ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/227955645

Efficacy and selectivity in flexible database docking

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · NOVEMBER 1999

Impact Factor: 2.63 · DOI: 10.1002/(SICI)1097-0134(19991115)37:3<334::AID-PROT3>3.0.CO;2-9

CITATIONS
10

READS
82

2 AUTHORS:

Ronald Knegtel
Vertex Pharmaceuticals (Europe) Ltd
53 PUBLICATIONS **1,713** CITATIONS
SEE PROFILE

Markus Wagener
Grünenthal GmbH
38 PUBLICATIONS **1,121** CITATIONS
SEE PROFILE

Available from: Ronald Knegtel
Retrieved on: 12 January 2016

PROTEINS: Structure, Function, and Genetics 37:334–345 (1999)

Efficacy and Selectivity in Flexible Database Docking

Ronald M.A. Knegtel* and Markus Wagener
Department of Molecular Design and Informatics, N.V. Organon, Oss, The Netherlands

ABSTRACT **Flexible database docking with DOCK 4.0 has been evaluated for its ability to retrieve biologically active molecules from a database of approximately 1,000 compounds with known activities against thrombin and the progesterone receptor. The retrieval of known actives and chemically similar but inactive molecules was monitored as a function of conformational and orientational sampling. The largest enrichment of actives among the 10% highest ranking molecules is obtained when only five conformations are used to seed the next round of ligand reconstruction and limited sampling is applied to place the base fragment in the binding site. The performance of energy and chemical scoring, as implemented in DOCK 4.0, was found to depend on the protein used for docking. For the progesterone receptor, energy scoring yields the largest enrichments (64%) in terms of actives retrieved among the 10% top scoring molecules, while chemical scoring performs best for thrombin (94%). With the exception of the application of energy scoring to the progesterone receptor, both energy-based scoring schemes applied in this study do not discriminate well between true actives and chemically similar but inactive compounds. In conclusion, flexible docking is able to effectively prioritize high-throughput screening databases, using less conformational sampling than normally required for appropriate reconstruction of protein-ligand complexes. The more subtle discrimination between chemically similar classes of active and inactive compounds remains, however, problematic. Proteins 1999:37:334–345.** © 1999 Wiley-Liss, Inc.

Key words: incremental construction algorithm; DOCK 4.0; thrombin; progesterone receptor; scoring; high-throughput screening

INTRODUCTION

The automated docking of small molecules to binding sites on protein surfaces finds useful applications in pharmaceutical lead discovery and optimization.[1–3] Currently available tools for molecular docking are either aimed at reconstituting the bound conformation of individual molecules[4,5] or focus on the retrieval of lead compounds from large three-dimensional (3D) databases.[6–9] The latter type of approach usually sacrifices some thoroughness of conformational sampling and scoring in order to speed up the calculations.[10] Due to the already computationally intensive task of screening large numbers of molecules, ligand flexibility has only recently been introduced in database docking. For the docking tools that are publicly available this involves the application of the incremental construction algorithm[9,11,12] although applications using evolutionary programming, genetic algorithms and flexibases have also been reported.[8,13–15] The incremental construction algorithm requires the presence of a rigid base fragment in the ligand. This base fragment is placed in the binding site first, after which the remainder of the ligand is added in a piecewise manner. This allows for the docking of flexible molecules in 1–3 minutes,[9–12] which represents a considerable increase in speed compared to earlier methods. Flexible docking using large 3D databases remains, however, a computationally highly intensive undertaking.

Another major bottleneck for the successful application of molecular database docking is the lack of a fast, universal and reliable scoring function (for reviews see Ajay and Murcko[16] and Knegtel and Grootenhuis[17]). Especially entropic contributions to the binding free energy are difficult to calculate and therefore many docking programs only estimate the binding enthalpy. The combined limitations in conformational sampling, scoring and predicting protein flexibility currently limit the capabilities of docking tools to reliably reconstruct protein-ligand complexes.[10,18,19] Two causes for the incorrect docking of ligands can be distinguished. Either the ligand can not be fitted to the protein conformation used for computational screening or the scoring function favors a conformation different from the native bound conformation.[10] Errors of the former type can usually not be dealt with by available database docking tools, due to limitations concerning the amount of conformational sampling per ligand and the use of a single protein conformation. The latter situation, in which other ligand conformers have more favorable scores than the true bound conformation is, however, not necessarily detrimental. Although the correct conformation of a molecule is not always ranked highest, the docking program is still able to identify at least one conformation that fits favorably into the binding site. Such molecules can thus be classified as potentially complementary to the binding site and be selected for experimental evaluation. Here we intend to establish the efficacy of flexible database docking in the retrieval of biologically active molecules from 3D databases.

In the current practice of drug discovery in pharmaceutical industry, computational database searches are compet-

*Correspondence to: Ronald Knegtel, Vertex Pharmaceuticals (Europe) Ltd., 88 Milton Park, Milton, Abingdon, Oxon, OX14 4RY, United Kingdom. E-mail: ronald.knegtel@vpharm.com
Received 23 February 1999; Accepted 10 June 1999

© 1999 WILEY-LISS, INC.

ing with experimental high-throughput screening.[20] In this setting, database docking can be applied to prioritize the compound database to be screened, rather than to select only a small subset of molecules from the database for testing. This implies that the correctness of the predicted activities is less critical than in the recent past, since larger numbers of compounds can be tested with less effort. An enrichment in the number of actives among the first few thousands compounds to be tested can, therefore, already be advantageous. Compounds that are found to be active can then be subjected to lead optimization at an earlier stage. In order to assess the feasibility of prescreening molecule databases by flexible docking for the purpose of prioritizing experimental screening efforts, the efficacy of flexible docking needs to be evaluated.

Evaluating the predictive power of database docking has been hampered, however, by the absence of databases for which of each member the activity with respect to the test system is known.[21] Most evaluations that have been reported in the literature focus on reproducing crystal structures or the identification of a small set of known actives from (subsets) of larger databases of molecules with unknown activity. In such cases the actual number of active molecules in the larger databases is unknown, as is the number of false positives. In addition, although the ability to reconstitute crystal structures is a first requirement for any docking approach, it is not a priori predictive of its efficacy in docking entirely new molecules to protein conformations that are not optimally adapted to them.[18] In this respect, more could be learned by validating docking protocols by docking ligands to protein conformations other than that taken from their own crystal structures. This would establish a validation protocol that mimics real-life applications. For database searches, the large body of screening results that is generated in pharmaceutical industry provides an abundant source of data suitable to assess the effectiveness of flexible database docking.

We have used the docking program DOCK 4.0[12,22,23] to dock small databases of molecules with known activities against thrombin and the progesterone receptor ligand-binding domain (PR-LBD). For these two systems both 3D structural information on the protein, as well as experimental test results for large numbers of molecules, were available within Organon. Thrombin has a solvent accessible active site of mixed polar/apolar character, with the S1 pocket binding positively charged functionalities of the ligand.[24] The progesterone-binding site of the PR-LBD completely encloses the ligand in a mostly hydrophobic pocket.[25] Both examples represent almost ideal cases for flexible docking. Placement of the base fragment is facilitated for thrombin by the need for charge complementarity in the S1 pocket and for the PR-LBD by the presence of a completely closed binding pocket. We are limited in the selection of test cases, however, to those systems that have been targeted by high-throughput screening within Organon and for which structural data on a preferably rigid binding site is available. The efficacy of flexible docking with DOCK energy scoring has already been demonstrated for ligands with uncharged base fragments binding to the solvent-accessible active site of dihydrofolate reductase.[12,14] Although this work showed that many known actives could be retrieved from a selection of the Available Chemicals Directory (ACD, MDL, San Leandro) the activities of other compounds selected by DOCK were unknown.

In order to further test the ability of flexible docking to select biologically active molecules from databases, a set of inactive compounds that are chemically similar to known actives was included in the test databases. These compounds are expected to offer some insight into the ability of flexible docking to discriminate between true actives and chemically similar inactives. The issue of selectivity has previously been addressed for dihydrofolate reductase,[26] however, using enzyme structures from two different organisms rather than two sets of related but selective ligands. In the work presented here a larger number of chemically diverse and inactive compounds forms the background from which true actives need to be extracted.

Our focus here is mainly on the capability of DOCK 4.0 to correctly identify active molecules from databases rather than the correct prediction of their bound conformations or binding affinity. Given the limitations of energy-based scoring, little correlation is expected between calculated energies and experimentally determined association constants. Here we present an analysis of the effects of applying various degrees of ligand conformational sampling and two different scoring functions implemented in DOCK 4.0 on the retrieval of active molecules from 3D databases. In addition, the capability of the docking program to discriminate between true actives and closely related inactive compounds is discussed.

MATERIALS AND METHODS

Protein Structures and Ligand Databases

For docking against thrombin, the crystal structure of bovine thrombin, solved at 2.3 Å resolution and complexed with the inhibitor NAPAP[24] (PDB code 1ETS), was used. A total of 11 spheres was generated using SPHGEN[6] in the S1 binding pocket, thus focusing docking of the base fragment in this region. For the progesterone receptor ligand binding domain (PR-LBD) the structure of the receptor complexed with progesterone at 1.8 Å resolution was used[25] (PDB code 1A28). A total of 22 spheres, covering the entire ligand-binding site, was generated with SPHGEN. In all cases all water molecules bound in the active site were deleted from the respective structures prior to sphere generation. Energy and chemical scoring grids of 0.3 Å spacing were generated using a unified atom model and a distance dependent dielectric of 4r with a 10 Å cut-off.[27]

For thrombin, the set of 32 inhibitors used in this study is identical to that reported earlier[18] and is listed in Figure 1A. In addition, a set of ten molecules was used that were synthesized during our anti-thrombosis program. They share structural and functional similarities with known inhibitors (cf., Fig. 1B) but were found to be inactive when tested. These compounds were added in order to investigate the capability of DOCK 4.0 to discriminate between active and chemically similar but inactive molecules. The test database was augmented with 1,000 compounds from

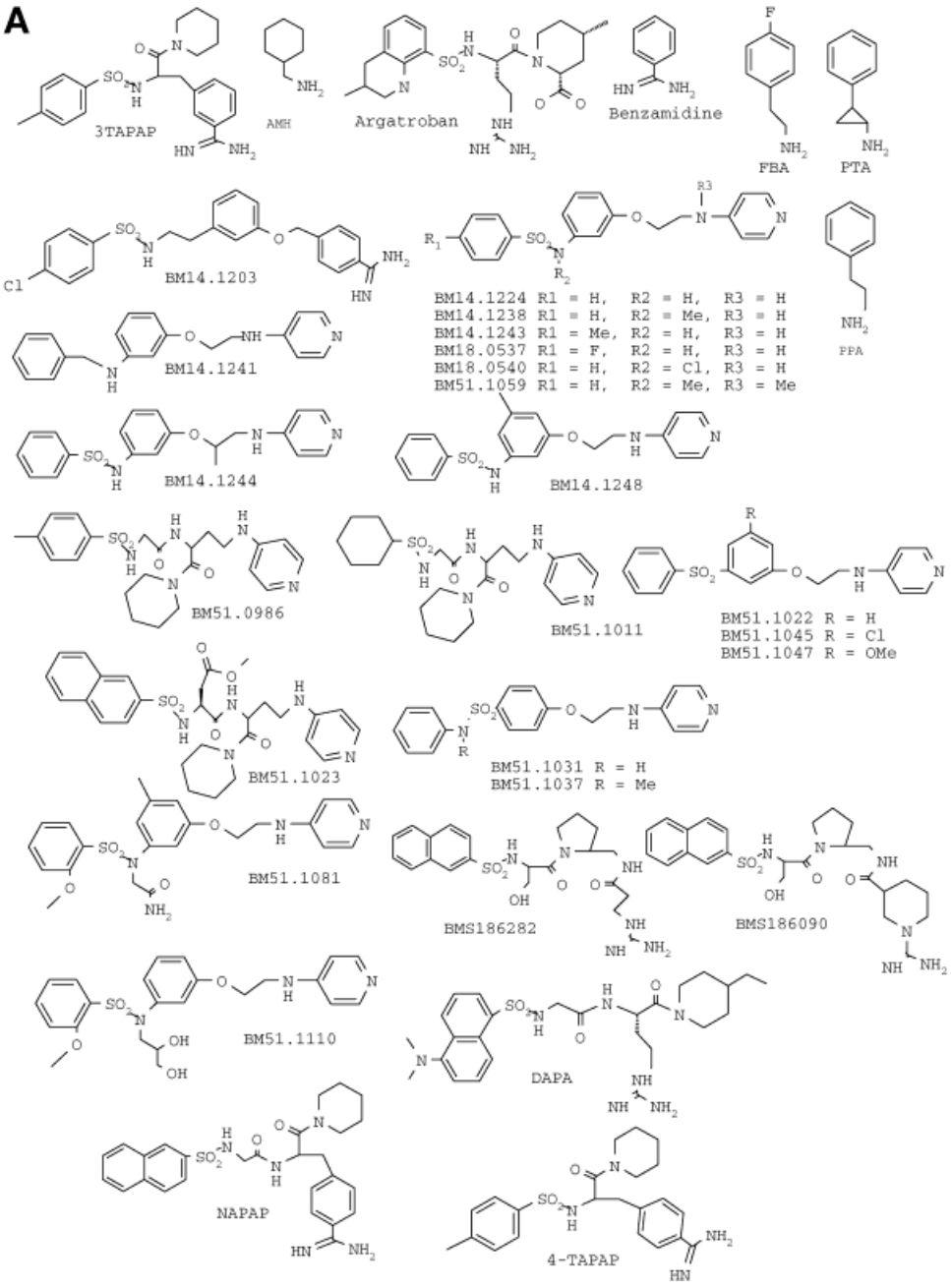

Fig. 1. Chemical structures of 32 known thrombin inhibitors **(A)** and ten chemically similar but inactive compounds **(B)** used for docking against thrombin. Aromatic nitrogens were protonated and carry a positive charge that enables them to bind to the thrombin S1 specificity pocket.

Organon's screening database with thrombin inhibition activities of less than 10%. Diverse compounds were selected using Ward clustering on the basis of BCI fingerprints.[28] These compounds form the background of diverse, inactive molecules from which active compounds are to be selected. In total, the database used for docking against thrombin contained 1,042 molecules.

For the PR-LBD, 28 molecules, which selectively act on the progesterone receptor, were taken from an in-house database of potential drug molecules reported in the scientific and patent literature. These molecules are listed in Figure 2A. In addition, a set of 20 estrogen receptor-specific molecules was added to the test database, including a number of non-steroidal compounds, as shown in Figure 2B. The database was augmented with 1,000 molecules displaying less than 10% activity in our progesterone receptor agonist assay at 10 μM. These molecules were chosen from Organon's screening database on the basis of their chemical diversity as established using BCI fingerprints and Ward clustering.[28] In total, the database used for docking against the PR-LBD contained 1,048 molecules. For both thrombin and the PR-LDB the three

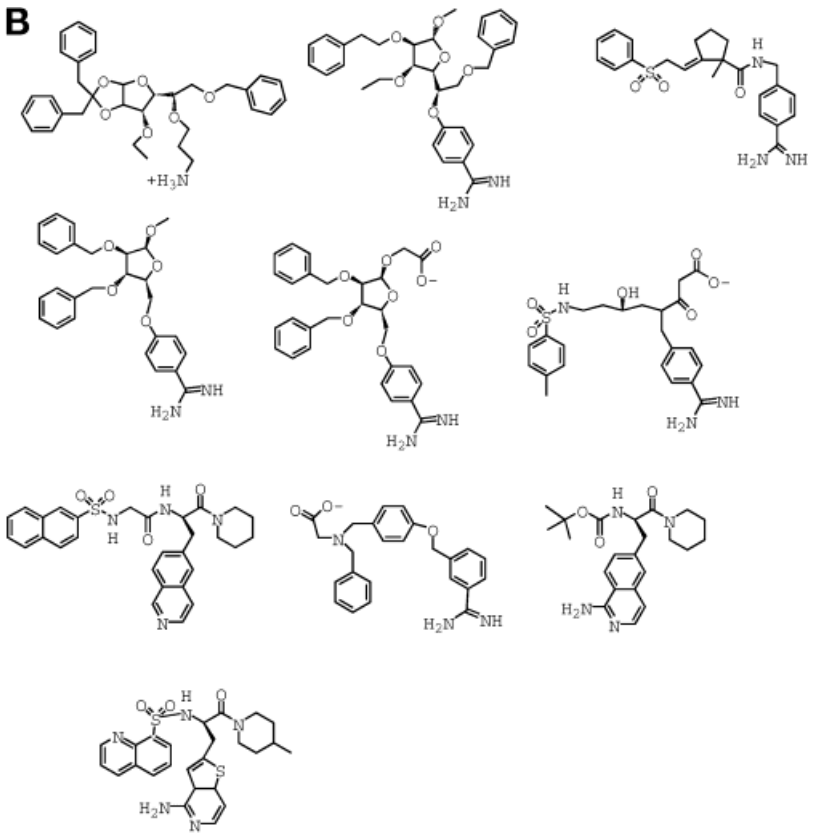

Figure 1. (Continued.)

different classes of ligands will be referred to as actives, similar inactives, and diverse inactives, respectively.

All molecules in the test database were processed from CORINA[29]-generated SDF files to Sybyl mol2 files with Gasteiger-Marsili charges[30] using Sybyl 6.4 (Tripos Inc., St. Louis, MO) and database conversion tools of DOCK 4.0. Special care was taken to assure the correct protonation states of the thrombin inhibitors since the automatic conversion tools failed to protonate the 4-amino pyridines and their analogues. All calculations were performed on an SGI Power Challenge with eight R10000 processors operating at 194 MHz.

Docking and Analysis

All rigid and flexible dockings were performed with DOCK 4.0 using energy and chemical scoring. Energy scoring applies a grid-based representation of the AMBER[31] force field inter-molecular energy. Chemical scoring applies weights to the attractive portion of the van der Waals contribution to the AMBER energy, depending on the chemical types of interacting ligand and protein atoms.[18,23] Five chemical groups (donor, acceptor, polar, hydrophobic, and neutral) are assigned to ligand and protein atoms on the basis of their Sybyl atom type and connectivity. The attractive van der Waals portion of contacts involving chemical groups that are not complementary (i.e., hydrophobic-polar, donor-donor etc.) are neglected while contacts between matching groups are unaffected.

The influence of increasing the conformational sampling on ligand retrieval was tested by using five different peripheral seeds: 0, 3, 5, 10, 25, and 50. The number of peripheral seeds defines the number of highest ranking ligand conformers that are used to seed consecutive ligand reconstruction steps when applying the incremental construction algorithm.[11,23] The use of 0 seeds implies the rigid body docking of a single CORINA generated ligand conformation. Matching of ligand atoms and spheres was performed using uniform sampling with a maximum of 100 orientations as default and a minimal base fragment size of 4 atoms. The number of base fragment placements generated with uniform sampling was increased from 100 to 500 at a constant peripheral seed of 5 to investigate the effect on the retrieval of actives. Maximally three steric clashes between protein and ligand were accepted. All docked ligands were energy/chemical score minimized in 100 cycles of the simplex minimizer implemented in DOCK 4.0 to a convergence of 0.1 kcal/mol.

For analysis, the ranks of active, similar inactive, and diverse inactive compounds among the top ranking 100 compounds (10% of our database) were binned in bins of ten compounds. In addition, enrichment factors (F) were calculated using formula (1):

$$F = \frac{Nact(p)/p}{Nact/N} \quad (1)$$

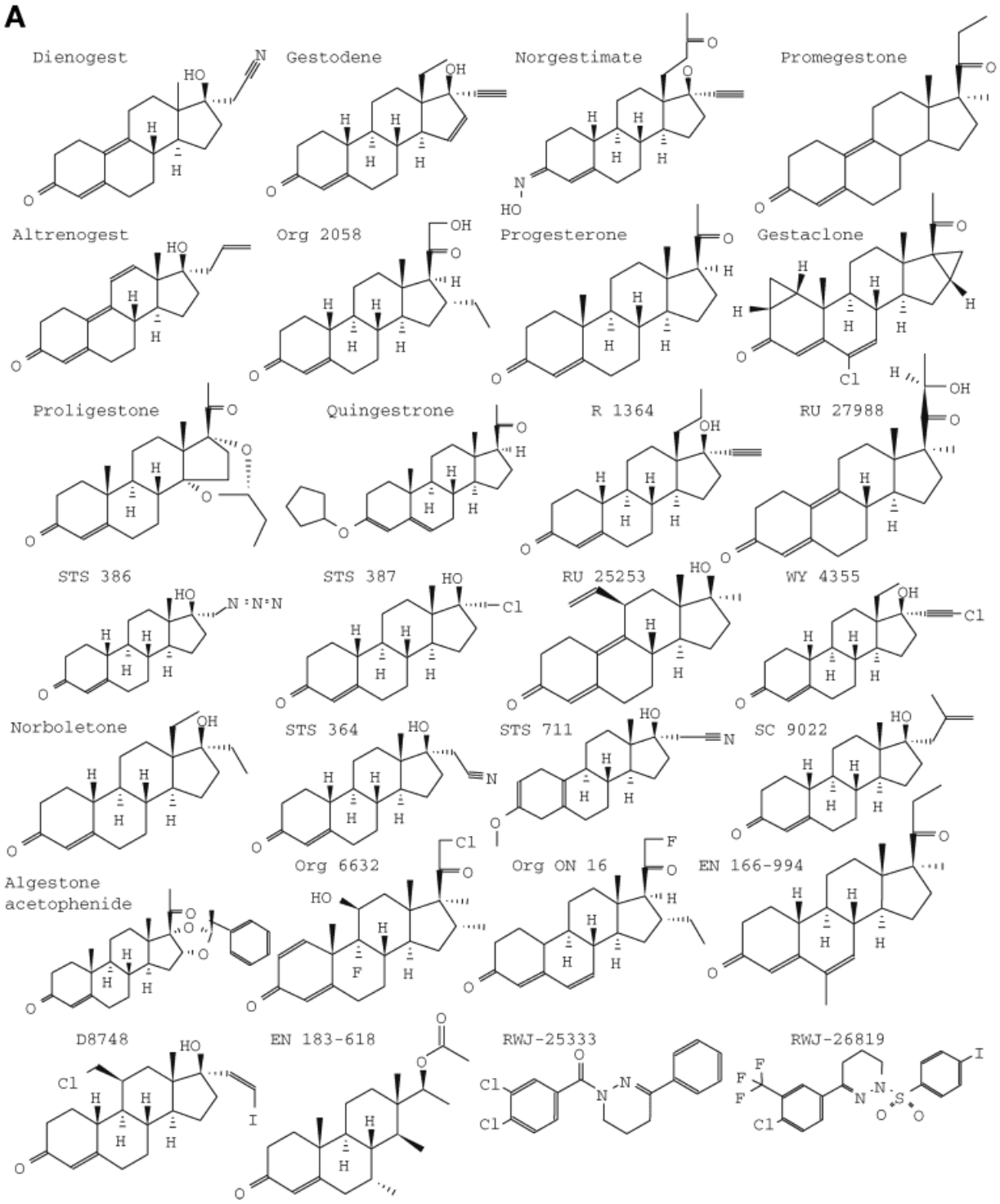

Fig. 2. Chemical structures of 28 known progesterone agonists **(A)** and 20 chemically similar estrogen receptor specific ligands **(B)** used for docking against the progesterone receptor ligand binding domain.

where $N_{act}(p)$ is the number of actives among the p highest ranking molecules, while N_{act} is the total number of actives and N the total number of molecules in the docking database. The same formula was used to calculate the enrichment in chemically similar inactive compounds. Inter-molecular energies and their van der Waals and electrostatic components were extracted from the docking output or obtained by rescoring.

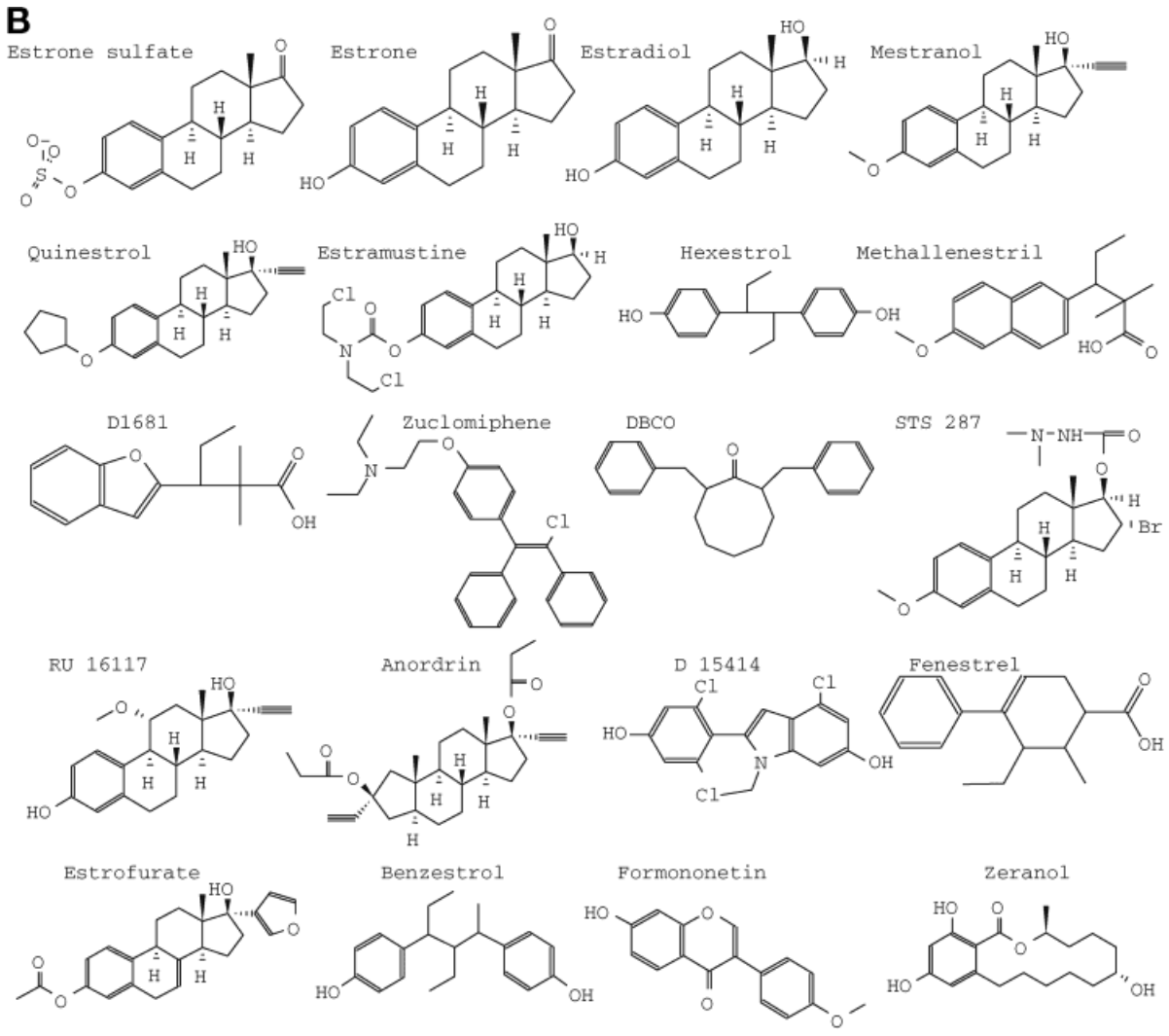

Figure 2. (Continued.)

RESULTS AND DISCUSSION

Thrombin

The enrichment factors for inhibitors of thrombin and similar inactives are shown in Figure 3. Interestingly, rigid docking (for which the number of peripheral seeds equals 0) still favors the true actives over inactives in terms of their ranking among the 100 highest scoring compounds. When flexibility is introduced, both actives and inactives are more effectively retrieved and become almost equally well represented among the top ranking compounds. The enrichment in actives of the 10% highest-ranking compounds reaches an apparent maximum at five peripheral seeds. Especially the smaller compounds, such as benzamidine, fail to rank among the 100 best scorers due to their smaller number of interactions with the protein. Chemical scoring consistently selects more thrombin inhibitors among the top ranking molecules than force field scoring. Apparently, the emphasis on chemical complementarity intrinsic to chemical scoring is quite effective in the context of the mixed polar/non-polar character of the thrombin active site.

When the number of attempted base fragment placements was increased from the default of 100 to 200 and 500 attempts, little effect was observed on the enrichment obtained using chemical scoring. Energy scoring, however, becomes less effective with the enrichment of actives decreasing from 8.1 to 3.9 at maximal sampling and the enrichment of similar inactives remaining almost constant. This effect is due to the improved ranking of diverse inactive compounds of which the number of representatives among the top 100 scorers increases from 61 to 80. Apparently, the application of more rigorous sampling causes many more low energy conformations to be identified for the diverse, inactive compounds. The ten highest-ranking diverse inactives are large molecules and peptides that apparently profit from more elaborate base fragment positioning.

In terms of selectivity, neither of the two scoring schemes is able to differentiate well between actives and inactives.

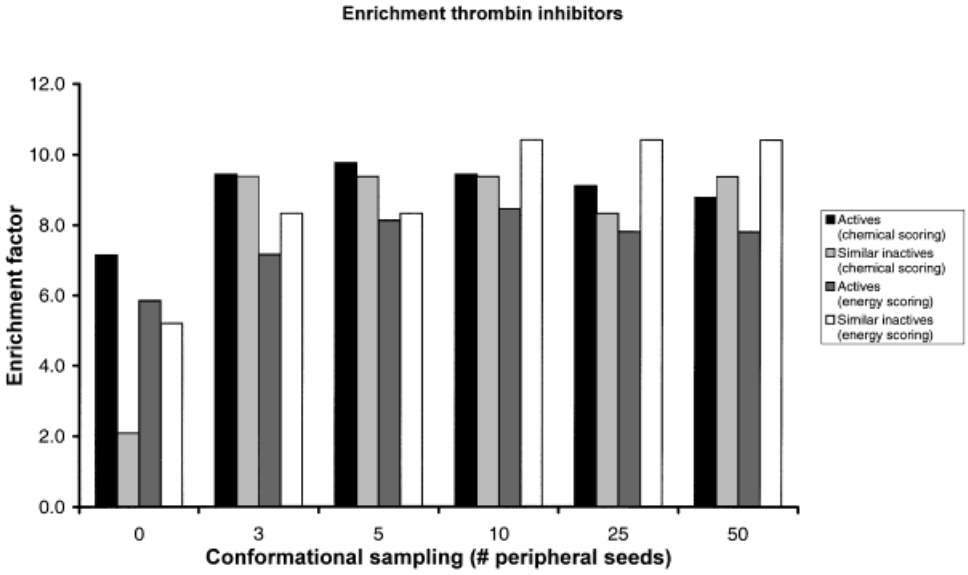

Fig. 3. Enrichment factors for thrombin inhibitors and similar, inactive molecules as a function of conformational sampling defined by the number of peripheral seeds. Results obtained with energy and chemical scoring are shown. The maximal enrichment factor for this database is 10.4.

The inactives that are similar to true inhibitors are enriched to roughly the same extent as the true thrombin inhibitors. In fact, when applying energy scoring, the top ranking 100 compounds become more enriched with similar inactives when sampling is increased. This can be explained by the fact that the true inhibitors are on average smaller in size (51 atoms) than the similar inactives (65 atoms on average). This causes the average van der Waals contribution to the total inter-molecular energy to be consistently more favorable than that of the actives when using energy scoring. This trend is illustrated in Figure 4. It is interesting to note that although the binding of thrombin inhibitors involves the formation of a salt-bridge between Asp189 and a positively charged group of the inhibitor, the total scores are still dominated by the van der Waals interactions.

For chemical scoring the situation is reversed. In this case the van der Waals contribution for the actives is more favorable, indicating that the weighting scheme that is applied successfully filters out contacts between non-complementary surface patches on ligand and protein. This apparently removes the bias introduced by the larger size of the similar but inactive ligands observed with energy scoring. Both the ranking of the various ligands and plotting of the different contributions to the final score indicate that equilibrium is reached at ten peripheral seeds. At only five peripheral seeds the scores are not yet converged but the ranking appears to be optimal. When more flexibility is applied, more of the diverse inactive compounds also acquire more favorable scores and compete for the top 100 of best scoring molecules.

A more detailed picture of how the ranking of docked ligands depends on the amount of sampling that is applied is shown in Figure 5. As sampling is increased when docking with chemical scoring, more actives accumulate in the higher-ranking bins. When energy scoring is applied the known inhibitors are almost evenly distributed over the top ranking 100 molecules. Chemical scoring clearly

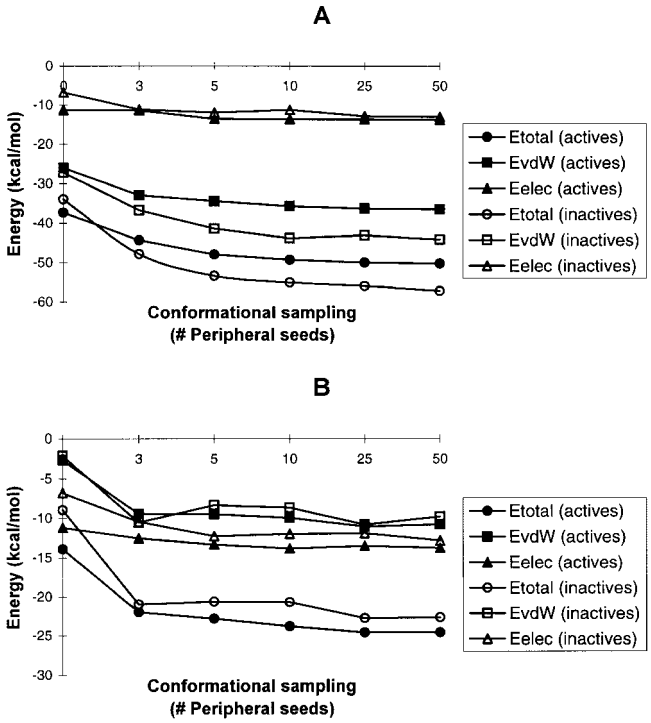

Fig. 4. The average total inter-molecular interaction energy (**A**, E_{total}), chemical score (**B**, E_{total}) and their van der Waals (E_{vdW}) and electrostatic components (E_{elec}) are plotted as a function of the number of peripheral seeds for the docking of active and similar inactive compounds against thrombin.

provides the largest enrichment, with 94% of the actives being ranked among the top 100 best scoring molecules and 78% in the top 50. Energy scoring performs less well with 78% actives ranked in the best scoring top 100 and 40.6 % in the top 50. Peripheral seeds larger than five do not appear to yield improved prioritization of the known actives. The inactive compounds (panels B and D in Figure

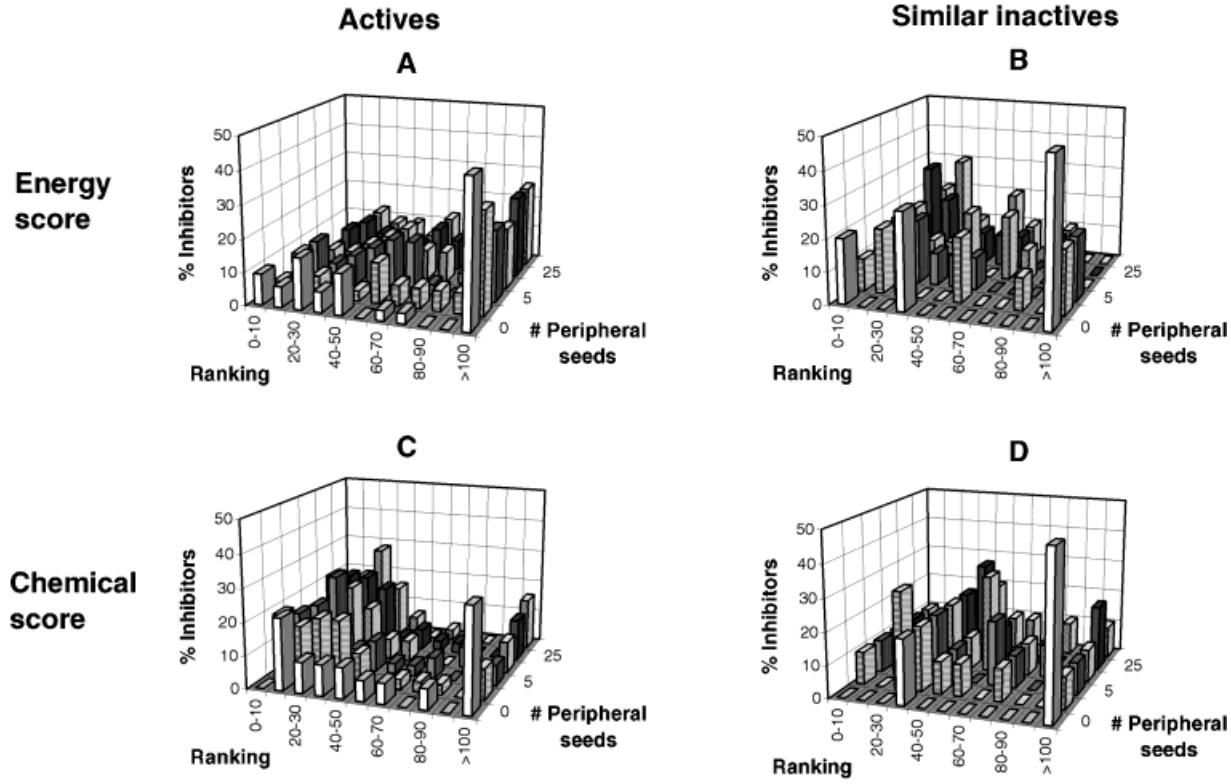

Fig. 5. The distribution of actives (**A** and **C**) and similar inactives (**B** and **D**) over the top 100 ranks after docking against thrombin with increasing amounts of sampling. Results are shown for dockings obtained with energy (A and B) and chemical scoring (C and D), respectively. Ranks are divided in bins of ten compounds and include a bin for molecules docked with ranks larger than 100.

5) appear to cluster in higher-ranking bins, independently of the scoring function that is applied.

The Progesterone Ligand Binding Domain

Figure 6 shows the enrichment in actives and similar inactives for the progesterone receptor using energy and chemical scoring. Both scoring methods are capable of some discrimination between true actives and inactives but opposed to the thrombin results, energy scoring achieves the largest enrichment. The discrimination between actives and similar inactives is partly due to the small difference in the average number of atoms for the active (54 atoms) and similar inactive molecules (48 atoms). Since the binding cavity is predominantly hydrophobic, the scoring can be expected to depend heavily on an accurate representation of the van der Waals interactions. Since chemical scoring involves scaling of the attractive part of the van der Waals interactions, and thereby introduces repulsive interactions, it may yield less accurate scorings in the PR-LBD case. This is exemplified by the fact that when chemical scoring is applied, four actives and five similar inactives can not be docked to yield negative interaction energies. For energy scoring only one active and one similar inactive of the same sets of ligands can not be docked to yield favorable interactions. These compounds do not fit into the conformation of the ligand-binding domain as used for docking. This may be related to the structural flexibility observed in the family of nuclear hormone receptors upon binding of agonists or antagonists. [32] When the compounds docked with chemical scoring and using five peripheral seeds are rescored and ranked using energy scoring the enrichment factor for actives increases from 4.1 to 6.7. This indicates that scoring rather than sampling defines the probability of retrieving actives in this system. As with thrombin, an optimum in enrichment is observed at five peripheral seeds.

Increasing the number of base fragment placements from 100 to 500 causes a slight reduction in the enrichment factors of actives with energy scoring (from 6.7 to 5.9). As with thrombin, this is due to a higher ranking of diverse inactive compounds although for the PR-LBD the effect is less pronounced. For chemical scoring both the enrichment and selectivity are observed to improve slightly. Due to its correct treatment of van der Waals interactions, energy scoring yields a better match between ligand and protein in terms of shape than chemical scoring. A tighter fit of the base fragment leaves more space for the remaining flexible part of the ligand to be fitted into the binding site. This explains why more sampling causes more inactive compounds to rank among the top 100.

As shown in Figure 7, the selectivity offered by energy scoring is mostly due to the difference in van der Waals interactions with the receptor. This difference is almost

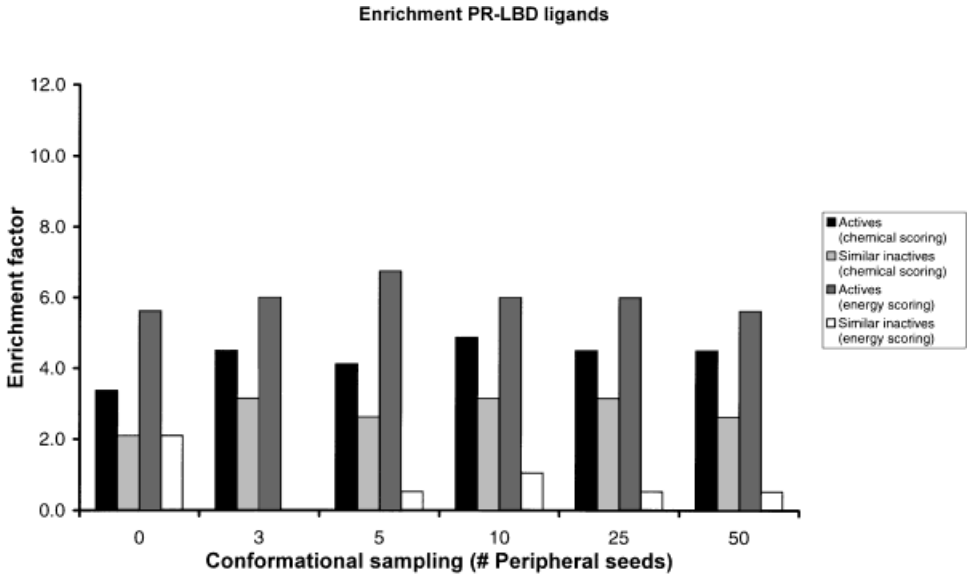

Fig. 6. Enrichment factors for progesterone receptor ligands and estrogen receptor specific ligands as a function of conformational sampling defined by the number of peripheral seeds. Results obtained with energy and chemical scoring are shown. The maximal enrichment factor for this database is 10.5.

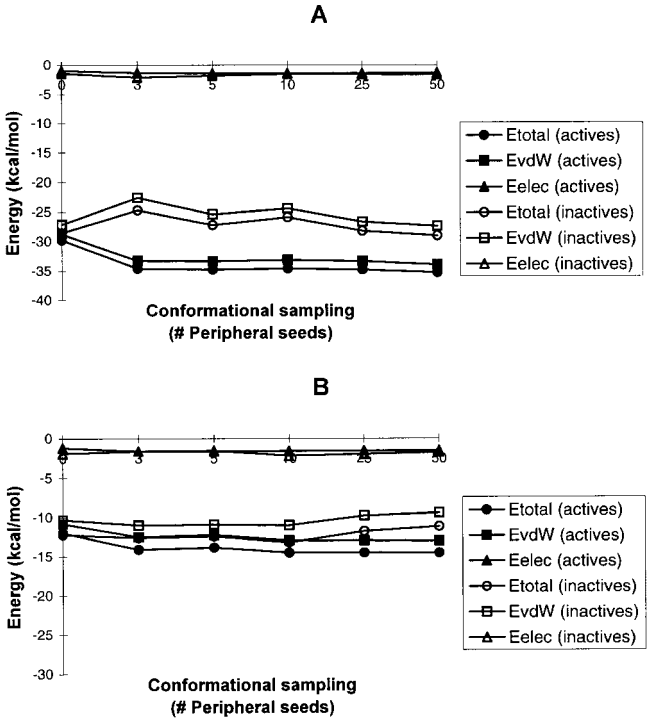

Fig. 7. The average total interaction energy (**A**, E_{total}), chemical score (**B**, E_{total}) and their van der Waals (E_{vdW}) and electrostatic components (E_{elec}) are plotted as a function of the number of peripheral seeds for the docking of active and similar inactive compounds against the progesterone ligand-binding domain.

completely obscured when chemical scoring is applied and what remains is due to the difference in size between the active and similar inactive ligand molecules. As for thrombin, the energies are close to convergence near five peripheral seeds. Increasing sampling does improve the energies somewhat but ultimately does so for all molecules in the database. This results in an improvement in ranking of also the diverse inactive compounds. It is interesting to note that although energy scoring performs best in terms of enrichment and selectivity, it causes progesterone (the only ligand of which the binding mode is known) to be docked in an orientation that is 180° rotated with respect to the known bound orientation. Chemical scoring yields the correct orientation. This result confirms that enrichment in terms of activity and correct reconstruction of bound ligand conformations are not necessarily correlated.

When the ranking of actives and similar inactives is analyzed in more detail (see Figure 8), it appears that with increased sampling the actives accumulate again in the top ranking bins. Energy scoring yields the largest enrichment with 64% of the known actives ranking among the top 100 and 50% among the top 50 using five peripheral seeds. Chemical scoring ranks only 39% among the 100 top scorers of which 36% are among the top 50. For peripheral seeds larger than five this accumulation is compromised somewhat due to the improved ranking of (diverse) inactive compounds. Although chemical scoring ranks fewer actives within the top 10% scorers, these molecules tend to cluster in the higher-ranking bins. The molecules retrieved with energy scoring are more evenly distributed over the 100 highest ranks, as was also observed with thrombin.

Implications for Flexible Database Docking

On the basis of the results presented here it appears that only a limited amount of conformational sampling is required to obtain a reasonable retrieval of actives from 3D

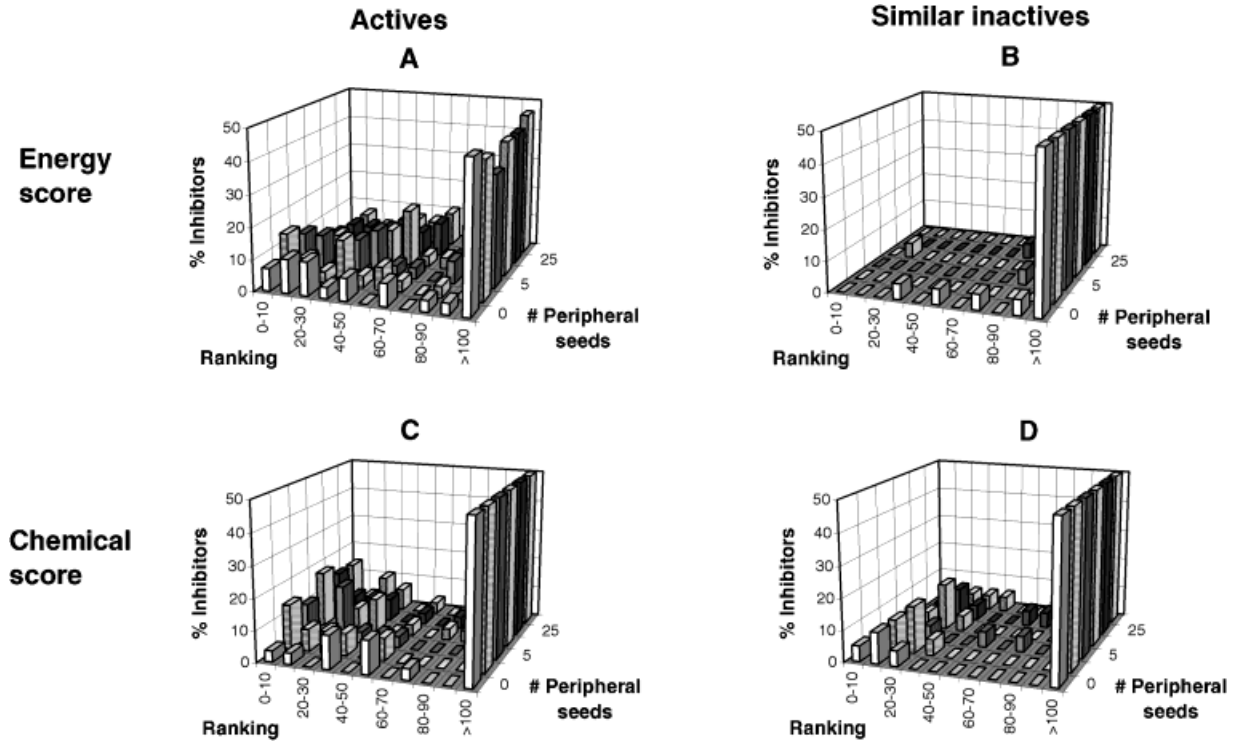

Fig. 8. The distribution of actives (**A** and **C**) and similar inactives (**B** and **D**) over the top 100 ranks after docking against the progesterone ligand-binding domain with increasing amounts of sampling. Results are shown for dockings obtained with energy (A and B) and chemical scoring (C and D), respectively. Ranks are divided in bins of ten compounds and include a bin for molecules docked with ranks larger than 100. The percentages of compounds with rankings larger than 100 have been truncated to 50% to allow for comparison with Figure 5.

molecule databases. It is surprising that for both the rather flexible thrombin inhibitors, as well as for the fairly rigid PR-LBD ligands, five peripheral seeds per incremental reconstruction step appear to be sufficient to identify the majority of the known ligands. In fact, applying more conformational sampling improves the scoring and ranking of inactive compounds at the expense of truly active molecules. Also, more intensive sampling of the base fragment placement reduces the retrieval of actives, especially when energy scoring is applied. Differences in flexibility among the three sets of molecules, that may cause the more flexible compounds to specifically obtain better scores with increased sampling, do not explain these results. The known thrombin inhibitors, for instance, have on average more flexible bonds than the diverse compounds, while this is reversed for the PR-LBD ligands (results not shown). Since in the examples presented here correct placement of the base fragment is facilitated by the nature of the binding sites, this could well be an important determinant for favorable ranking of active molecules. The presence of charge-charge interactions between thrombin inhibitors and the enzyme, for instance, clearly aids in obtaining a larger enrichment (94%) than for the uncharged PR-LBD ligands (64%) even though the total scores are dominated by van der Waals interactions. Alternatively, many small hydrophobic molecules may bind the PR-LBD but fail to activate the receptor in our assay. Although the test cases used in this study can be considered as close to ideal for (flexible) docking due to the nature of the respective binding sites, the results of others on dihydrofolate reductase[12,14] are comparable to ours.

Our results suggest that in practice flexible database docking may be performed with less sampling than would seem to be required for accurate reconstruction of protein-ligand complexes. Flexible docking of ligands to thrombin and the PR-LBD using five peripheral seeds and both scoring functions still requires on average 274 and 138 seconds per compound, respectively, on an SGI R10000 operating at 194 MHz. At 50 peripheral seeds, the average CPU time required per compound reaches a plateau near 1,000 seconds for both test systems. Although this remains computationally expensive, these numbers can roughly be divided by two if only one scoring function is used. This requires, however, that the optimal scoring function would be known in advance. In combination with pre-screening of molecule databases for compounds with a favorable profile in terms of drug-likeness, flexibility, oral availability and synthetic accessibility, this could bring flexible docking of larger numbers of compounds within reach.

Less conformational sampling also appears to yield a somewhat improved discrimination between true actives and inactives. When sampling is increased to 10–50 peripheral seeds, the scoring is improved for all compounds, resulting in more inactives acquiring high ranks. Apparently, both energy and chemical scoring have difficulties with separating true actives from chemically similar inac-

tives, with the exception of the application of energy scoring to the PR-LBD ligands. This is, however, not necessarily detrimental since the experimental testing of inactive compounds, that are chemically similar to true actives, may still provide useful structure-activity relationships.

Given a background of 1,000 unrelated molecules, the enrichment observed in actives and related inactives is encouraging. Although not all actives are retrieved as high-ranking dockings, the enrichment observed is sufficient to allow effective prioritization of screening databases. With 64–94% of the known actives being ranked among the 10% best scoring molecules, the majority of the biologically active compounds would be identified in the initial stages of screening. Although the scoring functions implemented in DOCK 4.0 favor other ligand conformations than the native-bound state in some cases,[18] this degeneracy does not appear to prevent the identification of potentially interesting lead compounds. For instance, DOCK has previously identified an inhibitor for HIV protease even though the predicted binding mode was in error.[33] Such errors are partially due to the fact that small changes in protein conformation or bound solvent and ions can have a large influence on the predicted binding mode of docked molecules.[18,33,34] Nevertheless, docking can identify molecules with intrinsic complementarity to ligand binding sites, even though the suggested bound conformations need not be correct as was discussed earlier by Blaney and Dixon.[34] The availability of high-throughput screening alleviates some of the need for docking protocols to rank all actives in a relatively small number of the highest-ranking compounds. In conclusion, flexible docking with DOCK 4.0 is able to select chemical classes of compounds that are potentially active but may fail to identify subtle features of importance to activity and selectivity.

A more disturbing result from the analysis presented here is that apparently different systems require different scoring functions for optimal results. For thrombin, the best enrichment is obtained by applying chemical scoring, while for the PR-LBD energy scoring performed best. This casts some doubts on the odds of deriving a truly universal scoring function. For instance, for a general forcefield-based scoring function, such as implemented in DOCK 4.0, similar results were expected for both test systems. In addition, some dependence of docking efficacy on the size of the docked molecules is observed. The inclusion of entropic effects such as desolvation and the reduction of internal degrees of freedom in the ligand and protein may compensate, however, for the differences observed with the energy-based scoring schemes implemented in DOCK 4.0. Rescoring with energy scoring of the compounds that were docked against the PR-LBD using chemical scoring yielded enrichments similar to those obtained when using energy scoring during the docking. This indicates that the actual docking may be performed with a faster and simpler scoring function, after which the final ranking can be performed using a more sophisticated scoring function.

The observation that different proteins require different scoring functions for optimal retrieval of actives from 3D databases clearly complicates the practical use of docking approaches. A broader comparison, including scoring functions that are not based on molecular mechanics energies and using the same conformational sampling protocol, would be required to establish whether the trend observed here is general. In practical applications, performing preliminary dockings and assays on a limited set of compounds could aid in choosing the optimal scoring function. Alternatively, a consensus score could be calculated by combining the rankings obtained with several different scoring functions. Our results do indicate that increased conformational sampling does not necessarily result in improved retrieval of actives. This effect is independent of the protein structure or scoring function used in this study. Increased sampling improves the scores of all molecules in the database and thereby complicates the differentiation between actives and inactives. On the basis of the data presented here, using the minimal amount of sampling that yields convergence for the average scores of a small test set of compounds is advantageous in terms of required computational time and enrichment.

CONCLUSIONS

The molecular docking program DOCK 4.0 was used to analyze the efficacy of flexible database docking in retrieving biologically active molecules from a database of compounds with known activity against thrombin and the progesterone receptor. The introduction of ligand flexibility in database docking improves the retrieval of known actives from 3D molecular databases compared to docking single, rule-based conformations. In the order of 64–94% of the known actives are retrieved within the top 10% of the ranked database. A limited amount of conformational sampling appears to be optimal, however, in retrieving actives and discriminating between actives and inactives. Increased conformational sampling results in an improvement of the scores of all compounds, resulting in competition between actives and inactives for the highest ranks. A similar effect is observed when the number of attempted base fragment placings is increased. This phenomenon can be interpreted as an “overfitting” of the conformational search with respect to the accuracy of the scoring functions and the appropriateness of the protein conformation used for docking. For DOCK 4.0, the use of five peripheral seeds per ligand incremental construction step was found to give an optimal enrichment of the top 10% of the ranked database. One may speculate that true ligands have some intrinsic properties that facilitate their identification compared to inactive compounds when using only limited conformational sampling. We conclude that effective database docking can be performed with less conformational sampling than would seem appropriate for correct reconstruction of the bound conformations.

In both our test cases, flexible docking with DOCK 4.0 has difficulty discriminating between true actives and chemically similar inactive compounds. Both sets of compounds are selected among the top 10% best scoring molecules to roughly the same extent. In practice, this is not necessarily detrimental since selecting the highest-

ranking compounds for experimental testing would still yield valuable structure-activity relationships on the true actives and related inactives.

The performance of chemical and energy scoring in the retrieval of actives and discriminating between actives and related inactives was found to depend on the test system at hand. For thrombin, chemical scoring yielded the largest enrichment while for the PR-LBD energy scoring performed best. Overall, chemical scoring appears to yield a sharper discrimination in terms of ranking between actives and chemically related inactives with respect to the bulk of unrelated inactive molecules. The observation that the two scoring functions implemented in DOCK 4.0 yield significantly different results depending on what protein is used for docking, is illustrative for the complexity of deriving universal scoring functions. Although the inclusion of entropic terms in the scoring function may compensate for this, it would be interesting to investigate the dependence of the retrieval of true actives on available scoring functions, protein class and conformation in trials of a larger scale, such as the CASP-2 docking trials.[10]

ACKNOWLEDGMENTS

The authors thank Dr. J. Kelder of N.V. Organon for selecting the PR-LDB ligands and the chemists, biologists and pharmacologists that have contributed to the screening results used in this study.

REFERENCES

1. Kuntz ID, Meng EC, Shoichet BK. Structure-based molecular design. Acc Chem Res 1985;27:117–123.
2. Kubinyi H. Combinatorial and computational approaches in structure-based drug design. Curr Opin Drug Disc Dev 1998;1:16–27.
3. Lengauer T, Rarey M. Computational methods for biomolecular docking. Curr Opin Struct Biol 1996;6:402–406.
4. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins 1990;8:195–202.
5. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
6. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161:269–288.
7. Welch W, Ruppert J, Jain AN. Hammerhead: fast, fully flexible automated docking of flexible ligands to protein binding sites. Chem Biol 1996;3:449–462.
8. Gehlhaar DK, Verkhivker GM, Rejto PA, et al. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 1995;2: 317–324.
9. 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.
10. Dixon JS. Evaluation of the CASP2 docking section. Proteins 1997;1:198–204.
11. Leach AR, Kuntz ID. Conformational analysis of flexible ligands in macromolecular receptor sites. J Comp Chem 1992;13:730–748.
12. Makino S, Kuntz ID. Automated flexible ligand docking method and its application to database search. J Comput Chem 1997;18: 1812–1825.
13. Clark DE, Westhead DR. Evolutionary algorithms in computer-aided molecular design. J Comput Aided Mol Des 1996;10:337–358.
14. Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles. Protein Sci. 1998;7:938–950.
15. Miller MD, Kearsley SK, Underwood DJ, Sheridan SP. FLOG: a system to select "quasi-flexible" ligands complementary to a receptor of known structure. J Comput Aided Mol Des 1994;8:153–174.
16. Ajay, Murcko MA. Computational methods to predict binding free energy in ligand-receptor complexes. J Med Chem 1996;38:4953–4967.
17. Knegtel RMA, Grootenhuis PDJ. Binding affinities and nonbonded interaction energies. In: Kubinyi H, Folkers G, Martin YC, editors. 3D QSAR in drug design. Volume 2. Ligand-protein interactions and molecular similarity. Dordrecht: Kluwer Academic Publishers; 1998. p 115–127.
18. Knegtel RMA, Bayada DM, Engh RA, von der Saal W, van Geerestein VJ, Grootenhuis PDJ. Comparison of two implementations of the incremental construction algorithm in flexible docking of thrombin inhibitors. J Comput Aided Mol Des 1999;13:167–183.
19. Sansom C. Extending the boundaries of molecular modeling. Nat Biotech 1998;16:917–918.
20. Hill DC. Trends in development of high-throughput screening technologies for rapid discovery of novel drugs. Curr Opin Drug Disc Dev 1998;1:92–97.
21. Charifson PS, Kuntz ID. Recent successes and continuing limitations in computer-aided drug design. In: Charifson PS, editor. Practical application of computer-aided drug design. New York: Marcel Dekker Inc.; 1997. p 1–37.
22. Ewing TJA, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comp Chem 1997;18:1175–1189.
23. Ewing T. DOCK Version 4.0 User Manual, San Francisco, California, Regents of the University of California, 1997.
24. Brandstetter H, Turk D, Hoeffken HW, et al. Refined 2.3 Å 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.
25. Williams SP, Sigler PB. Atomic structure of progesterone complexed with its receptor. Nature 1998;393:392–396.
26. Gschwend DA, Sirawaraporn W, Santi DV, Kuntz ID. Specificity in structure-based drug design: identification of a novel selective inhibitor of Pneumocystis carinii dihydrofolate reductase. Proteins 1997;29:59–67.
27. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. J Comput Chem 1992;13:505–524.
28. Barnard JM, Downes GM. Chemical fragment generation and clustering software. J Chem Inf Comput Sci 1997;37:141–142.
29. Sadowski J, Gasteiger J, Klebe G. Comparison of automatic three-dimensional model builders using 639 X-ray structures. J Chem Inf Comput Sci 1994;34:1000–1008.
30. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. Tetrahedron 1980;36:3219–3288.
31. Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all atom force field for simulations of proteins and nucleic acids. J Comp Chem 1986;7:230–252.
32. Brzozowski AM, Pike AC, Dauter Z, et al. Molecular basis of agonism and antagonism in the oestrogen receptor. Nature 1997; 389:753–758.
33. Rutenber E, Fauman EB, Keenan RJ, et al. Structure of a non-peptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design. J Biol Chem 1993;268: 15343–15346.
34. Blaney JM, Dixon JS. A good ligand is hard to find: automated docking methods. Perspect Drug Disc Des 1993;1:301–319.