New Approach to Molecular Docking and Its Application to Virtual Screening of Chemical Databases

Carol A. Baxter,* Christopher W. Murray, Bohdan Waszkowycz, Jin Li, Richard A. Sykes, Richard G. A. Bone, Tim D. J. Perkins, and William Wylie

Protherics Molecular Design Ltd., Beechfield House, Lyme Green Business Park, Macclesfield, Cheshire, SK11 0JL, U.K.

Received July 28, 1999

This paper describes the validation of a molecular docking method and its application to virtual database screening. The code flexibly docks ligand molecules into rigid receptor structures using a tabu search methodology driven by an empirically derived function for estimating the binding affinity of a protein−ligand complex. The docking method has been tested on 70 ligand−receptor complexes for which the experimental binding affinity and binding geometry are known. The lowest energy geometry produced by the docking protocol is within 2.0 Å root mean square of the experimental binding mode for 79% of the complexes. The method has been applied to the problem of virtual database screening to identify known ligands for thrombin, factor Xa, and the estrogen receptor. A database of 10 000 randomly chosen "druglike" molecules has been docked into the three receptor structures. In each case known receptor ligands were included in the study. The results showed good separation between the predicted binding affinities of the known ligand set and the database subset.

INTRODUCTION

Molecular docking is an important technique in structure-based drug design, providing an estimate of the binding mode and in some applications the binding affinity of a ligand−receptor complex. With an increasing number of known protein crystal structures available, the interest in molecular docking is high and much progress has been made in recent years.[1−6]

Early docking methods considered only translational and orientational degrees of freedom of the ligand with respect to the receptor.[7] However, this is an oversimplification as the molecule's geometry may change upon binding into the active site, so now most methods treat the ligand as being flexible. As allowing for ligand flexibility increases the number of degrees of freedom, soon the size of conformational search space becomes enormous and a highly efficient search procedure is required. Heuristic searching methods such as simulated annealing (SA),[8,9] evolutionary programming (EP),[10,11] and genetic algorithms (GA)[12−16] have all been applied to the docking problem. These methods produce accurate predictions of binding modes, but they are computationally expensive and unsuitable for docking large numbers of ligands. More efficient searches which have been successfully applied to database screening include incremental construction methods[17] used by FLEXX[18] and Hammerhead,[19] a fast shape matching algorithm similar to DOCK[26] with multiple conformations for the screened molecules,[20] and highly specialized sampling techniques such as QXP.[21]

An important feature of any docking method is an energy function that is capable of predicting binding modes. The minimum value of the function should correspond to the preferred binding mode(s) of the ligand. If a function scores a noncrystallographic ligand geometry the same as or better than the crystallographic one, its ability to identify the correct binding mode is questionable. In many applications an estimate of binding affinity[22−24] is used as the objective function to be minimized in the docking procedure.[18,19]

Virtual screening involves a number of different computational techniques that together can reduce the size of a virtual compound collection to a more manageable size.[25] In this paper we describe the use of molecular docking as a virtual screening tool. In this application a scoring function must be able to identify correctly docked structures and prioritize hits. Typically in a database screen a large number of molecules will successfully dock to the receptor's binding site. The job of the scoring function is to provide a ranking order of the molecules so that further analysis can be concentrated on the best scoring hits. Recently, Böhm[23] proposed a new, quick to evaluate function which is suitable for scoring thousands of ligand−receptor structures and has been implemented in the flexible docking algorithm FLEXX.[18] Other docking methods which have been successfully applied to the problem of database screening include DOCK,[26] which screened 10% of the Available Chemicals Directory (ACD) against dihydrofolate reductase (DHFR) and found 7 of the top 13 hits to be dihydrofolate or methotrexate derivatives. In similar studies, FLOG[20] picked out known DHFR inhibitors from a 7636 compound database and Hammerhead[19] screened 80 000 compounds from the ACD for binding to streptavidin and predicted the natural inhibitor biotin to be the top scoring hit.

Here we present the validation of the latest version of the automated docking program PRO_LEADS and its application to three-dimensional database screening. A detailed

* To whom correspondence should be addressed. E-mail: carol.baxter@protherics.com.

10.1021/ci990440d CCC: $19.00
Published on Web 01/26/2000

description of the algorithm has been presented elsewhere.[27] Briefly, PRO_LEADS flexibly docks ligand molecules into a rigid receptor. The extent to which the assumption of a rigid receptor biases the ability to dock a ligand and estimate its binding affinity has previously been investigated.[28] For a relatively rigid receptor robust results for the prediction of binding mode and binding affinity were obtained, although small changes in the receptor structure were observed to have a detrimental effect on the quality of the results.

A tabu search (TS) algorithm is used to sample the conformational space. The TS has been found to perform well in this application.[29] The modern form of the TS[30] imposes restrictions to prevent the search revisiting previous solutions and thus encouraging the exploration of new search space.

A simple empirically derived scoring function[31] is used to drive the search and estimate binding affinity. The function was optimized against a large training set of 82 protein−ligand complexes taken from the Protein Data Bank (PDB)[32] for which the binding affinities are known. The function reproduced the binding affinities of the complexes with a cross-validated error of 8.68 kJ/mol and performed well on a test set of 30 complexes. The optimized scoring function was augmented for application to PRO_LEADS by the addition of terms deemed to be important to molecular docking, e.g., a term to take into account ligand−receptor clashes. The energy evaluation is performed using a grid-based methodology. Implementation details of the modified scoring function have been reported elsewhere.[27]

In this paper, the most recent version of PRO_LEADS has been validated on a set of 70 protein−ligand complexes. Twenty complexes in the test set have not previously been studied with PRO_LEADS. The code has also been used to screen a subset of 10 000 molecules from the ChemBridge Prime Database[33] for binding to thrombin, factor Xa, and the estrogen receptor ligand binding domain (ER). Known receptor ligands were added to the 10 000 molecules to test the ability of PRO_LEADS to identify known ligands from a large set of randomly chosen molecules.

METHODS AND MATERIALS

PRO_LEADS allows the ligand full translational and orientational freedom subject to the constraint that no heavy atom moves outside a user-specified grid box defining the active site. The ligand is considered flexible through a list of rotatable bonds. Generally most sp^3-sp^3 and sp^3-sp^2 bonds are considered rotatable, and some sp^2-sp^2 bonds are defined rotatable where there is reason to expect that more than one conformer exists. The enzyme structures are treated as rigid and all water molecules are removed.

Tabu Search. The tabu search method implemented in PRO_LEADS maintains only one solution during the course of a search. For each docking experiment 1000 random starting positions of the ligand in the active site are generated and stored. The random solutions are scored using the energy function and ranked in order of energy. At the start of each run, and whenever a random start is required, a new solution is taken from the store and the lowest energy solution is taken first.

From the current solution, a user-defined number of "moves" is generated by a mutation-like procedure in which Cauchy random numbers are added to each of the docking variables. Each of the moves is then scored using the energy function and they are ranked in order, with the best energy move at the top of the rank. The TS maintains a tabu list, which stores a number of previously visited solutions. The moves are examined in rank order, and a move is considered "tabu" if it generates a solution that is not sufficiently different from the solutions stored. The threshold measure used in this work to determine whether a move is tabu is a root-mean-square (rms) deviation (measured over selected heavy atoms) of 0.75 Å or less between the two solutions being compared. The top ranking move will always be accepted, even if it is tabu, if its energy is lower than the solution with the lowest energy so far. Otherwise the algorithm chooses the lowest energy nontabu move. The accepted move becomes the new current solution and is added to the tabu list. Once the list is full the current solution replaces the solution with the longest residence in the tabu list. Once a new solution has been identified and stored, a new set of moves is generated from it and the search procedure continues with the next iteration.

A further mechanism which has been found to help the search escape from local minima is to restart the search from random if it is observed that the best solution has not changed after 100 iterations.

The tabu search continues for a user-defined number of iterations, after which it terminates and returns the best solution found. The best solution is also subjected to a local minimization using the Simplex algorithm.[34]

Energy Evaluation. In our earlier work, we developed an empirical scoring function which, given the geometry for a ligand−receptor complex, estimates the free energy of binding.[31] The empirical scoring function uses simple contact terms to estimate lipophilic and metal−ligand binding contributions, a simple explicit form for hydrogen bonds, and a term which penalizes flexibility. The function can be written in the form

$$\Delta G_{\text{binding}} = \Delta G_0 + \Delta G_{\text{hbond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lipo}} + \Delta G_{\text{rot}}$$

The form of the separate terms has been explained elsewhere.[31] The function has been trained using multiple linear regression on a set of 82 complexes for which the binding affinity is known experimentally. The robustness of the equation was tested against test sets comprising a further 30 ligand−receptor complexes and by using extensive cross-validation on the original training set. The function has a leave-one-out cross-validated q^2 of 0.658 and s_{press} value of 8.68 kJ/mol.

The ChemScore function has been extended to make it more applicable to docking.[27] Briefly, an elongated hydrogen bond function is employed during the beginning of the tabu search, in an attempt to compete with the long-range lipophilic term. The form of the hydrogen bond function has also been modified to include a second angle dependence to inhibit the generation of unrealistic ligand−receptor contacts. Terms to calculate the ligand internal energy, to penalize heavy atom clashes between the ligand and the receptor, and to penalize solutions that lie outside a user-defined active site have been included in the function. The addition of the new terms did not significantly affect the robustness of the ChemScore regression equation, and recalibration was not necessary.

Table 1. Details of the 70 Complexes in the Test Set and the Results of the Protocol 2 Docking Experiment[a]

PDB code	receptor name	no. of ligand rotatable bonds	exptl binding affinity (kJ/mol)	best docking energy (kJ/mol)	rms (Å)	PDB code	receptor name	no. of ligand rotatable bonds	exptl binding affinity (kJ/mol)	best docking energy (kJ/mol)	rms (Å)
1ABE	ABP	4	−40.33	−29.33	0.22	1QBT[b]	HIV-1 protease	10	−60.59	−66.10	0.67
1ABF	ABP	4	−30.92	−31.36	0.36	1QBU[b]	HIV-1 protease	10	−58.40	−54.92	0.72
1AOE[b]	DHFR	3	−55.10	−35.73	0.67	1RBP	retinol binding protein	2	−38.35	−43.37	1.09
1APU	penicillopepsin	16	−43.92	−28.43	8.25	1SLN[b]	stomelysin-1	13	−37.87	−36.30	6.77
1CBX	CPA	5	−36.21	−41.90	0.87	1STP	streptavidin	5	−71.44	−40.26	0.57
1DBB[c]	DB3	1	−51.35	−40.84	0.73	1TLP	thermolysin	12	−43.09	−42.29	1.76
1DBJ[c]	DB3	1	−43.80	−42.48	0.57	1TMN	thermolysin	13	−41.65	−48.63	1.77
1DMP[b]	HIV-1 protease	10	−54.50	−59.63	0.59	1TNG	trypsin	2	−16.73	−22.08	0.45
1DOG	glucoamylase II 471	5	−22.92	−27.00	3.80	1TNH	trypsin	2	−19.21	−21.50	1.11
1DWB[c]	thrombin	0	−16.66	−22.36	0.50	1ULB[c]	PNP	0	−11.71	−22.05	0.42
1EBG	enolase	2	−61.75	−44.73	1.57	1UVS[b]	thrombin	9	−30.80	−30.81	4.07
1EPO	endothiapepsin	18	−45.40	−43.60	1.34	1UVT[b]	thrombin	4	−43.58	−40.80	1.08
1ETR	thrombin	8	−42.21	−37.40	0.77	2CGR	immunoglobulin	5	−41.51	−35.55	0.70
1ETS	thrombin	7	−48.62	−35.75	1.19	2CPP	cytochrome p450cam	0	−34.63	−32.17	1.25
1ETT	thrombin	5	−35.30	−35.79	4.20	2CTC	carboxypeptidase A	4	−22.17	−31.04	0.52
1FAX[b]	factor Xa	7	−42.14	−36.02	1.22	2GBP	GBP	6	−43.36	−34.42	0.40
1HPV	HIV-1 protease	10	−52.61	−34.54	9.19	2H4N[b]	carbonic anhydrase II	1	−49.63	−23.71	1.18
1HSL	histidine-binding protein	4	−41.65	−30.55	0.46	2IFB	FABP	14	−30.99	−42.13	2.15
1HTF	HIV-1 protease	13	−46.18	−44.42	3.75	2PHH[c]	PHBH	1	−26.69	−31.56	1.64
1HVR	HIV-1 protease	10	−54.25	−66.70	0.91	2R04[c]	virus coat protein	9	−35.49	−42.34	1.14
1JAO[b]	neutrophil collagenase	8	−33.78	−36.06	1.74	2TMN	thermolysin	5	−33.58	−35.62	0.84
1JAP[b]	neutrophil collagenase	8	−26.93	−24.74	1.26	2TSC	thymidylate synthase	9	−48.62	−30.33	3.00
1MBI	myoglobin	0	−10.70	−21.65	0.49	2YPI	TP isomerase	3	−27.52	−23.16	0.38
1MMB[b]	neutrophil collagenase	13	−52.61	−42.75	1.80	3PTB	trypsin	0	−27.04	−25.81	0.41
1MNC	neutrophil collagenase	10	−51.35	−37.78	0.65	3TPI[b]	trypsinogen	7	−24.59	−30.06	0.53
1MTW[b]	trypsin	7	−31.35	−38.63	0.82	4DFR	DHFR	8	−55.33	−32.55	4.89
1NSD	neuraminidase	8	−30.24	−27.55	1.43	4TPI[b]	trypsinogen	6	−16.28	−27.58	1.29
1OKL[b]	carbonic anhydrase II	2	−34.41	−36.96	0.88	5ABP	ABP	6	−37.87	−31.34	0.31
1OKM[b]	carbonic anhydrase II	6	−33.16	−31.07	2.47	5CPP	cytochrome p450cam	0	−33.58	−32.27	1.22
1PGP	6-PGDH	11	−32.51	−23.03	5.96	5TLN	thermolysin	9	−36.32	−30.39	1.97
1PHF	cytochrome p450cam	0	−25.09	−29.38	0.59	6CPA	CPA	11	−65.74	−48.11	1.46
1PHG	cytochrome p450cam	2	−49.39	−37.57	1.27	7CPA	CPA	14	−79.64	−48.79	3.22
1PPC	trypsin	7	−36.83	−28.18	8.54	DFR4	DHFR	0	−34.23	−22.94	0.72
1PPH	trypsin	5	−35.50	−36.61	0.46	TMT1	thrombin	11	−39.70	−31.81	8.52
1QBR[b]	HIV-1 protease	10	−60.29	−56.56	0.95	TSC2	thymidylate synthase	4	−30.80	−33.83	0.61

[a] The lowest docking energy solution found is reported. A heavy atom rms (Å) is measured between the solution and the ligand crystal geometry. A rms < 2.0 Å is considered to be an acceptable estimate of binding mode. [b] Complex not used in work to derive scoring function. [c] Present in 30 complex set used to test scoring function.

For computational speed, lipophilic interactions and ligand−receptor clashes are precalculated and stored on two grids covering the active site. The value at a point in the grid is the sum of the term over receptor atoms when a ligand atom is placed at that grid point. Because the hydrogen bond interaction term has an angle dependency, it cannot be treated in the same way as the above terms and instead "neighbor lists" are used. In a neighbor list each grid point stores not an energy value but a list of indices of receptor atoms that are capable of forming a hydrogen bond with a complementary ligand atom if it were to be placed on the grid point. The hydrogen bond term is then determined by performing an explicit calculation for each receptor atom stored in the list. The neighbor list approach is exact and computationally efficient because the hydrogen bond term is short range.

VALIDATION EXPERIMENT

Test Set. A test set of 70 protein ligand complexes taken from the Brookhaven Databank[32] for which the binding affinity and binding geometry are known has been constructed. It contains only small noncovalently bound ligands of the type which de novo design or database searching might be expected to produce. We have excluded ligands with large cycles, large peptide ligands, and ligands containing several sugar monomers. It was decided to use only PDB structures, so the database contains no docked structures or structures from proprietary or privileged sources. Of these 70 complexes, 24 were not present in the original training set of 82 complexes used to calibrate the scoring function. Three additional structures (TMT1, TSC2, and DFR4) were obtained by removing part of the ligand in a PDB structure (1TMT, 2TSC, and 4DFR, respectively) as described previously.[31] The test set includes a variety of protein classes, and the ligand size ranges from 0 to 18 rotatable bonds and from 5 to 46 heavy atoms. Table 1 gives the PDB entry, the protein class, and the experimental binding affinity (kJ/mol) for each protein−ligand complex in the test set. The preparation of the complexes has been described elsewhere.[31] Because there is no explicit treatment of receptor flexibility and the scoring function requires the positions of the hydrogen atoms on the receptor, it is necessary to obtain good hydrogen atom positions for the receptor. Briefly, the crystal structures were extracted from the PDB, hydrogen atoms were added using InsightII,[35] and the ligands and receptors had atom types and partial charges assigned using the CVFF force field. The complexes were minimized with Discover[35] using tethers with large force constants to keep heavy atom positions close to the crystallographically

Table 2. Summary of Results for the 70 Complex Test Set Using Five Docking Protocols[a]

protocol	no. of tabu iterations	no. of tabu moves	no. of repeats	% of complexes with rms < 2.0 Å	% of complexes with rms < 1.5 Å	average time/ complex (min)	prediction of binding affinity	
							s (kJ/mol)	r^2
1	2000	100	100	84	76	55.2	9.4	0.53
2	2000	100	10	79	69	5.78	9.9	0.47
3	1000	100	10	67	63	3.04	12.1	0.22
4	1000	100	5	63	57	1.58	12.9	0.12
5	1000	50	5	60	56	0.94	13.8	0.00

[a] Each docking protocol involves a different number of scoring function evaluations. Details of the tabu search can be found in the text. A heavy atom rms (Å) is measured between the ligand crystal geometry and the solution found by the tabu search with the lowest docking energy. The times reported are in CPU minutes and all calculations were performed on one processor of an Origin2000.

determined positions. All water molecules were then removed from the structures. The ligands were separated from the receptor and minimized in vacuo using the CVFF force field in order to obtain unbiased geometries. Rotatable bonds were automatically assigned by a procedure based on recognition of chemical substructures using an in-house implementation of the SMILES[36] language. Before every docking run, all docking variables, rigid body variables, and flexible-ligand torsions were randomized.

Selection Criteria. In judging the performance of the method, we have concentrated mainly on the following three criteria.

1. A comparison between the lowest energies produced by the docking runs and the experimental binding affinities is made. The relevant quantities are the correlation coefficient, r^2, which reflects how well the trends in the activity are predicted, and the standard deviation *s*, which reflects how accurately any one value is predicted. If a docking method is to be applied to the problem of database searching, it must be capable of producing a ranked order of the molecules so that only the most favorable will be investigated. Prediction of binding affinity is a good indication that the ranking will be accurate.

2. The lowest energy solution produced by the program is referred to as the predicted conformation, and should be as close as possible to the crystallographic solution. In our earlier work we have used a heavy atom rms of within 1.5 Å as a definition of an acceptable solution. To compare results with the work of others, we also quote success rates for a heavy atom rms of less than 2.0 Å. If the lowest energy solution differs significantly from the crystallographic geometry, then the function will be unreliable at ranking different ligands in the same binding site, which will produce a large number of false positives in a screening application.

3. The time taken to complete the docking is considered. A slow docking method will not be applicable for database screening.

We believe the test set and the selection criteria form a good assessment of the performance of the program in both de novo design and database screening applications.

RESULTS

PRO_LEADS docked each minimized ligand into its receptor structure using five different docking protocols. The number of tabu iterations, the number of tabu moves, and the number of repeats, with each repeat starting from a random position, were varied. The docking protocols and results are summarized in Table 2. The binding mode prediction rate, the binding affinity prediction statistics, and the average time taken to dock one complex are also reported. The experiments were all performed on one processor (R10,000, 250 MHz) of an SGI Origin2000, and the times are quoted in CPU minutes.

Estimation of Binding Mode. In Table 2 the binding mode prediction rate is shown for two measures of success, i.e., within 2.0 Å and within 1.5 Å rms. Using the second docking protocol, which took on average 5.8 CPU min to dock one complex, the binding mode of 79% of the complexes was correctly predicted (when 2.0 Å was used as a measure of success). Using a rms of 1.5 Å, the percentage of complexes correctly predicted fell to 69%. The heavy atom rms of the 70 predicted conformations calculated using docking protocol 2 are reported in Table 1. The binding mode prediction rate degraded steadily as the amount of docking operations fell. With the fastest protocol, 60% of the complexes were correctly predicted and the docking took on average less than a minute per complex.

It is interesting to note that the ligands that are missed on the faster protocols tend to be the more flexible ligands. The average number of rotatable bonds per ligand across the 70 complex training set is 6.4. For the set of ligands that were correctly docked using protocol 2, but incorrectly docked using protocol 5, the average number of rotatable bonds per ligand is 10.1. At the short docking protocol the amount of conformational searching performed is insufficient and the global energy minima of large flexible molecules can be missed.

Estimation of Binding Affinity. The predicted binding energies of the 70 complexes using docking protocol 2 are given in Table 1, and have been plotted against the experimental binding affinities, Figure 1. The complex 1STP, streptavidin−biotin, is an outlier in the regression of the original scoring function,[31] so it was expected to be an outlier in this work. After omission of 1STP, the binding affinity prediction has a standard deviation, *s*, of 9.95 kJ/mol and a r^2 of 0.47. One other severe outlier can be identified in Figure 1. The binding affinity of 7CPA is poorly predicted because the correct binding mode has not been identified. With the omission of 1STP and 7CPA the standard deviation improves to 9.30 kJ/mol and the r^2 value improves to 0.54. The scoring function predicts the binding affinity of the 24 complexes not present in the original training set with a standard deviation of 8.66 kJ/mol and a r^2 of 0.59. One can therefore conclude that the original fitting process has not significantly biased the results. As with the prediction of binding mode, the prediction of the binding affinity degrades as the number of docking operations is reduced.

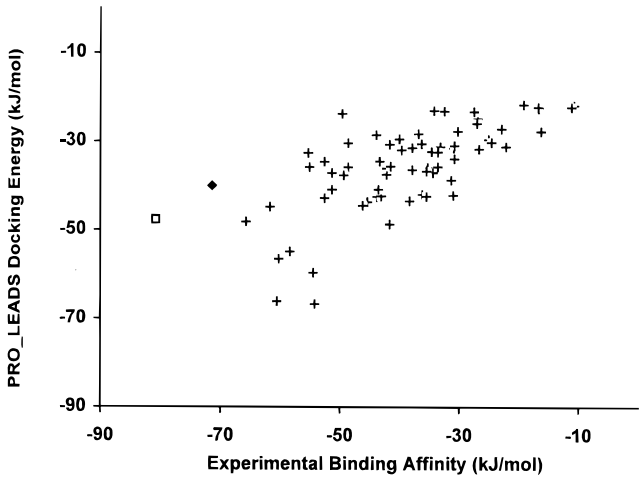

Figure 1. Scatter plot of predicted binding affinities for the 70 complex test set versus experimental binding affinities in kJ/mol. The predicted values were obtained using the PRO_LEADS scoring function. Outlier complex 1STP is shown as ◆. 7CPA is also shown, □.

Database Screening Experiment. The use of PRO_LEADS to screen a large virtual data set of "druglike" molecules (i.e., those with physicochemical properties representative of bioactive molecules[37]) is examined in three docking experiments against thrombin, factor Xa, and ER.

In each experiment a small set of ligands known to bind to the receptor was added to a subset of 10 000 randomly selected druglike molecules chosen from the ChemBridge Prime database. Each molecule was docked into the receptor 10 times, and the structure with the best docking energy structure was saved.

For the thrombin screening experiment the crystal structure of NAPAP in thrombin (PDB code 1DWD) was used as the receptor structure with a grid box of 7037 Å^3 defined. The grid box was chosen to encompass the known major binding pockets in the active site: in the case of thrombin, these included, for example, the S1 to S4 subsites as well as the catalytic residues. In the second experiment, the crystal structure of DX-9065a in factor Xa (PDB code 1FAX) was used as the receptor structure, with a grid box of 8012 Å^3 defined. The crystal structure of the antagonist raloxifene (PDB code 1ERR) was used as the receptor structure for ER, with a grid box of 4923 Å^3 defined, covering the binding site occupied by the antagonist.

The aim of the experiment was to show that PRO_LEADS could be used to screen a large number of molecules with minimal manual effort and in a practical time. A good result from the experiment would be to observe a separation in the docked energies for the two sets of molecules.

Data Set Preparation. Initial preparation of the ChemBridge[33] Prime database of 66 877 molecules involved the stripping of counterions and the addition of hydrogen atoms according to an extended SMILES[36] rule set. The complete database was reduced initially by applying filters of physical properties to remove outlier chemistries (e.g., those which are overfunctionalized, highly lipophilic, with a large number of rotatable bonds, or at the extremes of molecular weight) which would be less satisfactory as lead candidates. From the remaining set of compounds, which exhibited a more druglike profile than the complete database, 10 000 were randomly selected for the docking experiments and transformed into three dimensions using MSI's Converter[35] program.

Table 3. Summary Table of the Thrombin Database Screening Results[a]

no. of molecules with the best docking energies	% of database in the subset	no. of thrombin ligands in the subset	% of the subset that are thrombin ligands	enrichment factor
10 043	100	43	0.43	1.0
1000	10	36	3.6	8.4
500	5	31	6.2	14.41
100	1	14	13.9	32.3
50	0.5	11	21.9	50.9

[a] The enrichment factor is calculated by dividing the number of thrombin ligands in a subset by the number of thrombin ligands that would have been present if a random selection of molecules had been made.

The 10 000 molecule data set was enriched by the addition of ligands known to bind to the receptors. (The complete ligand data sets are available on request.) For the thrombin screening experiment, 43 noncovalent inhibitors of diverse chemistry were added to the data set, including classical amidine/guanidine-based inhibitors[38] (e.g., NAPAP, argatroban) as well as more recently published inhibitors with weakly basic or nonbasic P1 groups. The factor Xa test set comprised 30 mono- or dibasic ligands such as DX9065a[39] and BABCH.[40] For the ER screening experiment the data set comprised 12 ER antagonists[41] (e.g., raloxifene, tamoxifen) and 6 agonists[41] (e.g., estradiol, diethylstilbestrol), again chosen to encompass as broad a range of two-dimensional chemistry as possible. Note, there are significant structural differences between an ER−agonist complex and an ER−antagonist complex. The six ER agonists added to the data set were not expected to dock well to an estrogen receptor taken from a structure with a bound antagonist.

Docking Protocol. On the basis of the results of the PRO_LEADS validation exercise, the virtual screening experiments were run with docking protocol 2. Each molecule was docked into the active site of the receptor using a protocol of 200 000 energy function evaluations, the docking was repeated 10 times, and the lowest energy structure was saved.

The faster docking protocols failed to make a correct prediction of the binding modes of many of the serine protease complexes in the test set: some serine protease ligands are relatively large and flexible and therefore difficult to dock. Although acceptable binding mode prediction rates (greater than 60%) were achieved with faster docking protocols, it was felt that protocol 2 offered a compromise between accuracy and speed. The good binding affinity and binding mode predictions at this protocol gave confidence in the ability of the PRO_LEADS to correctly dock and rank the molecules.

Results. To test whether PRO_LEADS is capable of identifying known ligands, subsets of the best-scoring molecules were chosen by ranking the data set in order of the lowest predicted binding affinities. The number of known receptor ligands in each subset was determined and the results have been summarized in Tables 3−5. The enrichment factor is the number of known ligands in the subset divided by the number that would have been present in a random subset selection. The percentages of ligands in the subsets are also quoted in Tables 3−5.

Table 4. Summary Table of the Factor Xa Database Screening Results

no. of molecules with the best docking energies	% of database in the subset	no. of factor Xa ligands in the subset	% of the subset that are factor Xa ligands	enrichment factor
10 030	100	30	0.3	1.0
1000	10	26	2.6	8.6
500	5	22	4.4	14.7
100	1	13	13.0	43.3
50	0.5	11	22.0	73.3

Table 5. Summary Table of the Estrogen Receptor Screening Results

no. of molecules with the best docking energies	% of database in the subset	no. of ER ligands in the subset	% of the subset that are ER ligands	enrichment factor
10 018	100	18	0.18	1.0
1000	10	15	1.5	8.3
500	5	15	3.0	16.7
100	1	12[a]	12.0	66.7
50	0.5	12[a]	24.0	133.0

[a] Top 0.5% of molecules contains all 12 antagonists.

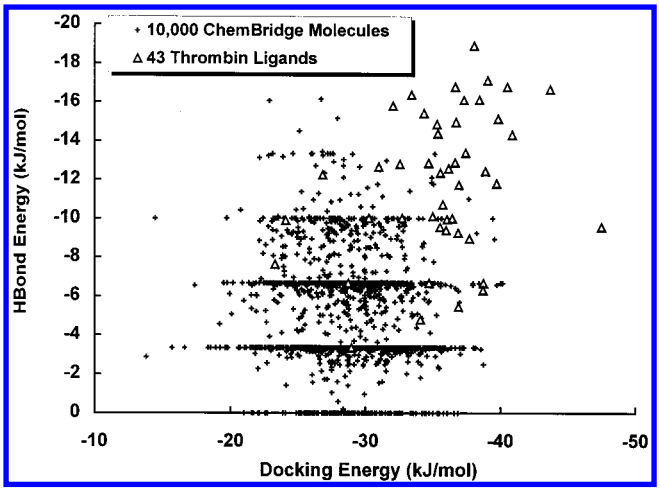

Figure 2. Scatter plot of docked energy vs hydrogen bond energy for the 10 000 random molecules and the 43 thrombin ligands docked against thrombin.

Thrombin. The results of the screening experiment are summarized in Table 3. The 1000 molecules with the best docking energies include 36 of the 43 thrombin ligands. If 1000 molecules had been chosen at random, the set would have included only 4.3 thrombin ligands, and therefore PRO_LEADS has increased the number of known thrombin ligands in the selected subset by an enrichment factor of 8.4. If a smaller set of molecules is chosen, the enrichment factor increases. The molecule with the lowest energy is a potent thrombin inhibitor with a measured binding affinity of −48.6 kJ/mol[42] and a PRO_LEADS predicted energy of −47.5 kJ/mol. Good separation between the two sets of molecules is seen when the docked energies are plotted against the contribution to the total docked energy from the hydrogen bond term, as shown in Figure 2. The hydrogen bond energy was chosen because hydrogen bonds in the region of the main specificity pockets are known to be important for good ligand binding to thrombin. In Figure 2 the thrombin ligands are concentrated in the top right quarter of the plot, which

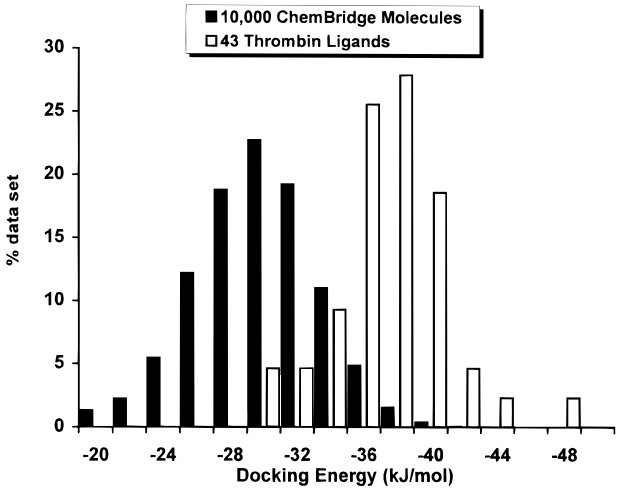

Figure 3. Histogram of docked energy for the 10 000 random molecules and the 43 thrombin ligands docked against thrombin, plotted as a function of the number of molecules in each set.

corresponds to a low total energy and a good hydrogen bond contribution. Even when only the docked energies are considered, good separation is observed: the frequency histogram in Figure 3 highlights the lower energies of the thrombin ligands.

Many of the random molecules achieved good predicted binding energies by making extensive lipophilic contacts with the receptor, particularly in the S2 pocket, bounded by Leu-99 and Tyr-60A. Only a minority of random compounds made good contacts in the S1 pocket: this is because the Prime data set does not contain many cations capable of binding to Asp-189 within the steric constraints of the S1 pocket, while general hydrophobic ligands achieve better energies in the less polar S2 pocket.

Factor Xa. The results of the screening experiment are summarized in Table 4. The 1000 molecules with the best docking energies include 26 of the 30 factor Xa ligands. If the 1000 molecules had been chosen at random, the set would have included only 3.0 ligands, and therefore PRO_LEADS has increased the number of factor Xa ligands in the subset by a factor of 8.6. Seven of the 10 top scoring molecules are from the factor Xa ligand set including the five molecules with the best docking energies. The energy separation between the two sets of molecules is shown by the histogram in Figure 4.

Factor Xa (FXa) is of interest in sharing some general structural similarities with thrombin while showing distinct substrate specificities due to the absence of the S2 pocket and the presence of a larger S4 pocket, and these features markedly affected the binding modes of the random data set. The majority of high-scoring random molecules achieved good lipophilic contacts with the large aromatic S4 pocket of FXa, which is capable of tolerating a more diverse set of ligands than thrombin's smaller S2 pocket.

Estrogen Receptor. The results of the screening experiment are summarized in Table 5. It is interesting to note that the 12 estrogen receptor ligands with the best docking energies are the 12 antagonists and that all these molecules are in the top 0.5% of the data set, giving an enrichment factor of 66.7. In fact all 12 antagonists were scored in the 25 top scoring molecules. The antagonist from this receptor structure, raloxifene, has the second lowest docking energy,

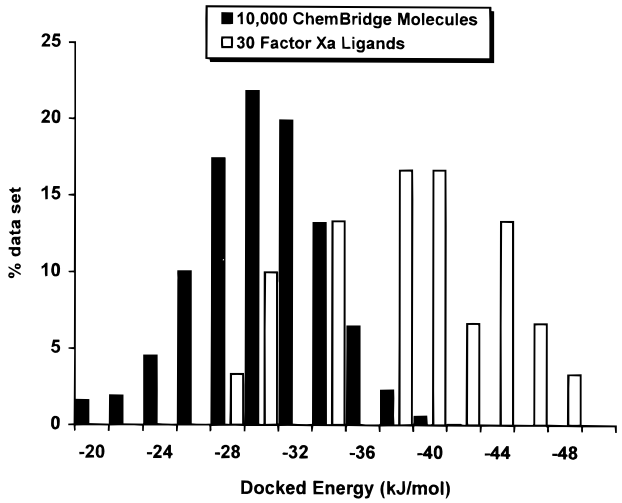

Figure 4. Histogram of the docked energy for the 10 000 random molecules and the 30 factor Xa ligands docked against factor Xa, plotted as a function of the number of molecules in each set.

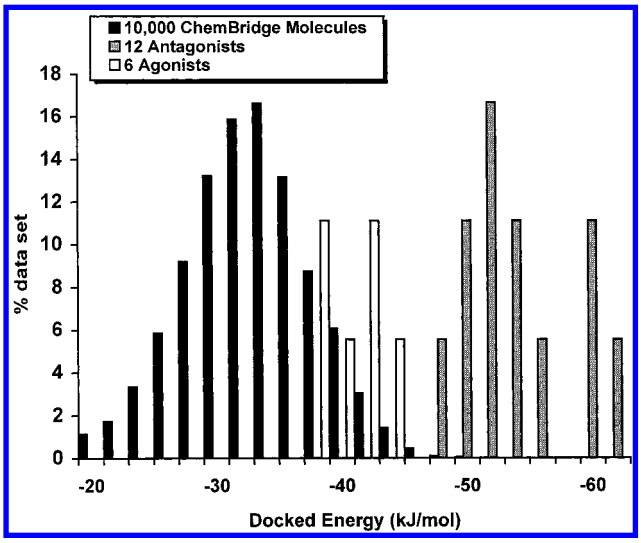

Figure 5. Histogram of the docked energy for the 10 000 random molecules and the 18 estrogen ligands (12 antagonists and 6 agonists) docked against estrogen receptor.

Figure 6. The lowest energy structure of raloxifene docked into estrogen receptor is displayed in gray, and the crystal structure is displayed in white.

−59.4 kJ/mol, and its predicted binding mode and crystal structure are shown in Figure 6. There is a clear separation between the docking energies of randomly chosen molecules, the agonists, and the antagonists, with three distinct peaks being observed in the energy histogram shown in Figure 5. It is interesting that the agonists do not score very highly but are still somewhat better than the random molecules.

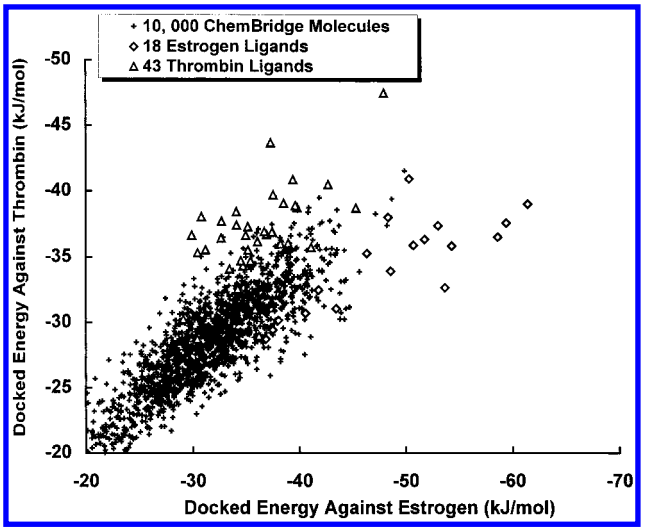

Figure 7. Scatter plot of docked energies of the 10 000 random molecules, the 43 thrombin ligands, and the 18 estrogen receptor ligands docked against the estrogen receptor and thrombin.

With regard to structure, ER is quite distinct from thrombin and FXa in presenting a completely enclosed binding site and is predominantly lipophilic in character. Random lipophilic ligands can therefore obtain moderate scores fairly readily, although on close inspection very few random ligands correctly fill out the classic steroid binding pocket.

In Figure 7 the PRO_LEADS docking energies for the 10 000 randomly chosen molecules, the 18 estrogen receptor ligands, and the 43 thrombin ligands docked into the estrogen receptor have been plotted against the energies of the same molecules docked into thrombin. The plot shows the random ligands in a broad band tending toward lower energies when docked into estrogen receptor and good separation between the two sets of ligands, each set showing selectivity for its own receptor.

CONCLUSIONS

The performance of the fast, flexible docking method PRO_LEADS in de novo design applications has been reported.[27] PRO_LEADS uses a novel tabu search algorithm to explore the extensive search space and an empirically derived scoring function that is capable of estimating binding affinities to rank the possible solutions. The method has been tested on a set of 70 protein−ligand complexes using a series of docking protocols. As expected, the protocol which involved the greatest number of scoring function evaluations gave the best prediction of binding affinity and the highest binding mode prediction rate. As the number of function evaluations was reduced, the quality of the results degraded. With time being an important consideration in both de novo design and virtual screening applications, a compromise between speed and accuracy was made, and it was decided that protocol 2 would be used for the virtual screening experiments.

Ten thousand druglike molecules from the ChemBridge Prime database and a small number of known ligands have been docked into thrombin, factor Xa, and ER. Good separation was observed between the docking energies of the two sets of molecules. Subsets containing molecules with low predicted binding energies were enriched with known ligands. Inevitably, using this technique, a small number of

ligands were incorrectly predicted, and subsequently, not included in the low-energy subsets. Another problem with the method is the large number of random molecules with moderate docking energies. PRO_LEADS, like other docking methods, uses a scoring function trained on a test set of complexes taken from the PDB.[32] There are no negative data available from the crystal structures of bound ligand complexes, and it is therefore difficult to include in the scoring function terms to penalize badly docked molecules. Inclusion in the scoring function of terms to describe misdocked complexes, for example lack of complementarity between polar and lipophilic surfaces, will help PRO_LEADS discriminate between molecules with moderate and poor docking energies.

In conclusion, PRO_LEADS is a fast and accurate docking method, which has been tested on a large set of protein−ligand complexes. A binding mode prediction of 60% was achieved with a fast docking protocol taking on average less than 1 min to dock each complex. The accuracy of both the binding mode prediction and the binding affinity prediction improved as the time spent docking increased. Good results were obtained using protocol 2. The binding mode prediction was 79%, and the binding affinity was estimated with a s value of 9.95 kJ/mol and an r^2 of 0.47. The docking time for this protocol was 5.8 CPU min per complex, which is an acceptable time for practical applications.

Despite the recognized limitations of the scoring function and the short docking protocol, PRO_LEADS successfully selected subsets of molecules, enriched with ligands that are known to bind to the receptor, from a 10 000 molecule druglike data set. PRO_LEADS screened molecules for activity against thrombin, factor Xa, and ER. No ER complexes were used to train the scoring function or to test PRO_LEADS. ER therefore provided an interesting and challenging example and excellent results were achieved.

PRO_LEADS is not as fast as some docking methods available,[18−20,26] but it provides a more reliable prediction of binding mode and binding affinity. Even with a docking time of, on average, 6 min per structure, PRO_LEADS can dock a 10 000 compound data set in less than 3 days on a 16 processor SGI Origin. Admittedly, it becomes more impracticable to screen very much larger compound collections. A more realistic and practical exercise would be to reduce a very large data set by the application of simple property and two-/three-dimensional filters prior to docking. PRO_LEADS would be a suitable docking tool in this application.

In the future we will be working on addressing some of the limitations of the method including the use of energy terms to penalize incorrectly docked structures either as part of the scoring function or as a postprocessing filter to aid in the final selection of molecules. We will continue to collect results by screening virtual libraries and other compound collections against different receptor structures.

REFERENCES AND NOTES

(1) Blaney, J. M.; Dixon, J. S. A Good Ligand is Hard to Find: Automatic Docking Methods. *Perspect. Drug Discov. Res.* **1993**, *15*, 301−319.

(2) Jones, G.; Willet, P. Docking Small-Molecule Ligands into Active Sites. *Curr. Opin. Biotechnol.* **1995**, *6*, 652−656.

(3) Lybrand, T. P. Ligand-Protein Docking and Rational Drug Design. *Curr. Opin. Struct. Biol*. **1995**, *5*, 224−228.

(4) Rosenfeld, R.; Vajda, S.; Delisi, C. Flexible Docking and Design. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 677−700.

(5) Gschwend, D. A.; Good, A. C.; Kuntz, I. D. Molecular Docking Towards Drug Discovery. *J. Mol. Recognit.* **1996**, *8*, 175−186.

(6) Lengauer, T.; Rarey, M. Computational Methods for Biomolecular Docking. *Curr. Opin. Struct. Biol.* **1996**, *6*, 402−406.

(7) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-ligand interactions. *J. Mol. Biol*. **1982**, *161*, 269−288.

(8) Goodsell, D. S.; Olson, A. J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins* **1990**, *8*, 195−202.

(9) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed Automated Docking of Flexible Ligands to Proteins: Parallel Applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293−304.

(10) Gehlhaar, D. K.; Verhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming. *Chem. Biol*. **1995**, *2*, 317−324.

(11) Gehlhaar, D. K.; Verhivker, G. M.; Rejto, P. A.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. In *Evolutionary Programming IV: Proceedings of the Fourth Annual Conference on Evolutionary Programming*; McDonnell, J. R., Reynolds, R. G., Fogel, D. B., Eds.; MIT Press: Cambridge, 1995; p 615.

(12) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.

(13) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol*. **1997**, *267*, 727−748.

(14) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M. A Genetic Algorithm Method for Molecular Docking Flexible Molecules. *J. Mol. Struct*. **1994**, *308*, 191−206.

(15) Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. Flexible Ligand Docking using a Genetic Algorithm. *J. Comput.-Aided Mol. Design* **1995**, *9*, 113−130.

(16) Clark, K. P.; Ajay. Flexible Ligand Docking without Parameter Adjustment Across Four Ligand−Receptor Complexes. *J. Comput. Chem.* **1995**, *16*, 1210−1226.

(17) Leach, A. R.; Kuntz, I. D. The Conformational Analysis of Flexible Ligands in Macromolecular Receptor Sites. *J. Comput. Chem.* **1992**, *13*, 730−748.

(18) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. Predicting Receptor−Ligand Interactions by an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(19) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: Fast, Full Automated Docking of Flexible Ligands to Protein Binding Sites. *Chem. Biol.* **1996**, *3*, 449−462.

(20) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. FLOG: A System to Select 'Quasi-Flexible' Ligands Complementary to a Receptor of Known Three-Dimensional Structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 153−174.

(21) McMartin, C.; Bohacek, R. S. QXP: Powerful, Rapid Computer Algorithms for Structure-Based Drug Design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333−344.

(22) Böhm, H.-J. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein−Ligand Complex of Known Three-Dimensional Structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.

(23) Böhm, H.-J. Prediction of Binding Constants of Protein−Ligands: A Fast Method for the Prioritization of Hits Obtained from De Novo Design or 3D Database Search Programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309−323.

(24) Jain, A. J. Scoring Noncovalent Protein−Ligand Interactions: A Continuous Differentiable Function Tuned to Compute Binding Affinities. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 427−440.

(25) Walters, W. P.; Stahl. M. A.; Murcko, M. A. Virtual Screening−An Overview. *Drug Discovery Today* **1998**, *3*, 160−178.

(26) Makino, S.; Kuntz, I. D. Automated Flexible Ligand Docking Method and Its Application for Database Search. *J. Comput. Chem*. **1997**, *18*, 1812−1825.

(27) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity. *Proteins* **1998**, *33*, 367−382.

(28) Murray, C. W.; Baxter, C. A. Application of Molecular Docking to Thrombin and Thermolysin. *J. Comput.-Aided Mol. Design*. **1999**, *13*, 547−562.

(29) Westhead, D. R.; Clark, D. E.; Murray, C. W. A Comparison of Heuristic Search Algorithms for Molecular Docking. *J. Comput.-Aided Mol. Design* **1997**, *11*, 209−228.

(30) Glover, F.; Laguna, M. In *Modern Heuristic Techniques for Combinatorial Problems*; Reeves, C. R., Ed.; Blackwell Scientific Publications: Oxford, 1993; p 70.

(31) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(32) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimamouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol*. **1997**, *112*, 535−542.
(33) ChemBridge Corporation, 16981 Via Tazon, Suite G, San Diego, CA 92127. http://www.chembridge.com.
(34) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. In *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, 1992; p 402.
(35) *InsightII/Discover*, v. 95.0; Molecular Simulations Inc.: San Diego, CA, 1996.
(36) Bone, R. G. A.; Firth, M. A.; Sykes, R. A. SMILES Extensions for Pattern Matching and Molecular Transformations: Applications in Chemoinformatics. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 846−860.
(37) Gillet, V. J.; Willett, P.; Bradshaw, J. Identification of Biological Activity Profiles Using Substructureal Analysis and Genetic Algorithms. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 165−1740.
(38) Brandstetter, H.; Turk, D.; Hoeffken, H. W.; Grosse, D.; Stuerzebecher, J.; Martin P. D.; Edwards, B. F. P.; Bode, W. Refined 2.3 Angstroms X-ray Crystal Structure of Bovine Thrombin Complexes Formed with the Benzamidine and Arginine-Based Thrombin Inhibitors/NAPAP, 4-TAPAP and MQPA: A Starting Point for Improving Antithrombotics. *J. Mol. Biol*. **1992**, *226*, 1085−1099.
(39) Brandstetter, H.; Kuhne, A.; Bode, W.; Huber, R.; von der Saal, W.; Wirthensohn, K.; Engh, R. A. X-ray Structure of Active Site-Inhibited Clotting Factor Xa−Implications for Drug Design and Substrate Recognition. *J. Biol. Chem*. **1996**, *271*, 29988−29992.
(40) Shaw, K. J.; Guilford, W. J.; Dallas, J. L.; Koovakkaat, S. K.; McCarrick, M. A.; Liang, A.; Light, D. R.; Morrissey, M. M. (Z,Z)-2,7-Bis(4-amidinobenzylidene)cyclo-heptan-1-one: Identification of a Highly Active Inhibitor of Blood Coagulation Factor Xa. *J. Med. Chem*. **1998**, *41*, 3551−3556.
(41) Grece, T. A.; Dodge, J. A.; Selective Estrogen Receptor Modulators. *Curr. Pharm. Des*. **1998**, *4*, 71−92.
(42) Tucker, T. J.; Lumma, W. C.; Lewis, S. D.; Gardell, S. J.; Lucas, B. J.; Sisko, J. T.; Lynch, J. J.; Lyle, E. A.; Baskin, E. P.; Waltmann, R. F.; Appleby, S. D.; Chen, I.-W.; Dancheck, K. B.; Naylor-Olsen, A. M.; Krueger, J. A.; Cooper, C. M.; Vacca, J. P. Synthesis of a Series of Potent and Orally Bioavailable Thrombin Inhibitors that Utilize 3,3-Disubstituted Propionic Acid Derivatives in the P_3 Position. *J. Med. Chem*. **1997**, *40*, 3687−3693.

CI990440D