J Comput Aided Mol Des (2010) 24:771–787
DOI 10.1007/s10822-010-9373-1

Homology modeling and metabolism prediction of human carboxylesterase-2 using docking analyses by *GriDock*: a parallelized tool based on AutoDock 4.0

Giulio Vistoli · Alessandro Pedretti · Angelica Mazzolari · Bernard Testa

Received: 8 March 2010 / Accepted: 28 June 2010 / Published online: 11 July 2010

Abstract Metabolic problems lead to numerous failures during clinical trials, and much effort is now devoted to developing in silico models predicting metabolic stability and metabolites. Such models are well known for cytochromes P450 and some transferases, whereas less has been done to predict the activity of human hydrolases. The present study was undertaken to develop a computational approach able to predict the hydrolysis of novel esters by human carboxylesterase hCES2. The study involved first a homology modeling of the hCES2 protein based on the model of hCES1 since the two proteins share a high degree of homology (≅73%). A set of 40 known substrates of hCES2 was taken from the literature; the ligands were docked in both their neutral and ionized forms using *GriDock*, a parallel tool based on the AutoDock4.0 engine which can perform efficient and easy virtual screening analyses of large molecular databases exploiting multi-core architectures. Useful statistical models (e.g., $r^2 = 0.91$ for substrates in their unprotonated state) were calculated by correlating experimental pK_m values with distance between the carbon atom of the substrate's ester group and the hydroxy function of Ser228. Additional parameters in the equations accounted for hydrophobic and electrostatic interactions between substrates and contributing residues. The negatively charged residues in the hCES2 cavity explained the preference of the enzyme for neutral substrates and, more generally, suggested that ligands which interact too strongly by ionic bonds (e.g., ACE inhibitors) cannot be good CES2 substrates because they are trapped in the cavity in unproductive modes and behave as inhibitors. The effects of protonation on substrate recognition and the contrasting behavior of substrates and products were finally investigated by MD simulations of some CES2 complexes.

Keywords Human esterases · hCES2 · Prodrugs · Metabolism prediction · Parallel computing

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9373-1) contains supplementary material, which is available to authorized users.

G. Vistoli (✉) · A. Pedretti · A. Mazzolari
Dipartimento di Scienze Farmaceutiche "Pietro Pratesi", Facoltà di Farmacia, Università degli Studi di Milano, Via Mangiagalli, 25, 20133 Milan, Italy
e-mail: giulio.vistoli@unimi.it

B. Testa
Department of Pharmacy, University Hospital Centre (CHUV), Rue du Bugnon, CH-1011 Lausanne, Switzerland

Introduction

The failure of more than half of drug candidates during clinical trials is due to an unsuitable pharmacokinetic profile. As a result, the ADMET profiling of new chemical entities is receiving an ever increasing attention in drug discovery and development, with the clear aim to select and develop only drug-like compounds showing an optimal pharmacokinetic profile. Such early pharmacokinetic profiling requires the assessment of a maximum number of useful and relevant molecular properties as prerequisites for a reliable prediction of drug-likeness even in large molecular libraries [1].

Predictive approaches for metabolic behavior belong to the most useful in silico methods aimed at predicting the pharmacokinetic profile of new chemical entities, given that inadequate metabolism represents one of the most problematic failures during clinical trials. Because the large variety of drug-metabolizing enzymes, the diversity of biotransformation reactions they catalyze and the many

influencing factors [2, 3] render global quantitative predictions of heterogeneous molecular datasets virtually impossible, there is a clear trend in the recent literature to develop quantitative computational models restricted to specific metabolizing enzymes [4, 5].

The metabolizing activity of the most relevant redox enzymes [6] (i.e., cytochromes P450 [7], monoamine oxidases [8], and alcohol dehydrogenases [9]) and those of some conjugating enzymes [10] (e.g., UDP-glucuronosyltransferases [11], sulfotransferases [12], methyltransferases [13] and glutathione-S-transferases [14]) have been investigated in great detail by computational techniques and can be successfully predicted by several approaches. Apart from our previous studies focused on hCES1 [15, 16], the hydrolyzing activity of the human hydrolases has been scantly analyzed in silico [17], although they play key roles in the hydrolytic metabolism of xenobiotics as well as in the activation of most carrier-linked prodrugs [18–21].

Carboxylesterases (CESs, EC 3.1.1.1) catalyze the hydrolysis of a wide variety of compounds featuring an ester, thioester, carbamate or amide group [22, 23]. It has been proposed that CES proteins can be classified into five isozymes labeled CES1 to CES5 according to their sequence homology even if the majority of CESs so far identified belong to the CES1 or CES2 families [24]. CESs are generally located in the endoplasmic reticulum of cells; while CES1 is expressed in many tissues with the highest levels in the liver, CES2 is expressed in a more restricted number of tissues [25]. The expression of CESs is moderately altered by xenobiotics or pathological conditions and increases with the age also showing a marked inter-individual variability. These findings may suggest that carboxylesterases are developmentally regulated, with for example the consequence that children are more vulnerable than adults to insecticides such as pyrethroids whose detoxification is mediated by CESs [26].

Carboxylesterases belong to the α/β hydrolase superfamily whose active site is characterized by a catalytic triad composed of Glu, His and Ser plus an oxyanion hole [27]. These enzymes use a two-step catalytic mechanism whereby the serine residue is acylated by the substrate and then hydrolyzed by water (or transesterified by ethanol). Despite their marked homology ($\cong$73%), CES1 and CES2 have some significant differences in their structure which impact on their substrate specificity (see below) and quaternary architecture since the experimentally resolved CES1 structures show a trimer–hexamer equilibrium, while CES2 should exist as monomer [24].

The substrate specificities of CES1 and CES2 are considered as significantly different. CES1 mainly hydrolyzes substrates with a smaller alkyl/aryl group and a larger acyl group, although its wide catalytic pocket also allows structurally diverse compounds with a large alkyl/aryl moiety to be recognized. In contrast, the CES2 enzyme recognizes substrates with larger alkyl/aryl groups and smaller acyl groups, and its specificity appears constrained by structural hindrance in the active pocket [28]. The metabolic hydrolysis of cocaine can clearly illustrate such differences since its methyl ester group is hydrolyzed primarily by CES1 while its benzoyl group is hydrolyzed mainly by CES2 [29].

The present study continues our ongoing effort to analyze structure–function relationships of human esterases and is focused on the development of a computational approach able to predict and rationalize the metabolic hydrolysis of novel esters by the human CES2 enzyme (hCES2, also termed hCES2A1). Since the hCES2 structure has not been resolved to date, the first step of the study involved the generation of a reliable hCES2 homology model which, given the homology between human carboxylesterases, was generated using the resolved structure of hCES1 in complex with coenzyme A as the template. The second step in the study involved docking analyses of the resulting hCES2 model using an exhaustive set of known substrates. Docking simulations were carried out by *GriDock*, a parallel software based on AutoDock 4.0 [30] which we have developed for efficient and easy virtual screening analyses of large molecular databases exploiting multi-core architectures. Finally, some complexes were submitted to molecular dynamics (MD) simulations to investigate the stability of the homology model, the reliability of docking results and the contrasting behavior of substrates and products.

Methods

GriDock implementation

Over the last years, multi-core architectures have become commonplace in the processor market, with almost all new computers now having at least a dual-core chip. This trend is well understandable considering that processors have reached a point where it is increasingly hard to markedly improve their performances without resorting to multi-core systems. Even if computational chemists are already accustomed to use parallelized codes on computer clusters and supercomputers, many chemical applications were unfortunately designed to run sequentially. Hence, there is an increasing plea for parallel code, and multi-core architectures now force programmers to parallelize old codes [31], as demonstrated by several molecular modeling packages which have been successfully parallelized using message-passing interface (MPI) or Parallel Virtual Machine (PVM).

On this ground, we have developed a new virtual screening tool called *GriDock* based on AutoDock4.0 and running on multi-core systems and grid architectures thanks to its parallel design. As schematized in Fig. 1, the program written in highly portable C++ code can screen very large ligand databases combining VEGA features [32] with the AutoDock engine [30]. The former manages ligand libraries and prepares the input files required by the latter, which performs the actual docking calculations. *GriDock* controls both programs, collects the results from each docking calculation and stores them in suitable output files.

Concerning the pre-docking processes for the receptor, VEGA includes a script which prepares the required input files starting from a protein file in any supported format. As shown in Fig. 1a, this script (a) deletes non-polar hydrogen atoms; (b) assigns the AMBER atom types and Gasteiger's charges, (c) converts the protein in the requested PDBQT format, and (d) executes AutoDock Tools (ADT) in order to compute the corresponding grid maps. The pre-docking preparation of the ligand input files is automatically performed by *GriDock* without any manual intervention (Fig. 1a). Exploiting VEGA functions, *GriDock* reads the input databases (supported formats: Zip archives, SDF and SQLite files) and performs for each ligand the following operations: (a) converting from 2D to 3D structure and adding the hydrogen atoms (if necessary); (b) assigning the AMBER atom types and Gasteiger's charges; (c) recognizing the flexible torsions; (d) converting the ligand file in the requested PDBQT format; (e) sending the so obtained input files to AutoDock engine. It should be noted that *GriDock* can also accept as input a single ligand in any supported format.

GriDock distributes the computational load between the available processors (CPUs), exploiting different parallelization protocols depending on the network resources and operative system. As depicted by Fig. 1b, *GriDock* can run locally on multi-core workstations using the symmetric multiprocessing (SMP) or remotely on network and grid systems through the MPICH2 technology [33]. Moreover, *GriDock* can run on both UNIX/Linux systems and Windows-based PCs recognizing if the exploited CPUs have a 32- or 64-bit architecture. When *GriDock* runs locally, it exploits the POSIX thread libraries under UNIX/Linux environment and the multithread application programming interfaces (APIs) under Windows operative systems. Irrespective of the operative system, *GriDock* runs remotely using the MPICH2 technology which is a widely portable implementation of the message passing interface (MPI).

When *GriDock* runs locally, the input files are shared by all nodes, whereas they are distributed to all remote workers by MPI library. In particular, the MPI master node controls the activity of each worker that screens one ligand at a time and sends the results to the master node. The master node collects the docking results from each node and writes three output files: (a) a log file to check the

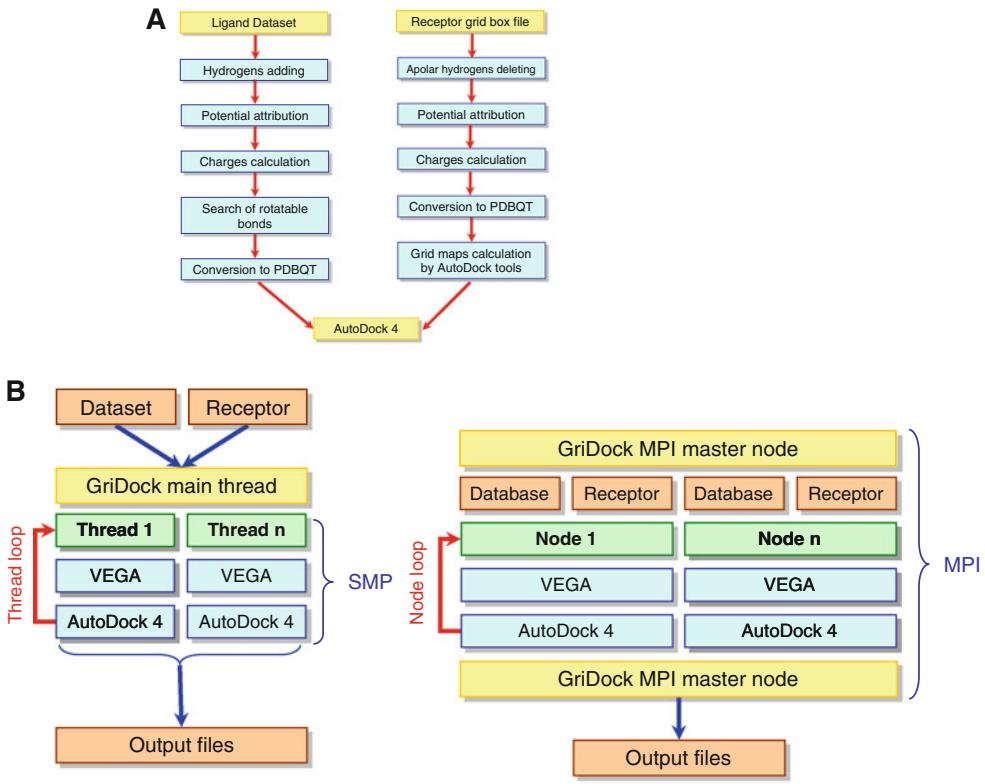

Fig. 1 Main flowcharts for the *GriDock* program representing (**a**) the pre-docking preparation of the receptor and ligand input files and (**b**) the parallelization schemes comparing the local architecture based on the symmetric multiprocessing (SMP), as provided by pthread library or Windows APIs, with the remote architecture based on MPICH2

calculation progress; (b) a table compiling the computed complexes ranked by docking score; (c) a zip archive including the structure of all complexes as generated by AutoDock 4. The complexes are saved in the AutoDock DLG files which are conveniently analyzed by VEGA as trajectory files. Gridock can be downloaded from www.vegazz.net and it is included in the VEGA suite of programs starting from 2.3.2 release.

Generation of hCES2 model

The amino acid sequence of human CES2 was retrieved from the Swiss-Prot database (entry code O00748, EST2_HUMAN). The primary transcription sequence was submitted to the Signal IP 3.0 Server [34] to predict the cleavage site of the signal peptide. Predictions show the highest probability of cleavage between Gly26 and Gln27. The primary sequence was submitted to Swiss-Model [35], a fully automated protein structure homology modeling server which selects suitable templates based on a Blast E-value limit providing truly reliable models when target and template share a significant percentage of identical residues. And indeed, the proposed model was generated using the resolved structure of hCES1 in complex with coenzyme A as the template (PDB Id: 2H7C, resolution 2.0 Å, identity hCES1 vs. hCES2 = 46.6%). The model quality appears satisfactory as assessed by: (a) the remarkable percentage of residues which fall in the allowed regions of Ramachandran plot (81.93%); (b) the lacking of relevant gaps apart from the first four and the last three residues which were manually added using the VEGA peptide builder; and (c) a very similar percentage of secondary motifs compared to hCES1 (as detailed under "Results and discussion").

With the full-length backbone constructed, side-chains and hydrogen atoms were added using VEGA. To remain compatible with physiological pH values, the side-chains of Arg, Lys, Glu, and Asp were ionized, while His and Cys residues were considered neutral by default. The complete model was carefully checked to avoid unphysical occurrences such as *cis* peptide bonds, wrong configurations or colliding side-chains. Then, the hCES2 model underwent an initial minimization until RMS gradient was equal to 1 kcal $mol^{-1}Å^{-1}$ to discard high-energy interactions. Finally, the model was optimized by a final minimization made up by two phases: first a minimization without constraints until RMS = 0.1 kcal $mol^{-1}Å^{-1}$ and then a second minimization with backbone fixed until RMS = 0.01 kcal $mol^{-1}Å^{-1}$ to preserve the predicted structure.

Ligand datasets

A ligand dataset of 40 known hCES2 substrates was collected from the literature [36–41] by selecting molecules whose hCES2-catalyzed hydrolysis was determined using purified recombinant hCES2 and expressed as K_m (in μM as compiled in Tables 1, 2, 3 and Fig. S1, Supporting Information). The ionizable substrates were docked considering the most probable ionization state at physiological pH, although all simulations were repeated with the basic groups in their unprotonated state. The conformational behavior of the compounds was investigated by a Monte-Carlo procedure (as implemented in the VEGA suite of programs [32]) which generated 1,000 conformers by randomly rotating the rotors. All geometries so obtained were stored and optimized to avoid high-energy rotamers. The 1,000 conformers were clustered according to their similarity to discard redundant ones; in this analysis two geometries were considered as non-redundant when they differed by more than 60° in at least one torsion angle. For each ligand, the so obtained lowest energy structure was then exploited in the following docking simulations. As described in the Results, the ligand dataset was then subdivided in the training ($n = 25$) and test ($n = 15$) sets. Even if randomly subdivided, the Tanimoto comparison of the chemical fingerprints [42] for substrates included in the two sets indicates that they have a similar average variability (0.27 vs. 0.28) suggesting that they equally cover the chemical space of hCES2 substrates.

Docking simulations

As previously described, docking simulations were performed by *GriDock*, a parallel tool based on the AutoDock4.0 engine. In detail, the grid box was set to include all residues within a 15 Å radius sphere around the catalytic Ser228 residue thus comprising the entire catalytic cavity. The resolution of the grid was 60 × 60 × 60 points with a grid spacing of 0.450 Å. Each substrate was docked into this grid with the Lamarckian algorithm as implemented in AutoDock. For the docking simulations, the flexible bonds of the ligand were automatically recognized by *GriDock* and left free to rotate. The genetic-based algorithm ran 20 simulations per substrate with 2,000,000 energy evaluations and a maximum number of generations of 27,000. The crossover rate was increased to 0.8, and the number of individuals in each population to 150. All other parameters were left at the AutoDock default settings [30].

The docking results were ranked considering both AutoDock scores and the closeness between Ser228 and the labile function. The best complexes were minimized keeping fixed all atoms outside a 15 Å radius sphere around the bound substrate to favor the mutual adaptability between ligand and enzyme. The optimized complexes were then used to re-calculate AutoDock docking scores, the VEGA energy scores and the Molecular Lipophilicity Potential Interaction Score (MLP_{InS} [16]) that we have

Table 1 Substrates and parameters used to develop Eq. 1 (as obtained by docking unprotonated basic ligands)

Substrates[a]	Exp K_m (μM)	Exp pK_m	Pred pK_m	Δ (exp-pred)	Distance Ser228 (Å)	MLP_{InS}	Volume (Å^3)	Ref.
(α*R*,1*R*)-A4	0.9	6.05	6.40	0.35	3.90	−0.65	342.7	[36]
(α*R*,1*R*)-*cis*-A3	0.9	6.05	6.23	0.18	3.10	−0.64	367.0	[36]
Irinotecan*	1.1	5.96	6.21	0.25	3.29	−0.52	513.1	[41]
(α*S*,1*R*)-*cis*-A3	1.4	5.85	6.14	0.29	3.98	−0.64	350.3	[36]
(α*R*,1*S*)-*cis*-A3	1.7	5.77	6.13	0.36	3.75	−0.63	346.1	[36]
(α*R*,1*S*)-*trans*-A3	1.9	5.72	6.07	0.35	3.98	−0.62	346.5	[36]
(α*R*,1*S*)-A4	2	5.70	6.15	0.45	3.95	−0.62	366.0	[36]
A6	2.1	5.68	6.32	0.65	3.77	−0.67	344.1	[36]
(α*R*,1*R*)-*trans*-A3	2.5	5.60	6.09	0.48	3.61	−0.60	345.2	[36]
(α*S*,1*S*)-*cis*-A3	2.5	5.60	6.26	0.66	3.17	−0.63	344.2	[36]
(α*S*,1*R*)-*trans*-A3	2.8	5.55	6.15	0.59	3.89	−0.63	350.3	[36]
NPC*	3.2	5.49	5.93	0.44	4.46	−0.60	337.1	[41]
(α*S*,1*S*)-*trans*-A3	3.8	5.42	6.23	0.81	3.72	−0.64	343.7	[36]
A8	5.4	5.27	6.00	0.74	3.92	−0.61	320.3	[36]
cis-Permethrin (*R,R*)	7.5	5.12	5.20	0.08	3.49	−0.38	336.0	[37]
trans-Permethrin (*R,S*)	8.6	5.06	5.17	0.11	3.57	−0.33	348.3	[37]
Cypermethrin	9.1	5.04	5.27	0.23	3.44	−0.39	336.0	[36]
(*R*)-Propranolol butyrate*	14.9	4.83	5.15	0.32	4.44	−0.39	326.3	[39]
(*S*)-Propranolol butyrate*	20.1	4.70	5.00	0.31	4.50	−0.36	326.3	[39]
p-Nitrophenyl acetate	31	4.51	4.21	0.29	2.99	−0.26	151.8	[36]
Dilazep*	89	4.05	4.69	0.64	3.35	−0.09	563.0	[38]
Butyl benzoate	110	3.96	4.12	0.16	2.92	−0.15	174.6	[40]
Quinapril*	122	3.91	3.97	0.05	4.79	−0.10	400.7	[38]
p-Nitrophenyl butyrate	131	3.88	3.88	0.00	3.16	−0.13	199.4	[40]
6-Monoacetyl morphine*	132	3.88	3.77	0.11	5.54	−0.20	292.2	[36]
4-Methylumbelliferyl acetate	150	3.82	3.94	0.11	2.61	−0.13	186.7	[36]
APC*	270	3.47	3.26	0.30	5.48	0.07	540.5	[41]
Aniracetam (exocyclic amide)	301	3.52	3.70	0.18	3.32	−0.08	198.8	[38]
Temocapril*	325	3.49	3.66	0.17	5.68	−0.11	385.8	[38]
Cocaine* (benzoyl ester group)	390	3.41	3.80	0.39	4.72	−0.16	280.2	[36]
Propyl benzoate	398	3.40	3.76	0.36	2.79	−0.07	162.1	[40]
Aniracetam (lactam ring)	412	3.39	3.92	0.54	4.01	−0.18	198.8	[38]
Benazepril*	785	3.11	4.01	0.90	2.98	−0.01	388.3	[38]
Capecitabine	1000	3.00	2.96	0.04	5.22	0.01	306.2	[40]
Oxybutin*	1124	2.95	3.75	0.80	5.88	−0.14	358.1	[38]
Ethyl benzoate	1198	2.92	3.56	0.64	2.78	−0.04	144.1	[40]
Cilazapril*	1349	2.87	4.48	1.61	3.99	−0.21	391.7	[38]
Aspirin	2270	2.64	3.19	0.55	3.39	−0.02	150.5	[38]
Camostat[b]	2700	2.57	2.93	0.36	6.07	0.00	351.2	[38]
Procaine*	3330	2.48	3.34	0.86	5.06	−0.09	233.9	[38]

[a] Asterisk denotes the basic compounds here considered in their unprotonated state

[b] Due to the strong basicity of guanidine function, camostat was simulated only in its protonated form

recently developed to account for hydrophobic interactions, the importance of which is well documented in ligand recognition [43]. In the statistical analyses, a small set of representative physicochemical properties of the bound ligands (e.g., virtual log *P*, PSA, SAS, volume, molecular weight and dipole moment) were also taken into account.

Table 2 Substrates and parameters used to develop Eq. 2 (as obtained by docking protonated basic ligands)

Substrates[a]	Exp K_m (μM)	Exp pK_m	Pred pK_m	Δ (exp-pred)	CVFF (kcal/mol)	MLP_{InS}	ASA (Å^2)
(α*R*,1*R*)-A4	0.9	6.05	5.39	0.66	−67.49	−0.64	534.4
(α*R*,1*R*)-*cis*-A3	0.9	6.05	5.72	0.32	−40.63	−0.65	569.1
Irinotecan*	1.1	5.47	4.92	1.03	−46.64	1.04	719.6
(α*S*,1*R*)-*cis*-A3	1.4	5.85	5.44	0.41	−58.35	−0.64	550.6
(α*R*,1*S*)-*cis*-A3	1.7	5.77	5.31	0.46	−55.68	−0.63	522.8
(α*R*,1*S*)-*trans*-A3	1.9	5.72	5.48	0.24	−54.83	−0.62	548.1
(α*R*,1*S*)-A4	2	5.70	5.61	0.09	−57.03	−0.62	574.2
A6	2.1	5.68	5.53	0.15	−55.37	−0.67	548.6
(α*R*,1*R*)-*trans*-A3	2.5	5.60	5.26	0.34	−57.72	−0.63	545.6
(α*S*,1*S*)-*cis*-A3	2.5	5.60	5.71	0.11	−41.96	−0.60	521.7
(α*S*,1*R*)-*trans*-A3	2.8	5.55	5.46	0.09	−57.25	−0.63	550.6
NPC*	3.2	5.49	5.14	0.35	−40.83	0.05	615.6
(α*S*,1*S*)-*trans*-A3	3.8	5.42	5.50	0.08	−58.71	−0.64	560.9
A8	5.4	5.27	5.32	0.05	−54.37	−0.61	523.5
cis-Permethrin (*R,R*)	7.5	5.12	4.84	0.28	−67.97	−0.38	534.5
trans-Permethrin (*R,S*)	8.6	5.06	5.36	0.29	−67.25	−0.33	616.5
Cypermethrin	9.1	5.04	5.28	0.24	−47.60	−0.39	534.5
(*R*)-Propranolol butyrate*	14.9	4.83	4.57	0.26	−48.29	0.49	578.8
(*S*)-Propranolol butyrate*	20.1	4.70	4.82	0.12	−31.87	0.50	566.5
p-Nitrophenyl acetate	31	4.51	3.15	1.36	−43.35	−0.26	224.7
Dilazep*	89	4.05	4.64	0.59	−140.80	−0.01	790.8
Butyl benzoate	110	3.96	4.07	0.11	−41.75	−0.15	375.1
Quinapril*	122	3.91	4.07	0.16	−46.34	0.90	567.9
p-Nitrophenyl butyrate	131	3.88	3.50	0.38	−54.15	−0.13	332.6
6-Monoacetyl morphine*	132	3.88	3.43	0.45	−65.69	0.26	424.1
4-Methylumbelliferyl acetate	150	3.82	3.84	0.02	−27.24	−0.13	301.5
APC*	270	3.57	2.95	0.62	−90.85	1.14	578.5
Aniracetam (exocyclic amide)	301	3.52	3.91	0.39	−39.14	−0.08	356.2
Temocapril*	325	3.49	4.39	0.90	−41.43	0.53	538.0
Cocaine* (benzoyl ester group)	390	3.41	3.33	0.08	−61.61	0.43	424.8
Propyl benzoate	398	3.40	3.85	0.45	−39.14	−0.07	349.4
Aniracetam (lactam ring)	412	3.39	4.22	0.84	−29.42	−0.18	356.2
Benazepril*	785	3.11	3.13	0.02	−121.88	0.46	585.4
Capecitabine	1000	3.00	3.81	0.81	−54.74	0.01	405.8
Oxybutin*	1124	2.95	4.37	1.42	−73.20	0.09	558.6
Ethyl benzoate	1198	2.92	3.51	0.58	−40.57	−0.04	307.3
Cilazapril*	1349	2.87	3.62	0.75	−110.05	0.14	568.6
Aspirin	2270	2.64	3.06	0.41	−41.50	−0.02	247.3
Camostat[b]	2700	2.57	2.69	0.12	−127.84	0.00	460.7
Procaine*	3330	2.48	1.58	0.89	−117.59	0.60	365.2

The references for biological data are reported in Table 1

[a] Asterisk denotes the basic compounds here considered in their protonated state

[b] Due to the strong basicity of guanidine function, camostat was simulated only in its protonated form

MD simulations

MD simulations involving hCES2 in complex with (*R*)-propranolol butyrate and (*R*)-propranolol were performed, with the ligands both in their neutral and protonated forms. The complexes of hCES2 with (*R*)-propranolol butyrate were taken from docking results, while the complexes of the enzyme with (*R*)-propranolol were generated by manually transforming the

Table 3 Substrates and parameters used to develop Eq. 3 (as obtained by averaging parameters from Tables 1, 2)

Substrates	Exp K_m (μM)	Exp pK_m	Pred pK_m	Δ (exp-pred)	pK[a]	Percent protonated	CVFF (kcal/mol)	MLP_{InS}	ASA (Å^2)
(α*R*,1*R*)-A4	0.9	6.05	5.12	0.48	–	0	−40.63	−0.65	534.4
(α*R*,1*R*)-*cis*-A3	0.9	6.05	5.35	0.20	–	0	−67.49	−0.64	569.1
Irinotecan*	1.1	5.47	5.26	0.24	9.3	0.988	−46.67	1.02	719.7
(α*S*,1*R*)-*cis*-A3	1.4	5.85	5.40	0.02	–	0	−58.35	−0.64	550.6
(α*R*,1*S*)-*cis*-A3	1.7	5.77	5.20	0.06	–	0	−55.68	−0.63	522.8
(α*R*,1*S*)-*trans*-A3	1.9	5.72	4.73	0.39	–	0	−54.83	−0.62	548.1
(α*R*,1*S*)-A4	2	5.70	5.32	0.26	–	0	−57.03	−0.62	574.2
A6	2.1	5.68	5.25	0.21	–	0	−55.37	−0.67	548.6
(α*R*,1*R*)-*trans*-A3	2.5	5.60	4.75	0.20	–	0	−41.96	−0.60	545.6
(α*S*,1*S*)-*cis*-A3	2.5	5.60	5.06	0.20	–	0	−57.72	−0.63	521.7
(α*S*,1*R*)-*trans*-A3	2.8	5.55	3.90	0.61	–	0	−57.25	−0.63	550.6
NPC*	3.2	5.49	4.56	0.51	9.9	0.997	−40.86	0.05	567.3
(α*S*,1*S*)-*trans*-A3	3.8	5.42	3.98	0.02	–	0	−58.71	−0.64	560.9
A8	5.4	5.27	5.00	1.08	–	0	−54.37	−0.61	523.5
cis-Permethrin (*R,R*)	7.5	5.12	3.32	0.56	–	0	−67.97	−0.38	534.5
trans-Permethrin (*R,S*)	8.6	5.06	3.37	0.51	–	0	−67.25	−0.33	616.5
Cypermethrin	9.1	5.04	3.76	0.07	–	0	−47.60	−0.39	534.5
(*R*)-Propranolol butyrate*	14.9	4.83	3.09	0.48	9.6 [49]	0.990	−48.37	0.48	578.9
(*S*)-Propranolol butyrate*	20.1	4.70	3.83	0.31	9.6 [49]	0.990	−32.19	0.49	566.7
p-Nitrophenyl acetate	31	4.51	3.32	0.17	–	0	−43.35	−0.26	224.7
Dilazep*	89	4.05	3.35	0.06	9	0.962	−138.81	−0.01	791.8
Butyl benzoate	110	3.96	3.77	0.37	–	0	−41.75	−0.15	375.1
Quinapril*	122	3.91	4.17	0.78	5.5	0.012	−71.12	−0.09	616.1
p-Nitrophenyl butyrate	131	3.88	3.62	0.52	–	0	−54.15	−0.13	332.6
6-Monoacetyl morphine*	132	3.88	3.73	0.73	9.3 [49]	0.988	−65.43	0.26	424.2
4-Methylumbelliferyl acetate	150	3.82	4.48	1.54	–	0	−27.24	−0.13	301.5
APC*	270	3.57	3.39	0.47	9.8	0.996	−90.72	1.14	579.0
Aniracetam (exocyclic amide)	301	3.52	3.99	1.12	–	0	−39.14	−0.08	356.2
Temocapril*	325	3.49	2.89	0.25	4.6 [15]	0.001	−122.93	−0.11	562.2
Cocaine* (benzoyl ester group)	390	3.41	2.33	0.23	8.6	0.941	−61.39	0.40	425.1
Propyl benzoate	398	3.40	1.39	1.08	–	0	−39.14	−0.07	349.4
Aniracetam (lactam ring)	412	3.39	5.12	0.48	–	0	−29.42	−0.18	356.2
Benazepril*	785	3.11	5.35	0.20	5.5	0.012	−115.59	−0.01	590.8
Capecitabine	1000	3.00	5.26	0.24	–	0	−54.74	0.01	405.8
Oxybutin*	1124	2.95	5.40	0.02	8	0.715	−70.76	0.02	560.8
Ethyl benzoate	1198	2.92	5.20	0.06	–	0	−40.57	−0.04	307.3
Cilazapril*	1349	2.87	4.73	0.39	5.5	0.012	−97.60	−0.21	555.2
Aspirin	2270	2.64	5.32	0.26	–	0	−41.50	−0.02	247.3
Camostat	2700	2.57	5.25	0.21	–	0	−127.84	0.00	460.7
Procaine*	3330	2.48	4.75	0.20	9.1 [49]	0.969	−115.09	0.58	365.3

The references for biological data are reported in Table 1

[a] For experimental pK values, the corresponding reference is reported in bracket. The undetermined pK values were predicted by the *Pharma Algorithms* method

enzyme complexed with (*R*)-propranolol butyrate. The complexes with (*R*)-propranolol were then minimized keeping fixed all atoms outside a 15 Å radius sphere around the bound product before performing the MD simulation.

The complexes so obtained were inserted into a 50 Å radius sphere of water molecules, and, after a preliminary minimization to optimize the relative position of solvent molecules, the systems underwent 10 ns of all-atoms MD

simulations with the following characteristics: (a) spherical boundary conditions were introduced to stabilize the simulation space; (b) Newton's equation was integrated using the r-RESPA method (every 4 fs for long-range electrostatic forces, 2 fs for short-range non bonded forces, and 1 fs for bonded forces); (c) the temperature was maintained 300 ± 10 K by means of Langevin's algorithm; (d) Lennard-Jones (L–J) interactions were calculated with a cut-off of 10 Å and the pair list was updated every 20 iterations; (e) a frame was stored every 5 ps, yielding 2,000 frames; and (f) no constraints were applied to the systems.

The simulations were carried out in two phases: an initial period of heating from 0 to 300 K over 6,000 iterations (6 ps, i.e., 1 K/20 iterations), and a monitored phase of simulation of 10 ns. Only the frames memorized during this last phase were considered. All described minimizations were performed using the conjugate gradient algorithm. The calculations described here were carried out on a 16 CPU Tyan-VX50 system using Namd2.6 [44] with the force field CHARMM and Gasteiger's atomic charges.

Results and discussion

Analysis of hCES2 homology models

Figure 2a shows the cartoon structure of the proposed hCES2 model as colored according to secondary structure types and evidences the α/β topology typical of the serine hydrolase superfamily. Figure 2b depicts the superimposition of hCES2 *vs.* hCES1 (as performed with the MATRAS software [45]) and confirms a global similarity between the two considered isozymes as underlined by rmsd values (1.56 Å as computed considering Cα atoms only). The small differences mainly involve the arrangement of α-helices, while the folding of β-sheets is perfectly conserved. Globally, the two structures show similar percentages of secondary motifs (hCES1 = 32% for α-helices, 13% for β-sheets and 35% for structured loops; hCES2 = 33% for α-helices, 16% for β-sheets and 31% for structured loops).

A more careful analysis of the hCES2 homology model reveals that the N-terminal region is mainly organized in antiparallel β-sheets and includes the catalytic domain which contains the highly conserved catalytic triad (i.e., Ser228, Glu345 and His457) at the base of the active site gorge. The entrance to the substrate binding gorge is flanked by α-helices and surrounded by both polar and apolar residues. A key difference between the catalytic domains of hCES2 and hCES1 (as clearly shown by Fig. 2b) is the deletion in the former of a large loop near the substrate binding site of hCES1 (from Asp307 to Thr321). As depicted by Fig. 2b, such a loop is located at the entrance of the catalytic cavity and acts as a lid modulating enzyme accessibility. Hence, its deletion in hCES2 renders the catalytic cavity wider and more accessible to the solvent and may explain the observed differences in substrate specificity due to a different architecture of the active pocket.

The C-terminal domain is composed by several α-helices, and, in analogy with hCES1, it should define the Z-site which modulates the trimer-hexamer equilibrium and influences the catalytic efficiency [46]. However, since hCES2 should be active as monomer [24], one may postulate that the putative Z-site in hCES2 functions in substrate screening and allosteric regulation only. A detailed comparison of the main residues facing the Z-site in the two isozymes reveals that some relevant polar residues in the hCES1 Z-site are not conserved in the corresponding region of hCES2. For example, the three consecutive lysine

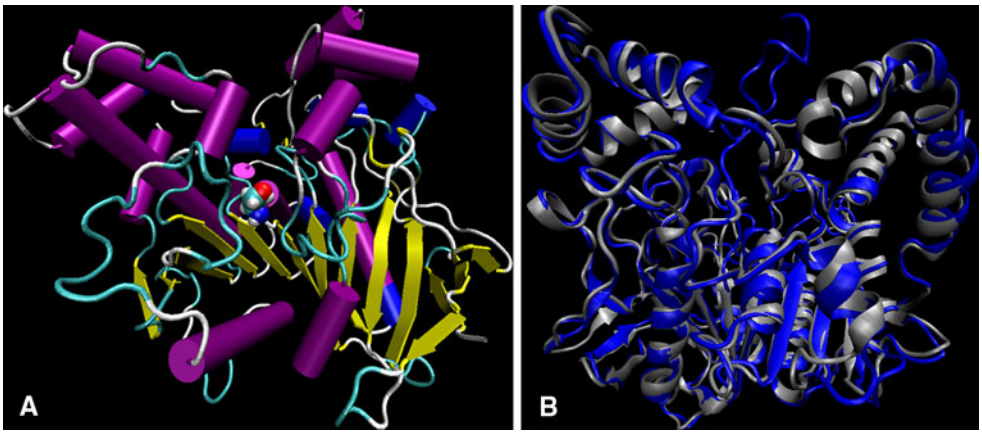

Fig. 2 Homology modelled hCES2 structure. **a** Cartoon structure for the generated model as colored by secondary structure with the catalytic Ser228 residue represented as CPK (color legend: *yellow* β-sheets; *purple* α-helices; *blue* 3.10 helices; *azure* turns; *white* coils). **b** Superimposition of hCES1 (*blue*) and hCES2 (*silver*) as represented by a solid ribbon

residues which characterize the Z-site of hCES1 (Lys412, Lys413, and Lys414) are not conserved in hCES2 (Thr402, Leu403 and Gln404). Similarly, other three important polar residues lining the hCES1 Z-site (Ser369, Glu370, Gln372) are not precisely conserved in hCES2 (Gln360, Lys361, Glu362), while the key apolar residues facing the hCES1 Z-site (e.g., Phe355, Trp357, Pro360 and Tyr366) are highly conserved in hCES2 (Phe346, Trp348, Pro351 and Tyr356). These findings suggest that the Z-site binding features in hCES2 can be significantly different from those of hCES2 even if it should be conceivable that the hCES2 Z-site conserves the ligand promiscuity which parallels the catalytic site promiscuity and characterizes similar accessory binding sites in other metabolizing enzymes. Again, the non-conservation of some key ionized residues in the hCES2 Z-site may explain why hCES2 does not assemble in oligomeric quaternary structures as seen by hCES1.

A precise conservation is conversely observed when analyzing the key residues facing the so called side door [47]. This is an alternative exit route which is exploited by esterases to push out small products from the catalytic gorge. The possibility to remove small molecules through the side door plays a key role in determining catalytic efficiency toward some substrates as demonstrated by the effect of mutations on the hydrolysis of irinotecan [48]. These studies unveiled the pivotal role of four apolar residues lining the side door of hCES1 (Leu251, Met425, Phe426 and Phe546). Such hydrophobic residues are exactly conserved also in hCES2 (Leu258, Met415, Phe416, and Phe537) suggesting that also here they can be involved in promoting the release of small products.

Docking results

Prediction of hCES2 activity

In a previous modeling study of hCES1 biocatalysis [16], we showed that the protonation state of basic substrates plays a key role in their recognition by the enzyme and in modulating the efficiency of the catalytic turnover. In particular, the simulations revealed that protonated substrates are bound and trapped in the catalytic pocket by strong ionic bonds suggesting that the neutral substrates are better recognized by hCES1 than the cations. With a view to analyzing the role of protonation also in hCES2, docking analyses involved two sets of calculations since the substrates with basic groups were simulated in both unprotonated and protonated states. Here, we first describe correlation studies aimed at identifying the best set of docking calculations and ultimately at focusing our analysis of docking results on the most predictive ones.

Table 4 reports the correlation equations (Eqs. 1, 2) obtained for the dataset of 40 substrates in the two sets of docking simulations (Tables 1, 2). All correlations included only three independent variables to guarantee their statistical robustness. The correlations reveal that the substrates with unprotonated basic groups (Eq. 1, Table 1) correlate markedly better than the protonated ones (Eq. 2, Table 2) probably because the latter are biased by ionic contacts. Similarly to that observed for hCES1 [16], this result may suggest that the neutral forms are better recognized by the enzyme than the cations.

Both equations include the MLP_{InS} parameter, confirming the crucial role of hydrophobic contacts also in hCES2 binding. Although there is a fair and expected intercorrelation between MLP_{InS} scores and virtual log P values as computed by the MLP algorithm ($r^2 = 0.55$), the substitution of log P instead of MLP_{InS} in Eqs. 1, 2 markedly worsened their statistical parameters (e.g., r^2 decreased from 0.91 to 0.57 in Eq. 1). This emphasizes that MLP_{InS} scores account for the mutual hydrophobic contacts between substrate and enzyme, an information not encoded in log P values, and one that plays a pivotal role in the equations reported.

Notably, both equations include two docking-based parameters and one ligand-based descriptor which is the molecular volume of unprotonated ligands and the apolar surface area (ASA) of protonated ones. Apolar surface further emphasizes the beneficial role of hydrophobic contacts even if there is not cross-correlation between MLP_{InS} scores and ASA values thus justifying their simultaneous use in such relationship. Since the apolar surface can be seen also as a descriptor of molecular size, both ligand-based parameters may indicate that hCES2 can efficiently hydrolyze also large substrates. The second docking-based parameter included in the equations for the protonated substrates (Eq. 2) confirms that ionic contacts play a detrimental role as indicated by the positive sign of CVFF score which accounts for polar interactions computed by the CVFF force field.

It is interesting to observe that the correlation for the compounds with unprotonated basic groups (Eq. 1, Table 1) includes the distance between the carbon atom of substrate ester group and the hydroxy function of Ser228 ($Dist_{Ser228}$). This distance codes for the ability of a molecule to assume a binding mode conducive to catalysis and is well understandable considering that docking calculations can simulate only the recognition phase between enzyme and substrates to the exclusion of the catalytic phase. In other words, a given compound can be found to elicit good binding interactions and high docking scores with an enzyme, yet fail to possess an appropriate target group or fail to allow its binding in a catalytically productive mode. Interestingly, the correlation for the compounds with protonated basic groups (Eq. 2, Table 2) does not include such parameter suggesting that docking

Table 4 Best correlations as derived from docking simulations

Eq.	Ionization[a]	Equation[b,c,d,e]	Statistics
1	N	$pK_m = -4.17(\pm0.23)\ MLP_{InS}$	$n = 40$;
		$-\ 0.25(\pm0.071)Dist_{Ser228}$	$r^2 = 0.91$;
		$+\ 2.12 \cdot 10^{-3}(\pm6.35 \times 10^{-4})$ Volume	$s = 0.36$;
		$+\ 3.43(\pm0.301)$	$F = 122.09$
2	N+	$pK_m = -1.46(\pm0.19)\ MLP_{InS}$	$n = 40$;
		$+\ 1.87 \times 10^{-2}(\pm3.61 \times 10^{-3})$ CVFF	$r^2 = 0.76$;
		$+\ 6.83 \times 10^{-3}(\pm8.37 \times 10^{-4})$ ASA	$s = 0.57$;
		$+\ 1.93(\pm0.38)$	$F = 39.04$
3	N/N+	$pK_m = -0.93(\pm0.11)\ MLP_{InS}$	$n = 40$;
	(weighted parameters)	$+\ 2.49 \times 10^{-2}\ (\pm3.71 \times 10^{-3})$ CVFF	$r^2 = 0.80$;
		$-\ 7.53 \times 10^{-3}(\pm8.73 \times 10^{-4})$ ASA	$s = 0.55$;
		$+\ 2.05(\pm0.41)$	$F = 46.41$
4	N	$pK_m = -4.15(\pm0.37)\ MLP_{InS}$	$n = 25$;
	(training set)	$-\ 0.23(\pm0.093)Dist_{Ser228}$	$r^2 = 0.90$;
		$-\ 1.34 \times 10^{-3}(\pm6.91 \times 10^{-4})$ Volume	$s = 0.38$;
		$+\ 3.51(\pm0.43)$	$F = 61.92$

[a] Ionization state of basic substrates: N = neutral forms; N + = protonated forms

[b] MLP_{InS} = Molecular Lipophilicity Potential (MLP) Interaction Score

[c] $Dist_{Ser228}$ = distance between the carbon atom of substrate's ester group and the hydroxy function of Ser228

[d] ASA = apolar surface area

[e] CVFF = non-bond interaction energy as computed by the CVFF force-filed including an electrostatic term based on a distance dielectric function and a Van der Waals term based on the Lennard-Jones (LJ-6-12) equation

simulations for protonated substrates are so biased by ionic interactions that they are no more able to accommodate the labile function proportionally near to Ser228.

With a view to verifying whether protonated and unprotonated states can mutually contribute to hCES2 biorecognition depending on ionization constants, Table 3 compiles the averaged parameters weighting the descriptors from both sets of docking simulations (Tables 1, 2) in function of the relative abundance of protonated and unprotonated species at physiological pH. The pK values (see Table 3, [49]), for the substrates whose ionization constants were not reported in literature, were predicted using the *Pharma Algorithms* method as implemented in the *VCCLAB* website (www.vcclab.org). The parameters collected in Table 3 were exploited to develop a new correlation (Eq. 3, Table 4) which includes the same descriptors of Eq. 2 and shows a slightly better predictive power than that of Eq. 2, but far below from that of Eq. 1 as obtained for unprotonated substrates.

A comparison of the descriptors collected in Tables 1, 2, 3 evidences that the weighted values (Table 3) are very similar to those compiled in Table 2 for protonated forms apart from the parameters for ACE inhibitors which are quite superimposable to those computed for unprotonated forms (Table 1) due to the very low basicity of their amino group as experimentally determined for temocapril [15] and calculated for the other analogues. Stated differently, the ionization constants of the basic substrates render the protonated forms largely predominant at physiological pH apart from the mentioned ACE inhibitors, a fact explaining why Eq. 3 shows statistics slightly better than Eq. 2 but clearly poorer than Eq. 1 which remains the best relationship to predict the stability of new medicinal esters.

To further assess the predictive power of Eq. 1, the dataset was randomly subdivided into a training ($n = 25$) and a test ($n = 15$) set. The correlation as computed for the training set only (Eq. 4, Table 4) confirms the reliability of Eq. 1 since these two equations (Eqs. 1, 4) have quite identical coefficients suggesting that Eq. 1 is a stable correlation and reasonably independent on the number of considered ligands. The relation between experimental and predicted pK_m values (Eq. 5), as computed by Eq. 4 for the external test set, affords an encouraging validation for the predictive power of Eq. 1. Equation 5 is indeed a satisfactory one given its statistics, its slope being close to 45° ($\alpha = 43.7°$) and the intercept close to zero:

$$pK_m = 0.956pK_m\text{pred} + 0.021$$
$$n = 15;\ r^2 = 0.92;\ SE = 0.29;\ F = 212.29 \quad (5)$$

A more in-depth analysis of the pK_m values, as predicted by Eq. 1 for the entire dataset, reveals that only one

substrate shows a difference between experimental and predicted pK_i values greater than 1 (cilazapril, $\Delta pK_m = 1.15$), whereas for 27 substrates out of 40 the difference is below 0.5. The analysis of residuals shows that there is no correlation with the pK_m values suggesting that Eq. 1 is equally predictive for good and poor hCES2 substrates.

Analysis of the putative complexes

As a preamble, we recall that AutoDock 4.0 proved successful in accommodating the substrates in the hCES2 cavity in a pose conducive to catalysis. For the vast majority of docked substrates, indeed, the lowest energy complex was that with the hydrolyzable group nearest to Ser228. Again, AutoDock scores, even if not involved in the reported equations, conveniently accounted for the substrate regioselectivity seen with aniracetam and cocaine.

Figure 3 reports the main residues lining the hCES2 cavity, showing its noteworthy abundance in hydrophobic residues and mainly alkyl side-chains. As observed for hCES1 [16], the negatively charged residues in the catalytic cavity (i.e., Glu105, Asp200, Glu227 and Glu345) have the plausible role of favoring the egress of enzymatic products and can also explain why the binding modes of the substrates with protonated basic groups are greatly influenced by ionic contacts as discussed above.

Similarly to that observed for the hCES1 isozyme [16], the catalytic Ser228 residue subdivides the hCES2 cavity into two subpockets of quite different polarity. Indeed, the region that accommodates the alkyl/aryl group clearly appears more polar than the region which contains the acyl moiety and is fully surrounded by hydrophobic residues apart from Arg449 which is relatively close to the catalytic

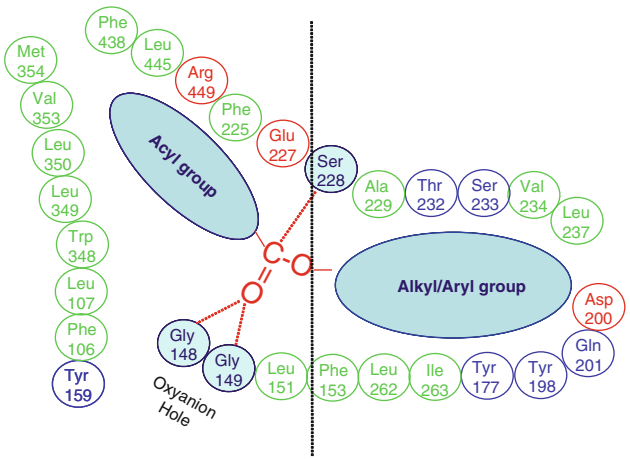

Fig. 3 Two-dimensional scheme illustrating the main residues lining the catalytic pocket. The following color code was used: apolar residues *green*; H-bonding residues *blue* (including the catalytic residue Ser228 and oxyanion hole Gly148 and Gly149); ionized residues *red*

triad. The main difference between the catalytic sites of hCES1 and hCES2 is the size of the two subpockets since the cavity for alkyl/aryl groups in hCES2 is wider than the cavity for acyl moieties, the latter appearing constrained by the deletion of a large and flexible loop (see above) and probably also by the mentioned arginine residue which restrains the cavity by stabilizing strong interactions with the surrounding residues.

Conversely, the subpocket harboring the alkyl/aryl portion appears wider and flanked by a set of conserved polar residues (Tyr177, Asp200, Gln201, Thr232 and Ser233). These observations may explain why hCES2 prefers substrates with less bulky acyl moieties and bulkier alkyl/aryl moieties, and they also suggest that an optimal feature of hCES2 recognition should be an alkyl/aryl moiety relatively more polar than the acyl moiety and allowing interactions with the polar residues. Globally, these considerations underline that the two subpockets of hCES2 conserve polarity profiles similar to those in hCES1, but prefer opposite optimal sizes for acyl and alkyl/aryl moieties, thus explaining the different substrate specificities of the two esterases.

The above considerations find compelling evidences when analyzing the putative enzyme-substrate complexes. Thus, 4-methylumbelliferyl acetate has its labile group conveniently close to Ser228, while the chromen-2-one moiety stabilizes H-bonds with Tyr177 and Ser233 and the acetyl portion elicits hydrophobic contacts with Leu107. Similarly, the aryl group in p-nitrophenyl acetate and p-nitrophenyl butyrate generates H-bonds with Tyr177, while the acyl group contacts Leu107 and Leu349. Again, the aryl portion of aspirin stabilizes H-bonds with Tyr177 and elicits π–π stacking with Tyr177 and His147.

The alkyl groups of the docked benzoyl esters elicit hydrophobic interactions with Trp145 and Ile146, while their aromatic acyl moiety stabilizes π–π stacking with Phe153 and Tyr159 plus a set of apolar contacts with Ile349, Ile350 and Val 353. Again, the benzoyl group of cocaine is suitably close to Ser228 and interacts also with Tyr159 and Ser108 while its tropane ring stabilizes H-bonds with Tyr177 and Ser233 and its methyl ester moiety elicits apolar contacts with Ile146 and Val234. Similarly, the benzoyl group of procaine approaches Phe153 and the para-amino function interacts with Glu227, while the aminoalkyl portion generates H-bonds with Thr232.

The complexes for pyrethroids are as expected for good hCES2 substrates. As evidenced in Fig. 4a for (αS,1*R*)-*cis*-A3, (a) the chrysanthemoyl moiety elicits a very rich set of hydrophobic interactions with Ala150, Ala229, Leu349, Ile350, Val 353 and Met354; (b) the ester group suitably approaches Ser 228; (c) the nitrile function contacts Phe153; and (d) the methoxy group stabilizes H-bonds with Tyr177, Gln201, and Ser233, while the naphthyl portion

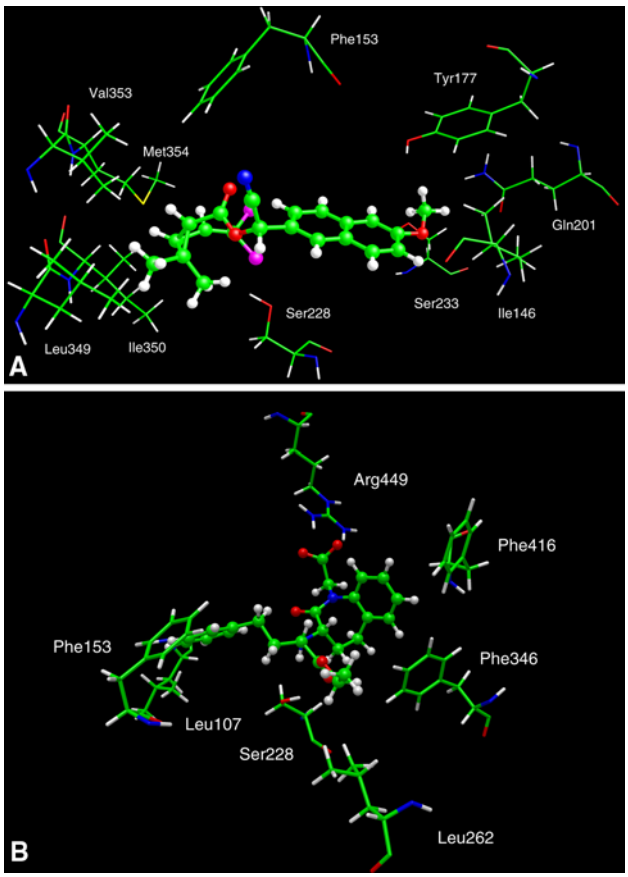

Fig. 4 Main interactions stabilizing the putative complex of hCES1 with the pyrethroid (α*S*,1*R*)-*cis*-A3 (**a**) and benazepril (**b**). One can note in both complexes the catalytic interaction with Ser228. The complex for pyrethroid (**a**) shows the H-bond between the aryl substituents and Ser233 and Tyr177, while the acyl groups mainly elicit apolar contacts. The complex for benazepril (**b**) is clearly stabilized by a rich set of π–π interactions and somewhat biased by the ion-pair with Arg449

generates π–π stacking with His147 and Tyr177 plus a set of apolar contacts with Ile146 and Leu151.

All complexes for ACE inhibitors are characterized by a strong ion-pair between Arg449 and the ligand's carboxylate group. For example, Fig. 4b reports the main residues stabilizing the putative complex for benazepril. Along with the mentioned ionic interactions, the acyl portion realizes a rich set of apolar contacts since the benzodiazepine moiety approaches Phe346, Leu349, Ile350, Pro351 and Phe416, while the phenyl alkyl portion interacts with Phe106, Leu107 and Phe153; the ester function suitably approaches Ser228 and the small alkyl group contacts Ala150 and Leu262. Despite the vast pattern of apolar interactions and the low basicity of their amino group (Table 3), the strong salt bridge that all ACE inhibitors elicit with Arg449 may explain their modest K_m values (Tables 1, 2, 3) since they remain trapped into the catalytic pocket hampering the catalytic turnover.

The complexes of propranolol esters assume a quite different pose since the naphthyl moiety elicits π–π stacking with Trp145 and Tyr159 and the ligand's ether function stabilizes H-bond with Ser108. The 2-aminopropyl group interacts with Tyr177, Phe225 and Val234 and the acyl portion contacts Leu107 and Leu349. The two enantiomers of propranolol butyrate show a quite similar binding mode thus justifying the loss of stereoselectivity as evidenced with hCES1. The ability of exploiting lateral subpockets to accommodate large aryl/alkyl portions is further confirmed by the complexes of irinotecan and its metabolites (NPC and APC) since the camptothecin portion also exploits a quite lateral subcavity where the lactone function stabilizes reinforced H-bonds with Lys308, and the remaining apolar rings generate hydrophobic interactions with Leu259, Leu262, Ile310, Pro311, Leu378, Met415 and Phe416. The bispiperidine moiety contacts a set of apolar residues (Leu107, Trp145, Ile146 and Val234), while the amino group of NPC generates H-bonds with Tyr159. Also, the amino group of the APC metabolite contacts Tyr159, while the carboxy function interacts with Gln109 and the strong polarity of such zwitterionic substructure can well explain the modest K_m value for APC.

A rapid analysis of the docking results as computed for the protonated basic substrates reveal that the resulting complexes are in general quite similar to those obtained for unprotonated substrates, the major difference involving the protonated function which can stabilize strong ion-pairs with the vicinal anionic residues such as Glu105, Asp200, Glu227 and Glu345. Among these residues, a strong salt bridge with Glu227, which can occur when the basic group is significantly close to the labile function, is probably the most detrimental due to significant closeness between Glu227 and Ser228, a fact that reduces the accessibility of the catalytic Ser228 hydroxy group and hampers the correct catalytic mechanism.

The detrimental effect of ion-pairs with the vicinal anionic residues is well illustrated by meperidine, a molecule which is not hydrolyzed by hCES2 but behaves as competitive inhibitor. Docking results for the protonated form evidenced that it is not able to assume a catalytically productive pose. Indeed, the resulting complex appears largely influenced by the ion-pair between the ligand's basic group and Glu228, and is further stabilized by significant π–π stacking involving the ligand's phenyl ring on the one hand and Phe106 plus Phe110 on the other hand. Such a pose, despite involving good binding interactions, is completely biased by ionic interactions; furthermore, the ester function is too distant from Ser228 (8.41 Å), thus explaining why meperidine acts as a competitive inhibitor. This result may also indicate that protonated compounds may be involved in enzyme recognition and are not hydrolyzed when the biasing effect of the ionic interactions dominates.

With a view to investigate the main factors governing CES1/CES2 selectivity, Table 5 collects a small set of investigated substrates ($n = 15$) for which selectivity values ($\Delta pKm_{CES1/CES2}$) were computed from the available hCES1 and hCES2 Km data. Along with the enzymatic parameters, Table 5 also compiles the log P values as well as the fragmental log P and volume values of the alkyl/aryl and acyl moieties used to obtain the Eq. 6.

$$\begin{aligned} \Delta pK_m = {} & 0.231(\pm 0.074)\log P - 2.54 \times 10^{-3} \\ & \left(\pm 1.14 \times 10^{-3}\right)Vol_{acyl} - 8.59 \times 10^{-3} \\ & \left(\pm 1.77 \times 10^{-3}\right)Vol_{alkyl} + 0.376(\pm 0.041) \end{aligned} \quad (6)$$

$$n = 15;\ r^2 = 0.74;\ s = 0.52;\ F = 10.29;\ p < 0.002$$

Such a relation, even if far from having a reliable predictive power due to the small number of considered compounds, is indicative of two trends, namely (a) that CES1 prefers more lipophilic substrates without a clear difference between acyl and alkyl components, while CES2 prefers bulkier substrates; and (b) that the relevance of the alkyl moiety is markedly stronger than that of the acyl moiety, as indicated by their corresponding regression coefficient, thus confirming that CES2 prefers substrates with an alkyl group larger than the acyl one. While the effects of fragmental volumes on CES1/CES2 selectivity are well known and quite expected, the log P effects can be explained remembering that the CES2 enzymatic gorge includes many more water molecules than the CES1 cavity. This renders the CES2 binding site somewhat less apolar than that of CES1 while maintaining a clear preference for lipophilic substrates due mainly to the numerous conserved apolar residues that also line the CES2 enzymatic gorge.

MD simulations

Four molecular dynamics simulations were carried out on hCES2 complexed with the substrate (*R*)-propranolol butyrate and its product (*R*)-propranolol. Such a substrate was chosen for three reasons: (a) it possesses an intermediate affinity thus aptly represents hCES2 substrates; (b) it is the enantiomer with the higher affinity even if the binding stereoselectivity is markedly less pronounced than that observed with hCES1; and (c) it is a basic substrate whose enzymatic product, namely (*R*)-propranolol, contains an amino group, thus allowing the influence of protonation on catalytic turnover to be better investigated.

Both the substrate and the product were simulated separately in their neutral and protonated forms, our objectives being to investigate the stability of both homology hCES2 model and computed complexes, the pathways of product egress, and the influence of protonation on the in situ behavior of ligands. The resulting trajectories were analyzed by monitoring how the distance between Ser228 and the target ester group in (*R*)-propranolol butyrate or the hydroxy function in (*R*)-propranolol varied during the duration of the simulations (10 ns). A preliminary analysis had examined the stability of the protein as assessed by the percentage of residues that remained in the allowed regions of the Ramachandran plot. It was found that the simulations did not undermine the folding stability of the protein since the percentage remained constantly around 75% during all MD runs (plot not shown). Globally, such a result confirms the reliability of the generated hCES2 homology model and suggests that the differences in behavior described below are

Table 5 Substrates and parameters used to rationalize the CES1/CES2 selectivity ($\Delta pKm_{CES1/CES2}$) and to develop Eq. 6

Substrates	pKm CES1	pKm CES2	ΔpKm	Log *P*	Log *P* acyl	Log *P* alkyl	Volume acyl (Å^3)	Volume alkyl (Å^3)
A3Rcis	6.52	6.05	0.47	6.06	3.29	2.77	142.6	188.8
A8	5.82	5.27	0.55	6.02	3.25	2.77	118.9	188.8
Permethrin Cis SR	5.01	5.12	−0.11	8.02	3.29	4.73	142.6	180
Quinapril	3.87	3.91	−0.04	4.75	3.02	1.73	346.9	45.2
p-Nitrophenyl butyrate	3.89	3.88	0.01	5.67	2.82	2.85	77.6	104.4
Dilazep	3.81	4.05	−0.24	5.87	3.54	2.33	386.3	160.2
Benazepril	3.13	3.11	0.02	4.32	2.59	1.73	326.3	45.2
4-Methylumbelliferyl acetate	3.10	3.82	−0.72	0.47	1.09	−0.62	28.9	144.6
Cilazapril	2.89	2.87	0.02	4.78	3.05	1.73	333.5	45.2
Irinotecan	2.84	5.47	−2.63	2.78	0.97	1.81	189	334.3
Monoacetylmorphine	2.08	3.88	−1.80	1.80	1.09	0.71	28.9	258.2
Ethyl benzoate	3.89	2.92	0.96	2.11	1.29	0.82	112.5	53.6
Propyl benzoate	3.49	3.40	0.09	2.71	1.29	1.42	112.5	70.7
Butyl benzoate	3.45	3.96	−0.51	3.07	1.29	1.78	112.5	87.7
Cypermerithrin	4.70	5.05	−0.35	5.76	1.80	3.96	141	195

due to the dynamic response of the enzyme and bound ligands rather than random structural distortions.

Figure 5 shows the distance profile between Ser228 and the two electrical forms of (*R*)-propranolol butyrate. Clearly, the ligand remained close to the catalytic site over the entire simulation time irrespective of the ligand electrical state. This result confirms the stability of the computed complexes and suggests that the behavior of this substrate is not markedly influenced by its ionization state. A careful analysis of these plots unveils that the protonated form shows a greater mobility than the neutral one (standard deviation equal to 1.63 vs. 1.21). This fact may be explained considering that the protonated form is influenced by both the catalytic interaction between the ester function and Ser228 and the biasing ion-pair between the protonated amine group and Glu253; as a result, it repeatedly bounces between these two somewhat different poses. Conversely, the neutral form is not hampered by detrimental salt-bridges and remains in a pose conducive to catalysis for the entire simulation time.

While the two distance profiles for the substrate are comparable enough to suggest a substantial stability of the monitored complexes, the two MD runs for the product (*R*)-propranolol afforded truly contrasting results. Indeed, Fig. 6 clearly shows that the protonated form remained stable within the catalytic cavity, whereas the neutral product moved away from Ser228 and made a progressive exit from the enzymatic cavity. Interestingly, the protonated product showed a very low mobility (standard deviation = 0.78) during the simulation time, a result easily explained considering that protonated (*R*)-propranolol was no longer influenced by the interaction with Ser228 and remained stably tethered by the ionic bond with Glu253. These MD runs clearly indicate that a protonated ligand can remain trapped in the catalytic site and behaves as an

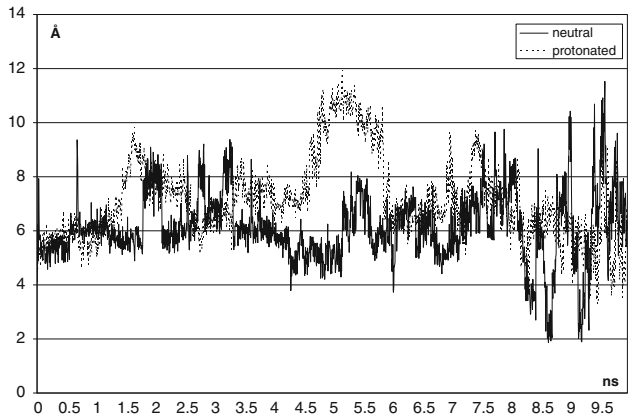

Fig. 5 The dynamic profile for the distance between Ser228 and the ester group of propranolol butyrate as obtained in 10 ns MD simulations (*black line* neutral substrate; *dashed line* protonated substrate)

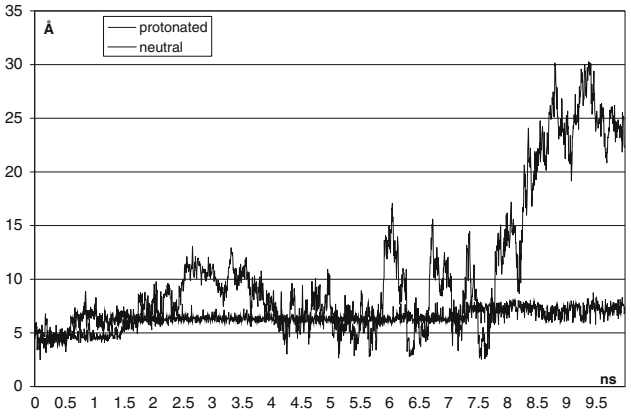

Fig. 6 The dynamic profile for the distance between Ser228 and the hydroxy group of propranolol as obtained in 10 ns MD simulations (*black line* neutral product; *dashed line* protonated product)

inhibitor by forming ionic bonds with some acidic residue(s). One can assume such an entrapment to be particularly strong when the basic group is borne by the alkyl/aryl moiety as in (*R*)-propranolol, whereas the effects of a positive charge in the acyl moiety can be partially counteracted by the free carboxylate function liberated by the hydrolytic reaction.

Conversely, the MD simulation for neutral (*R*)-propranolol shows that the non-protonated enzymatic product can freely move away from the catalytic site allowing for an efficient catalytic turnover. This result deserves further discussion. First, it confirms that MD simulations are able to successfully discriminate between substrates and products even when the structural difference between the two is modest. More generally, one can argue that such simulations are also able to discriminate between substrates (which remain close to Ser228) and non-substrates (which leave the catalytic cavity), providing a useful even if time-demanding tool to investigate at atomic level the metabolic fate of new medicinal esters.

A comparison between different ligands and different enzymes, for example the MD runs for hCES2 complexed with neutral (*R*)-propranolol and those of hCES1 complexed with neutral morphine [16], may even suggest that the product egress in hCES2 is slower than in hCES1. Indeed, product exit in hCES1 was practically complete after 5 ns, whereas the simulation with hCES2 had to be extended to 10 ns to observe the exit of (*R*)-propranolol from the catalytic cavity. The different velocities of product egress can impact on catalytic efficiency as confirmed by the kinetic parameters for hydrolysis of p-aminobenzoate esters by hCES1 and hCES2, with V_{max} values (expressed as nmol/min/mg protein) which were lower in hCES2 than in hCES1 for all considered substrates and irrespective of affinity values [40]. Many other factors might also account partly for such difference in kinetic data

(e.g., the velocity of the release of the acyl-enzyme intermediate), but it seems to us that it may be explained also by a slowing down effect caused by the greater abundance of water molecules in the hCES2 cavity (see the "Introduction"), an abundance which hampers product exit by stabilizing a rich network of H-bonds. Furthermore, one can suppose that the presence of water solvent may also slow down the substrate ingress thus markedly influencing all ligand movements.

Finally, the simulation with neutral (*R*)-propranolol affords a clear picture of the pathway covered by the metabolite during its exit, allowing the recognition of hCES2 residues which facilitate product exit. As observed with the hCES1 isozyme, the hCES2 product pathway is largely surrounded by apolar side-chains (Leu259, Leu262, Ile263, Trp348, Leu349, and Ile350) with very few polar residues to enter transient H-bonds with the ligand; only when the product reaches the hCES2 surface does it approach some ionized residues which have the twofold role of attracting the product and promoting its exit. Even supposing that the monitored exit pathway is somewhat specific for (*R*)-propranolol, the key residues involved in promoting product egress might even be of value in designing non-competitive inhibitors which bind far from the catalytic center.

Conclusion

In this study, a reliable homology model for human CES2 was generated and used to carry out both docking analyses on a dataset of 40 heterogeneous substrates, and MD simulations on some of the complexes. The study has two main objectives along with the validation of the new parallelized tool GriDock. First, it continues our ongoing project aimed at developing computational approaches to predict and rationalize the metabolic hydrolysis of novel esters by human esterases. Such a project, which has so far involved the most important human carboxylesterases (i.e., hCES1 and hCES2), is planned to continue with in silico investigations of the hydrolytic activity of human cholinesterase in order to complete a suitable toolkit of computational methods able to predict the stability and metabolic fate of new medicinal esters in large molecular datasets. Second, the study represents a further validation of the MLP_{InS} score which, by allowing hydrophobic interactions to be conveniently accounted for, affords a straightforward algorithm whose interest is confirmed by its presence in all correlations. Considering the potential applications of the MLP_{InS} score and in order to foster its wide utilization, an on-line web resource was developed (http://nova.colombo58.unimi.it/score.htm) with which the users can freely compute such score function for a given complex without installing any program, also testing several distance functions.

Taken globally, the described correlations, the analyzed complexes and the MD simulations underline the key role of hydrophobic interactions especially with the acyl moiety of substrates. Moreover, the simulations afford evidence that hCES2 preferentially recognizes basic substrates in their neutral form. Indeed, (a) docking results revealed that the binding modes and binding stability of ligands with positively charged groups are markedly affected by electrostatic interactions, so much so that they do not reflect the differences in reactivity of the substrates, as evidenced by poor correlations for protonated substrates; and (b) MD simulations show that the protonated products cannot leave the catalytic cavity and must behave as inhibitors. In general, these results suggest that ligands which interact too strongly by ionic bonds, as also observed for the ion-pair between Arg449 and the carboxylate group of ACE inhibitors, cannot be good CES2 substrates because they are trapped in the cavity in unproductive modes and behave as inhibitors. This observation is in line with the promiscuity of the enzyme, which is known to interact with structurally diverse substrates, a fact consistent with the present rationalization that a variety of hydrophobic interactions plus some weak polar interactions are enough to make a compound behave as a good hCES2 substrate.

References

1. Vistoli G, Pedretti A, Testa B (2008) Assessing drug-likeness—what are we missing? Drug Discov Today 13:285–294
2. Krämer SD, Testa B (2008) The biochemistry of drug metabolism—an introduction. Part 6: Inter-individual factors affecting drug metabolism. Chem Biodivers 5:2465–2578
3. Krämer SD, Testa B (2009) The biochemistry of drug metabolism—an introduction. Part 7: Intra-individual factors affecting drug metabolism. Chem Biodivers 6:1477–1660
4. Testa B, Krämer SD (2006) The biochemistry of drug metabolism—an introduction. Part 1: Principles and overview. Chem Biodivers 3:1053–1101
5. Czodrowski P, Kriegl JM, Scheuerer S, Fox T (2009) Computational approaches to predict drug metabolism. Exp Opin Drug Metab Toxicol 5:15–27
6. Testa B, Krämer SD (2007) The biochemistry of drug metabolism–an introduction. Part 2: Redox reactions and their enzymes. Chem Biodivers 4:257–405
7. Stjernschantz E, Vermeulen NP, Oostenbrink C (2008) Computational prediction of drug binding and rationalisation of selectivity towards cytochromes P450. Exp Opin Drug Metab Toxicol 4:513–527
8. Edmondson DE, Mattevi A, Binda C, Li M, Hubálek F (2004) Structure and mechanism of monoamine oxidase. Curr Med Chem 11:1983–1993
9. Carballeira JD, Quezada MA, Alvarez E, Sinisterra JV (2004) High throughput screening and QSAR-3D/CoMFA: useful tools to design predictive models of substrate specificity for biocatalysts. Molecules 9:673–693

10. Testa B, Krämer SD (2008) The biochemistry of drug metabolism—an introduction. Part 4: Reactions of conjugation and their enzymes. Chem Biodivers 5:2171–2336
11. Smith PA, Sorich MJ, Low LS, McKinnon RA, Miners JO (2004) Towards integrated ADME prediction: past, present and future directions for modelling metabolism by UDP-glucuronosyltransferases. J Mol Graph Model 22:507–517
12. Najmanovich RJ, Allali-Hassani A, Morris RJ, Dombrovsky L, Pan PW, Vedadi M, Plotnikov AN, Edwards A, Arrowsmith C, Thornton JM (2007) Analysis of binding site similarity, small-molecule similarity and experimental binding profiles in the human cytosolic sulfotransferase family. Bioinformatics 23:104–109
13. Sipilä J, Taskinen J (2004) CoMFA modeling of human catechol *O*-methyltransferase enzyme kinetics. J Chem Inf Comput Sci 44:97–104
14. Soffers AE, Ploemen JH, Moonen MJ, Wobbes T, van Ommen B, Vervoort J, van Bladeren PJ, Rietjens IM (1996) Regioselectivity and quantitative structure-activity relationships for the conjugation of a series of fluoronitrobenzenes by purified glutathione S-transferase enzymes from rat and man. Chem Res Toxicol 9:638–646
15. Vistoli G, Pedretti A, Mazzolari A, Bolchi C, Testa B (2009) Influence of ionization state on the activation of temocapril by hCES1: a molecular-dynamics study. Chem Biodivers 6:2092–2100
16. Vistoli G, Pedretti A, Mazzolari A, Testa B (2010) In silico prediction of human carboxylesterase-1 (hCES1) metabolism combining docking analyses and MD simulations. Bioorg Med Chem 18:320–329
17. Testa B, Mayer JM (2003) Hydrolysis in drug and prodrug metabolism–Chemistry, Biochemistry and Enzymology. Wiley-VCH, Weinheim
18. Hatfield JM, Wierdl M, Wadkins RM, Potter PM (2008) Modifications of human carboxylesterase for improved prodrug activation. Exp Opin Drug Metab Toxicol 4:153–165
19. Ettmayer P, Amidon GL, Clement B, Testa B (2004) Lessons learned from marketed and investigational prodrugs. J Med Chem 47:2393–2404
20. Testa B (2009) Prodrugs: bridging pharmacodynamic/pharmacokinetic gaps. Curr Opin Chem Biol 13:338–344
21. Testa B, Krämer SD (2009) The biochemistry of drug metabolism—an introduction. Part 5. Metabolism and bioactivity. Chem Biodivers 6:591–684
22. Imai T (2006) Human carboxylesterase isozymes: catalytic properties and rational drug design. Drug Metab Pharmacokinet 21:173–185
23. Hosokawa M (2008) Structure and catalytic properties of carboxylesterase isozymes involved in metabolic activation of prodrugs. Molecules 13:412–431
24. Sanghani SP, Sanghani PC, Schiel MA, Bosron WF (2009) Human carboxylesterases; an update on CES1, CES2 and CES3. Protein Pept Lett 16:1207–1214
25. Satoh T, Hosokawa M (1998) The mammalian carboxylesterases: from molecules to functions. Annu Rev Pharmacol Toxicol 38:257–288
26. Yang D, Pearce RE, Wang X, Gaedigk R, Wan YJ, Yan B (2009) Human carboxylesterases HCE1 and HCE2: ontogenic expression, inter-individual variability and differential hydrolysis of oseltamivir, aspirin, deltamethrin and permethrin. Biochem Pharmacol 77:238–247
27. Testa B, Krämer SD (2007) The biochemistry of drug metabolism—an introduction. Part 3: Reactions of hydrolysis and their enzymes. Chem Biodivers 4:2031–2122
28. Satoh T, Hosokawa M (2006) Structure, function and regulation of carboxylesterases. Chem Biol Interact 162:195–211
29. Pindel EV, Kedishvili NY, Abraham TL, Brzezinski MR, Zhang J, Dean RA, Bosron WF (1997) Purification and cloning of a broad substrate specificity human liver carboxylesterase that catalyzes the hydrolysis of cocaine and heroin. J Biol Chem 272:14769–14775
30. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and and empirical binding free energy function. J Comput Chem 19:1639–1662
31. Donald J, Martonosi M (2006) An efficient, practical parallelization methodology for multicore architecture simulation. IEEE Comput Architect Lett 5:2
32. Pedretti A, Villa L, Vistoli G (2002) VEGA: a versatile program to convert, handle and visualize molecular structure on Windows-based PCs. J Mol Graph Model 21:47–49
33. Gropp W, Thakur R (2007) Thread safety in an MPI Implementation: requirements and analysis. Parallel Comput 9:595–604
34. Dyrløv Bendtsen J, Nielsen H, Von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795
35. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL Workspace: a web-based environment for protein structure homology modeling. Bioinformatics 22:195–201
36. Nishi K, Huang H, Kamita SG, Kim IH, Morisseau C, Hammock BD (2006) Characterization of pyrethroid hydrolysis by the human liver carboxylesterases hCE-1 and hCE-2. Arch Biochem Biophys 445:115–123
37. Ross MK, Borazjani A, Edwards CC, Potter PM (2006) Hydrolytic metabolism of pyrethroids by human and other mammalian carboxylesterases. Biochem Pharmacol 71:657–669
38. Takai S, Matsuda A, Usami Y, Adachi T, Sugiyama T, Katagiri Y, Tatematsu M, Hirano K (1997) Hydrolytic profile for ester- or amide-linkage by carboxylesterases pI 5.3 and 4.5 from human liver. Biol Pharm Bull 20:869–873
39. Taketani M, Shii M, Ohura K, Ninomiya S, Imai T (2007) Carboxylesterase in the liver and small intestine of experimental animals and human. Life Sci 81:924–932
40. Imai T, Taketani M, Shii M, Hosokawa M, Chiba K (2006) Substrate specificity of carboxylesterase isozymes and their contribution to hydrolase activity in human liver and small intestine. Drug Metab Dispos 34:1734–1741
41. Sanghani SP, Quinney SK, Fredenburg TB, Davis WI, Murry DJ, Bosron WF (2004) Hydrolysis of irinotecan and its oxidative metabolites, 7-ethyl-10-[4-N-(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin and 7-ethyl-10-[4-(1-piperidino)-1-amino]-carbonyloxycamptothecin, by human carboxylesterases CES1A1, CES2, and a newly expressed carboxylesterase isoenzyme, CES3. Drug Metab Dispos 32:505–511
42. Godden JW, Xue L, Bajorath J (2000) Classification of biologically active compounds by median partitioning. J Chem Inf Comput Sci 40:163–166
43. Testa B, Carrupt PA, Gaillard P, Billois F, Weber P (1996) Lipophilicity in molecular modeling. Pharm Res 13:335–343
44. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 16:1781–1802
45. Kawabata T (2003) MATRAS: a program for protein 3D structure comparison. Nucleic Acids Res 31:3367–3369
46. Bencharit S, Edwards CC, Morton CL, Howard-Williams EL, Kuhn P, Potter PM, Redinbo MR (2006) Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1. J Mol Biol 363:201–214
47. Streit TM, Borazjani A, Lentz SE, Wierdl M, Potter PM, Gwaltney SR, Ross MK (2008) Evaluation of the 'side door' in carboxylesterase-mediated catalysis and inhibition. Biol Chem 389:149–162

48. Wierdl M, Morton CL, Nguyen NK, Redinbo MR, Potter PM (2004) Molecular modeling of CPT-11 metabolism by carboxylesterases (CEs): use of pnb CE as a model. Biochemistry 43:1874–1882

49. Eckert F, Klamt A (2006) Accurate prediction of basicity in aqueous solution with COSMO-RS. J Comput Chem 27:11–19