

Design and Synthesis of DDR1 Inhibitors with a Desired Pharmacophore Using Deep Generative Models

Atsushi Yoshimori,^{*[a]} Yasunobu Asawa,^[a] Enzo Kawasaki,^[b] Tomohiko Tasaka,^[c] Seiji Matsuda,^[d] Toru Sekikawa,^[d] Satoshi Tanabe,^[d] Masahiro Neya,^[d] Hideaki Natsugari,^[c] and Chisato Kanai^[b]

Discoidin domain receptor 1 (DDR1) inhibitors with a desired pharmacophore were designed using deep generative models (DGMs). DDR1 is a receptor tyrosine kinase activated by matrix collagens and implicated in diseases such as cancer, fibrosis and hypoxia. Herein we describe the synthesis and inhibitory activity of compounds generated from DGMs. Three compounds were found to have sub-micromolar inhibitory activity. The most potent of which, compound **3** (*N*-(4-chloro-3-((pyridin-3-yloxy)methyl)phenyl)-3-(trifluoromethyl)benzamide), had an IC₅₀ value of 92.5 nM. Furthermore, these compounds were predicted to interact with DDR1, which have a desired pharmacophore derived from a known DDR1 inhibitor. The results of synthesis and experiments indicated that our *de novo* design strategy is practical for hit identification and scaffold hopping.

Deep generative models (DGMs) have been successfully applied to image generation,^[1] language translation,^[2] and others.^[3] In recent years, chemical structure generation using DGMs is receiving a lot of attention in *de novo* drug design.^[4] Successful examples of hit identification with DGMs have been reported by several groups.^[5,6] Although DGMs are able to generate molecules with desired properties, most of the properties don't have 3D information, such as shape and pharmacophore.^[4] It has been known that properties originating from 3D shape

and/or pharmacophore are very useful in the drug design process.^[7] Thus, we constructed DGMs for generating molecules with a desired pharmacophore. The procedure for DGM construction has been published elsewhere.^[8] Briefly, the method has three steps consisting of prior network construction, agent network construction, and structure sampling (Figure 1). First, the prior network is trained using SMILES strings^[9] from ChEMBL.^[10] After the training, the prior network generates valid SMILES strings. Next, the agent network is trained using reinforcement learning. The training shifts the probability distribution from that of the prior network towards a distribution modulated by a pharmacophore score. In the third step, the trained agent network generates SMILES strings, which are likely to have a desired pharmacophore. The prior and agent network constructions were implemented using REINVENT^[11] and the pharmacophore score was calculated using LigandScout 4.4.^[12]

Discoidin domain receptor 1 (DDR1) is a collagen-activated receptor tyrosine kinase and a potential therapeutic target for a wide range of human diseases, such as cancer,^[13] fibrosis^[14] and hypoxia.^[15]

To construct a desired pharmacophore for DDR1, the crystal structure of DDR1 kinase domain in complex with ponatinib (PDB: 3ZOS) is used (Figure 2A).^[16] Here, eight pharmacophore features of the inhibitor were identified as three aromatic (Ar1, Ar2, Ar3), two hydrophobic (Hy1, Hy2), two hydrogen acceptor (HA1, HA2) and one hydrogen donor (HD1) features (Figure 2B). In addition, an ensemble of exclusion volume spheres obtained from the crystal structure was used. For pharmacophore scoring, the scoring function was set to 'Relative Pharmacophore-Fit'; maximum number of omitted features is set to '1'

[a] Dr. A. Yoshimori, Y. Asawa
Institute for Theoretical Medicine, Inc.
26-1, Muraoka-Higashi 2-chome, Fujisawa
Kanagawa 251-0012 (Japan)
E-mail: yoshimori@itmol.com

[b] E. Kawasaki, Dr. C. Kanai
INTAGE Healthcare, Inc.
79, Kankoboko-cho, Shimogyo-ku
Kyoto, 600-8009 (Japan)

[c] T. Tasaka, Dr. H. Natsugari
Affinity Science Corporation
Aios Gotanda-ekimae, 1-11-1 Nishi-Gotanda, Shinagawa
Tokyo 141-0031 (Japan)

[d] S. Matsuda, Dr. T. Sekikawa, S. Tanabe, Dr. M. Neya
KNC Laboratories Co., Ltd.
7-1-19, Minatojiminamimachi, Chuo-ku, Kobe
Hyogo, 650-0047 (Japan)

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cmdc.202000786>

© 2020 The Authors. ChemMedChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

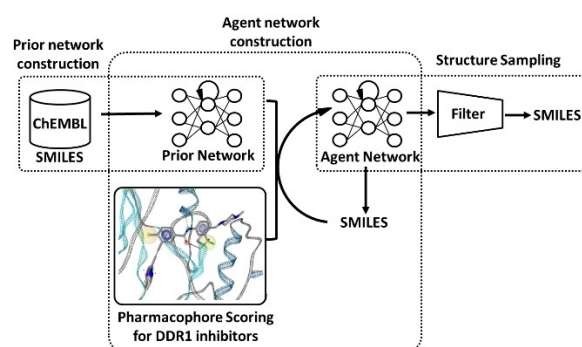


Figure 1. Workflow of our strategy for identification of DDR1 inhibitors.

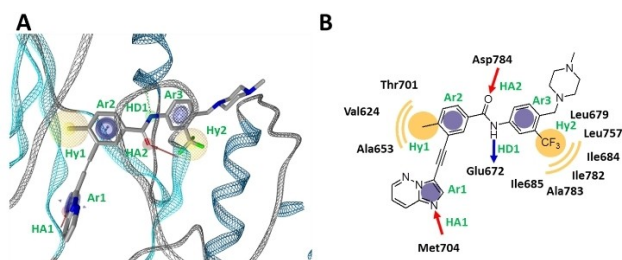


Figure 2. A) Pharmacophore model for DDR1 inhibitors. B) 2D depiction of the pharmacophore model for DDR1 inhibitors.

and the Hy1, Hy2, Ar3 and HD1 features were set as 'optional feature'. The Relative Pharmacophore-Fit ($rel.S_{FCR}$) was defined in Equation (1).^[8,12]

$$rel.S_{FCR} = S_{FCR} / (10 \times (N_{MFP} + 1)) \quad (1)$$

$$S_{FCR} = 10 \times N_{MFP} + S_{RMS}$$

$$S_{RMS} = 9 - 3 \times \text{Min}(RMS_{FP}, 3)$$

where S_{FCR} is the feature count/RMS distance score, N_{MFP} is the number of geometrically matched feature pairs, S_{RMS} is the matched feature pair RMS distance score in the range [0,9], and RMS_{FP} is the RMS of the matched feature pair distances. The parameter of conformer generation of generated structures from agent network was set as 'iCon Fast' option for idbgen tool^[17] provided with LigandScout 4.4.

The agent network was trained based on the desired pharmacophore and prior network. Training of the agent network was done with a batch size of 64 using the Adam optimizer for 10,000 steps. All other parameters were set to default values in REINVENT. After the training, structure sampling was performed. 570,542 valid SMILES were generated during the structure sampling of 640,000 SMILES strings. In the same way, 588,240 valid SMILES were generated from the prior network. The pharmacophore scores of the generated structures were calculated using LigandScout 4.4. The distribution of pharmacophore scores is displayed as a histogram in Figure 3. The 137,790 structures having pharmacophore scores ≥ 0.5

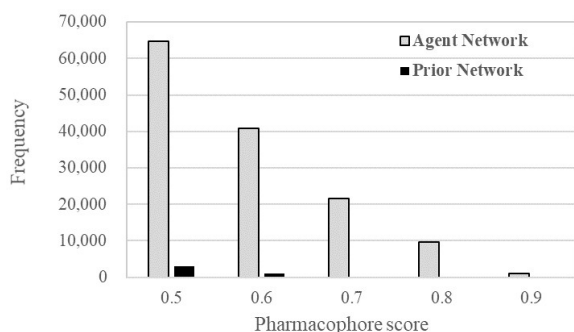


Figure 3. Distribution of Pharmacophore scores generated structures from the Agent and Prior network. Axis labels of pharmacophore score < 0.5 were omitted because all of hit compounds have pharmacophore score ≥ 0.5 .

were confirmed among the valid SMILES strings from the agent network. On the other hand, there are only 4,306 structures having pharmacophore scores ≥ 0.5 among the valid SMILES strings from the prior network. This result indicates that the agent network can generate structures fulfilling the desired pharmacophore of DDR1 inhibitor with high frequency. In order to perform filtering of the generated structures from agent network which have pharmacophore scores greater than 0.8 (10,694 structures), binding affinity scores were calculated using iaffinity module implemented in LigandScout 4.4. Compounds having binding affinity scores less than -37 kJ/mol were selected (4731 compounds). The selected compounds were inspected visually to determine which compounds to synthesize, taking into account their pharmacophore scores, binding scores and synthesis accessibilities. Consequently, 9 compounds (1–9) were selected which are illustrated in Figure 4. During the visual inspection, two compounds (7, 8) were modified by removing halogen atoms, that do not contribute much to binding interaction with DDR1. Compound 9 was modified by changing the position of pyridinyl nitrogen from para to meta to form hydrogen bond with the hinge region of DDR1. Accordingly, we synthesized 9 compounds (1–6 and 7a, 8a, 9a). Synthesis of the nine compounds is summarized in Schemes 1–9 (see Supporting Information).

The synthesized compounds were evaluated for their inhibitory activity against DDR1. The kinase assays were performed using Off-chip Mobility Shift Assay which were carried out via a kinase profiling service (Carna Biosciences, Inc., Kobe, Japan) (see Supporting Information). The results are summarized in Table 1. Among the tested compounds, compound 3 exhibited interesting double-digit nanomolar inhibitory activity against DDR1 ($IC_{50} = 92.5$ nM). The binding interaction of compound 3 derived from pharmacophore matching is shown in Figure 5A. Compound 3 fulfills all of the pharmacophore features of a DDR1 inhibitor, although there are slightly misaligned features.

Compounds 4 and 7a were also found to have potent inhibitory activities (IC_{50} values: 186.7 and 171.3 nM, respec-

Table 1. DDR1 inhibitory activity of the synthesized compounds.

Compound	Pharmacophore score ^[a]	Binding affinity score [kJ/mol] ^[b]	IC_{50} [nM] ^[c]
1	0.96	-50.51	1005.9
2	0.95	-52.67	2239.4
3	0.83	-47.13	92.5
4	0.96	-51.97	186.7
5	0.86	-37.29	> 30,000
6	0.84	-40.83	> 30,000
7	0.85	-54.38	NT ^[d]
8	0.85	-51.78	NT
9	0.85	-49.46	NT
7a	0.85	-53.06	171.3
8a	0.85	-51.15	1244.3
9a	0.85	-54.83	1111.0

[a] Calculated using Relative Pharmacophore-Fit score in LigandScout 4.4. [b] Calculated using iaffinity module in LigandScout 4.4. [c] The compound concentration required for 50% inhibition (IC_{50}) was determined from semi-logarithmic dose-response plots, and the results represent the mean of duplicated samples. [d] NT = not tested.

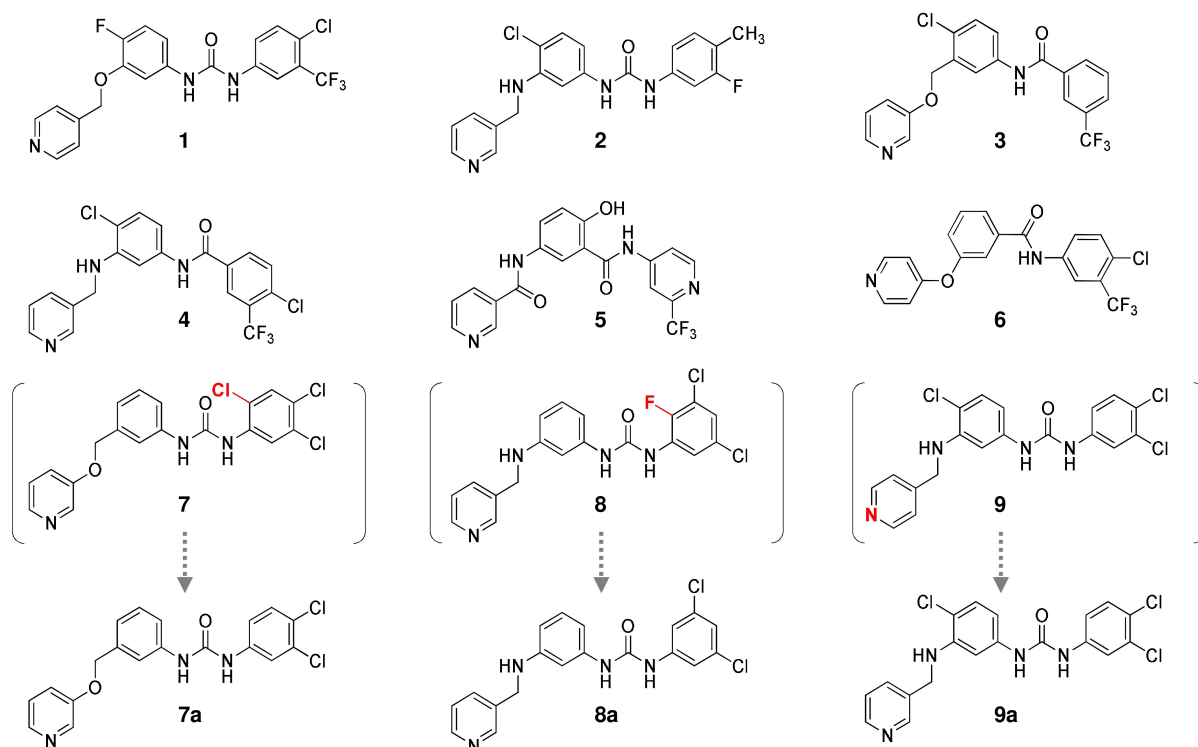


Figure 4. Synthesized compounds evaluated as DDR1 inhibitors.

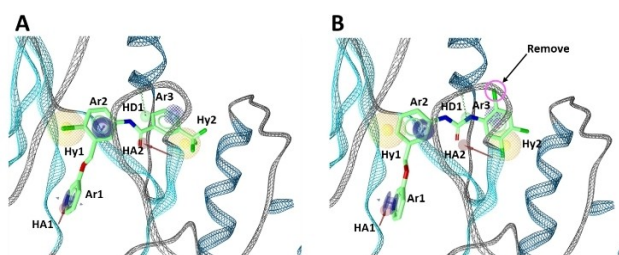


Figure 5. Binding interactions of A) compound 3 and B) compound 7.

tively). Compound 7a was designed by removing a Cl atom from 1,2,4-trichlorobenzene in compound 7 as illustrated in Figure 5B. Compounds 1, 2, 8a, 9a were found to moderately inhibit DDR1 activity (IC_{50} values: 1005.9, 2239.4, 1244.3 and 1111.0 nM, respectively). Compounds 5 and 6 did not exhibit any inhibitory activity at all. Binding affinity scores of the two compounds are high values (-37.29 , and -40.83 kJ/mol), indicating low binding affinities. These results indicate that our strategy of using DGMs has worked efficiently to design DDR1 inhibitors.

To check if the generated structures (compound 1–6, 7a–9a) have already been registered in certain databases, structure search was performed in ChEMBL^[10] and PubChem.^[18] We have found that compound 3 is registered in PubChem with CID 58614959 and is annotated as Raf kinase and p38 MAP kinase inhibitor.

In conclusion, we were able to design DDR1 inhibitors with a desired pharmacophore using DGMs. Compound 3 showed

potent inhibitory activity with an IC_{50} value of 92.5 nM against DDR1. In general, in order to predict inhibitory activities of generated compounds from DGMs, many experimental inhibitory data are needed to construct accurate prediction models. However, our strategy needs only pharmacophore information to design inhibitors against a target protein. Therefore, our strategy can be used in the early stage of drug discovery process. Ponatinib is a drug used to treat chronic myeloid leukemia and inhibits DDR1 with a K_d value of 1.3 nM.^[16] In this study, our pharmacophore is derived from the crystal structure of DDR1 kinase domain in complex with ponatinib. The scaffolds of the synthesized compounds (Figure 4) were found to be different from that of ponatinib (Figure 2). Thus, our strategy can also be used for scaffold hopping.

In order to determine which compounds to synthesize, it is important to filter generated structures from the agent network efficiently. We are now trying to construct more practical filtering methods that include criteria such as drug-likeness score,^[19] ADMET properties^[20] and synthesis accessibility.^[21] We believe that this pharmacophore-based DGM strategy can be applied to various drug discovery campaigns in the future.

Acknowledgements

We thank Dr. Hirofumi Nakano for his insightful comments and suggestions.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: deep generative model · pharmacophore model · *de novo* design · DDR1

- [1] H. Huang, P. S. Yu, C. Wang, arXiv:1803.04469v2 **2018**.
- [2] M. Johnson, M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F. Viégas, M. Wattenberg, G. Corrado, M. Hughes, J. Dean, *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 339–351.
- [3] R. de Bem, A. Ghosh, T. Ajanthan, O. Miksik, A. Boukhayma, N. Siddharth, P. Torr, *Int. J. Comput. Vis.* **2020**, *128*, 1537–1563.
- [4] D. C. Elton, Z. Boukouvalas, M. D. Fugea, P. W. Chung, *Mol. Syst. Des. Eng.* **2019**, *4*, 828–849.
- [5] D. Merk, L. Friedrich, F. Grisoni, G. Schneider, *Mol. Inf.* **2018**, *37*, 1700153.
- [6] A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev, Y. Volkov, A. Zholus, R. R. Shayakhmetov, A. Zhebrak, L. I. Minaeva, B. A. Zagribelnyy, L. H. Lee, R. Soll, D. Madge, L. Xing, T. Guo, A. Aspuru-Guzik, *Nat. Biotechnol.* **2019**, *37*, 1038–1040.
- [7] Q. Gao, L. Yang, Y. Zhu, *Curr. Comput.-Aided Drug Des.* **2010**, *6*, 37–49.
- [8] A. Yoshimori, E. Kawasaki, C. Kanai, T. Tasaka, *Chem. Pharm. Bull.* **2020**, *68*, 227–233.
- [9] D. J. Weininger, *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- [10] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, J. P. Overington, *Nucleic Acids Res.* **2014**, *42*, D1083–D1090.
- [11] a) REINVENT: Molecular *de novo* design using recurrent neural networks and reinforcement learning <https://github.com/marcusolivecrona/reinvent>; b) M. Olivecrona, T. Blaschke, O. Engkvist, H. Chen, *J. Cheminf.* **2017**, *9*, 1.
- [12] a) G. Wolber, T. Langer, *J. Chem. Inf. Model.* **2005**, *45*, 160–169; b) G. Wolber, A. A. Dornhofer, T. Langer, *J. Comput.-Aided Mol. Des.* **2007**, *20*, 773–788.
- [13] J. Quan, T. Yahata, S. Adachi, K. Yoshihara, K. Tanaka, *Int. J. Mol. Sci.* **2011**, *12*, 971–982.
- [14] S. Moll, A. Desmoulière, M. J. Moeller, J. Pache, L. Badi, F. Arcadu, H. Richter, A. Satz, S. Uhles, A. Cavalli, F. Drawnel, L. Scapozza, M. Prunotto, *BBA Mol. Cell Res.* **2019**, *1866*, 118474.
- [15] S. Li, Z. Zhang, J. Xue, X. Guo, S. Liang, A. Liu, *Med. Sci. Monit.* **2015**, *21*, 2433–2438.
- [16] P. Canning, L. Tan, K. Chu, S. W. Lee, N. S. Gray, A. N. Bullock, *J. Mol. Biol.* **2014**, *426*, 2457–2470.
- [17] G. Poli, T. Seidel, T. Langer, *Front. Chem.* **2018**, *6*, 229.
- [18] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton, *Nucleic Acids Res.* **2019**, *47*, D1102–D1109.
- [19] G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins, *Nat. Chem.* **2012**, *4*, 90–98.
- [20] J. Wenzel, H. Matter, F. Schmidt, *J. Chem. Inf. Model.* **2019**, *59*, 1253–1268.
- [21] P. Ertl, A. Schuffenhauer, *J. Cheminf.* **2009**, *1*, 8.

Manuscript received: October 6, 2020

Revised manuscript received: November 27, 2020

Accepted manuscript online: December 1, 2020

Version of record online: January 15, 2021