Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 17 (2007) 5361–5365
Bioorganic & Medicinal Chemistry Letters

Virtual screening application of a model of full-length HIV-1 integrase complexed with viral DNA

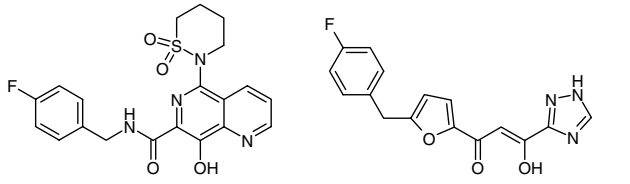
Chenzhong Liao,[a] Rajeshri G. Karki,[a,†] Christophe Marchand,[b] Yves Pommier[b] and Marc C. Nicklaus[a,*]

[a]*Laboratory of Medicinal Chemistry, Center for Cancer Research, National Cancer Institute, National Institutes of Health, DHHS, Frederick, MD 21702, USA*
[b]*Laboratory of Molecular Pharmacology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, DHHS, Bethesda, MD 20892, USA*

Received 16 April 2007; revised 3 August 2007; accepted 6 August 2007
Available online 11 August 2007

Abstract—To address the absence of experimental data on the full-length structure of HIV-1 integrase and the way it binds to viral and human DNA, we had previously [Karki, R. G.; Tang, Y.; Burke, T. R., Jr.; Nicklaus, M. C. *J. Comput. Aided Mol. Des.* **2004**, *18*, 739] constructed models of full-length HIV-1 integrase complexed with models of viral and human DNA. Here we describe the discovery of novel HIV-1 integrase strand transfer inhibitors based on one of these models. Virtual screening methods including docking and filtering by predicted ADME/Tox properties yielded several μM level inhibitors of the strand transfer reaction catalyzed by wild-type HIV-1 integrase.

The *pol* gene of the human immunodeficiency virus type 1 (HIV-1) encodes three essential enzymes, protease (PR), reverse transcriptase (RT), and integrase (IN). About two dozen anti-HIV drugs have been approved by the FDA,[1] however, they all target only two of these enzymes, PR and RT. Several factors including the emergence of multidrug-resistant HIV strains, drug toxicity, and patients' ability to comply with the prescribed therapy make it highly desirable to develop novel drugs that target other viral replication processes. IN is one such target. However, no approved drugs based on IN inhibition are currently available. Although a large number of integrase inhibitors have been published over the years,[2–6] only a handful have progressed to clinical trials, and further development of two promising candidates, the diketo acids or diketo bioisostere compounds L-870,810 (Merck) and S-1360 (Shionogi) (Fig. 1) was halted after unacceptable liver and kidney cell toxicity was found in dogs.[7] Still, there are very recent promising reports[8] that IN may finally join the other two viral enzymes as a target for which an anti-HIV drug with high efficiency and low adverse effects (Raltegravir, Fig. 1) is available. All this testifies to the fact that IN is still a both difficult and worthwhile target for the development of efficient anti-HIV drugs.



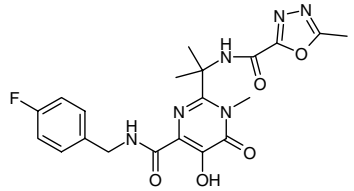


Figure 1. Structures of three strand transfer inhibitors of integrase that have progressed to clinical trials.

Keywords: HIV-1 integrase strand transfer inhibitors; Computer-aided drug design; Virtual screening; Drug discovery.
* Corresponding author. Tel.: +1 301 846 5903; fax: +1 301 846 6033; e-mail: mn1@helix.nih.gov
† Present address: Global Discovery Chemistry-CADD, Novartis Institutes for Biomedical Research, Inc., 3160/100 Technology Square, Cambridge, MA 02139, USA.

0960-894X/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.bmcl.2007.08.011

IN, a 32-kDa protein encoded at the 3′-end of the HIV *pol* gene, is essential for viral replication and catalyzes the insertion of the reverse-transcribed viral DNA into the host chromosome.[9,10] The integration reaction is carried out in two steps: 3′-end processing (3′-P) and strand transfer (ST). Key factors that make IN a difficult viral enzyme to target include[11,12]: (1) absence of a full-length experimental structure of the enzyme (crystal or NMR) and information on the way it binds to viral and human DNA (crystallization has been hampered mainly by poor solubility); (2) insufficient information concerning the biochemical mechanism of integration; and (3) lack of information regarding structural requirements for compounds to achieve IN inhibition in vitro while retaining antiviral activity in vivo.

To address the first point in a computational way, we have previously used the available experimental results to construct three different plausible models of full-length HIV-1 integrase complexed with models of viral and host DNA by employing molecular modeling techniques.[13] Among these, the model having two magnesium ions coordinating the three catalytic residues (Asp 64, Asp 116, and Glu 152) and best reproducing key protein–DNA interactions was found to fulfill best the requirements for binding of known diketo-acid inhibitors, and was therefore chosen for the present docking study. Although several such models[14–21] have been proposed for the interaction of viral DNA with a multimer of IN, to the authors' knowledge, this is the first model used in virtual screening to find inhibitors of the strand transfer reaction catalyzed by wild-type HIV-1 integrase.

Key features of this model that were important for our virtual screening efforts include: (1) it is a full-length tetramer complexed with the U3 and U5 ends of the viral DNA and human DNA (Fig. 2); (2) it was modeled to represent the state after 3′-P but before ST; (3) most importantly, this model is consistent with the reported biochemical findings regarding enzyme–DNA interactions.

Figure 3 shows the binding site of the model that was used for virtual screening (generated using SiteMap 2.0,[22] Schrödinger). This structure shows that the binding site includes not only residues of the protein (Asp 64, Cys 65, Thr 66, His 67, Glu 92, Asp 116, Asn 117, Gly 118, Asn 120, and Glu 152), but also several nucleotide residues of the viral DNA (mainly C17 and A18).

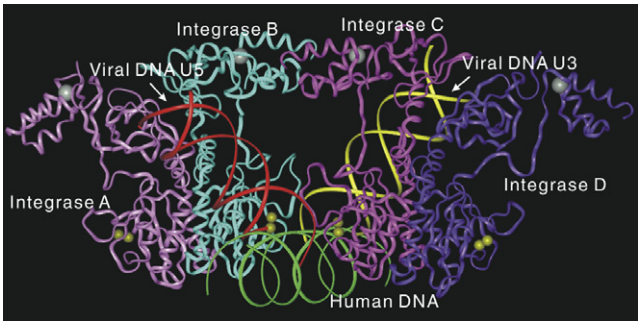


Figure 2. Structural model of the full-length integrase tetramer bound with models of viral DNA ends and human DNA[13]. Gray spheres: Zn^{2+} ions; yellow spheres: Mg^{2+} ions.

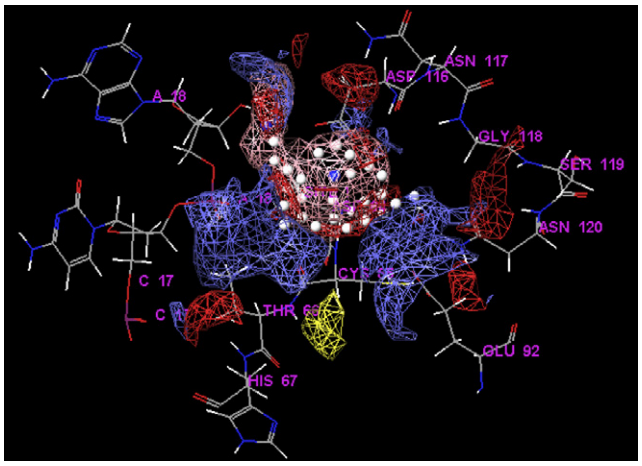


Figure 3. Site points (white spheres), hydrophobic (yellow mesh), hydrogen bond donor (blue mesh), hydrogen bond acceptor (red mesh), and metal binding maps (pink maps) for the model of full-length HIV-1 integrase complexed with viral DNA used for virtual screening.

We generated both ligand-based and structure-based pharmacophore models using the programs Catalyst 4.9[23] (Accelrys) (conformations for the ligands were generated using both Catalyst 4.9 and Macromodel 9.0[24] from Schrödinger) and Structure-Based Focusing (SBF) in Cerius^2[25] (Accelrys), respectively. To generate the ligand-based pharmacophore models, 50 different IN inhibitors with particular emphasis on selective ST inhibition were used. These 50 IN inhibitors are mostly 4-aryl-2-hydroxy-4-oxo-2-butenoic acids and their isosteric tetrazoles.[26,27] In total, we produced 30 different pharmacophore models. Figure 4 shows one of them as an example.

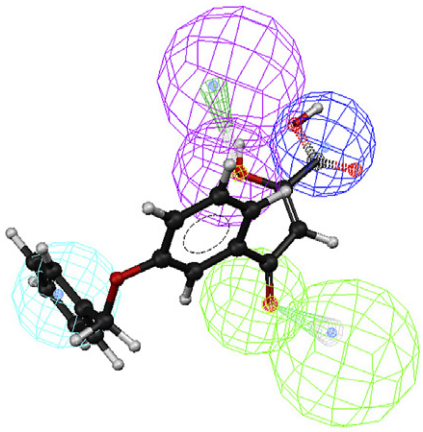


Figure 4. One of the pharmacophores used to search the ChemNavigator database. Green: hydrogen bond acceptor; magenta: hydrogen bond donor; blue: negatively charged group; cyan: hydrophobic aromatic group.

The first step of the virtual screening was a search across a pre-existing multi-conformation Catalyst database of the ChemNavigator[28] *iResearch Library* (*iRL*), which comprised 13.5 million purchasable samples at the time of the study. Each of the pharmacophore models generated was used as a query to search the *iRL* using the catSearch feature in Catalyst. A total of 234,894 compounds were identified as hits. Because most compounds in the training set used for developing the pharmacophores were acids (mainly diketo acids), all of the hits contained a carboxylic acid group.

From the above hit list, duplicate molecules, molecules violating Lipinski's 'Rule of Five', and inorganic compounds were deleted using a Pipeline Pilot 5.0[29] (SciTegic) protocol shown in Figure 5.

This resulted in 167,479 compounds. All these were docked into our integrase model using Glide versions 3.5[30] and 4.0[31] (Schrödinger) utilizing Standard Precision mode (a more powerful Extra Precision mode could not be used in this case because it is not optimized for metal-containing systems). When we had docked the diketo acid inhibitors in this model, we had found that the docking poses of their neutral states were able to explain their structure–activity relationship much better than their charged states. Therefore, all 167,479 compounds were docked in their neutral state. Details about how we used this model to dock small molecules have been described elsewhere.[13]

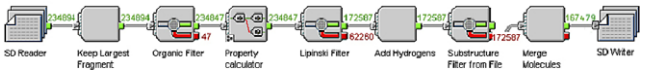


Figure 5. Pipeline Pilot protocol used to process the compounds obtained in the pharmacophore filtering step.

The 500 compounds with the top scores obtained with Glide 3.5 and the top 1000 compounds obtained from Glide 4.0 were analyzed visually. To pass this test as a potentially good inhibitor, a compound had to: (1) be docked into the active site; (2) form at least two coordination-type interactions with at least one Mg^{2+} metal ion; (3) form at least one hydrogen bond with the residues of Glu 92, Asp 116, Asn 117, Gly 118 or Asn 120 (and sometimes also His 67 and Cys 65).

Combining evaluation of previously computed ADME/Tox properties for the 1500 manually pre-screened compounds (calculated with both ADMET Predictor 1.2.4[32] (Simulations Plus) and QikProp 2.5[33] (Schrödinger)) with our evaluation of the docking poses, and taking into account sample availability, 88 compounds were finally selected for purchase.

Table 1. Integration inhibition for compounds **1**–**9**

Compound	G-Score (kJ/mol)[a]	IC_{50} BioVeris (μM)	IC_{50} 3′-P (μM)	IC_{50} ST (μM)
1	−10.44	37	>333	>333
2	−9.79	113	65	49
3	−9.66	780	155	120
4	−9.64	141	120	70
5	−9.86	121	87	57
6	−9.75	182	>333	>333
7	−9.69	292	204	114
8	−9.84	670	295	N/A[b]
9	−9.87	580	121	54

[a] Glide Standard Precision docking score.
[b] N/A, not determined.

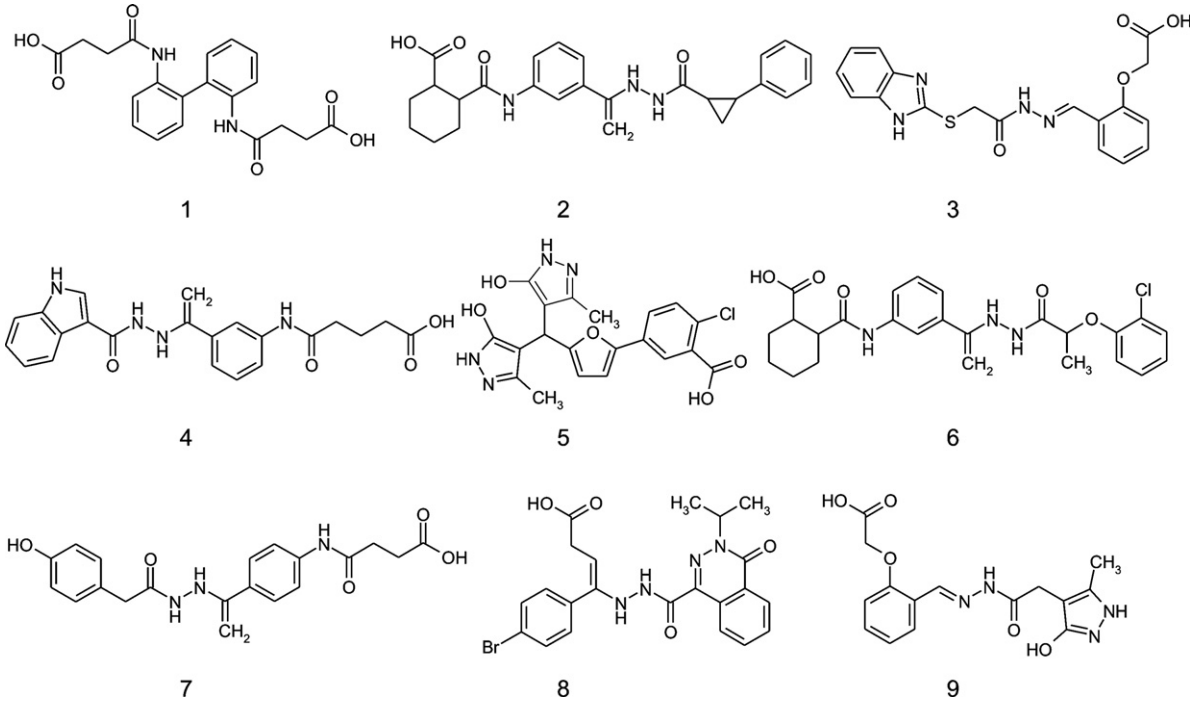


Figure 6. Structures of the nine active compounds.

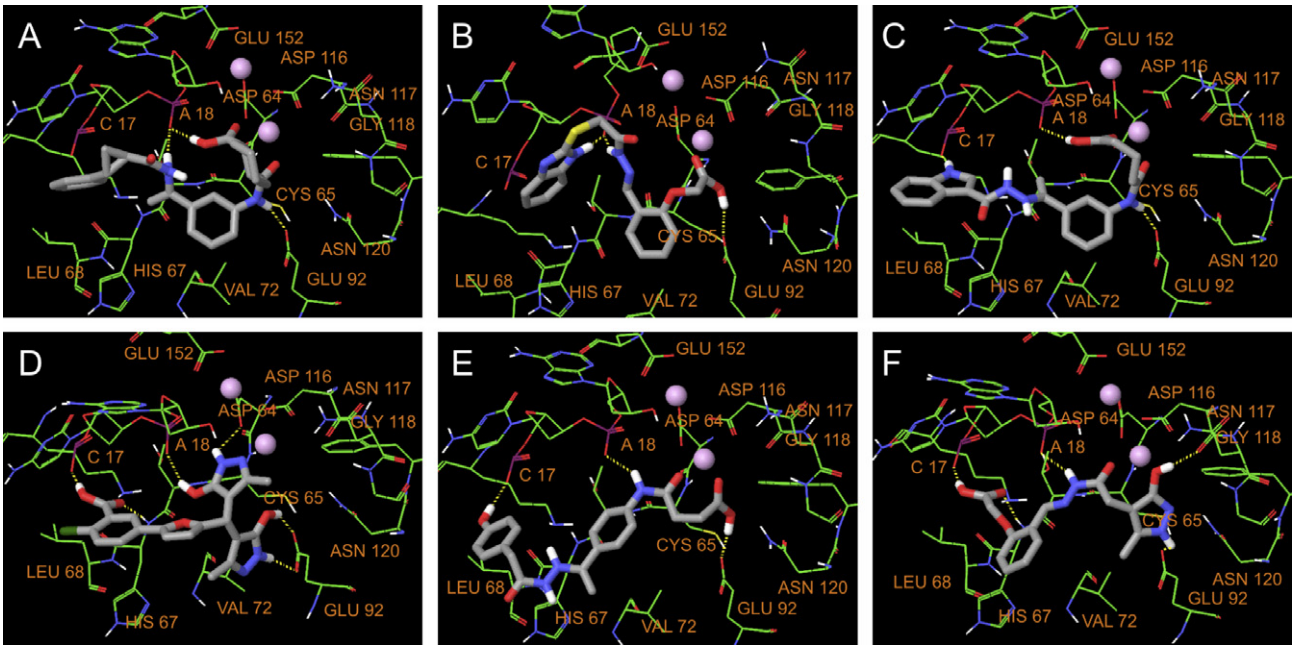


Figure 7. Putative binding modes of the six most active hits. (A)–(F) correspond to binding modes of compounds **2**–**5**, **7**, and **9**, respectively. Hydrogen bonds are represented by yellow dotted lines. For clarity, only polar hydrogen atoms are shown here.

The purchased compounds were tested for their abilities to inhibit HIV-1 integrase in vitro. We used an electrochemiluminescent high-throughput screening assay developed by BioVeris, Corp. (Gaithersburg, MD), which is based on the use of ruthenium labels. The 88-compound series was screened in the presence of Mg^{2+} in the BV™ ST assay. Positive hits were then confirmed in a gel-based assay in the presence of Mg^{2+} and IC_{50} values for 3′-P and ST were determined. Details about the experimental methods of the assays are given elsewhere.[4,34,35]

Among our 88-compound series, nine compounds showed inhibition of HIV-1 IN in the BV™ ST assay (Fig. 6). Six of them were confirmed active in the gel-based assay with IC_{50} values for ST between 49 and 120 μM (Table 1). The IC_{50} values for ST inhibition of these six hits are lower than the corresponding IC_{50} values for 3′-P inhibition, indicating that these compounds do have the desired selectivity over ST, although not at an overall high level of inhibition.

Figure 7 shows the putative binding modes of the most active six hits as they were produced by the docking runs. They all are somewhat similar to each other, in that all inhibitors chelate one magnesium ion, and form hydrogen bonds with residue Glu 92 in the protein and A18 in the viral DNA. Except compounds **2** and **3**, all compounds form hydrogen bonds with C17 in the viral DNA. This supports the assumption that viral DNA is quite likely crucially involved in the binding of ST selective IN inhibitors.

Although the biological activities of these nine ST inhibitors are not very high in absolute terms, they represent novel chemical structures which can be considered as starting points for further developments. We have shown that our IN models can be successfully employed in structure-based drug design toward integrase, an important target for antiviral drug discovery. Application of this model yielded several novel IN ST inhibitors out of a relatively small number of assayed compounds.

Acknowledgments

This research was supported in part by the Intramural Research Program of the NIH, NCI, in part by grant ISTC/BTEP #3197/111. We are grateful for helpful discussions with Prof. Vladimir Poroikov, members of his team, and additional collaborators in the above grant in Russia.

References and notes

1. Richman, D. D. *Nature* **2001**, *410*, 995.
2. Dayam, R.; Neamati, N. *Curr. Pharm. Des.* **2003**, *9*, 1789.
3. Johnson, A. A.; Marchand, C.; Pommier, Y. *Curr. Top. Med. Chem.* **2004**, *4*, 1059.
4. Pommier, Y.; Johnson, A. A.; Marchand, C. *Nat. Rev. Drug Discov.* **2005**, *4*, 236.
5. Neamati, N.; Marchand, C.; Pommier, Y. *Adv. Pharmacol.* **2000**, *49*, 147.
6. HIV-1 Integrase Inhibitors: Update and Perspectives. http://www.discover.nci.nih.gov/pommier/pommier.htm.
7. Rosemond, M. J.; St. John-Williams, L.; Yamaguchi, T.; Fujishita, T.; Walsh, J. S. *Chem. Biol. Interact.* **2004**, *147*, 129.
8. URL: http://www.retroconference.org/2007/data/files/webpage_for_CROI.htm.
9. Chiu, T. K.; Davies, D. R. *Curr. Top. Med. Chem.* **2004**, *4*, 965.
10. Craigie, R. *J. Biol. Chem.* **2001**, *276*, 23213.

11. Neamati, N. *Expert. Opin. Investig. Drugs* **2001**, *10*, 281.
12. Chen, I. J.; Neamati, N.; MacKerell, A. D., Jr. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 217.
13. Karki, R. G.; Tang, Y.; Burke, T. R., Jr.; Nicklaus, M. C. *J. Comput. Aided Mol. Des.* **2004**, *18*, 739.
14. Gao, K.; Butler, S. L.; Bushman, F. *EMBO J.* **2001**, *20*, 3565.
15. Chen, J. C.; Krucinski, J.; Miercke, L. J.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8233.
16. Wang, J. Y.; Ling, H.; Yang, W.; Craigie, R. *EMBO J.* **2001**, *20*, 7333.
17. Heuer, T. S.; Brown, P. O. *Biochemistry* **1998**, *37*, 6667.
18. Podtelezhnikov, A. A.; Gao, K.; Bushman, F. D.; McCammon, J. A. *Biopolymers* **2003**, *68*, 110.
19. De Luca, L.; Pedretti, A.; Vistoli, G.; Barreca, M. L.; Villa, L.; Monforte, P.; Chimirri, A. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 1083.
20. Wang, L. D.; Liu, C. L.; Chen, W. Z.; Wang, C. X. *Biochem. Biophys. Res. Commun.* **2005**, *337*, 313.
21. Adesokan, A. A.; Roberts, V. A.; Lee, K. W.; Lins, R. D.; Briggs, J. M. *J. Med. Chem.* **2004**, *47*, 821.
22. SiteMap, version 2.0, Schrödinger, LLC, New York, NY, 2005.
23. Accelrys Software Inc., Catalyst, Release 4.9, Accelrys Software Inc.: San Diego, 2003.
24. Macromodel 9.0, Schrödinger, LLC, New York, NY, **2005**.
25. Accelrys Software Inc., Cerius^2 Modeling Environment, Release 4.7, Accelrys Software Inc.: San Diego, 2003.
26. Pais, G. C.; Zhang, X.; Marchand, C.; Neamati, N.; Cowansage, K.; Svarovskaia, E. S.; Pathak, V. K.; Tang, Y.; Nicklaus, M.; Pommier, Y.; Burke, T. R., Jr. *J. Med. Chem.* **2002**, *45*, 3184.
27. Zhang, X.; Pais, G. C.; Svarovskaia, E. S.; Marchand, C.; Johnson, A. A.; Karki, R. G.; Nicklaus, M. C.; Pathak, V. K.; Pommier, Y.; Burke, T. R. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1215.
28. URL: http://www.chemnavigator.com/.
29. Pipeline Pilot, version 5.0, SciTegic, Inc.: San Diego, CA, 2005.
30. Glide, version 3.5, Schrödinger, LLC, New York, NY, 2005.
31. Glide, version 4.0, Schrödinger, LLC, New York, NY, 2005.
32. ADMET Predictor, version 1.2.4, Simulations Plus, Inc.: Lancaster, CA, 2005.
33. QikProp, version 2.5, Schrödinger, LLC, New York, NY, 2005.
34. Semenova, E. A.; Johnson, A. A.; Marchand, C.; Davis, D. A.; Yarchoan, R.; Pommier, Y. *Mol. Pharmacol.* **2006**, *69*, 1454.
35. Di Santo, R.; Costi, R.; Roux, A.; Artico, M.; Lavecchia, A.; Marinelli, L.; Novellino, E.; Palmisano, L.; Andreotti, M.; Amici, R.; Galluzzo, C. M.; Nencioni, L.; Palamara, A. T.; Pommier, Y.; Marchand, C. *J. Med. Chem.* **2006**, *49*, 1939.