Journal of
CHEMICAL INFORMATION
AND MODELING

Article

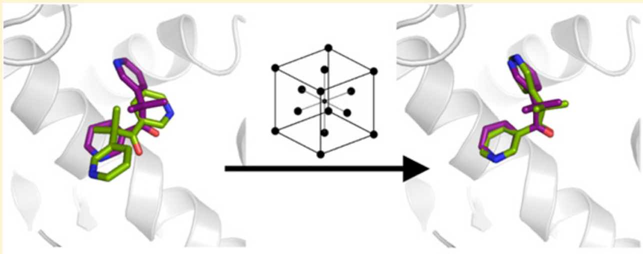
pubs.acs.org/jcim

Use of Experimental Design To Optimize Docking Performance: The Case of LiGenDock, the Docking Module of Ligen, a New De Novo Design Program

Claudia Beato,[†] Andrea R. Beccari,[†,‡] Carlo Cavazzoni,[§] Simone Lorenzi,[‡] and Gabriele Costantino[†,*]

[†]Dipartimento di Farmacia, Università degli Studi di Parma, Viale Area delle Scienze, 27/A, 43124 Parma, Italy
[‡]Dompé S.p.A., Via Campo di Pile, 67100 L'Aquila, Italy
[§]CINECA, Via Magnanelli 6/3, 40033 Casalecchio di Reno (BO), Italy

S Supporting Information



ABSTRACT: On route toward a novel de novo design program, called LiGen, we developed a docking program, LiGenDock, based on pharmacophore models of binding sites, including a non-enumerative docking algorithm. In this paper, we present the functionalities of LiGenDock and its accompanying module LiGenPocket, aimed at the binding site analysis and structure-based pharmacophore definition. We also report the optimization procedure we have carried out to improve the cognate docking and virtual screening performance of LiGenDock. In particular, we applied the design of experiments (DoE) methodology to screen the set of user-adjustable parameters to identify those having the largest influence on the accuracy of the results (which ensure the best performance in pose prediction and in virtual screening approaches) and then to choose their optimal values. The results are also compared with those obtained by two popular docking programs, namely, Glide and AutoDock for pose prediction, and Glide and DOCK6 for Virtual Screening.

INTRODUCTION

Computational approaches are more and more often integrated into drug discovery programs, aiming at accelerating hit identification and lead optimization.[1] Because the number of available protein three-dimensional structures has rapidly increased during the last past years, a widely used approach is molecular docking, which aims to predict the three-dimensional disposition assumed by a ligand in protein−ligand complexes. From a more formal point of view, molecular docking searches a global optimum in an energy landscape defined by the scoring function, protein, ligand, and degrees of freedom to be explored.[2] To address this complex problem, all docking programs are divided into two main parts, namely, the search of the ligand disposition, guided by the docking algorithm, and the scoring function, that tries to make a correct prediction of the energetics of interaction and, thus, of the biological activity.[3] In the development of a new docking algorithm, there are two key features of docking simulations that one should always keep in mind: speed and accuracy.[4] The main objective is to obtain a fast method that is able to screen molecular libraries to discover novel lead compounds (in virtual screening) and also to reproduce experimental ligand conformation (i.e., crystallographic ligand pose).

The docking process begins with the application of the docking algorithm to find a pose of the ligand molecule in the active site. Sampling the degrees of freedom is not a trivial task because even relatively small organic molecules can contain many conformational degrees of freedom, and the process must be performed with a certain accuracy to identify the conformation that best matches the receptor active site.[1] Then the different proposed ligand−receptor complexes are evaluated and ranked by the scoring function.

The first molecular docking programs were used to treat both the protein and ligand molecule as rigid bodies, fixing all the internal degrees of freedom, except for the three translations and three rotations.[5] Within this concept, the only way to address the conformational flexibility of the ligands is to pregenerate a library of ligand conformations that are formally treated as separated molecular entities. Examples are FRED[6] and DOCK4.0.[7,8] These approaches were quickly replaced by algorithms able to explicitly take into account the conformational degrees of freedom during docking. Several ligand flexibility algorithms have been proposed and can be divided into three main families according to the type of search: systematic, stochastic, and deterministic or simulation methods.[1,9]

To the first family belong those algorithms that try to explore all the degrees of freedom in a molecule, as QXP, that carries out full conformational searches for flexible cyclic and acyclic molecules[10] with extremely good results.[11] The main issue of this kind of approach is the combinatorial explosion of ligand conformations number. Therefore, ligands are generally first

Received: February 1, 2013

divided into fragments and then incrementally grown within the binding pocket;[12] examples are FlexX,[13] Surflex,[14] and eHITS.[15]

Stochastic-based algorithms make random changes, usually changing one degree of freedom at time;[3] examples are Monte Carlo (MC) methods and evolutionary algorithms applied, for example, by ICM[16] and AutoDock.[17]

Molecular dynamic is the most used deterministic approach. However, the main concern regarding this kind of method is that often molecular dynamic is not able to cross high-energy barriers within short simulation time, and therefore, the system gets trapped in local minima.[1,3] To avoid this problem, several attempts have been proposed, such as starting the simulation from different ligand positions or simulate the system at different temperatures; however, these efforts are quite expensive in terms of calculation time, limiting the application of molecular dynamic to docking only one or few compounds.

Regardless of the way in which the docking process is handled, all the available docking software share at least the feature that the docking process is controlled by the values of several user-adjustable parameters, and an appropriate choice of these parameters is a prerequisite to obtain meaningful results.[18] However, in many applications, users tend to adopt default settings, assuming that they will yield reasonably good results, regardless of the specific problem in which they are involved. Thus, providing the best ensemble of "default settings" is crucial for optimizing the program's output under standard conditions. Furthermore, having an optimized set of default parameters enables benchmarking of the performance of the program at the best of its possibility. As previously reported (see, for example, Andersson et al.[19]), experimental designs provide the best way to find the optimal ensemble of parameters by changing all of them simultaneously in a controlled and systematic way.

On route toward a novel suite of programs for de novo design, which we called LiGen (LigandGenerator) and presented in another paper,[20] we developed a docking protocol based on the definition of pharmacophore models of the binding site. In more detail, LiGen consists of a set of modules that work sequentially or as standalone tools, according to the user's need. It contains four main modules: LiGenPass for binding site recognition, LiGenPocket for binding site analysis and structure-based pharmacophore definition, LiGenDock for docking and virtual screening, and LiGenBuilder for de novo design.

In this paper, we report a description of the main functionalities of LiGenPocket and LiGenDock, the two modules constituting the docking engine of LiGen, as well as the optimization procedure we have carried out to improve the cognate docking and virtual screening performances. Design of experiments (DoE) methodology was used first to identify which user-adjustable parameters are influencing the results most and then to choose their optimal values. The results of LiGenDock have then been compared with those obtained by two popular docking programs, namely, Glide and AutoDock for pose prediction, and Glide and DOCK6, for virtual screening.

■ MATERIALS AND METHODS

LiGen consists of a set of tools that can be combined in a user-defined manner to generate project-centric workflows. The main algorithm as well as a detailed description of how LiGen works are reported in another article.[20]

LiGenPocket. LiGenPocket computes volume, shape, and physicochemical properties (donor, acceptor, hydrophobic, etc.) of the binding pocket and proposes a pharmacophore model based on these characteristics.

LiGenPocket accepts as input a three-dimensional structure of the protein of interest in PDB or MOL2 file format. The basic algorithm of LiGenPocket is a variant of the one proposed in 2000 by Wang et al.[21] Briefly, LiGenPocket creates a regular Cartesian grid (grid spacing 0.5 Å) around the co-crystallized ligand, if there is one, or around an active site point (ASP) generated by LiGenPass, indicating the center of mass of the binding cavity, as described in another paper.[20] In the case of the co-crystallized ligand, the software first defines a sphere around the ligand (with a user defined radius) and then creates the grid inside it. In the first step, a hydrogen atom is used as a probe to check the accessibility of the grid points. If the probe bumps into the protein, that grid point will be labeled as "not free"; otherwise, it will be labeled as "free". A bump is counted when the interatomic distance is less than the sum of the van der Waals radii reduced by 0.5 Å. In the second step, the possible interaction sites will be derived using three types of probes to map the main interactions usually occurring in binding sites: a positively charged sp3 nitrogen atom (ammonium cation), representing a hydrogen bond donor; a negatively charged sp2 oxygen atom (as in a carboxyl group), representing a hydrogen bond acceptor; and a sp3 carbon atom (methane), representing a hydrophobic group. A score representing the binding energy between the probe and the protein will be calculated. To calculate the scores, LiGen uses an in-house developed scoring function based on the paper of Wang et al.[22] as reported in another paper.[20] In this way, all the grid points are mapped and assigned three scores, representing the three binding energies of the interactions with the three kinds of probes. In general, however, not all of them would be worthy of being considered for the pharmacophore model definition. For this reason, only those grid points having at least one of the three score higher than the average score for that kind of interaction are retained. Then, the survived grid points are labeled as "H-bond donor", "H-bond acceptor", or "hydrophobic", according to the highest score reached. The number of "neighbors", defined as the number of grid points with the same label falling within a certain user-defined distance, is computed for every survived grid point. The average number of neighbors of the same type is calculated for each grid point, and only those having a number of neighbors higher than the average are retained for the further step and defined as "key sites". Finally, the survived grid points forming the key sites are clustered, and the geometric center of each cluster represents a pharmacophoric feature.

In the binding site analysis, several parameters can be tuned by users according to their own needs. For example, the minimal distance among the pharmacophoric features, or the grid spacing, can be modified through the grid accuracy parameter, selecting the desired degree of accuracy in the pocket description. All the user adjustable pocket parameters are reported in Table 1, along with a brief explanation of their functionalities.

LiGenDock. The main feature of LiGenDock is the use of the pharmacophore scheme generated by LiGenPocket as the driving force for the docking procedure, including a non-systematic flexible docking algorithm. A simple description of the framework of the docking algorithm is the following:

1. One ligand is taken into account, and ligand features (e.g., H-bond donor site) are computed.

Table 1. LiGenPocket Parameters Definition

parameter	abbreviation	notes
minimal feature distance	Min F dist	sets the minimal distance between the pharmacophore features identified in the binding pocket. (1 to 5 Å)
maximal feature number	Max F num	specifies maximal number of features that can be considered in the pharmacophore that describes the binding pocket
distance cutoff	Dist CO	sets the cutoff radius to search for pocket atoms around a ligand (when pocket is computed around a co-crystallized ligand)
van der Waals bumps	Prot VDW B	defines how much the binding site surface is smoothed (represents the fraction of the van der Waals radius considered to tag the grid points)
grid accuracy	Grid Acc	specifies the grid spacing used in grid generation through the expression: grid spacing = 1.0 Å/grid accuracy
ligand neighbor threshold	Lig Neig Thr	represents the ligand−ligand distance threshold used to count ligand neighbor atoms; used for coarse graining ligand atoms
score distance threshold	Score Dist Thr	assigns a score to the identified pharmacophore points; specifies the maximum distance to take into account for the grid points around the pharmacophore point; sum of each single grid point score gives the score of the pharmacophore point
grid distance threshold	Grid Dist Thr	defines for every grid point the area of where to count the number of grid points of the same type; number of grid points found is used to compute the score of the grid point taken into account
coarse grain ligand		enables to apply a filter to coarse grain the ligand or the cluster of probes generated by LiGen-Pass to speed up the calculation
include H bumps		allows to consider hydrogen atoms during the calculation of grid points that bump the receptor
include water		allows to include water molecules in the calculation of the receptor grid

Table 2. LiGenDock Parameters Definition

parameter	abbreviation	notes
neighbor threshold	Neig Thr	number of neighbors of grid points of the same type necessary for a pharmacophore point to be considered as a candidate for ligand docking
distance threshold	Dist Thr	maximum distance to consider a ligand functional group superposing on a pharmacophore feature
number of poses		defines number of output poses
pose overlap	Pose Over	maximum degree of overlap between two poses
pose diversity		sets a limit to the number of poses of a molecule
hydrophobic threshold	Hyd Thr	value of atomic LogP above which an atomic site is considered hydrophobic; used in scoring the interaction between the receptor and ligand
angle delta	Ag Delta	defines the extent of the angle used to rotate the ligand inside the pocket around the axis of the two matched pharmacophoric features
conformer van der Waals bumps	Conf VDW B	defines degree of ligand volume smoothing; represents the fraction of the van der Waals radius to be considered when computing bumping between fragment during conformer generation.
conformer angle delta	Conf Ag Delta	defines extent of the angle used to rotate rotatable ligand bonds during conformer generation

2. The docking process starts matching a ligand's feature (i.e., a hydrogen bond donor site) with the previously identified pharmacophoric features of the same type.
3. The docked ligand is rotated by an appropriate angle to match a second pharmacophore feature with a second ligand's feature. Because it is unlikely that the second pair will overlap perfectly, a user defined tolerance cutoff is used to evaluate the goodness of the match.
4. The ligand is then rotated by an appropriate angle around the axis passing between the two pharmacophore features trying to match a third feature (not necessary).
5. Then torsional angles are unlocked, and ligand conformers are generated in situ trying to match as many features as possible (some torsional angles may be selectively locked by the user).
6. At every step, the pose's score, related to the estimated binding energy of the ligand-protein complex, is computed and compared with the scores of previously generated poses. If this actual score is better than the worst score of the already generated poses, the new pose is retained instead of the previous worst pose (the one with the lowest score). The risk of getting trapped in a local minimum can be minimized by imposing a high RMSD difference between two poses to be considered different and further processed.
7. Finally, the score is optimized with a simple score minimization algorithm that treats the docked ligand as a rigid body inside the pocket. The ligand is rotated in space around the docking pose until a minimum score, representing the most favorable binding energy, is reached. This minimization is the steepest descendent minimization during which the ligand position inside the binding site is changed by a discrete value of 0.25 Å in seven directions of the three-dimensional space (3 axes and 4 quadrant bisectors), and then the direction of the diminishing score is taken.

Several parameters allow users to tune the docking algorithm according to their needs. For example, the distance threshold parameter allows the user to change the tolerance radius for considering a ligand feature matching a pharmacophore feature, or the angle delta parameter allows the user to define the degrees of the angle used to rotate the ligand inside the binding site. So, by using a smaller angle, one performs a more exhaustive exploration of the possible ligand orientations inside binding pocket. Also, some constraints are implemented in the algorithm, as the maximum degree of overlap among different poses of the same ligand, that can be set by the users. A complete list of the parameters with a brief explanation of their meaning is given in Table 2. The values of these parameters can have a large influence on the outcome of the docking process and a suitable set of parameters is necessary for gaining good results.

Docking outputs are a collection of ligand poses in mol2, pdb, or sdf file format, and a table summarizing scores and ligand/pharmacophore feature matches.

Table 3. DUD Complexes Used in This Study

protein	PDB code	resolution (Å)	no. ligands	no. decoys
Nuclear Hormone Receptors				
ERagonist	1l2i	1.9	67	2361
ERantagonist	3ert	1.9	39	1399
GR	1m2z	2.5	78	2804
MR	2aa2	1.9	15	535
PPARg	1fm9	2.1	81	2910
PR	1sr7	1.9	27	967
RXRa	1mvc	1.9	20	708
Kinases				
CDK2	1ckp	2.1	50	1780
EGFr	1m17	2.6	416	14914
FGFr1	1agw	2.4	118	4216
HSP90	1uy6	1.9	24	861
P38 MAP	1kv2	2.8	234	8399
SRC	2src	1.5	162	5801
TK	1kim	2.1	22	785
Serine Proteases				
FXa	1f0r	2.7	142	5102
thrombin	1ba8	1.8	65	2294
trypsin	1bju	1.8	43	1545
Metalloenzymes				
ACE	1o86	2.0	49	1728
COMT	1h1d	2.0	12	430
PDE5	1xp0	1.8	51	1810
Folate Enzymes				
DHFR	3dfr	1.7	201	7150
GART	1c2t	2.1	21	753
Other Enzymes				
AChE	1eve	2.5	105	3732
ALR2	1ah3	2.3	26	920
AmpC	1xgj	2.0	21	734
COX-1	1p4g	2.1	25	850
COX-2	1cx2	3.0	349	12491
GPB	1a8i	1.8	52	1851
HIVPR	1hpx	2.0	53	1888
HIVRT	1rt1	2.6	40	1439
HMGR	1hw8	2.1	35	1242
InhA	1p44	2.7	85	3043
NA	1a4g	2.2	49	1745
PARP	1efy	2.2	33	1178
PNP	1b8o	1.5	25	884
SAHH	1a7a	2.8	33	1159

Data Set/Benchmark Composition. During this study, three different data sets were considered.

For the optimization of the cognate docking, i.e., the reproduction of the crystallographic ligand conformation, we used as a training set 100 crystallographic complexes taken from the CCDC Astex clean data set, consisting of 224 entries.[23] Proteins were selected according to their relative family abundance in the PDB database.[24] First, the 224 complexes were grouped according to the protein family they belong to, and then the percentage abundance of those families in the whole PDB database was calculated. Finally, in agreement with the family representation percentage in the PDB, a hundred protein were selected out of the 224.

To test the optimized parameters for cognate docking, as well as to compare LiGen performance with those of commonly available docking software, we selected 171 complexes taken from the PDBbind CORE SET database (2010 release),[25,26] excluding entries containing a ligand with molecular weight higher than 500 Da.[27]

The third benchmark, used to optimize and test the VS ability of the LiGenDock algorithm, was obtained by selecting 36 high-quality crystal structures from the directory of useful decoys (DUD).[28] All complexes taken into account in this study are listed in Table 3. In the present study, four targets from DUD were excluded from the selection: (1) the human vascular endothelial growth factor receptor 2 kinase domain (VEGFr2, PDB code: 1vr2) because it lacks a co-crystallized ligand that we used to center the binding site grid (even though we could have used LiGenPass to define the binding site location, we preferred to use the same approach, i.e., using the co-crystallized ligand to center all the binding site grids; therefore, we excluded this complex), (2) the platelet-derived growth factor receptor kinase (PDGFrb) because it is an homology model, (3) the androgen receptor (AR, PDB code: 1xq2) because it supersedes in the PDB by 2ao6, and (4) the adenosine deaminase (ADA, PDB code: 1vdw) because there was a mismatch with the PDB code published in the original paper. In the optimization phase, we randomly select 10 targets among those selected from the DUD to reduce the computational efforts needed (Table S1, Supporting Information). To evaluate the improvements gained through the optimization procedure, the experiments were performed with default and optimized parameters, using all 36 targets selected from the DUD.

Protein Preparation. Water molecules were removed from all the considered proteins. The protein protonation state of complexes taken from the CCDC Astex database was retained as provided by Astex. Protein structures of complexes taken from PDBbind and DUD were prepared using Protein Preparation Wizard,[29] contained in the Maestro suite, undergoing the following preparation steps: (a) hydrogen atoms were added according to the protonation state at pH 7.0, (b) ions and crystallization cofactors were removed, (c) atom and bond types were assigned, and (d) an energy minimization in OPLS2005 was run to refine the structure.

Ligands Preparation. Ligand molecules were prepared using LigPrep of Maestro.[30] Cognate docking involves the redocking of the co-crystallized ligand to see whether the docking process is able to reproduce the crystallographic conformation (also called self-docking). Using a starting conformation different from the co-crystallized one is therefore important to evaluate the ability of the software to reproduce the crystallographic ligand conformation. Hence, for every ligand of the selected CCDC Astex and PDBbind CORE SET complexes, a set of conformers, not containing the co-crystallized conformation, was generated using ConfGen,[31] also included in Maestro suite. Then, a conformer was randomly chosen from this set as a starting conformation for docking. For rigid ligands, no conformers can be generated, so the only option was to assign 3D coordinates different from those in the PDB.

Experimental Design. In this study, we were interested in analyzing how the docking performance of LiGen is influenced by the different sets of parameters and in finding an optimal set of parameter that allows users to obtain the best results, both in pose prediction and virtual screening. Our goal was first to find out which parameters affect the results most and then to establish the values for optimal docking performance. Therefore, we needed at first to screen LiGen parameters to find out which of them were important for the outcome, and then we went into a second-phase of optimization to assign the optimal values to the previously identified parameters. The general workflow applied is summarized in Figure 1.

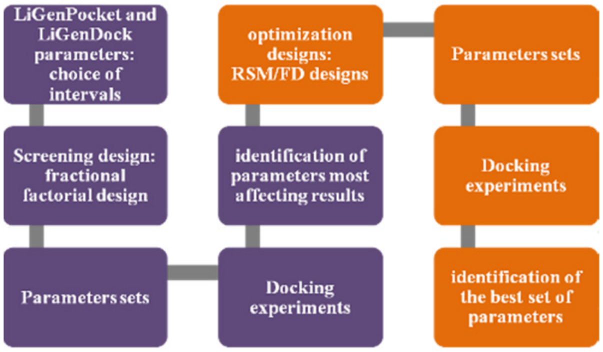

LiGenPocket and LiGenDock parameters: choice of intervals
optimization designs: RSM/FD designs
Parameters sets
Screening design: fractional factorial design
identification of parameters most affecting results
Docking experiments
Parameters sets
Docking experiments
identification of the best set of parameters

Figure 1. Schematic flowchart of the study: first five blocks (purple) represent the steps applied during the screening procedure to identify parameters most affecting results. The second four blocks (orange) represent the steps of the optimization protocol for cognate docking and virtual screening.

The docking algorithms are generally used for pose prediction and for virtual screening. Pose prediction and virtual screening have different goals: the goal of the first one is to predict how a ligand may bind, assuming that ligand can bind, whereas the aim of the second one is to predict whether a ligand can bind or not.[32] Because they have different aims, the parameters to use can be slightly different; therefore, we decided to optimize parameters for the two main docking applications independently.

Statistical experimental designs represented a good solution for our needs because they offer the possibility to vary all the parameters under investigation at the same time. Responses evaluated in the models were, for cognate docking, the number of poses with RMSD form the co-crystallized conformation less than 2 Å, whereas for virtual screening, the early enrichment, assessed by the value of the area under the receiver operating characteristic curve (ROC) measured at 1% of the screened database (ROC(1%)).

Multiple linear regression (MLR) was used to investigate the relationship between the docking parameters (independent variables or predictors) and the results (dependent variables or predictand). MLR is based on least-squares:[33] the model is fitted such that the sum-of-squares of differences of observed and predicted values is minimized. The general equation of the model is

$$\underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{y} = \underbrace{\begin{pmatrix} f_1(x_1)\ldots f_p(x_1) \\ \vdots \quad \ddots \quad \vdots \\ f_1(x_n)\ldots f_p(x_n) \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{\varepsilon}$$

in which y_i is the response or dependent variable, x_i is the input or independent variable and $f_j(x_i)$ is a function of the input variable x_i (sometimes the variable x_i can be a function of the data). The global response y is expressed as a linear combination of model terms $f_j(x)$ $(j = 1,...,p)$ at each of the observations (x_1, y_1), ..., (x_n, y_n), β is the coefficient of the parameters, and ε is the residual term associated to the experiments. The function $f_1(x) = 1$ is included among the f_j, so that the model contains a constant term (the intercept). Coefficients resulting from the design model were used to interpret parameters' influence on the docking performance.

For our first aim, the identification of the most relevant parameters affecting the docking results, we applied a fractional factorial design (FFD) (Figure 2B). All experimental designs were performed using MATLAB software.[34] FFDs are experimental designs in which only an adequate chosen subset (fraction) of the full factorial experiment is selected to be run,[35] especially those of resolution III that do not estimate the interaction effects are particularly useful in a first screening phase, where the

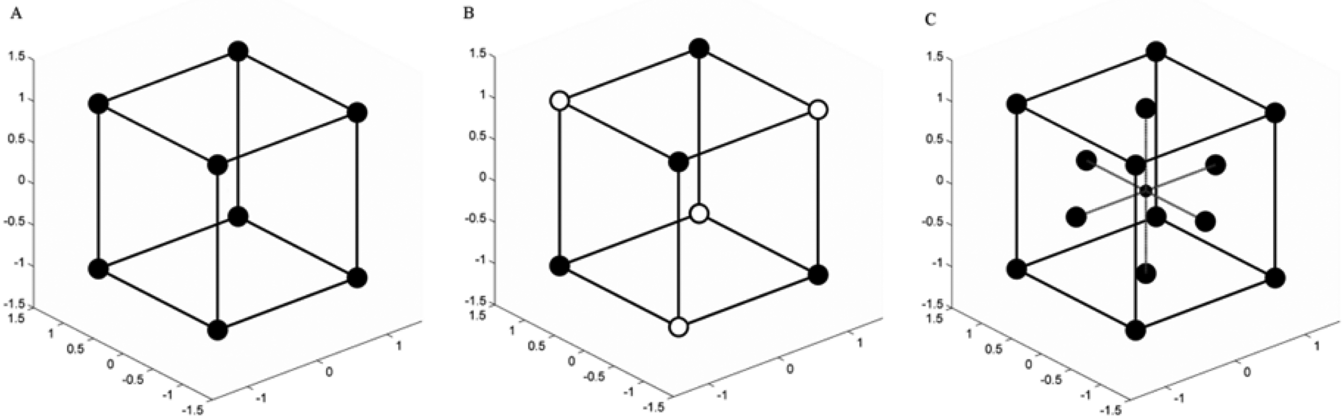


Figure 2. General representation of experimental designs applied in this study. (A) A 2^3 two-level full factorial design (FD); the exponent represents the number of factors investigated, 2 is the number of level for each factor (−1 or low, +1 or high). Every circle in the picture represents a run (or experiment) of the design, characterized by a unique combination of three factors. (B) A 2^{3-1} fractional factorial design (FFD); 2 is the number of levels, the first part of the exponent (3) is the number of factors, and that is the same exponent of the corresponding full factorial design; the second part (−1) represents the degree of fractionation. The resulting number of the subtraction (4) is the number of runs of the design, represented in the picture with filled black circles. (C) Faced central composite designs (CCD) for two parameters with three levels each contains a full factorial design (or a fractional factorial) with center points, and external points that allow estimation of curvature are the center of each face of the factorial space. Each circle point represents an experiment.[35]

most significant set of factors are sought.[35−37] For every quantitative parameter (i.e., parameters for which a numerical value can be assigned), high and low values were selected, together with a center point, representing the three levels in the design procedure. The range of values for the parameters' intervals was selected in order to be large enough to be sure to capture the effect of the parameter, if there is one. We decided to exclude qualitative parameters (i.e., parameters for which on/off values can be assigned) from our analysis.

After the key factors were identified, an optimization phase was performed using full factorial design FD (in the case of virtual screening optimization) or response surface method (RSM) design (in the case of cognate docking optimization; Figure 2A,C, respectively). FD is used to study the effect of each single parameter on the response variable and the effects of the interaction between parameters on the measured response.[35] RSM is a collection of mathematical and statistical techniques based on the fit of empirical models to experimental data obtained in relation to the experimental designs.[38,39] Studying parameters' variables at least at three levels allows us to determine first- and second-order effects and possibly also critical points (maximum, minimum, or saddle). For our study, we used a central composite design (CCD). CCD was first presented by Box and Wilson[40] and is commonly used in optimization procedures. A CCD consists of the following parts: (1) a full factorial or fractional factorial design, (2) an additional design, often a star design, in which experimental points are at α distance from its center, and (3) a central point. There are three types of CCD, depending on where the points of the star design ("star points") are located: (1) circumscribed, with star points located outside the factorial space, (2) inscribed, with star points located inside the factorial design space, used when points of the factorial design are real experimental limits, and (3) faced, with star points located on each face of the factorial space.[35] The first two types require five levels for each parameter, while the last one requires three levels. For our work, we chose a faced CCD because for some of the parameters only three levels were possible in the desired design space (Figure 2C).

During model fitting, statistics of the models were computed to evaluate the accuracy of the model[33]

- R^2 value, which represents the explanatory power of the regression model, computed from the sum-of-squares as

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

where SST is the total sum of squares, SSE is the sum of error squares, and SSR is the sum of squares due to the regression computed respectively as

$$\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\text{SSE} = \sum_{i=1}^{n} \hat{e}_i^2$$

$$\text{SSR} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

- F-ratio expressed by

$$F = \frac{\text{MSR}}{\text{MSE}}$$

where MSR is mean squared regression, and MSE the residual mean square. F-ratio represents the explanatory power of the model, but the advantage over R^2 is that F-ratio takes into account the degrees of freedom, which depend on the sample size and the number of predictors in the model. In this way, F-ratio incorporates sample size and number of independent variables in the assessment of significance of the relationship.

- *Adjusted R^2* attempts to compensate for the fact that R^2 for a regression can be made arbitrarily high by including more predictors in the model. *Adjusted* R^2 is given by

$$\bar{R}^2 = 1 - \frac{\text{MSE}}{\text{MST}}$$

where MST is total mean square.

- *p value* was calculated for each parameter, and its value represents the level of significance of the parameter. A value smaller than 0.05 means that the coefficient calculated by the model is significant with a confidence interval of 95%

Docking with Glide and AutoDock. The performances of LiGenDock in cognate docking were compared to the ones of two commonly used programs, namely, Glide and AutoDock. In the comparison, both accuracy and speed were considered and analyzed.

Docking with Glide[41] was performed using standard precision (SP) mode, using the default set of parameters, except for the number of required poses, which was changed to 10, so that the same number of poses was used for all programs (Glide, LiGen, and AutoDock).

AutoDock uses the Lamarkian version of the genetic algorithm to generate the ligand poses inside the protein active site.[17] In our test, we used the default parameter set.

Both in Glide and AutoDock, as well as with LiGen, the active site selection was based on the position of the native ligand in the crystallographic complex.

Evaluation of Pose Prediction and Virtual Screening Results. Along with the increased number of scientific papers reporting new docking software and/or docking software evaluation,[12,32,42−44] some articles with recommended guidelines for docking evaluation appeared[2,45] in the past years. Consistent with those recommendations, we used as response during the optimization of pose prediction the percentage of best-predicted poses with RMSD less than 2.0 Å from the experimental-solved ligand structure. In comparing LiGen results with those of other software, we calculated also the percentage of poses with RSMD less than 3.0 Å, both for best-predicted pose (the one with the lowest RMSD, irrespective of the ranking position) and best-scoring pose (the pose ranked first by the scoring function). Moreover, the computational time needed for docking was also calculated to better evaluate software performance. In the present study, to evaluate virtual screening performance, we used the area under the receiver operating characteristic (ROC) curve to measure the global enrichment; to evaluate the early enrichment, we used the values of the AUC under the ROC curve at 1%, 2%, 5%, and 20% of the x-axis, referred to hereafter as ROC(1%), ROC(2%), ROC(5%), and ROC(20%), respectively, as suggested by Repasky et al.[46] BEDROC, with a value of 20 for parameter α, as suggested by other works[47,48] was also calculated;

■ RESULTS AND DISCUSSION

Pose Prediction Optimization. As stated above, the two-fold aim of our study was to identify the subset of user-modifiable

Table 4. Original Results and after Optimization[a]

PDB code	RMSD original results	RMSD after optimization	PDB code	RMSD original results	RMSD after optimization	PDB code	RMSD original results	RMSD after optimization
1a4g	5.61	2.37	1fkg	6.07	5.01	1tph	2.12	1.79
1a9u	8.20	1.15	1frp	7.58	3.07	1tpp		0.73
1acj		0.76	1ghb	10.65	5.03	1trk		4.31
1acm	2.45	1.74	1glp	7.42	6.21	1tyl	4.5	1.16
1apu		5.13	1gpy	5.16	2.37	1ukz	7.11	1.46
1aqw	8	2.1	1hdc	13.99	2.18	1ulb	5.11	1.21
1ase	6.29	1.24	1hfc	6.01	2.52	1ydr		1.26
1b59	9.58	2.18	1imb	1.93	1.47	1yee	7.96	2.17
1bgo	13.28	3.3	1ivb	4.98	2.1	2ak3	6.6	2.6
1bl7	6.43	2.29	1ivq	12.83	5.84	2cht	1.98	0.98
1blh	4.03	1.74	1ldm	5.45	0.92	2cmd	2.99	0.73
1bmq	8.75	6.49	1mld	2.59	1.85	2cpp	1.47	1.09
1byb		4.15	1 mmq	4.87	3.02	2dbl	2.91	2.72
1byg	9.63	0.71	1okl	4.22	2.27	2fox	6.09	1.99
1cbs	10.99	1.41	1pbd	5.91	0.36	2h4n	9.53	1.83
1cdg		3.9	1pdz	1.58	1.73	2phh	4.05	0.26
1cil	6.1	2.16	1pgp	4.8	4.33	2qwk	9.61	2.53
1cle	23.54	3.25	1phd	4.32	1.4	2r07	11.91	2.17
1coy	1.12	1.15	1phg	5.78	0.72	2tsc	6.52	2.65
1cqp	8.95	2.09	1ppi		4.87	2yhx	5.88	4.35
1cvu		2.42	1pso	12.88	6.77	3cla	6.07	2.69
1d4p	12.33	1.58	1qbr	14.35	4.28	3cpa	6.65	2.62
1dd7	6.59	5.71	1rbp	12.62	2.22	3ert	11.25	1.82
1dhf	7.83	2.4	1rds	7.87	3	3hvt		3.28
1die	3.66	2.35	1rob	10.41	1.97	4aah		1.06
1dy9	7.9	3.88	1rt2	12.16	1.42	4cox	10.41	4.14
1ejn	11.42	6.71	1slt	4.99	1.9	4cts	1.04	1.05
1elc	9.5	4.15	1snc	7.12	1.8	4er2	13.1	8.61
1eta	4.77	2.33	1tdb	3.04	2.21	4fab	4.52	1.17
1ets	8.35	5.1	1tka		2.42	4phv	14.25	1.53
1ett	8.11	5.43	1tmn	9.45	5.08	4tpi	6.74	1.16
1f0s	5.2	1.65	1tng		4.95	5abp	2.05	2.83
1fen	12.32	2.09	1tni		2.94	7tim	10.07	1.39
1fgi		1.96						

	RMSD < 1 Å	RMSD < 2 Å	RMSD <3 Å	missing results	time (s)
original results	0%	6%	12%	15	26.49
after optimization	9%	41%	70%	0	27.23

[a]Results for the 100 proteins taken from the Astex Data set.[23] Original results and results after the optimization are the RMSD of the best predicted pose with respect to the crystallized ligand. In the last part of the table is reported a summary of the results before and after the optimization together with the number of eventually missing results and the time spent for the docking process (seconds).

parameters, which affect the docking results most, and to establish their optimal values. As a first step before proceeding in the optimization, we assessed the original LiGen performance by using the set of parameters assigned during the code development. As shown in Table 4 (column "RMSD original results"), results with the original set of parameters were not very good. Only in six cases out of 100 the best predicted pose has an RMSD less than 2 Å from the co-crystallized ligand. In 15 cases, LiGen failed to find a pose. Visual inspection of results showed that many poses are located outside the binding site, for example, for proteins 1bgo, 1bmq, 1cqp. Moreover, in many cases, also the pharmacophore models were not completely contained within the binding site, but pharmacophoric features were also present on the protein surface outside the cavity (1cbs, 1d4p, 1ett, 4tpi, etc.), highlighting the need for "smoothing" the molecular surface by scaling the van der Waals radii of receptor atoms (tuning the value of the parameter *van der Waals bumps*) to reduce penalties for ligand–protein close contacts (Figure 3A). In other experiments, for example, for 2phh and 1fkg, we found that even if the pharmacophore was placed inside the binding site the functional groups needed for binding were not or not completely positioned correctly (Figure 3B).

All the quantitative LiGen's parameters involved in the binding site characterization and in the docking process itself were selected to perform an experimental design to assess which among them had the most impact on results. A total of 15 parameters, eight from LiGenPocket and seven from LiGenDock, were selected. Selected parameters are those reported in italics in Table1 and Table 2. In building the binding site grid, nonpolar hydrogen atoms were not considered; therefore, the parameter *include H bumps* was excluded from the experimental design, as well as the coarse grain ligand parameter, which allows to speedup the grid-defining process (it was not included in the study because the investigation of the speedup process was beyond the scope of this paper). The 15 selected parameters were used to generate a

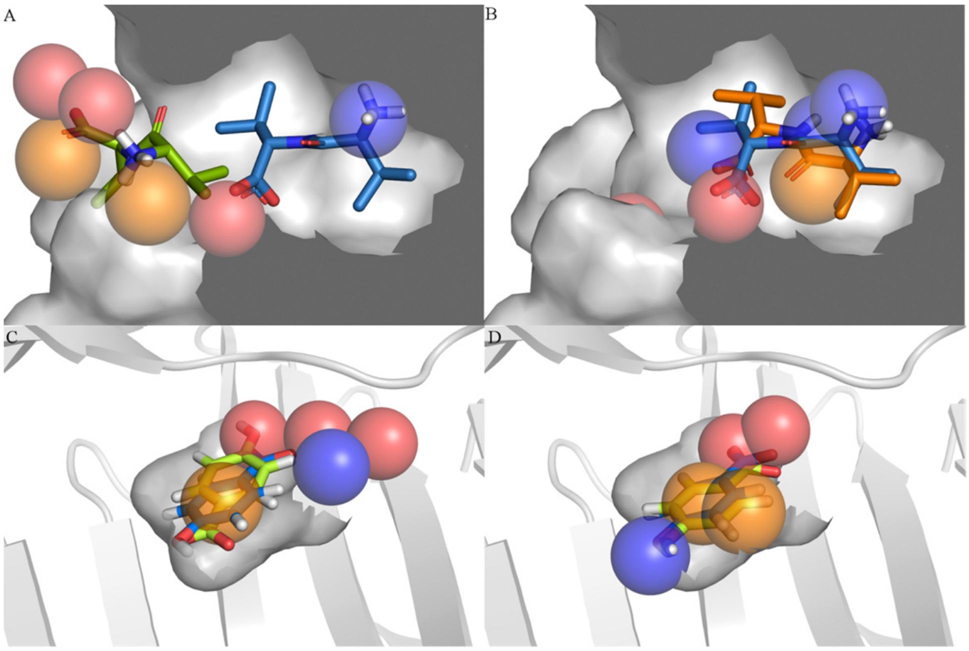


Figure 3. Examples of improvement in pose prediction. (A) Using the original parameters pose (lime green) placed outside the binding site with respect to the experimentally determined ligand pose (blue, complex PDB code: 4tpi). (B) With the optimized set of parameters, especially decreasing a little the van der Waals volume of the protein atoms forming the binding site, it is possible to produce a pose (orange) that overlaps quite well the crystallographic one. (C) The beginning pose (light gray) is flipped with respect to the crystal complex (cyan, PDB code 2phh) due to a nonoptimal position of the H bond donor/acceptor features of the pharmacophore, whereas (D) after optimization, the pharmacophore allows to generate a pose that completely overlaps the original one. Figures are prepared with PyMol.[56]

Table 5. Design Statistics

	R^2	R^2 adjust	F
design 1 (FFD)	0.7113	0.6727	184.007
design 2 (RSM)	0.9450	0.8971	197.258

design 1(FFD)	p value	design 2 (RSM)	p values		p values
Min F Dist	2.63×10^{-04}	Min F Dist	0.607	Vdw B Prot * Grid Acc	0.776
Max F Num	3.02×10^{-08}	Max F Num	0.001	Vdw B Prot * Dist Thr	0.13
Dist C.O.	0.984	Vdw B Prot	0.344	Vdw B Prot * Neib Thr	0.89
Vdw Bumps P	2.95×10^{-05}	Grid Acc	0.047	Grid Acc * Dist Thr	0.78
Grid Acc	7.87×10^{-28}	Dist Thr	0.336	Grid Acc * Neib Thr	1.11×10^{-12}
Lig Neib Thr	0.984	Neib Thr	0.157	Dist Thr * Neib Thr	0.89
Score Dist Thr	0.953	Min F Dist * Max F Num	0.002	Min F Dist * Min F Dist	0.91
Gris Dist Thr	0.116	Min F Dist * Vdw B Prot	0.323	Max F Num * Max F Num	0.01
Hyd Thr	0.381	Min F Dist * Grid Acc	0.396	Vdw B Prot * Vdw B Prot	0.27
Dist Thr	0.022	Min F Dist * Dist Thr	0.887	Grid Acc * Grid Acc	2.77×10^{-03}
Pose Over	0.800	Min F Dist * Neib Thr	0.125	Dist Thr * Dist Thr	0.67
Ag Delta	0.521	Max F Num * Vdw B Prot	0.396		
Conf Vdw B	0.953	Max F Num * Grid Acc	0.479		
Neib Thr	0.066	Max F Num * Dist Thr	0.054		
Conf Ag Delta	0.682	Max F Num * Neib Thr	0.015		

fractional factorial design of resolution III, with 128 experiments, plus one center point. The parameter intervals were chosen in order to have the value of the central point set at the original value for all parameters, except for those having as original value the lowest or the highest possible value. The resulting experiments, with the set of parameters and results, are reported in Table S3 of the Supporting Information. We obtained a significant model, whose statistics are reported in Table 5. The model is not of outstanding quality due to the fact that many combinations of parameters did not yield any results. A deeper analysis of these results correlates experiments with no or very poor results to the lowest grid accuracy value. However, some important conclusions can be drawn from statistics in Table 4.

It suggests that parameters that influence the cognate docking experiments most, are i) the minimal distance between two pharmacophoric features (minimal feature distance), ii) the maximal number of features identified in the binding site (maximal feature number), iii) how much the protein's van der Waals volume is smoothed (van der Waals bumps), iv) the grid spacing (grid accuracy) and v) the tolerance in considering a ligand functional group superposed to the feature (distance threshold). All these parameters have a p value lower than 0.05 (Table 4). The p value for neighbor threshold, that indicates which pharmacophoric points should be considered during docking, is just a bit above 0.05, the threshold for being statistically meaningful, so since it is on the borderline it should probably be taken into account. For this parameter and also for the other parameters having low influence according to the regression model, the experimental design was repeated investigating two extra-levels outside the values range of the first design, to ensure the low impact was not due to the previously selected ranges. The new parameters' ranges were chosen by extending the previous extreme values by 25% of the difference between them. Results obtained from the two extended levels were compared to those obtained with the original high and low ones respectively, using the non parametric Kolmogorov−Smirnov test with a 0.05 level of significance (Table 6).[49,50] The test confirms that among the excluded values only neighbor threshold influences the quality of the results, as we already supposed, and should be considered for future analysis.

Table 6. Kolmogorov-Smirnov Test[a]

	settings		
parameter	design	expanded	*d* value
ligand neighbor threshold	0.5	–	–
	3	3.875	
grid distance threshold	1	0.5	–
	3	3.5	0.167
distance cut off	4	3.5	0.187
	6	6.5	0.100
score distance threshold	1	0.5	0.029
	3	3.5	0.000
conformer angle delta	3	1	0.111
	10	12	0.167
neighbor threshold	50	25	0.292
	150	175	0.281
hydrophobic threshold	0.1	0.05	0.000
	0.3	0.35	0.222
angle delta	10	–	–
	50	60	0.114
conformer van der Waals Bumps	0.5	0.125	0.081
	1	–	–
pose overlap	0.5	0.125	0.220
	1	–	–

[a] Results of the Kolmogorov−Smirnov test from additional dockings using parameter values outside the design range (expanded setting) compared with the high and low parameter values in the design (design setting). Difference in docking results using the expanded settings compared to the designed ones was defined as Kolmogorov−Smirnov *d* value ≥ 0.240, which corresponds to a 0.05 level of significance for a sample size of 64.[49,50]

The six parameters thus identified were then used to perform a RSM to investigate additional value levels and to study the interaction effect between parameters. The other parameters, which the previous analysis showed to be less influential were assigned the central value of the screening design. A faced CCD was identified as the design that best suits our needs. Parameters and results of this experimental design are given in Table S4 of the Supporting Information. With respect to the FFD, there was a significant enhancement in the percentage of poses with RMSD less than 2 Å for all the experiments. Parameters having the greatest influence on results were (i) the maximal number of identified pharmacophoric features, (ii) grid accuracy, i.e., how fine is the grid spacing used in the analysis of the binding site, and (iii) the interaction between the previous two with neighbor threshold (the number of grid points needed to consider a pharmacophore feature suitable for docking). Best experiments produced poses with RMSD < 2 Å in 43 cases (experiments 12, 34, and 37; Table S4, Supporting Information) and with RMSD < 3 Å for 70 complexes out of 100 (experiments 10 and 28; Table S4, Supporting Information). An exhaustive analysis of results revealed that bad results occur more frequently in cases of nondrug-like ligands. As shown in the scatter plot reported in Figure 4, in most cases, the poses with higher values of RMSD involve ligands with molecular weight higher than 500 and more than 10 rotatable bonds. Among the exceptions of badly predicted drug-like compounds, localized in the upper left part of the scatter plot, are ligands that feature bad pharmacophore−ligand matching due to unsampled binding conformations, as in the cases of 1ejn and 4cox or due to a binding site partially exposed to the solvent, as in the cases of 1tng and 1 mmq (Figure S1, Supporting Information).

These results indicate that this first round of parameter optimization allowed us to significantly improve the performance of LiGenDock with respect to the initial set of parameters. Yet, the obtained results may seem not exceptional in terms of absolute metrics, given the number of poses predicted within 2 Å from the co-crystallized ligand. However, it should be noticed that LiGenDock has been originally derived within a de novo design suite of programs, where the main objective is the identification of novel chemotypes able to interact with the partner macromolecule. The pharmacophore driven docking procedure used by LiGenDock, based on a nonsystematic conformational sampling, results in very high speed performing docking experiments in an average time of only 27 s per protein. Apparently, the cost for speed is paid by reduced accuracy, although deep visual inspection of the results strongly suggests that poses within 3A from the co-crystallized ligands are still quite accurate. The high RMSD value is due to a different position of some ligand functional groups respect to those of the crystallized ligand; however, this slightly different orientation is justified by matching a pharmacophoric feature not matched by the ligand in the crystallographic complex.

To further test the performance of LiGenDock with an optimized set of parameters, we decided to validate on a different data set the set of parameters from experiment number 28 of the RSM (Table S5, Supporting Information). The choice of this set of parameters was due to the following reasons: (i) it is one of the few sets of parameters for which we were able to obtain poses for all the complexes, thus indicating it is suitable for complexes having different characteristics, for example, it is good both for very small (1pbd, ligand MW 137.38, RMSD 0.36) and big ligand size (1eta, ligand MW 776.87, RMSD 2.33) (Figure S2, Supporting Information) and (ii) with this set, we were able to obtain the highest number of poses with RMSD less than 3 Å (70%). The number of poses with RMSD less than 2 Å was a little smaller than with other best performing set of parameters, but visual inspection of results suggested that the small differences among best performing experiments in terms of RMSD < 2 Å and < 3 Å should not be emphasized too much, and the ability of producing an higher number of poses was the consideration that guided our choice. The validation was carried out by using the CORE PDBbind database as an external data set (not used during the optimization study), excluding those complexes having a ligand molecule with molecular weight higher than 500 Da. The same test set was also used to perform docking with two other docking programs, namely,

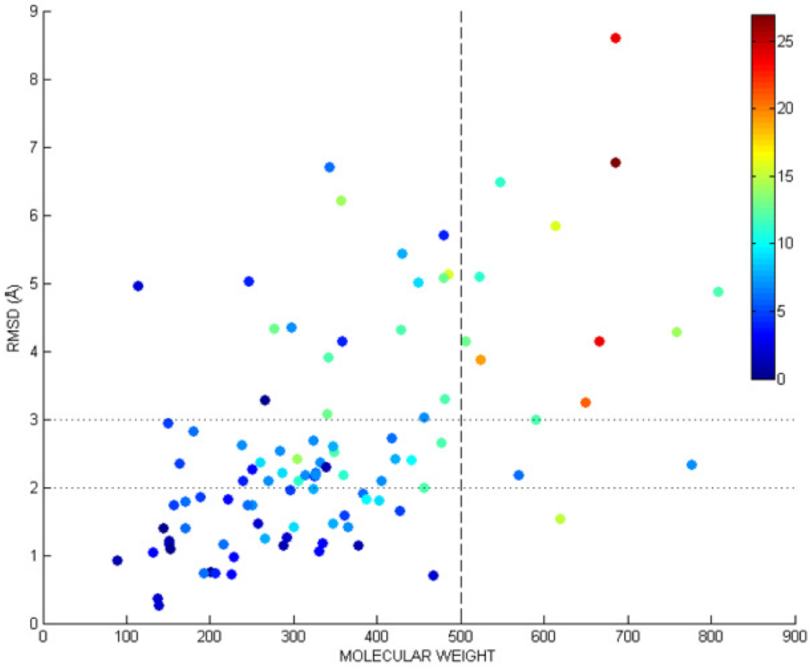


Figure 4. Scatter plot of the RMSD of the best predicted poses of the best experiment of the RSM (experiment 28). Abscissa shows the molecular weight of the ligands. The color encodes the number of rotable bonds. The best results were observed for drug-like molecules.

Table 7. Comparison of LiGen, AutoDock, and Glide Results Using the PDBbind Data Set

	best predicted pose			best scoring pose		
	LiGen	AutoDock	Glide	LiGen	AutoDock	Glide
n results	170	170	170	170	170	170
n RMSD < 2 Å	95	109	128	56	69	98
n RMSD < 3 Å	145	139	146	96	96	117
% rmds < 2	54.7	64.1	75.3	32.9	40.6	57.6
% rmsd < 3	84.7	81.8	85.9	56.5	56.5	68.8

	LiGen	AutoDock	Glide
% (number) best predicted pose = best scoring pose	26% (44)	12% (21)	30% (51)
pocket time (s)	5.46	9.83	161.42
dock time (s)	20.30	397.08	18.24
total time (s)	25.76	407.63	179.66

Glide and AutoDock. Results are detailed in Table S6 of the Supporting Information and summarized in Table 7. Poses predicted within a RMSD range of 3 Å from the crystallographic pose are 85%, more than AutoDock and only one point less than Glide. As expected, the number of poses with a RMSD within 2 Å is smaller for LiGen with respect to the others (55.3% LiGen, 64.1% AutoDock, and 75.3% Glide, respectively). The same trend of results can be observed if we consider the best scoring pose. However, it is worth mentioning that the best pose corresponds to the best scoring pose in 26% of the experiments in the case of LiGenDock and in 30% of the experiments performed with Glide but only in 12% in the case of AutoDock. Analysis of the poses with RMSD higher than 3 Å from the experimentally solved ligand revealed that in many cases they involve ligands with more than 10 rotatable bonds, as in the cases of 1b11 and 1gni (Figure S3A, Supporting Information). In other cases, LiGen failed in defining a good pharmacophore, especially for those proteins presenting a solvent exposed binding site, as for example, in the case of 1nhu, 1tyr, 1v2o, and 2g8r (Figure S3B, Supporting Information). Other badly predicted poses are found for zinc-dependent metalloproteins, as 1ndy, 1zs0, 1zvx, or 8cpa, indicating that some improvement must be introduced for scoring ligand−metal interactions (Figure S3C, Supporting Information).

From the comparison reported in Table 7, it is clear that Glide performs better than LiGen and AutoDock; the LiGen performance is very similar to AutoDock in terms of poses predicted within an RMSD of 3 Å. It should be noted, however, another interesting aspect coming from Table 7, i.e., the difference on the average speed of the three programs. This is not particularly relevant for pose prediction but can be for virtual screening and makes LiGenDock attractive for this kind of application. In particular, LiGen requires less than 30 s on average to produce 10 poses, whereas Glide needs roughly 3.5 min on average and AutoDock about 7 min on average.

Small differences in times and RMSD values with respect to those reported elsewhere in the literature[46,51] are due to different starting conformations in our test conditions with respect to the ones used by others.

Given the good performance in predicting poses with RMSD less than 3 Å in very short time, this parameter set used during self-docking validation was also chosen as default setting for routine pose prediction experiments.

Table 8. Virtual Screening Results of Preliminary Test with Original Set of Parameters (left) and with Parameter Set Optimized for Cognate Docking (right)

	Results with original parameters					Results with parameters optimized for cognate docking				
PDB code	ROC	% ROC (1%)	% ROC (2%)	% ROC (5%)	% ROC (20%)	ROC	% ROC (1%)	% ROC (2%)	% ROC (5%)	% ROC (20%)
1a7a	0.30	3.00	3.00	3.00	3.00	0.95	7.00	11.00	18.00	33.00
1agw	0.55	0.00	1.70	4.20	11.90	0.63	0.00	2.00	4.00	33.00
1b8o	0.82	8.00	20.00	28.00	64.00	0.77	0.00	0.00	1.00	12.00
1f0r	0.67	2.80	3.50	10.60	45.10	0.62	6.00	7.00	11.00	36.00
1hw8	0.55	0.00	2.90	2.90	17.10	0.76	1.00	1.00	3.00	13.00
1kim	0.31	13.60	27.30	27.30	27.30	0.54	0.00	0.00	0.00	3.00
1uy6	0.46	0.00	0.00	0.00	16.70	0.39	0.00	0.00	0.00	0.00
1xgj	0.31	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00
3ert	0.58	0.00	0.00	2.60	23.10	0.68	2.00	2.00	5.00	20.00
1m2z	0.30	1.30	1.30	1.30	1.30	0.20	0.00	0.00	0.00	0.00
mean	0.49	2.87	5.97	7.99	20.95		0.57	1.60	2.30	4.20
median	0.51	0.65	2.30	2.95	16.90		0.63	0.00	0.50	2.00
sd	0.17	4.30	9.07	10.23	19.35		0.24	2.54	3.55	5.65

Table 9. Comparison of VS Results[a]

	original						after optimization					
	ROC	ROC (1%)	ROC (2%)	ROC (5%)	ROC (20%)	BEDROC $\alpha = 20.0$	ROC	ROC (1%)	ROC (2%)	ROC (5%)	ROC (20%)	BEDROC ($\alpha = 20.0$)
entire DUD												
mean	0.54	1.30	2.88	4.27	18.29	0.05	0.73	8.07	12.10	20.47	41.82	0.18
median	0.58	0.00	1.10	2.60	16.30	0.03	0.73	2.00	3.80	10.20	31.80	0.09
sd	0.26	2.90	5.33	5.87	16.59	0.05	0.17	14.91	19.17	23.90	28.01	0.20
self-decoys												
mean	0.56	0.92	1.85	4.8	22.38	0.06	0.71	5.94	9.58	16.75	39.07	0.16
mediana	0.61	0.00	0.75	3.00	22.80	0.05	0.69	1.70	3.05	8.05	33.95	0.12
sd	0.24	1.45	2.36	5.15	16.84	0.07	0.16	11.27	16.73	21.6	25.18	0.15

[a] In the first part of the table, the entire DUD database, and in the second part of the table, the "own decoys" subset of DUD. Results are reported for the original set of parameters (left) and for the optimized ones (right).

Virtual Screening Optimization. Virtual screening experiments are conceptually different from pose prediction experiments, even though they are both based on the same molecular docking approach. With a virtual screening experiment, we want to discriminate ligands that can bind to a receptor from those that are expected not to bind (the decoys), while in the case of pose prediction, we expect to reproduce the known binding mode of a ligand to a receptor.[32] When the docking process is guided by pharmacophore models, as in the case of LiGenDock, an important issue is the strictness of the pharmacophore model. Indeed, on the one hand, very strict settings would lead to poor structural diversity in the compounds retrieved from VS, whereas on the other, a very fuzzy pharmacophore is more likely to return a large number of false positives.[52] As previously done with the optimization of parameters involved in cognate docking, we sought a set of parameters with optimized values for VS. Because a preliminary trial with parameters optimized for pose prediction gave modest results, performing slightly better than original parameters in terms of global enrichment but a little worse in case of early enrichment (Table 8), we performed a full factorial design (reported in Table S7, Supporting Information) with parameters that were previously shown to be important (the importance of the four parameters varied in the full factorial design has been confirmed by a previous FFD, data not shown): (1) grid accuracy (related with grid spacing), (2) the maximal number of pharmacophoric features because these two parameters came out as the most important ones from the first part of this study, (3) the minimal distance between two features, and (4) angle delta, the angle used in rotating ligand inside the binding site to match as many pharmacophoric features as possible. The maximum number of features was allowed to vary between 8 and 15, a range lower than the one used in the optimization of cognate docking; these values were chosen to avoid recognizing too many decoys as good binders. This choice was driven by the allowed partial ligand–pharmacophore match during virtual screening. In this sense, a too large number of pharmacophoric features will have permitted to generate features also for less important characteristics of the binding site or, for example, for the boundary regions of the pocket. Thus, in principle, some decoys matching only unimportant, or low-important, features could have been retrieved. We decided to include also the minimum distance between the features to ensure the pharmacophore resembles in an accurate way the binding site. Angle delta was included in the design because we wanted to exclude the possibility that lack of recognition of some active compounds (as seen with the original set of parameters) was due to a bad (not fine enough) ligand–pharmacophore match. Parameters not under investigation were set to the optimal values found previously. Because the number of experiments to run with all the complexes of the DUD database would have required too much computational time, we randomly selected 10 complexes, reported in Table S1 of the Supporting Information, to run the optimization procedure, whereas for the evaluation of the improvements resulting from the optimization procedure, the optimized performance was assessed using all the 36 selected targets. Results of LiGen's virtual screening performance, reported in Tables S8 and S9 of the Supporting Information and summarized in Table 9, with the original set of parameters were modest, especially regarding the early enrichment: the average ROC(1%) and ROC(5%) were, respectively, 1.30% and 4.27%. Also the global AUC presents a mean value just above random (0.54 with respect to 0.50 for random performance), with several structures showing very low ligand recognition and in two cases actives were completely discarded (PDB code 1sr7 and 1mvc). Parameters and results of the 16 experiments of the full factorial design are listed in Table S7 of the Supporting Information. Average ROC values have previously been used to assess virtual screening performance using the DUD database.[53,54] Thus, the average ROC(1%) was used to fit the design, and design statistics are given in Table 10. A deep analysis of the results showed that the best results both in terms of global and early enrichment were obtained with parameters of experiment number 1. The high standard deviation is due to the very low values of some virtual screening experiments. In particular, the AmpC β-lactamase (AmpC, PDB code 1xgj), the thymidine kinase (TK, PDB code 1kim), and the human heat shock protein 90 (HSP90, PDB code 1uy6) showed early enrichment (ROC(1%) and ROC(2%)) almost always equal to zero, regardless of the values assigned to parameters. Very poor enrichment for these structures was already reported in literature in a comparable experiment by Repasky and Murphy.[46] Virtual screening experiments using parameters of the first experiment of the full factorial design were run also for the other structures selected from the DUD database against all the decoys and against self-decoys (only decoys with scaffolds similar to active ligands), and the results are reported in Tables S8 and S9 of the Supporting Information, respectively. The optimized set of parameters improved results for almost half of the structures of the data set, both for early and global enrichment, as shown in Figure 5. Poor early enrichment are found in case of some kinases, like TK (PDB code 1kim), and HSP90 (PDB code 1uy6). In particular, TK is reported to be a challenging target due to receptor flexibility, solvent exposed binding site, and the importance of water bridge interactions,[28] not taken into account in our experiments. To have a more complete view of LiGenDock performance, we report in Table 11 a comparison between LiGen, Glide, and DOCK6 VS results, using for the last two programs, data published in other papers that appeared when this study was being performed.[46,53] As shown in Table 11, the global enrichment of the three programs is good, performing all better than random. Glide is the best one, having the average and the median AUC value of 0.80 and 0.82,

Table 10. VS Design Statistic

	virtual screening		
	R^2	R^2 adjust	F
design 3	0.7929	0.7176	10.5281
parameters	***p* value**		
grid accuracy	0.0397		
maximal features number	0.2796		
minimal features distance	0.0002		
angle delta	0.0371		

Table 11. Comparison between LiGen, Glide, and DOCK6 VS Results[46,53a]

	ROC	SD	median	max	min	ROC (1%)	ROC (2%)
LiGen	0.73	0.17	0.73	0.99	0.37	8.07	12.10
Glide	0.80	0.14	0.82	0.98	0.42	25.18	33.64
DOCK6	0.60	0.17	0.56	0.96	0.29	4.99	20.19

[a] With ROC is indicated the AUC of the ROC curve; SD is the standard deviation; max is the highest ROC value; min is the lowest ROC value found.

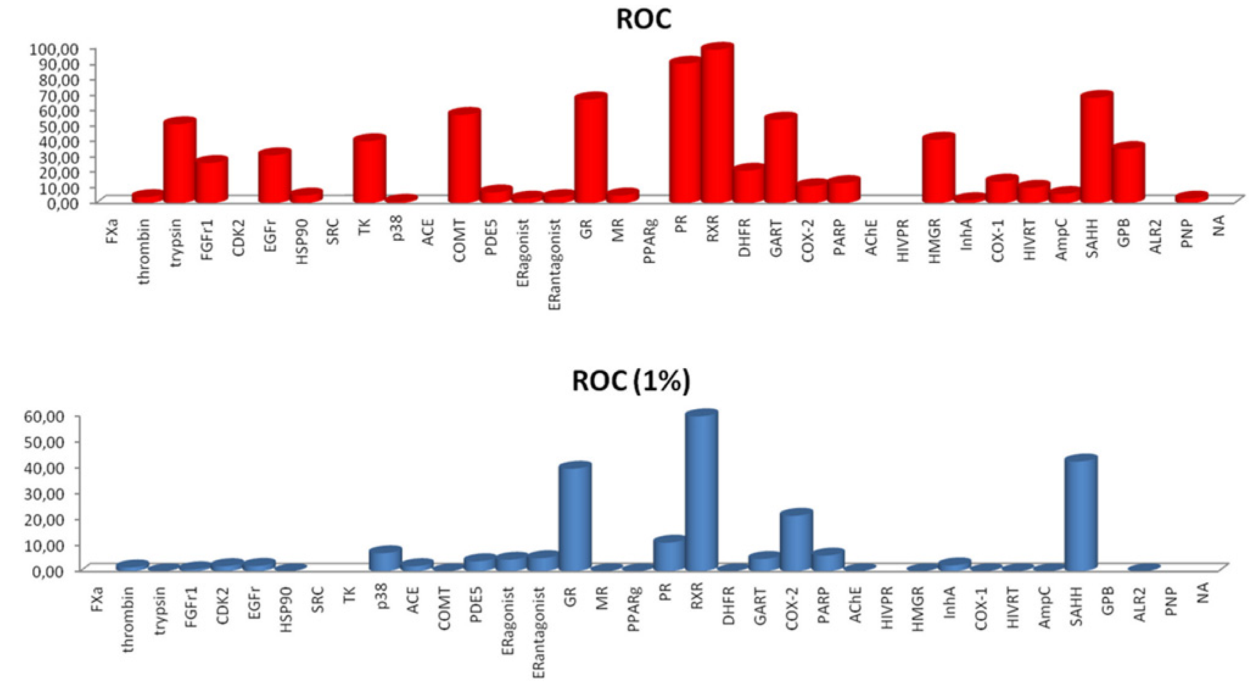


Figure 5. Histogram plot to show the improvements of the ROC and of ROC(1%) values for all the DUD complexes considered in the study. The bars represent the percentage of improvement gained through the optimization procedure. Complexes with missing bars are those for which no improvement was registered.

respectively. LiGenDock, with average and median AUC values of 0.73 in both cases, performs slightly worse than Glide but better than DOCK6, whith average and median AUC of 0.60 and 0.56, respectively. The last two columns of Table 11 report the average percentage of ROC(1%) and ROC(2%). As evident from this table, Glide is the best performing program. LiGenDock performance is better than DOCK6 in the first 1% of the screened database (ROC(1%) value) and is definitely better than random performance (random performance at 1% of the screened database is 0.5%), confirming the goodness of the LiGen approach also for VS. Notably, even though also after the optimization the average early enrichment (ROC(1%)) is not outstanding, there was a significant improvement with respect to the original set of parameters, corroborating the application of the optimization protocol. A large enhancement was also registered for global enrichment (AUC) for ROC(5%). For us, values of the early enrichment metrics were not a surprising outcome. LiGen docking process is driven by the pharmacophore generated inside the binding site, so the suggested binding pose should not be seen as a results of an extensive conformational search for the global energy minimum but as a result of a reduced/constrained conformational search to match the highest number of pharmacophoric features. In this sense, a higher number of decoys can be recognized during the docking/virtual screening process. Moreover, in a real application of virtual screening, this can be an advantage because it allows for the recognition of possibly new and diverse scaffolds. An example of poor benchmarking results but good outcomes in a real life application can be found in the work of Löwer et al.[55] The LiGen docking algorithm has been also developed for use in fragment docking in de novo design, so the recognition of new scaffolds is of primary importance, definitely more than high enrichment in recognition of already know ligands. The high speed of LiGen's code allows virtual screening with the entire DUD database in less than 105 h, that means about 3 s for each ligand. This represents very good results in term of time needed to screen large databases because, for example, for Glide, is reported a time of 10 s per ligand.[46]

■ CONCLUSIONS

The primary goal of this study was the optimization of LiGen docking performance using a procedure based on experimental designs. LiGen is a de novo design suite of programs, presented in another paper,[20] consisting of a set of modules: LiGenPass for binding site recognition, LiGenPocket for binding site analysis and structure-based pharmacophore definition, LiGenDock for docking and virtual screening, and LiGenBuilder for de novo design. In this paper, we focused on LiGenPocket and LiGenDock, which constitute the docking engine of the program. A number of parameters controlling the docking procedure were varied according to statistical experimental designs. First, the most influential parameters were identified through a fractional factorial design, yielding parameter sets that covered the selected interval of parameter values. The parameter sets thus designed were then applied in a docking study using a set of 100 protein—ligand complexes taken from the CCDC Astex data set. The number of poses presenting an RMSD less than 2 Å between the best predicted docking poses, and the corresponding crystallographic ligands were considered as response for fitting the design. A significant regression model between the docking runs using the designed parameter sets and the docking results (number of poses with RMSD less than 2 Å) was established, thus shedding light on the parameters with large influence on docking results. The most relevant parameters were the minimum distance between two pharmacophoric features (minimum feature distance), the maximum number of features identified in the binding site (maximum feature number), the degree of smoothing of the protein van der Waals volume (van der Waals bumps), the grid spacing (grid accuracy), the tolerance in considering a ligand functional group superposed on the pharmacophoric feature (distance threshold), and the threshold indicating which pharmacophoric points should be considered during docking (neighbor threshold). Furthermore, a response surface model was developed using these parameters to find the optimal parameters' set. As shown in Table 3, with the optimized set of parameters, we obtained a number of poses with RMSD less than 2 Å almost seven times higher compared to the original set (41% of the optimized set with respect to the 6% of the original parameters' set). This gain in the accuracy of pose prediction was not followed by an increment in time consumed by the docking process because the difference of the average time spent is less than 1 s. It should be noticed that the LiGenDock algorithm has been originally derived within a de novo design suite of programs, where the main objective is the identification of novel chemotypes able to interact with the partner macromolecule. Thus, the LiGen's pharmacophore-based approach partially suffers in terms of precision in exactly reproducing experimental binding poses, although a deep visual inspection of the results suggests that poses within 3 Å from the co-crystallized ligands are still quite accurate. Moreover, the comparison between LiGen docking results with AutoDock and Glide using a data set extracted from the PDBbind database confirms the quality of LiGen approach, even though Glide was the best performing program. As reported in Table 7, the number of LiGen's predicted poses within 3 Å from the co-crystallized ligand is similar to those predicted by Glide and a little better than by AutoDock (poses within 3 Å from the crystallized ligands: LiGen 84.7%, AutoDock 81%, Glide 85.9%).

Investigation of the influence of parameters on the VS results was also performed using experimental design to find an optimal parameters set for virtual screening experiments. Global enrichment, represented by the mean ROC values of 0.73 after the optimization procedure, is consistent with values obtained with other software and reported in literature[46,53,54] and also summarized in Table 11. The not particularly excellent performance in terms of early enrichment should not be considered as a negative result; actually, a higher number of decoys can be recognized during the virtual screening process because LiGen's approach has been originally developed for de novo design purposes, i.e., to recognize possibly new and diverse scaffolds. In this sense, there are already examples in the literature of pharmacophore-based virtual screening studies with modest benchmarking results but good outcomes in real life applications, as shown by the paper by Löwer and coworkers.[55] Furthermore, the high speed reached, screening the entire DUD database in about 105 h, makes LiGen very attractive for virtual screening applications. It should be commented that simultaneous optimization of both virtual screening and pose prediction performance would carry out a two-properties optimization, possibly through the definition of a desiderability function. This can be done, but we anticipated that the results cannot be better than those described in the paper. The resulting set of parameters might possibly be seen as "general purposes" parameters, but our data, reported in the present paper, clearly indicate that the optimization of the two properties diverges, so that optimization of an "averaged" desireability function must necessarily afford "averaged" results.

Finally, but most importantly, the results presented in the previous sections highlight the usefulness of experimental designs for optimization purposes also in the field of computational drug discovery, even though this approach is only seldom applied in this field. Using experimental designs, we were able to define two optimal sets of parameters: one for cognate docking experiment and the other for virtual screening.

ASSOCIATED CONTENT

S Supporting Information

List of proteins used for virtual screening optimization (Table S1). List of PDB codes, ligand molecular weight and number of rotatable bonds (Table S2). Experimental design tables used in pose prediction optimization (Table S3–S4). List of the parameters chosen as default set (Table S5). LiGen, Glide, and AutoDock results using the PDBbind database (Table S6). Experimental design table used for VS optimization (Table S7). VS results for all the DUD complexes using the entire DUD (Table S8) and only the "own-decoys" subset (Table 9). Additional docking examples (Figure S1–S3). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gabriele.costantino@unipr.it.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS:

ASP, active site point; DUD, directory of useful decoys; FD, full factorial design; DoE, design of experiment; FFD, fractional factorial design; CCD, central composite design; ROC, receiver operating characteristic; SST, total sum of squares; SSE, sum of error squares; SSR, regression sum of squares; MSR, mean squared regression; MSE, residual mean square; VS, virtual screening; ACE, angiotensin-converting enzyme; AChE, acetylcholinesterase; ADA, adenosine deaminase; ALR2, aldose reductase; AmpC, AmpC β-lactamase; AR, androgen receptor; CDK2, cyclin-dependent kinase 2; COMT, catechol O-methyltransferase; COX-1, cyclooxygenase-1; COX-2, cyclooxygenase-2; DHFR, dihydrofolate reductase; EGFr, epidermal growth factor receptor; ER, estrogen receptor; FGFr1, fibroblast growth factor receptor kinase; FXa, factor Xa; GART, glycinamide ribonucleotide transformylase; GPB, glycogen phosphorylase β; GR, glucocorticoid receptor; HIVPR, HIV protease; HIVRT, HIV reverse transcriptase; HMGR, hydroxymethylglutaryl-CoA reductase; HSP90, human heat shock protein 90; InhA, enoyl ACP reductase; MR, mineralocorticoid receptor; NA, neuraminidase; P38 MAP, P38 mitogen activated protein; PARP, poly(ADP-ribose) polymerase; PDE5, phosphodiesterase 5; PDGFrb, platelet derived growth factor receptor kinase; PNP, purine nucleoside phosphorylase; PPARg, peroxisome proliferator activated receptor γ; PR, progesterone receptor; RXRa, retinoic X receptor α; SAHH, S-adenosyl-homocysteine hydrolase; SRC, tyrosine kinase SRC; TK, thymidine kinase; VEGFr2, vascular endothelial growth factor receptor; ATP, adenosine-5′-triphosphate; β-GAR, β-glycinamide ribonucleotide; NAD(P)(H), nicotinamide adenine dinucleotide (phosphate)(reduced); PLP, pyridoxal-5′-phosphate

REFERENCES

(1) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.

(2) Jain, A. N. Bias, reporting, and sharing: Computational evaluations of docking methods. *J. Comput.-Aided Mol. Des.* **2008**, 22.

(3) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.

(4) Dias, R.; de Azevedo, W. F. Molecular docking algorithms. *Curr. Drug Targets* **2008**, *9*, 1040–1047.

(5) Halperin, I.; Ma, B. Y.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 409–443.

(6) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76–90.

(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.

(8) Shoichet, B. K.; Kuntz, I. D. Matching chemistry and shape in molecular docking. *Protein Eng.* **1993**, *6*, 732–733.

(9) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein–ligand docking: Current status and future challenges. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 15–26.

(10) McMartin, C.; Bohacek, R. S. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333–344.

(11) Stouten, P. F. W.; Kroemer, R. T. Core Concepts and Methods: Target Structure-Based Docking and Scoring. In *Comprehensive Medicinal Chemistry II*; Taylor, J. B., Triggle, D. J., Eds.; Elsevier: Oxford, UK, 2006; Vol. *4*, pp 255–281.

(12) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L. Comparative study of several algorithms for flexible ligand docking. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 755–763.

(13) 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.

(14) Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499–511.

(15) Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S.; Johnson, A. P. eHITS: An innovative approach to the docking and scoring function problems. *Curr. Protein Pept. Sci.* **2006**, *7*, 421–435.

(16) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM: A new method for protein modeling and design. Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488–506.

(17) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(18) Antes, I.; Merkwirth, C.; Lengauer, T. POEM: Parameter optimization using ensemble methods: Application to target specific scoring functions. *J. Chem. Inf. Model.* **2005**, *45*, 1291–1302.

(19) Andersson, C. D.; Thysell, E.; Lindstrom, A.; Bylesjo, M.; Raubacher, F.; Linusson, A. A multivariate approach to investigate docking parameters' effects on docking performance. *J. Chem. Inf. Model.* **2007**, *47*, 1673–1687.

(20) Beccari, A.; Cavazzoni, C.; Beato, C.; Costantino, G. LiGen: High performance workflow for chemistry driven de novo design. *J. Chem. Inf. Model* **2013**.

(21) Wang, R.; Gao, Y.; Lai, L. H. LigBuilder: A multi-purpose program for structure-based drug design. *J. Mol. Model.* **2000**, *6*, 498–516.

(22) Wang, R.; Lai, L. H.; Wang, S. M. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11–26.

(23) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of

protein−ligand interaction. *Proteins: Struct., Funct., Bioinf.* **2002**, *49*, 457−471.

(24) Protein Data Bank (PDB). http://www.rcsb.org/pdb/home/home.do (accessed April 20, 2013).

(25) Wang, R.; Fang, X. L., Y.; Yang, C.-Y.; Wang, S. The PDBbind database: Methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111−4119.

(26) Wang, R.; Fang, X. L.; Wang, Y.; PDBbind, S. Database: Collection of binding affinities for protein−ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.

(27) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−26.

(28) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.

(29) *Protein Preparation Wizard; Epik version 2.0; Impact version 5.5; Prime version 2.1*; Schrödinger, LLC: New York, 2009.

(30) *LigPrep*, version 2.3; Schrödinger, LLC: New York, 2009.

(31) *ConfGen*, version 2.1: Schrödinger, LLC: New York, 2009.

(32) Onodera, K.; Satou, K.; Hirota, H. Evaluations of molecular docking programs for virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1609−1618.

(33) Box, G. E. P.; Hunter, J. S.; Hunter, W. G. Basics: Probability, Parameters and Statistics. In *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2005; pp 17−65.

(34) *MATLAB and Statistics Toolbox*, Release 2010; MathWorks: Natick, MA, 2010.

(35) Process Improvement. In *NIST/SEMATECH e-Handbook of Statistical Methods*; National Institute of Standards and Technology (NIST): Washington, DC, 2012.

(36) Lundstedt, T.; Seifert, E.; Abramo, L.; Thelin, B.; Nystrom, A.; Pettersen, J.; Bergman, R. Experimental design and optimization. *Chemom. Intell. Lab. Syst.* **1998**, *42*, 3−40.

(37) Box, G. E. P.; Hunter, J. S.; Hunter, W. G. Factorial Designs at Two Levels: Advantages of Experimental Design. Fraction Factorial Designs: Economy in Experimentation. In *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2005; pp 173−279.

(38) Bezerra, M. A.; Santelli, R. E.; Oliveira, E. P.; Villar, L. S.; Escaleira, L. A. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta* **2008**, *76*, 965−977.

(39) Box, G. E. P.; Hunter, J. S.; Hunter, W. G. Modelling Relationships, Sequential Assembly: Basics for Response Surface Methods. In *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2005; pp 437−487.

(40) Box, G. E. P.; Wilson, K. B. On the experimental attainment of optimum. *J. R. Stat. Soc.* **1951**, *13*, 1−45.

(41) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(42) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: Comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.

(43) Li, X.; Li, Y.; Cheng, T.; Liu, Z.; Wang, R. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *J. Comput. Chem.* **2010**, *31*, 2109−2125.

(44) Plewczynski, D.; Lazniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **2011**, *32*, 742−755.

(45) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput. Aided Mol. Des.* **2008**, *22*, 133−139.

(46) Repasky, M. P.; Murphy, R. B.; Banks, J. L.; Greenwood, J. R.; Tubert-Brohman, I.; Bhat, S.; Friesner, R. A. Docking performance of the Glide program as evaluated on the Astex and DUD datasets: A complete set of Glide SP results and selected results for a new scoring function integrating WaterMap and Glide. *J. Comput. Aided Mol. Des.* **2012**, *26*, 787−799.

(47) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, 48.

(48) Truchon, J.-F.; Bayly, C. I. Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **2007**, *47*, 488−508.

(49) Kirkman, T. W. Statistics To Use. http://www.physics.csbsju.edu/stats/ (accessed January 2012).

(50) Massey, F. J. The Kolmogorov−Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68−78.

(51) Trott, O.; Olson, A. J. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.

(52) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing pitfalls in virtual screening: A critical review. *J. Chem. Inf. Model.* **2012**, *52*, 867−881.

(53) Brozell, S. R.; Mukherjee, S.; Balius, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 749−773.

(54) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455−1474.

(55) Loewer, M.; Geppert, T.; Schneider, P.; Hoy, B.; Wessler, S.; Schneider, G. Inhibitors of *Helicobacter pylori* protease HtrA found by 'virtual ligand' screening combat bacterial invasion of epithelia. *PLoS One* **2011**, *6*, e17986.

(56) DeLano, W. L. *PyMol Molecular Graphic System*, version1.2r1; DeLano Scientific: South San Francisco, CA, 2009.