Evaluation of drug–human serum albumin binding interactions with support vector machine aided online automated docking

Ferenc Zsila[1], Zsolt Bikadi[2], David Malik[3], Peter Hari[3], Imre Pechan[4], Attila Berces[5] and Eszter Hazai[2],*

[1]Department of Molecular Pharmacology, Institute of Biomolecular Chemistry, Chemical Research Center, H-1025 Budapest, Pusztaszeri út 59-67., [2]Virtua Drug, Ltd., H-1015 Budapest, Csalogány st. 4C, [3]Delta Services, Ltd., H-1033 Budapest, Szentendrei st. 39 – 53, [4]Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest and [5]Omixon Zrt, Szuk utca 9, H-9082 Nyul, Hungary

ABSTRACT

Motivation: Human serum albumin (HSA), the most abundant plasma protein is well known for its extraordinary binding capacity for both endogenous and exogenous substances, including a wide range of drugs. Interaction with the two principal binding sites of HSA in subdomain IIA (site 1) and in subdomain IIIA (site 2) controls the free, active concentration of a drug, provides a reservoir for a long duration of action and ultimately affects the ADME (absorption, distribution, metabolism, and excretion) profile. Due to the continuous demand to investigate HSA binding properties of novel drugs, drug candidates and drug-like compounds, a support vector machine (SVM) model was developed that efficiently predicts albumin binding. Our SVM model was integrated to a free, web-based prediction platform (http://albumin.althotas.com). Automated molecular docking calculations for prediction of complex geometry are also integrated into the web service. The platform enables the users (i) to predict if albumin binds the query ligand, (ii) to determine the probable ligand binding site (site 1 or site 2), (iii) to select the albumin X-ray structure which is complexed with the most similar ligand and (iv) to calculate complex geometry using molecular docking calculations. Our SVM model and the potential offered by the combined use of *in silico* calculation methods and experimental binding data is illustrated.

Contact: eszter.hazai@virtuadrug.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Human serum albumin (HSA), a highly soluble negatively charged protein, greatly augments the transport capacity of blood plasma where it is present at a high concentration (∼0.6 mM). Its effect is exerted by reversible binding of a vast array of chemically diverse exogenous and endogenous compounds (Carter and Ho, 1994; Curry, 2009; Kratochwil *et al.*, 2002). This astonishing binding capacity, which often seriously impacts pharmacokinetic properties of therapeutic drugs is encoded in the secondary structure of HSA.

Namely, the protein is composed of ∼70:30% α-helix:random coils with no β-sheets, which gives the protein a high degree of conformational flexibility (Carter and Ho, 1994). The 66.5 kDa single polypeptide chain of HSA consists of 585 amino acids and the individual helices are connected by 17 disulfide bonds to form 9 structural loops. Each 190-residue domain, labeled I, II and III from the N terminus, contains 10 helices connected by turns and are structurally very similar to one another (Carter and Ho, 1994). The three homologous domains can be further divided into six-helix and four-helix subdomains termed A and B, which have a common four antiparallel α-helix core. Competitive displacement studies of Sudlow *et al.* using fluorescent molecular probes enabled the identification of two specific drug binding sites termed in the literature as site 1 (subdomain IIA) and site 2 (subdomain IIIA) (Sudlow *et al.*, 1975, 1976). Typically, HSA ligands are accommodated primarily to one of the two high-affinity sites with typical binding association constants in the range of $10^4 – 10^6$ M^{-1} (Carter and Ho, 1994; Kratochwil *et al.*, 2002). It became evident from the experimentally resolved complex structures (Supplementary Table S1) that these binding regions are located in subdomain IIA and IIIA. Structural evaluation of compounds bound specifically to these HSA sites revealed that site 1 ligands appear to be bulky heterocyclic compounds with a negative, often delocalized charge near to the center of a largely nonpolar molecule (Supplementary Table S1) (Buttar *et al.*, 2010; Ghuman *et al.*, 2005; Petitpas *et al.*, 2001; Ryan *et al.*, 2010; Yang *et al.*, 2007). On the other hand, site 2 (also called the indole-benzodiazepine site) preferably accommodates stick-like aromatic carboxylic acids with a negative charge located at the alpha carbon, distant from the hydrophobic region of the molecule (Supplementary Table S1; Ghuman *et al.*, 2005). However, these structural features are not strict prerequisites for site 1 and site 2 binding since numerous ligands are known to bind to both drug binding sites though with different affinities [e.g. L-thyroxine (Petitpas *et al.*, 2003), indoxyl sulphate (Ghuman *et al.*, 2005), dansyl-L-asparagine (Ryan *et al.*, 2010), or ibuprofen (Ghuman *et al.*, 2005)].

High resolution crystal structures of HSA complexed with various ligands resolved stereochemical details of its preformed drug binding sites and enables the understanding of their binding specificity (Fig. 1). Site 1 is a large, flexible, multi-chamber cavity within the core of subdomain IIA that comprises all six helices of the subdomain and additional residues from subdomains IB, IIB

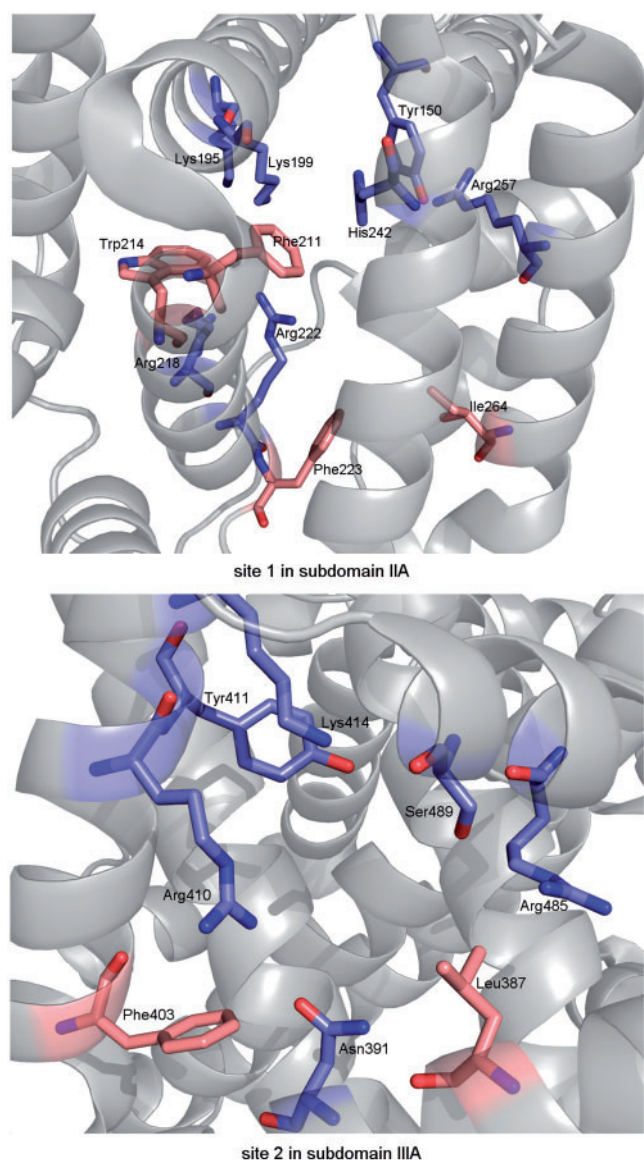site 1 in subdomain IIA

site 2 in subdomain IIIA

Fig. 1. (**A**) Representation of polar side-chains of the drug binding site 1 of HSA in subdomain IIA being involved in forming ionic/H-bonding interactions with ligand molecules (prepared using the 2BXD pdb file of the warfarin–HSA complex). Side-chain atoms are colored by atom type (carbon, purple or pink; nitrogen, blue; oxygen, red). Note the two hydrophilic clusters of residues around the entrance (Lys195, Lys199, Arg218, Arg222) and at the middle/end (Tyr150, His242, Arg257) of the binding pocket. Some apolar residues are also shown. (**B**) Polar key residues interacting with ligands bound at drug binding site 2 of HSA in subdomain IIIA (prepared using the 2BXG pdb file of the ibuprofen–HSA complex).

and IIIA (Ghuman *et al.*, 2005; Petitpas *et al.*, 2001; Ryan *et al.*, 2010) (Fig. 1). Residues 148–153 from subdomain IB close off the top of the site which has an extension toward the interdomain cleft stabilized by residues from subdomain IIB (Val343, Leu347) and IIIA (Leu481), respectively. The binding pocket can be accessed through an open entrance of ∼10 Å in diameter between helices h1 and h2, which faces subdomain IIIA and thus is partially shielded

from bulk solvent. The opening is surrounded by four basic residues (Lys195, Lys199, Arg218, Arg222). Although the interior of the binding crevice is mainly apolar, there is an additional cluster of polar residues toward the middle/end of the pocket (Tyr150, His242, Arg257). The main volume of the pocket is separated by Ile264 into two hydrophobic sub-chambers. Just inside the hydrophilic entrance, there is a third chamber formed by Phe211, Trp214, Ala215, Arg218 and Leu219. Trp214 is an important structural element of site 1 since rotation of its indole ring enables certain ligands to be accommodated at the binding site and to form π–π stacking interactions (Buttar *et al.*, 2010; Ghuman *et al.*, 2005). The large central compartment of site 1 is typically occupied by planar aromatic/heterocyclic rings of the ligands, which are inserted snugly between the 'ceiling' and 'floor' of the pocket represented by Ala291 and Leu219/238, respectively. The enclosure of the guest molecules within the pocket is driven by formation of multiple hydrophobic contacts with nonpolar site residues. On the other hand, acidic or electronegative peripherial groups of the ligands are oriented in the sub-chamber formed around Trp214, or toward the solvent accessible mouth of the cavity where they can form intermolecular H-bonds with the basic residues (Ghuman *et al.*, 2005; Petitpas *et al.*, 2001). The additional basic patch on the middle/opposite side of the pocket (Tyr150, His242, Arg257) also enables H-bond formation with ligand substituents. X-ray data of compounds crystallized with HSA show in most cases H-bonding with the phenolic –OH of Tyr150 suggesting the important contribution of this residue in ligand binding at site 1 (Buttar *et al.*, 2010; Ghuman *et al.*, 2005; Petitpas *et al.*, 2001; Ryan *et al.*, 2010). Appropriately sized molecules possessing two anionic functions on opposite ends such as L-thyroxine (Petitpas *et al.*, 2003) and the uremic toxin 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid (CMPF; Ghuman *et al.*, 2005) can form multiple H-bonds at both sides of the pocket resulting high-affinity binding. The site 1 has ample binding room so that most compounds examined to date by X-ray method do not fill it completely. Due to this fact, site 1 is able to accommodate multiple ligands simultaneously as has been demonstrated by crystal structures of HSA with co-binding pairs of azapropazone-phenylbutazone, azapropazone-indomethacin and indomethacin-myristic acid (Ghuman *et al.*, 2005). Fatty acid (FA) binding induced conformational changes of HSA alter the structure of the site 1 in the sense that it diverts Tyr 150 and, to a lesser extent, Arg 257 out of the pocket. The carboxylate head of FAs bound in the long and narrow channel located between subdomains IA and IIA (FA site 2; Simard *et al.*, 2006) forms H-bonds with Tyr150, Arg257 and Ser287 so these residues will be no longer available to interact with site 1 ligands (Ghuman *et al.*, 2005). Except for oxyphenbutazone, however, these changes have only a minor impact on the location and orientation of the bound ligands. The site 1 itself can also anchor a fatty acid molecule but due to its low-affinity binding drug ligands of the site displace the lipid molecule.

Site 2 in subdomain IIIA (indole-benzodiazepin site) is topologically similar to site 1 in a sense that both are composed of six-helices of the corresponding subdomains, which are arranged to form largely apolar cavities with defined polar features. In contrast to site 1, the entrance to site 2 is exposed to solvent and its inner cavity is smaller and more rigid which features account for the pronounced stereoselectivity observed for bound compounds (e.g. L-tryptophan shows a 100-fold greater HSA affinity than the D-enantiomer; 2,3 benzodiazepines exhibit conformational selectivity; Ghuman *et al.*,

2005). Furthermore, site 2 is less complex in its structure since only few residues from subdomain IIB contribute to its opening. The interior of the binding pocket is hydrophobic, with a single polar patch near to the entrance centered on Arg410 and Tyr411 but also including Lys414 and Ser489 (Fig. 1). This arrangement explains the ligand specificity of the site featured by mostly extended, stick-like nonsteroidal anti-inflammatory agents (NSAIDs) with a peripheral carboxylic group (e.g. ibuprofen, flurbiprofen, diflunisal). In spite of its relative rigidity, site 2 also possesses a degree of conformational adaptability as shown by its capacity to bind two molecules of long-chain fatty acids simultaneously (Curry, 2003). The aliphatic chain of one fatty acid is accommodated in extended conformation by a long, narrow hydrophobic tunnel and its carboxylic head is involved in ionic (Arg410) as well as H-bonding (Tyr411, Ser489) interactions (FA site 4, high-affinity). The second fatty acid makes salt bridges with Arg348 and Arg485, and a H-bond with Ser342 while its methylene tail occupies the central pocket of site 2 (FA site 3, low-affinity). Accordingly, both fatty acid ligands of subdomain IIIA compete directly with the binding of site 2 specific compounds (Bhattacharya *et al.*, 2000; Curry, 2009; Ghuman *et al.*, 2005; Petitpas *et al.*, 2003). However, under normal physiological conditions only 0.1–2 molecules of FA are bound per a HSA molecule (Curry, 2003) which are distributed primarily among the three high-affinity FA sites in subdomain IA-IIA, IIIA and IIIB (Simard *et al.*, 2006).

Besides the two primary drug binding sites in subdomain IIA and IIIA, HSA possesses some additional ligand binding pockets, which can function as secondary sites for agents that bind preferentially to site 1 or 2 (Supplementary Table S2). Moreover, the hydrophobic, L-shaped cavity in subdomain IB termed here as site 3 is also the primary binding locus of some compounds including hemin (Zunszain *et al.*, 2003), bilirubin (Zunszain *et al.*, 2008), the steroid antibiotic fusidic acid (Zunszain *et al.*, 2008) and a sulphonamide derivative (Buttar *et al.*, 2010). Secondary binding sites of other substances have also been identified in a shallow trench at the interface between subdomain IIA and IIB that overlaps the FA site 6, in subdomain IIIB (oxyphenbutazone, propofol) and in the cleft between domains I and III where iodipamide and L-thyroxine can be bound (Bhattacharya *et al.*, 2000; Ghuman *et al.*, 2005; Ryan *et al.*, 2010). In addition, some molecules owing site 1 as primary binding site have been observed to possess a secondary binding locus at site 2 in subdomain IIIA and vice versa (Ghuman *et al.*, 2005).

Given the importance of HSA, knowledge of the exact binding site of albumin is of potential importance in drug research, e.g. in the investigation of drug–drug interactions. Co-administered drugs may exert a competition for the binding to the same albumin binding site, free fraction of the low-affinity drug is enhanced. Similarly, competition for a certain binding site may occur between drugs and endogenous substances (Tesseromatis and Alevizou, 2008). This competition has a potential clinical consequence (Bird and Carmona, 2008).The availability of a vast amount of experimental binding data, and a number of X-ray structures enable the development of reliable *in silico* prediction models.

In silico prediction methods of protein–ligand interactions are divided into ligand-based methods and structure-based methods. Structure-based methods i.e. molecular docking gives insights into the ligand–protein interactions in atomic detail. Molecular docking methods are capable of predicting interactions with ligands sharing low or no similarity with known ligands. However, the calculation of free enthalpy of the ligand–protein interaction possesses low correlation with experimental data. Therefore, docking calculations alone cannot be used as a tool to define binding ability and the binding site of a ligand.

In ligand-based methods such as quantitative structure-activity relationships (QSAR) and support vector machine (SVM), the potential ligand binding is predicted based on similarity with the known ligand structure and its physicochemical properties. A number of studies have been published that use *in silico* tools in order to predict HSA binding of various ligands (Colmenarejo, 2003). These studies utilize different *in silico* methods such as QSAR (Kaliszan *et al.*, 1992; Colmenarejo *et al.*, 2001; Hall *et al.*, 2003), or SVM (Xue *et al.*, 2004). The drawback of ligand-based methods is that they do not provide information on interactions with the protein at an atomic level. Therefore, in our study, both ligand (SVM) and structure-based (molecular docking) *in silico* methods were combined in order to yield accurate prediction of ligand–HSA binding and complex geometry.

It is important to note that the outcome of the available *in silico* studies cannot be simply utilized by experimental chemists, and therefore it does not sufficiently contribute to aiding experimental work.

Thus, the goal of our work was—besides the development of *in silico* model that accurately predicts HSA binding—to create a web service that enables the *in silico* prediction of albumin binding that could be used by experimental researchers. More specifically the web-based platform utilizing our SVM model enables the users to predict: (i) the capability of HSA to bind the query ligand; (ii) the probable ligand binding site, and, moreover albumin-ligand complex geometry is estimated with molecular docking tools. Several examples are presented here that demonstrate the capability of our method to predict HSA binding.

2 METHODS

2.1 SVM calculations

In the course of our SVM method development, molecules are characterized with numerous descriptors and then are presented in the multidimensional space based on the calculated descriptors. The compounds are classified into three categories based on experimental data: (i) non-ligand; (ii) site 1 ligands; and (iii) site 2 ligands. SVM constructs a hyperplane in this high-dimensional space having the maximum margin between the three classes. In the course of prediction molecular descriptors of a given molecule are calculated and then is classified in its position by this multidimensional space.

A total of 163 small molecules with available experimental data on their HSA binding properties were collected from the literature (Supplementary Tables S3 and S4). The collection contains 62 compounds that bind to HSA binding site 1 (site 1 ligands), 38 site 2 ligands and 63 molecules that exhibit no or low-affinity HSA binding. The structures of the ligands were downloaded from the PubChem Database (http://pubchem.ncbi.nlm.nih.gov). All molecules were subjected to geometry optimization using Molconvert (Chemaxon) software applying Dreiding molecular mechanics force field (Mayo *et al.*, 1990). Gasteiger partial charges were calculated (Gasteiger and Marsili, 1980). The Dragon software (www.talete.mi.it) was used to calculate a total of 3250 molecular descriptors for each molecule. The descriptors were filtered using following preprocessing steps: (i) those with >80% zero values and (ii) those that have too small standard deviation values (3%). Libsvm (LIBSVM: a library for support vector machines; software available at www.csie.ntu.edu.tw/~cjlin/libsvm) was employed as the SVM algorithm in

our work. A radial basic function (RBF) was chosen as the kernel function, where 'r' is the parameter. The other parameters were set to default values.

In the training process two parameters, the regularization parameter 'C' and the kernel width parameter 'g' were optimized by using a grid search approach. The feature selection tool fselect (part of the Libsvm package) was employed to measure the relative importance of the features. Descriptors with a correlation larger than 0.9 were filtered and descriptors with higher F-score were kept for further SVM calculations.

As external validation set 63 molecules were selected. The remaining 100 molecules were randomly divided to training (60%) and test (40%) sets. 100 SVM model built using randomly chosen training set was trained and validated by cross-validation, and the performance of the model was evaluated by the corresponding test set.

Accuracy of the method was calculated by the following equation: Accuracy = (true positive + true negative)/(true positive + false positive + true negative + false negative). The SVM model with the highest accuracy on the test set was selected. The model was validated based on the external dataset.

2.2 Molecular docking

Molecular docking calculations were carried out using Autodock Vina software (Trott and Olson, 2010). Ligand structures were optimized using Dreiding force field (Mayo *et al.*, 1990) in Molconvert program of Chemaxon. Gasteiger partial charges (Gasteiger and Marsili, 1980) were calculated on ligand atoms. X-ray structures of the crystallized small molecule-HSA complexes (Supplementary Tables S1 and S2) were downloaded from the Protein Data Bank (http://www.rcsb.org). Polar hydrogen atoms were added to the protein and Gasteiger partial charges were calculated using Autodock Tools. Water molecules, heteroatoms were removed from the structures. Simulation boxes were centered on the originally crystallized ligands. $20 \times 20 \times 20$ Å simulation box was used in each docking calculations, using an exhaustiveness option of 8 (average accuracy).

2.3 Web service

A free web service was built that is capable of predicting the albumin binding site and geometry of ligands (http://albumin.althotas.com) using the combined SVM docking-based method described here. Autodock Vina software is also integrated to the web service for complex geometry calculation. The service has been built in PHP-MySQL and utilizes several external programs and methods in the workflow. The workflow of the service is summarized in Figure 2. The structure of the query ligand can be uploaded or drawn in using the built-in Chemaxon Marvin Java applet. The platform is also connected to Pubchem database, so drug molecules can be directly downloaded from Pubchem using text search. The structural conversions and 3D geometry optimization by Dreiding method are carried out using the Molconvert software. The 2D and 3D molecular descriptors are calculated using DragonX software. Our built-in SVM model presented here is used to predict the binding site of a given substance. Molecular similarity between the query structure and the X-ray HSA ligands is calculated in order to aid the selection of the appropriate PDB structure to be used in complex geometry calculation. In the final step the complex geometry and binding affinity are predicted by automatic molecular docking calculations using Autodock Vina algorithm (Trott and Olson, 2010). In order to increase the docking speed, a special implementation of Autodock on Fpga was developed and is in the testing phase in this server. This feature is already freely available in our web service, Docking Server (http://www.dockingserver.com) (Bikadi and Hazai, 2009).

3 RESULTS

The goal of the current work was to develop a model that accurately predicts HSA binding site of a ligand and the resulting complex
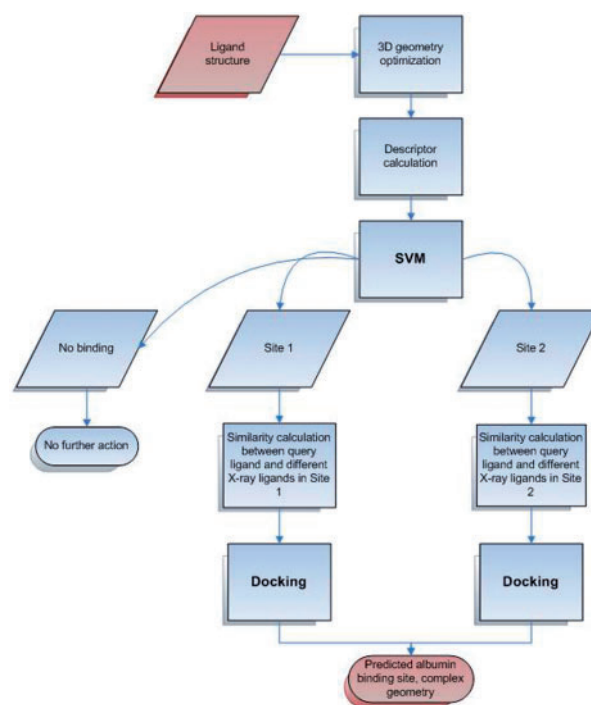

Fig. 2. Workflow of the web service.

geometry. A web service has been established that utilizes the combined method for prediction of HSA binding. A 2D structure of a ligand is used as input and HSA binding characteristics are predicted.

3.1 Ligand-based (SVM) approach for determination of ligand binding to a potential HSA binding site

A number of studies utilize different *in silico* methods for HSA binding prediction (Colmenarejo *et al.*, 2001; Hall *et al.*, 2003; Xue *et al.*, 2004). While both previous and our study are aimed at *in silico* prediction of albumin binding, there is a significant difference between our and the previously published models which makes direct comparisons impossible. Namely, the published models have limited global applicability in contrast to our model because of the followings: (i) The published models do not distinguish between site 1 and site 2 ligands. This is a major drawback of these studies, as valid QSAR calculations should be based on structurally related molecules binding at the same site of the protein. (ii) The models were designed to quantitatively predict the binding affinity of compounds, which bind to HSA, while our method is capable to classify nonalbumin and albumin ligands. (iii) The applied experimental methods (binding constant calculated from the retention time in a HPLC column with immobilized HSA in Colmenarejo, 2001) are not accurate and the experimental conditions (using 15% acetonitril in the eluent, thus ligand binding is investigated in a partly organic buffer) do not reflect the real physiological environment of albumin.

In order to develop a globally applicable model for albumin binding, we collected a unique dataset containing 163 ligands categorized into site 1, site 2 or nonalbumin ligands. In our view, datasets collected from the literature cannot be used for quantitative prediction because of the great influence of experimental conditions on affinity values (Weisiger, 2001). However, for an SVM classification model, the quantitative binding data are unnecessary and the gathered data can be used for building a generally applicable model.

The SVM approach has been shown to be a very effective classification tool in the fields of computational biology and chemistry (Ivanciuc, 2007). In our study, a three states classification method was utilized, where ligands were divided into the following classes: class 0, no or low affinity to HSA; class 1, site 1 ligand; and class 2, site 2 ligands (see Supplementary Tables S3 and S4).

Because of the lack of sufficient amount of experimental binding data on site 3, it was not considered in the SVM model. One of the key points in SVM model building is the numerical representation (descriptor calculation) of the chemical structure. The model performance and the accuracy of the results strongly depend on the descriptor calculation and optimal selection. In the course of our study, 3250 descriptors were calculated for each molecule. After careful selection, 45 descriptors were used in the final model, which are shown in Supplementary Table S5.

Several of the 45 selected descriptors in our model reflect physicochemical properties, which have been previously characterized as a contributing factor to albumin binding. As can be seen, logP (Colmenarejo *et al.*, 2001; Hersey *et al.*, 1991) is one of the most important physicochemical parameter in our model, as several selected descriptors are based on logP and the aromacity of the compounds (Supplementary Table S5). Anti-inflammatory drugs (NSAIDs) are known to bind to HSA with high affinity (Ghuman *et al.*, 2005; Kratochwil *et al.*, 2002; Sudlow *et al.*, 1976; Urien *et al.*, 1984)—and in accordance with these, anti-inflammatory-like index is included in our model. Similarly, carboxylic acid containing compounds bind strongly to HSA, which is reflected in our model by the presence of 'number of carboxylic acids' descriptor. A number of 3D descriptors were also selected reflecting the difference in site 1 and site 2 geometries. The selected model yielded 100% accuracy on training and 78% accuracy on test set. This model was validated on our external dataset and 81% of the ligands were accurately predicted (for detailed results see Table 1). The external dataset included 18 compounds that possessed low or no affinity toward HSA. From these, 15 compounds were accurately predicted, 1 compound was predicted to be site 1 ligand, whereas 2 compounds were predicted as site 2 ligands. Among the 27 site 1 ligands, 22 were accurately predicted, 4 was identified as site 2, whereas 1 as ligand possessing low or no affinity toward HSA. Among the 18 site 2 ligands in the external dataset 14 compounds were accurately predicted, whereas 3 compounds were predicted as site 1 and 1 compound as site 2 ligand. These results indicate that only 5 compounds were misclassified for their capability of binding to HSA.

3.2 Structure-based approach for determination of complex geometry

In the course of our work, molecular docking calculations were used in order to obtain complex geometry at the binding site

previously predicted by our SVM method. As the selection of the most appropriate protein geometry is crucial for accurate complex geometry prediction, the available X-ray structures of ligand–HSA complexes were analyzed in detail in order to determine possible differences at the binding sites. Comparison of X-ray data of ligand-HSA complexes shows some degree of side-chain variability at the binding sites. Depending on the size and chemical constitution of the accommodated ligand molecule, these changes at site 1 involve the rotation of aromatic (Tyr150, Trp214) and basic residues (Lys199, Arg218, Arg222) (Ghuman *et al.*, 2005; Zunszain *et al.*, 2008). Distinctly from other site 1 ligands, accommodation of indomethacin, iodipamide (Ghuman *et al.*, 2005), dansyl-L-phenylalanine (Ryan *et al.*, 2010) and carboxylic indole derivatives (Buttar *et al.*, 2010) occurs in a partially distinct compartment in subdomain IIA that is formed by the rotation of the Trp214 side-chain. In these cases, the indole ring is involved in π–π stacking interaction with the aromatic moiety of the actual ligand molecule. Ligand induced side-chain rearrangements can also be observed at site 2 which binds negatively charged (e.g. ibuprofen, indoxyl sulphate) and neutral compounds (e.g. propofol, diazepam) too (Bhattacharya *et al.*, 2000; Ghuman *et al.*, 2005). Comparison of the binding environment of ibuprofen and diazepam bound at site 2 reveals important ligand-induced side-chain alterations. The carboxylic group of ibuprofen makes salt bridges with Lys414 and Arg410 (PDB id: 2BXG). In contrast, binding of diazepam having no acidic functional group does not require the accommodation of these basic residues and thus they occupy different steric position (PDB id: 2BXF). Moreover, the binding of diazepam induces large rotation of Leu387 and Leu453, which allows the phenyl ring of the drug to access the rear right-hand sub-chamber of the pocket (Ghuman *et al.*, 2005). Thus, it can be concluded that the careful selection of the X-ray structure is crucial for accurate docking calculation.

3.3 Organization of the web service

A web service has been developed which enables the user to predict ligand binding to HSA (http://albumin.althotas.com).

The molecule of interest can be uploaded in pdb, mol, mol2, hin or smiles format or drawn in. Moreover, Pubchem database is directly connected to our platform and can be searched by compound name. After submitting the compound of interest, the followings will be calculated: (i) SVM prediction on binding of the ligand to HSA (no or low-affinity HSA ligand, site 1 or site 2 ligand); (ii) physicochemical parameters of the ligand which were found to play an important role in HSA binding: molecular weight, logP, Ghose–Viswanadhan–Wendoloski anti-inflammatory-like index, number of carboxylic acids, number of substituted phenyl rings in the ligand molecule. The Tanimoto similarity of the molecule of interest with the X-ray ligands is also calculated and presented in a table in order to advice the user to select the appropriate X-ray structure for docking calculations. By clicking on the menus on the left panel molecular docking calculations will be automatically started. The calculated complex geometry, docking energy, binding affinity, calculated interaction surface and literature reference of the X-ray structure actually used are shown. These results aid experimental researchers on analyzing ligand–HSA interaction. However, it should be kept in mind that the standard error of Autodock Vina is 2.85 kcal/mol (Trott and Olson, 2010), which

Table 1. SVM prediction results

Compound name	exp	pr
Training set		
(Orthochlorophenyl-4phenyl)acetic acid	2	2
2,4-Dichlorophenoxyacetic acid	2	2
3-Acetylcoumarin	1	1
3-Carboxy-4-methyl-5-propil-2-furan-propionic acid	1	1
4-Chromanol	1	1
6-*n*-Benzyl-2-thiouracil	1	1
Aciclovir	0	0
Allopurinol	0	0
Amikacin sulfate	0	0
Amphetamine	0	0
Azapropazone	1	1
Bumetanide	1	1
Busulfan	0	0
Cefepime	0	0
Chlordiazepoxide	2	2
Chlorothiazide	1	1
Compound1	1	1
Compound2	1	1
Dansyl-L-proline	2	2
Decitabine	0	0
Diazepam	2	2
Diiodisalicylic acid	1	1
Dopamine	0	0
Ethosuximide	0	0
Felbamate	0	0
Flufenamic acid	2	2
Flurbiprofen	2	2
Galantamine	0	0
Ganciclovir	0	0
Indomethacin	1	1
Indoxyl sulphate	2	2
Iodipamide	1	1
Ioxilan	0	0
Ketamine	0	0
Levocetirizine	2	2
Minoxidil	0	0
Naproxen	2	2
Nicotine	0	0
Oseltamivir	0	0
Phenprocoumon	1	1
Phenylbutazone	1	1
Propofol	2	2
Rizatriptan	0	0
Ropinirole	0	0
Salicylazosulphapyridine	1	1
Sulfamethoxypyridazine	1	1
Sulindac sulfide	2	2
Sulphafurazol	1	1
Sulphamethizole	1	1
Sumatriptan	0	0
Thyroxine	2	2
Tolbutamide	1	1
Tramadol	0	0
Tranexamic acid	0	0

Compound name	exp	pr
Tryptophan	2	2
Valproic acid	1	1
Zanamivir	0	0
Zidovudine	0	0
Zolmitriptan	0	0
Test set		
5-Fluoroacil	0	0
Acetylsalicylic acid	1	1
Benoxaprofen	2	2
Bethanidine	0	0
Captopril	0	0
Chlorpropamide	1	1
Chromone-2-carboxylic acid	1	1
Cytarabine	0	0
Daidzein	1	1
Dansyl-L-sarcosine	2	2
Diflunisal	2	2
Erythromycylamine	0	0
Flucloxacillin	1	1
Fluconazole	0	0
Gemcitabine	0	0
Genistein	1	1
Guanethidine	0	0
Iopanoic acid	1	1
Levetiracetam	0	0
Meptazinol	0	0
Methimazole	0	0
Morphine	0	0
Oxazepam	2	2
Paracetamol	0	0
Pentostatin	0	0
Pramipexole	0	0
Pranofen	2	2
Qercetin	1	1
Salicylosalicylic acid	1	1
Thiotepa	0	0
Tolazamide	1	1
....		
Azidocillin	1	0
Dansyl-L-arginine	1	0
Glafenic acid	1	2
Ketoprofen	2	1
Niflumic acid	1	2
Propiomazine	2	1
Thyroxine	1	2
Topotecan	0	1
Methyldopa	0	2
External set		
(Orthochlorophenyl-4phenyl)hydroxyacetic acid	2	2
4-Hydroxycoumarin	1	1
6-Mercaptopurine	0	0
Amisulpride	0	0
Argatroban	0	0
Atovaquone	1	1
Baicalein	1	1

Compound name	exp	pr
Bleomycin	0	0
Carprofen	2	2
Chlorazepate	2	2
Chloro-2' *p*-biphenylcarboxylic acid	2	2
Cyclophosphamide	0	0
Dansyl-L-norvaline	2	2
Dansyl-L-phenylalanine	2	2
Diclofenac	2	2
Dicoumarol	1	1
Didanosine	0	0
Fenbufen	2	2
Fludarabine phosphate	0	0
Glibenclamide	1	1
Isoxicam	1	1
Itanoxone	2	2
Lisinopril	0	0
Mefenamic acid	2	2
Meloxicam	1	1
Melphalan	0	0
Metformin	0	0
Nalidixic acid	1	1
Naringenin	1	1
Nelarabine	0	0
Neostigmine	0	0
Ochratoxin A	1	1
Oxyphenbutazone	1	1
p-Biphenylyl-4 oxo-4methyl-2 butyric acid	2	2
Phenytoin	1	1
Piroxicam	1	1
R-warfarin	1	1
Salicylamide	1	1
S-Ibuprofen	2	2
Spiramycin	0	0
Sulfinpyrazone	1	1
Sulphadimethoxine	1	1
Suprofen	2	2
Suprofen methyl ester	1	1
S-warfarin	1	1
Tenoxicam	1	1
Thymoquinone	1	1
Topiramate	0	0
Triiodobenzoic acid	1	1
Venlafaxine	0	0
....		
5-Dimethylaminonaphthalene-1-sulfonamide	1	2
Cefazolin	1	0
Cloxacillin	2	1
Dansyl-L-asparagine	1	2
Dansyl-L-glutamate	1	2
Dicloxacillin	2	1
Probenecid	2	1
Sulphamoxole	1	2
Topotecan	0	1
Clofibrate	2	0
Iohexol	0	2
Ioxaglic acid	0	2

Class 0, non-ligand; 1, site 1 ligand; 2, site 2 ligand; exp, experimental; pred, predicted. Dotted lines separate the correctly classified and misclassified compounds.

means that binding affinities practically cannot be quantitatively predicted using docking calculations. This problem is not limited to Vina software. Namely, a critical assessment study of docking softwares and scoring functions concluded that none of the docking programs or scoring functions were capable of predicting ligand binding affinity (Warren *et al.*, 2006).

3.4 Evaluation of the platform on its ability to predict HSA binding interactions of site 1 and site 2 specific drug molecules

Some pharmaceutical substances were selected whose binding sites have previously been reported to coincide with site 1 or site 2. In order to test the usefulness of the system, predictions were carried out for these molecules (these molecules were part of the external set in our SVM model). Site 1 ligands were all accurately predicted by our SVM model (Supplementary Table S6). Site 2 ligands were all predicted to bind to albumin, and with the exception of probenecid and ketoprofen the binding site was correctly identified (Supplementary Table S7). X-ray structures of HSA used in the docking procedures were chosen by considering the molecular similarity score between X-ray ligands of HSA and the molecules to be docked. The docking results were evaluated in correlation with the experimental binding data by considering the geometry of the docked compound and the nature/number of noncovalent interactions involved in stabilization of the complex.

In accordance with the X-ray data, the H-bonding pattern of compounds docked to site 1 (Supplementary Table S6) highlights the role of the two polar side-chain clusters located at the entrance and at the middle/closed end of the pocket (Buttar *et al.*, 2010; Ghuman *et al.*, 2005; Ryan *et al.*, 2010). The polar hydroxynaphthoquinone moiety of atovaquone is engulfed in the central/back compartment of the cavity where it forms H-bonds while its hydrophobic ring substituents extend along the horizontal axis of the pocket. In full agreement with the X-ray data (Buttar *et al.*, 2010; Ghuman *et al.*, 2005; Ryan *et al.*, 2010), Tyr150 was found to be involved in H-bonding with all molecules docked to site 1 emphasizing its central role in ligand binding at site 1.

In the course of docking calculations performed on site 2, apolar moieties of the ligand molecules were found to be buried in the hydrophobic cleft of subdomain IIIA. Hydrophilic head of the ligands interact with the single main polar patch of the site 2 constituted by Arg410, Tyr411, Lys414 and Ser489 (Supplementary Table S7). In every case, except for diclofenac, there is a H-bonding with the hydroxyl group of Tyr411 showing its equivalent role to Tyr150 at site 1. Similarly to a number of other ligands of HSA (Buttar *et al.*, 2010) displacement data indicated the binding of atovaquone in subdomain IIIA too (Zsila and Fitos, 2010). In the docked complex, hydophobic tail of the molecule is deeply inserted into the apolar void of the site while the oxygen atoms of the hydroxynaphthoquinone ring are positioned at the mouth of the cavity and make extensive H-bonds with the proximal polar residues.

The structure of the NSAID drug oxaprozin is reminiscent to phenylbutazone so at the first sight its primary binding at site 1 would be expected. Displacement studies using marker compounds have demonstrated that oxaprozin binds with a significantly lower affinity than at site 2 (Aubry *et al.*, 1995). Both oxaprozin and phenylbutazone are ionized at physiological pH but they markedly differ in the distribution of the negative charge: in phenylbutazone

the charge is delocalized on the carbonyl oxygens of the pyrazolidine ring, in oxaprozin, however, it is localized at the carboxylic end-group of the molecule. Taking the structural similarity between oxaprozin and diazepam into account, docking calculations were performed by using the X-ray template of the diazepam–HSA complex. The docked oxaprozin molecule fits on the bound position of diazepam projecting its one phenyl ring into the rear right-hand sub-chamber of the pocket formed by rearrangement of Leu387 and Leu453 side-chains (Ghuman *et al.*, 2005) (Supplementary Fig. S1). The carboxylic group can form three H-bonds with the side-chains of Tyr411, Lys414 and Ser489, respectively.

The extensive plasma protein binding of diclofenac can be ascribed to its interaction with HSA. HSA binding of diclofenac is characterized by two classes of sites, a high-affinity ($K_a = 5 \times 10^5$ M^{-1}) and a low-affinity one ($K_a = 0.6 \times 10^5$ M^{-1}) locus (Chamouard *et al.*, 1985). Sign of the induced circular dichroism (CD) bands of diclofenac bound to site 1 exhibits complete reversal when it is transferred to site 2 (Chamouard *et al.*, 1985; Yamasaki *et al.*, 2000) suggesting that chiral conformations of the drug molecule adopted at these sites are in mirror image relation. The induced CD spectrum of diclofenac molecule accommodated in an asymmetric protein binding environment can be accounted for the nonplanar conformation of its diphenylamine chromophore. The results obtained with diclofenac docked into the binding pocket of subdomain IIA and IIIA are in full agreement with the above assumption showing nonplanar, nearly mirror-image conformations of the molecule (Supplementary Fig. S2). At site 1 diclofenac is positioned in the binding room of warfarin and its carboxylic group points toward the entrance of the pocket within H-bonding distance to Lys199, Arg218 and Arg222. Similarly, the carboxylic group also forms H-bonds with basic side-chains near to the entrance of the site 2 cavity (Arg410, Lys414) while the difluorophenyl ring of the drug is inserted into the hydrophobic interior of the binding pocket (data not shown).

4 CONCLUSIONS

The goal of the presented study was to develop an SVM model that accurately predicts HSA binding site of a ligand and the resulting complex geometry using molecular docking calculations. A web service has been developed that utilizes the combined method for prediction of HSA binding. The goal of the web service was to make our combined method available for scientists to aid experimental research. Our SVM method accurately predicted potential albumin binding and the binding site of the ligand in 81% of the cases in the external dataset. Moreover, the users are advised which X-ray structure would potentially yield the most appropriate result based on similarity between the X-ray and query ligands. Autodock Vina implemented in our web-based platform enables to perform minutes scale ligand–HSA docking calculations for researchers who conduct HSA binding experiments. The results can be obtained by this way help to evaluate which kind of interactions (hydrophobic, H-bond, ionic) can act between protein residues and the guest compound docked to a specific HSA binding site. This approach also facilitates our understanding of the conformational features of ligand molecules in HSA-bound state as well as experimental displacement data by comparison of docked ligand–HSA structures with experimentally resolved complexes. In

addition, the accumulating number of X-ray structures of ligand–HSA adducts allows continuous future improvement of the online platform to enhance its efficacy in correlating docking results with experimental binding data.

ACKNOWLEDGEMENTS

REFERENCES

Aubry,A.F. *et al.* (1995) The effect of co-administered drugs on oxaprozin binding to human serum albumin. *J. Pharm. Pharmacol.*, **47**, 937–944.

Bhattacharya,A.A. *et al.* (2000) Binding of the general anesthetics propofol and halothane to human serum albumin. High resolution crystal structures. *J. Biol. Chem.*, **275**, 38731–38738.

Bikadi, Z. and Hazai, E. (2009) Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminform.*, **1**, 15.

Bird, J., and Carmona, C. (2008) Probable interaction between warfarin and torsemide. *Ann. Pharmacother.*, **42**, 1893–1898.

Buttar,D. *et al.* (2010) A combined spectroscopic and crystallographic approach to probing drug-human serum albumin interactions. *Bioorg. Med. Chem.*, **18**, 7486–7496.

Carter,D.C. and Ho,J.X. (1994) Structure of serum albumin. *Adv. Protein Chem.*, **45**, 153–203.

Chamouard,J.M. *et al.* (1985) Diclofenac binding to albumin and lipoproteins in human serum. *Biochem. Pharmacol.*, **34**, 1695–1700.

Colmenarejo,G. (2003) In silico prediction of drug-binding strengths to human serum albumin. *Med. Res. Rev.*, **23**, 275–301.

Colmenarejo,G., *et al.* (2001) Cheminformatic models to predict binding affinities to human serum albumin. *J. Med. Chem.*, **44**, 4370–4378.

Curry,S. (2003) Plasma albumin as a fatty acid carrier. *Adv. Mol. Cell Biol.*, **33**, 29–46.

Curry,S. (2009) Lessons from the crystallographic analysis of small molecule binding to human serum albumin. *Drug Metab. Pharmacokinet.*, **24**, 342–357.

Curry,S. *et al.* (1998) Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nat. Struct. Biol.*, **5**, 827–835.

Gasteiger,J. and Marsili,M. (1980) Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron*, **36**, 3219–3228.

Ghuman,J. *et al.* (2005) Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.*, **353**, 38–52.

Hall,L.M. *et al.* (2003) Modeling drug albumin binding affinity with E-state topological structure representation. *J. Chem. Inf. Comput. Sci.*, **43**, 2120–2128.

Hanai,T. *et al.* (2002) Prediction of human serum albumin-drug binding affinity without albumin. *Anal. Chim. Acta*, **454**, 101–108.

Hersey,A. *et al.* (1991) A quantitative structure-activity relationship approach to the minimization of albumin binding. *J. Pharm. Sci.*, **80**, 333–337.

Ivanciuc,O. (2007) *Applications of support vector machines in chemistry*. John Wiley & Sons, Inc., Weinheim.

Kaliszan,R. *et al.* (1992). Quantitative structure-enantioselective retention relationships for the chromatography of 1,4–benzodiazepines on a human serum albumin based HPLC chiral stationary phase: An approach to the computational prediction of retention and enantioselectivity. *Chromatographia*, **33**, 546–550.

Kratochwil,N.A. *et al.* (2002) Predicting plasma protein binding of drugs: a new approach. *Biochem. Pharmacol.*, **64**, 1355–1374.

Mayo,S.L. *et al.* (1990) DREIDING: a generic force-field for molecular simulations. *J. Phys. Chem.*, **94**, 8897–8909.

Petitpas,I. *et al.* (2001) Crystal structure analysis of warfarin binding to human serum albumin: anatomy of drug site I. *J. Biol. Chem.*, **276**, 22804–22809.

Petitpas,I. *et al.* (2003) Structural basis of albumin-thyroxine interactions and familial dysalbuminemic hyperthyroxinemia. *Proc. Natl Acad. Sci. USA*, **100**, 6440–6445.

Ryan,A.J. *et al.* (2011) Structural basis of binding of fluorescent, site-specific dansylated amino acids to human serum albumin. *J. Struct. Biol.*, **174**, 84–91.

Simard,J.R. *et al.* (2006) Location of high and low affinity fatty acid binding sites on human serum albumin revealed by NMR drug-competition analysis. *J. Mol. Biol.*, **361**, 336–351.

Sudlow,G. *et al.* (1975) The characterization of two specific drug binding sites on human serum albumin. *Mol. Pharmacol.*, **11**, 824–832.

Sudlow,G. *et al.* (1976) Further characterization of specific drug binding sites on human serum albumin. *Mol. Pharmacol.*, **12**, 1052–1061.

Tesseromatis,C. and Alevizou,A. (2008). The role of the protein-binding on the mode of drug action as well the interactions with other drugs. *Eur. J. Drug Metab. Pharmacokinet.*, **33**, 225–230.

Trott,O. and Olson,A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **31**, 455–461.

Urien,S. *et al.* (1984) The binding of aryl carboxylic acid derivatives to human serum albumin - A structure-activity study. *Biochem. Pharmacol.*, **33**, 2283–2289.

Xue,C.X. *et al.* (2004) QSAR models for the prediction of binding affinities to human serum albumin using the heuristic method and a support vector machine. *J. Chem. Inf. Comput. Sci.*, **44**, 1693–1700.

Yamasaki,K. *et al.* (2000) Circular dichroism simulation shows a site-II-to-site-I displacement of human serum albumin-bound diclofenac by ibuprofen. *AAPS PharmSciTech*, **1**, E12.

Yang,F. *et al.* (2007) Effect of human serum albumin on drug metabolism: structural evidence of esterase activity of human serum albumin. *J. Struct. Biol.*, **157**, 348–355.

Warren,C.G. *et al.* (2006) A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, **49**, 5912–5931.

Weisiger,R.A. *et al.* (2001) Affinity of human serum albumin for bilirubin varies with albumin concentration and buffer composition: results of a novel ultrafiltration method. *J. Biol. Chem.*, **276**, 29953–29960.

Zsila,F. and Fitos,I. (2010) Combination of chiroptical, absorption and fluorescence spectroscopic methods reveals multiple, hydrophobicity-driven human serum albumin binding of the antimalarial atovaquone and related hydroxynaphthoquinone compounds. *Org. Biomol. Chem.*, **8**, 4905–4914.

Zunszain,P.A. *et al.* (2003) Crystal structural analysis of human serum albumin complexed with hemin and fatty acid. *BMC Struct. Biol.*, **3**, 6.

Zunszain,P.A. *et al.* (2008) Crystallographic analysis of human serum albumin complexed with 4Z,15E-bilirubin-IXa. *J. Mol. Biol.*, **381**, 394–406.