Electronic Supplementary Material (ESI) for Medicinal Chemistry Communications

Electronic supplementary Information

Identification of novel inhibitors of p53-MDM2 interaction facilitated by pharmacophore-based virtual screen combining molecular docking strategy

Weisi Wang[a], Xiaolei Zhu[a], Xueqin Hong[a], Lin Zheng[b], Hong Zhu[b], Yongzhou Hu[a]*

[a]ZJU-ENS Joint Laboratory of Medicinal Chemistry, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China
[b]Institute of Pharmacology and Toxicology, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China.

*To whom the correspondence should be addressed:
Yongzhou Hu, Prof.
E-mail: huyz@zju.edu.cn
Tel/Fax: 86 571 8820 8460.
ZJU-ENS Joint Laboratory of Medicinal Chemistry
College of Pharmaceutical Sciences
Zhejiang University
Hangzhou 310058, China

Contents

PHARMACOPHORE ESTABLISHMENT AND VALIDATION

Materials and methods

For the pharmacophore modeling studies, a set of 98 compounds were selected from literatures and patents in view of their chemical structural diversity and certain coverage of activity range.[1-28] Their MDM2 binding affinity data spanned over 6 orders of magnitude (from 0.002μM to 480μM). 2D chemical structures of 29 compounds in training set are shown in Fig. S1. The remaining 69 compounds with maximal 3D diversity and continuous bioactivity magnitude constituted a test set, as shown in Fig. S2.

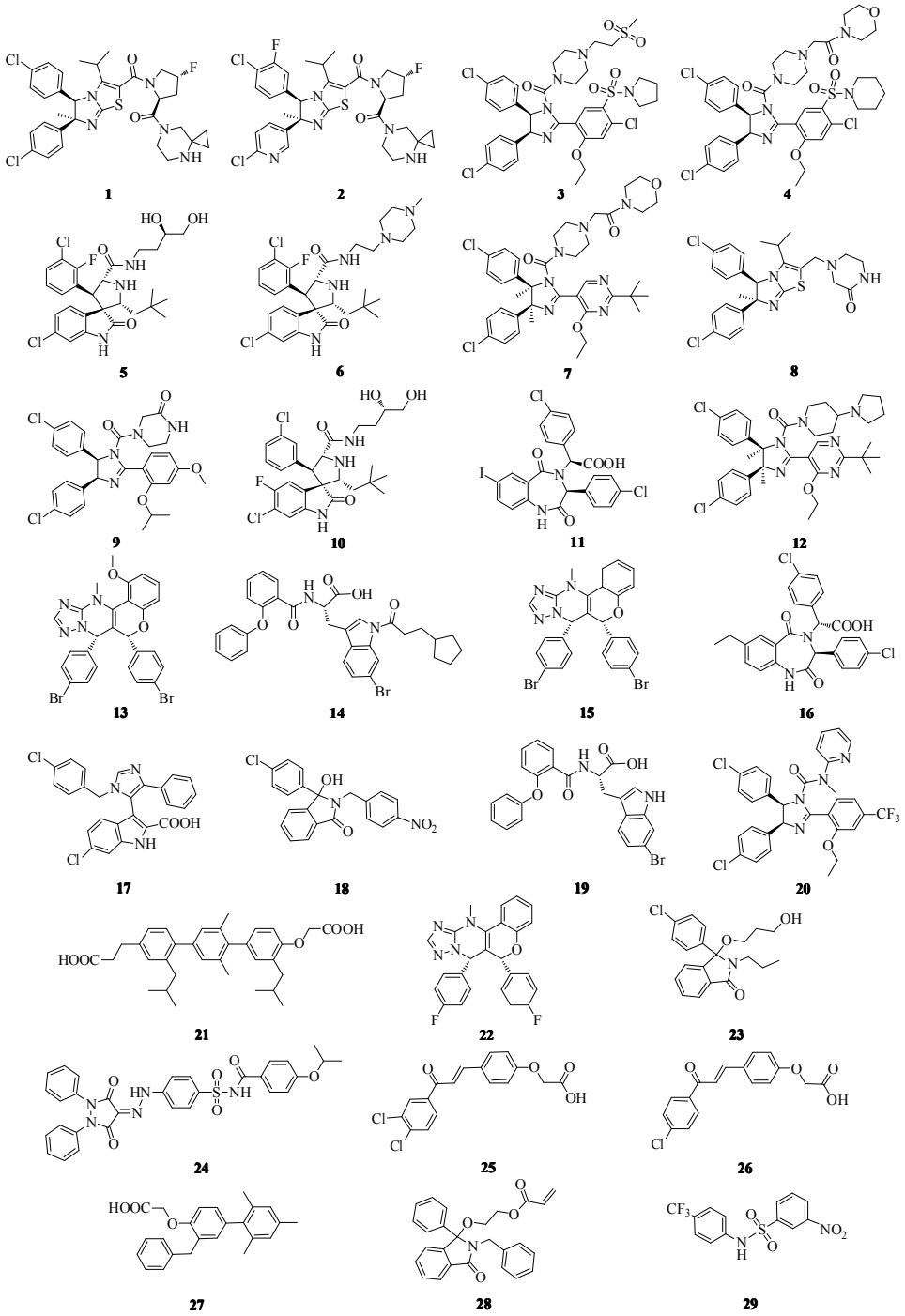

Figure S1. Chemical structures of 29 molecules in training set.

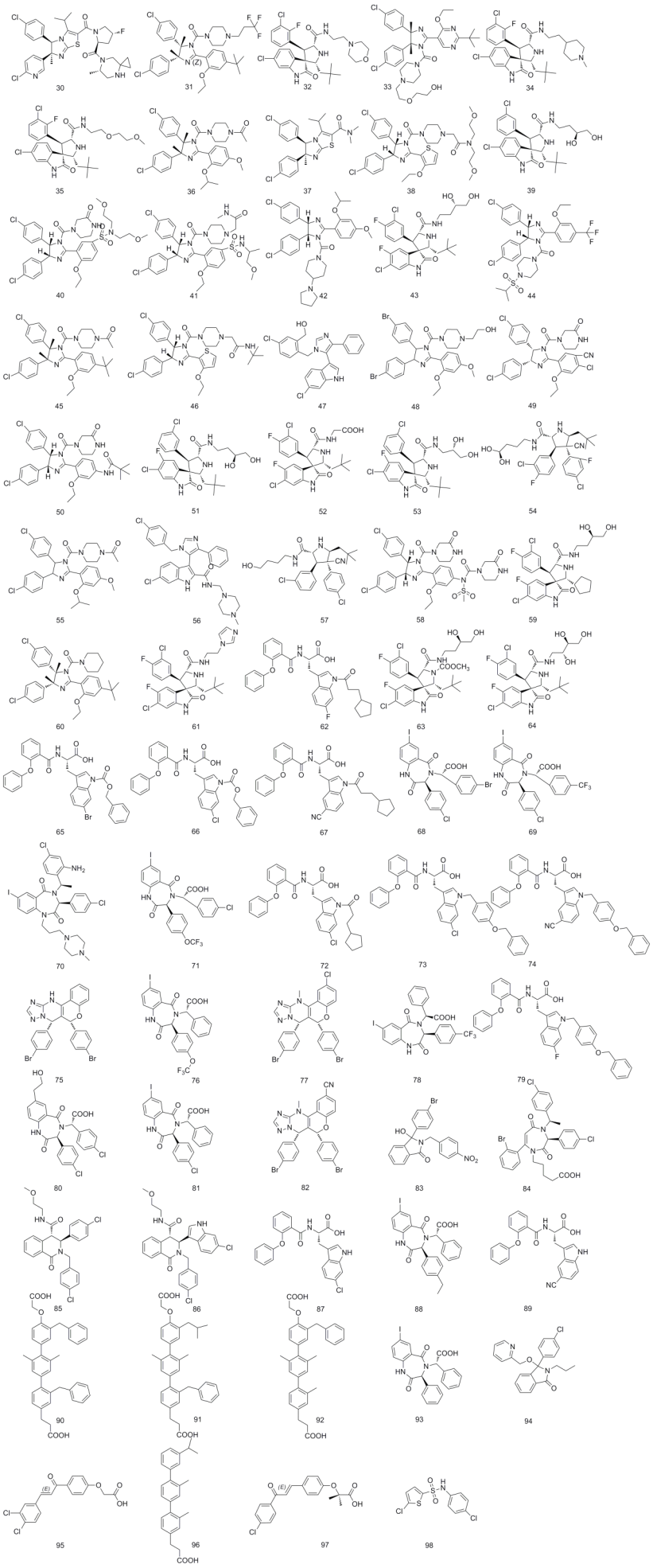

Figure S2. Chemical structures of 69 molecules in test set.

For the program setting, all training and test set conformations were generated using the CHARMm-like force field.[29] The 'Best Quality conformer generation' option was used and the maximum number of conformers for each molecule was 250. The uncertainty value was set to 2.0. Default settings were used for other parameters. The molecules associated with their conformational models were then submitted to catalyst hypothesis generation. Catalyst software package (version 4.11, Accelrys Inc., San Diego, CA) on a Silicon GraphicImage Origin 3800 workstation was used in this study to generate all pharmacophore models.

Generation and validation of pharmacophore hypotheses

The top 10 hypotheses generated are presented in Table S1 together with their statistical parameters. Fixed cost and null cost represent the cost of two theoretical hypotheses, one in which the error cost is minimal (all compounds fall along a line of slope = 1), and one in which the error cost is high (all compounds fall along a line of slope = 0). The two models can be considered the upper and lower bounds for the training set. In general, if a returned hypothesis has a cost that differs from the null cost by 40-60 bits, there is a high probability it has a 75-90% chance of representing a true correlation in the data. And if this difference is greater than 60 bits, there is an excellent chance (> 90%) the model represents a true correlation. RMS and Correlations are measures of the regression derived by Catalyst in fitting the estimated activity to the measured activity. The RMS factor represents the deviation of the log (estimated activities) from the log (measured activities) normalized by the log (uncertainties). Correl factor is the linear regression derived from the geometric fit index. The Configuration cost is a measure of the magnitude of the hypothesis space for a given training set. If the Config value exceeds 18, there are more degrees of freedom in training set than Catalyst can properly deal with and the hypothesis results may not be useful. Among the 10 hypotheses, Hypo_1 is characterized by the highest cost difference of 313.36, the lowest RMSD of 2.044, and the highest correlation coefficient of 0.920. Noticeably, the cost range between Hypo_1 and the fixed cost is 62.83, while that between the null hypothesis and Hypo_1 is 313.36, showing that the model has >90% probability of representing a true correlation in the data.[30] The configuration cost for the Hypo_1 is 17.024. Fig. S3A shows the topological features of Hypo_1. The experimental and estimated p53-MDM2 binding inhibitory activities (IC_{50} values) derived from the Hypo_1 hypothesis for the training set molecules are listed in Table S2 and the plots of predicted versus experimental values are recorded in Fig. S3B. The superimposition of Hypo_1 with Nutlin-3[9] (a representative inhibitor with high potency, Fig. S3A) shows that Nutlin-3 shares all of the features of Hypo_1 quite well (predicted IC_{50}=390nM, actual IC_{50}=90nM). While in the case of **26**[25], the HY1 feature cannot fit well to the corresponding pharmacophore feature while the HY2 and RA1 features are totally missed (Fig. S3A), leading to a poor predicted activity but close to its actual activity (predicted IC_{50}=140,000nM, actual IC_{50}=206,000nM).

Table S1. Information of statistical significance and predictive power presented in cost values measured for top 10 hypotheses as a result of HypoGen generation process.

Hypo. No.	Total cost	Δ cost	RMSD	Correl.
1	165.16	313.36	2.044	0.920
2	168.19	310.33	2.123	0.913
3	180.14	298.39	2.311	0.896
4	184.31	294.22	2.372	0.891
5	186.91	291.61	2.415	0.886
6	187.19	291.33	2.408	0.887
7	187.37	291.16	2.420	0.886
8	188.06	290.47	2.429	0.885
9	188.85	289.68	2.433	0.885
10	189.54	288.99	2.401	0.887

Null cost of 10 top-scored hypotheses is 478.53, fixed cost value is 102.33 and configuration cost is 17.02.

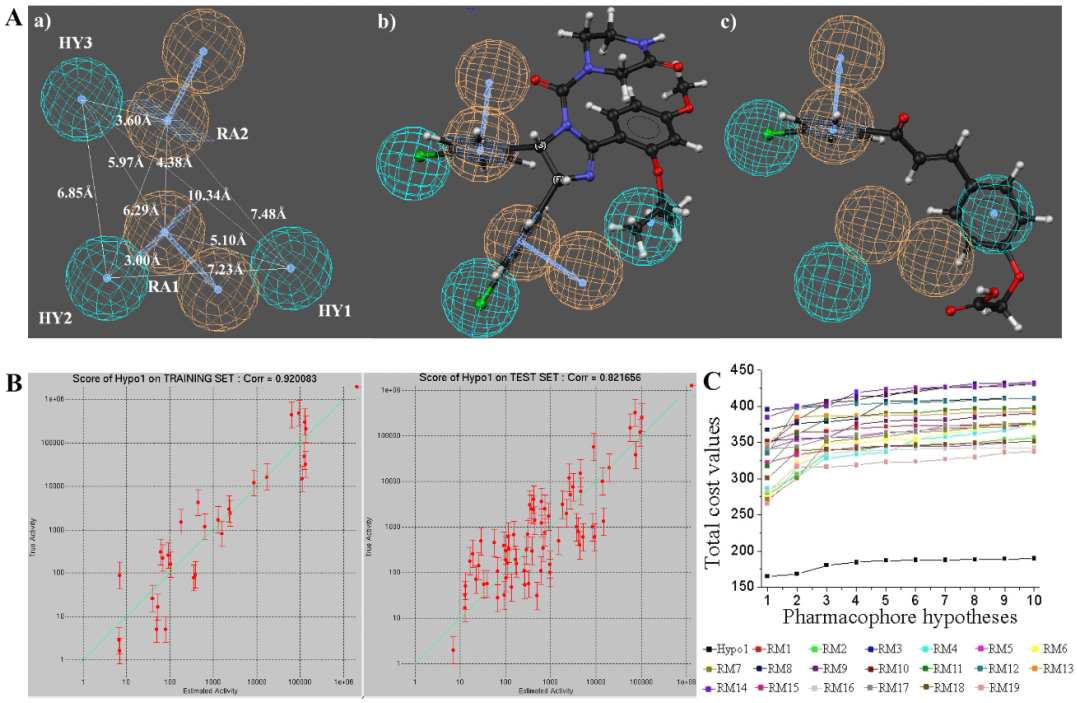

Figure. S3. A) Hypo_1 and its alignment to representative compounds. a) The topological features of Hypo_1: five pharmacophoric features, including two ring aromatic (RA, orange) and three hydrophobic group (HY, cyan). (b) Hypo_1 aligned to Nutlin-3. (c) Hypo_1 aligned to compound **26.** (B) The predictive ability of Hypo_1. a) The regression of actual versus predicted activities by the Hypo_1 for the training set. b) The regression of actual versus predicted activities by the Hypo_1 for the test set. (C) The difference in costs between the HypoGen runs (Hypo_1) and the scrambled runs (RM1 w RM19).

A total of 69 compounds with distinct chemical structures and diverse activity classes were

assembled as a test set (Table S2). As shown in Fig. S3B, there is a good line correlation between the experimental and estimated MDM2 binding affinities of the test set compounds with a correlation coefficient of 0.822, indicating that Hypo1 has a good predictive ability.

The validation of the Hypo_1 hypothesis was further demonstrated by Cat-Scramble test based on Fischer's randomization test.[31] For the Cat-Scramble program, the experimental activities of compounds in training set were scrambled randomly, and the resulting data set were used for HypoGen generation. All parameters were kept at the original HypoGen calculation. A total of 19 random spreadsheets were created to generate pharmacophore hypotheses which fitted a confidence level of 95%. The total costs of Hypo_1 and 19 randomized runs are shown in Figure S3C. None of the outcome hypotheses has a lower cost score than the original hypothesis (Hypo_1), suggesting that there is a strong correlation between chemical structures and biological activities for Hypo_1 and it is not generated by chance.

Another commonly used method for pharmacophore model validation is Enrichment Factor (EF) at a given percentage.[32] EF is defined as: $EF = \frac{TP}{TP+FP}\frac{N}{n}$ where TP is the number of true positives, FP is the number of false positives, N is the number of the total compounds and n is the number of the total positive compounds. 1981 molecules retrieved randomly from Available Chemicals Directory (ACD) database (negative compounds) and 19 inhibitors of p53-MDM2 (positive compounds) composed a test database. The EF of the test database screened by Hypo_1 at 2%, 5% and 10% are 66.7, 18.8 and 8.6, respectively. The results of the Enrichment Factor study suggested that Hypo_1 would be a valuable and reliable tool for identifying novel inhibitors of p53-MDM2 interaction in virtual screening.

Table S2. The experimental data and Hypo_1-predicted IC_{50} values of training set (**1-29**) and test set (**30-98**).

Cmpd.	MDM2 binding IC_{50} (nM)		Error factor	Ref.	Cmpd.	MDM2 binding IC_{50} (nM)		Error factor	Ref.
	Exp.	Pred.				Exp.	Pred.		
1	1.6	7	4.4	2	**50**	160	180	1.1	5
2	2.8	6.8	2.4	2	**51***	162	110	-1.5	11
3	5	49	9.9	5	**52**	180	17	11	10
4	5	78	16	5	**53***	189	100	-1.8	11
5	16.4	51	3.1	11	**54***	192	170	-1.1	27
6	25.7	40	1.6	11	**55**	260	19	-14	9
7	77	360	4.6	3	**56***	300	380	1.3	18
8	87	6.9	-13	1	**57***	309	100	-3	27
9	90	390	4.4	9	**58**	310	290	-1.1	5
10	157	100	-1.5	11	**59**	340	120	-2.9	10
11	220	67	-3.3	15	**60***	344	680	2	26
12	252	92	-2.7	3	**61**	390	93	-4.2	10
13	300	61	-4.9	16	**62**	400	4400	11	17
14	800	1600	2	17	**63**	450	57	-7.9	10
15	1170	650	-1.8	16	**64**	490	29	-17	10

16	1500	180	-8.3	14	**65**	500	1500	3	17
17	1710	1300	-1.3	19	**66**	600	5300	8.8	17
18	2400	2500	1	21	**67**	600	9500	16	17
19	3000	2300	-1.3	17	**68**	620	110	-5.4	14
20	4260	450	-9.6	7	**69***	670	150	-4.4	14
21	12000	8700	-1.4	22	**70***	704	310	-2.2	13
22	14800	110000	7.7	16	**71**	760	750	-1	14
23	16400	17000	1	20	**72**	800	4000	5	17
24	32000	140000	4.2	23	**73**	1000	3700	3.7	17
25	49000	130000	2.6	25	**74**	1000	8500	8.5	17
26	206000	140000	-1.5	25	**75**	1230	620	-2	16
27	300000	130000	-2.3	22	**76**	1320	15000	11	14
28	439000	66000	-6.7	20	**77***	1400	450	-3.1	16
29	482000	96000	-5	24	**78**	1700	920	-1.9	15
30*	2	7.1	3.5	2	**79**	2000	2200	1.1	17
31	17	13	-1.3	26	**80**	2400	410	-5.8	14
32	28	67	2.4	11	**81**	2500	370	-6.8	14
33*	31	510	16	3	**82**	3060	350	-8.8	16
34	32	96	3	11	**83**	3100	1800	-1.7	20
35*	32	13	-2.5	11	**84**	3600	630	-5.7	15
36	48	140	2.9	26	**85**	4000	420	-9.4	28
37*	51	13	-3.9	1	**86**	5000	2700	-1.9	28
38	54	270	5	6	**87**	6000	4600	-1.3	17
39*	56	33	-1.7	11	**88**	7500	3100	-2.4	14
40	57	330	5.8	5	**89**	10000	14000	1.4	17
41	57	40	-1.4	5	**90**	12000	2600	-4.6	22
42	71	22	-3.2	8	**91**	15000	4500	-3.3	22
43*	76	100	1.4	10	**92***	20000	20000	-1	22
44	100	960	9.6	7	**93**	38000	77000	2	14
45*	107	67	-1.6	26	**94**	57000	8900	-6.4	20
46	110	260	2.4	6	**95***	117000	96000	-1.2	25
47*	110	610	5.6	18	**96**	150000	58000	-2.6	22
48	140	26	-5.4	9	**97**	250000	100000	-2.4	25
49	150	970	6.4	5	**98**	320000	74000	-4.3	24

*The compounds were used as active compounds (positive) in Enrichment Factor studies.

DOCKING SIMULATIONS

Docking simulations were carried out by using CDOCKER (Discovery Studio, version 2.1; Accelrys, San Diego, CA, USA, 2008). The X-ray crystal structure of MDM2 bound to the transactivation domain of p53 (PDB ID:1YCR) was used for the docking calculation. After removing the ligand and solvent molecules, the CHARMm-force field was applied to the protein. And the area around the p53 peptide was chosen as the active site with a radius set as 10 Å. Each compound was generated random conformation using CHARMm-based molecular dynamics

(1000 steps), and then docked into the defined MDM2 binding site. The other parameters were set as default. Each docking simulation was carried out at least three independent calculations, and the best CDOCKING ENERGY were listed in Table S3.

VIRTUAL SCREENING RESULTS

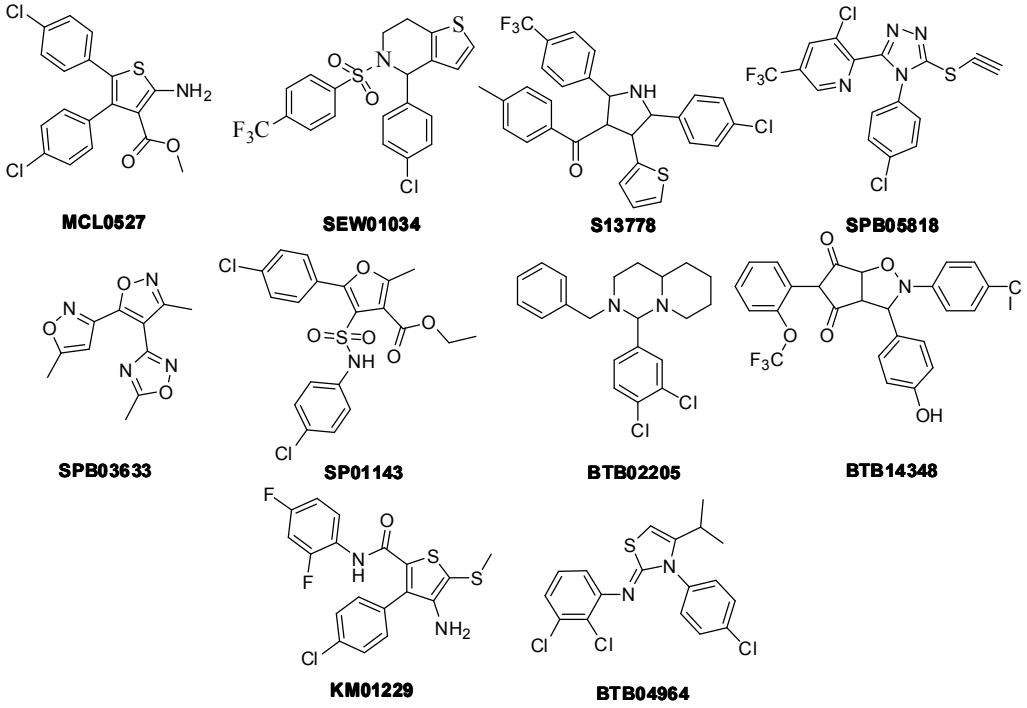

Figure S4. 2D structures of 10 preferable screening hits.

Table S3. The Fit-values, predicted MDM2 binding affinities and docking energy of 10 hits identified by virtual screening.

Compd.	Fit-value[a]	MDM2 binding Predicted IC_{50} (μM)	CDOCKING ENERGY (Kcal/mol)
MCL0527	11.56	0.79	-32.26
SEW01034	10.81	0.91	-23.02
S13778	10.08	1.1	-14.06
SPB05818	9.17	1.7	-26.58
SPB03633	8.53	2.7	-6.83
SP01143	7.96	6.7	-11.62
BTB02205	6.24	12	-7.50
BTB14348	5.58	18	-18.64
KM01229	3.78	58	-5.50
BTB04964	3.39	76	-2.35
Nutlin-3	11.98	0.39	-36.81

[a] Fit-value indicates how well the features in the pharmacophore overlap the chemical features in the molecule.

References:

1. *WO Pat*, 2008072655, 2008.
2. *WO Pat*, 2010082612, 2010.
3. *WO Pat*, 2009047161, 2009.
4. *WO Pat*, 2007063013, 2007.
5. *WO Pat*, 2006097261, 2006.
6. *US Pat*, 20070167437, 2007.
7. *US Pat*, 20040259884, 2004.
8. *US Pat*, 20050282803, 2005.
9. L. T. Vassilev, B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi and E. A. Liu, *Science.*, 2004, **303**, 844-848.
10. *WO Pat*, 2008036168, 2008.
11. S. Yu, D. Qin, S. Shangary, J. Chen, G. Wang, K. Ding, D. McEachern, S. Qiu, Z. Nikolovska-Coleska, R. Miller, S. Kang, D. Yang and S. Wang, *J Med Chem.*, 2009, **52**, 7970-7973.
12. I. R. Hardcastle, J. Liu, E. Valeur, A. Watson, S. U. Ahmed, T. J. Blackburn, K. Bennaceur, W. Clegg, C. Drummond, J. A. Endicott, B. T. Golding, R. J. Griffin, J. Gruber, K. Haggerty, R. W. Harrington, C. Hutton, S. Kemp, X. Lu, J. M. McDonnell, D. R. Newell, M. E. Noble, S. L. Payne, C. H. Revill, C. Riedinger, Q. Xu and J. Lunec, *J Med Chem.*, 2011, **54**, 1233-1243.
13. H. K. Koblish, S. Zhao, C. F. Franks, R. R. Donatelli, R. M. Tominovich, L. V. LaFrance, K. A. Leonard, J. M. Gushue, D. J. Parks, R. R. Calvo, K. L. Milkiewicz, J. J. Marugan, P. Raboisson, M. D. Cummings, B. L. Grasberger, D. L. Johnson, T. Lu, C. J. Molloy and A. C. Maroney, *Mol Cancer Ther.*, 2006, **5**, 160-169.
14. D. J. Parks, L. V. Lafrance, R. R. Calvo, K. L. Milkiewicz, V. Gupta, J. Lattanze, K. Ramachandren, T. E. Carver, E. C. Petrella, M. D. Cummings, D. Maguire, B. L. Grasberger and T. Lu, *Bioorg Med Chem Lett.*, 2005, **15**, 765-770.
15. P. Raboisson, J. J. Marugan, C. Schubert, H. K. Koblish, T. Lu, S. Zhao, M. R. Player, A. C. Maroney, R. L. Reed, N. D. Huebert, J. Lattanze, D. J. Parks and M. D. Cummings, *Bioorg Med Chem Lett.*, 2005, **15**, 1857-1861.
16. J. G. Allen, M. P. Bourbeau, G. E. Wohlhieter, M. D. Bartberger, K. Michelsen, R. Hungate, R. C. Gadwood, R. D. Gaston, B. Evans, L. W. Mann, M. E. Matison, S. Schneider, X. Huang, D. Yu, P. S. Andrews, A. Reichelt, A. M. Long, P. Yakowec, E. Y. Yang, T. A. Lee and J. D. Oliner, *J Med Chem.*, 2009, **52**, 7044-7053.
17. R. Zhang, T. Mayhood, P. Lipari, Y. Wang, J. Durkin, R. Syto, J. Gesell, C. McNemar and W. Windsor, *Anal Biochem.*, 2004, **331**, 138-146.
18. *WO Pat*, 2008119741, 2008.
19. G. M. Popowicz, A. Czarna, S. Wolf, K. Wang, W. Wang, A. Domling and T. A. Holak, *Cell Cycle.*, 2010, **9**, 1104-1111.
20. I. R. Hardcastle, S. U. Ahmed, H. Atkins, G. Farnie, B. T. Golding, R. J. Griffin, S. Guyenne, C. Hutton, P. Kallblad, S. J. Kemp, M. S. Kitching, D. R. Newell, S. Norbedo, J. S. Northen, R. J. Reid, K. Saravanan, H. M. Willems and J. Lunec, *J Med Chem.*, 2006, **49**, 6209-6221.
21. I. R. Hardcastle, J. Liu, E. Valeur, A. Watson, S. U. Ahmed, T. J. Blackburn, K. Bennaceur, W. Clegg, C. Drummond, J. A. Endicott, B. T. Golding, R. J. Griffin, J. Gruber, K. Haggerty, R. W.

Harrington, C. Hutton, S. Kemp, X. Lu, J. M. McDonnell, D. R. Newell, M. E. Noble, S. L. Payne, C. H. Revill, C. Riedinger, Q. Xu and J. Lunec, *J Med Chem.*, 2011, **54**, 1233-1243.

22. L. Chen, H. Yin, B. Farooqi, S. Sebti, A. D. Hamilton and J. Chen, *Mol Cancer Ther.*, 2005, **4**, 1019-1025.

23. P. S. Galatin and D. J. Abraham, *J Med Chem.*, 2004, **47**, 4163-4165.

24. *WO Pat*, 2004005278, 2004.

25. R. Stoll, C. Renner, S. Hansen, S. Palme, C. Klein, A. Belling, W. Zeslawski, M. Kamionka, T. Rehm, P. Muhlhahn, R. Schumacher, F. Hesse, B. Kaluza, W. Voelter, R. A. Engh and T. A. Holak, *Biochemistry-Us.*, 2001, **40**, 336-344.

26. *US Pat*, 2007129416, 2007.

27. *WO Pat*, 2010031713, 2010.

28. U. Rothweiler, A. Czarna, M. Krajewski, J. Ciombor, C. Kalinski, V. Khazak, G. Ross, N. Skobeleva, L. Weber and T. A. Holak, *ChemMedChem.*, 2008, **3**, 1118-1128.

29. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J Comput Chem.*, 1983, **4**, 187-217.

30. A. K. Debnath, *J Med Chem.*, 2002, **45**, 41-53.

31. S. Sakkiah, S. Thangapandian, S. John, Y. J. Kwon and K. W. Lee, *Eur J Med Chem.*, 2010, **45**, 2132-2140.

32. P. C. Hawkins, A. G. Skillman and A. Nicholls, *J Med Chem.*, 2007, **50**, 74-82.