Fragment-Based de Novo Ligand Design by Multiobjective Evolutionary Optimization

Fabian Dey and Amedeo Caflisch*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

GANDI (*G*enetic *A*lgorithm-based de *N*ovo *D*esign of *I*nhibitors) is a computational tool for automatic fragment-based design of molecules within a protein binding site of known structure. A genetic algorithm and a tabu search act in concert to join predocked fragments with a user-supplied list of fragments. A novel feature of GANDI is the simultaneous optimization of force field energy *and* a term enforcing 2D-similarity to known inhibitor(s) or 3D-overlap to known binding mode(s). Scaffold hopping can be promoted by tuning the relative weights of these terms. The performance of GANDI is tested on cyclin-dependent kinase 2 (CDK2) using a library of about 14 000 fragments and the binding mode of a known oxindole inhibitor to bias the design. Top ranking GANDI molecules are involved in one to three hydrogen bonds with the backbone polar groups in the hinge region of CDK2, an interaction pattern observed in potent kinase inhibitors. Notably, a GANDI molecule with very favorable predicted binding affinity shares a 2-N-phenyl-1,3-thiazole-2,4-diamine moiety with a known nanomolar inhibitor of CDK2. Importantly, molecules with a favorable GANDI score are synthetic accessible. In fact, eight of the 1809 molecules designed by GANDI for CDK2 are found in the ZINC database of commercially available compounds which also contains about 600 compounds with identical scaffolds as those in the top ranking GANDI molecules.

INTRODUCTION

The search for drug molecules with computational methods is often performed by high-throughput docking or to a lesser extent by de novo drug design approaches.[1] While virtual screening relies on pre-existing compounds, de novo design approaches generate novel molecules out of building blocks consisting of single atoms or fragments. Because of the difficulty to predict synthetic accessibility, de novo drug design tools often generate molecules that are demanding to synthesize.[2] Approaches that are commonly employed to improve synthetic accessibility include the use of connection rules to join building blocks [3−5] or the build-up of molecules from fragments obtained by prior decomposition of existing compounds.[6,7] Connection rules are either derived from organic synthesis reactions[3,4,8] or they are based on the knowledge of the occurrence of certain bonds in existing molecules.[9,10] Several statistics of building blocks occurring in known molecules have been collected.[11−13] The RECAP procedure by Lewell et al.[14] is particularly interesting for de novo drug design, because building blocks are generated by decomposition of virtual libraries based on cleavage rules inferred from organic synthesis reactions.

Global optimization algorithms are usually employed to search the chemical space. Genetic algorithms[15,16] (in ADAPT,[17] LEA3D,[7] and LigBuilder[10]) and Monte Carlo-based procedures (in SKELGEN,[18] SMoG,[19] and SPROUT[6]) are able to sample rugged scoring function surfaces efficiently. Genetic algorithms are based on the theory of natural selection,[20] evolving multiple solutions concomitantly. A new offspring is created with mutation and crossover operators and is subject to selection subsequently. Monte Carlo based procedures use the Metropolis criterion[21] to accept or reject transitions from one state to another.

Methods for estimating the free energy of binding of a small molecule to a target protein, referred to as the scoring problem in both docking and de novo design, can be roughly divided into three families:[22,23] force field-based, knowledge-based, and empirical approaches. As for high-throughput docking, all of the three aforementioned methods have been implemented in de novo drug design approaches (e.g., force field-based in GroupBuild[9] and CCLD,[24] knowledge-based in SMoG,[19] and empirical in LigBuilder[10]). Additional factors besides estimates of the binding strength are often considered to ensure that drug- or leadlike molecules are generated. The use of multiple scoring functions or the additional application of constraints on the design process is referred to as multiobjective optimization. Among the various ways of combining individual scores, the weighted-sum approach[25] is often used.[8,17,18] Individual scores are multiplied by user-defined coefficients and summed up to yield the total score. Applications of de novo drug design programs resulting in the identification of novel inhibitors against different proteins have been reported.[26−29]

Here, a novel approach for *G*enetic *A*lgorithm-based de *N*ovo *D*esign of *I*nhibitors (GANDI) is presented. GANDI is a fragment-based method that generates molecules by joining predocked fragments with linkers. A parallel genetic algorithm[30] employing the simultaneous evolution of multiple populations is used in GANDI to search for feasible solutions. To increase the selection pressure a different form of elitism is implemented, where both the parent and the children populations compete for their survival. Only the predocked fragments are encoded by the genetic algorithm, while suitable linker fragments are efficiently evaluated with a tabu search[31−33] using look-up tables. The cost function in

* Corresponding author phone: (+41 44) 635 55 21; fax: (+41 44) 635 68 62; e-mail: caflisch@bioc.uzh.ch.
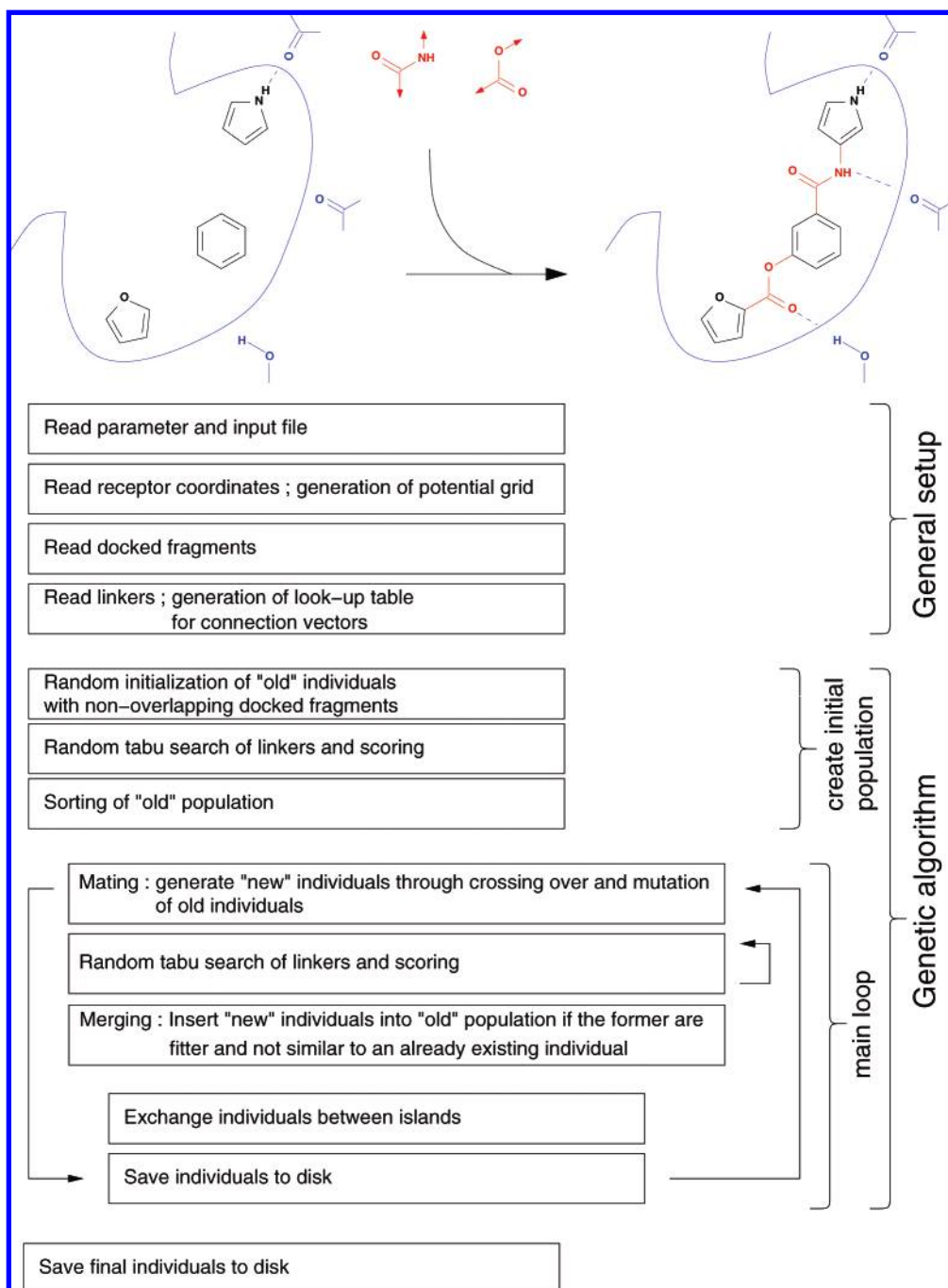
Figure 1. (Top) Schematic illustration of GANDI. Molecules are generated by linking predocked fragments (black) using a list of user-supplied linker fragments (red). The protein surface and its polar groups are in blue. (Bottom) Flow chart of GANDI. The flow is top to bottom and includes two iterative procedures, which are the main loop of the genetic algorithm and the random tabu search embedded within it (arrows on the right).

GANDI is a novel combination of force field energy *and* a measure of the 2D-similarity to known inhibitor(s) or 3D-overlap to known binding mode(s). The energy term consists of intra- and intermolecular contributions. The 2D-similarity term is the Tanimoto coefficient[34] with 2D-fingerprints as variables. The measure of 3D-overlap enforces a spatial distribution of the atoms of the designed molecule similar to the one in the known binding mode of the inhibitor(s) without explicitly considering the covalent structure. Scaffold hopping (that is to say design of isofunctional molecular structures with significantly different backbones) is facilitated by overweighting the 3D-overlap measure and at the same time maintaining structural diversity in the population of

molecules during genetic algorithm optimization. GANDI is evaluated on cyclin-dependent kinase 2 (CDK2), presenting a complex optimization problem due to the large number of fragments used. Different optimization algorithm setups are analyzed for their search efficiency. Notably, GANDI is able to suggest molecules with new scaffolds or substituents that, at the same time, preserve the main binding interaction motifs of known inhibitors of CDK2.

METHODS

Overall Procedure. To connect predocked fragments with linker fragments (Figure 1, top), GANDI uses a combination of two stochastic search procedures, a genetic algorithm,[15,16]

GANDI: Fragment-Based de Novo Ligand Design Program

J. Chem. Inf. Model., Vol. 48, No. 3, 2008 **681**

and a tabu search.[31−33] Heavy atom−hydrogen atom vectors are the connection points, which can be selected by the user. Covalent bonds generated by GANDI for linking fragments are single bonds. The scoring function is a linear combination of three terms with multiplicative parameters as input. A flow chart of GANDI is given in Figure 1, bottom.

Genetic Algorithm for Global Optimization. The genetic algorithm in GANDI is an island (or parallel) model, using the simultaneous, noninteracting evolution of multiple populations at the same time.[30] The exchange of genetic material is performed after a certain amount of optimization iterations by swapping individuals (i.e., molecules) between neighboring, all, or randomly picked islands (Supporting Information Figure S1). Every individual contains a single chromosome consisting of multiple genes. Contrary to classic genetic algorithms, the implementation in GANDI uses integers as gene values encoding indexes of docked fragments. Hence, the value of each gene ranges from one to the number of docked poses. Individuals are selected from the parent population with binary tournament selection[35] to undergo modifications. One-point crossing over is accomplished by swapping genes at a single randomly chosen chromosome position. Genes are selected and mutated by a random integer derived from a normal distribution with adjustable width. The encoded fragments of the "children" are examined for clashes, and individuals with unfavorable interactions are immediately removed from the population.

Tabu Search for Efficient Linking. The next step involves linking the encoded fragments for each individual by a tabu search.[31−33] For efficiency reasons GANDI builds a look-up table containing all distances and angles of all pairs of linker fragment connection vectors. Using cutoff values and the look-up table, all possible connections of fragment pairs of an individual are generated. A connection solution is randomly picked, and the two fragments are joined with the linker defined therein (Supporting Information Figure S2). By employing a breadth-first search strategy[36] fragments of an individual are divided into two sets: a connected set, which are the newly merged fragments and all fragments connected previously, and a not-connected set. A new connection solution is picked with at least one docked fragment being part of the not-connected set, while omitting already occupied connection vectors. The procedure continues until all fragments are connected or a maximal number of connection trials has been exceeded. The score is calculated for the built-up molecule (see below), and the merging procedure is repeated for a user-defined amount of iterations, storing only the docked fragments−linkers assembly with the lowest score for every individual. A tabu list of previously visited solutions is used during these iterations to avoid repeated calculation of scores.

Selection of Fittest Individuals. Children are inserted into the parent population if no structurally similar parent has a more favorable score (see below for scoring function). The 3D-similarity between molecules A and B is measured by[37,38]

$$\text{Sim}_{3D}(A, B) = \frac{S_{AB}}{\max(S_{AA}, S_{BB})} \quad (1)$$

$$S_{XY} = \sum_{i \in X} \sum_{j \in Y} w_{t_i t_j} e^{-\gamma r_{ij}^2} \quad (2)$$

where r_{ij} is the distance between two atoms ($i \in$ molecule X, $j \in$ molecule Y), $w_{t_i t_j}$ is a matrix whose coefficients reflect the similarity between element types (an unit matrix is currently used), and γ is a coefficient which acts on the broadness of the distribution of the positions. The 3D-similarity Sim_{3D} does not explicitly consider the covalent structure of molecules but relies on the arrangement of atoms in space. In this way significantly different binding modes of the same molecule (or of two very similar molecules, e.g., differing by only a halogen atom) are kept in the population and evolved further if they have a good score. In the present application to CDK2 the maximal similarity Sim_{3D} of two individuals within a population was set to 0.8 and γ to 0.9. Individuals of the old population similar to new ones, but with less favorable score, are given an arbitrarily high score to reduce the likelihood of their selection in a subsequent mating step. The size of the old population is adjusted after merging the old and the new population by removing individuals with the least favorable score.

Scoring. The scoring function implemented in GANDI is a linear combination, i.e., a weighted-sum, of three terms: a force field-based binding energy (E_{ff}) and two measures of similarity (Sim_{3D} and Sim_{2D}) to a user-supplied target structure (e.g., a known inhibitor)

$$S_{\text{total}} = w_{ff} E_{ff} - w_{3D} \text{Sim}_{3D} - w_{2D} \text{Sim}_{2D} \quad (3)$$

where the multiplicative parameters w_{ff}, w_{3D}, and w_{2D} are input values, and usually either w_{3D} or w_{2D} is set to zero. The minus signs for the similarity terms are used because optimization is performed by minimization of S_{total} and $\text{Sim}_{3D,2D}$ grow with increasing similarity. Based on preliminary test calculations, the following values of the parameters were used in the application to CDK2: $w_{ff} = 0.02$, $w_{3D} = 1$, and $w_{2D} = 0$. These values of w_{ff} and w_{3D} yield a good balance of favorable binding energy and 3D-structural similarity to the binding mode of a known oxindole inhibitor of CDK2 (PDB code 1KE5). The term for 2D-similarity was not taken into account because the knowledge of the binding mode (from the X-ray structure) was preferred to a ligand-based design.

The force field-based energy function consists of van der Waals and electrostatic terms. Both intraligand (*intra*) and ligand/receptor (*inter*) interactions are taken into account

$$E_{ff} = E_{vdW}^{inter} + E_{elec}^{inter} + E_{vdW}^{intra} + E_{elec}^{intra} \quad (4)$$

The van der Waals energy is calculated with a 6−12 Lennard-Jones potential and the electrostatic energy with a distance-dependent dielectric model using CHARMm atom types, partial charges of protein atoms, and van der Waals parameters[39] (see below for partial charges of compounds). Intrafragment and intralinker interactions as well as fragment−linker interactions between atoms separated by one or two covalent bonds are not evaluated. The potential of the receptor is calculated and stored on a grid[38] and is used only for the linkers. The energies of the fragment poses are read in from the MOL2-files to save computational time.

The 3D-structural similarity Sim_{3D} between the newly assembled molecule and a user-supplied template molecule is evaluated by eq 1.
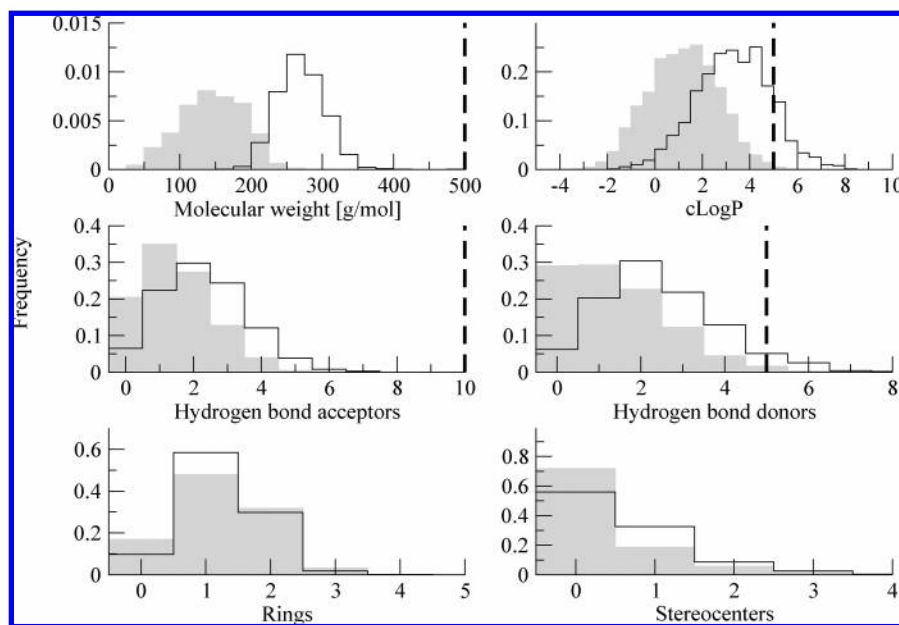
Figure 2. Normalized distributions of molecular properties calculated by DAIM.[45] Histograms for the 13 788 fragments docked by SEED[38] are in gray and shaded, and those for the 1809 unique molecules obtained by the 10 GANDI runs with 4 islands and exchange of individuals are black and empty. The vertical thick dashed lines represent the thresholds defined by Lipinski's rule of five.[52] The number of stereocenters was evaluated with the STERGEN-module in CORINA,[41] which considers tetrahedral chiral centers as well as cis/trans isomerism.

The third term in the scoring function is a fingerprint-based 2D-measure of similarity Sim_{2D}. To reduce computational cost and allow the user to decide on the fingerprint type, the fingerprints of the template, linker, and docked fragments are precalculated and read in from the input MOL2 files. The fingerprint can be any string of numbers and is user-defined, but each entry of the string has to be additive because the individual entries of the fingerprint of the assembled molecule are calculated by summing up the corresponding values of the linkers and docked fragments. The fingerprint similarity between the target A and the source molecule B is calculated with the Tanimoto coefficient[34]

$$Sim_{2D} = \frac{\sum_{i=1}^{n} x_{iA} x_{iB}}{\sum_{i=1}^{n} x_{iA}^2 + \sum_{i=1}^{n} x_{iB}^2 - \sum_{i=1}^{n} x_{iA} x_{iB}} \qquad (5)$$

where x_{iA} denotes the ith fingerprint entry of molecule A.

Protein Preparation. The crystallographic structure of CDK2 cocrystallized with an oxindole-based inhibitor[40] was downloaded from the Brookhaven database (accession code 1KE5). The bound inhibitor and all water molecules were removed, and hydrogens were added to side chains and termini of the protein according to pH 7. CHARMm atom types[39] were assigned, and hydrogens were minimized in the absence of the oxindole-based inhibitor with the CHARMm force field (Accelrys Inc.) to a gradient of the energy of 0.01 kcal mol^{-1} Å^{-1}. A distance-dependent dielectric function of $\epsilon(r) = 4r$ and default nonbonding cutoffs of 14 Å were used.

Preparation of the Oxindole-Based Inhibitor as a Target for Structural Similarity. The crystallographic ligand was prepared as described for the protein but was minimized without any constraints inside the rigid protein binding site using $\epsilon(r) = 4r$.

Preparation of Fragment Library. The library of fragments, from which the molecules were constructed, was obtained from Molinspiration Cheminformatics (www.molinspiration.com, November 2006 accession date). The library consisted of 20 000 fragments with one and 20 000 fragments with two connection points occurring in bioactive molecules. Both sets were converted from SMILES-strings to MOL2 format with CORINA[41] and Babel,[42] adding hydrogens with Babel according to a pH of 7 and calculating partial charges with the MPEOE approach.[43,44] CHARMm atom types were assigned, and all fragments were subject to minimization. The connection points defined in the source SMILES were used as connection vectors of the fragments, using all possible heavy atom−hydrogen atom vectors. Additionally, geometrically identical vectors or pairs of vectors were searched for by superpositioning all original vectors or pairs of vectors with all possible heavy-atom−hydrogen atom vectors and measuring the 3D-similarity Sim_{3D} of the two structures with eq 1. The original and superimposed fragments were deemed identical if the similarity was larger than 0.95.

Docking of Fragments. Of the 40 000 fragments in the library only the 13 788 containing less than four rotatable bonds were used. Their properties are shown in Figure 2 (gray histograms). Of these, 6906 and 6882 have one and two connection vectors, respectively. They were docked into the receptor binding site with SEED, a program for docking mainly rigid fragments with evaluation of protein-fragment energy and electrostatic desolvation.[38] Note that the 6882 fragments with two connection vectors were used also as linker fragments in GANDI (see the **GANDI** subsection). A maximal number of 20 cluster representative positions of each fragment type were kept, using also a cutoff value of −5 kcal mol^{-1} for the SEED[38] energy. To discard docked fragments with vector directions pointing toward the protein surface, hydrogen atoms of connection vectors were indi-

GANDI: Fragment-Based de Novo Ligand Design Program

J. Chem. Inf. Model., Vol. 48, No. 3, 2008 **683**
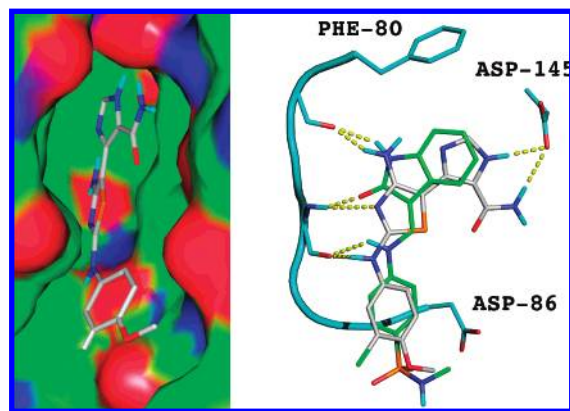
Figure 3. Binding mode of GANDI molecule with diaminothiazole scaffold (last molecule in Table 3). (Left) Surface model of the ATP-binding site of CDK2 (PDB structure 1KE5) with regions corresponding to carbon, nitrogen, and oxygen atoms in green, blue, and red, respectively. (Right) Overlap of GANDI molecule (carbon atoms in gray) and the known oxindole inhibitor (carbon atoms in green). The hinge region and three side chains of CDK2 are shown with carbon atoms in cyan. The yellow dashed lines indicate intermolecular hydrogen bonds. The figure was prepared with PyMOL (DeLano Scientific, San Carlos, CA).

vidually modified into a methyl group, followed by minimization of the methyl group inside the protein binding site and in absence of the protein. Fragments were discarded if the difference in energy for any vector between bound and unbound states was larger than 5 kcal mol^{-1}. This procedure yielded a total of 8506 docked fragments with 72 822 poses. Since some fragments contained multiple equivalent vectors or pairs of vectors (e.g., six for benzene), which were defined as separate poses in the GANDI input file, the final number of poses increased to a total of 100 027.

GANDI. The setup of GANDI included three different parameter settings with 400 individuals in a single island, 4 islands of 100 individuals each, and 4 islands with 100 individuals exchanging 5% of all individuals every 20th iteration with a randomly selected island. Calculations were repeated 10 times with distinct random seed numbers for 1000 iterations of the genetic algorithm and 20 iterations of the tabu search per individual with all three setups. The minimized oxindole-based inhibitor cocrystallized with the protein (PDB code 1KE5) was used as a target structure. The coefficients of the scoring function terms were set to $w_{ff} = 0.02$, $w_{3D} = 1$, and $w_{2D} = 0$.

The 13 788 Molinspiration-library fragments with less than four rotatable bonds were used in each run of GANDI. The 6882 Molinspiration-library fragments with two connection vectors and up to three rotatable bonds were used as linker fragments with a total of 17 372 unique vector pairs. To prevent the generation of unstable molecules, bonds between nitrogen, oxygen, and sulfur atoms were avoided by using a list of forbidden connections (S−S, S−O, S−N, O−O, O−N, N−N).[9,10] To maintain diversity throughout the optimization procedure, the maximal 3D-similarity (Sim$_{3D}$, eq 1) of two individuals within an island was set to 0.8 and coefficient γ to 0.9. All molecules were stored to disk every 20th iteration of the optimization procedure.

A single run with 1000 iterations took 49.5 ± 1.2 min, 55.8 ± 3.0 min, and 58.5 ± 1.7 min for the GANDI setups with 4 islands without exchange, 4 islands including exchange, and a single island, respectively, on a AMD Opteron 252 (single CPU, 2.6 GHz). The 1809 unique molecules obtained by the 10 optimization runs with 4 islands with exchange, for which CHARMm force field parameters were available, were analyzed and processed further. Only
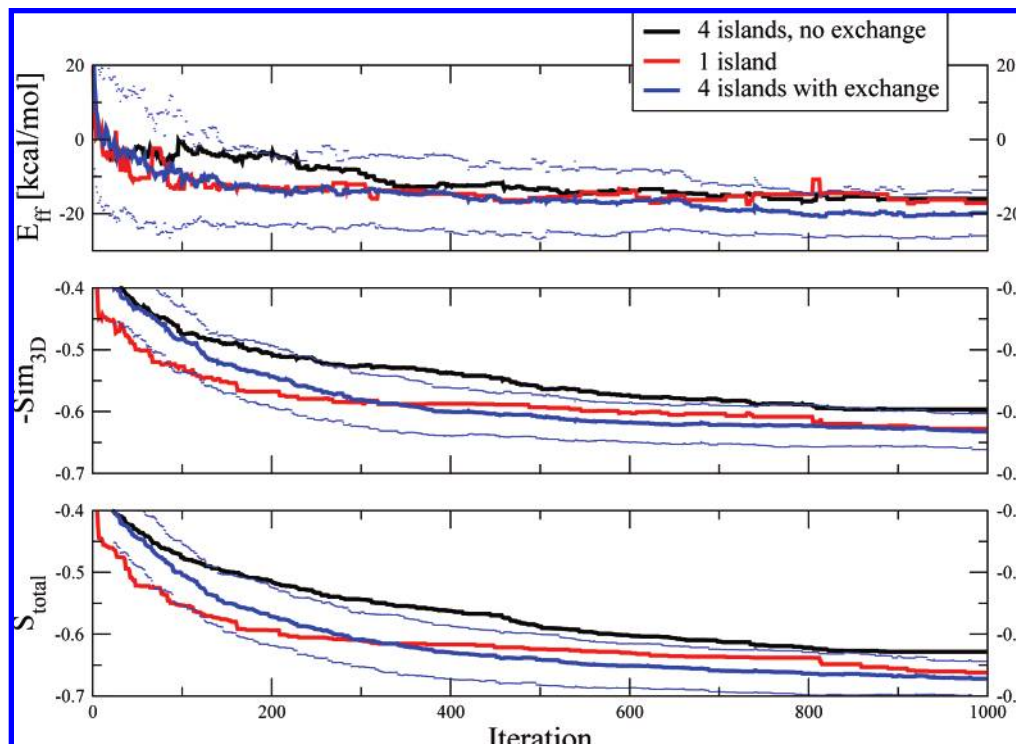
Figure 4. Evolution of the GANDI scoring function terms and the total score of the best individual in each island. The values at each iteration step were averaged over 10 runs. Optimization was performed by minimization of $S_{total} = 0.02E_{ff} - Sim_{3D}$. The bold lines are averages, while the thin blue lines are the standard deviation of the run with 4 islands including exchange.
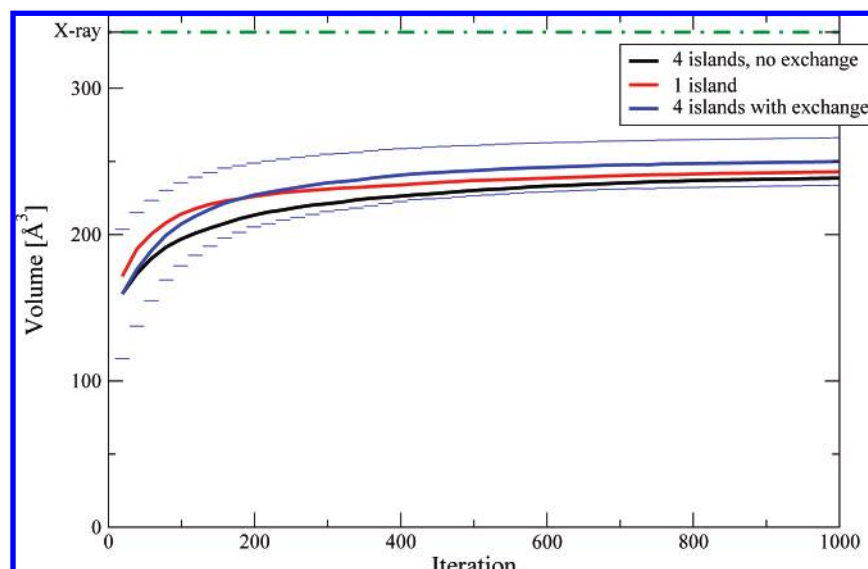
Figure 5. Evolution of the volume overlap between the crystallographic inhibitor and the molecules produced by GANDI. Every 20th iteration step the volume overlap is averaged over all individuals and all 10 runs. The thin blue lines are the standard deviation of the run with 4 islands including exchange of individuals. The volume of the minimized crystallographic inhibitor (338.37 Å^3) is shown in green. The volume was calculated with the COOR-module in CHARMM.[49]

one of two poses of the same molecule was retained if the DAIM-fingerprint[45] consisting of 17 entries was identical and the 3D-similarity Sim$_{3D}$ according to eq 1 was larger than 0.99.

Strain Energy. Atom types and partial charges were updated for the generated molecules and minimized with the CHARMm force field inside and in the absence of the rigid protein to determine the strain energy and the free energy of binding. The strain was calculated with a distance-dependent dielectric of $\epsilon(r) = 4r$ with

$$\Delta E_{\text{Strain}} = E_{\text{inside}}^{\text{compound}} - E_{\text{outside}}^{\text{compound}} \qquad (6)$$

where inside corresponds to the conformation of the molecule obtained by minimization inside the rigid receptor binding site, and outside corresponds to the conformation minimized in absence of the receptor.

Calculated Binding Energy. Upon ligand minimization inside the rigid protein with $\epsilon(r) = 4r$, the free energy of binding was calculated with a 2-parameter LIECE (linear interaction energy with continuum electrostatics) model[46,47]

$$\Delta G_{\text{bind}}^{\text{LIECE}} = \alpha \Delta E_{\text{vdW}} + \beta \Delta G_{\text{elec}} \qquad (7)$$

where ΔE_{vdW} is the ligand/protein van der Waals interaction energy, and ΔG_{elec} is the sum of the ligand/protein Coulombic energy in vacuo ($\epsilon = 1$) and the change in solvation energy of inhibitor and protein upon binding. The Coulombic energy in vacuo was calculated with infinite cutoff and neglecting 1−2 and 1−3 interactions. The electrostatic solvation energy was calculated with the finite difference Poisson approach implemented in CHARMM[48,49] using a final grid spacing of 0.4 Å. The α and β parameters had been derived by fitting $\Delta G_{\text{bind}}^{\text{LIECE}}$ to IC$_{50}$ values of 73 known inhibitors in a previous study, yielding $\alpha = 0.2866$ and $\beta = 0.0520$.[47]

Docking by DAIM/SEED/FFLD. An in-house developed suite of programs for flexible ligand docking was used to assess the quality of the binding modes generated by GANDI. First GANDI molecules were decomposed into fragments
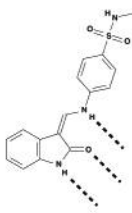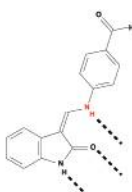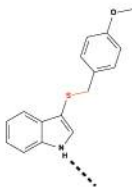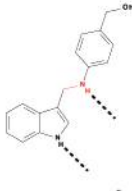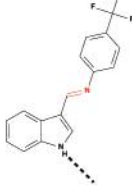
Table 1. 2D-Similarity Sim$_{2D}$ between Molecules Generated by GANDI and Known Inhibitors

inhibitor set	sorting	N_{best}[a]	N_{GANDI}	N_{KI} [%]
Bramson et al.[40]	$\Delta G_{\text{bind}}^{\text{LIECE}}$	25	5	60
(50 cpds with		50	10	64
nanomolar affinity)		100	13	64
	S_{total}	25	2	28
		50	4	66
		100	7	66
Gibson et al.[53]	$\Delta G_{\text{bind}}^{\text{LIECE}}$	25	0	0
(23 cpds with		50	4	91
micromolar affinity)		100	6	91
	S_{total}	25	2	57
		50	4	74
		100	5	74

[a] Number of GANDI molecules with most favorable $\Delta G_{\text{bind}}^{\text{LIECE}}$ or S_{total} considered for Sim$_{2D}$ evaluation. Values of Sim$_{2D}$ between the known inhibitors and the GANDI molecules were calculated based on normalized DAIM-fingerprints[45] (Supporting Information Table S1). Note that Sim$_{2D}$ was not used for optimization in GANDI. N_{GANDI} is the number of GANDI molecules with Sim$_{2D} > 0.9$ to at least one molecule from the respective inhibitor set. N_{KI} is the percentage of known inhibitors with Sim$_{2D} > 0.9$ to at least one GANDI molecule. The oxindole-based inhibitor used as a target structure (PDB code 1KE5) belongs to the nanomolar CDK2 inhibitors.

by DAIM.[45] The DAIM decomposition is fully automatic (i.e., it does not require any manual intervention) and proceeds in four phases: ring identification, initial fragment definition, functional group merging, and completion of the valences. DAIM prioritizes also the resulting fragments according to their suitability as anchors for the docking program FFLD.[50] In FFLD an efficient evaluation of binding energy (consisting of van der Waals and hydrogen bond terms) is used as the cost function of a genetic algorithm optimization in ligand dihedral space. The 1640 GANDI molecules which yielded at least three fragments upon decomposition with DAIM[45] were redocked into the rigid protein binding site by FFLD. These molecules were first minimized in the absence of the protein to remove any conformational bias introduced by GANDI. The fragments

GANDI: Fragment-Based de Novo Ligand Design Program

J. Chem. Inf. Model., Vol. 48, No. 3, 2008 **685**

Table 2. Molecules with Favorable GANDI Score S_{total} Generated in the 10 GANDI Runs with 4 Islands Including Exchange of Individuals[e]

Structure	Similarity to oxindole inhibitor (PDB code 1KE5)		GANDI			
	$Sim_{2D}{}^a$	$Sim_{3D}{}^b$	$S_{total}{}^c$	$N_{Sub}{}^d$	ΔG_{bind}^{LIECE} [kcal/mol]	MW [g/mol]
Target						
(X-ray 1KE5)	(1)	(1)	–	0	-10.4	329
GANDI molecules with optimal score S_{total}						
	0.855	0.705	-0.721	0	-9.6	264
	0.584	0.666	-0.694	6	-9.6	269
	0.800	0.659	-0.687	207	-9.3	252
	0.391	0.648	-0.682	285	-9.5	288

[a] Values of Sim_{2D} calculated according to eq 5 using normalized DAIM-fingerprints[45] (Supporting Information Table S1). [b] Values of Sim_{3D} calculated by GANDI using eq 1. [c] Values of S_{total} calculated by GANDI using eq 3. [d] Number of molecules found by DAIM[45] in the ZINC library containing the scaffold of the GANDI molecule (ring substituents and apolar hydrogens were removed to yield the query scaffold). [e] Atoms and bonds in red are linker fragments, while fragments in black were docked with SEED.[38] Dashed lines represent intermolecular hydrogen bonds to the protein backbone in the hinge region.

obtained by decomposition with DAIM were then minimized after updating the partial charges and atom types and docked into the receptor binding site with SEED.[38] The resulting anchor points of the fragments were used to run FFLD[50] three times with distinct random seed numbers. Each FFLD run consisted of 50 cycles of the genetic algorithm which is

sufficient to generate a diverse set of poses. For each molecule docked by FFLD the 300 energy-sorted poses with the lowest FFLD-energies were clustered with the leader algorithm using eq 1 as a similarity measure and a similarity cutoff of 0.6 yielding 16 820 poses (10.26 ± 7.08 per molecule). The cluster representatives were minimized with

Table 3. Molecules with Favorable ΔG_{bind}^{LIECE} Generated in the 10 GANDI Runs with 4 Islands Including Exchange of Individuals[a]

		Similarity to oxindole inhibitor (PDB code 1KE5)		GANDI			
	Structure	$Sim_{2D}{}^a$	$Sim_{3D}{}^b$	$S_{total}{}^c$	$N_{Sub}{}^d$	ΔG_{bind}^{LIECE} [kcal/mol]	MW [g/mol]
Target							
(X-ray 1KE5)		(1)	(1)	–	0	-10.4	329
GANDI molecules with optimal ΔG_{bind}^{LIECE}							
		0.809	0.463	-0.499	37	-10.7	278
		0.910	0.501	-0.515	51	-10.4	297
		0.777	0.515	-0.526	0	-10.3	317
		0.725	0.474	-0.497	0	-10.2	365

[a] The last molecule in this table is shown in Figure 3. See the legend of Table 2 for explanations.

the CHARMm force field in the rigid protein binding site, and the LIECE binding free energy[47] was calculated for all poses as detailed above for the molecules generated by GANDI.

2D-Similarity Analysis. In the present application to CDK2, 2D-similarity was used only for analysis and not during GANDI runs. 2D-similarity between molecules was assessed with the Tanimoto coefficient[34] (Sim_{2D}, eq 5) based on normalized DAIM-fingerprints.[45] Normalization was performed by dividing the regular fingerprint values by the median DAIM-fingerprint values of all unique molecules

from the ZINC library[51] (version 5, Supporting Information Table S1).

RESULTS

The serine/threonine kinase CDK2 used in the present study shares the common topology of the ATP binding site found in most kinases. Intermolecular hydrogen bonds are key interactions found between most known inhibitors of CDK2 and the hinge region connecting the N- and C-terminal lobes of the protein (Figure 3, right). A hydrophobic pocket resides at the buried end (pocket in green in the top part of

GANDI: Fragment-Based de Novo Ligand Design Program

J. Chem. Inf. Model., Vol. 48, No. 3, 2008 **687**

Figure 3, left) near the conserved DFG motif of the binding site. By using a diverse library of fragments and due to the relatively small size of the ATP-binding site, key interactions are expected to be fulfilled by sets of nonoverlapping fragments.

Optimization. The 3D-similarity term Sim_{3D} in the scoring function seems to dominate the total score given the coefficients of the weighted-sum approach described in the Methods section (Figure 4). Yet, the force field-based term E_{ff} is essential for optimizing both binding interactions and intraligand interactions (eq 4). In fact, the 3D-similarity to the target molecule decreases almost steadily during optimization (Figure 4, middle), while the force field-based term (Figure 4, top) decreases fast initially and fluctuates within a certain range afterward. The single island setup displays the fastest decrease in the total score of the fittest individual, while the genetic algorithm with 4 islands including exchange decreases to similar final values. On the other hand, running the genetic algorithm with 4 noninteracting islands results in worse scores throughout. Preliminary GANDI runs with 800 individuals in a single or 100 individuals in 8 islands with exchange of individuals displayed even a faster decrease in the overall score per iteration at a 2-fold increase in computational cost (data not shown).

Similarly, the overlapping volumes between the target structure and the molecules generated by GANDI increase fast initially and only change marginally afterward (Figure 5). The increasing volume overlap is a direct effect of the 3D-similarity based scoring function term Sim_{3D}, which penalizes deviations of the GANDI molecule and its binding mode from the target inhibitor.

Taken together, the results indicate that the size of the population predetermines the amount of search space to be sampled and thus has a direct effect on the quality of the solutions found. Hence, the larger the number of individuals which can exchange genetic material, the faster the decrease in total scores. But these enlarged populations are associated with an increase in CPU time, which might offset the overall improvement. The migration of individuals between islands successfully prevents the algorithm from getting trapped in less favorable regions of the search space as observed in the case of the genetic algorithm setup with noninteracting islands.

Druglikeness. Most of the molecules generated by GANDI fulfill Lipinski's rule of five[52] and do not have stereocenters (Figure 2). This result is not surprising because the size of the binding site, the target structure, and the set of fragments predetermine the features of the molecules designed by GANDI.

Similarity to Known CDK2 Inhibitors. Several of the 100 GANDI molecules with the most favorable S_{total} or ΔG_{bind}^{LIECE} are similar (Sim_{2D}) to either the oxindole or another series of CDK2 inhibitors[53] (Table 1). A comparison with the crystal structure of CDK2 in the complex with the oxindole inhibitor[40] shows that GANDI molecules include key motifs of the target structure, e.g., the two ring systems joined by a linker (Tables 2 and 3, Figure 3). Note that the oxindole inhibitor used as target for the 3D-similarity (PDB code 1KE5) is not generated by GANDI because benzyl-sulfonamide is not present in the Molinspiration library. Scaffold or linker hopping is observed in some of the
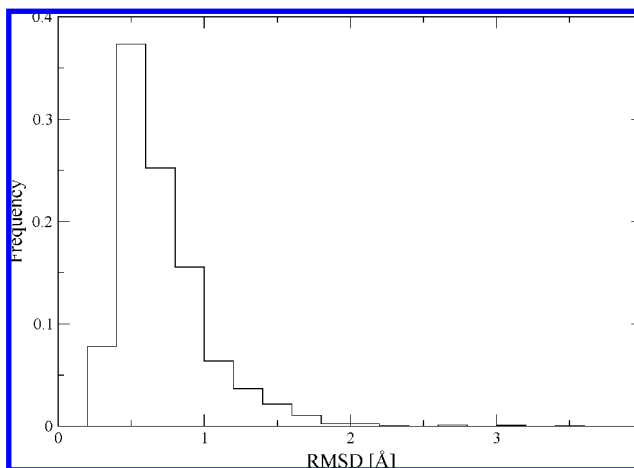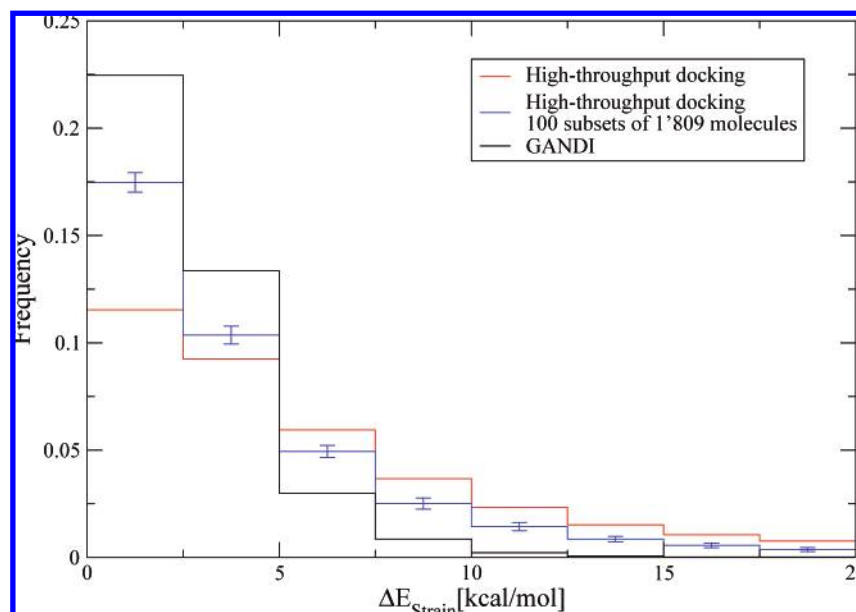

Figure 6. Normalized distribution of heavy atom rmsd between the conformations generated by GANDI and their CHARMm-minimized counterparts.

compounds generated by GANDI (Tables 2 and 3) and is a consequence of the enforced 3D-structural diversity within populations during genetic algorithm optimization. Despite the remarkable similarity with known inhibitors some of the GANDI molecules are likely to be inactive, e.g., the compound with the $-CH_2-NH-$ linker which lacks a hydrogen bond acceptor for the backbone NH of the hinge region (Table 2).

One striking result is the GANDI compound in the bottom of Table 3 which contains a 2-N-phenyl-1,3-thiazole-2,4-diamine moiety. The three nitrogen atoms of its diaminothiazole scaffold are involved in three hydrogen bonds with the hinge region in the pose generated by GANDI (Figure 3, right). An anonymous referee pointed out that diaminothiazoles are known inhibitors of CDK2. In particular the 2-N-phenyl-1,3-thiazole-2,4-diamine moiety is present in a 220 nM inhibitor of CDK2 (compound **15** in the review article[54]).

Binding Mode and Free Energy of Binding. The binding poses of the molecules generated by GANDI display low rmsd values from their CHARMm-minimized counterparts (Figure 6), indicating that the former are structurally close to a local minimum conformation of the CHARMm force field (Accelrys Inc.). The distribution of strain energies of all molecules generated by GANDI is also more favorable compared with a distribution derived from a previously reported high-throughput docking campaign[47] (Figure 7). Since molecular weight distributions of the two sets differ, with the GANDI molecules generally being smaller, a set of poses with molecular weight distribution as the GANDI molecule set was randomly selected 100 times from the library of molecules used in the high-throughput docking campaign. Removal of the size bias results in a shift of the corresponding distribution to smaller strain energies. Nonetheless the molecules from GANDI still display lower strain energies.

The distribution of the calculated binding energy versus the molecular weight reveals that most of the molecules designed by GANDI are located between two series of known inhibitors with micromolar[53] and nanomolar[40] activity (Figure 8). The redocking of GANDI molecules with DAIM, SEED, and FFLD yields an average of 10 poses per compound. Comparing the poses with the lowest ΔG_{bind}^{LIECE} of all redocked molecules with the binding modes produced

Figure 7. Normalized distribution of strain energies of all molecules generated by GANDI (black) and, as a comparison, all poses resulting from a high-throughput docking campaign[47] (red). The blue histogram is the average over 100 randomly picked subsets of the latter with equal molecular weight distribution (bin size of 10 g mol^{-1}) as the set of GANDI molecules to rule out any size dependency of the strain energy.
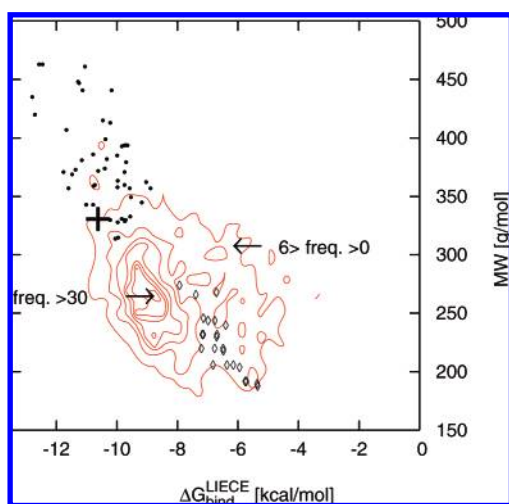
Figure 8. Two-dimensional histogram of calculated free energy of binding versus molecular weight of the 1809 GANDI molecules (bin sizes of 0.29 kcal mol^{-1} and 9.57 g mol^{-1} for the X- and Y-axis, respectively). Isofrequency lines of the histogram (red) are shown every frequency-interval of five units. The plus symbol denotes the oxindole inhibitor present in the crystal structure, and the dots and diamonds are two series of known inhibitors of CDK2 with nanomolar[40] and micromolar[53] activity, respectively, which had been used in a previous study.[47]

by GANDI revealed that 32% of the former show a distinct binding mode (RMSD > 2 Å) with more favorable free energy of binding than the latter. Furthermore, only 9% of the redocked molecules have a ΔG_{bind}^{LIECE} more favorable by more than 1 kcal mol^{-1} than the GANDI pose. These results indicate that GANDI mostly generates optimal poses.

Availability of Molecules Designed by GANDI. Analysis of the existence of GANDI molecules in the ZINC library[51] (version 5) reveals that eight out of 1809 GANDI molecules are commercially available. Additionally, 586 molecules with a scaffold identical to one of the eight GANDI compounds shown in Tables 2 and 3 are retrieved from the ZINC library

by performing a substructure search with DAIM[45] (Supporting Information Tables S2–S6). On the other hand, no molecules containing the scaffold of the oxindole-based inhibitor are present in ZINC.

CONCLUSIONS

GANDI is a new computer program for fragment-based de novo ligand design. A combination of genetic algorithm and tabu search is used in GANDI for the simultaneous optimization of force field energy *and* (2D or 3D) similarity to known inhibitor(s). Therefore, the design is both binding site-based and ligand-based. Importantly, the relative importance of these two *driving forces* can be modulated by the user.

In an application to the CDK2 kinase 1809 molecules were generated by GANDI within the ATP-binding site in less than 10 h on a low-cost PC using a library of 14 000 fragments with up to three rotatable bonds. The binding mode of these molecules show low strain compared to poses of commercially available compounds of similar size originating from a previous high-throughput docking campaign.[47] In addition the binding modes are structurally close to a local minimum of the CHARMm force field, and their predicted free energy of binding (calculated using continuum electrostatics solvation) is within the range of known inhibitors of CDK2. Notably, molecules similar to those generated by GANDI are commercially available providing further evidence of the usefulness of GANDI for de novo drug design.

GANDI can generate molecules similar to known kinase inhibitors. Importantly, by enforcing diversity throughout the optimization and by using a 3D-similarity-based scoring function term Sim$_{3D}$, which does not rely on a covalent structure of the compared molecules, scaffold or linker hopping was observed, retaining the common binding motifs of known kinase inhibitors. A striking example is the GANDI compound on the bottom of Table 3 which shares a 2-N-phenyl-1,3-thiazole-2,4-diamine moiety with a 220 nM

GANDI: FRAGMENT-BASED DE NOVO LIGAND DESIGN PROGRAM

J. Chem. Inf. Model., Vol. 48, No. 3, 2008 **689**

inhibitor of CDK2.[54] Its central diaminothiazole scaffold is involved in the same intermolecular hydrogen-bonding pattern with the hinge region as several known inhibitors of CDK2 (Figure 3, right).

Because a large variety of ATP-binding site inhibitors of kinases form hydrogen bonds with the hinge region, in the present application of GANDI the 3D-similarity to the *binding mode* of a known nanomolar inhibitor of CDK2 was enforced (using Sim_{3D}). For protein targets with known inhibitors of uncertain (or unknown) binding mode it might be appropriate to bias the GANDI design by the 2D-similarity to known inhibitors (using Sim_{2D}). For this purpose, the relative importance of the force field binding energy and 2D-information (i.e., 2D ligand-based design) can be tuned by the multiplicative parameter of Sim_{2D}. Furthermore, chimeric molecules can be designed by GANDI using two or more known inhibitors as targets for the 2D- or 3D-similarity evaluation. Chimerization takes place spontaneously because during optimization the genetic algorithm enforces 3D-diversity among molecules in the population. This ligand-based design approach and other applications of GANDI are currently being investigated in our research group.

Availability of GANDI. The program GANDI as well as its documentation and test cases are available for free to not-for-profit institutions.

ACKNOWLEDGMENT

Supporting Information Available: ZINC molecules retrieved using GANDI molecules as query, DAIM-fingerprint[45] normalization values, genetic algorithm exchange modes, linker attachment, evolution of structural similarity, Sim_{2D} matrices and distributions, and $\Delta G_{bind}^{LIECE} - S_{total}$ scatter plot. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug. Discovery* **2005**, *4*, 649−663.

(2) Böhm, H.-J.; Banner, D. W.; Weber, L. Combinatorial docking and combinatorial chemistry: Design of potent non-peptide thrombin inhibitors. *J. Comput.-Aided Mol. Des.* **1999**, *V13*, 51−56.

(3) Schneider, G.; Lee, M.-L.; Stahl, M.; Schneider, P. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 487−494.

(4) Vinkers, H.; de Jonge, M.; Daeyaert, F.; Heeres, J.; Koymans, L.; van Lenthe, J.; Lewi, P.; Timmerman, H.; Van Aken, K.; Janssen, P. SYNOPSIS: SYNthesize and OPtimize System in Silico. *J. Med. Chem.* **2003**, *46*, 2765−2773.

(5) Firth-Clark, S.; Todorov, N.; Alberts, I.; Williams, A.; James, T.; Dean, P. Exhaustive de novo Design of Low-Molecular-Weight Fragments Against the ATP-Binding Site of DNA-Gyrase. *J. Chem. Inf. Model.* **2006**, *46*, 1168−1173.

(6) Gillet, V. J.; Myatt, G.; Zsoldos, Z.; Johnson, A. P. SPROUT, HIPPO and CAESA: Tools fo de novo structure generation and estimation of synthetic accessibility. *Perspect. Drug Discovery Des.* **1995**, *3*, 34−50.

(7) Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S. LEA3D: A Computer-Aided Ligand Design for Structure-Based Drug Design. *J. Med. Chem.* **2005**, *48*, 2457−2468.

(8) Fechner, U.; Schneider, G. Flux (2): Comparison of Molecular Mutation and Crossover Operators for Ligand-Based de Novo Design. *J. Chem. Inf. Model.* **2007**, *47*, 656−667.

(9) Rotstein, S. H.; Murcko, M. A. GroupBuild: A fragment-based method for de novo drug design. *J. Med. Chem.* **1993**, *36*, 1700−1710.

(10) Wang, R.; Gao, Y.; Lai, L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *J. Mol. Model.* **2000**, *6*, 498−516.

(11) Murcko, M. A.; Bemis, G. W. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.

(12) Murcko, M. A.; Bemis, G. W. Properties of Known Drugs. 2. Side Chains. *J. Med. Chem.* **1999**, *42*, 5095−5099.

(13) Ertl, P. Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374−380.

(14) Lewell, X.; Judd, D.; Watson, S.; Hann, M. RECAP-Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Model.* **1998**, *38*, 511−522.

(15) Holland, J. H. *Adaptation in natural and artificial systems*; University of Michigan Press: Ann Arbor, MI, 1975.

(16) Goldberg, D. E. *Genetic Algorithms in Search Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.

(17) Pegg, S. C.; Haresco, J. J.; Kuntz, I. D. A genetic algorithm for structure-based de novo design. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 911−933.

(18) Todorov, N. P.; Dean, P. M. Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 175−192.

(19) DeWitte, R.; Ishchenko, A.; Shakhnovich, E. SMoG: de novo design method based on simple, fast, and accurate free energy estimates. 2. Case studies in molecular design. *J. Am. Chem. Soc.* **1997**, *119*, 4608−4617.

(20) Darwin, C. *On The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*; John Murray: London, 1859.

(21) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.

(22) Apostolakis, J.; Caflisch, A. Computational ligand design. *Comb. Chem. High Throughput Screening* **1999**, *2*, 91−104.

(23) Coupez, B.; Lewis, R. A. Docking and Scoring - Theoretically Easy, Practically Impossible? *Curr. Med. Chem.* **2006**, *13*, 2995−3003.

(24) Caflisch, A. Computational combinatorial ligand design: Application to human α-thrombin. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 372−396.

(25) Bäck, T.; Fogel, D. B.; Michalewicz, Z. *Handbook of Evolutionary Computation*; Institute of Physics Publishing and Oxford University Press: 1997, pp C4.5:2.

(26) Schneider, G.; Clément-Chomienne, O.; Hilfiger, L.; Schneider, P.; Kirsch, S.; Böhm, H.-J.; Neidhart, W. Virtual Screening for Bioactive Molecules by Evolutionary De Novo Design. *Angew. Chem., Int. Ed.* **2000**, *39*, 4130−4133.

(27) Rogers-Evans, M.; Alanine, A.; Bleicher, K.; Kube, D.; Schneider, G. Identification of novel cannabinoid receptor ligands *via* evolutionary de novo design and rapid parallel synthesis. *QSAR Comb. Sci.* **2004**, *23*, 426−430.

(28) Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D. Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 663−667.

(29) Firth-Clark, S.; Willems, H.; Williams, A.; Harris, W. Generation and Selection of Novel Estrogen Receptor Ligands Using the *De Novo* Structure-Based Design Tool, SkelGen. *J. Chem. Inf. Model.* **2006**, *46*, 642−647.

(30) Grosso, P. B. Computer Simulations of Genetic Adaptation: Parallel Subcomponent Interaction in a Multilocus Model, Ph.D. Thesis, Computer and Communication Sciences Dept, University of Michigan, 1985.

(31) Glover, F. Future paths for integer programming and links to artificial intelligence. *Comput. Oper. Res.* **1986**, *13*, 533−549.

(32) Glover, F. Tabu Search− Part I. *ORSA J. Comput.* **1989**, *1*, 190−206.

(33) Glover, F. Tabu Search− Part II. *ORSA J. Comput.* **1990**, *1*, 4−32.

(34) Tanimoto, T. *IBM Internal Report*; IBM Technical Report Series; Nov 17, 1957.

(35) Brindle, A. Genetic algorithms for function optimization, Ph.D. Thesis, Computer Science Dept, University of Alberta, 1981.

(36) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to Algorithms*; MIT Press and McGraw-Hill: 2001; pp 531−539.

(37) Kearsley, S.; Smith, G. An alternative method for the alignment of molecular structures. Maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615−633.

(38) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 88−105.

(39) Momany, F. A.; Rone, R. Validation of the general purpose QUANTA 3.2/CHARMm force field. *J. Comput. Chem.* **1992**, *13*, 888−900.

(40) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T.; Harris, P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-Based Inhibitors of Cyclin-Dependent Kinase 2 (CDK2): Design, Synthesis, Enzymatic Activities, and X-ray Crystallographic Analysis. *J. Med. Chem.* **2001**, *44*, 4339−4358.

(41) Sadowski, J.; Gasteiger, J. From atoms and bonds to three-dimensional atomic coordinates: automatic model builders. *Chem. Rev.* **1993**, *93*, 2567−2581.

(42) Guha, R.; Howard, M.; Hutchison, G.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. The Blue Obelisk-Interoperability in Chemical Informatics. *J. Chem. Inf. Model.* **2006**, *46*, 991−998.

(43) No, K.; Grant, J.; Scheraga, H. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4732−4739.

(44) No, K.; Grant, J.; Jhon, M.; Scheraga, H. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 2. Application to ionic and aromatic molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4740−4746.

(45) Kolb, P.; Caflisch, A. Automatic and Efficient Decomposition of Two-Dimensional Structures of Small Molecules for Fragment-Based High-Throughput Docking. *J. Med. Chem.* **2006**, *49*, 7384−7392.

(46) Huang, D.; Caflisch, A. Efficient Evaluation of Binding Free Energy Using Continuum Electrostatics Solvation. *J. Med. Chem.* **2004**, *47*, 5791−5797.

(47) Kolb, P.; Huang, D.; Dey, F.; Caflisch, A. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J. Med. Chem.* **2008**, *51*, 1179−1188.

(48) Im, W.; Beglov, D.; Roux, B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59−75.

(49) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(50) Budin, N.; Majeux, N.; Caflisch, A. Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.* **2001**, *382*, 1365−1372.

(51) Irwin, J.; Shoichet, B. ZINC − a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.

(52) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and developmental settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.

(53) Gibson, A. E.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Curtin, N. J.; Davies, T. G.; Endicott, J. A.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Whitfield, H. J. Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O6-substituted guanine derivatives. *J. Med. Chem.* **2002**, *45*, 3381−3393.

(54) Toogood, P. L. Cyclin-dependent kinase inhibitors for treating cancer. *Med. Res. Rev.* **2001**, *21*, 487−498.

CI700424B