ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 297–303

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

COMPARE: a web accessible tool for investigating mechanisms of cell growth inhibition

Daniel W. Zaharevitz*, Susan L. Holbeck, Christopher Bowerman, Penny A. Svetlik

Information Technology Branch, Developmental Therapeutics Program, National Cancer Institute, EPN Room 811, 6130 Executive Blvd., Bethesda, MD 20892, USA

Received 13 April 2001; accepted 13 April 2001

Abstract

For more than 10 years the National Cancer Institute (NCI) has tested compounds for their ability to inhibit the growth of human tumor cell lines in culture (NCI screen). Work of Ken Paull [J. Natl. Cancer Inst. 81 (1989) 1088] demonstrated that compounds with similar mechanism of cell growth inhibition show similar patterns of activity in the NCI screen. This observation was developed into an algorithm called COMPARE and has been successfully used to predict mechanisms for a wide variety of compounds. More recently, this method has been extended to associate patterns of cell growth inhibition by compounds with measurements of molecular entities (such as gene expression) in the cell lines in the NCI screen. The COMPARE method and associated data are freely available on the Developmental Therapeutics Program (DTP) web site (http://dtp.nci.nih.gov/). Examples of the use of COMPARE on these web pages will be explained and demonstrated. Published by Elsevier Science Inc.

Keywords: Cell-based screening; Correlation analysis; COMPARE; Drug discovery; Antitumor drugs

1. Introduction

Advances in structural and molecular biology have allowed modern drug discovery efforts to focus on finding molecules that specifically interact with particular targets. Assays are developed that can measure this specific interaction and the assays are then used to identify lead compounds and monitor development of analogs. While there is no doubt that this approach is enormously useful in a wide variety of projects, there are limitations; chiefly that drug discovery efforts cannot start until a particular target has been identified. This is in contrast to empirical approaches where some complex biological endpoint is measured in an intact system and compounds that achieve the desired effect by any mechanism, known or unknown, could potentially be discovered.

For almost 50 years, the National Cancer Institute (NCI) has used an empirical approach for general screening of compounds for potential anticancer activity. For more than 10 years, the screen has been a measurement of growth inhibition of 60 human tumor cell lines in culture [1–3]. Early in the development of this screen, Ken Paull recognized that the differential sensitivities of the cell lines to a particular compound formed a distinctive pattern and that compounds that had similar patterns of activities tended to have similar mechanisms of cell growth inhibition. He developed this observation into the COMPARE algorithm [4,5], thus allowing this empirical assay to gain at least some of the advantages of more targeted assays. This work will describe types of questions that COMPARE has been found useful for investigating and how to access the COMPARE calculations and associated data on the Developmental Therapeutics Program (DTP) web site.

2. COMPARE methodology

The details of the COMPARE methodology have been described [5]. Briefly, it consists of calculating the linear correlation coefficient between the data over all cell lines for the pattern of interest (called the seed) and all sets of data in the database to be searched. The results are sorted by the correlation coefficient. The method was originally implemented in the statistical program, SAS. In bringing the method to the web, the calculation of the correlation coefficient was implemented in a C program that uses the Common Gateway Interface (CGI) to accept parameters from a web form. Running the calculation involves three steps: (1) choosing

* Corresponding author. Tel.: +1-301-496-8747; fax: +1-301-480-4808.
E-mail address: zaharevitz@dtpax2.ncifcrf.gov (D.W. Zaharevitz).

1093-3263/02/$ – see front matter. Published by Elsevier Science Inc.
PII: S1093-3263(01)00126-7

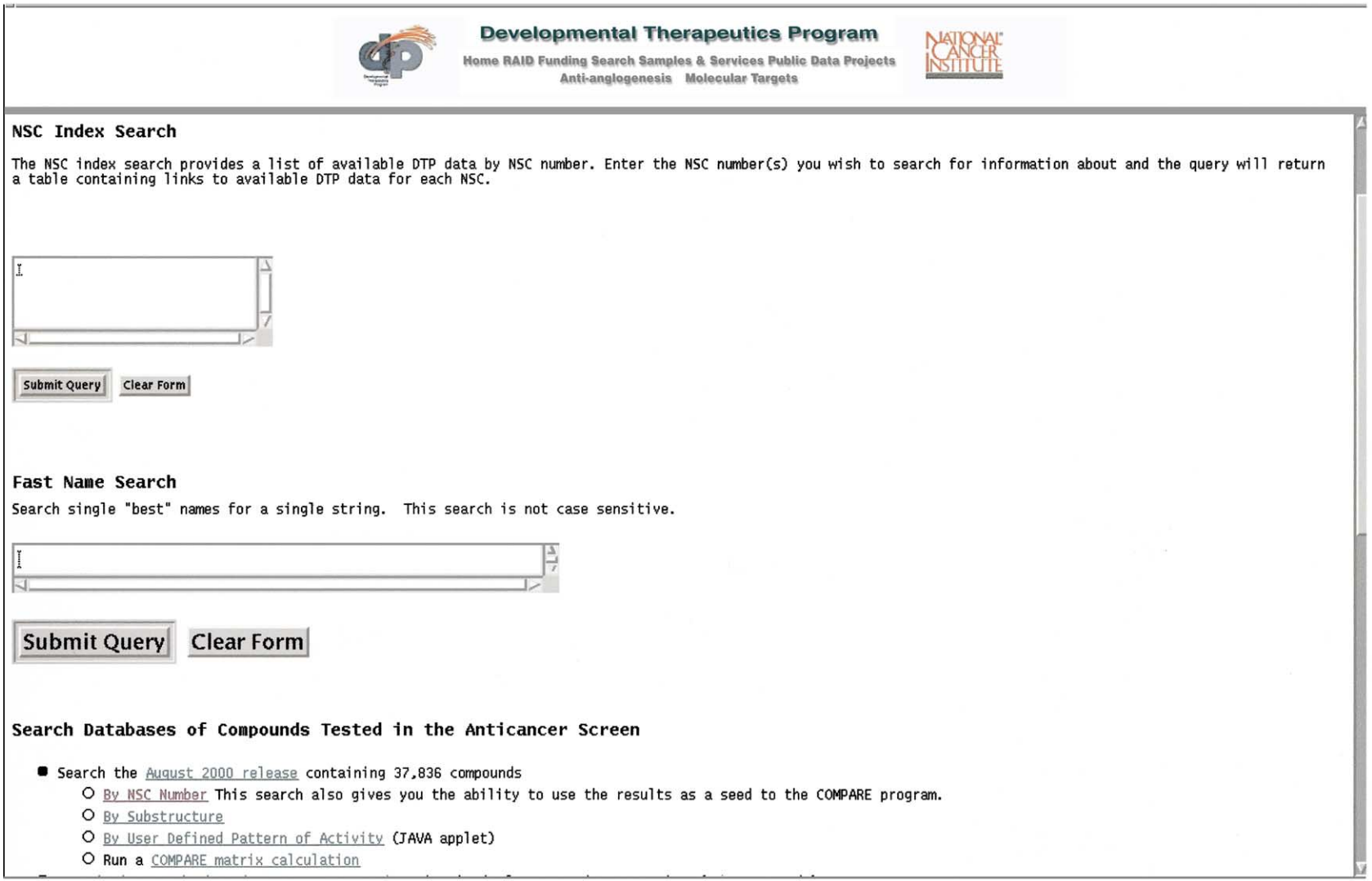

Fig. 1. Search page on the DTP web site. Searching by NSC number, chemical name and chemical substructure is available.

the seed; (2) choosing the parameters; and (3) choosing the database.

The most common choices of seed are the results from the NCI screen for a particular compound. Various methods for finding compounds are available on the DTP search page (http://dtp.nci.nih.gov/docs/dtp_search.html) (Fig. 1). Searches can be performed on NSC number (the NCI internal identification number), chemical name (although only a small fraction of the tested compounds has a name), or chemical substructure. Measurement of some molecular target (discussed more fully below) is also a potential seed and these data can be searched from http://dtp.nci.nih.gov/mtargets/mt_index.htm. There is also a JAVA applet that allows users to create an arbitrary pattern and use it as a seed.

The COMPARE calculation can be modified by a number of user controlled parameters (Fig. 2). The calculation can be set to return either the largest positive correlations (standard COMPARE) or the largest negative correlations (reverse COMPARE). The number of correlations to report can be specified as well as the growth inhibition measurement (GI_{50}; concentration required for 50% inhibition of growth, TGI; concentration required for total inhibition of growth, and LC_{50}; concentration required for 50% cell kill) to use. The correlations are calculated by default using all data for all cell lines that appear in both the seed and the test set. In cases where some data has failed quality control checks, the calculation can include less than the full 60 cell line panel. The user can set a value for the minimum number of cell lines necessary for the calculation to be reported. The user can also specify the minimum variance in a test set for the calculation to be reported. This parameter is useful to exclude data where the variation in the test set is very much smaller than the seed. The default of 0.02 is near the lowest reasonable value, but larger values could be useful depending on the seed. Finally there are a number of ways the user can modify the choice of cell lines used in the calculation. There are several subgroups of cell lines that can be chosen via a radio button. At the present time, these include cells with low expression of the multi-drug resistance (MDR) protein, cells that are wild type for *p53* and cells that are *p53* mutant. This selection gives the user the option to minimize the confounding influence of these conditions on the correlations. There is also an option to specify the handling of each individual cell line. For each cell line the user can choose whether data for this cell line is required for a calculation to be reported, used if present (but not required), or ignored in the calculation.

At the present time there are six databases that can be searched with a COMPARE calculation. The “Synthetics” database consists of cell growth inhibition data for

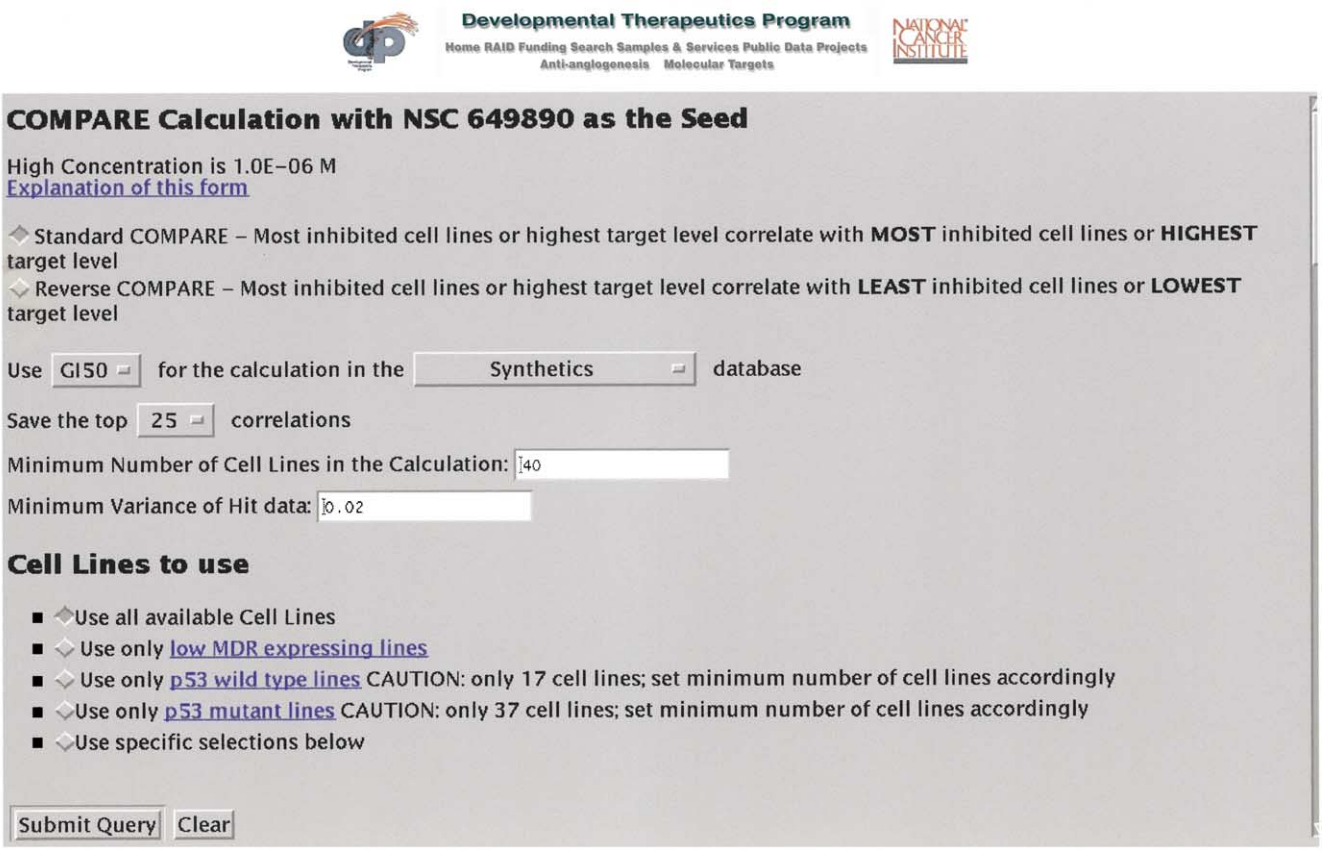

Fig. 2. COMPARE parameters. The seed is already chosen and is NSC 649890, flavopiridol. The parameter choices are explained in the text.

approximately 38,000 synthetic or pure natural product compounds. The "Standard Agents" are a set of 175 compounds that have been clinically tested in humans for anticancer activity [6]. The "Diversity Set" is a set of 1990 compounds that are available from the NCI on 96-well plates [7]. The latter two databases are subsets of the "Synthetics" database. The "Natural Product Extracts" contains screening results for approximately 15,000 crude natural product extracts [8]. Lastly there are two databases of molecular characterizations of the cell lines in the screen. The "Molecular Targets" database contains data for almost 200 target levels and activity measurements made by a number of different workers [9]. The "NCI60 Microarray" database contains the DNA chip measurements previously reported [10,11]. The data contained in all these databases are available for downloading in bulk [12].

The format of the results returned depends on the database searched. Searches run in databases of cell growth inhibition results return (Fig. 3) the rank order, NSC number, chemical name (if any), highest concentration tested, average value of the measurement over all the cell lines, the number of cell lines used in the calculation, the variance of the data used, the number of experiments that were averaged to get the data, the correlation coefficient, and a link to get a display of both the seed pattern and the test pattern on the same page. The chemical name is linked to a PubMed text search on the name. The NSC number is linked to a set of 3D coordinates for the compound. For natural product extracts, phylum and family name replaces chemical name. For searches run in databases of molecular target data, the molecular target ID replaces NSC number and this is linked to a description page for the target that contains the name(s) of the principal investigator(s), links to publications describing the measurement of the target, a brief description of the methodology, links to a display of the target data for all cell lines and links to other sites with possibly useful information, including Genecards [13], OMIM [14], and UniGene [15]. A COMPARE report run against the microarray database is similar to the molecular target report, but it also has a link for 3′ accession number. This is the GenBank accession number for the sequence from the 3′ end of the clone used to produce the cDNA on the microarray. Clicking on the link will take the user to the GenBank sequence for this clone. The description column in the report provides information on each clone, based on information from UniGene [15] clustering. All gene assignments in this column

COMPARE Results for S649890, Flavopiridol

Values for the National Cancer Institute Human Tumor Cell Line Screen

This calculation provided by the Developmental Therapeutics Program, DCTD, NCI

GI_{50} Results from the Synthetic Compound (August 2000 Release) Database
Searched 38680 samples
Minimum Number of Cell Lines in the Calculation 40
Minumum Variance of Hit data: 0.02
Cell Lines with the HIGHEST input values are the most inhibited

Rank	NSC	Compound Name	High Conc.	Average GI_{50}	# cell lines	Variance	# of Exp	PCC
1	649890	Flavopiridol	1.0E−06 M	7.5E−08 M	60	0.081	2	1.000
2	649890	Flavopiridol	1.0E−04 M	4.2E−08 M	60	0.059	9	0.696
3	666096	Olomoucine	1.0E−04 M	5.2E−05 M	60	0.036	5	0.670
4	700694		1.0E−04 M	7.4E−06 M	60	0.074	2	0.660
5	85561		1.0E−04 M	2.2E−05 M	51	0.189	1	0.655

Fig. 3. Typical result page from a COMPARE calculation.

should be considered tentative. The description is updated periodically to incorporate new clustering information from UniGene.

3. Use of COMPARE

One way to use COMPARE is to suggest possible mechanism(s) of action for compounds of interest. The compound is tested in the NCI screen and the pattern of activity is used as a seed to search for correlations with compounds with known mechanism of action. One interesting example of this use of COMPARE is the identification of MAP–kinase–kinase as a substrate for the proteolytic activity for anthrax lethal factor [16]. When the pattern of activity for anthrax lethal factor (NSC 678519) was used as a seed, the most striking correlation was to PD 098059 (NSC 679828, Fig. 4), a known inhibitor of MAP–kinase–kinase [17]. This suggested the hypothesis that anthrax lethal factor was eliminating MAP–kinase–kinase activity by degrading the enzyme, a hypothesis that was confirmed in the laboratory [16]. Important information about the mechanism of action of azatoxin (Fig. 5) was also found by using COMPARE in this manner. Azatoxin was rationally designed to inhibit topoisomerase II activity [18], but COMPARE calculations suggested substantial activity due to disruption of tubulin polymerization, which was then demonstrated experimentally [19]. In subsequent work, COMPARE calculations were used to help guide development of analogs with only the topoisomerase inhibition activity [20].

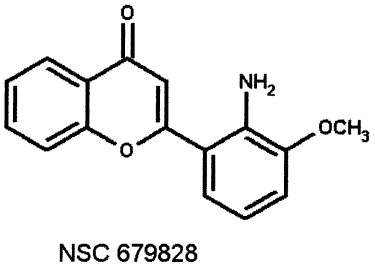

Fig. 4. Chemical structure of PD 098059.

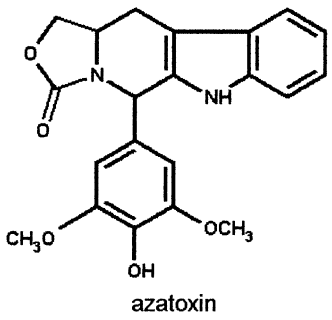

Fig. 5. Chemical structure of azatoxin.

COMPARE can also be used to find novel compounds that act by a mechanism of interest. This is accomplished by running COMPARE with data from a compound with the mechanism of interest as the seed and looking in the results for compounds with novel structures that can be tested for the activity of interest. The first example of this approach was the identification of a number of structurally diverse inhibitors of tubulin polymerization [21]. This approach has also been successful in finding novel inhibitors of IMP dehydrogenase [22], dihydroorotate dehydrogenase [23,24], topoisomerase I [25], and cyclin-dependent kinases (Fig. 6) [26]. These results required testing not more than a few dozen compounds in the specific target assay. Use of COMPARE allowed useful results to be obtained despite testing less than 1/1000 of the available compounds.

A number of laboratories have measured the level or activity of some molecular target in the cell lines in the NCI screen, then used COMPARE to identify novel compounds that interact with the target. Wosikowski et al. [27] measured epidermal growth factor receptor (EGFR) mRNA, which encodes a receptor tyrosine kinase, in each of the 60 cell lines. This pattern was used as a COMPARE seed to query the compound database. A number of compounds were identified for which sensitivity to the compound correlated with high levels of EGFR mRNA. These compounds were then tested for their ability to inhibit EGFR autophosphorylation in vitro. From the 15 candidate compounds tested, four were shown to inhibit EGFR kinase activity. Lee et al. [28] and Alvarez et al. [29] used functional and mRNA assays to characterize MDR-1/P-glycoprotein in the 60 cell panel. Running COMPARE with this data as a seed identified a large number of compounds for which high levels of expression of this drug efflux pump correlated with resistance to the compound. Further experiments confirmed that these compounds were indeed substrates for the P-glycoprotein efflux pump, providing evidence for a broader range of substrates than previously known. Low expression of the *nm23* gene has been associated with a highly metastatic phenotype and Freije et al. [30] measured Nm23 RNA levels in the cell lines in the NCI screen and used COMPARE to identify a number of compounds that preferentially inhibit cells with low Nm23 expression. This is an example of how COMPARE allows drug discovery efforts to begin even when the structure and function of the target are not clearly known.

4. Summary

Even with the advances in structure-based discovery and target-based assays, empirical assays can be very valuable in drug discovery efforts. The use of the COMPARE algorithm to analyze data from the NCI human tumor cell line screen has been an important part of a variety of drug discovery and development projects. The method and all associated data are freely available on the web and thus can widely serve the drug discovery community, both as a tool and

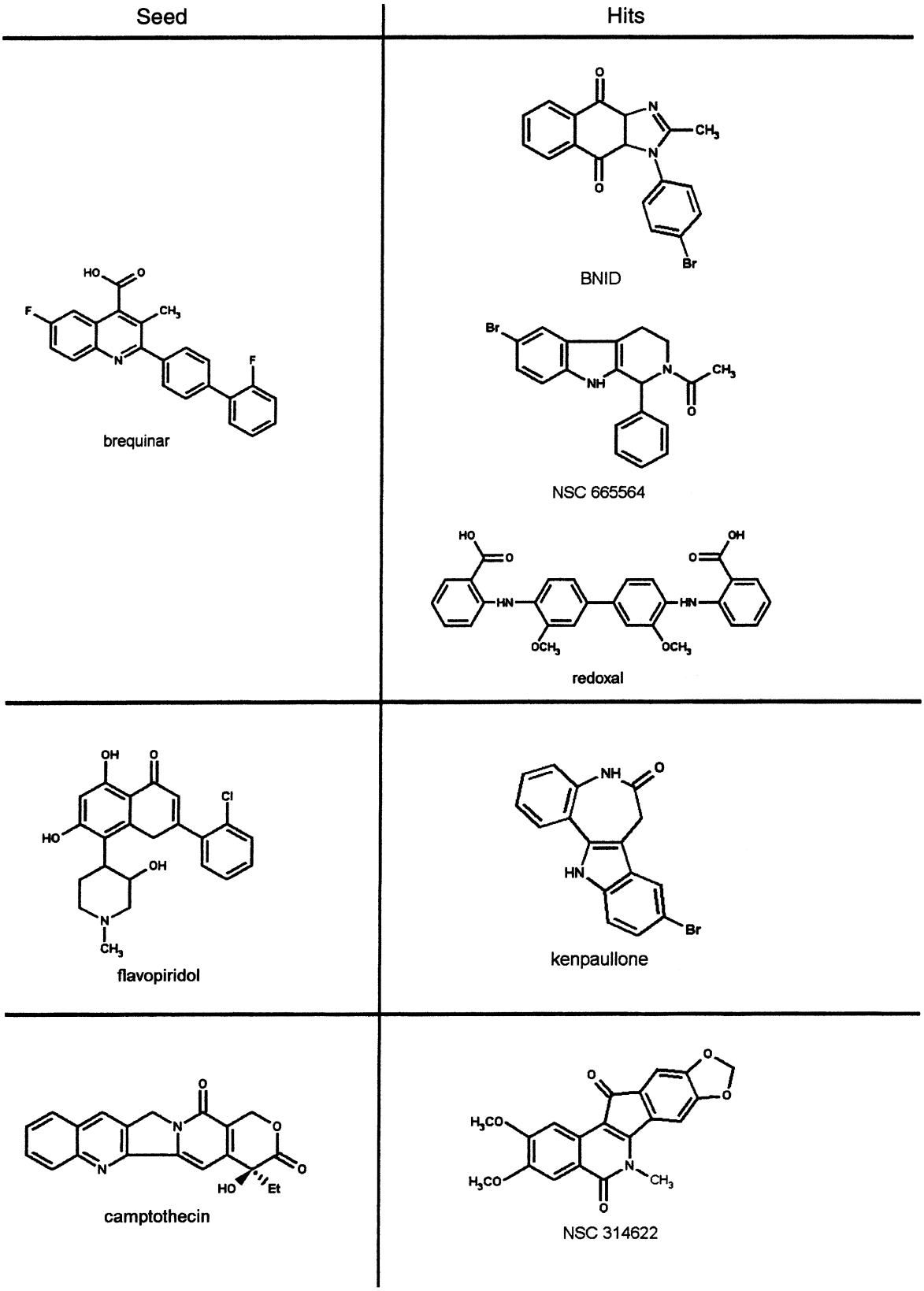

Fig. 6. Chemical structures of novel inhibitors identified in several COMPARE projects. The cell screening results for the seed structure were used to search the database and the hits are novel inhibitors of dihydrooratate dehydrogenase (brequinar), cyclin-dependent kinase (flavopiridol) and topoisomerase I (camptothecin).

as a starting point for further development of data mining strategies.

References

[1] P. Skehan, R. Storeng, D. Scudiero, J.B. McMahon, D. Vistica, J.T. Warren, H.R. Bokesch, S. Kenney, M.R. Boyd, New colorimetric cytotoxicity assay for anticancer-drug screening, J. Natl. Cancer Inst. 82 (1990) 1107–1112.

[2] A. Monks, D. Scudiero, P. Skehan, R. Shoemaker, K. Paull, D. Vistica, C. Hose, J. Langley, P. Cronise, A. Vaigro-Wolff, M. Gray-Goodrich, H. Campbell, M.R. Boyd, Feasibility of a high-flux anticancer drug screen utilizing a diverse panel of human tumor cell lines in culture, J. Natl. Cancer Inst. 83 (1991) 757–766.

[3] M.R. Boyd, K. Paull, L.R. Rubinstein, Data display and analysis strategies for the NCI disease-oriented in vitro antitumor drug screen, in: F.A. Valeriote, T. Corbett, L. Baker (Eds.), Cytotoxic Anticancer Drugs: Models and Concepts for Drug Discovery and Development, Kluwer Academic Publishers, Hingham, MA, 1992, pp. 11–34.

[4] K. Paull, R.H. Shoemaker, L. Hodes, A. Monks, D.A. Scudiero, L. Rubinstein, J. Plowman, M.R. Boyd, Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm, J. Natl. Cancer Inst. 81 (1989) 1088–1092.

[5] K. Paull, E. Hamel, L. Malspeis, The prediction of biochemical mechanism of action from the in vitro antitumor screen of the National Cancer Institute, in: W.O. Foye (Ed.), Cancer Chemotherapeutic Agents, American Chemical Society Books, Washington, DC, 1995, pp. 9–45, http://dtp.nci.nih.gov/docs/compare/compare.html.

[6] http://dtp.nci.nih.gov/docs/cancer/searches/standard_agent.html.

[7] How the compounds were selected and policies and procedures for requesting copies of this library, http://dtp.nci.nih.gov/branches/dscb/diversity_explaination.html.

[8] Description of the natural product repository and policies and procedures, http://dtp.nci.nih.gov/branches/npb/repository.htm.

[9] http://dtp.nci.nih.gov/mtargets/mt_index.html.

[10] U. Scherf, D.T. Ross, M. Waltham, L.H. Smith, J.K. Lee, L. Tanabe, K.W. Kohn, W.C. Reinhold, T.G. Myers, D.T. Andrews, D.A. Scudiero, M.B. Eisen, E.A. Sausville, Y. Pommier, D. Botstein, P.O. Brown, J.N. Weinstein, A gene expression database for the molecular pharmacology of cancer, Natl. Genet. 24 (3) (2000) 36–44.

[11] D.T. Ross, U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J.C. Lee, D. Lashkari, D. Shalon, T.G. Myers, J.N. Weinstein, D. Botstein, P.O. Brown, Systematic variation in gene expression patterns in human cancer cell lines, Natl. Genet. 24 (3) (2000) 227–235.

[12] http://dtp.nci.nih.gov/webdata.html.

[13] http://bioinformatics.weizmann.ac.il/cards/.

[14] http://www.ncbi.nlm.nih.gov/Omim/.

[15] http://www.ncbi.nlm.nih.gov/UniGene/index.html.

[16] N.S. Duesbery, C.P. Webb, S.H. Leppla, V.M. Gordon, K.R. Klimpel, T.D. Copeland, N.G. Ahn, M.K. Oskarsson, K. Fukasawa, K. Paull, G.F. Vande Woude, Proteolytic inactivation of MAP–kinase–kinase by anthrax lethal factor, Science 280 (5364) (1998) 734–737.

[17] D.T. Dudley, L. Pang, S.J. Decker, A.J. Bridges, A.R. Saltiel, A synthetic inhibitor of the mitogen-activated protein kinase cascade, Proc. Natl. Acad. Sci. U.S.A. 92 (17) (1995) 7686–7689.

[18] F. Leteurtre, J. Madalengoitia, A. Orr, T.J. Guzi, E. Lehnert, T. Macdonald, Y. Pommier, T.J. Cuzi, Rational design and molecular effects of a new topoisomerase II inhibitor, azatoxin, Cancer Res. 52 (16) (1992) 4478–4483.

[19] E. Solary, F. Leteurtre, K. Paull, D. Scudiero, E. Hamel, Y. Pommier, Dual inhibition of topoisomerase II and tubulin polymerization by azatoxin: a novel cytotoxic agent, Biochem. Pharmacol. 45 (12) (1993) 2449–2456.

[20] F. Leteurtre, D.L. Sackett, J. Madalengoitia, G. Kohlhagen, T. MacDonald, E. Hamel, K. Paull, Y. Pommier, Azatoxin derivatives with potent and selective action on topoisomerase II, Biochem. Pharmacol. 49 (9) (1995) 1283–1290.

[21] K. Paull, C.M. Lin, L. Malspeis, E. Hamel, Identification of novel antimitotic agents acting at the tubulin level by computer-assisted evaluation of differential cytotoxicity data, Cancer Res. 52 (1992) 3892–3900.

[22] K. Gharehbaghi, K. Paull, J.A. Kelley, J.J. Barchi Jr., V.E. Marquez, D.A. Cooney, A. Monks, D. Scudiero, K. Kohn, H.N. Jayaram, Cytotoxicity and characterization of an active metabolite of benzamide riboside: a novel inhibitor of IMP dehydrogenase, Int. J. Cancer 56 (6) (1994) 892–899.

[23] E.S. Cleaveland, A. Monks, A. Vaigro-Wolff, D.W. Zaharevitz, K. Paull, K. Ardalan, D.A. Cooney, H. Ford Jr., Site of action of two novel pyrimidine biosynthesis inhibitors accurately predicted by the COMPARE program, Biochem. Pharmacol. 49 (7) (1995) 947–954.

[24] E.S. Cleaveland, D.W. Zaharevitz, J.A. Kelley, K. Paull, D.A. Cooney, H. Ford Jr., Identification of a novel inhibitor (NSC 665564) of dihydroorotate dehydrogenase with a potency equivalent to brequinar, Biochem. Biophys. Res. Commun. 223 (3) (1996) 654–659.

[25] G. Kohlhagen, K. Paull, M. Cushman, P. Nagafuji, Y. Pommier, Protein-linked DNA strand breaks induced by NSC 314622: a novel non-camptothecin topoisomerase I poison, Mol. Pharmacol. 54 (1998) 50–58.

[26] D.W. Zaharevitz, R. Gussio, M. Leost, A.M. Senderowicz, T. Lahusen, C. Kunick, L. Meijer, E.A. Sausville, Discovery and initial characterization of the paullones: a novel class of small-molecule inhibitors of cyclin-dependent kinases, Cancer Res. 59 (11) (1999) 2566–2569.

[27] K. Wosikowski, D. Schuurhuis, K. Johnson, K. Paull, T.G. Myers, J.N. Weinstein, S.E. Bates, Identification of epidermal growth factor receptor and c-erbB2 pathway inhibitors by correlation with gene expression patterns, J. Natl. Cancer Inst. 89 (20) (1997) 1505–1515.

[28] J.-S. Lee, K. Paull, M. Alvarez, C. Hose, A. Monks, M. Grever, A.T. Fojo, S.E. Bates, Rhodamine efflux patterns predict P-glycoprotein substrates in the National Cancer Institute drug screen, Mol. Pharmacol. 46 (4) (1994) 627–638.

[29] M. Alvarez, K. Paull, A. Monks, C. Hose, J.-S. Lee, J. Weinstein, M. Grever, S. Bates, T. Fojo, Generation of a drug resistance profile by quantitation of MDR-1/P-glycoprotein in the cell lines of the National Cancer Institute Anticancer Drug Screen, J. Clin. Invest. 95 (5) (1995) 2205–2214.

[30] J.M. Freije, J.A. Lawrence, M.G. Hollingshead, A. De la Rosa, V. Narayanan, M. Grever, E.A. Sausville, K. Paull, P.S. Steeg, Identification of compounds with preferential inhibitory activity against low Nm23 expressing human breast carcinoma and melanoma cell lines, Natl. Med. 3 (4) (1997) 395–401.