

Dissecting celastrol with machine learning to unveil dark pharmacology

Tiago Rodrigues,^a Bernardo P. de Almeida,^a Nuno L. Barbosa-Morais^a and Gonalo J. L. Bernardes^{a,b,*}

Received 00th January 20xx
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

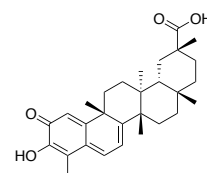
Coalescing bespoke machine learning and bioinformatics analyses with cell-based assays we unveil pharmacology of celastrol. Celastrol is a direct modulator of the progesterone and cannabinoid receptors, and its effects correlate with the antiproliferative activity. We demonstrate how *in silico* methods may drive systems biology studies for natural products.

Introduction

Natural products (NPs) are a privileged source of protein binding motifs that have been curated by evolutionary pressure over millions of years.^{1,2} NPs offer diverse and biologically-relevant chemotypes, which ought to be explored in chemical biology and drug discovery programs.³ The significant expansion of the synthetic chemistry toolbox has played a paramount role in enabling the exploration of such molecules. Still, NP research is experiencing a steady decline in recent years.⁴ It has been argued that the lack of knowledge of binding counterparts is an important contributor to such a decline as well as a major bottleneck in early discovery chemistry.⁵ As target identification method of choice, chemical proteomics offers an elegant solution to unravel the underlying biology of NPs, but such methods persist as both laborious and time consuming.⁶ Also, the need for chemical derivatization of the NP of interest may drastically disrupt its affinity towards relevant on-/off-targets.⁵ Moreover, pull down of membrane proteins, *e.g.* G-protein coupled receptors, is unlikely, as is the identification of proteins with low cellular expression. As an alternative, we and others have shown that machine intelligence harnessing the increasing wealth of currently available biological and chemical data is a viable and complementary solution to chemical proteomics, while mitigating several shortcomings of the latter.⁷⁻⁹

Celastrol (**1**, Fig. 1) is a quinone methide triterpene that exhibits anti-obesity and antiproliferative activities, through disparate mechanisms of action.¹⁰ For example, it has been shown that **1** inhibits the accumulation of nuclear hypoxia-inducible factor-1 α (HIF-1 α), suppresses HSP90 and reduces the transcriptional activity of

VEGF genes.¹¹ However, despite the clinical interest – including anti-inflammation, cancer and neurodegenerative disorders^{12, 13} – the network pharmacology of **1** remains poorly understood, which may hamper an effective preclinical development. Herein we shed light on the dark pharmacology of **1** by using machine intelligence to confidently and accurately quantify the predicted ligand–target relationships. Employing this approach we identified **1** as a potent cannabinoid receptor-1/2 (CB1/2) agonist and as a moderate progesterone receptor antagonist. Importantly, by analysing the gene expression of a panel of cancer cell lines, we associated the *CB1* gene with the antiproliferative activity of the NP in specific cell lines. The result is relevant, as the low *CB1* transcript levels in the analyzed cell lines would render the discovery inaccessible through mainstream chemical proteomic approaches.



Celastrol (**1**)

Fig. 1 Structure of the quinone methide triterpene celastrol.

Results and discussion

As part of our ongoing research program aiming to clarify modes of action for NPs of clinical interest we set out to investigate **1**. We used SPiDER¹⁴ to predict potential drug targets for **1**, given the extensive prior validation of the software tool in NP chemical space.¹⁵⁻¹⁸ In short, SPiDER builds independent self-organizing maps based on topological pharmacophore¹⁹ and physicochemical small molecule descriptors (MOE, Chemical Computing Group, Canada) prior to defining a consensus qualitative (bind / does not bind) prediction. The exploration of the chemical feature space through self-organizing

^a Instituto de Medicina Molecular Joo Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028, Lisboa, Portugal.

^b Department of Chemistry, University of Cambridge, Lensfield Road CB2 1EW, Cambridge, UK.

* Corresponding author e-mail: gbernardes@medicina.ulisboa.pt, gb453@cam.ac.uk.

Electronic Supplementary Information (ESI) available: Methods, supplementary data and figures. See DOI: 10.1039/x0xx00000x

maps offers an intuitive means for data visualization and interpretation, as co-clustering of query compounds with any given reference ligands suggests a similar ligand–target relationship.¹⁴ Additionally, a *p*-like value is calculated to rank order the SPiDER output. Through this approach, several targets were predicted for **1**, including G-protein coupled receptors, ion channels, nuclear receptors and enzymes (Table 1).

To curate SPiDER predictions, we had previously built DEcRyPT, an orthogonal machine learning method that infers affinity values (*pAffinity*, $-\log[\text{Affinity}]$) for the query ligand against drug targets of interest.⁷ DEcRyPT employs the random forest (RF) regression technology and type-scaled CATS2 topological pharmacophore descriptors (MOE implementation), which provide a “fuzzy” molecular representation and have been deemed important to leverage confident target predictions in NP space.¹⁶ Despite a prior successful case study,⁷ the utility of DEcRyPT remained limited given that only 236 drug targets had been made known to the method. Therefore, we here considerably expanded the range of predictable targets to 1026 (DEcRyPT 2.0). An average mean absolute error of 0.541 ± 0.167 *pAffinity* units in 10-fold cross-validation studies suggests the general utility of the RF models (*cf.* ESI). Importantly, it has been suggested that adversarial controls should become commonplace in current machine learning applications, to rule out the exploitation of data/experimental artifacts and validate the applicability of the employed descriptors.^{20, 21} To that end, we built alternate models by randomizing the target variable to assess the relevance of DEcRyPT 2.0 and the importance/correlation of the CATS2 descriptors with the reported bioactivities (*cf.* ESI). Our data fully supports the appropriateness of our target prediction technology, as all adversarial controls provided notorious low accuracy in stratified 10-fold cross-validation studies. With these RF models in hand, and to provide an additional layer of confidence to the affinity predictions, we created an ancillary Score value based on the background *pAffinity* distribution for each of the predictable targets. To compute such a background, we used 139,352 molecules previously identified as dark chemical matter,²² assuming that these entities are essentially inactive against drug targets included in DEcRyPT 2.0, *i.e.* afford low (basal) *pAffinity* values for each target. As implemented in DEcRyPT 2.0, higher Score values denote a higher distance to the background *pAffinity* distribution (Table 1), thus providing a quantitative metric for assay prioritization.

Table 1. Drug target predictions for celastrol using SPiDER and DEcRyPT 2.0.

SPiDER		DEcRyPT 2.0	
Target	<i>p</i> value	<i>pAffinity</i> ^a	Score
Mineralocorticoid	0.012	6.4	-0.4
PTP ^{b,c}	0.017	4.1–5.5	-204.2–44.8 ^d
Progesterone	0.018	7.5 ^f	132.2
Glutamate ionotropic	0.022	4.7–5.7	-78.9–20.3 ^d
Liver X	0.023	5.5 (α), 6.2 (β)	-68.3, 43.3
TRP channel	0.024	6.1 (TRPV1)	12.4
Phospholipase ^c	0.027	5.3 (A2)	86.1
Smoothened	0.028	7.6	114.3
GPCR19	0.030	5.7	-79.4
Polymerase	0.031	5.5	-54.3
Farnesoid X	0.033	5.0	-89.8
Vitamin D	0.034	6.5 ^e	143.0
Estrogen ^c	0.037	5.2 (α), 5.3 (β)	14.6, -21.5

Microtubules	0.038	n/a	n/a
Androgen	0.040	6.8	38.9
Glucocorticoid	0.041	5.6	-58.3
Aromatase	0.041	5.1	-91.6
Integrins ^c	0.044	7.0	46.8
Cannabinoid	0.045	6.5 ^g (CB1), 7.3 ^h (CB2)	59.0, 62.2
Acid glycoprotein	0.046	n/a	n/a

^a*pAffinity* = $-\log(\text{EC}_{50}, K_{D/i})$; ^bProtein tyrosine phosphatase; ^cBoldface denotes targets whose association to celastrol had been previously reported in the literature, but remained unknown to the algorithms; ^dRange denotes that several targets within the same protein family are considered. ^e*pAffinity* variance = 2.4. ^f*pAffinity* variance = 2.0. ^g*pAffinity* variance = 1.2. ^h*pAffinity* variance = 1.5. n/a, not available due to insufficient data to build a random forest model.

Applying to **1**, both machine learning methods presented complementary vantage points and confidence regarding their output (Table 1). Indeed, while the predictions suggest that **1** remains unexplored as binding counterpart for a number of cases, both methods confidently predicted targets whose association to **1** was unknown to the algorithms, but already reported in the literature, *e.g.* protein tyrosine phosphatase 1B,²³ phospholipase A2,²⁴ estrogen receptor,²⁵ and integrins.²⁶ Indirect evidence of modulation of cannabinoid receptor-2 had also been reported.²⁷ Having obtained encouraging results from this pseudo-prospective evaluation of **1**, we then prioritized the cannabinoid 1/2, vitamin D and progesterone receptors for screening. These had not been previously probed, and a range of predictions highly distinct from the background had been made by DEcRyPT 2.0. Additionally, confirmation of activity would significantly expand the ligand space – as assessed by uncertainty sampling based on computed *pAffinity* variances²⁸ – and binding and/or functional assays were readily accessible. Moreover, all the prioritized targets had been previously implicated in cancer,^{29–31} which can potentially explain the phenotypic effects of **1** both *in vitro* and *in vivo*.

A primary, single-concentration (20 μM or *pAffinity* = 4.7, corresponding to 50% effect at the tested concentration) screen against the selected targets (Fig. 2a, b) showed a range of functional effects by **1**. Notoriously, **1** was inactive against the vitamin D receptor (VDR; predicted *pAffinity* = 4.0 vs. experimental *pAffinity* < 4.7), and presented agonistic and antagonistic effects against the cannabinoid and progesterone receptors, respectively (Fig. 2a,b), as confidently suggested by the *pAffinity* values computed by DEcRyPT 2.0.

Next, we confirmed the screening data in concentration–response curve studies (Fig. 2c), where **1** presented potent agonistic effects for cannabinoid receptors – EC_{50} (CB1) = $0.8 \mu\text{M} \pm 0.2$ log units (*pAffinity* = 6.1); EC_{50} (CB2) = $1.1 \mu\text{M} \pm 0.1$ log units (*pAffinity* = 6.0) – and modest antagonism of the progesterone receptor (K_B = $7.1 \mu\text{M}$; *pAffinity* = 5.1). Radioligand displacement assays further corroborated the direct interaction with the receptors (*cf.* ESI). Moreover, we ruled out **1** as a small colloiddally aggregating molecule (SCAM) disrupting cannabinoid receptor signaling by protein sequestration, partial denaturation or protein scaffold mimicry – the major sources of false positive readouts in screening assays³² – at concentrations below 10 μM through dynamic light scattering measurements (*cf.* ESI). Most importantly, the obtained *pAffinity* data is in line with the DEcRyPT 2.0 predictions (Table 1), which

validates the RF models and the designed Score value as assistants in the prioritization of biochemical assays.

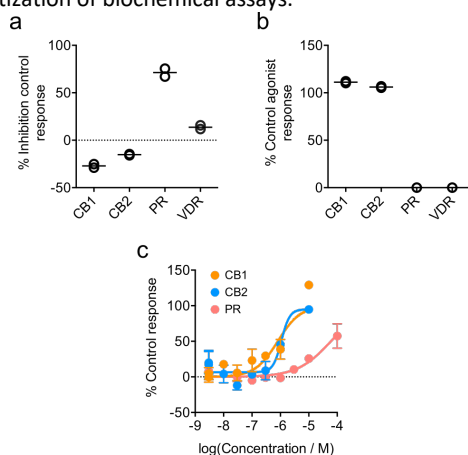


Fig. 2 Celastrol, **1**, is a modulator of the progesterone (PR), and cannabinoid 1/2 (CB1 and CB2), but not of the vitamin D (VDR) receptors. a) Screening of **1** (20 μ M) for functional antagonism; $n = 2$. b) Screening of **1** (20 μ M) for functional agonism; $n = 2$. c) Concentration–response curves of **1** against the CB1/2 and progesterone receptors; $n = 2$; dots depict the average value and whiskers correspond to the range of values. EC_{50} (CB1) = 0.8 μ M \pm 0.2 log units; EC_{50} (CB2) = 1.1 μ M \pm 0.1 log units; IC_{50} (PR) = 60 μ M \pm 0.1 log units, $K_B = 7.1$ μ M.

Key for this successful application of SPiDER/DEcRyPT 2.0 to **1** was the learning of data patterns and relationships between ligands/biology beyond simple ligand similarity measures. For example, through a common similarity-based search using the ECFP4-like Morgan fingerprint (radius 2, 2048 bits), we obtained an average Tanimoto index < 0.10 when comparing **1** to ligands of CB1/2 and the progesterone receptor in ChEMBL v24. In any case had the nearest neighbor a Tanimoto similarity higher than 0.19. For similar ligands this index approaches 1. Conversely, if the ligands are substantially different the result is close to 0. Thus, the similarity search results would have disqualified screening of **1** against the cannabinoid and progesterone receptors, given the dissimilarity between ligands. The results further attest this NP as a new chemotype for the modulation of the receptors reported herein. Noteworthy, the association of cannabinoid and progesterone biology by ligand structure / pharmacophore features remained non-obvious as no cross-screening between these two receptor families had been reported, according to ChEMBL. Thus, simple database analyses cannot rationalize the performed assays, which highlights the potential of our machine intelligence.

Intrigued by the obtained data we then projected the cannabinoid and progesterone receptor ligands, together with **1**, to the plane using manifold learning of CATS2 descriptors (Fig. 3). Our results show that, besides a new chemotype, compound **1** also presents an unusual topological pharmacophore feature arrangement as it does not cluster in the bulk of CB1/2 and progesterone receptor ligands. The obtained projection corroborates the output by DEcRyPT 2.0, as a ligand dissimilarity may explain the high prediction variance values and fulfils the initial selection criteria of expanding ligand space. Altogether, our data show the potential and highlight the uniqueness

of our machine intelligence method for identifying subtle patterns in sparse chemical datasets and unravelling dark pharmacology of small bioactive molecules.

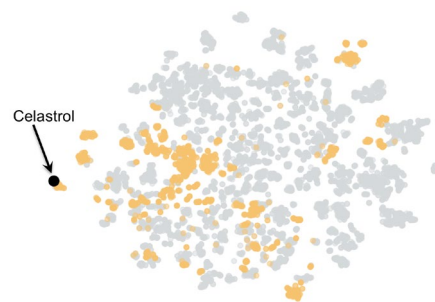


Fig. 3 Projection (t -distributed stochastic neighborhood embedding) of topological pharmacophore descriptor (CATS2) space. Grey: Cannabinoid receptor ligands; Orange: progesterone receptor ligands. Data shows that celastrol (black dot) presents an unusual pattern among ligands of the studied receptors.

To assert the relevance of inhibition of CB1/2 in the antiproliferative activity of **1**, we analyzed the publically available screening data against the NCI-60 panel of cancer cell lines.³³ We sought for a correlation between GI_{50} values of **1** against 60 different cancer cell lines and their CB1/2 gene expression profiles,³⁴ hypothesising that **1** will have higher activity in cell lines highly expressing the predicted target genes. Compound **1** displayed a range of activity values across the different NCI-60 tissues of origin (p value = 0.005, Kruskal-Wallis rank sum test; Fig. 4a), showing higher inhibitory activity in blood and kidney cell lines and lower activity in lung cancer (Fig. 4a and ESI). Indeed, although CB1 and CB2 are generally lowly expressed in the NCI-60 panel (average expression in 16 and 33 percentiles among all 23,060 genes, respectively; Fig. 4b), the expression of CB1 associates with the antiproliferative activity of **1** (Spearman's $\rho = 0.29$, p value = 0.03). Conversely, no significant correlation was found for CB2 ($\rho = -0.05$, p value = 0.72), nor for other candidate targets predicted by SPiDER / DEcRyPT 2.0 (Table 1), or reported in the literature (*cf.* ESI). The result reinforces the significance of our finding and the putative role of CB1 receptors in the antiproliferative activity of **1**.

Next, we analysed the expression levels of CB1/2 across 9,721 tumour and 725 matched-normal human samples of 32 cancer types from The Cancer Genome Atlas (TCGA; <https://cancergenome.nih.gov/>; *cf.* ESI), aiming at identifying tumour types where the impact of **1** should be higher. CB1 was differently expressed (False Discovery Rate (FDR) < 0.05 , Wilcoxon rank-sum test) between tumour and matched-normal samples in 14 out of 15 cancer types (lowerly and higherly expressed in tumour samples of 10 and 4 cancer types, respectively – *cf.* ESI). CB2 was differently expressed in 7 cancer types (FDR < 0.05 , Wilcoxon rank-sum test) (*cf.* ESI). Despite the varying degrees of activity of **1** against the NCI-60 cell lines, overexpression of cannabinoid receptors potentiates the activity of the NP (Fig. 4c). It is apparent that other mechanisms of action are at play, and that phenotype changes induced by **1** are a result of multi-target engagement. Nonetheless, tuning pharmacology, such as activity against cannabinoid receptors, by molecular design may prove an important strategy to enhance the antiproliferative activity of celastrol analogues in

appropriate cancer subtypes. Critically, molecular design should also focus on improving the oral absorption of **1**, which has been shown suboptimal.³⁵ Alternatively, to increase the concentration of **1** to levels relevant to modulation of the cannabinoid signalling, tailored drug delivery systems may be designed.³⁶

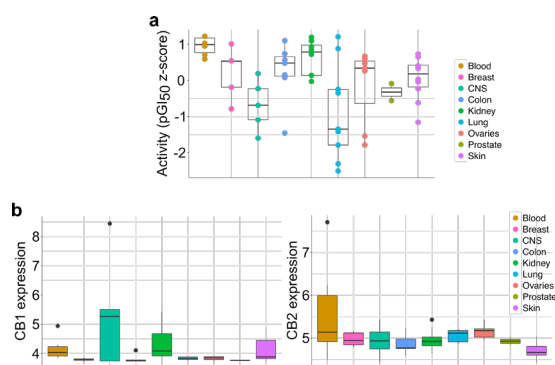


Fig. 4 Expression of the cannabinoid 1 receptor (CB1) contributes to the antiproliferative activity of celestrol. **a)** Activity of celestrol against the NCI-60 panel of cancer cell lines, grouped by tissue. **b)** Cannabinoid receptors have low expression in different types of cancer cell lines.

Conclusions

In summary, we have built a new and widely applicable machine intelligence tool to unravel dark pharmacology of NPs. We show how such a method counter-intuitively integrates uncertainty sampling and a scoring function to assess prediction reliability. By applying DECrypT 2.0 to study **1**, we rediscovered known targets but, most importantly, illuminated hitherto unknown biology. Significantly, the discovered targets could be correlated with the antiproliferative activity of **1**, thus offering new avenues for the development of celestrol-inspired chemical matter and further exploit the cannabinoid-mediated signaling in cancer.

Notes and references

We thank the Royal Society (URF\R\180019), H2020 (ERC StG, GA no. 676832 and 743640) and FCT Portugal (IF/00624/2015 and 02/SAICT/2017, Grant 28333) for funding. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 807281.

1. T. Rodrigues, D. Reker, P. Schneider and G. Schneider, *Nat. Chem.*, 2016, **8**, 531.
2. D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2012, **75**, 311.
3. S. Wetzler, R. S. Bon, K. Kumar and H. Waldmann, *Angew. Chem. Int. Ed.*, 2011, **50**, 10800.
4. E. Patridge, P. Gareiss, M. S. Kinch and D. Hoyer, *Drug Discov. Today*, 2016, **21**, 204.
5. T. Rodrigues, *Org. Biomol. Chem.*, 2017, **15**, 9275.
6. L. Laraia and H. Waldmann, *Drug Discov. Today Technol.*, 2017, **23**, 75.
7. T. Rodrigues, M. Werner, J. Roth, E. H. G. da Cruz, M. C. Marques, P. Akkapeddi, S. A. Lobo, A. Koeberle, F. Corzana, E. N. da Silva Junior, O. Werz and G. J. L. Bernardes, *Chem. Sci.*, 2018, **9**, 6899.

8. T. Rodrigues, F. Sieglitz, V. J. Somovilla, P. M. Cal, A. Galione, F. Corzana and G. J. Bernardes, *Angew. Chem. Int. Ed.*, 2016, **55**, 11077.
9. P. Schneider and G. Schneider, *Angew. Chem. Int. Ed.*, 2017, **56**, 11520.
10. R. Cascao, J. E. Fonseca and L. F. Moita, *Front. Med.*, 2017, **4**, 69.
11. L. Huang, Z. Zhang, S. Zhang, J. Ren, R. Zhang, H. Zeng, Q. Li and G. Wu, *Int. J. Mol. Med.*, 2011, **27**, 407.
12. A. C. Allison, R. Cacabelos, V. R. Lombardi, X. A. Alvarez and C. Vigo, *Prog. Neuropsychopharmacol. Biol. Psychiatry*, 2001, **25**, 1341.
13. D. Kashyap, A. Sharma, H. S. Tuli, K. Sak, T. Mukherjee and A. Bishayee, *Crit. Rev. Oncol. Hematol.*, 2018, **128**, 70.
14. D. Reker, T. Rodrigues, P. Schneider and G. Schneider, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, 4067.
15. G. Schneider, D. Reker, T. Chen, K. Hauenstein, P. Schneider and K. H. Altmann, *Angew. Chem. Int. Ed.*, 2016, **55**, 12408.
16. D. Reker, A. M. Perna, T. Rodrigues, P. Schneider, M. Reutlinger, B. Monch, A. Koeberle, C. Lamers, M. Gabler, H. Steinmetz, R. Muller, M. Schubert-Zsilavec, O. Werz and G. Schneider, *Nat. Chem.*, 2014, **6**, 1072.
17. T. Rodrigues, D. Reker, J. Kunze, P. Schneider and G. Schneider, *Angew. Chem. Int. Ed.*, 2015, **54**, 10516.
18. L. Friedrich, T. Rodrigues, C. S. Neuhaus, P. Schneider and G. Schneider, *Angew. Chem. Int. Ed.*, 2016, **55**, 6789.
19. M. Reutlinger, C. P. Koch, D. Reker, N. Todoroff, P. Schneider, T. Rodrigues and G. Schneider, *Mol. Inf.*, 2013, **32**, 133.
20. K. V. Chuang and M. J. Keiser, *Science*, 2018, **362**, pii: eaat8603.
21. K. V. Chuang and M. J. Keiser, *ACS Chem. Biol.*, 2018, **13**, 2819.
22. A. M. Wassermann *Et al.*, *Nat. Chem. Biol.*, 2015, **11**, 958.
23. L. M. Scott, L. Chen, K. G. Daniel, W. H. Brooks, W. C. Guida, H. R. Lawrence, S. M. Sebt, N. J. Lawrence and J. Wu, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 730.
24. V. Joshi, S. H. Venkatesha, C. Ramakrishnan, A. N. Nanjaraj Urs, V. Hiremath, K. D. Moudgil, D. Velmurugan and B. S. Vishwanath, *Pharmacol. Res.*, 2016, **113**, 265.
25. S. Y. Jang, S. W. Jang and J. Ko, *Cancer Lett.*, 2011, **300**, 57.
26. H. Zhu, X. W. Liu, T. Y. Cai, J. Cao, C. X. Tu, W. Lu, Q. J. He and B. Yang, *J. Pharmacol. Exp. Ther.*, 2010, **334**, 489.
27. L. Yang, Y. Li, J. Ren, C. Zhu, J. Fu, D. Lin and Y. Qiu, *Int. J. Mol. Sci.*, 2014, **15**, 13637.
28. D. Reker, G. J. L. Bernardes and T. Rodrigues, *ChemRxiv*, 2018, <https://doi.org/10.26434/chemrxiv.7291205.v7291201>.
29. J. Thorne and M. J. Campbell, *Proc. Nutr. Soc.*, 2008, **67**, 115.
30. A. R. Daniel, C. R. Hagan and C. A. Lange, *Expert Rev. Endocrinol. Metab.*, 2011, **6**, 359.
31. B. Chakravarti, J. Ravi and R. K. Ganju, *Oncotarget*, 2014, **5**, 5852.
32. D. Reker, G. J. L. Bernardes and T. Rodrigues, *Nature Chem.*, 2019, **11**, DOI: 10.1038/s41557-019-0234-9.
33. R. H. Shoemaker, *Nat. Rev. Cancer*, 2006, **6**, 813.
34. W. H. Gmeiner, W. C. Reinhold and Y. Pommier, *Mol. Cancer Ther.*, 2010, **9**, 3105.
35. J. Zhang, C. Y. Li, M. J. Xu, T. Wu, J. H. Chu, S. J. Liu and W. Z. Ju, *J. Ethnopharmacol.*, 2012, **144**, 195.
36. L. Guo, S. Luo, Z. Du, M. Zhou, P. Li, Y. Fu, X. Sun, Y. Huang and Z. Zhang, *Nat. Commun.*, 2017, **8**, 878.