Interpretation of Scoring Functions Using 3D Molecular Fields. Mapping the Diacyl-Hydrazine-Binding Pocket of an Insect Ecdysone Receptor

B. Bordas,*,† I. Belai,† A. Lopata,‡ and Z. Szanto‡

Plant Protection Institute, Hungarian Academy of Sciences, P.O. Box 102, H-1525 Budapest, Hungary, and CheMicro Ltd., Károly körút 1, H-1075 Budapest, Hungary

Received July 27, 2006

A method is presented for the interpretation of receptor docking score values (rough measures of binding affinities) of ligands in terms of 3D molecular field interaction contributions. The FlexX and FlexX-Pharm methods were used to dock the structures of designed sets of ligands into the ligand-binding pocket of a selected receptor. In the next step the relationship was investigated between the FlexX and CScore scores and 3D molecular fields obtained for the docked conformations of the ligands, using the CoMFA (Comparative Molecular Field Analysis) and CoMSIA (Comparative Molecular Similarity Indices Analysis) methods. The approach yielded highly significant CoMFA and CoMSIA models demonstrating that a high portion of the variance in the docking score values of the ligands can be explained by steric, electrostatic, hydrophobic, and hydrogen bond donor and acceptor molecular field interaction contributions. The approach was exemplified by using the crystal structure of the ligand-binding domain of the ecdysone receptor (EcR) of the moth *Heliotis virescens* as well as virtual molecule libraries of analogues of known diacyl-hydrazine (DAH) type ecdysteroid agonists. By docking appropriately designed virtual compound libraries into the DAH binding pocket of EcR followed by CoMFA and CoMSIA of the docked conformations, hitherto unexplored regions of the receptor cavity could be mapped. By mapping the significant molecular field interaction contributions onto the model of the receptor−ligand complex, important receptor−ligand interactions could be highlighted that may help the design of novel highly scored receptor ligands. An advantage of the method is that no experimental biological activity data are required to exhaustively map the receptor-binding site.

INTRODUCTION

New screening methodologies have contributed significantly to the efficiency of the drug discovery process. The throughput of conventional screening methods of compound libraries has been significantly surpassed by the in silico screening methods allowing the evaluation of large virtual compound libraries.[1−4] Ligand- and receptor-based methods are used to construct predictive models for lead optimization and de novo drug design. Receptor-based methods rely on prior knowledge of a pharmacological target (e.g., protein structure) to derive novel ligands, while ligand-based methods are used when no target structure is available. These methods can be combined in the framework of in silico lead generation and optimization methodologies to increase performance of the drug design process. This paper presents a novel approach for the integration of the receptor- and ligand-based methods for efficient drug and pesticide design.

Docking algorithms employ various scoring functions to estimate the binding affinity for a receptor−ligand complex of known 3D structure.[5] The score values for a compound set are represented by an 1D numerical vector, the meaning of which can be interpreted in terms of general receptor−ligand interactions only by examining the scoring function itself. The output of a set of scoring functions can be combined to produce a consensus score that is supposed to be more accurate and robust than any single function in itself. However, the mechanistic interpretation of a consensus score is even more difficult than elucidation of the meaning of a single scoring function.

Visualizations and analysis of the details of receptor−ligand interactions for each ligand in a large scale screening project can be rather tedious. In contrast, the output of several ligand-based design methods, e.g., CoMFA,[6] CoMSIA,[7] or DISCO,[8] representing the receptor−ligand interactions of a set of compounds, is more informative for the researcher. For instance, the popular ligand-based 3D QSAR methods CoMFA and CoMSIA can visualize the correlation between a biological activity and the relevant molecular fields around the aligned ligands by colored 3D isocontours, giving an immediate overall view of the receptor−ligand interactions. DISCOtech, the successor of the DISCO module in the Sybyl suite of programs,[9] generates and visualizes 3D pharmacophore hypotheses comprising the features of the ligands important for binding to a receptor. Receptor-based drug design methods do not necessarily need experimental biological activity data (e.g., receptor docking procedures or de novo drug design methods), whereas ligand-based design methods (e.g., classical QSAR, CoMFA, CoMSIA, DISCOtech, and pattern recognition approaches) require biological activity data.

* Corresponding author phone: 36-1-48-77-542; fax: 36-1-48-77-555; e-mail: bbor@nki.hu.
† Plant Protection Institute.
‡ CheMicro Ltd.

10.1021/ci600317v CCC: $37.00
Published on Web 12/07/2006

In this paper, a method is presented for the interpretation of receptor docking score values of ligands using 3D molecular field interaction contributions. The FlexX and FlexX-Pharm methods[10,11] were used to dock the structures of designed sets of ligands into the ligand-binding pocket of a selected receptor. In the next step the relationship was investigated between the FlexX and CScore[12] scores and 3D molecular fields obtained for the docked conformations of the ligands, using the CoMFA (Comparative Molecular Field Analysis) and CoMSIA (Comparative Molecular Similarity Indices Analysis) methods. Docking-based alignments are often used in CoMFA and CoMSIA studies.[13] In the present approach the complementary use of receptor-based docking methodologies and ligand-based 3D QSAR methods, employing the docking score values of the ligands as dependent variables, allows the CoMFA and CoMSIA analysis of any virtual compound library without available experimental biological activity data. By means of the proposed approach, various creative applications can be envisaged. Here we will show that by using appropriately designed virtual compound libraries, receptor sites can be systematically mapped and novel receptor−ligand interaction sites identified by locating the 3D position and nature of significant molecular field interaction contributions.

High statistical significance of the developed CoMFA and CoMSIA models indicated that in most cases a major portion of the variance in the docking score values (most conventional r^2 values fell into the range of 0.95 through 1) was explained by steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor molecular field interaction contributions.

The usefulness of the proposed approach for the design of highly scored receptor ligands is demonstrated using the recently reported crystal structure of the ligand binding domain (LBD) of the ecdysone receptor (EcR) of the moth *Heliotis virescens* (pdb code: 1R20)[14] and virtual molecule libraries of analogues of known diacyl-hydrazine (DAH) type ecdysteroid agonists compiled for the purpose of mapping the receptor.

MATERIALS AND METHODS

Computational Methods. All calculations were performed using the Sybyl 7.1 molecular modeling program package[9] running under Irix 6.5 operating system on a Silicon Graphics Octane2 R12000 workstation. Structures of ligands for the docking experiments were built manually in Sybyl followed by molecular mechanics geometry optimization using the Tripos force field with Gasteiger−Hückel charges, the conjugate gradient minimization method, and a gradient of 0.01 kcal·mol^{-1}·Å^{-1} as termination criterion.

The optimized ligands were docked into the DAH binding pocket of EcR using the Sybyl FlexX and FlexX-Pharm modules.[9] The cocrystal file of the LBD of the EcR/USP heterodimer from the moth *H. virescens* and a synthetic ecdysteroid agonist BYI06830 (**1**) was retrieved from the Protein Data Bank (pdb code: 1R20). The EcR protein has different but partially overlapping binding cavities for the ecdysteroid and DAH ligands.[14] The water molecules were removed from the crystal structure, and hydrogens were added to the protein residues by utilizing the Sybyl Biopolymer module. The receptor site was defined in FlexX containing all residues within 6.5 Å radius of the bound ligand.

Sybyl FlexX is a docking program that rapidly docks a conformationally flexible ligand into a receptor binding site while keeping the receptor structure rigid, using an incremental construction algorithm that builds the ligand in the site.[9,10] The resulting docked conformations are scored based on the strength of receptor−ligand interactions. The Sybyl CScore module was also employed that uses multiple types of scoring functions to rank the affinity of ligands bound to the active site of a receptor.[9,12] CScore is a consensus scoring program that integrates multiple, well-known, and extensively applied scoring functions from the scientific literature: a GOLD-like function (G Score), a DOCK-like function (D Score), ChemScore (ChemScore), a PMF function (PMF Score), and FlexX Score (F Score). In the docking experiments the program Sybyl FlexX-Pharm was also used that allows pharmacophore-type constraints to be used in FlexX to guide ligand docking.[9,11] Two primary types of constraints are available in FlexX-Pharm: receptor/ligand-interactions and spatial constraints. FlexX uses FlexX-Pharm constraints as filters prior to docking in a rapid prescreen to ensure that a ligand can possibly satisfy the set of pharmacophore constraints.

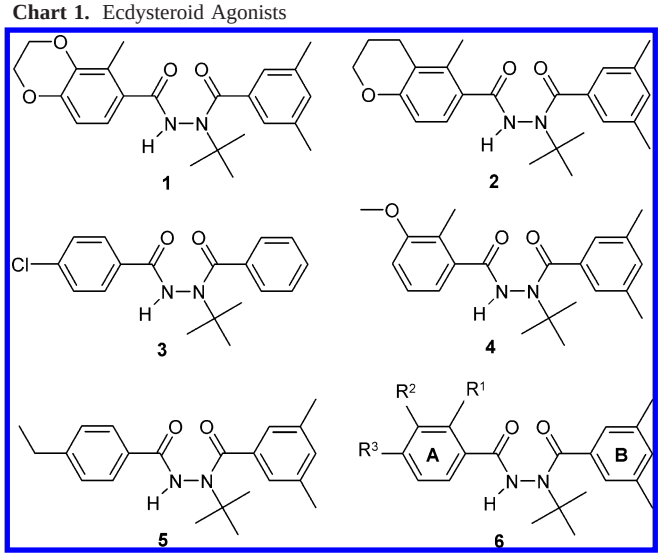
In our FlexX-Pharm docking experiments the common dibenzoyl-hydrazine skeleton of the analogue structures was anchored in the receptor cavity by a pharmacophore constraint consisting of three hydrogen bonds. These bonds bind the benzoylcarbonyl oxygen next to the **A** aromatic ring to the alcoholic hydroxyl group of Thr343 (H-bond donor), the benzoylcarbonyl oxygen next to the **B** aromatic ring to the side-chain amide group of Asn504 (H-bond donor), and the N−H group of the hydrazine moiety to the phenolic hydroxyl group of Tyr408 (H-bond acceptor). In the FlexX-Pharm pharmacophore all three interaction parameters were defined as Essential.

The FlexX-Pharm docked conformations referring to five different score values [F Score (FlexX Score), ChemScore, G Score, D Score, and PMF Score] for each compound set were separately saved into Sybyl Mol2 molecule databases and tables employing the CScore module. Then Gasteiger−Hückel atom charges on the FlexX-Pharm docked conformations were calculated using an in house SPL script.

CoMFA and CoMSIA (implemented in the QSAR module of Sybyl)[6,7,9] of each conformation set were performed using the five different score values as dependent variables. All ligands were used in their docked conformation without further alignment.

CoMFA was employed to correlate the score values of the ligands with their docked 3D structures represented by a steric (Lennard-Jones) and electrostatic (Columbic) molecular field sampled at the intersections of a 3D lattice. The applied grid spacing was 2 Å. The probe atom was an sp^3 hybridized carbon atom with a charge of +1, the dielectric function was $1/r$, and the dielectric constant was $\epsilon = 1$. Partial least squares of latent variables (PLS) analysis was carried out to obtain a linear QSAR expression. Cross-validation was performed using the leave-one-out (LOO) cross-validation technique. Column filtering with a minimum σ value of 2.0 kcal·mol^{-1} was applied. The number of accepted PLS components was based on the first local maximum of q^2.

Chart 1. Ecdysteroid Agonists



CoMSIA is a 3D QSAR method similar to CoMFA, where the molecular fields are expressed in terms of similarity indices between the compounds of interest calculated via interaction with a common probe atom at intersections of a regularly spaced lattice. CoMSIA takes into account the following molecular properties that are assumed to have significant contributions to ligand binding affinity: (i) steric contributions by the third power of the atomic radii, (ii) electrostatics by atomic charges, (iii) hydrophobicities by atom-based parameters, (iv) hydrogen bond donor properties; and (v) hydrogen bond acceptor properties by suitably placed pseudoatom probes with 1.0 Å radius, +1.0 charge, as well as hydrophobicity and hydrogen bond property values of +1. The value of the attenuation factor α was 0.3 for the Gaussian-type function. The other settings were the same as in the CoMFA calculations.

In CoMFA the calculated steric and electrostatic fields and in CoMSIA the calculated steric, electrostatic, hydrophobic, and hydrogen bond donor and hydrogen bond acceptor fields around the investigated ligands were visualized as scalar products of coefficients and standard deviations (StDev*Coeff) set for favored and disfavored levels for the molecular fields. The colored 3D isoenergy contours indicate the importance and positive or negative contributions of the molecular fields to the dependent variable, in our case to the docking score value.

The conventional r^2 of the CoMFA and CoMSIA models indicates what portion of the variance in the dependent variable is explained by the model, whereas the cross-validated r^2 (q^2)[6] value indicates how well the dependent variable is predicted for each compound by the model calculated using the rest of the analogues in the data set. In a summary of the criteria of a well-conducted CoMFA analysis, Martin et al.[15] proposed that q^2 should be >0.3 such that the possibility of chance correlation is <5%.

Compounds. In this study the synthetic ecdysteroid agonist BYI06830 (**1**) and four registered agonists (insecticides against lepidopteras)[16,17] were considered: chromafenozide (**2**); halofenozide (**3**); methoxyfenozide (**4**); and tebufenozide (**5**). In addition, a general structure **6** of the diacyl-hydrazine analogues was also used to generate virtual molecular libraries for the docking experiments (Chart 1).

Chart 1 shows the synthetic ecdysteroid agonist BYI06830, *N′-tert*-butyl-*N′*-(3,5-dimethylbenzoyl)-5-methyl-2,3-dihydro-1,4-benzodioxine-6-carbohydrazide (**1**), four registered DAH type agonists: chromafenozide, *N′-tert*-butyl-*N′*-(3,5-dimethylbenzoyl)-5-methylchromane-6-carbohydrazide (**2**), halofenozide, *N′*-benzoyl-*N′-tert*-butyl-4-chlorobenzohydrazide (**3**), methoxyfenozide, *N′-tert*-butyl-*N′*-(3,5-dimethylbenzoyl)-3-methoxy-2-methylbenzohydrazide (**4**), tebufenozide, *N-tert*-butyl-*N′*-(4-ethylbenzoyl)-3,5-dimethylbenzohydrazide (**5**), and a general structure of the diacyl-hydrazine analogues used in docking experiments (**6**).

Data Sets. Four FlexX-Pharm docked conformation sets (sets 1−4) were generated for demonstrating the docking score interpretation procedure. For each docked conformation set, five subsets were generated with CScore referring to five docking score evaluation methods (e.g., for set 2, the set 2_1, set 2_2, set 2_3, set 2_4, and set 2_5 terms denote the subsets containing the docked conformations of **6** according to the F Score, ChemScore, G Score, D Score, and PMF Score values, respectively).

Set 1: 53 docked conformations of the four registered ecdysteroid agonists (**2**−**5**) into the DAH binding pocket of EcR, including 7 successfully docked conformations of **2**, 10 successfully docked conformations of **3**, 18 successfully docked conformations of **4**, and 18 successfully docked conformations of **5** (for each agonist all successfully docked conformations were considered);

Set 2: 24 docked conformations of 24 molecules of general structure **6**, where R^1, R^2, and $R^3 = C_1-C_8$ normal alkyl substituent, with the proviso that only one of R^1, R^2, and R^3 is other than H (for each molecule the first ranked docked conformation was considered for each scoring function);

Set 3: 24 docked conformations of 24 molecules of general structure **6**, where R^1, R^2, and $R^3 = (CH_2)_nOCH_3$, where $n = 0-7$, with the proviso that only one of R^1, R^2, and R^3 is other than H (for each molecule the first ranked docked conformation was considered for each scoring function);

Set 4: 24 docked conformations of 24 molecules of general structure **6**, where R^1, R^2, and $R^3 = (CH_2)_nOH$, where $n = 0-7$, with the proviso that only one of R^1, R^2, and R^3 is other than H (for each molecule the first ranked docked conformation was considered for each scoring function).

RESULTS

First we extracted the agonist BYI06830 (**1**) from the DAH receptor complex (pdb code: 1R20) and then successfully redocked it into the same receptor site using FlexX (rms deviation between the original and redocked structures was 0.54 Å) (Figure 1). This way we checked the quality of the receptor structure prepared for docking (water molecules removed, hydrogen atoms added) as well as the FlexX docking procedure that proved to be appropriate. The FlexX algorithm found the three highly conserved hydrogen bonding interactions between the BYI06830 agonist and the EcR, that are also present in the crystal structure of the receptor−ligand complex. These interactions were used subsequently for setting up a pharmacophore constraint set in FlexX-Pharm to guide the FlexX docking procedure.

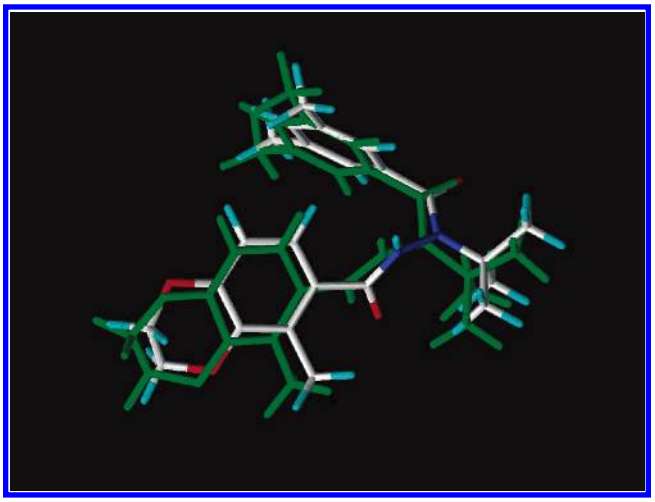


Figure 1. Overlay of the FlexX docked conformation of the agonist BYI06830 and its crystal structure (green lines) in the DAH binding site of EcR (rmsd = 0.54 Å).

The CoMFA and CoMSIA calculations of sets 1_1−4_5 using the respective docking score values as dependent variables yielded CoMFA and CoMSIA models with high statistical significance (Table 1).

DAHs **2**−**5** were docked into the DAH binding pocket of EcR using FlexX and FlexX-Pharm, yielding altogether 53 successfully docked conformations. Five compound sets, each including the 53 docked conformations (sets 1_1−1_5), were compiled employing CScore on the basis of the five score values obtained for each conformation (F Score, ChemScore, G Score, D Score, and PMF Score). CoMFA and CoMSIA were performed for each set using the corresponding score values as dependent variables. The best model with the smallest number of PLS components (N = 3) and the highest q^2 value (0.902) was obtained with CoMFA using the G Score values (r^2 = 0.956) (Table 1, CoMFA set 1_3). Figure 2 shows the contour plot of the electrostatic and steric field contributions ("StDev*Coeff") obtained for this CoMFA model. In Figure 2 four ligands (the first ranked docked conformation of each of the four registered ecdysteroid agonists **2**−**5**, in wire-frame model) as well as 3 amino acid residues of EcR (Thr343, Asn504 and Tyr408, in ball-and-stick model) are shown for illustration. Blue and red isocontours (60 and 40% contributions, respectively) denote regions in 3D where negative electrostatic fields decrease or increase, respectively, the magnitude of the score value. Green and yellow isocontours (60 and 40% contributions, respectively) denote regions in 3D where the presence of steric bulk increases or decreases, respectively, the magnitude of the score value. Hydrogen bonding interactions of the two benzoylcarbonyl oxygens and the hydrazine N−H group of the docked conformers with the Thr343, Asn504, and Tyr408 residues, respectively, enforced by the pharmacophore constraint defined in FlexX-Pharm, are shown with dashed lines.

Twenty-four DAHs of type **6** with a C_1−C_8 normal alkyl substituent in the ortho-, meta-, or para-position of the **A** aromatic ring were docked into the DAH binding pocket of EcR using FlexX and FlexX-Pharm. This way, the available space in the receptor pocket was mapped for substituents around the A aromatic ring. Five compound sets were compiled employing CScore, each including the first ranked docked conformation for each DAH and each scoring function (sets 2_1−2_5). CoMFA and CoMSIA were performed for each set using the corresponding score values as dependent variables. The best model with a small number of PLS components (N = 4) and a high q^2 value (0.851) was obtained with CoMFA using the ChemScore values (r^2 = 0.960) (Table 1, CoMFA set 2_2). It should be noted that the large contributions of the steric and hydrophobic molecular fields obtained for the CoMFA and CoMSIA models of set 2 (see Table 1) are in accordance with our original intention that this compound set was designed to map the available space around the **A** ring in the DAH binding pocket of EcR. In Figure 3 the green isocontour plot denotes the steric field contributions ("StDev*Coeff") (60% contribution) obtained for the best CoMFA model. In Figure 3 all 24 ligands included in the analysis (in wire-frame model) as well as 3 amino acid residues of EcR (Thr343, Asn504, and Tyr408, in ball-and-stick model) are shown for illustration.

Twenty-four DAHs of type **6** substituted with a $(CH_2)_n$-OCH_3 group (n = 0−7) in the ortho-, meta-, or para-position of the **A** aromatic ring were docked into the DAH binding pocket of EcR using FlexX and FlexX-Pharm. This way, the receptor pocket was searched around the **A** aromatic ring for amino acid residues with hydrogen bond donor ability. Five compound sets were compiled employing CScore, each including the first ranked docked conformation for each DAH and each scoring function (sets 3_1−3_5). CoMFA and CoMSIA were performed for each set using the corresponding score values as dependent variables. The best CoMSIA model with the smallest number of PLS components (N = 3) and the highest q^2 value (0.906) was obtained with the ChemScore values (r^2 = 0.982) (Table 1, CoMSIA set 3_2). It should be noted that the large contributions of the hydrogen bond acceptor and electrostatic molecular fields obtained for the CoMSIA models of set 3 (see Table 1) are in accordance with our original intention that this compound set was designed to search the receptor pocket around the **A** ring for hydrogen bond donating amino acid residues. Figure 4 shows the contour plot of the hydrogen bond acceptor field contributions ("StDev*Coeff") (60% contribution) obtained for the best CoMSIA model. The side-chain amide moiety of the Gln503 residue was identified as a hydrogen bond donor interacting with the ether oxygen of the substituent on the **A** aromatic ring of 15 (out of all 24) docked conformers as acceptors. The interaction with the amide proton of Gln503 was found to be significant (magenta contours). Hydrogen bonding interactions are shown with dashed lines. In Figure 4 15 ligands (i.e., those forming a hydrogen bond with Gln503, in wire-frame model) as well as 4 amino acid residues of EcR (Thr343, Asn504, Tyr408, and Gln503, in ball-and-stick model) are shown for illustration.

Twenty-four DAHs of type **6** substituted with a $-(CH_2)_n$-OH group (n = 0−7) in the ortho-, meta-, or para-position of the **A** aromatic ring were docked into the DAH binding pocket of EcR using FlexX and FlexX-Pharm. This way, the receptor pocket was searched around the **A** aromatic ring for amino acid residues with hydrogen bond acceptor and donor abilities. Five compound sets were compiled employing CScore, each including the first ranked docked conformation for each DAH and each scoring function (sets 4_1−4_5). CoMFA and CoMSIA were performed for each set

Table 1. CoMFA and CoMSIA Results for Sets 1_1−4_5 Using the Respective Docking Score Values as Dependent Variables[l]

set	score	n^a	N^b	$q^{2\,c}$	$r^{2\,d}$	SEE[e]	F^f	S^g	E^h	H^i	D^j	A^k
						CoMFA Set 1						
1_1	F_Score	53	3	0.596	0.880	1.194	120.1	0.713	0.287			
1_2	ChemScore	53	4	0.834	0.965	0.473	333.9	0.657	0.343			
1_3	**G Score**	**53**	**3**	**0.902**	**0.956**	**6.462**	**358.6**	**0.688**	**0.312**			
1_4	D Score	53	3	0.897	0.959	2.617	383.3	0.640	0.360			
1_5	PMF Score	53	5	0.745	0.961	1.792	230.1	0.655	0.345			
						CoMSIA Set 1						
1_1	F_Score	53	7	0.488	0.946	0.835	113.1	0.178	0.221	0.288	0.170	0.142
1_2	ChemScore	53	5	0.829	0.956	0.539	204.0	0.189	0.206	0.296	0.183	0.125
1_3	G Score	53	4	0.891	0.947	7.212	213.8	0.241	0.278	0.332	0.047	0.101
1_4	D Score	53	3	0.892	0.935	3.288	236.9	0.206	0.294	0.330	0.090	0.079
1_5	PMF Score	53	7	0.711	0.940	2.257	101.4	0.204	0.209	0.297	0.121	0.170
						CoMFA Set 2						
2_1	F_Score	24	3	0.788	0.928	0.718	85.6	0.696	0.304			
2_2	**ChemScore**	**24**	**4**	**0.851**	**0.960**	**0.745**	**114.4**	**0.796**	**0.204**			
2_3	G Score	24	4	0.814	0.961	8.186	115.8	0.801	0.199			
2_4	D Score	24	5	0.786	0.985	2.391	233.3	0.872	0.128			
2_5	PMF Score	24	5	0.610	0.980	0.968	177.1	0.865	0.135			
						CoMSIA Set 2						
2_1	F_Score	24	1	0.790	0.831	1.045	108.5	0.126	0.378	0.144	0.144	0.208
2_2	ChemScore	24	5	0.901	0.979	0.552	170.2	0.408	0.105	0.371	0.048	0.068
2_3	G Score	24	4	0.868	0.953	8.973	95.6	0.353	0.082	0.344	0.075	0.146
2_4	D Score	24	5	0.850	0.962	3.794	90.5	0.374	0.085	0.351	0.071	0.119
2_5	PMF Score	24	3	0.690	0.896	2.095	57.6	0.293	0.106	0.275	0.189	0.136
						CoMFA Set 3						
3_1	F_Score	24	2	0.634	0.926	0.753	131.1	0.664	0.336			
3_2	ChemScore	24	4	0.911	0.996	0.312	1131.3	0.644	0.356			
3_3	G Score	24	3	0.806	0.975	5.379	265.4	0.723	0.277			
3_4	D Score	24	3	0.771	0.979	2.740	311.8	0.746	0.254			
3_5	PMF Score	24	4	0.787	0.996	0.568	1199.0	0.659	0.341			
						CoMSIA Set 3						
3_1	F_Score	24	3	0.674	0.947	0.654	118.3	0.120	0.272	0.136	0.232	0.240
3_2	**ChemScore**	**24**	**3**	**0.906**	**0.982**	**0.633**	**360.9**	**0.209**	**0.233**	**0.192**	**0.120**	**0.246**
3_3	G Score	24	4	0.759	0.982	4.780	253.6	0.248	0.207	0.208	0.080	0.257
3_4	D Score	24	4	0.766	0.982	2.629	254.8	0.243	0.214	0.228	0.068	0.247
3_5	PMF Score	24	4	0.865	0.995	0.642	935.4	0.124	0.205	0.135	0.324	0.212
						CoMFA Set 4						
4_1	F_Score	24	5	0.576	0.994	0.196	628.3	0.553	0.447			
4_2	ChemScore	24	4	0.875	0.996	0.249	1086.3	0.752	0.248			
4_3	G Score	24	4	0.853	0.993	3.648	650.6	0.705	0.295			
4_4	D Score	24	3	0.881	0.986	2.489	470.8	0.710	0.290			
4_5	PMF Score	24	3	0.642	0.973	1.275	240.5	0.530	0.470			
						CoMSIA Set 4						
4_1	F_Score	24	5	0.572	0.984	0.332	216.4	0.093	0.273	0.141	0.236	0.256
4_2	ChemScore	24	3	0.759	0.977	0.554	287.4	0.219	0.218	0.185	0.185	0.194
4_3	G Score	24	5	0.816	0.996	2.699	953.9	0.240	0.175	0.167	0.208	0.209
4_4	**D Score**	**24**	**5**	**0.862**	**0.994**	**1.744**	**579.9**	**0.192**	**0.176**	**0.154**	**0.238**	**0.240**
4_5	PMF Score	24	1	0.625	0.789	3.399	82.3	0.143	0.249	0.159	0.194	0.254

[a] Number of compounds. [b] Optimal number of PLS components. [c] Cross-validated r^2. [d] Non-cross-validated r^2. [e] Standard error of the estimate. [f] Fisher test value. [g] Contribution of the steric field (fraction). [h] Contribution of the electrostatic field (fraction). [i] Contribution of the hydrophobic field (fraction). [j] Contribution of the hydrogen bond donor field (fraction). [k] Contribution of the hydrogen bond acceptor field (fraction). [l] The best models are set in boldface.

using the corresponding score values as dependent variables. The best CoMSIA model with five PLS components and the highest q^2 value (0.862) was obtained with the D Score values ($r^2 = 0.994$) (Table 1, CoMSIA set 4_4). It should be noted that the large contributions of the hydrogen bond donor, hydrogen bond acceptor, and electrostatic molecular fields obtained for the CoMSIA models of set 4 (see Table 1) are in accordance with our original intention that this compound set was designed to search the receptor pocket around the **A** ring for hydrogen bond accepting and donating amino acid residues. Figure 5 shows the contour plot of the hydrogen bond donor field contributions ("StDev*Coeff") (60% contribution) obtained for the best CoMSIA model. The carbonyl groups of the Tyr415, Val416, Asp419, Leu500, and Gln 503 residues were identified as hydrogen bond acceptors interacting with the alcoholic hydroxyl group of the substituent on the **A** aromatic ring of 18 (out of all 24) docked conformers as donors. Among these, Val416 and Gln503 established the most significant interactions, Val416 forming a hydrogen bond with 7 docked conformers and Gln503 with 5 docked conformers. Hydrogen bonding interactions are shown with dashed lines. In Figure 5 18 ligands (i.e., those forming a hydrogen bond with one of the Tyr415, Val416, Asp419, Leu500, and Gln503 residues, in wire-frame model) as well as 8 amino acid residues of EcR (Thr343, Asn504, Tyr408, Tyr415, Val416, Asp419,

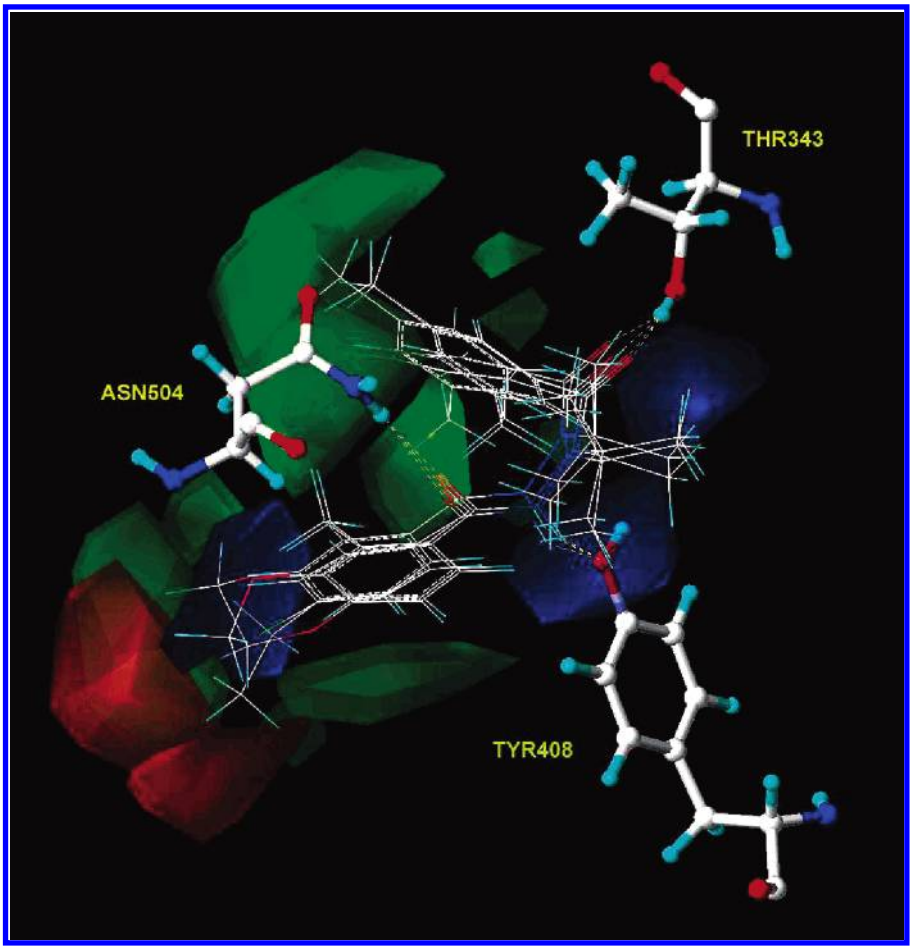


Figure 2. Contour plot of the steric and electrostatic field contributions obtained for the CoMFA model of set 1_3 using the G Score values as the dependent variable.

Leu500, and Gln503, in ball-and-stick model) are shown for illustration.

Figure 6 shows the contour plot of the hydrogen bond acceptor field contributions ("StDev*Coeff") (60% contribution) obtained for the CoMSIA model of set 4_4. The backbone amide moieties of Asp419, Leu420, and Met507 were identified as hydrogen bond donors interacting with the alcoholic hydroxyl group of the substituent on the **A** aromatic ring of 3 (out of all 24) docked conformers as acceptors. The hydrogen bonding interactions are shown with dashed lines. Twenty-one docked conformers did not establish hydrogen bond acceptor interaction with the EcR receptor. In Figure 6 three ligands (i.e., those forming a hydrogen bond with one of the Asp419, Leu420, and Met507 residues, in wire-frame model) as well as 6 amino acid residues of EcR (Thr343, Asn504, Tyr408, Asp419, Leu420, and Met507, in ball-and-stick model) are shown for illustration.

DISCUSSION

Since the discovery of DAHs their substitution pattern has been exhaustively explored, and several patent applications have been filed claiming potent DAH agonists in absence of the experimental EcR structure. Since 2003 the availability of the experimental EcR structure (pdb code: 1R20) has opened up new possibilities for designing novel DAH agonists in a more rational manner. The proposed method of correlating the receptor docking score values of DAHs with CoMFA and CoMSIA molecular field interaction contributions allows the mapping of hitherto unexplored regions of the EcR pocket in terms of molecular fields by employing appropriately designed ligand sets. In this study, first the registered DAHs **2**−**5** were docked into the EcR yielding 53 successfully docked conformations (set 1). The pharmacophore constraint defined in FlexX-Pharm forced the docked conformations to adopt positions locked by three highly conserved hydrogen bonding interactions between the DAH analogues and the receptor pocket. The best CoMFA model was obtained using the G Score values (set 1_3) (Table 1, Figure 2). The CoMFA model of set 1_3 is highly significant, explaining 95.6% of the variance in the G Score values; still the model is not very informative due to the low structural variation in the altogether 53 docked conformations of the registered DAHs **2**−**5**. Thus, the CoMFA molecular field contributions calculated and visually presented in Figure 2 show mainly the conformational variability of the docked DAHs.

It is apparent that the proposed method is suitable for searching novel receptor−ligand binding possibilities in the DAH binding pocket of EcR by systematically substituting the diacyl-hydrazine skeleton with diverse substituents and docking the ligand set so obtained. In this study we restricted our search to the region around the **A** ring of the DAHs, while maintaining the dibenzoyl-hydrazine scaffold. First the boundary of the DAH binding pocket of EcR was explored by docking 24 DAH analogues of type **6** substituted with a

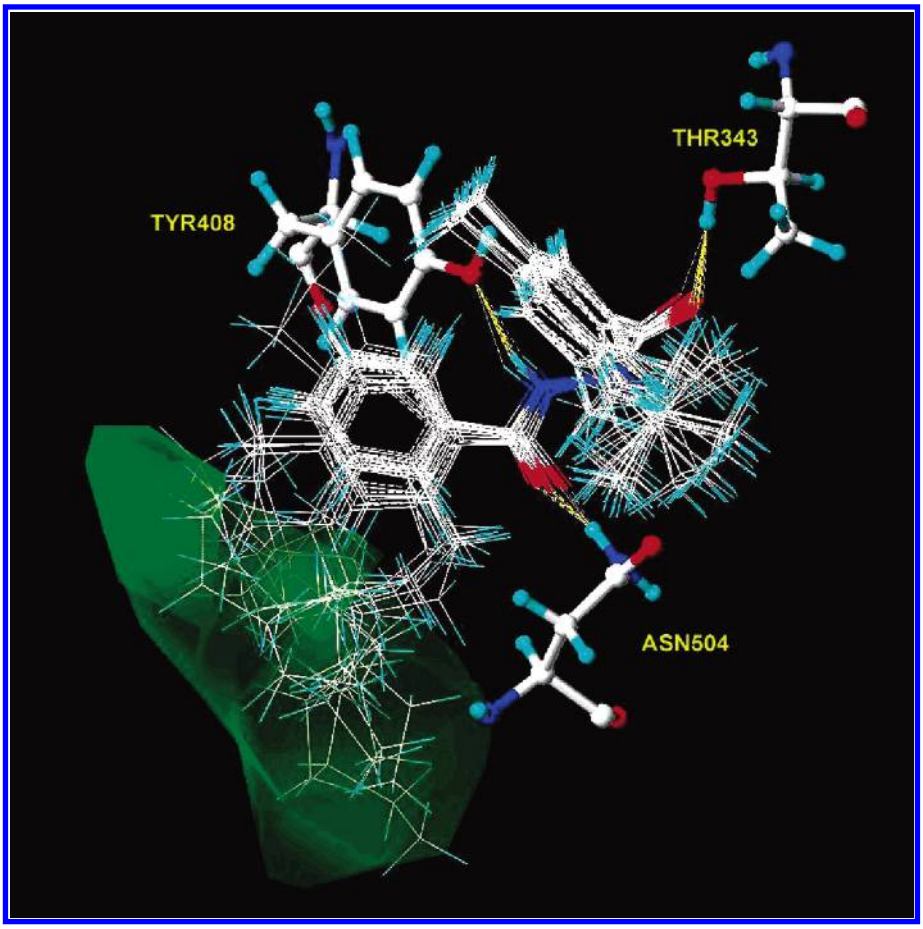


Figure 3. Contour plot of the steric field contributions obtained for the CoMFA model of set 2_2 using the ChemScore values as the dependent variable.

C_1-C_8 normal alkyl group in the ortho-, meta-, or para-position of the **A** ring (set 2) into the EcR. The best CoMFA model was obtained using ChemScore values as the dependent variable, explaining 96.0% of the variance in the score values (set 2_2) (Table 1, Figure 3). The model reveals the presence of a receptor channel leading to the surface of the protein, capable of accommodating long substituents in the ortho-, meta-, or para-position of the **A** aromatic ring (Figure 3). The consequences of simultaneous substitution of several positions on the ring could also be explored by using appropriate combinatorial libraries to obtain a more detailed picture of the available space for substituents in the region around the **A** ring.

In the next step, amino acid residues capable of hydrogen bond donor interactions within the identified receptor channel were searched for by docking 24 DAH analogues of type **6** carrying a substituent of general structure $-(CH_2)_nOCH_3$ ($n = 0-7$) in the ortho-, meta-, or para-position of the **A** ring (set 3). The ether oxygen at the end of the substituent is capable of establishing hydrogen bonds with the receptor-binding site. CoMSIA of the 24 successfully docked DAHs and the ChemScore values as the dependent variable yielded the best model, explaining 98.2% of the variance in the ChemScore values (set 3_2) (Table 1, Figure 4). In Figure 4 the CoMSIA isoenergy contour (magenta) highlights the region where the ether oxygens of 15 docked conformers are located as proton acceptors interacting with the side-chain amide group of Gln503. The 15 docked conformers were those members of set 3 that contain a long enough $-(CH_2)_nOCH_3$ substituent on the **A** ring with $n > 2$. Thus novel hydrogen bonding interaction possibilities, readily available for the substituents on the **A** ring, have been located. Each hydrogen bond enhances the receptor binding affinity of the ligand significantly, thus introduction of structural moieties capable of interacting with the side-chain amide group of Gln503 might be considered in the design of novel EcR agonists showing high docking score values.

In the next step, CoMFA and CoMSIA were employed to identify amino acid residues of EcR capable of interacting with the docked DAH derivatives as hydrogen bond acceptors or donors. For this purpose a small library of 24 DAHs of type **6** was built. In the 24 DAHs the substituent on the **A** ring was a $-(CH_2)_nOH$ group ($n = 0-7$) in the ortho-, meta-, or para-position (set 4). The best CoMSIA model of the 24 successfully docked DAHs using D Score values as the dependent variable explained 99.4% of the variance in the D Score values (set 4_4) (Table 1, Figures 5 and 6). Figure 5 shows the hydrogen bonding interactions of the alcoholic hydroxyl groups of 18 docked conformers acting as proton donors with the carbonyl groups of Tyr415, Val416, Asp419, Leu500, and Gln503 as proton acceptors. Among these interactions those of the carbonyl group of Val416 and Gln503 were found to be the most significant (cyan contour) as Val416 established a hydrogen bond with 7 docked conformers and Gln503 with 5 docked conformers. Some of the cyan contour regions in Figure 5 are apparently not related to existing hydrogen bonds.

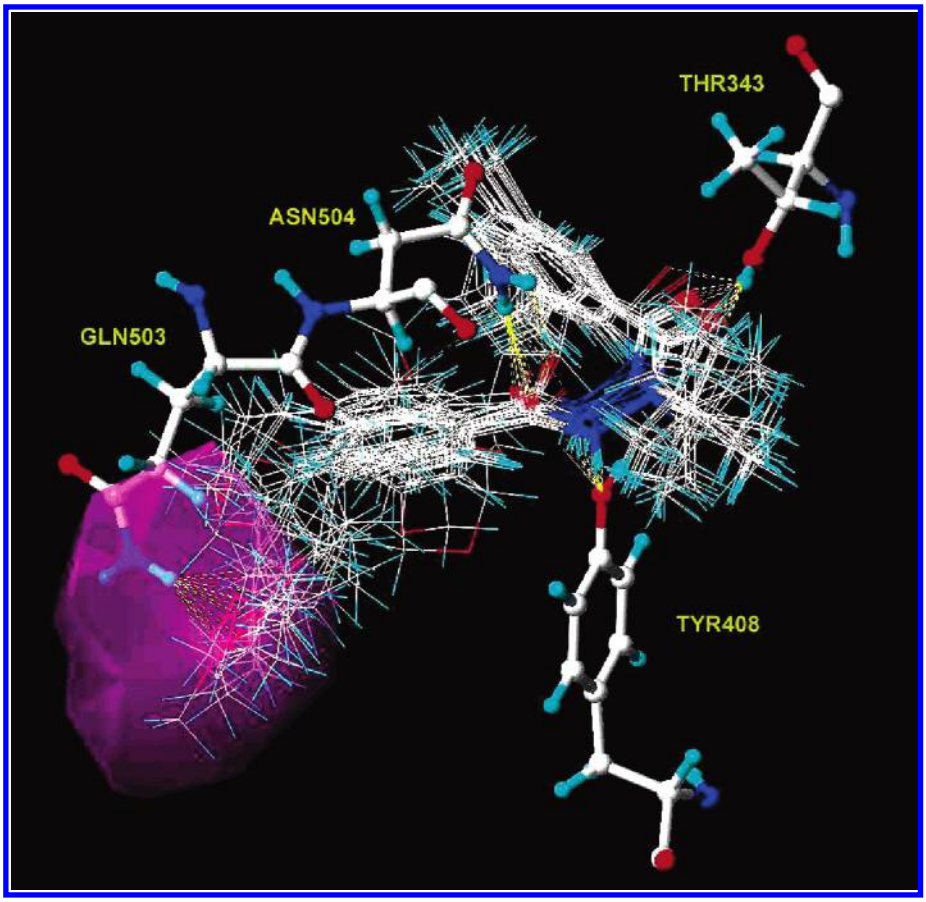


Figure 4. Contour plot of the hydrogen bond acceptor field contributions obtained for the CoMSIA model of set 3_2 using the ChemScore values as the dependent variable.

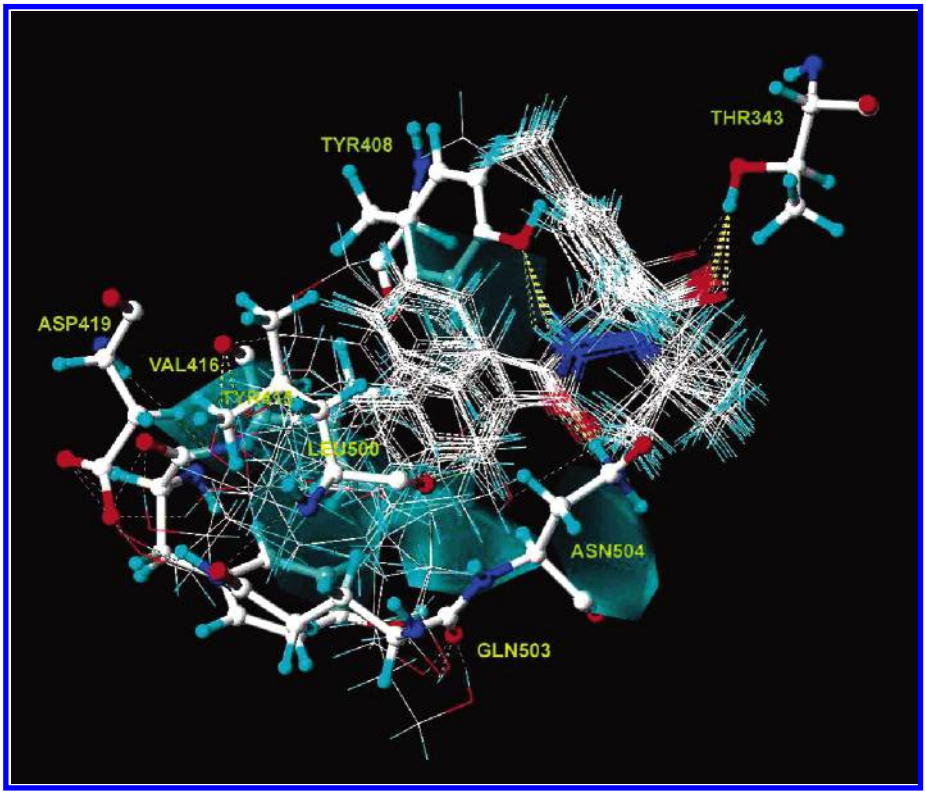


Figure 5. Contour plot of the hydrogen bond donor field contributions obtained for the CoMSIA model of set 4_4 using the D Score values as the dependent variable.

Based on the same CoMSIA model of set 4_4, Figure 6 shows the hydrogen bonding interactions of the alcoholic hydroxyl groups of 3 docked conformers acting as proton acceptors with the backbone amide moieties of Asp419,

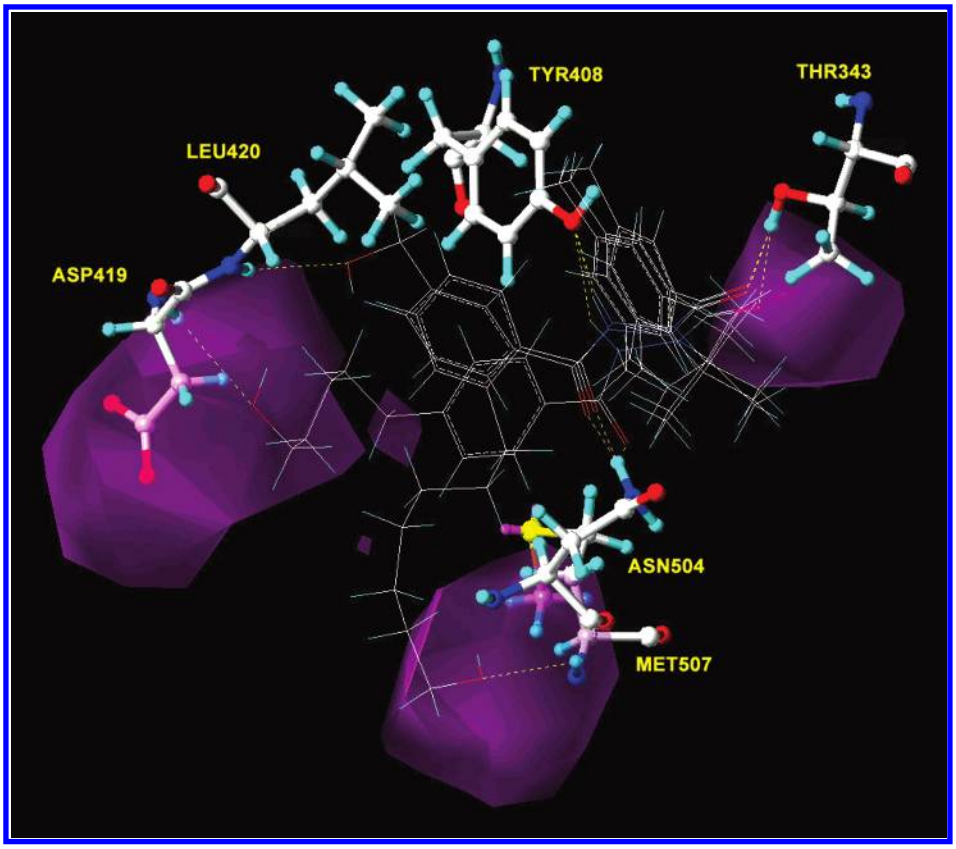


Figure 6. Contour plot of the hydrogen bond acceptor field contributions obtained for the CoMSIA model of set 4_4 using the D Score values as the dependent variable.

Leu420, and Met507 as proton donors. It should be noted that relatively few, only 3, docked conformers established hydrogen bond donor interactions with the EcR receptor.

The ligand-based CoMFA and CoMSIA methods and the receptor-based FlexX and FlexX-Pharm docking methods are all well established modeling approaches. The finding that in many cases a major portion of the variance in a selected docking score can be explained by molecular field interaction contributions mutually validates both modeling concepts. This is an important insight, as the selection of the proper scoring function is often a difficult problem. Regarding the correlation between docking score values and biological activity values only crude approximations are available.[4]

Ligand- and receptor-based design methods are often combined to give a more informative picture of the receptor−ligand interactions. For example, the ligand structures for CoMFA calculations can be aligned on the basis of the available X-ray crystal structure of a receptor−ligand complex, and the CoMFA isoenergy contours can be mapped onto the model of the receptor−ligand complex, if it is available, to highlight receptor−ligand interactions in terms of molecular fields. Both ligand- and receptor-based design methods have their inherent advantages and drawbacks. In case of the ligand-based CoMFA and CoMSIA methods two potential sources of errors are the biological activity data (e.g., ADME effects, or experimental errors of the measurements) and the methodology itself (e.g., incorrect ligand alignment based on a hypothetical receptor bound conformation of the template structure and the linear nature of the PLS analysis). In case of the receptor-based methods, important sources of errors are the quality of the experimental or homology receptor model used, the approximate nature of the docking algorithm (e.g., rigid or flexible docking), and the quality and appropriateness of the scoring function. In our combined approach some of these error sources have been effectively eliminated (e.g., ADME effects, incorrect ligand alignment, etc.). The novel feature of the proposed approach is that it allows the identification of significant receptor−ligand interactions in 3D in terms of molecular field interactions from docking score values.

CONCLUSIONS

The immediate objective of an in silico screening project is to find high scoring compounds in molecule libraries. We have shown that in most cases a highly significant portion of the variance in the receptor docking score values of ligands can be interpreted in terms of 3D molecular field interaction contributions using CoMFA and CoMSIA methods and the docking score values as dependent variables. This has been demonstrated for FlexX and CScore docking score values using the EcR and small virtual molecule libraries of DAH agonists yielding highly significant CoMFA and CoMSIA models. The existence of such highly significant 3D QSAR models validates the underlying rationale of both the CoMFA type ligand-based and the FlexX type receptor-based approaches that are both established molecular modeling methodologies. An advantage of the proposed approach is that CoMFA and CoMSIA can be performed for any docked molecule set using docking score values as dependent variables and no experimental biological activity data are required. By docking appropriately designed virtual molecule libraries the receptor can be exhaustively explored to detect new receptor−ligand interaction sites. The generated CoMFA and CoMSIA isoenergy contour plots can be mapped onto

the docked receptor−ligand complex model, thus visualizing the significant receptor−ligand interaction sites. Our approach is a useful tool for use in in silico screening of molecule libraries to search the target receptor for receptor−ligand interaction sites that are significant in terms of molecular field interactions. This information may help the search for novel, high scoring compounds that interact with the mapped receptor site.

ACKNOWLEDGMENT

This study was financially supported by the Hungarian Research Fund (OTKA) No. K 67737 and the Economic Competitiveness Operative Program, Project No. GVOP-3.2.1.-2004-04-0068/3.0.

REFERENCES AND NOTES

(1) Schwardt, O.; Kolb, H.; Ernst, B. Drug discovery today. *Curr. Top. Med. Chem.* **2003**, *3*, 1−9.
(2) Barril, X.; Hubbard, R. E.; Morley, S. D. Virtual screening in structure-based drug discovery. *Mini-Rev. Med. Chem.* **2004**, *4*, 779−791.
(3) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.
(4) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Rev. Drug Discovery* **2004**, *3*, 935−949.
(5) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 235−249.
(6) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. J. Comparative Molecular-Field Analysis (CoMFA).1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(7) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indexes in a comparative-analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.
(8) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; Delazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83−102.
(9) *Sybyl 7.1*; Tripos Inc.: St. Louis, MO.
(10) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(11) Hindle, S. A.; Rarey, M.; Buning, C.; Lengauer, T.: Flexible docking under pharmacophore type constraints. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 129−149.
(12) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Modell.* **2002**, *20*, 281−295.
(13) Muegge, I.; Podlogar, B. L. 3D-quantitative structure activity relationships of biphenyl carboxylic acid MMP-3 inhibitors: Exploring automated docking as alignment method. *Quant. Struct.-Act. Relat.* **2001**, *20*, 215−222.
(14) Billas, I. M. L.; Iwema, T.; Garnier, J. M.; Mitschler, A.; Rochel, N.; Moras, D. Structural-adaptability in the ligand-binding pocket of the ecdysone hormone receptor. *Nature* **2003**, *426*, 91−96.
(15) Martin, Y. C.; Kim, K.-H.; Lin, C. T. Comparative Molecular Field Analysis: CoMFA. In *Advances in Quantitative Structure−Property Relationships*; Charton, M., Ed.; JAI Press: Greenwich, CT, 1996; Vol. 1, pp 1−52.
(16) Yanagi, M. Development of a novel lepidopteran insect control Agent, Chromafenozide (Matric), and Prospect of Success. *Agrochem. Jpn.* **2000**, *76*, 16−18.
(17) Retnakaran, A.; Krell, P.; Reng, Q.; Arif, B. Ecdysone Agonists: Mechanism and Importance in Controlling Insect Pests of Agriculture and Forestry. *Arch. Insect Biochem. Physiol.* **2003**, *54*, 187−199.

CI600317V