with Glide [32] using the single precision (SP) protocol on the AR crystal structure (PDB code 3S79). A van der Waals (vdW) radius scaling factor of 0.80 Å was used for the protein and ligands atoms possessing a partial charge less than 0.15 in order to mimic protein flexibility. A metal constraint was taken into account to obtain binding poses in which the nitrogen atom was at coordination distance from the heme iron atom, as performed in a previous study [19]. Finally, the possibility of tautomerism of the imidazole ring of the inhibitors has been explicitly considered.

Docking calculations on AR allosteric binding pockets (Sites 1 and 2) were performed for compounds **3a-3h** and **2b** following the protocol detailed in ref [17] to monitor the possibility that these molecules could also exert a non-active site directed inhibition mechanism.

4.3.2 Classical Molecular Dynamics (MD) simulations

Next, we performed classical MD simulations of AR in complex with the selected docked inhibitors (**3a-3h** and **2b**) using a protocol adopted in previous studies [31]. The most likely protonation states under physiological conditions were obtained using the webserver H++ [33], while Asp309 was considered its neutral form, as reported in literature [34, 35]. For the MD simulations we have employed Parm99SB AMBER force field for the protein [36], the Shahrokh *et al.* parameters for the heme moiety and Cys437 [37] and the general Amber FF (GAFF) for the inhibitors [38]. Electrostatic Potential (ESP) charges were obtained by performing geometry optimization of the inhibitors at Hartree-Fock level of theory with a 6-31G* basis set using the Gaussian 09 software and were then converted in Restricted Electrostatic potential (RESP) charges using Antechamber module of Ambertools 18 [39]. The obtained models were solvated using TIP3P waters leading to a total of 63,230 atoms. The topologies were finally built with Ambertools 18, and later converted to a GROMACS format with the software acpype.

GROMACS 2020.2 [40] was used to perform classical MD simulations using an integration time step of 2 fs. All covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm. MD simulations were done in the NPT ensemble, at a temperature of 300 K, using a velocity-rescaling thermostat [41]. An initial energy minimization step was done with the steepest descend algorithm. An equilibration of the models was performed for 20 ns with the protein and ligand atoms

harmonically restrained using a force constant of 1000 kJ mol^{-1} nm^{-2}. Next, the constraints were gradually released (1000, 500, 250 and 50 kJ mol^{-1} nm^{-2}) in four subsequent 20 ns runs, while leaving the side chains free of constraint. Thus, to relax the AR structure to the presence of the ligand, while preserving the stability of the coordination bond between the heme iron and the imidazole ring of the selected inhibitors, each system underwent a total of 100 ns-long MD simulations.

4.3.3 QM/MM Molecular Dynamics Simulations

Because of the known difficulties in accurately describing coordination bonds using classical FF [42, 43], QM/MM Born Oppenheimer MD simulations were performed using CP2K 7.1 program [44]. The QM region, including the heme group, Cys437 and the imidazole ring of the inhibitors (54 atoms), was simulated in a cubic box having sides of 20 Å. Density Functional Theory (DFT) with the BLYP [45, 46] exchanges correlation functional and dual Gaussian-type/Plane Waves basis set (GPW) was used [47]. In particular, we employed a double ζ (MOLOPT) basis set, an auxiliary PW basis set with a density cutoff of 400 Ry and Goedecker-Teter-Hutter (GTH) pseudopotentials [48]. This protocol has been previously used in many QM/MM MD simulations of biomolecules, including the aromatase enzyme [31, 49, 50]. The dangling bonds between the quantum and classical layers were saturated with capping hydrogens atoms. An integration step of 0.5 fs was used in all the QM/MM MD simulations using a NVT ensemble. All the investigated systems were optimized and equilibrated at 300 K without constraints for 2 ps, using a Nosé-Hoover thermostat [51], followed by 8 ps of simulation once the stable formation of the coordination bond has occurred. A doublet spin state was used for the heme iron, consistently with previous studies [19]. The MM region was described using the same force field employed in the classical MD simulations.

4.3.4 Analysis

Cluster analysis of the MD trajectories was carried out using g_cluster tool, using the Daura et. al algorithm [52] of the GROMACS 2018.2 package. The Amber 18 tool MM_PBSA.py [53] was used to perform Molecular Mechanics Generalized Born Surface Area (MM-GBSA) free energy calculation, taking 100 frames from the last 5 ps of the QM/MM MD trajectories.

In particular, a generalized born solvation model was employed (igb=8) [54], using a salt concentration of 0.1 M. The conformational entropic component of the free energy was not considered, since it was previously suggested that this term does not improve the quality of the results

using the MM-G(P)BSA method [55, 56]. Visualization of the MD trajectories and images were done using the VMD program [57].

References

[1] American Cancer Society. Cancer Facts & Figures 2022. Atlanta, 2022.

[2] Feng, Y.; Spezia, M.; Huang, S.; Yuan, C.; Zeng, Z.; Zhang, L.; Ji, X.; Liu, W.; Huang, B.; Luo, W.; Liua, B.; Lei, Y.; Du, S.; Vuppalapati, A.; Luu, H.H.; Haydon, R. C.; He, T.; Rena, G. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes Dis* **2018**, *5* (2), 77-106. DOI: 10.1016/j.gendis.2018.05.001.

[3] Lewis-Wambi, J. S.; Jordan, V. C. Treatment of Postmenopausal Breast Cancer with Selective Estrogen Receptor Modulators (SERMs). *Breast Dis* **2005**, *24*, 93-105. DOI: 10.3233/bd-2006-24108.

[4] Patel, H. K.; Bihani, T. Selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders (SERDs) in cancer treatment. *Pharmacol Ther* **2018**, *186*, 1-24. DOI: 10.1016/j.pharmthera.2017.12.012.

[5] Ghosh, D.; Lo, J.; Egbuta, C. Recent Progress in the Discovery of Next Generation Inhibitors of Aromatase from the Structure-Function Perspective. *J Med Chem* **2016**, *59* (11), 5131-5148. DOI: 10.1021/acs.jmedchem.5b01281.

[6] Spinello, A.; Ritacco, I.; Magistrato, A. The Catalytic Mechanism of Steroidogenic Cytochromes P450 from All-Atom Simulations: Entwinement with Membrane Environment, Redox Partners, and Post-Transcriptional Regulation. *Catalysts* **2019**, *9* (1), Review. DOI: 10.3390/catal9010081.

[7] Simpson, E. R.; Clyne, C.; Rubin, G.; Boon, W. C.; Robertson, K.; Britt, K.; Speed, C.; Jones, M. Aromatase--a brief overview. *Annu Rev Physiol* **2002**, *64*, 93-127. DOI: 10.1146/annurev.physiol.64.081601.142703.

[8] Dutta, U.; Pant, K. Aromatase inhibitors: past, present and future in breast cancer therapy. *Med Oncol* **2008**, *25* (2), 113-124. DOI: 10.1007/s12032-007-9019-x.

[9] Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature* **2009**, *457* (7226), 219-223. DOI: 10.1038/nature07614.

[10] Fanning, S. W.; Mayne, C. G.; Dharmarajan, V.; Carlson, K. E.; Martin, T. A.; Novick, S. J.; Toy, W.; Green, B.; Panchamukhi, S.; Katzenellenbogen, B. S.; Tajkhorshid, E.; Griffin, P. R.; Shen, Y.; Chandarlapaty, S.; Katzenellenbogen, J. A.; Greene, G. L. Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. *Elife* **2016**, *5*. DOI: 10.7554/eLife.12792.

[11] Pavlin, M.; Spinello, A.; Pennati, M.; Zaffaroni, N.; Gobbi, S.; Bisi, A.; Colombo, G.; Magistrato, A. A Computational Assay of Estrogen Receptor α Antagonists Reveals the Key Common Structural Traits of Drugs Effectively Fighting Refractory Breast Cancers. *Sci Rep* **2018**, *8* (1), 649. DOI: 10.1038/s41598-017-17364-4.

[12] Spinello, A.; Ritacco, I.; Magistrato, A. Recent advances in computational design of potent aromatase inhibitors: open-eye on endocrine-resistant breast cancers. *Expert Opin Drug Discov* **2019**, *14* (10), 1065-1076. DOI: 10.1080/17460441.2019.1646245.

[13] Caciolla, J.; Bisi, A.; Belluti, F.; Rampa, A.; Gobbi, S. Reconsidering Aromatase for Breast Cancer Treatment: New Roles for an Old Target. *Molecules* **2020**, *25* (22). DOI: 10.3390/molecules25225351.

[14] Lu, W. J.; Desta, Z.; Flockhart, D. A. Tamoxifen metabolites as active inhibitors of aromatase in the treatment of breast cancer. *Breast Cancer Res Treat* **2012**, *131* (2), 473-481. DOI: 10.1007/s10549-011-1428-z.

[15] Sgrignani, J.; Bon, M.; Colombo, G.; Magistrato, A. Computational approaches elucidate the allosteric mechanism of human aromatase inhibition: a novel possible route to Small-molecule regulation of CYP450s activities? *J Chem Inf Model* **2014**, *54* (10), 2856-2868. DOI: 10.1021/ci500425y.

[16] Magistrato, A.; Sgrignani, J.; Krause, R.; Cavalli, A. Single or Multiple Access Channels to the CYP450s Active Site? An Answer from Free Energy Simulations of the Human Aromatase Enzyme. *J Phys Chem Lett* **2017**, *8* (9), 2036-2042. DOI: 10.1021/acs.jpclett.7b00697.

[17] Spinello, A.; Martini, S.; Berti, F.; Pennati, M.; Pavlin, M.; Sgrignani, J.; Grazioso, G.; Colombo, G.; Zaffaroni, N.; Magistrato, A. Rational design of allosteric modulators of the aromatase enzyme: An unprecedented therapeutic strategy to fight breast cancer. *Eur J Med Chem* **2019**, *168*, 253-262. DOI: 10.1016/j.ejmech.2019.02.045.

[18] Ghosh, D.; Lo, J.; Morton, D.; Valette, D.; Xi, J.; Griswold, J.; Hubbell, S.; Egbuta, C.; Jiang, W.; An, J.; Davies, H. M. L al. Novel aromatase inhibitors by structure-guided design. *J Med Chem* **2012**, *55* (19), 8464-8476. DOI: 10.1021/jm300930n.

[19] Caciolla, J.; Spinello, A.; Martini, S.; Bisi, A.; Zaffaroni, N.; Gobbi, S.; Magistrato, A. Targeting Orthosteric and Allosteric Pockets of Aromatase via Dual-Mode Novel Azole Inhibitors. *ACS Med Chem Lett* **2020**, *11* (5), 732-739. DOI: 10.1021/acsmedchemlett.9b00591.

[20] Gobbi, S.; Zimmer, C.; Belluti, F.; Rampa, A.; Hartmann, R. W.; Recanatini, M.; Bisi, A. Novel highly potent and selective nonsteroidal aromatase inhibitors: synthesis, biological evaluation and structure-activity relationships investigation. *J Med Chem* **2010**, *53* (14), 5347-5351. DOI: 10.1021/jm100319h.

[21] Gobbi, S.; Hu, Q.; Foschi, G.; Catanzaro, E.; Belluti, F.; Rampa, A.; Fimognari, C.; Hartmann, R. W.; Bisi, A. Benzophenones as xanthone-open model CYP11B1 inhibitors potentially useful for promoting wound healing. *Bioorg Chem* **2019**, *86*, 401-409. DOI: 10.1016/j.bioorg.2019.01.066.

[22] Gobbi, S.; Cavalli, A.; Negri, M.; Schewe, K. E.; Belluti, F.; Piazzi, L.; Hartmann, R. W.; Recanatini, M.; Bisi, A. Imidazolylmethylbenzophenones as highly potent aromatase inhibitors. *J Med Chem* **2007**, *50* (15), 3420-3422.

[23] Brynmor, J. The Halogenation of Phenolic Ethers and Anilides. Part VIII. Alkoxy- and Dialhxy-benmphenones and Dialkmydiphenylsulphones. *J. Chem. Soc.* **1936**, 1231-1234.

[24] Egbuta, C.; Lo, J.; Ghosh, D. Mechanism of inhibition of estrogen biosynthesis by azole fungicides. *Endocrinology* **2014**, *155* (12), 4622-4628. DOI: 10.1210/en.2014-1561.

[25] Zhou, D.; Wang, J.; Chen, E.; Murai, J.; Siiteri, P. K.; Chen, S. Aromatase gene is amplified in MCF-7 human breast cancer cells. *J Steroid Biochem Mol Biol* **1993**, *46* (2), 147-153. DOI: 10.1016/0960-0760(93)90289-9.

[26] Santner, S. J.; Chen, S.; Zhou, D.; Korsunsky, Z.; Martel, J.; Santen, R. J. Effect of androstenedione on growth of untransfected and aromatase-transfected MCF-7 cells in culture. *J Steroid Biochem Mol Biol* **1993**, *44* (4-6), 611-616. DOI: 10.1016/0960-0760(93)90267-z.

[27] Spinello, A.; Borišek, J.; Pavlin, M.; Janoš, P.; Magistrato, A. Computing Metal-Binding Proteins for Therapeutic Benefit. *ChemMedChem* **2021**, *16* (13), 2034-2049. DOI: 10.1002/cmdc.202100109.

[28] Palermo, G.; Spinello, A.; Saha, A.; Magistrato, A. Frontiers of metal-coordinating drug design. *Expert Opin Drug Discov* **2021**, *16* (5), 497-511. DOI: 10.1080/17460441.2021.1851188.

[29] Caciolla, J.; Martini, S.; Spinello, A.; Pavlin, M.; Turrini, E.; Simonelli, F.; Belluti, F.; Rampa, A.; Bisi, A.; Fimognari, C.; Zaffaroni, N.; Gobbi, S.; Magistrato, A. Balanced dual acting compounds targeting aromatase and estrogen receptor α as an emerging therapeutic opportunity to counteract estrogen responsive breast cancer. *Eur J Med Chem* **2021**, *224*, 113733. DOI: 10.1016/j.ejmech.2021.113733.

[30] Massova, I.; Kollman, P. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in Drug Discovery and Design* **2000**, *18*, 113-135, Review. DOI: 10.1023/A:1008763014207.

[31] Spinello, A.; Pavlin, M.; Casalino, L.; Magistrato, A. A Dehydrogenase Dual Hydrogen Abstraction Mechanism Promotes Estrogen Biosynthesis: Can We Expand the Functional Annotation of the Aromatase Enzyme? *Chemistry* **2018**, *24* (42), 10840-10849. DOI: 10.1002/chem.201802025.

[32] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, E. D.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004**, *47* (7), 1739-1749. DOI: 10.1021/jm0306430.

[33] Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **2012**, *40* (Web Server issue), W537-541. DOI: 10.1093/nar/gks375.

[34] Di Nardo, G.; Breitner, M.; Bandino, A.; Ghosh, D.; Jennings, G. K.; Hackett, J. C.; Gilardi, G. Evidence for an elevated aspartate pK(a) in the active site of human aromatase. *J Biol Chem* **2015**, *290* (2), 1186-1196. DOI: 10.1074/jbc.M114.595108.

[35] Sgrignani, J.; Magistrato, A. Influence of the membrane lipophilic environment on the structure and on the substrate access/egress routes of the human aromatase enzyme. A computational study. *J Chem Inf Model* **2012**, *52* (6), 1595-1606. DOI: 10.1021/ci300151h.

[36] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950-1958. DOI: 10.1002/prot.22711.

[37] Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E. Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J Comput Chem* **2012**, *33* (2), 119-133. DOI: 10.1002/jcc.21922.

[38] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **2004**, *25* (9), 1157-1174. DOI: 10.1002/jcc.20035.

[39] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **2006**, *25* (2), 247-260. DOI: 10.1016/j.jmgm.2005.12.005.

[40] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, *26* (16), 1701-1718. DOI: 10.1002/jcc.20291.

[41] Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **2007**, *126* (1), Article. DOI: 10.1063/1.2408420.

[42] Sgrignani, J.; Casalino, L.; Doro, F.; Spinello, A.; Magistrato, A. Can multiscale simulations unravel the function of metallo-enzymes to improve knowledge-based drug discovery? *Future Med Chem* **2019**, *11* (7), 771-791. DOI: 10.4155/fmc-2018-0495.

[43] Spinello, A.; Magistrato, A. An omics perspective to the molecular mechanisms of anticancer metallo-drugs in the computational microscope era. *Expert Opin Drug Discov* **2017**, *12* (8), 813-825. DOI: 10.1080/17460441.2017.1340272.

[44] VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Computer Physics Communications* **2005**, *167* (2), 103-128, Article. DOI: 10.1016/j.cpc.2004.12.014.

[45] Becke, A. Density-Functional Exchange-Energy Approximation With Correct Asymptotic-Behavior. *Physical Review a* **1988**, *38* (6), 3098-3100, Note. DOI: 10.1103/PhysRevA.38.3098.

[46] Lee, C.; Yang, W.; Parr, R. Development Of The Colle-Salvetti Correlation-Energy Formula Into A Functional Of The Electron-Density. *Physical Review B* **1988**, *37* (2), 785-789, Article. DOI: 10.1103/PhysRevB.37.785.

[47] VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *Journal of Chemical Physics* **2007**, *127* (11), Article. DOI: 10.1063/1.2770708.

[48] Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Physical Review B* **1996**, *54* (3), 1703-1710, Article. DOI: 10.1103/PhysRevB.54.1703.