Supporting Information

Computational Approaches Elucidate the Allosteric Mechanism of Human Aromatase Inhibition: A Novel Possible Route to Small-Molecule Regulation of CYP450s Activities?

Jacopo Sgrignani[1⊥], Marta Bon[2⊥], Giorgio Colombo[1*], Alessandra Magistrato[2*]

1. Istituto di Chimica del Riconoscimento Molecolare, CNR , Via Mario Bianco 9, 20131 Milano, Italy.
2. CNR-IOM-Democritos National Simulation Center c/o SISSA, via Bonomea 265, 34165 Trieste, Italy.

⊥ J.S and M.B. equally contributed to this work.

Corresponding Authors:

Alessandra Magistrato
CNR-IOM- Democritos National Simulation Center
Via Bonomea 265
34165 Trieste, Italy.
Email: alessandra.magistrato@sissa.it

Giorgio Colombo
Istituto di Chimica del Riconoscimento Molecolare, CNR .
Via Mario Bianco 9,
20131 Milano, Italy.
Email: g.colombo@icrm.cnr.it

Details about SiteMap Analysis

After SiteMap analysis eight parameters are reported for every site: (1) SiteScore, (2) size, (3) exposure score, (4) enclosure score, (5) contact, (6) hydrophobic/hydrophilic character, and (7) donor/acceptor character and (8) drug score (Dscore).

In 2009 Halgren reported[1] about an extensive analysis of the SiteMap performances in predicting ligand binding site in protein considering a set of 538 complexes which structure was experimentally solved.

This evaluation has permitted to better relate the SiteMap scores with the properties of the predicted binding site. In particular a SiteScore near 1 is usually assigned to sites able to bind their molecular counterpart with an affinity in the micro-molar range, while a Dscore > 1 could indicate the propensity of a binding site to be targeted with drug like molecules. This analysis also highlighted that an exposure score below 0.52, an enclosure score higher than 0.76 and contact (hydrophobic/hydrophilic) scores near 1 are ideal for identifying promising binding site candidates.

Table S1. Score assigned to the five site identified by the SiteMap analysis.

Title	Site_1	Site_2	Site_3	Site_4	Site_5
SiteScore	1.00	0.98	0.94	0.86	0.77
Size	128	119	81	73	41
Dscore	0.95	1.02	0.96	0.77	0.47
Volume	255	284	144	167	66
Exposure	0.56	0.65	0.48	0.64	0.43
Enclosure	0.70	0.62	0.69	0.65	0.75
Contact	0.92	0.79	1.02	0.82	1.14
Phobic	0.17	0.61	1.34	0.26	0.00
Philic	1.26	0.88	0.92	1.28	1.79
Balance	0.14	0.69	1.43	0.20	0.00
Don/acc	0.67	0.54	0.79	0.44	1.28

Table S2. Residues forming the three putative sites considered in the work.

	Residues forming the site
Site_1	Arg192; Val 214; Val 215; Ile 217; Gln 218; Gly 219; Phe 221; Asp 222; Pro 308; Ash 309; Ser 312; Val 313; Ile 474; His 480; Pro 481; Asp 482; Glu 483; Thr 484
Site_2	Ile 350; Glu 351; Lys 354; Glu 357; Tyr 361;Asn 421; Val 422; Tyr 424 Phe 427; Gln 428; Pro 429; Phe 430; Gly 431; Phe 432; Lys 440; Tyr 441 Met 444; Lys 448
Site_3	Lys 150; Ala 151; Ser 153; Pro 155; Gly 156; Arg 159; Leu 202; Phe 203 Glu 273; Glu 274; Met 276; Asp 277; Phe 278; Glu 281

Table S3. Interaction energies (ΔE_{int}) (kcal/mol) between HA and endoxifen (both the stereoisomers) calculated after 18 ns of productive run. Standard deviations are reported. Interaction energy for Z-endoxifen in site_2 was not reported because this complex dissociated after 8ns.

	ΔE_{int} HA/E-endoxifen	ΔE_{int} HA/Z-endoxifen
Site_1	-55 (7)	-69 (5)
Site_2	-24 (5)	
Site_3	-67 (14)	-69 (12)

Table S4. Interaction energies (ΔE_{int}) (kcal/mol) between the drugs and HA calculated over the production run of 100 ns. Standard deviations are reported in parenthesis. Interaction energies of Z-NDT inside site_1 and Z-tamoxifen in site_3 are not reported because these inhibitors dissociated from HA after few ns. a

	Coulomb (kcal/mol)	Lennard-Jones (kcal/mol)	ΔE_{int} (kcal/mol)
Site_1			
E-endoxifen	-31 (5)	-26 (2)	-57 (5)
Z-endoxifen	-29 (12)	-31 (5)	-60 (14)
E-NDT	-22 (5)	- 33 (2)	-55 (5)
E-tamoxifen	-19 (5)	-33 (5)	-52 (5)
Z-tamoxifen	-26 (2)	-38 (5)	-64 (5)
Site_3			
E-endoxifen	-26 (12)	-33 (2)	-59 (12)
Z-endoxifen	-33 (12)	-33 (5)	-66 (12)
E-NDT	-2 (2)	-43 (2)	-45 (5)
Z-NDT	-7 (10)	-24 (5)	-31 (12)
E-tamoxifen	-5 (2)	-43 (2)	-48 (5)

Table S5. H-bonds between tamoxifen metabolites and HA over MD trajectories. Only H-bond with persistency larger than 10% are reported. For E-NDT, Z-NDT and E-tamoxifen in site_3 no H-bonds were detected.

E-endoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Glu483@ Oε1	EndoE@ H18	EndoE@ N1	75,15	2,887 (0,18)	34,29 (15,77)
Glu483@ O□2	EndoE@ H18	EndoE@ N1	66,08	2,894 (0,19)	30,04 (14,26)
Gln218@ Oε1	EndoE@ H17	EndoE@ N1	12,14	2,843 (0,14)	27,86 (13,43)
Z-endoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
His480@Nδ	EndoZ@ H18	EndoE@ N1	23,55	3,030 (0,15)	36,3 (13,29)
Lys243@ O	EndoZ@ H13	EndoE@ O1	22,03	2,747 (0,14)	23,46 (11,65)
Asp482@ Oδ1	EndoZ@ H13	EndoE@ O1	21,03	2,695 (0,21)	20,06 (13,16)
Asp482@ Oδ2	EndoZ@ H13	EndoE@ O1	20,41	2,702 (0,23)	20,27 (13,35)
E-NDT site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Ser478@Oγ	NdesE@H17	NdesE@N1	41,21	2,943 (0,16)	26,49 (13,60)
Leu479@O	NdesE@H17	NdesE@N1	17,15	2,851 (0,13)	30,39 (13,14)
Ser478@Oγ	NdesE@H13	NdesE@N1	15,66	2,960 (0,17)	29,67 (15,05)
Z-tamoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
His480@Nδ	TamE@H16	TamE@N1	13,19	3,115 (0,16)	24,49 (11,57)
E-tamoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Ser478@Oγ	TamZ@H16	TamZ@N1	63,25	2,934 (0,14)	34,01 (11,92)
Ser478@O	TamZ@H16	TamZ@N1	22,32	3,048 (0,19)	23,71 (10,31)
E-endoxifen site_3					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Glu281@ Oε2	EndoE@ H13	EndoE@O1	68,89	2,662 (0,13)	15,83 (8,42)
Glu281@ Oε1	EndoE@ H13	EndoE@O1	25,05	2,688 (0,17)	16,48 (9,83)
Z-endoxifen site_3					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Glu281@Oε1	EndoZ@H13	EndoZ@O1	50,01	2,629 (0,11)	14,49 (7,90)
Glu281@Oε2	EndoZ@H13	EndoZ@O1	49,89	2,629 (0,11)	14,85 (8,17)
Glu273@O	EndoZ@H18	EndoZ@N1	27,38	2,796 (0,11)	20,71 (11,22)
Glu273@O	EndoZ@H17	EndoZ@N1	25,80	2,822 (0,12)	22,89 (11,56)
E-NDT site_3					
No H-bonds detected					
Z-NDT site_3					
No H-bonds detected					
E-tamoxifen site_3					
No H-bonds detected					

Table S6. H-bonds between HA and ASD during MD simulations with tamoxifen metabolites bound both in site_1 and site_3.

E-endoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,47	2,976 (0,15)	19,87 (9,65)
Asd@O1	Asp309@HD2	Asp309@Od2	6,11	2,798 (0,17)	26,07 (11,45)
Z-endoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	98,78	2,987 (0,16)	21,29 (10,15)
Asd@O1	Asp309@HD2	Asp309@Od2	10,14	2,974 (0,26)	40,12 (14,37)
E-NDT site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,76	2,923 (0,13)	20,14 (9,43)
E-tanoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,92	2,926 (0,13)	19,31 (9,37)
E-tamoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,76	2,923 (0,13)	20,14 (9,43)
Z-tamoxifen site_1					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,64	2,934 (0,13)	22,00 (11,62)
Asd@O1	Thr310@HG1	Thr310@Og1	23,88	3,068 (0,24)	32,78 (15,79)
Asd@O1	Asp309@HD2	Asp309@Og2	17,06	2,803 (0,17)	27,27 (11,41)
E-endoxifen site_3					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,84	2,945 (0,14)	19,32 (9,52)
Asd@O1	Asp309@HD2	Asp309@OD2	7,26	2,891 (0,22)	27,73 (14,21)
Z-endoxifen site_3					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,60	2,971 (0,14)	19,94 (10,07)
Asd@O1	Asp309@HD2	Asp309@Od2	47,58	2,824 (0,18)	20,10 (12,67)
E-NDT site_3					
	Acceptor H	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	98,82	3,009 (0,16)	20,77 (9,48)
Z-NDT site_3					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,70	2,947 (0,14)	20,72 (9,53)
Z-tamoxifen site_3					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	99,52	2,966 (0,14)	20,74 (10,16)
Asd@O1	Asp309@HD2	Asp309@Od2	28,53	2,806 (0,17)	16,83 (10,84)
Free HA					
Donor	*Acceptor H*	*Acceptor*	*%occupied*	*distance*	*angle*
Asd@O2	Met374@H	Met374@N	81,47	3,152 (0,17)	17,38 (9,59)

Table S7. Summary of the residues, which targeted by mutagenesis studies, resulted to have an impact on HA function. The biological effect induced by their mutation i salso mentioned.

Residue/residues	
Pro308, Arg115, Ile133, Phe134, Phe221, Trp224, Ala306, Thr310, Val370, Val373, Met374, Leu477	They make van der Waals contacts with androstenedione.
Arg435, Cys437, Phe430, Gly439	Interaction with the heme group.
Asp309	Makes a hydrogen bond with the 3-keto moiety of the bond Asd - O1
Lys99, Lys108, 389/390, Lys420, Arg425 Lys108 Tyr361	Positively charged amino acids which interact with negatively charged residues on the CPR. Interacts with Asn175, Thr177 residues of CPR. Key role in increasing aromatase activity in cancer cells.
21-42, 49-71 67-68, 69,80	Interactions with the membrane. Embedded in the membrane. They are supposed to play an important role in the access of the substrate
229-231 (F-G loop) Arg192, Ile217, Gln218, Phe221, Asp222, Trp224, Ala225, Thr310, Ser312, Val313, Thr320, Val369, Ile474, Ser478, Leu479, His480, Pro481, Ala 212, Ile217, Tyr220, Tyr221, Ile225, Val257, Arg264, Asp309, Tyr310, Glu483, Thr484.	Undergoes an open/close motion that allows the stereoids to enter into or leave from the active site through access channel. Putative channels for the entrance/egress of Asd to/from the catalytic site.
Leu190, Leu191, Arg192, Arg193, Met195, Thr198, Met446, Ile450, Glu483	Putative channels for the entrance/egress of O2 to/from the catalytic site.
Val178,Thr179, Asn180, Arg180, Glu183, Tyr184, Val185, Asp186, Lys440 (D-E loop).	Important role in the self-association of aromatase monomers, which leads to an oligomeric assembly.

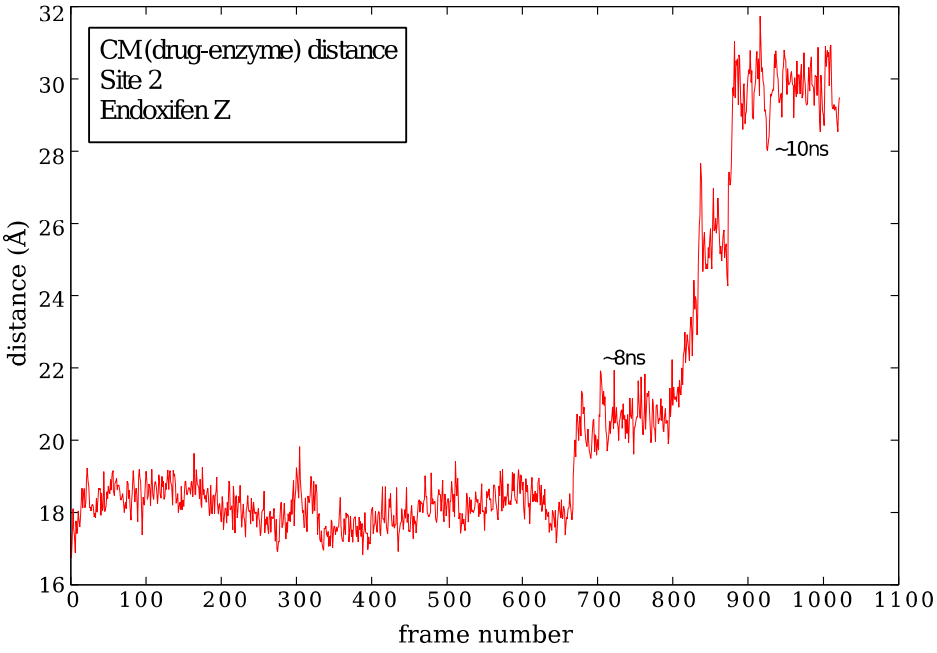

Figure S1. Site 2: Distance (Å) between the center of masses of the Z-endoxifen and of HA vs simulation time.

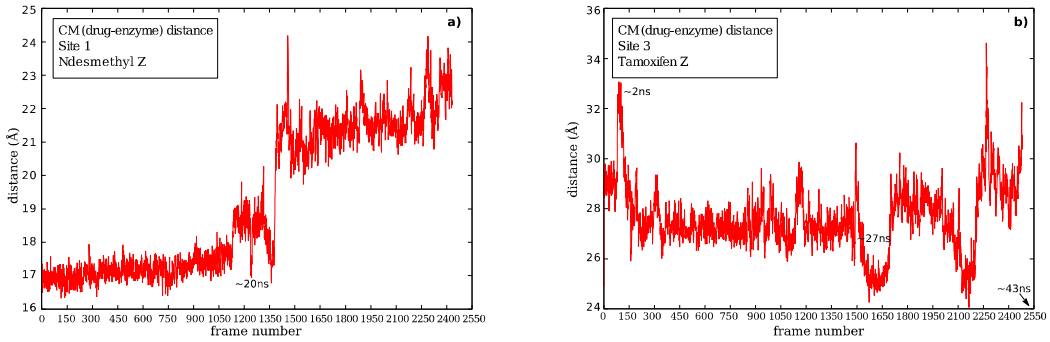

Figure S2. a) Site 1: Distance (Å) between the center of masses of the Z-N-desmethyl-tamoxifen and of HA. b) Site 3: Distance (Å) between the center of masses of the Z-tamoxifen and of HA.

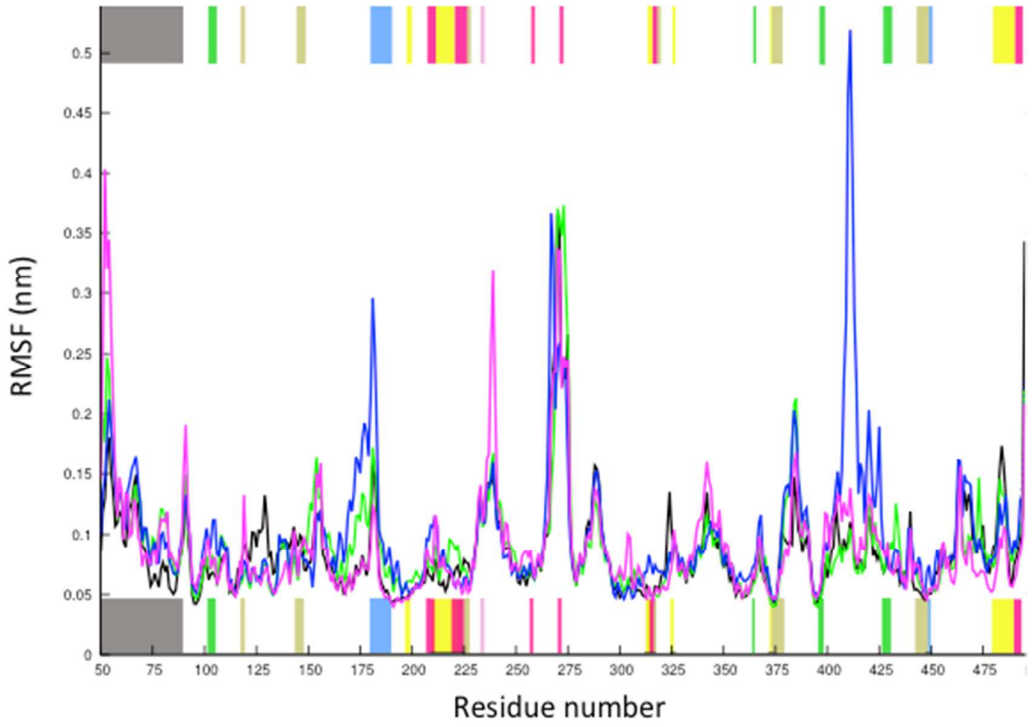

Figure S3. RMSF per residue for HA in complex with with E-NDT (green) and E-tamoxifen (blue), Z-tamoxifen (Purple) bound in site_1 and for the aromatase without inhibitors (black). With the different colors at the top and at the bottom of the picture we indicate the zones to which key-residues (listed in Table S5) belong: i) tan for the active site residues; ii) green for the residues, which interact with CPR-reductase; iii) pink for the F-G loop; iv) yellow for the crystallographic channel of access for ASD; v) purple for the channel of access for ASD in presence of the membrane found by some of us[2]; vi) light blue for the D-E loop.

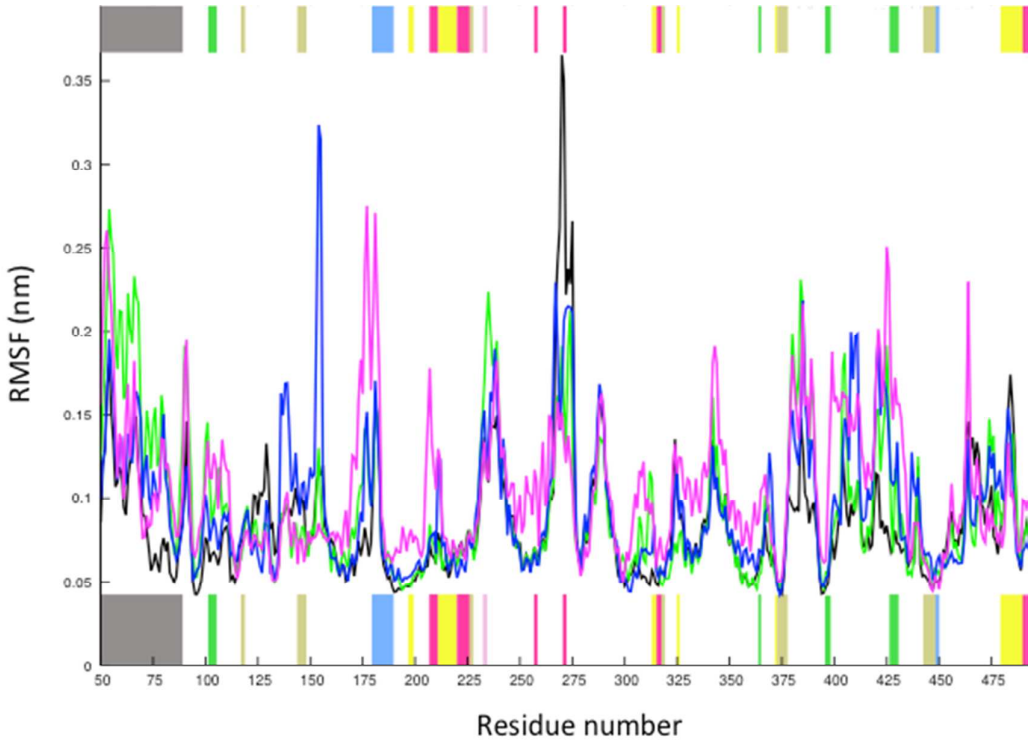

Figure S4. RMSF per residue for HA in complex with with E-NDT (green) and Z-NDT (blue), E-tamoxifen (Purple) bound in site_3 and for the aromatase without inhibitors (black). With the different colors at the top and at the bottom of the picture we indicate the zones to which key-residues (listed in Table S5) belong: i) tan for the active site residues; ii) green for the residues, which interact with CPR-reductase; iii) pink for the F-G loop; iv) yellow for the crystallographic channel of access for ASD; v) purple for the channel of access for ASD in presence of the membrane found by some of us[2]; vi) light blue for the D-E loop.

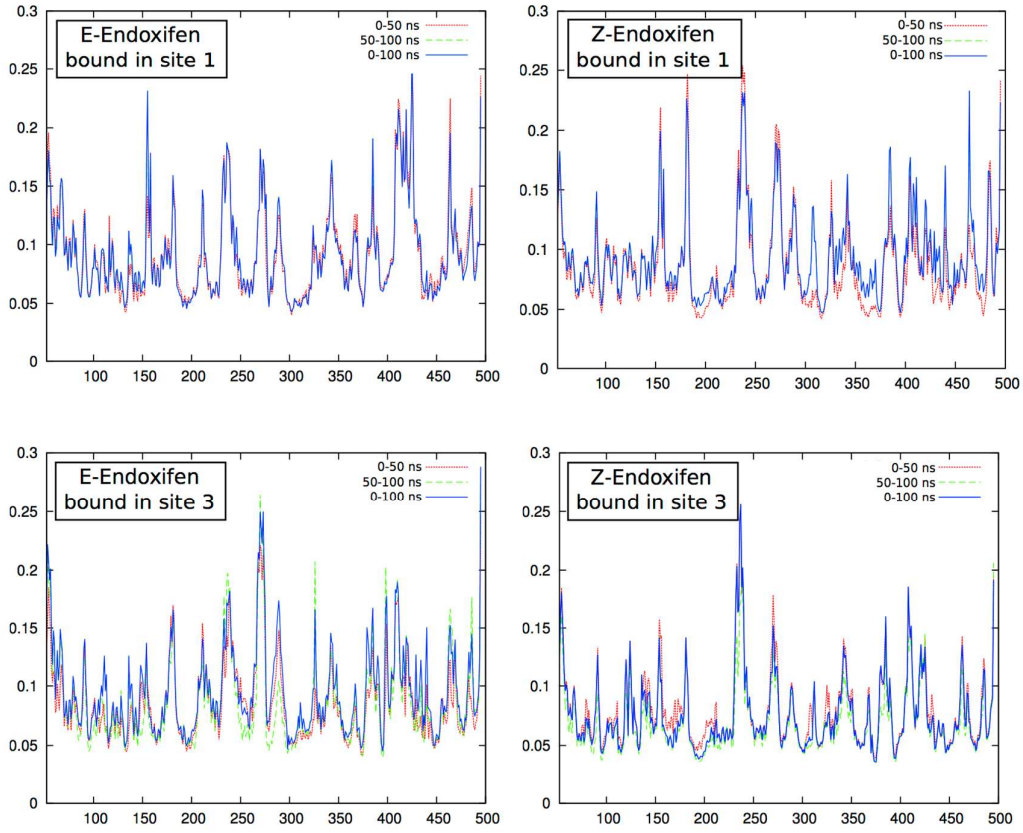

Figure S5. RMSF per residue for HA/endoxifen adducts calculated considering different portions of MD trajectory.

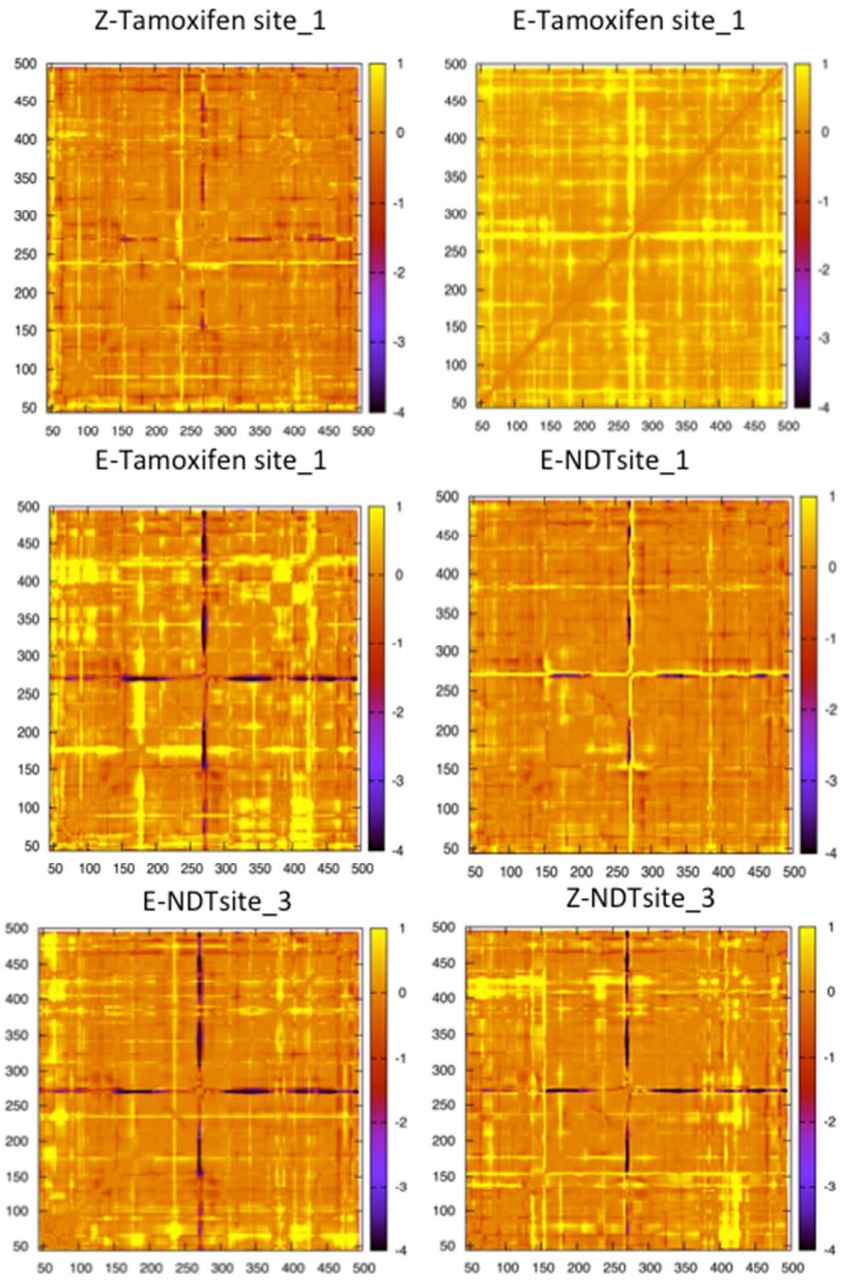

Figure S6. Difference between RMS values of residue-residue distance fluctuation matrixes for system not reported in the main text. The difference is calculated subtracting from the RMS value of each drug/HA adduct the RMSF value corresponding to HA without inhibitor. Fluctuation scale is expressed in angstrom.

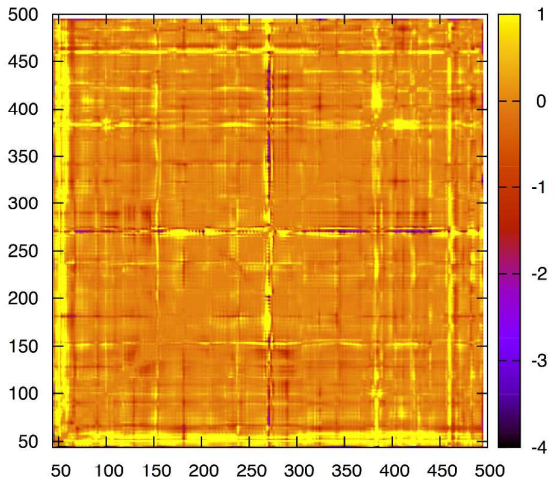

Figure S7. Difference between RMS value of residue-residue distance fluctuation matrixes for the HA simulation on the membrane surface. The difference is calculated subtracting from RMS of HA in membrane that of the HA in water. Fluctuation scale is expressed in angstrom.

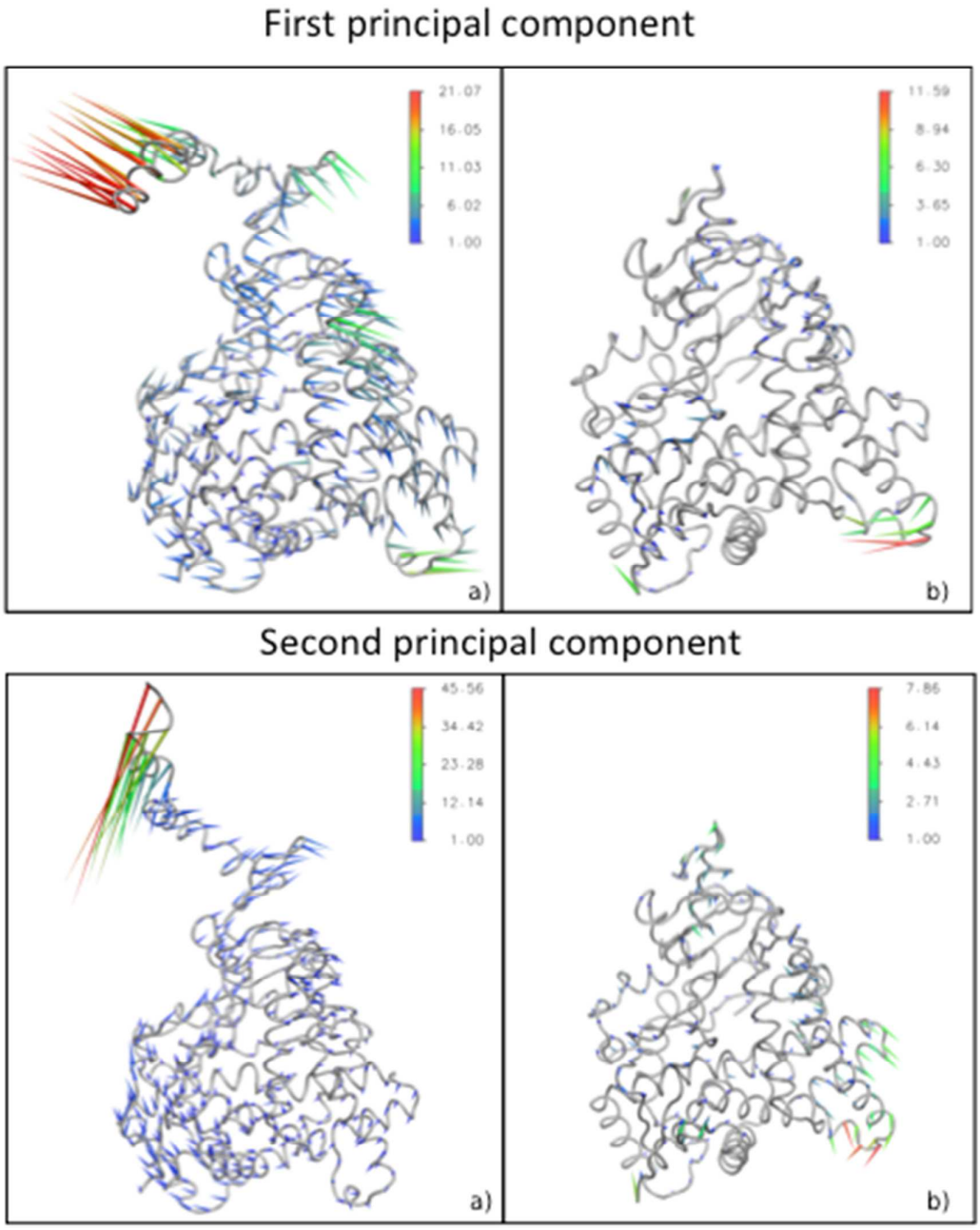

Figure S8. Porcupine plots of free aromatase when simulated on the membrane surface (a) or in water (b).

Perfomance of force field paramenters in describing the Heme group geometry.

Ability of the considered force field parameters in correctly reproducing the geomentrical properties of the Heme group has been verified by inspection of classical MD trajectories.

In particular the HEMe ring geometry was well reproduced being, similarly to the X-ray structure (pdb code 3EQM), the average Fe-N distances measured during the MD simulations betwen 2 and 2.10 **Å** and the average N-Fe-N angle ~89°.

Finally visual comparison of selected MD snaphots and the X-ray strcuture confirmed us the correct description of the HEME ring plane.

References

(1) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability *J. Chem. Inf. Mod.* **2009**, 49, 377-389.

(2) 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, 1595-1606