Exhaustive de novo Design of Low-Molecular-Weight Fragments Against the ATP-Binding Site of DNA-Gyrase

Stuart Firth-Clark,* Nikolay P. Todorov, Ian L. Alberts, Anthony Williams, Timothy James,† and Philip M. Dean

De Novo Pharmaceuticals Ltd., Compass House, Vision Park, Histon, Cambridge CB4 9ZR, U.K.

Received August 22, 2005

We present a de novo design approach to generating small fragments in the DNA-gyrase ATP-binding site using the computational drug design platform SkelGen. We have generated an exhaustive number of structural possibilities, which were subsequently filtered for site complementarity and synthetic tractability. A number of known active fragments are found, but most of the species created are potentially novel and could be valuable for further elaboration and development into lead-like structures.

INTRODUCTION

Fragment-based approaches[1−9] identify small molecular fragments (MW <300 Da) that show weak activity against a particular target. The active fragments consist only of structural elements that complement the corresponding active site and can subsequently be "grown" into larger ligands using medicinal chemistry techniques.[1] Fragment-based methods are particularly useful when "traditional" techniques such as High-Throughput Screening (HTS) fail to identify a hit. This is because structural information on the target is used to guide computational and medicinal chemists into an area of chemical space that has greater probability in providing that hit. A successful implementation of the fragment-based approach was carried out[11] against DNA-gyrase, a potential antibacterial target responsible for catalyzing the process of DNA supercoiling.[10] Boehm et al. used virtual screening to select potentially active fragments and subsequently determined the presence or absence of fragment binding to the ATP-binding site of the enzyme using a series of biophysical techniques.[11]

The value of fragment-based approaches is evidently dependent on finding appropriate starting fragments. Fragmentation techniques[12] and analysis[13−15] have been applied to known drugs, highlighting potential fragments that could be subsequently screened against chosen targets. The advantage of using a de novo design program is the ability to combine fragments together, thus providing an opportunity to generate species that may not be present in the initial data set. Herein, we describe a de novo design strategy that is used to find potentially novel fragments for the ATP-binding site of DNA-gyrase. To gain as many potential fragments as possible, we exhaustively generate structures in the binding site and show that we approach the total number of possible solutions available under simulation constraints. We then apply a filtering protocol that removes fragments which do not complement the binding site well or are not commercially available or synthetically tractable. From this procedure, we identified all of the active fragments reported by Boehm et al.[11] as well as advocate a number of novel alternative structures for experimental testing. This protocol therefore provides a set of fragments that includes commercially available compounds and potentially novel species that require simple synthesis. The fragments generated in this study are prime candidates for further 'growth' and chemical modification in the design of lead-like ligands.

METHODS

We used the de novo design method SkelGen,[16−18] which generates ligand structures by linking together specified input fragments through a set of bonding rules. Compounds from the World Drug Index (WDI) database were fragmented to generate a number of input fragments. This ensures a nonspecific set of input fragments that are appropriate for ligand design across a wide range of targets. These fragments were divided into two sections: the first section contained 12 acyclic fragments, allowing up to 10 to be included in each designed ligand. The second section comprised around 2000 cyclic fragments, where only one was allowed per generated ligand. By restricting growth in this way, the molecular weight of the output was kept below 300. Structure generation was driven by simulated annealing[19] of a penalty function that includes terms for inter- and intramolecular interactions, pharmacophore constraints, and chemical composition. Conformational adjustments along exo-cyclic bonds are permitted as structures are grown, allowing full rotation around single bonds, and 0/180 degree torsion changes across double bonds. Solutions were accepted when all constraints defined on the ligand structure were satisfied. The ligands were truncated to include only atoms interacting with pharmacophore points while keeping ring systems intact. Therefore, each solution represented a bare fragment with no substituents, facilitating structure classification and enumeration.

Protein Structure. The crystal structure of 1EI1[20] was used in which the B-chain, water, and ligands were removed. Prior to deleting the solvent and ligand, hydrogen atoms were

* Corresponding author phone: +44 1223 238000; fax: +44 1223 238088; e-mail: stuart.firth-clark@denovopharma.com.

† Present address: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

10.1021/ci050338i CCC: $33.50
Published on Web 01/12/2006

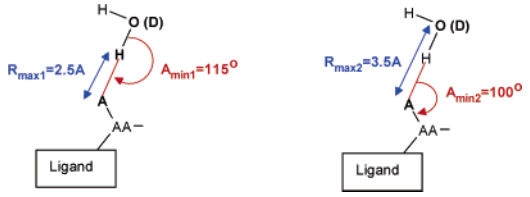


Figure 1. Interaction limits for the directed donor—the ligand acceptor is placed at a maximum distance of 2.5 Å from the exposed water hydrogen, with a corresponding minimum D−H−A angle of 115°. In addition, the ligand acceptor must be placed within a distance of 3.5 Å of the water oxygen while maintaining an AA−A−H angle of at least 100°. The atom, AA, refers to an atom that is attached to the ligand acceptor.

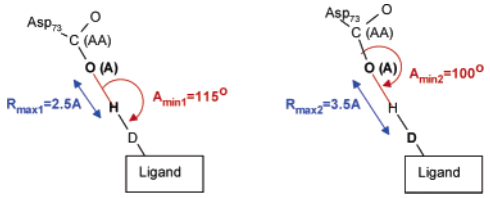


Figure 2. Interaction limits for the directed acceptor—the ligand-donor hydrogen is placed at a maximum distance of 2.5 Å from the OD2 acceptor atom of Asp-73, while a corresponding minimum D−H−A angle of 115° must be retained. In addition, the ligand donor must be placed within a distance of 3.5 Å of the Asp-73-OD2 acceptor atom while maintaining an AA−A−H angle of at least 100°. The atom, AA, refers to the carboxylate carbon (CD1) of the Asp-73.

added and subsequently minimized if any atom of the corresponding residue was within 15 Å of the ligand. Minimization was done using the conjugate gradients algorithm in InsightII/Discover-3 (Accelrys Inc., San Diego, CA) to a maximum gradient of 0.001 kcal mol^{-1} $Å^{-1}$. Heavy atoms were fixed to their crystal structure coordinates.

Fragment Generation. SkelGen was used to generate fragments within the ATP binding site of DNA-gyrase that satisfied two pharmacophoric restraints, supplied as input: a directed donor defined the interaction between the wat-1601 ligand-exposed OH and the ligand acceptor. The interaction limits were defined by minimum angles and maximum distances as shown in Figure 1. A directed acceptor restraint (Figure 2) defined the interaction limits between the Asp-73 carboxylate OD2 hydrogen-bond acceptor and the ligand, applying the same restrictions to that of the directed donor restraint. A rectangular box was also supplied that defined the "SkelGen Active Site"—the volume in the protein active site in which structure generation took place. This was defined using the original ADPNP crystallized ligand, **1**. In total, 100 000 structures were requested.

RESULTS

Generation Rate of Unique Structures. SkelGen generates structures stochastically,[21] and as the number of solutions generated increases, duplicates start to appear. To model the accumulation of unique solutions, let $u(t)$ be the total number of solutions at step t. We make the assumptions that (i) the probability of finding each unique solution is the same and (ii) the rate of generation of unique solutions, $du(t)/dt$, is proportional to the number of unique solutions that have not yet been discovered. In a linear model

$$du(t)/dt = a(b - u(t))$$

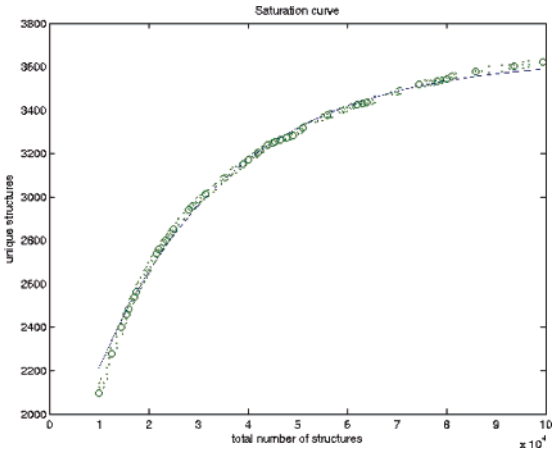


Figure 3. First-order exponential model prediction (dashed line) compared to SkelGen generated data (green circles).

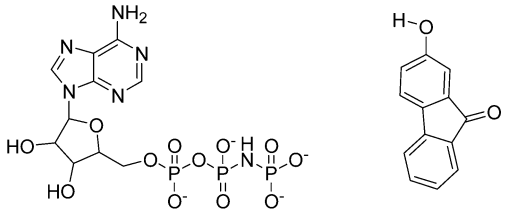
where b is the total number of unique solutions for the problem. The number of unique structures u, when t structures are generated can be expressed as

$$u(t) = b + c \cdot \exp(a \cdot t)$$

From data points (t_i, u_i) for solution generation and using a simplex simulated annealing global optimization technique,[22] a, b, and c can be estimated.

When analyzed, the 100 000 structures were found to contain 3600 different solutions. For the linear model, we estimated $a = -3.7 \times 10^{-5}$, $b = 3641.3$, and $c = -2069.1$. The predicted maximum number of structures potentially produced in this experiment is therefore 3641. However, the model underestimates the numbers of structures generated at later times (Figure 3) to within 1.5% of the actual numbers generated. Assuming this still holds beyond $t = 100\,000$, the upper limit for the total number of unique structures generated was estimated to be about 3700.

Filtering the Fragments. The 3600 fragments were filtered using the ScreenScore function that has been shown to work well in virtual screening exercises.[23] Better ScreenScore values can be obtained with larger ligands, thereby biasing the filtered output toward these species. Therefore, a more appropriate property was computed—the "ScreenScore per heavy atom" (SSHA)—which estimates ligand complementarity[24] in the binding site. Initially, the SSHA was determined for the adenine moiety (Figure 4g) in the ADPNP crystal ligand, **1**, and fragments that did not score as well were removed from the data set. It was found that the vast majority of multicycles (e.g. **2**, fused tricycles, and greater) did not complement the active site sufficiently to be accepted. We also applied an amine interaction filter that removed fragments using an unprotonated amine to satisfy both Asp-73 and water pharmacophoric constraints. In total about 500 fragments were eliminated from the original 3600.



1 **2**

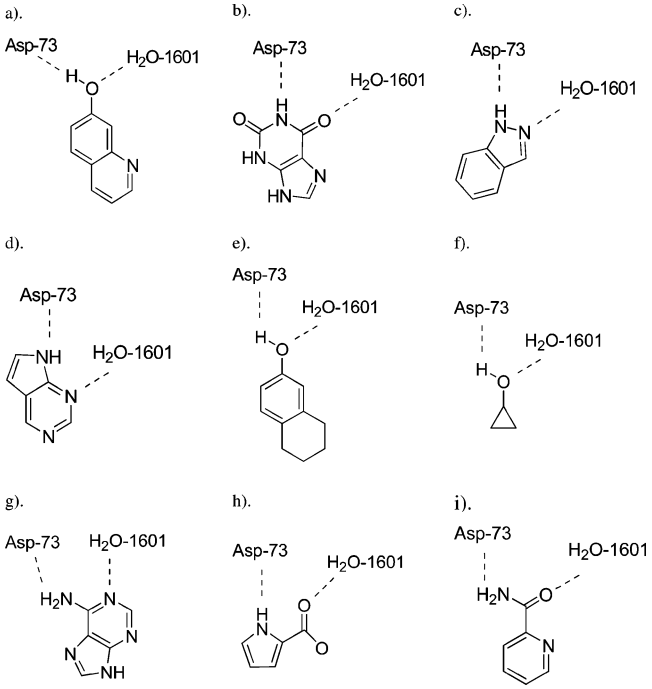


Figure 4. Example fragments generated by SkelGen.

The remaining 3100 fragments were passed through a synthesis-filtering protocol that involved searching the ACD[25] and retrosynthetically backtracking each fragment to its constituent components. This used a list of 107 possible chemical transformations, including simple functional group chemistry, nucleophilic substitution, aromatic substitution, heterocycle formation, and condensations. We only accepted fragments that were present in the ACD database or could be synthesized from reactants listed in the ACD (regardless of the number of synthesis steps required). Using this scheme, 2100 structures were removed but of the remaining 1000, nearly 400 were found in the ACD. A further 250 of the 600 noncommercial fragments could potentially be made from a one-step synthesis, increasing the number of practically useful fragments generated to 650.

Analyzing the Output. Using standard analysis tools (PCA, Ward) within Cerius2 (Accelrys, Inc., San Diego, CA), it was possible to identify a number of general structural features within the filtered data set of 1000 fragments. These features are summarized by a number of diverse examples taken from this data set (Figure 4).

Interactions Using External Groups Only. These involve either hydroxyl (e.g. Figure 4a,e,f) or amide (e.g. Figure 4i) groups to satisfy the pharmacophoric constraints.

Interactions Using Ring Atoms Only. Ring systems such as indazoles, Figure 4c, can provide adjacent donor and acceptor nitrogen atoms to satisfy the Asp-73 and water constraints. Another possibility is for the donor and acceptor to be located in different rings of a fused ring system (e.g. Figure 4d).

Interactions Using Both Ring and External Atoms. Multifunctional rings such as xanthines, Figure 4b, are able to use the amide function to satisfy the pharmacophore. In adenine (Figure 4g), the ring provides a nitrogen acceptor, while the external NH_2 acts as the donor. In some fragments, the converse is true—the ring provides the donor while an external carbonyl-oxygen (perhaps from a carboxylate, keto, or aldehyde group) acts as an acceptor (e.g. Figure 4h).

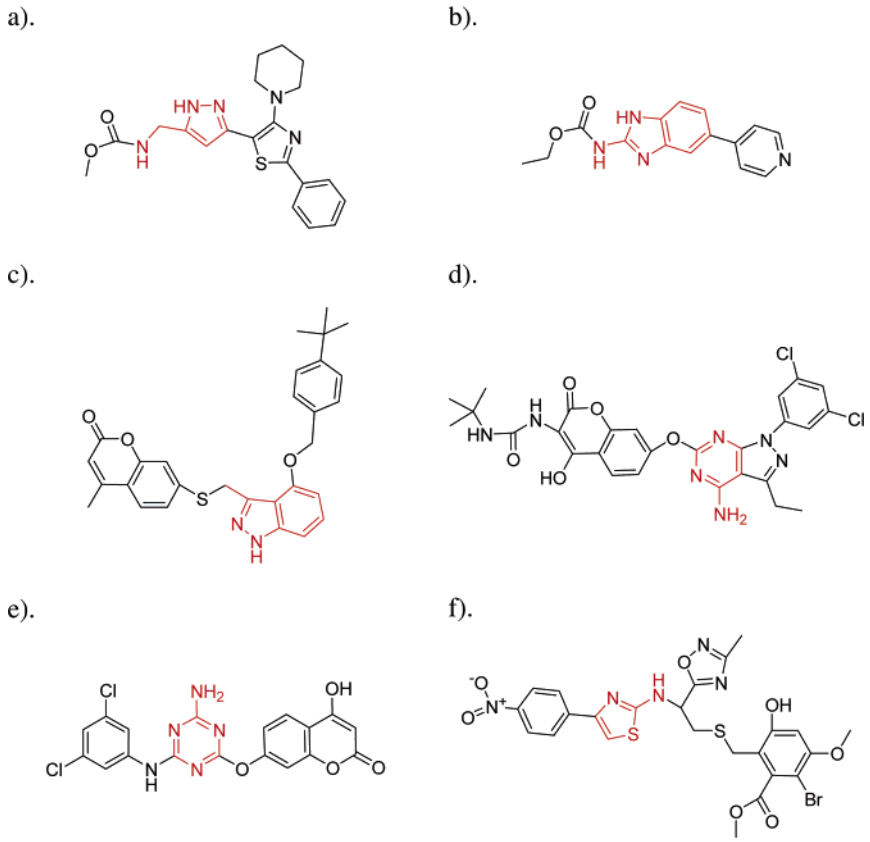


Figure 5. A selection of known active DNA-gyrase compounds showing the corresponding fragment in orange: a). substituted pyrazole [WD-2001-011450], b). 2-aminobenzimidazole [WD-2002-013059], c). indazole [WD-2000-010054], d). amino pyrimidine [WD-2000-004714], e). amino triazine [WD-97-011110], and f). amino thiazole [WD-2001-002916].

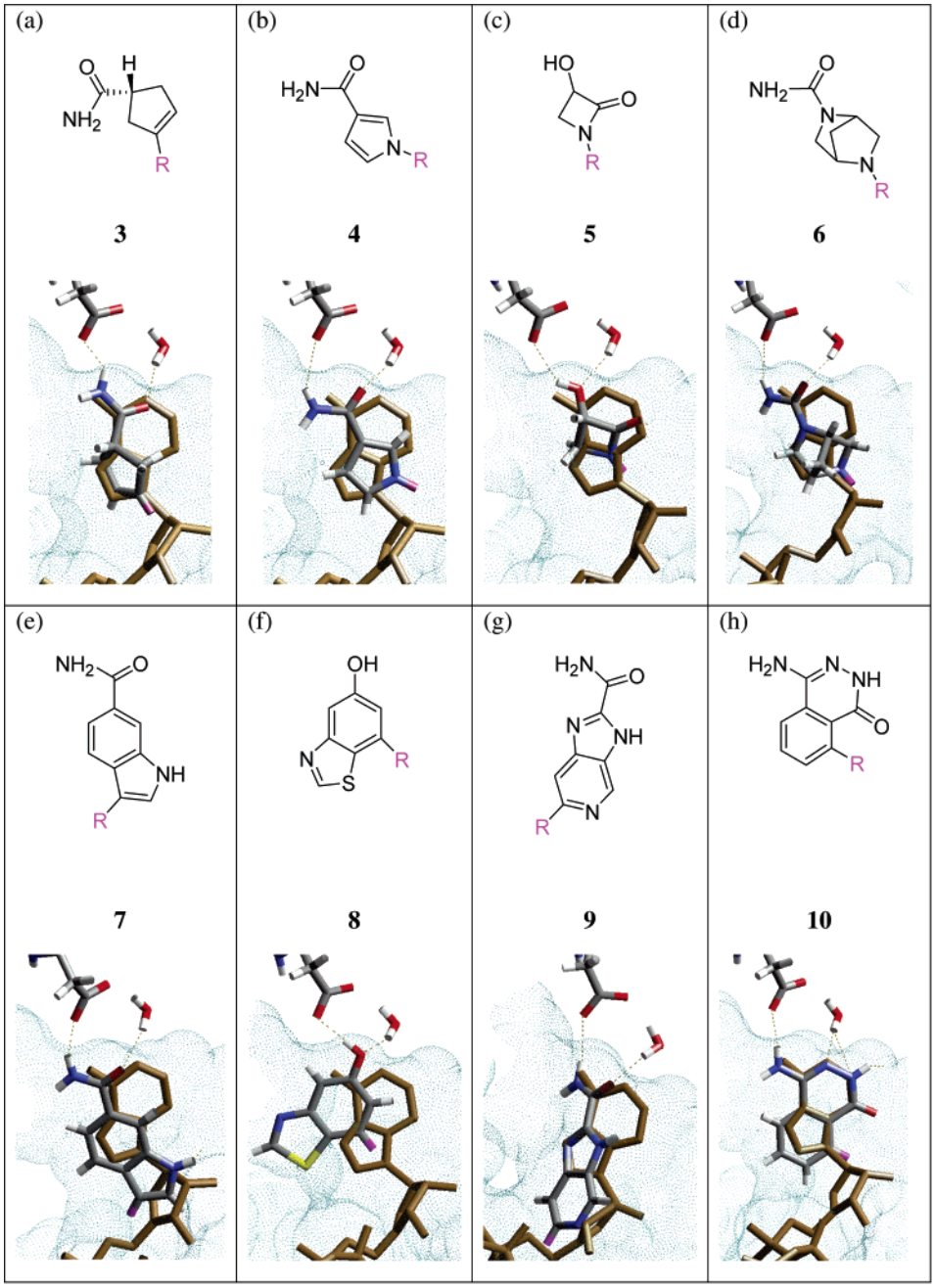


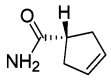
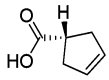
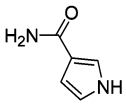
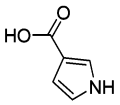
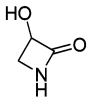
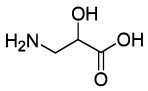
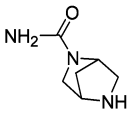
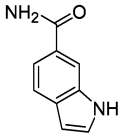
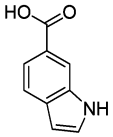
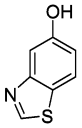
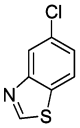
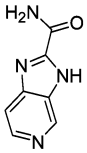
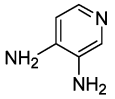
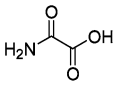
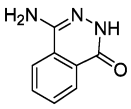
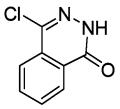
Figure 6. 2D and 3D representations of eight example fragments that have been selected based on the requirement for synthesis, potential novelty, and appropriate exit vector. The 2D representation highlights the exit vector with a purple "R", while this is represented in the 3D model as a magenta atom. For comparison, the ADPNP ligand found in the crystal structure is included in the 3D images (light brown). Yellow dashed lines in the 3D pictures represent potential hydrogen bonds.

Rings. A common motif was planar 5,6-membered fused rings such as adenine (Figure 4g), indazole, (Figure 4c), pyrrolopyrimidine (Figure 4d), and xanthine (Figure 4b). 6,6-membered rings were also obtained such as quinoline (Figure 4a). Additionally, some fragments contained a fused ring system in which one of the rings was planar, and the other was not (Figure 4e). Monocycles were also obtained ranging from three-membered rings (Figure 4f) up to six-membered species (Figure 4i), with few examples of larger rings.

Finding Known Gyrase Fragments. Boehm et al.[11] identified seven classes of active DNA-Gyrase fragments, which included phenols (similar to Figure 4a,e), indazole (Figure 4c), and pyrrolopyrimidines (Figure 4d). All seven of the fragment classes were found in the SkelGen output.

Finding Potentially Novel DNA-Gyrase Fragments. An assessment of novelty was carried out using a combination of Daylight tools to search the World Drug Alerts File.[26,27] Each fragment was converted into a SMART[28] query, while retaining the hydrogen-bonding pattern on the heteroatoms.[29] A subset of the WDA was created in Xvmerlin[27] using a combination of query strings for "DNA gyrase" and "antibacterial activity". This identified a subset of 66 known DNA gyrase inhibitors that was subsequently searched with the filtered SkelGen fragments using the Daylight toolkit program SMARTMATCH.[27] From this search, 34 fragments were identified in known DNA gyrase inhibitors, a selection of which can be found in Figure 5. Thus the remaining fragments—the majority of the SkelGen generated output—

Table 1. The Retrosynthetic Transforms Used to Break Down the Eight Selected Fragments Shown in Figure 6[a]

Fragment	Product	Reactant 1	Reactant 2	Transform	Penalty
3			NH_3	Amide Formation	0
4			NH_3	Amide Formation	0
5			-	Amide Formation	1
6			$N\equiv C-O^-$	Terminal Urea Formation	1
7			NH_3	Amide Formation	0
8			H_2O	Nucleophilic Aromatic Substitution	0
9				Benzimidazole Formation	0
10			NH_3	Nucleophilic Aromatic Substitution	0

[a] A penalty is added when protecting group chemistry is required.

were considered potential novel DNA gyrase fragments, examples of which are shown in Figure 4b,f,h,i.

Fragment Selection. In selecting a number of example fragments that could be potentially developed into lead-like compounds (using a further in silico design stage or just medicinal chemistry techniques alone), we considered a number of requirements. (1) The fragments chosen would require some synthesis, thus showing that SkelGen is not restricted to just generating purchasable species, (2) that chosen fragments do not exist as substructures within known DNA-gyrase compounds, thereby potentially providing a degree of novelty, and (3) that an appropriate, "chemically accessible" exit vector is available, allowing development of the fragment along the same vector as that taken by the crystallized ADPNP ligand in the binding site. 2D and 3D representations of eight examples are shown in Figure 6, which have been picked with the above criteria in mind and show some structural diversity. All the examples make the required hydrogen bonds to Asp-73 and water-1601, while there are a couple of structures that show extra hydrogen-bonding capabilities. Fragments **7** and **9** could respectively use the indole nitrogen or "imidazole" nitrogen as a hydrogen-bond donor toward the Tyr-109 side chain (not shown).

The exit vectors shown in Figure 6 are those closest to the vector of the crystallized ADPNP ligand, **1**. There may indeed be other positions of interest, but, for simplicity, these have not been highlighted.

Table 1 shows the simple, one-step reaction transforms used to break down each of the fragments into their constituent, commercially available reagents. Half of the selected structures had an amide transform applied, showing that they could be created from the reaction between a carboxylic acid and ammonia (or amine). This includes fragment **5**, where the transform is applied to break the ring, rather than breaking down the fragment into daughter molecules. If there is more than one point in the reactant molecule that is susceptible to the same reaction, a penalty of 1.0 is set for every position that may require some degree of protection. For example, fragment **6** undergoes the

terminal urea transform to separate the amide from the diazabicycloheptane ring. Starting from the reactants, it is possible to see that both chemically equivalent nitrogen atoms are susceptible to the reaction; this highlights the need to alter the stoichiometry of the reactants to encourage the formation of the desired product. However, the advantage of selecting fragments such as **5** and **6** for further study is that they can be readily substituted at their N-atom attachment points (as is also the case with fragment **4**).

CONCLUSION

SkelGen is a useful tool for designing appropriate fragments against a given target. The described in silico fragment generation method has advantages over current fragment screening methods in generating diverse, potentially novel chemotypes that could overcome IP restrictions. We have shown for DNA-gyrase, that SkelGen is able to generate and select known active fragments as part of the 3600 different chemotypes generated. We have found that 400 of the filtered fragments were available commercially, while a further 250 could be generated in a "one-step" synthesis. From the 250 fragments requiring chemistry, eight examples were selected based on potential novelty and providing an appropriate exit vector for further development. We have shown that SkelGen is able to generate a rich, diverse choice of fragments that can be 'grown' into lead-like ligands.

ACKNOWLEDGMENT

We thank William Harris and Henriëtte Willems for helpful discussions.

REFERENCES AND NOTES

(1) Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem.* **2004**, *47*, 3463−3482.
(2) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-Based Lead Discovery Using X-ray Crystallography. *J. Med. Chem.* **2005**, *48*, 403−413.
(3) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531−1534.
(4) Hajduk, P. J.; Sheppard, G.; Nettesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M., Jr.; Marcotte, P. A.; Severin, J.; Walter, K.; Smith, H.; Gubbins, E.; Simmer, R.; Holzman, T. F.; Morgan, D. W.; Davidsen, S. K.; Summers, J. B.; Fesik, S. W. Discovery of Potent Nonpeptide Inhibitors of Stromelysin Using SAR by NMR. *J. Am. Chem. Soc.* **1997**, *119*, 5818−5827.
(5) Hajduk, P. J.; Boyd, S.; Nettesheim, D.; Nienaber, V.; Severin, J.; Smith, R.; Davidson, D.; Rockway, T.; Fesik, S. W. Identification of Novel Inhibitors of Urokinase via NMR−Based Screening. *J. Med. Chem.* **2000**, *43*, 3862−3866.
(6) Hajduk, P. J.; Gomtsyan, A.; Didomenico, S.; Cowart, M.; Bayburt, E. K.; Solomon, L.; Severin, J.; Smith, R.; Walter, K.; Holzman, T. F.; Stewart, A.; McGaraughty, S.; Jarvis, M. F.; Kowaluk, E. A.; Fesik, S. W. Design of Adenosine Kinase Inhibitors from the NMR−Based Screening of Fragments. *J. Med. Chem.* **2000**, *43*, 4781−4786.
(7) . Hajduk, P. J.; Shuker, S. B.; Nettesheim, D. G.; Craig, R.; Augeri, D. J.; Betebenner, D.; Albert, D. H.; Guo, Y.; Meadows, R. P.; Xu, L.; Michaelides, M.; Davidsen, S. K.; Fesik, S. W. NMR-Based Modification of Matrix Metalloproteinase Inhibitors with Improved Bioavailability. *J. Med. Chem*. **2002**, *45*, 5628−5639.
(8) Deprez, P.; Mandine, E.; Gofflo, D.; Meunier, S.; Lesuisse, D. Small Ligands Interacting with the Phosphotyrosine Binding Pocket of the Src SH2 Protein. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1295−1298.
(9) Lesuisse, D.; Lange, G.; Deprez, P.; Bénard, D.; Schoot, B.; Delettre, G.; Marquette, J.-P.; Broto, P.; Jean-Baptiste, V.; Bichet, P.; Sarubbi, E.; Mandine, E. SAR and X-ray. A New Approach Combining Fragment-Based Screening and Rational Drug Design: Application to the Discovery of Nanomolar Inhibitors of Src SH2. *J. Med. Chem.* **2002**, *45*, 2379−2387.
(10) Reece, R. J.; Maxwell, A. DNA Gyrase: Structure and Function. *Crit. Rev. Biochem. Mol. Biol.* **1991**, *26*, 335−375.
(11) Boehm, H.-J.; Boehringer, M.; Bur, D.; Gmuender, H.; Huber, W.; Klaus, W.; Kostrewa, D.; Kuehne, H.; Luebbers, T.; Meunier-Keller, N.; Mueller, F. Novel Inhibitors of DNA Gyrase: 3D Structure Based Biased Needle Screening, Hit Validation by Biophysical Methods, and 3D Guided Optimization. A Promising Alternative to Random Screening. *J. Med. Chem.* **2000**, *43*, 2664−2674.
(12) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP−Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511−522.
(13) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(14) Bemis, G. W.; Murcko, M. A. Properties of Known Drugs. 2. Side Chains. *J. Med. Chem.* **1999**, *42*, 5095−5099.
(15) Clark, M. Generalized Fragment-Substructure Based Property Prediction Method. *J. Chem. Inf. Model.* **2005**, *45*, 30−38.
(16) Stahl, M.; Todorov, N. P.; James, T.; Mauser, H.; Boehm, H.-J.; Dean, P. M. A Validation Study on the Practical Use of Automated *de novo* Design. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 459−478.
(17) Todorov, N. P.; Dean, P. M. A Branch-and-Bound Method for Optimal Atom-Type Assignment in *de novo* Ligand Design. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 335−349.
(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) Kirkpatrick, S.; Gellatt, C. D., Jr.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671−680.
(20) Brino, L.; Urzhumtsev, A.; Mousli, M.; Bronner, C.; Mitschler, A.; Oudet, P.; Moras, D. Dimerization of *Escherichia coli* DNA-gyrase B Provides a Structural Mechanism for Activating the ATPase Catalytic Center. *J. Biol. Chem.* **2000**, *275*, 9468−9475.
(21) Garland, S. L. New Computational Methods for the Design of Peptidomimetics. Ph.D. Thesis, University of Cambridge, U.K., 1996.
(22) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes*, 2nd ed.; Cambridge University Press: 1992.
(23) Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(24) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead Selection. *Drug Discovery Today* **2004**, *9*, 430−431.
(25) Available Chemicals Directory is available from MDL Information Systems Inc., San Leandro, CA, 94577, and contains specialty bulk chemicals from commercial sources. Website: www.mdli.com.
(26) Derwent World Drug Alerts, Derwent Information, London, 2002.
(27) Xvmerlin a Daylight database system and SMARTMATCH a Daylight toolkit program, Daylight Chemical Information Systems Inc., http://www.daylight.com.
(28) SMARTS is substructure chemical searching language used to query chemical structures as described in the following: James, C. A.; Weininger, D.; Delany, J. Daylight Theory Manual; Daylight Chemical Information Systems Inc., April 2005; http://www.daylight.com/dayhtml/doc/theory/theory.toc.html.
(29) This procedure was automated using a modified version of the smiles2smarts tool kit program kindly provided by John Bradshaw, Daylight Chemical Information Systems.

CI050338I