Structural Determinants for Transport across the Intestinal Bile Acid Transporter Using C-24 Bile Acid Conjugates

Rana Rais,[†] Chayan Acharya,[†] Alexander D. MacKerell, Jr.,* and James E. Polli*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201, United States

Abstract: The human apical sodium dependent bile acid transporter (hASBT) reabsorbs gram quantities of bile acid daily and is a potential prodrug target to increase oral drug absorption. In the absence of a high resolution hASBT crystal structure, 3D-QSAR modeling may prove beneficial in designing prodrug targets to hASBT. The objective was to derive a conformationally sampled pharmacophore 3D-QSAR (CSP-SAR) model for the uptake of bile acid conjugates by hASBT. A series of bile acid conjugates of glutamyl chenodeoxycholate were evaluated in terms of K_m and normalized V_{max} (normV_{max}) using hASBT-MDCK cells. All monoanionic conjugates were potent substrates. Dianions, cations and zwitterions, which bound with a high affinity, were not substrates. CSP-SAR models were derived using structural and physicochemical descriptors, and evaluated via cross validation. The best CSP-SAR model for K_m included two structural and two physiochemical descriptors, where substrate hydrophobicity enhanced affinity. A best CSP-SAR model for K_m/normV_{max} employed one structural and three physicochemical descriptors, also indicating hydrophobicity enhanced efficiency. Overall, the bile acid C-24 region accommodated a range of substituted anilines, provided a single negative charge was present near C-24. In comparing uptake findings to prior inhibition results, increased hydrophobicity enhanced activity, with dianions and zwitterions hindering activity.

Keywords: Bile acid; transporter; substrate; Michaelis−Menten; affinity; capacity; hydrophobicity; structure−activity relationship

Introduction

The human apical sodium dependent bile acid transporter (SLC10A2, hASBT) is the major transporter involved in uptake and transport of bile acids from the intestine. It also plays an important role in the enterohepatic circulation of bile acids, enabling efficient recycling and conservation of bile acids to maintain the bile acid pool in the body.[1−4] The transport of both conjugated and unconjugated bile acids with

over 95% efficacy emphasizes hASBT's tremendous potential as a prodrug target.[5] The design of successful prodrug targets to hASBT is impeded by a lack of understanding of the molecular mechanisms of uptake, and the absence of a high resolution crystal structure. Consequently, there is a need to identify the important structural and functional characteristics that underlie the transport mechanisms for successful prodrug targeting of hASBT. Despite the extensive studies on the functional and molecular properties hASBT and its interac-

---

* Corresponding authors. J.E.P.: University of Maryland, School of Pharmacy, Department of Pharmaceutical Sciences, 20 Penn Street, HSF2 Room 623, Baltimore, MD 21201; tel, 410-706-8292; fax 410-706-5017; e-mail, jpolli@rx.umaryland.edu. A.D.M.: University of Maryland, School of Pharmacy, Department of Pharmaceutical Sciences, 20 Penn Street, HSF2 Room 629, Baltimore, MD 21201; tel, 410-706-7442; fax, 410-706-5017; e-mail, amackere@rx.umaryland.edu.

[†] Contributed equally to the work.

(1) Dawson, P. A.; Lan, T.; Rao, A. Bile acid transporters. *J. Lipid Res.* **2009**, *50*, 2340–57.

(2) St-Pierre, M. V.; Kullak-Ublick, G. A.; Hagenbuch, B.; Meier, P. J. Transport of bile acids in hepatic and non-hepatic tissues. *J. Exp. Biol.* **2001**, *204*, 1673–86.

(3) Hagenbuch, B.; Dawson, P. The sodium bile salt cotransport family SLC10. *Pfluegers Arch.* **2004**, *447*, 566–70.

(4) Alrefai, W. A.; Gill, R. K. Bile acid transporters: structure, function, regulation and pathophysiological implications. *Pharm. Res.* **2007**, *24*, 1803–23.

(5) Balakrishnan, A.; Polli, J. E. Apical sodium dependent bile acid transporter (ASBT, SLC10A2): a potential prodrug target. *Mol. Pharmaceutics* **2006**, *3*, 223–30.

tions with various bile acid analogues, an understanding of the substrate requirements for hASBT is limited. In particular, the influence of the C-24 chemistry space attributes that enhance or impede interaction of bile acid analogues with hASBT are not clear.[6] Notably, no 3D-QSAR uptake model has been developed for hASBT substrates.

Prior to cloning of hASBT, Lack et al. demonstrated that conjugation improved transport efficacy and proposed that a negative charge around the C-24 region is essential for interaction with hASBT.[7] Most studies further established that a single negative charge around C-24 was important for transport activity. For instance, Swaan and colleagues examined the transport of four C-24 bile acid conjugates across Caco-2 monolayers, each providing a negative charge near C-24, of which two conjugates were rapidly transported.[8] Further analysis suggested that the length of the C-24 side chain can be 14 Å or longer for translocation, and that large hydrophobic moieties enhanced binding to hASBT. Baringhaus et al. developed a 3D-QSAR inhibition model with a set of 17 inhibitors of the rabbit ileal Na/bile acid cotransporter. The developed pharmacophore included five essential features: one hydrogen bond donor, one hydrogen bond acceptor and three hydrophobic features.[9] Previously, in our laboratory, native and C-24 modified bile acid conjugates were studied in stably transfected hASBT-MDCK monolayers. These studies suggested that the natural bile acid chenodeoxycholic acid (CDCA) can be coupled to glutamic acid as a linker (denoted glu-CDCA) in order to maintain a single negative charge around C-24, to which various probes can be conjugated.[6,10] Additionally, acyclovir bioavailability was increased by conjugating acyclovir to CDCA via a valine linker at the bile acid's C-24 position, demonstrating negative charge could be replaced by a hydrogen bond accepting group.[8]

The conformationally sampled pharmacophore (CSP) method is a novel approach[11−14] that can be combined with physiochemical parameters to yield 3D-QSAR models (denoted CSP-SAR). CSP is well suited for compound classes with high structural flexibility, such as bile acid conjugates. CSP considers all accessible molecular conformations to ensure that the bioactive conformer of each compound is included for model development. The inhibition requirements of hASBT using anilinyl conjugates of glu-CDCA[15] and aminopiperidine conjugates were studied previously using CSP-SAR.[14] In this study, we focus on delineating the uptake requirements of hASBT using a previously reported series of glu-CDCA conjugates of benzene and aminopiperidine analogues,[14,15] where benzene and aminopiperidine represent drug scaffolds. Transport data in these bile acid analogues were used to build a CSP-SAR model that could be of utility to predict potential prodrugs, where CDCA would be conjugated to the drug via a glutamic acid linker. Overall, this work attempts to elucidate hASBT substrate requirements and derive 3D-QSAR uptake models, which have promised to develop a rational prodrug design strategy to target ASBT.

Experimental Section

Materials. [³H]-Taurocholic acid (10 μCi/mmol) was purchased from American Radiolabeled Chemicals, Inc. (St. Louis, MO). Taurocholate was obtained from Sigma Aldrich (St. Louis, MO). CDCA was obtained from TCI America (Portland, OR). Geneticin, fetal bovine serum (FBS), trypsin, and DMEM were purchased from Invitrogen (Rockville, MD). All other reagents and chemicals were of the highest purity commercially available.

Synthetic Procedures. A series of substituted anilinyl[15] and aminopiperidine conjugates glu-CDCA were previously synthesized.[14] Briefly, almost all conjugates employed coupling α-benzyl glutamic acid to CDCA via either *N*-hydroxysuccinimide (OSU) ester or benzotriazole (OBT) ester. Various substituted probes were then coupled to bile acid via glutamic acid by stirring either at RT or at 60 °C using *O*-benzotriazol-1-yl-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) as the activating agent and triethylamine (TEA) as the base. The resulting neutral compounds were then subjected to hydrogenation in a Parr shaker for 1−2 h in ethanol and 10% palladium to remove the α-benzyl group, yielding the mono- and dianionic targets. Neutral

(6) Balakrishnan, A.; Wring, S. A.; Coop, A.; Polli, J. E. Influence of charge and steric bulk in the C-24 region on the interaction of bile acids with human apical sodium-dependent bile acid transporter. *Mol. Pharmaceutics* **2006**, *3*, 282–92.

(7) Lack, L. Properties and biological significance of the ileal bile salt transport system. *Environ. Health Perspect.* **1979**, *33*, 79–89.

(8) Swaan, P. W.; Szoka, F. C., Jr.; Oie, S. Molecular modeling of the intestinal bile acid carrier: a comparative molecular field analysis study. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 581–8.

(9) Baringhaus, K. H.; Matter, H.; Stengelin, S.; Kramer, W. Substrate specificity of the ileal and the hepatic Na(+)/bile acid cotransporters of the rabbit. II. A reliable 3D QSAR pharmacophore model for the ileal Na(+)/bile acid cotransporter. *J. Lipid Res.* **1999**, *40*, 2158–68.

(10) Balakrishnan, A.; Wring, S. A.; Polli, J. E. Interaction of native bile acids with human apical sodium-dependent bile acid transporter (hASBT): influence of steroidal hydroxylation pattern and C-24 conjugation. *Pharm. Res.* **2006**, *23*, 1451–9.

(11) Bernard, D.; Coop, A.; MacKerell, A. D., Jr. 2D conformationally sampled pharmacophore: a ligand-based pharmacophore to differentiate delta opioid agonists from antagonists. *J. Am. Chem. Soc.* **2003**, *125*, 3101–7.

(12) Bernard, D.; Coop, A.; MacKerell, A. D., Jr. Quantitative conformationally sampled pharmacophore for delta opioid ligands: reevaluation of hydrophobic moieties essential for biological activity. *J. Med. Chem.* **2007**, *50*, 1799–809.

(13) Bernard, D.; Coop, A.; MacKerell, A. D., Jr. Conformationally sampled pharmacophore for peptidic delta opioid ligands. *J. Med. Chem.* **2005**, *48*, 7773–80.

(14) Gonzalez, P. M.; Acharya, C.; Mackerell, A. D., Jr.; Polli, J. E. Inhibition requirements of the human apical sodium-dependent bile acid transporter (hASBT) using aminopiperidine conjugates of glutamyl-bile acids. *Pharm. Res.* **2009**, *26*, 1665–78.

(15) Rais, R.; Acharya, C.; MacKerell, A. D.; Polli, J. E. Molecular switch controlling the binding of anionic bile acid conjugates by hASBT. *J. Med. Chem.* **2010**, *53*, 4749–60.

compound intermediates were purified by column chromatography using a gradient of hexane and ethyl acetate. Final target compounds were obtained as solids after deprotection. Identity and purity were confirmed by TLC, MS, NMR, and elemental analysis.[15]

Cell Culture. Stably transfected hASBT-MDCK cells were cultured as described previously.[10,16] Briefly, cells were grown at 37 °C, 90% relative humidity, and 5% CO_2 atmosphere and fed every 2 days. The media for cell culture comprised DMEM supplemented with 10% FBS, 50 U/mL penicillin, and 50 μg/mL streptomycin. Geneticin at 1 mg/mL was used to maintain selection pressure. Cells were passaged every 4 days or after reaching 90% confluence.

Conjugate Uptake into hASBT-MDCK Cells. Uptake studies of 32 conjugates of glu-CDCA were performed in 12-well cell culture polystyrene plates (Corning Costar; Corning, NY). Of the 32 conjugates, 29 were anilinyl conjugates of glu-CDCA,[15] that were either mono- or dianionic, while three were aminopiperidinyl conjugates of glu-CDCA.[14] Aminopiperidinyl conjugates were either cationic or zwitterionic. The 12-well plates were seeded with hASBT-MDCK cells at a density of 1.5×10^6 cells/well and induced on day 4 with 10 mM sodium butyrate for 12−15 h prior to uptake experiments. Conjugate concentration varied from 0 to 100 μM for most monoanions and up to 5000 μM for dianions. Uptake studies were initiated by washing the cells three times with warm Hanks balanced salt solution (HBSS) [pH 6.8] and exposing cells to donor solution for 10 min at 37 °C and 50 rpm in an orbital shaker. Cells were washed three times with chilled sodium free buffer. Parallel identical studies were performed without sodium, in order to measure passive conjugate uptake. Additionally, taurocholate V_{max} was measured on each occasion as previously described,[14] as taurocholate is a reference hASBT substrate. For each conjugate, conjugate V_{max} was normalized to observed taurocholate V_{max} to yield substrate normalized V_{max}, denoted normV_{max}. normV_{max} accommodates variable hASBT protein expression across occasions.

Quantification of Conjugates by LC/MS/MS. Cells were lysed and subjected to extraction as described previously.[17] Briefly, at the end of uptake experiment, cells were lysed by addition of 300 μL of acetonitrile. Acetonitrile was evaporated at RT over 2−3 h. Extraction of conjugate from cell lysate was performed by adding 1 mL of 50:50 acetonitrile:water spiked with 500 nM internal standard; plates were then sealed with parafilm. After 30 min, extract was placed into silanized vials and stored in −80 °C until analyzed. Recovery experiments using the extraction protocol have been previously described using ^{3}[H]-taurocholate as

a standard substrate for hASBT.[17] Taurocholate uptake studies were performed as a standard control; taurocholate kinetic parameters were determined from LC/MS/MS analysis using the methods described below except detection was achieved under negative ion electrospray tandem MS using the [M − H]$^-$ peaks and compound **1** as the internal standard. The multiple reaction monitoring (MRM) transition used for taurocholate was 514.4>123.66 and run time was 8 min.

Conjugate uptake was quantified by LC/MS/MS using a Finnigan Surveyor HPLC system equipped with an autosampler with a Finnigan TSQ quantum Discovery Max mass spectrometer (Thermo Fisher Scientific Inc., Waltham MA). The column was a Phenomenex Luna C8, 50 × 2 mm, 5 μ column, heated to 40 °C. The flow rate was 0.4 mL/min. The gradient of 5−95% acetonitrile was employed over 0.2−4 min and the overall run time for most conjugates was 6−8 min. The starting mobile phase gradient for most conjugates was 50:50 acetonitrile:water (Table S1a in the Supporting Information) up to 0.2 min, followed by increase in organic phase to 95% up to 3−4 min. It was maintained at 95% organic phase for 1 min and then reduced back to 50% organic phase by 6−8 min. The mobile phase included 0.1% formic acid as a modifier for all monoionic and dianionic conjugates. For zwitterionic and cationic conjugates, 0.1% ammonium hydroxide was used as the modifier. Injection volume was 10 μL. Detection was achieved under positive ion electrospray tandem MS using the [M + H]$^+$ peak, as positive mode provided the greatest sensitivity, except for the standard taurocholate where detection was achieved under negative ion mode. The unsubstituted aniline conjugate, which shares a structure similar to all studied conjugates, was used as the internal standard. MRM transitions for the conjugates are listed in Table S1a in the Supporting Information.

Kinetic Analysis. Conjugate kinetic parameters V_{max}, K_m and P_P were obtained by fitting uptake data to modified the Michaelis−Menten model as previously described using WinNonlin 5.2 (Pharsight, Mountain View, CA).[17] Equation 1 is a modified version of the classical Michaelis−Menten which accounts for the presence of an apical boundary layer; P_{ABL} was 1.5×10^{-4} cm^2/s.[18] Data from sodium free studies were fitted to eq 2, which represents the passive transport.

$$J = \frac{P_{ABL}\left(\dfrac{V\mathrm{max}}{K_m + S} + P_P\right)}{P_{ABL} + \dfrac{V_{max}}{K_m + S} + P_P} S \quad (1)$$

$$J = \frac{P_{ABL} P_P S}{P_{ABL} + P_P} \quad (2)$$

Computational Methods. The twenty-seven anilinyl conjugates of glu-CDCA that were substrates were used to develop CSP-SAR substrate models, individually for K_m and K_m/normV_{max}. None of the aminopiperidine conjugates were substrates, and hence not used in CSP-SAR analysis. K_m reflects substrate affinity for the transporter. K_m values less

(16) Balakrishnan, A.; Sussman, D. J.; Polli, J. E. Development of stably transfected monolayer overexpressing the human apical sodium-dependent bile acid transporter (hASBT). *Pharm. Res.* **2005**, *22*, 1269–80.

(17) Rais, R.; Gonzalez, P. M.; Zheng, X.; Wring, S. A.; Polli, J. E. Method to screen substrates of apical sodium-dependent bile acid transporter. *AAPS J.* **2008**, *10*, 596–605.

than 0.1 μM were assigned a value of 0.1 μM, due to the limited ability to accurately estimate potent K_m values.[18] V_{max} is the transporter capacity for the substrate. normV_{max} is V_{max} normalized against observed taurocholate V_{max} in order to accommodate variable hASBT protein expression across occasions. K_m/normV_{max} describes the reciprocal of the transporter efficiency. Model development for normV_{max} and normV_{max}/K_m were not successful, as described below.

Compound descriptors for CSP-SAR modeling were previously generated.[15] Briefly, *in silico* models of conjugates were built using the program CHARMM[19] with the all-atom CHARMM general force field (CGenFF).[20] Each molecule was subjected to 1000 steps of steepest descent (SD) and 500 steps of Newton−Raphson (NRAP) energy minimization in the gas phase to a gradient of 10^{-4} kcal/mol/Å. Structures minimized using CHARMM were subjected to replica exchange molecular dynamic (MD) simulations involving 20 ns simulations with four replicas of each molecule between 300K and 400K, using an exponential scale (300K, 330K, 363K, and 400K).[21–23] Exchange of replicas was attempted after every 250 MD steps. MD simulations were performed using Langevin dynamics[24] with an integration time step of 0.002 ps, and the aqueous solvation was modeled implicitly via the generalized Born continuum solvent model (GBMV).[25,26] SHAKE was applied to all covalent bonds involving hydrogens.[27] Conformations saved every 20 ps and obtained from all 4 replicas were used for the analysis. The protonation states of the ionizable groups present in the molecules were determined based on the experimental pH of 6.8. All free acids were assumed to be protonated.

As previously reported, several physicochemical descriptors were calculated for each conjugate by averaging over the full trajectories.[15] Those descriptors that manifested here in models are briefly described here. The free energy of solvation of each compound in water (ΔG_w) was calculated using the semiempirical quantum chemical program AM-SOL6.8 by invoking the AM1-SM5.42R continuum solvent model.[28] ΔG_w was calculated by taking the difference between the heat of formation plus solvation energy of the solvated system in its relaxed electronic state and that for the gas-phase system. Average solvent accessible surface area (SASA) and polar surface area (PSA) were calculated using the Lee and Richards method as implemented in CHARMM.[29] The radius of the solvent molecule was taken as 1.4 Å, which approximates the radius of a water molecule.[30] PSA for each compound was calculated by adding the contribution of polar atoms (N, O, F, Cl, Br and hydrogen atoms covalently attached to any of these atoms). For calculation of PSA, halogens were included as they could potentially take part in polar interactions with other polar atoms.[31–34] SASA and PSA were also calculated exclusively for the region beyond C-24 of the bile acid conjugates ($SASA_{side}$ and PSA_{side}) to access the contribution of the substituents alone toward the interaction with hASBT. Surface area terms were obtained as averages over the conformations saved from the MD simulations. Molar refractivity (MR) and flexibility parameter (KierFlex) were calculated using the program MOE.[35] The number of

(18) Balakrishnan, A.; Hussainzada, N.; Gonzalez, P.; Bermejo, M.; Swaan, P. W.; Polli, J. E. Bias in estimation of transporter kinetic parameters from overexpression systems: Interplay of transporter expression level and substrate affinity. *J. Pharmacol. Exp. Ther.* **2007**, *320*, 133–44.

(19) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. V.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–614.

(20) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. CHARMM General Force Field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31* (4), 671–90.

(21) Feig, M.; Karanicolas, J.; Brooks, C. L., 3rd. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377–95.

(22) Garcia, A. E.; Sanbonmatsu, K. Y. Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins* **2001**, *42*, 345–54.

(23) Sugita, Y.; Okamoto, Y. Replica-exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–51.

(24) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon Press: Oxford, 1987; p 385.

(25) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. *J. Comput. Chem.* **2003**, *24*, 1348–56.

(26) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116*, 10606–14.

(27) Ryckaert, J. P.; Ciccotti, G.; Berendsen, J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–41.

(28) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Zhu, T.; Thompson, J. D.; Winget, P. e. a. AMSOL Version 6.8; University of Minnesota: Minneapolis, MN.

(29) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379–400.

(30) Chothia, C. Structural invariants in protein folding. *Nature* **1975**, *254*, 304–8.

(31) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16789–94.

(32) Panigrahi, S. K.; Desiraju, G. R. Strong and weak hydrogen bonds in the protein-ligand interface. *Proteins* **2007**, *67*, 128–41.

(33) Berski, S.; Ciunik, Z.; Drabent, K.; Latajka, Z.; Panek, J. Dominant role of C-Br---N halogen bond in molecular self-organization. Crystallographic and quantum-chemical study of schiff-base-containing triazoles. *J. Phys. Chem. B* **2004**, *108*, 12327–32.
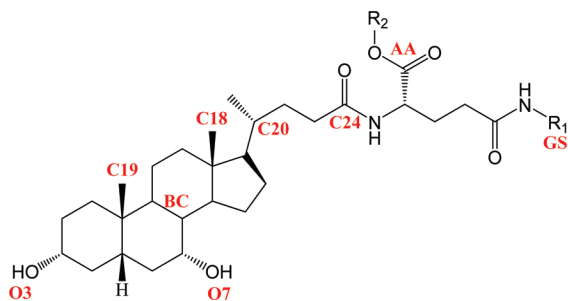
Figure 1. Structural features in anilinyl conjugates of glu-CDCA used for pharmacophore development. Notation is as follows: O3, position of 3-OH; O7, position of 7-OH; BC, centroid of B and C rings of steroidal nucleus; C18, location of C-18; C19, location of C-19; C20, location of C-20; C-24 is the location of the charge on unconjugated native bile acids; AA, centroid of carboxylic acid oxygens on α-acid; GS, centroid of γ-substituent (R_1); R_2 is the substituent on the α-acid ($-H$ or $-CH_2C_6H_5$).

hydrogen bond donors (HD) for each molecule was calculated using Discovery Studio (Accelrys Inc.).[36]

Conformational properties of molecules required for CSP model development included the 20 pharmacophore distances and 26 pharmacophore angles reported previously[15] as shown in Figure 1. Overlap coefficients (OC) for each compound were calculated with respect to the most potent substrate in the data set (compound **13**, $K_m = 0.1$ μM). Greater OC values reflect greater structural similarity to the most potent compound.

Regression Analysis. Using the statistical program ***R***, regression analysis was applied to each K_m and $K_m/\text{norm}V_{max}$ to develop CSP-SAR substrate models. Molecular descriptors were subjected to single variable as well as multivariable regression against each of the dependent variables. The structural and physicochemical descriptors were then combined to make a complete set of molecular descriptors for model development. All molecular descriptors were subjected to multivariable regression analysis against both K_m and $K_m/\text{norm}V_{max}$ separately.

Molecular descriptors were subjected to multivariable regression in all possible combinations of two. However, any combination having correlation between each other (*r*) greater than 0.8 were not considered. Table S2 in the Supporting Information lists correlation values between independent descriptors. All combinations of descriptors with *p*-values of the independent variables, as well as the intercept, less

than 0.05 were selected for further analysis. Additional descriptors were sequentially added to each selected pair of descriptors. This approach was applied until all possible combinations of descriptors yielded *p*-values greater than 0.05. All combinations of descriptors with *p*-values less than 0.05 formed the set of candidate models. The standardized coefficients of the independent variables[37] of the candidate models were also calculated,[38] as previously described.[15] Model evaluation was performed using the modified Akaike information criterion (AIC_C)[39,40] analysis and leave-one-out cross-validation method, as described previously.[14,15] In calculating AIC_C, the residual sum of squares (RSS) was obtained from fitting all compounds.

Results and Discussion

Results from uptake experiments into hASBT-MDCK cells are presented, followed by CSP-SAR substrate models. CSP-SAR models are for substrate affinity (i.e., K_m) and for reciprocal of efficiency (i.e., $K_m/\text{norm}V_{max}$).

Conjugate Uptake into hASBT-MDCK Cells. A congeneric series of 32 aniline and aminopiperidine bile acid conjugates previously synthesized and tested for hASBT inhibition[15] were evaluated for uptake by hASBT. Figure 2 shows the uptake profiles of taurocholate ($K_m = 4.84$ μM and $V_{max} = 0.699$ pmol/cm^2/s), which is a known potent substrate, and the unsubstituted anilinyl conjugate of glu-CDCA (compound **1**). Compound **1** kinetics was similar to that of taurocholate. Like taurocholate, compound **1** was a potent hASBT substrate ($K_m = 0.823$ μM and norm$V_{max} = 1.19$). Total uptake (in presence of sodium) was markedly higher than passive uptake (absence of sodium).

Kinetics of hASBT-mediated transport was assessed for each conjugate from transport studies employing a range of donor concentrations for each conjugate. Table 1 lists 27 compounds (compounds **1−27**) that were hASBT substrates. Table S1b in the Supporting Information provides stereochemical SMILES of the compounds used in this study. All were monoanionic, except one dianionic compound. Table 2 lists five compounds (compounds **28−32**) that were not hASBT substrates, and includes two dianions, one cation, and two zwitterionic compounds. Tables 1 and 2 indicate compound kinetic parameters K_m, normV_{max}, and P_P. Monoanionic conjugates were potent substrates of hASBT with activity ranging from 0.1 μM to 17 μM. This K_m range is the same as the range for native bile acids, which is from

(34) Acharya, C.; Seo, P. R.; Polli, J. E.; MacKerell, A. D., Jr. Computational model for predicting chemical substituent effects on passive drug permeability across parallel artificial membranes. *Mol. Pharmaceutics* **2008**, *5*, 818–28.

(35) Hanke, P. R.; Timm, P.; Falk, G.; Kramer, W. Behavior of different suture materials in the urinary bladder of the rabbit with special reference to wound healing, epithelization and crystallization. *Urol. Int.* **1994**, *52*, 26–33.

(36) Discovery Studio, Accelrys Inc. (http://www.accelrys.com/dstudio).

(37) Schroeder, L. D.; Sjoquist, D. L.; Stephan, P. E. *Understanding regression analysis: an introductory guide*; Sage Publications: 1986.

(38) Olobatuyi, M. E. *A user's guide to path analysis*; University Press of America: 2006.

(39) Akaike, H. Likeliood of a Model and Information Criteria. *J. Econometrics* **1981**, *16*, 3–14.

(40) McQuarrie, A. D. R.; Tsai, C. L. *Regression and Time Series Model Selection*; World Scientific: Singapore, 1998.
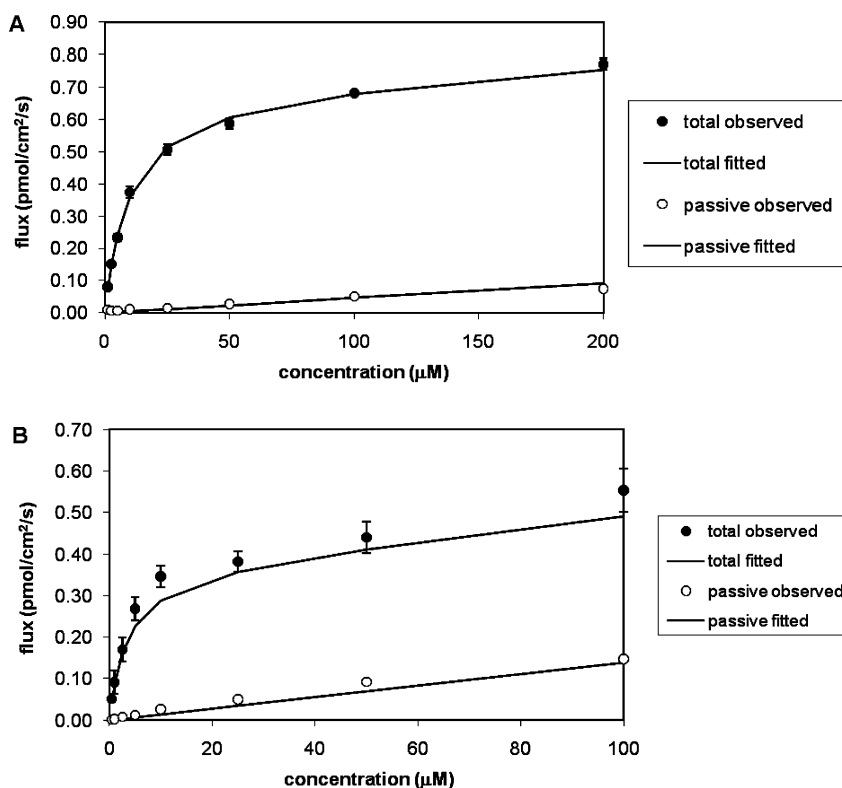
Figure 2. Concentration-dependent uptake profiles. (A) Taurocholate and (B) unsubstituted aniline conjugate were actively translocated by hASBT. Uptake was greater in the presence of sodium (●) than in the absence of sodium (○), indicating them to be hASBT substrates.

<0.1 to $16\,\mu$M.[41] These results suggest the potential of further exploiting C-24 conjugation to design prodrugs that can be targeted to hASBT. Among the substrates, ring substituents varied in electronic character, size, and hydrophobicity. Increased polarity led to a slight decrease in affinity (i.e., increase in K_m) as seen by esters, amine and boc group substituents on the benzene ring. Electronic character did not significantly impact affinity, as electron donating (e.g., methyl, methoxy) and electron withdrawing substituents (e.g., chloro, flouro) were potent substrates.

Regarding compound charge, substrate results differed from inhibition results. For inhibition, a single negative charge was not an essential requirement.[6] However, charge played the dominant role in translocation across hASBT. Only monoanions were potent substrates. Compared to other monoanions, the location of the negative charge was moved distal from C-24 to the benzene ring by protecting the α-acid on glutamic acid, resulting in loss of activity (compound **26**, $K_m = 495\,\mu$M). This result suggests that the negative charge should preferentially be within the vicinity of C-24 for active hASBT uptake. None of the three aminopiperidine conjugates (one cation and two zwitterions) were substrates. Only one of the three dianionic conjugates was a substrate (compound **27**), but was a very weak substrate. Previous studies have

(41) Balakrishnan, A.; Wring, S. A.; Polli, J. E. Interaction of native bile acids with human apical sodium-dependent bile acid transporter (hASBT): Influence of steroidal hydroxylation pattern and C-24 conjugation. *Pharm. Res.* **2006**, *23*, 1451–9.

also shown that dianions are weak inhibitors of hASBT,[15] but that cations and zwitterions bound with moderate to good affinity.[6,14] Translocation is a more elaborate and complex phenomenon than binding, such that compounds which were strong inhibitors of hASBT were not found to be strongly translocated if not bearing a single negative charge.

Figure 3 (panel A) illustrates the correlation between normV_{max} and substrate K_m for compounds **1−25**. Compounds **26** and **27** were weak substrates and not included. An inverse relationship was observed where higher capacity was weakly associated with substrate affinity (i.e., lower K_m) (linear $r^2 = 0.391$). This association points toward high binding affinity to generally promote transport capacity.

Figure 3 also plots the relationships of K_m and normV_{max} versus inhibitory K_i for compounds **1−25** (panels B and C, respectively). K_i values were reported previously[15] and reflect inhibition potency. Each K_m and normV_{max} was weakly associated with K_i (linear $r^2 = 0.360$ for K_m, and linear $r^2 = 0.310$ for normV_{max}). These monoanionic conjugates were potent inhibitors (K_i ranged from 0.889 to 16.3 μM) and potent substrates (K_m ranged from 0.1 to 17.5 μM). K_i and K_m values were qualitatively similar, reflecting high inhibition potency and high substrate affinity. Like native bile acids,[10] these data indicate that substrate binding was the rate limiting step for translocation of monoanions.

Computational Modeling. The compounds were subjected to CSP-SAR analysis. K_m, normV_{max}, normV_{max}/K_m, and $K_m/$normV_{max} were separately considered, with the aim

Table 1. Structures, K_m, and Normalized V_{max} Values of Anilinyl Conjugates of Glu-CDCA[c]
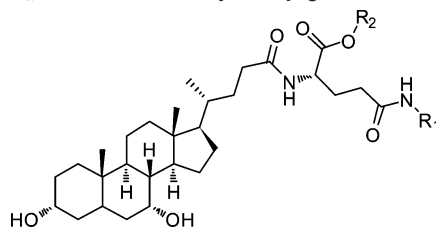
Compound	R$_1$	R$_2$	Charge	K_m (µM)[a]	Normalized V_{max}	$Pp \times 10^6$ (cm/s)[b]
1		H	-1	0.823 ± 0.320	1.19±0.05	1.40±0.11
2		H	-1	0.100[a] ± 0.773	1.51±0.18	4.51±0.05
3		H	-1	2.22 ± 1.71	1.30±0.18	6.03+0.72
4		H	-1	1.11 ± 0.35	1.01±0.03	0.643±.040
5		H	-1	0.450 ± 0.611	0.686±0.075	2.54±0.16
6		H	-1	0.100[a] ± 0.330	1.09±0.01	2.08+0.15
7		H	-1	1.69 ± 0.71	1.72±0.14	2.74±0.19
8		H	-1	0.919 ± 0.622	1.04±0.11	1.62±0.26
9		H	-1	0.100[a] ± 0.500	1.48±0.17	2.44±0.10
10		H	-1	3.87 ± 0.55	0.978±0.087	1.21±0.08
11		H	-1	1.21 ± 0.58	0.553±0.058	2.43±0.13
12		H	-1	6.56 ± 0.98	0.903±0.105	0.666±0.086
13		H	-1	0.100[a] ± 5.302	2.32±0.887	9.98±0.09
14		H	-1	3.42 ± 1.23	0.375±0.070	1.71±0.01

Table 1. Continued

Compound	R₁	R₂	Charge	K_m (µM)[a]	Normalized V_{max}	Pp x 10⁶ (cm/s)[b]
15		H	-1	7.27 ± 2.80	1.47±0.31	4.08±0.28
16		H	-1	2.54 ± 0.69	0.800±0.110	3.39±0.12
17		H	-1	10.3 ± 3.1	0.203±0.017	1.59±0.07
18		H	-1	17.5 ± 4.7	0.250±0.022	1.73±0.07
19		H	-1	9.89 ± 1.15	0.762±0.241	0.667±0.031
20		H	-1	14.9 ± 1.2	0.535±0.081	0.610±0.029
21		H	-1	7.91 ± 1.94	0.336±0.043	0.586±0.045
22		H	-1	0.549 ± 0.240	1.07±0.06	1.00+0.05
23		H	-1	7.82 ± 2.06	0.546±0.057	1.45+0.06
24		H	-1	2.53 ± 0.48	0.989±0.066	1.42+0.02
25		H	-1	11.4 ± 1.5	0.441±0.047	0.451+0.017
26			-1	495 ± 222	1.34±0.412	1.32+0.39
27		H	-2	607 ± 476	0.639±0.190	0.183±0.008

[a] K_m was assigned a value of 0.1 µM when regressed value was less than 0.1 µM; regressed K_m values for compounds **2**, **6**, **9**, and **13** were 0.010 ± 0.773 µM, 0.011 ± 0.331 µM, 0.007 ± 0.551 µM, and 0.008 ± 5.301 µM, respectively. [b] P_P is the passive permeability of the conjugates. [c] This table only includes conjugates that were substrates.

to identify structural and physicochemical descriptors that explain the biological activity. Results for normV_{max} and normV_{max}/K_m did not yield predictive models. The single variable regression analysis yielded poor correlation for

Table 2. Conjugates That Were Nonsubstrate for ASBT[a]

Compound	R_1	R_2	Charge	Pp x 10^6 (cm/s)
28	(2-methylbenzoic acid, COOH)	H	-2 (Dianion)	0.216±0.012
29	(HOOC-phenyl-CH2)	H	-2 (Dianion)	0.104±0.005
30	(HN-piperidine)	$CH_2C_6H_5$	+1 (Cation)	4.71±0.09
31	(HN-piperidine-CH2)	H	zwitterion	0.204±0.023
32	(benzyl-N-piperidine)	H	zwitterion	0.336±0.017

[a] R_1 and R_2 refer to substituents on the same general structure in Table 1.

normV_{max} and normV_{max}/K_m ($r^2 \leq 0.3$), and further fitting was not attempted. However, successful models were obtained for K_m and K_m/normV_{max}, and these results are discussed below.

CSP-SAR hASBT Substrate Model Development: Quantitative Model. Two data sets were examined to obtain substrate CSP-SAR models. Set 1 included all compounds that were substrates, while set 2 excluded compounds **26** and **27**. The K_m values for **26** and **27** were around 400−600-fold higher than that of other compounds, indicating that specific features of **26** and **27** led to significantly lower activities. **27** was the only dianionic conjugate that showed any measurable translocation. **26** was similar to **27** except that the α-acid on the glutamic acid linker was protected by the benzyl group, providing a distal charge to C-24. As shown below, these compounds dominated model fitting when included in data set (i.e., set 1), such that the sensitivity of the model to properties of the other compounds was diminished. Table S3 in the Supporting Information shows various molecular descriptor values used to describe K_m and K_m/normV_{max} of the anilinyl conjugates of glu-CDCA. Aminopiperidine conjugates were not substrates and were not considered in the model development.

CSP-SAR Model for K_m: Single Variable Regression. For each data set 1 and set 2, two best CSP-SAR models for K_m were obtained. Table S4 in the Supporting Information presents intermediate analysis results from single variable regression of molecular descriptors against K_m for set 1. Table S5 in the Supporting Information represents the same results for set 2. Molecular descriptors included both structural (e.g., angle, distance) and physicochemical descriptors where physicochemical descriptors provided higher correlations to K_m for single variable regression.

In the single regression analysis, the physiochemical descriptors provided higher correlations to K_m than the structural terms. For set 1, ΔG_w yielded the highest correlation against K_m among all the molecular descriptors (Table S4 in the Supporting Information), while HD yielded the highest correlation with K_m for set 2 (Table S5 in the Supporting Information). Log P and SASA$_{side}$ yielded the second highest correlation with K_m for set 1 and set 2, respectively. The negative coefficients of ΔG_w and log P for both sets indicated that increased compound hydrophobicity enhanced affinity (i.e., decreased K_m). Consistently, HD yielded a positive coefficient for set 2, further indicating that polar character reduced affinity.

Among the structural descriptors, OC of the angles C19−GS−O3 and O7−AA−C18 yielded the two highest correlations for set 1. Figure 4 (panels A and B) shows the probability distributions of these two angles for compounds **13** (most potent substrate among all compounds) and **27** (weakest substrate in set 1). All conformational descriptors, except angles GS−O3−AA and GS−O7−AA, exhibited negative linear coefficients for the OCs in set 1, indicating that conformational similarity to **13** promoted lower K_m value. Angles GS−O3−AA and GS−O7−AA had positive coefficients. However, further analysis showed almost no change in the angle distributions by compounds (Table S4 in the Supporting Information) and poor r^2, suggesting that these descriptors were not predictive of K_m.

Structural descriptors did not yield high correlation ($r^2 \leq 0.114$) with K_m for set 2 as shown in Table S5 in the Supporting Information. Figure S1 in the Supporting Information shows the probability distributions of angles O7−C18−GS and O7−C19−GS for **13** (most potent substrate among all compounds) and **18** (weakest substrate in
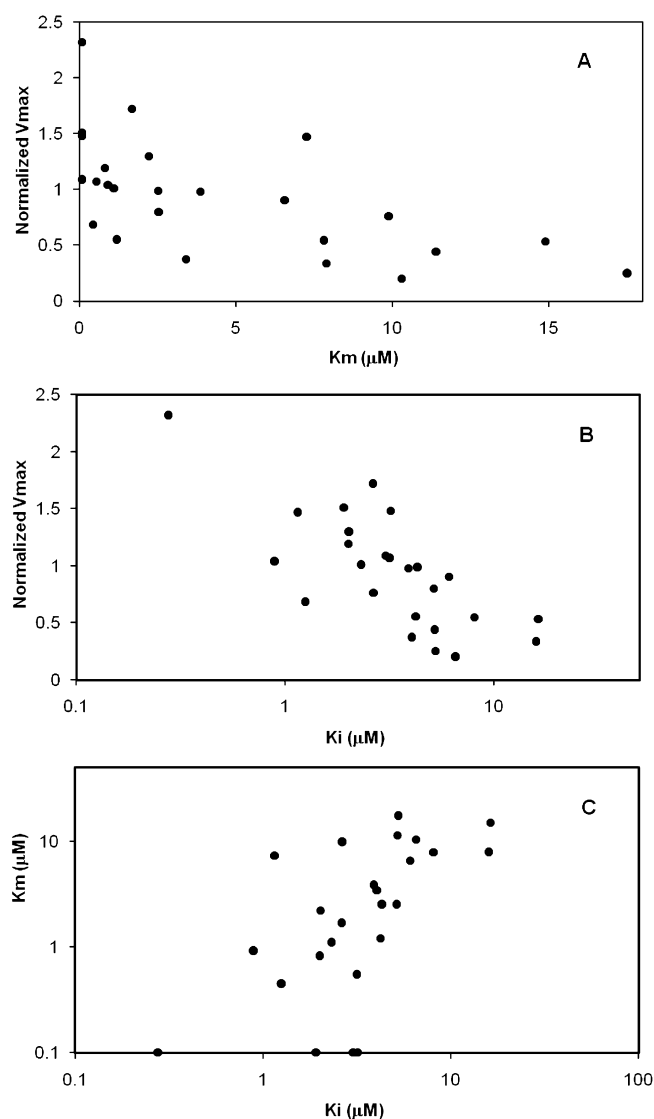
Figure 3. Plotted relationships between substrate and inhibition parameters for conjugates **1**−**25** (i.e., set 1). (A) The relationship between normalized V_{max} and K_m is plotted. Compounds with higher affinity (i.e., lower K_m) were associated with higher capacity. (B) The relationship between K_m and K_i is plotted. Substrate affinity and inhibitory potency were generally similar, implicating binding as the rate limiting step in overall conjugate translocation. (C) The relationship between normalized V_{max} and K_i is plotted. Inhibitor potency and capacity for each conjugate were generally positively associated, as lower K_i (i.e., higher binding) was associated with higher translocation capacity. The trend between K_i and normalized V_{max} was similar to the trend between K_m and normalized V_{max} in panel A.

set 2). These angles were the structural descriptors in single variable regression for set 2 that exhibited the two best correlations with K_m, but correlations are still relatively low.

CSP-SAR Model for K_m: Multivariable Regression. For set 1 and set 2, 860 and 168 candidate models, respectively, were identified via multivariable regression. Tables 3 and 4 presents the two top models for sets 1 and 2, respectively.
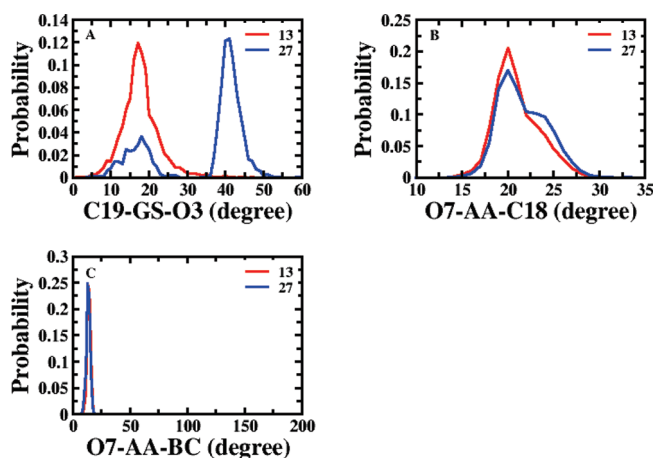

Figure 4. Probability distributions of structural descriptors for K_m modeling of set 1. (A) Angle C19−GS−O3 and (B) angle O7−AA−C18 displayed the two highest single variable correlations among all structural descriptors. (C) Angle O7−AA−BC represents spatial orientation of 7-OH, α-acid and the steroidal nucleus. The red line represents the distribution of **13** (most potent substrate of all compounds) and the blue line for **27** (weakest substrate in set 1). Panels B and C depict the conformational distributions of the topological descriptors in the top two multivariable CSP-SAR models for set 1.

AIC was used to rank model fit quality. The best model for set 1 yielded $r^2 = 0.926$ and $Q^2 = 0.630$, and showed 10.8% probability of being the best model. The best model set 2 yielded $r^2 = 0.821$ and $Q^2 = 0.683$, and had 13.5% probability of being the best model. The regression plots of the predicted vs observed K_m values are drawn in Figure 5 for set 1 and set 2.

ΔG_w and MR appeared in the two top models for set 1 with negative and positive coefficients, respectively, indicating that hydrophobicity favors transport affinity (Table 3). In these models, coefficients of angles O7−AA−C18 and O7−AA−BC were negative, indicating conformational similarity to **13**, promotes activity. Angle O7−AA−C18 is the spatial relationship between 7-OH, α-acid, and C18; angle O7−AA−BC characterizes the spatial orientation of 7-OH, α-acid, and the steroidal nucleus. Figure 4 (panels B and C) depicts the conformational distribution of these two structural descriptors for **13** and **27** (weakest substrate in set 1). The standardized coefficients for distances BC−GS$_{min}$ or O7−GS$_{min}$ indicated that these distances had the least effect on substrate activity.

ΔG_w appeared in 395 of the 860 candidate models for set 1, in part due to the dianionic compound **27**. This compound's weak affinity is explained by its extreme aqueous solvation free energy. For the top two models, the standardized coefficients of ΔG_w indicate this descriptor to be the most influential. The role of ΔG_w in determining K_m is consistent with inhibition results for these conjugates, where ΔG_w was also the most significant factor for binding affinity (i.e., K_i).

Table 3. Results from Multivariable Regression and AIC$_C$ Analysis for the Selected CSP-SAR Models for K_m for Set 1 [a]

CSP-SAR models	r^2	Q^2	F	W_i	K (μM)	A (μM/kcal/mol)	B (μM)	C (μM/m³/mol)	D (μM)	standardized A (μM/kcal/mol)	standardized B (μM)	standardized C (μM/m³/mol)	standardized D (μM)
ΔG$_w$, O7–AA–C18, MR, BC–GS$_{min}$	0.926	0.630	66.1	10.8	−1460 ± 217	−6.59 ± 0.58	−630 ± 137	72.4 ± 11	327 ± 104	−0.780	−0.456	0.390	0.329
ΔG$_w$, O7–AA–BC, MR, O7–GS$_{min}$	0.926	0.633	60.3	10.4	−1488 ± 214	−6.89 ± 0.60	−591 ± 129	73.1 ± 11	295 ± 103	−0.816	−0.413	0.394	0.288

[a] A, B, C and D represent the coefficient of the independent variables (molecular descriptors); K represents the coefficient of the intercept; Q² represents the cross-validated r²; F is the test statistic value from the F-test regarding the significance of the model; W$_i$ is percent probability of the model being the best model; set 1 includes all compounds that were substrates.

Table 4. Results from Multivariable Regression and AIC$_C$ Analysis for the Selected CSP-SAR Models for K_m for Set 2 [a]

CSP-SAR model	r^2	Q^2	F	W_i	K (μM)	A (μM)	B (μM)	C	D	standardized A (μM)	standardized B (μM)	standardized C	standardized D
HD, KierFlex, ΔG$_w$, GS–O3–C18	0.821	0.683	21.7	13.5	−99.3 ± 17	8.03 ± 1.3	5.10 ± 1.1	−0.556 ± 0.184 (μM/kcal/mol)	−23.2 ± 6.7 (μM)	0.662	0.457	−0.397 (μM/kcal/mol)	−0.438 (μM)
HD, GS–O7–C18, KierFlex, ΔG$_w$	0.809	0.673	20.1	6.35	−96.7 ± 17	7.74 ± 1.4	−18.6 ± 5.8	5.10 ± 1.2 (μM)	−0.483 ± 0.18 (μM/kcal/mol)	0.639	−0.401	0.456 (μM)	−0.344 (μM/kcal/mol)

[a] A, B, C and D represent the coefficient of the independent variables (molecular descriptors); K represents the coefficient of the intercept. Set 2 includes all compounds that were substrates, except **26** and **27**, which were weak substrates.
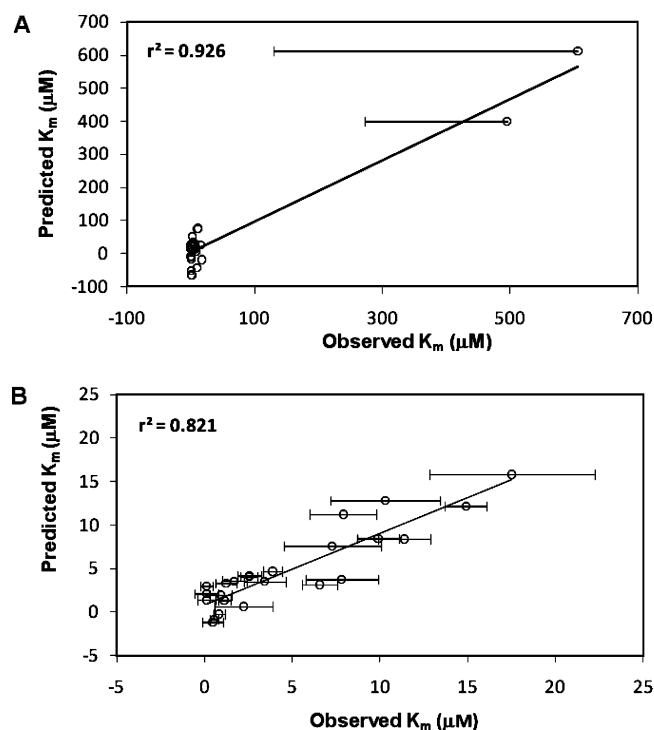
Figure 5. Regression plots of predicted vs observed K_m values for anilinyl conjugates of glu-CDCA using CSP-SAR. (A) Using set 1, the model yielded $r^2 = 0.926$ and $Q^2 = 0.630$. (B) Using set 2, which excludes compounds **26** and **27** which were the two least potent substrates, the model yielded $r^2 = 0.821$ and $Q^2 = 0.683$.
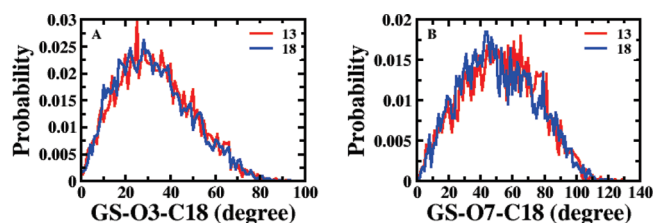

Figure 6. Probability distributions of structural descriptors for K_m modeling of set 2. (A) Angle GS−O3−C18 and (B) angle GS−O7−C18 were the structural descriptors in the best and second best models for set 2, respectively. Compounds **13** and **18** were the most and least potent substrates in set 2, respectively.

The two top models for set 2 included HD, KierFlex and ΔG_w (Table 4). The positive coefficient of HD indicated that increased polar character diminished activity. Consistently, the negative coefficient of log *P* indicated that hydrophobicity favors activity. The positive coefficient of KierFlex indicated that substrate flexibility disfavored affinity. The models also contained angle GS−O3−C18 or angle GS−O7−C18. Each showed negative coefficients, indicating that conformational similarity to **13** promoted activity. GS−O3−C18 is the angle between the γ-substituent, 3-OH, and C-18, while GS−O7−C18 represents the spatial relationship between the γ-substituent, 7-OH, and C-18. Figure 6 (panels A and B) depicts the conformational distributions for **13** and **18** (weakest substrate in set 2). While ΔG_w was perhaps the dominant parameter
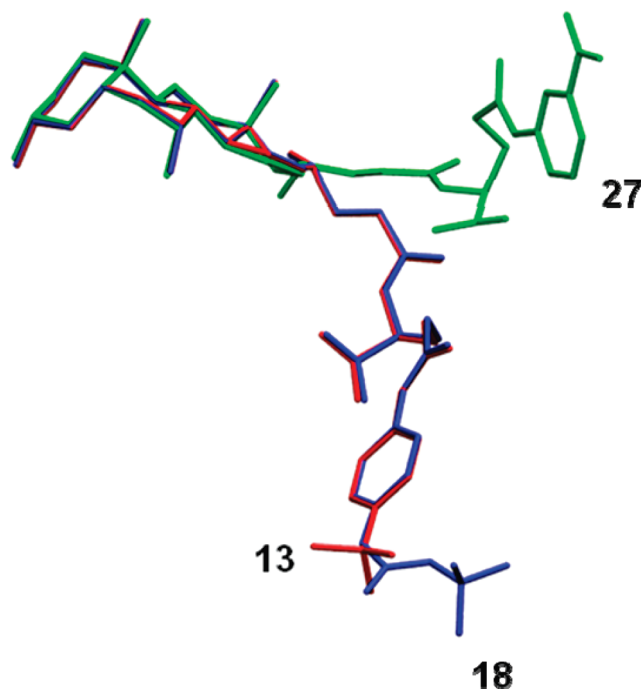

Figure 7. Representative conformations of compounds **13**, **18**, and **27**. The most potent substrate (**13**) is shown in red. The least potent substrate (**27**) of all compounds is shown in green. Compound **18** is shown in blue and is the least potent substrate in set 2. The γ-substituent of compound **27** sampled distinctive conformations compared to other compounds. Regarding structural conformation, compound **18** was more similar to **13** than was **27**.

for set 1, in part due to compound **27** all descriptors in set 2 contributed more equally to explain the K_m data, as judged by the similar magnitude coefficients. The predicted K_m values of the conjugates in both sets using their respective best model are shown in Table S8 in the Supporting Information.

Figure 7 presents representative conformers for compounds **13**, **18**, and **27** (red, blue, and green, respectively). The relative orientation of **27** (weakest substrate) is distinctly different from **13** (most potent substrate). Compound **18** is the least potent substrate in set 2 and was more similar to **13** than was **27**.

CSP-SAR Model for $K_m/normV_{max}$: Single Variable Regression. $K_m/normV_{max}$ describes the reciprocal of the transporter efficiency. The more familiar form $normV_{max}/K_m$ reflects transporter efficiency, but was not successfully modeled. Compared to $K_m/normV_{max}$, $normV_{max}/K_m$ was not successful due to the regression effects of assigned $K_m = 0.1$ μM to the most potent substrates, as K_m is in the denominator of $normV_{max}/K_m$ but not $K_m/normV_{max}$.

Interestingly, single variable regression results for $K_m/normV_{max}$ were similar to such results for K_m. ΔG_w yielded the highest correlation in set 1 (Table S6 in the Supporting Information), while HD yielded the highest correlation for set 2 (Table S7 in the Supporting Information). log *P* and SASA$_{side}$ yielded the second highest correlation with $K_m/$
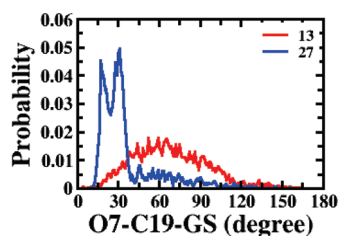
Figure 8. Probability distribution of angle O7−C19−GS. O7−C19−GS was the best structural description from single variable analysis for $K_m/norm V_{max}$ for set 1. The red line represents the distribution of **13**; the blue line represents the same for **27**.

$norm V_{max}$ for sets 1 and 2, respectively. For each ΔG_w, log P, and HD, compound hydrophobicity favored efficiency. The positive coefficient of $SASA_{side}$ indicates that bulky groups beyond the C-24 region disfavor activity, in agreement with prior findings.[6] OCs of the angles C19−GS−O3 and O7−C19−GS yielded the two highest correlations among the structural descriptors for set 1. Figure 4 (panel A) and Figure 8 present these angle distributions for **13** and **27**. Similar to the results for K_m, all conformational descriptors except angles GS−O3−AA and GS−O7−AA exhibited negative linear coefficients for the OC, as expected since similarity to **13** promoted lower $K_m/norm V_{max}$ value. Angles GS−O3−AA and GS−O7−AA had positive coefficients in set 1, but showed almost no change in compound sampling patterns (Table S6 in the Supporting Information) and poor r^2.

Among the structural descriptors, OCs of angles O7−C19−GS and O7−C18−GS yielded the two highest correlations in single variable regression analysis for set 2. Angle distributions for **13** and **18** are shown in Figure S1 in the Supporting Information. All structural descriptors, except for C19−GS−O7, AA−GS−O3 and GS−O3−AA, yielded negative coefficients, indicating that structural similarity with **13** favors activity (Table S7 in the Supporting Information). The poor r^2 (<0.005) of C19−GS−O7, AA−GS−O3 and GS−O3−AA indicated that these angles did not influence conjugate uptake efficiency.

CSP-SAR Model for $K_m/norm V_{max}$: Multivariable Regression. Multivariable regression identified 500 candidate models for set 1 and 465 candidate models for set 2. In Table 5, the best model for set 1 yielded $r^2 = 0.975$ and $Q^2 = 0.895$, and showed 28.9% probability of being the best model. In Table 6, the best model for set 2 yielded $r^2 = 0.937$ and $Q^2 = 0.877$, and showed 24.7% probability of being the best model. The regression plots of the predicted vs observed $K_m/norm V_{max}$ values are shown in Figure 9 for set 1 and set 2.

Each ΔG_w and SASA appeared in the two top models for set 1 with negative and positive coefficients, respectively. ΔG_w indicated that lipophilicity favored transport efficiency. SASA suggested that bulky ligands disfavored efficiency. Both models also included angle O7−AA−C18, which is the angle involving 7-OH, α-acid and C18. Its negative coefficients indicated that conformational similarity to **13**

Table 5. Results from Multivariable Regression and AIC$_C$ Analysis for the Selected CSP-SAR Models for $K_m/norm V_{max}$ for Set 1[a]

CSP-SAR model	r^2	Q^2	F	W_i	K (μM)	A (μM/kcal/mol)	B (μM)	C (μM/Å²)	D (μM/Å²)	standardized A (μM/kcal/mol)	standardized B (μM)	standardized C (μM/Å²)	standardized D (μM/Å²)
ΔG_w, O7−AA−C18, PSA$_{side}$, SASA	0.975	0.895	207	28.9	−1210 ± 140	−9.61 ± 0.42	−535 ± 77	0.600 ± 0.25	0.835 ± 0.13	−0.868	−0.296	0.087	0.249
ΔG_w, O7−AA−C18, SASA, PSA	0.974	0.867	197	15.1	−1218 ± 148	−9.62 ± 0.44	−556 ± 82	0.828 ± 0.14	0.554 ± 0.26	−0.869	−0.307	0.247	0.081

[a] A, B, C and D represent the coefficient of the independent variables (molecular descriptors); K represents the coefficient of the intercept. Set 1 includes all compounds that were substrates.

Table 6. Results from Multivariable Regression and AIC$_C$ Analysis for the Selected CSP-SAR Models for K_m/normV_{max} for Set 2[a]

CSP-SAR model	r^2	Q^2	F	W_i	K (μM)	A (μM)	B (μM/Å²)	C (μM/Å²)	D (μM)	standardized A (μM)	standardized B (μM/Å²)	standardized C (μM/Å²)	standardized D (μM)
HD, SASA$_{side}$, SASA, O7−AA−C19	0.937	0.877	71.1	24.7	−71.9 ± 29	26.7 ± 2.6	0.454 ± 0.061	−0.346 ± 0.077	70.7 ± 23	0.629	1.24	−0.976	0.432
HD, SASA$_{side}$, SASA, C19−AA	0.936	0.884	69.6	19.4	−83.2 ± 28	27.8 ± 2.7	0.422 ± 0.054	−0.313 ± 0.069	63.5 ± 21	0.656	1.15	−0.884	0.379

[a] A, B, C and D represent the coefficient of the independent variables (molecular descriptors); K represents the coefficient of the intercept. Set 2 includes all compounds that were substrates, except **26** and **27**, which were weak substrates.
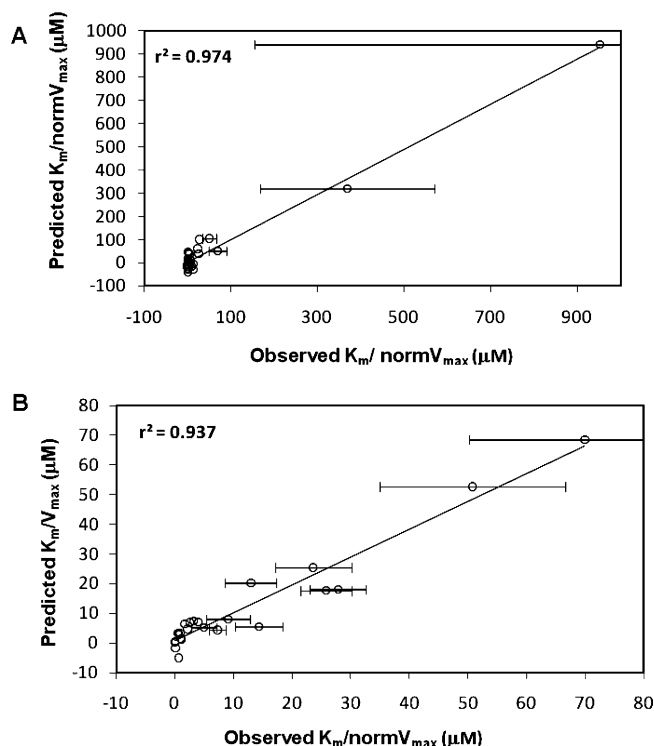

Figure 9. Regression plots of predicted vs observed K_m/normV_{max} values for anilinyl conjugates of glu-CDCA using CSP-SAR. (A) Using set 1, the model yielded $r^2 = 0.975$ and $Q^2 = 0.895$. (B) Using set 2, the model yielded $r^2 = 0.937$ and $Q^2 = 0.877$.

promoted efficiency. Figure 4 (panel B) depicts O7−AA−C18 distribution for **13** and **27**. The two top models also included either PSA$_{side}$ and PSA (Table 5). Higher conjugate polar surface area disfavored efficiency. Similar to K_m models for set 1, the standardized coefficients of both models suggested that ΔG_w predominated over other descriptors to determine transport efficiency.

The two top models for set 2 included HD, SASA, and SASA$_{side}$ (Table 6). The positive coefficient of HD indicated that increase in polar character diminished efficiency. The positive coefficient of SASA$_{side}$ indicated that bulkier C-24 side chain disfavored efficiency. In contrast to SASA$_{side}$, the negative coefficient of SASA indicated that overall solvent accessible surface area favored efficiency. However, the standardized coefficients indicate that SASA did not impact efficiency as much as SASA$_{side}$. Each model included either angle O7−AA−C19 or distance C19−AA. Each exhibited a positive coefficient, indicating that similarity with **13** disfavored efficiency, although amongst all the descriptors these structural properties had the least amount of contribution to the models indicating that activity was mainly explained by physicochemical properties. The predicted K_m/normV_{max} values of the conjugates in both sets using their respective best model are shown in Table S9 in the Supporting Information.

Logarithmic transformed data was also analyzed [i.e., log K_m and log(K_m/normV_{max})]. Model regression results for log K_m for set 1 and set 2 are presented in Tables S10 and S11

in the Supporting Information, respectively. Model regression results for $\log(K_m/\text{norm}V_{max})$ for set 1 and set 2 are presented in Tables S12 and S13 in the Supporting Information, respectively. Results from logarithmic transformed data agreed with nonlogarithmic transformed data analysis, in that increased hydrophobicity enhanced activity. Figure S2 in the Supporting Information shows predicted versus observed log K_m for set 1 and set 2. Figure S3 in the Supporting Information shows predicted versus observed $\log(K_m/\text{norm}V_{max})$ for set 1 and set 2.

Conclusion

hASBT is a prodrug target to increase the absorption of drug candidates that otherwise are insufficiently orally absorbed. The objective was to derive a CSP-SAR model for hASBT uptake, since 3D-QSAR models for ASBT substrates are lacking. A congeneric series of 32 aniline and aminopiperidine bile acid conjugates were evaluated. Twenty-seven compounds were substrates. Five were nonsubstrates. Monoanionic conjugates were potent substrates, with K_m in the same range as K_m of native bile acids, and thus supporting C-24 conjugation as a prodrug approach. Among the substrates, ring substituents that increased polarity led to a slight decrease in affinity. Electronic character did not significantly impact affinity. Only monoanions were potent substrates, with preference for the negative charge to be proximal to C-24 for active hASBT uptake. None of the three aminopiperidine conjugates (one cation and two zwitterions) were substrates, although these three conjugates were potent inhibitors. Only one of the three dianionic conjugates was a substrate, but it was a weak substrate. Translocation across hASBT is a more elaborate and complex phenomenon than binding to hASBT, such that compounds which were strong inhibitors of ASBT were not found to be strongly translocated if not bearing a single negative charge.

CSP-SAR models successfully explained substrate affinity (i.e., K_m) and $K_m/\text{norm}V_{max}$, which is the inverse of substrate efficiency. Since K_m data for compounds **26** and **27** were around 400−600-fold higher than the rest of the compounds,

two parallel sets of analysis were performed; one set included all substrates (set 1), while the other set excluded **26** and **27** (set 2). Single variable linear regression analysis showed that physicochemical properties of these compounds yielded better correlation with the various kinetic parameters than the conformational descriptors. Both K_m and $K_m/\text{norm}V_{max}$ yielded successful CSP-SAR models for both sets of compounds. The best CSP-SAR model for K_m included two structural and two physiochemical descriptors, where substrate hydrophobicity enhanced affinity. The best CSP-SAR model for $K_m/\text{norm}V_{max}$ in both set 1 and set 2 employed one structural and three physicochemical descriptors, indicating hydrophobicity enhanced efficiency. Notably, the difference in the quality of model fit and the predictive power of the best models between set 1 and set 2 was less than 0.105 showing that the CSP-SAR method was able to develop robust models for both sets of compounds. CSP-SAR models for the substrate requirements were overall guided by conjugate hydrophobic and polar characteristics.

Abbreviations Used

hASBT, human apical sodium-dependent bile acid transporter; CDCA, chenodeoxycholic acid; CSP, conformationally sampled pharmacophore; 3D-QSAR, three-dimensional structure acivity relationship; MD, molecular dynamics; PSA, polar surface area; HD, hydrogen bond donor; OC, overlap coefficient; SASA, solvent accessible surface area; MR, molar refractivity; MDCK, Madin Darby canine kidney.