ELSEVIER

European Journal of Cancer 40 (2004) 785–793

European Journal of Cancer

www.ejconline.com

Update on NCI *in vitro* drug screen utilities

S.L. Holbeck*

National Cancer Institute, Developmental Therapeutics Program, Information Technology Branch, Rockville, MD, USA

Received 1 October 2003; received in revised form 20 October 2003; accepted 17 November 2003

Abstract

Development of new anti-cancer drugs is a costly and risky proposition. The Developmental Therapeutics Program (DTP) of the National Cancer Institutes of the United States (U.S.) facilitates the drug development process by providing access to preclinical screening services. Since the early 1990's, DTP has screened tens of thousands of compounds against a panel of 60 human tumour cell lines representing nine tissue sites. At the same time, DTP began to accumulate information on the expression of molecular entities in the same 60 cell line panel. Many of these data are freely available to the public at http://dtp.nci.nih.gov. More recently, additional, more focused screens have entered the picture, with data also available through the web site. These include screening of roughly 100 000 compounds against a panel of yeast mutants, and screening of the NCI Diversity Set in assays designed to detect effects on Molecular Targets of interest.
Published by Elsevier Ltd.

Keywords: Neoplasms; Antineoplastic agents; Computational biology; Molecular biology; Proteomics; Genomics; Pharmacogenetics; Human Genome Project; Genetic Research; (Therapies, Investigational); *Saccharomyces cerevisiae*

1. Introduction

The development of anti-cancer drugs is an expensive and time-consuming process. The Developmental Therapeutics Program (DTP) of the United States (U.S.) National Cancer Institute (NCI) reduces the risks in this process by providing *in vitro* and *in vivo* screening services, as well as access to pharmacological and formulation resources. Just as valuable is the publicly available information on the data derived from these screens. This review will focus on the data and information analysis tools that DTP provides for the *in vitro* screens. Other articles in this issue will focus on *in vivo* testing and late preclinical resources provided by DTP.

Compound screening at DTP has focused on the response of a panel of 60 human tumour cell lines, with data on tens of thousands of compounds. An ongoing programme characterises expression of molecular targets within this panel. Nearly 100 000 compounds were analysed in a collaboration with the Fred Hutchinson Cancer Research Center for their ability to inhibit the growth of a panel of yeast strains with alterations in cancer-relevant genes. Screening campaigns were conducted for compounds affecting several molecular targets of interest. All of these data are freely available through a web site maintained by the NCI-DTP at http://dtp.nci.nih.gov/.

2. Compounds submitted to NCI-DTP

The acquisition of compounds for screening by the NCI began in 1955, and continues to this day, with over 500 000 compounds currently registered. This collection contains compounds from a large number of suppliers, including scientists in academia or government laboratories, as well as small biotechnology companies and large pharmaceutical companies. Researchers from over 100 countries have submitted compounds to the NCI screening programmes. Roughly half of the compounds were submitted under NCI's discreet screening agreement, which precludes NCI from disseminating data on these compounds. For the remainder, the data is publicly available through the DTP web site (http://dtp.nci.

* National Cancer Institute, Developmental Therapeutics Program, Information Technology Branch, 6130 Executive Blvd. Room 8014, Bethesda, MD 20892-7444, USA. Tel.: 301-435-9178; fax: 301-480-4808.

E-mail address: holbecks@mail.nih.gov (S.L. Holbeck).

0959-8049/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.ejca.2003.11.022

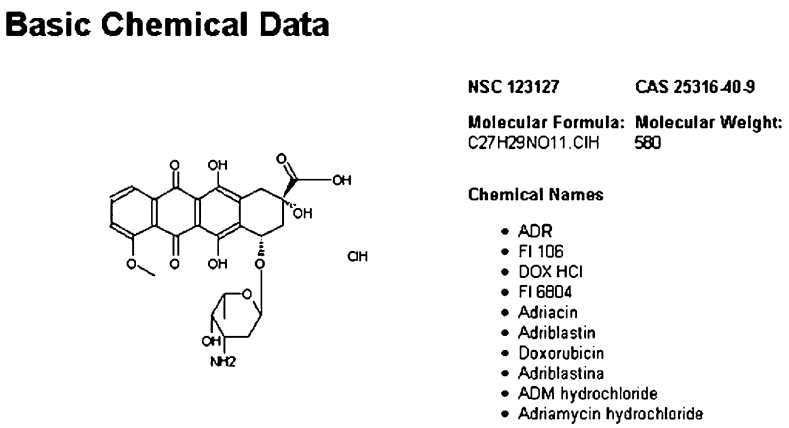

Fig. 1. The results of a search at http://dtp.nci.nih.gov/docs/dtp_search.html for chemical information on doxorubicin (Adriamycin). The web page displays a two-dimensional structure, NSC number, CAS number, Molecular Formula, Molecular Weight, and chemical names by which a compound is known.

nih.gov/), including compound structures, data from the NCI 60 human tumour cell line screen, the NCI acquired immunodeficiency syndrome (AIDS) screen and the NCI Yeast Anticancer Drug Screen. Those interested in submitting compounds for testing may do so using DTP's on-line submission form (http://dtp.nci.nih.gov/docs/misc/common_files/submit_compounds.html).

While the DTP compound collection contains many compounds with an identified mechanism of action, for the vast majority the target remains to be identified. Many were designed as chemically interesting structures, or are purified natural products. Some compounds were designed to interact with particular cellular targets. Many others are structural analogues of compounds with a known mechanism. The resulting collection is diverse, with a broad range of chemotypes and biological activities represented. Several plated sets have been developed to allow researchers to exploit this diversity in novel drug screens. The Diversity Set is a group of 1990 compounds selected to represent a broad range of chemotypes. The utility of this set is demonstrated by the success of several novel targeted screens [4,35,8,23]. The Mechanistic Diversity Set is comprised of 879 compounds chosen based on diversity of activity in the 60 cell line screen. Recently, a plated set of 235 purified natural products was developed. A large plated set contains 140 000 compounds. Before investing significant resources in lead development, researchers may wish to verify the structure, as DTP does not routinely perform chemical analyses on compounds, and some samples may degrade during storage.

The DTP web site provides basic chemical data on over 250 000 compounds. This includes (if available) two-dimensional and three-dimensional structures, CAS number, molecular formula, molecular weight and chemical names. Researchers may search for compound data using the NSC number (the NCI's identification number for a compound), by CAS number, by chemical name or by chemical structure at http://dtp.nci.nih.gov/docs/dtp_search.html. Fig. 1 displays the results of a search for doxorubicin (Adriamycin) (NSC 123127).

3. NCI 60 Human tumour cell line panel

In 1989, the NCI-DTP initiated an *in vitro* screen for potential anti-cancer drugs utilising a panel of 60

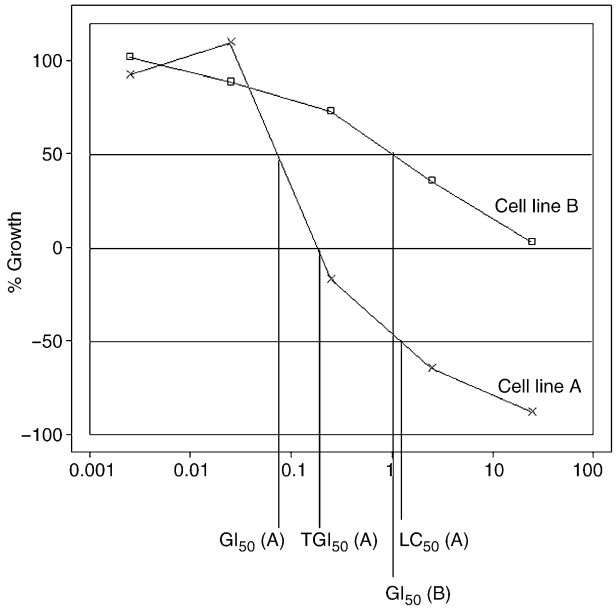

Fig. 2. Three endpoints (negative $\log_{10}$ of the concentration inhibiting the growth of 50% of the cells (GI_{50}), total growth inhibition (TGI) and negative $\log_{10}$ concentration need to kill 50% of the cells (LC_{50}) are calculated from 5-log dose response curves for compounds tested in the National Cancer Institute (NCI) 60 human tumour cell line screen. In this example, cell line A is much more sensitive than cell line B, with a GI_{50} value roughly 10-fold lower. Cell line B never reaches the TGI and LC_{50} endpoints in this concentration range. For the purposes of COMPARE, these endpoints are assigned the maximum concentration tested.

human tumour cell lines derived from various tissue types [27]. Compounds are tested over a 5-log concentration range against each of the 60 cell lines for their ability to inhibit the growth of, or to kill, the cells in a 2-day assay. Fig. 2 displays a simplified dose-response curve showing the response of two cell lines to doxorubicin. To facilitate analysis of the data, three endpoints are calculated for each cell line. The GI_{50} value is the negative $\log_{10}$ of the concentration required to inhibit the growth of that cell line by 50% (relative to untreated cells). TGI is the negative $\log_{10}$ minimum concentration that causes total growth inhibition, and LC_{50} reflects the negative $\log_{10}$ concentration needed to kill 50% of the cells. Paull and colleagues [34] developed the Mean Graph as a way of easily visualising the results from all 60 cell lines at once. A sample is shown in Fig. 3. To generate these graphs, the mean of each of the endpoints across all 60 cell lines is calculated. For each cell line, the difference between the GI_{50} for that cell line and the mean GI_{50} across all cell lines is calculated. When these differences are graphed, it becomes apparent at a glance which cell lines are more sensitive (those with bars deflecting to the right of the mean), and which cell lines are less sensitive (bars deflecting to the left). This screening service is available free of charge. To date, roughly 85 000 compounds have been screened. 60 cell line data for 43 000 compounds are currently available to the public through the DTP web site.

4. COMPARE

Compounds with similar mechanisms of actions tend to have similar patterns of growth inhibition in the 60 cell line screen, i.e. the same set of cell lines will tend to be more sensitive to both compounds, with a different subset being less sensitive to both. This can be visualised as compounds having similar Mean Graphs. To capitalise on this, Paull and colleagues [34] developed the COMPARE algorithm, which can be thought of as quantitating the similarity of Mean Graphs from different compounds. This algorithm takes the pattern for a "seed" compound and calculates Pearson correlation coefficients (PCCs) for it against each of the thousands of other compounds in the database, then returns a list of the highest correlations. Thus, one can start with a compound of unknown mechanism and determine whether it behaves similarly to compounds of defined mechanism that have been through the screen previously. This approach has been utilised to identify

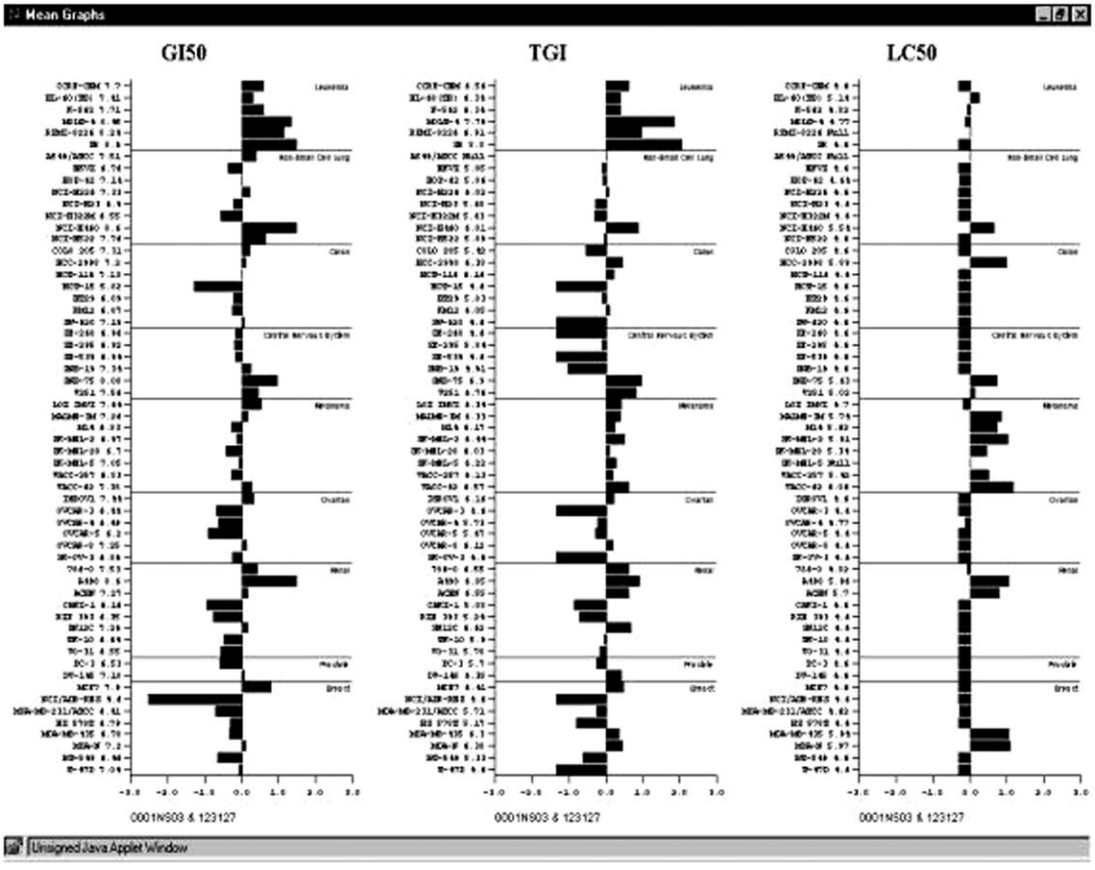

Fig. 3. A mean graph for doxorubicin is shown. The midline of each portion of the graph represents the mean for that endpoint, calculated across all 60 cell lines. This mean value is then subtracted from the value for each individual cell line and plotted. Cell lines more sensitive to doxorubicin are visualised as bars deflecting to the right, while more resistant cell lines have bars extending to the left of the mean.

novel tubulin-interacting compounds, topoisomerase poisons, vacuolar-type (H+)-adenosine triphosphatase (ATPase), and dihydroorotate dehydrogenase inhibitors [6,54,10,25,40]. Results may be unexpected, as with a series of Lavendustin A analogues submitted to the screen as having kinase inhibition activity, but were suggested by COMPARE to be possible tubulin inhibitors, a result that was experimentally verified [29]. Alternatively, one may wish to identify new chemotypes with a mechanism similar to that of a pharmaceutically intractable compound, in the hopes of identifying agents that might lack the undesirable features. In this case, one uses the compound of known mechanism as a seed, and COMPARE will return a list of compounds with the best correlations—these become candidates for testing to determine whether they do indeed share a common mechanism with the seed. Such an approach was used to identify a novel class of cyclin-dependent kinase (CDK) inhibitors, the paullones, starting with flavopiridol as a seed [57].

COMPARE can be accessed through the DTP web site. Currently, the old COMPARE interface is still operational, although no longer supported. The new COMPARE interface may be found at http://itbwork.nci.nih.gov/PublicServer/servlet/CompareServer, and allows users to utilise either compound or Molecular Target data (described in detail in the next section) as a seed for running COMPARE against either compound or Molecular Target databases. In addition, users can create their own seed data. Since all 60 cell line screening data, even on compounds not covered by a confidentiality agreement, is kept confidential for 2 years, the ability to create one's own seed data allows suppliers access to COMPARE for recently screened compounds. In addition, as data for compounds that are covered by a confidentiality agreement is not available through the public web site, this allows suppliers to create a seed to access COMPARE for their own confidential compounds. One might also create a seed to create composite patterns, such as co-expression of several molecular targets.

5. Molecular Targets in the 60 cell line panel

The sensitivity of a cell line to a compound is necessarily determined by the cellular environment—which genes are being expressed, which signalling pathways are turned on or off, whether various repair pathways are operational, etc. For some compounds, a single component may be a major determinant of sensitivity, while for others many components contribute to the response. In order to address this, DTP has an ongoing programme to characterise "Molecular Targets" within the 60 cell line panel. Molecular Targets in this context are used to denote measurable entities in the cell lines. The inclusion of a particular Target is driven by interest from the cancer research community, with most of the data provided by researchers outside of DTP. Researchers apply to measure a target (or targets) of interest. If approved, DTP provides them with 60 cell line materials at no cost, with the proviso that data may be posted on the DTP web site once the researcher has had an opportunity to publish the results. Interested parties can find more information on this programme at http://dtp.nci.nih.gov/mtargets/mt_index.html.

Data currently available include mutation status of genes important in cancer (including *p53* and the *Ras* genes), as well as quantitation of proteins within the cells (e.g. cyclins and CDKs), RNA levels (e.g. for many tyrosine kinases and phosphatases), and enzyme activity (e.g. DT-diaphorase and multidrug resistant (MDR) activities). Table 1 summarises the Molecular Targets for which data are currently available to the public. Microarray data, measuring the baseline expression of thousands of genes, is also available from two separate experiments. The first utilised a cDNA array to analyse baseline expression of 9706 entities in each of the 60 cell lines, relative to expression in a pool of 12 of these cell lines [36,38]. A second microarray experiment performed at Millennium Pharmaceuticals analysed baseline expression of 5365 features in each of the 60 cell lines utilising oligonucleotide arrays.

The COMPARE algorithm, in additional to identifying compounds with similar growth inhibitory patterns, can be used with Molecular Target data. Target data can be visualised in a Mean Graph, just as for a compound. COMPARE can be used to identify positive correlations, where cell lines with higher levels of a Target tend to be more sensitive to a compound. While negative correlations are not generally used when comparing compounds with other compounds, it can make sense in comparing compounds to Targets. This would identify cases where cell lines with lower levels of the Target tend to be more sensitive to the compound, or, stated differently, where cell lines with higher levels of the Target tend to be resistant to the compound. COMPARE can also be used to identify Targets that are co-expressed (positively correlated) with other Targets, as well as those that tend not to be expressed simultaneously (negative correlations).

The value of using COMPARE to identify compound mechanisms has been validated by several success stories. Kubo and colleagues [22] analysed p16 status in the 60 cell lines and tested compounds that COMPARE determined to have the highest correlations. They demonstrated that one of those compounds had activity against CDK4, an enzyme that is regulated by p16. Wosikowski and colleagues [52] measured RNA levels for several genes in the epidermal growth factor receptor (EGFR) and ErbB2 pathways. By screening a small number of compounds suggested by COMPARE, they

identified 14 compounds with demonstrable activity against these pathways. Sensitivity of cells to an inhibitor of brain glycogen phosphorylase correlates with expression of this target [39]. Cluster analysis with homoharringtonine against the microarray database suggested a correlation with EGFR expression, which was verified experimentally [13].

In most cases, the correlations between compound and target data is more modest than between two compounds. This may be because few Molecular Targets are

Table 1
Molecular targets measured in the NCI 60 cell line panel

88 KINASES, INCLUDING	**CELL CYCLE CONTROL**	**REDOX**
ABL	Cdc2	ALDH1 & ALDH3
AXL	Cdc25A, Cdc25B, Cdc25C	aldose reductase
CSF1R	cdc7	Cytochrome b5 reductase
CSK	Cdks 2, 4, 5, 6	Cytochrome P450 activity
EGFR	Chk1	Cytochrome P450 reductase
EphA1-8	Cyclins A, B, D1, D2, D3, E	Dihydrodiol dehydrogenase
EphB1-5	MDM2	DT-diaphorase
ErbB2-4	p16	GGCS
FGFR1-4	p19	GGT
FGR	p53	Glutathione
FLT1	Rb	GSTs A1, M1a, M1b, M2, M3,-mu,-pi
FLT4	Waf-1	HSI reductase
FYN		Thioredoxin
IGF1R		Thioredoxin reductase
INSR		
JAK1-2	**APOPTOSIS**	**SIGNAL TRANSDUCTION**
KDR		
KIT	AIF	FGF2
LCK	Bad	K-Ras, K-Ras, N-Ras
LYN	Bag1	TGF-alpha
MET	Bax	
PDGFRa	Bcl2	
PDGFRb	Bcl-xl	
RET	Bid	
SRC	c-IAP	
TRKA-C	FAP-1	
YES	Survivin	
ZAP70	x-IAP	
35 PHOSPHATASES, INCLUDING	**DNA REPAIR**	**OTHERS**
BAS	ERCCs 1, 2, 3	Actins beta & gamma
BDP1	Gadd45	BCRP
CD45	MGMT	Carcinoembryonic antigen
DEP-1	Mlh1	c-myc
ESP	Msh2	Desmin
GLEPP1	PCNA	Dihydropyrimidine dehydrogenase
IA2	Topo II alpha & beta	F1ATPase
IAR	XPA	Glycogen phosphorylase
MEG		HSC70
Meg2	**TRANSPORT**	HSP60
PCP-2		HSP90
PEST	ABC2	Laminin B
PTP	ABCAs 5, 6,12, 13	nm23
PTP-1B	ABCG5	Protein disulphide isomerase
SAP	ABCG8	Thymidine Kinase
SHP1	LRP	Thymidylate Synthase
SHP2	MDR activity	TRAG-3
STEP	MDR1	Tubulin beta
TC-PTP	MRP	Vimentin
ZPEP	RFC1	

PCNA, proliferating cell nuclear antigen; MDR1, multidrug resistance protein 1, TGF-alpha, transforming growth factor-alpha. This table summarises the results of Molecular Target measured in the NCI 60 human tumor cell line panel, not including data from microarray experiments. Targets are grouped according to function. Much of the data for kinase [7,9,52], cell cycle [3,22,32], apoptosis [3,19,21,26,43–45], DNA repair [3,31,46,55], transport [1,2,18,24,28,37,48,53], redox [12,41,47,51,56], signal transduction [9,20] and other targets [11,15,17,31,39,42] have been published. For all Targets, experimental details may be found by searching http://dtp.nci.nih.gov/mtargets/mt_index.html.

the single major influence on sensitivity. An exception to this is seen for Targets that influence the amount of active compound within the cell. For example, compounds that are substrates for the P-glycoprotein (Pgp) efflux pump show very strong negative correlations with Pgp activity, i.e. cells with high levels of Pgp activity are resistant to these drugs [1,2,53]. Thus, cellular components that influence the uptake or efflux of compounds may be found by a COMPARE analysis. A second category of Targets with a major influence on the amount of active compound within a cell are enzymes that can alter a compound, either activating a pro-drug, or destroying an active compound. An example of this is the positive correlations found between activity of the quinone-metabolising enzyme DT-diaphorase [14] and the sensitivity of the 60 cell line panel to EO9 and related compounds that are activated by this enzyme.

For cases where multiple cellular components influence drug sensitivity, COMPARE can still be of use. COMPARE with Diethyldithiocarbamate (NSC 4857), a reported inhibitor of nuclear factor (NF)-kappa B [30], as a seed yields modest correlations (0.35 to 0.51) with multiple genes involved in interleukin-1 (IL-1) signalling, including the IL-1 receptor and phospholipase C. Activation of the IL-1 pathway leads to the nuclear translocation of NF-kappa B.

Occasionally, the actual Molecular Target of a compound will be found by COMPARE. Sensitivity to an ErbB2 immunotoxin is dependent on expression of the gene, and yields a fairly high PCC of 0.54. Pasquale and colleagues [33] recently determined the ability of the 60 cell line panel to be infected by adeno-associated virus 5 (AAV5), and used those data as a seed to run COMPARE against the microarray dataset. One of the genes whose expression correlated with infectivity was the platelet-derived growth factor alpha, which they demonstrated to be the cellular receptor for the AAV5 virus.

6. NCI yeast anticancer drug screen

In the mid-1990's, a novel anti-cancer drug screen was initiated by Leland Hartwell and Stephen Friend at the Fred Hutchinson Cancer Research Center in Seattle, utilising a panel of *Saccharomyces cerevisiae* strains altered in DNA damage repair or cell cycle control genes. This project was begun as an NCI field station, and later converted into a contract managed by NCI-DTP. A more detailed review of this screen will be published elsewhere in Ref. [16]. Nearly 100 000 compounds were subjected to the initial stage (Stage 0) of this screen, against six strains at a single dose. Compounds meeting activity criteria were selected for further testing at two doses (Stage 1). The most promising compounds were selected for testing at five doses against 13 yeast strains (Stage 2). The data from all stages of the yeast screen are available at http://dtp.nci.nih.gov/yacds/index.html. One can search for data by NSC number, or browse Stage 0 and Stage 1 data by patterns of activity. Stage 2 data can be searched by NSC number, or one may browse through summaries of compounds with selective activity against a particular yeast strain. For example, compounds highly selective for strains with mutations in *rad50* and *rad52*, genes needed for repair of double-stranded DNA breaks, include known topoisomerase poisons such as 9-amino camptothecin (NSC 603071), which inhibits only these two strains over the ~2 log range of concentrations tested. In addition to compounds expected to introduce DNA breaks, several novel structures showed selectivity for these two strains. Two such compounds were demonstrated to interact with topoisomerase I [12]. Many of the Stage 2-tested compounds showed selectivity for the mitotic spindle checkpoint mutant *bub3*. While the DTP repository has a large number of tubulin-interactive compounds, which would be expected to trigger the spindle checkpoint, many compounds that interact with mammalian tubulin fail to bind yeast tubulin [5]. Thus, many of the *bub3*-selective hits are candidates for interfering with other components of the mitotic apparatus.

7. Downloadable datasets

For those interested in finding information on one or a few compounds, the DTP web site provides tools to retrieve and analyse data, as described in the preceding sections. Other researchers are interested in mining the large datasets. Most of the data on the DTP web site is available as downloadable datasets, and are in comma-delimited format that users can import into relational databases or spreadsheets and manipulate as desired. Thus, users can download 60 cell line data for approximately 43 000 compounds. This is organised into three files, each comprised of data for a single endpoint (GI_{50}, TGI or LC_{50}). Three additional files contain the data for just the Diversity Set. Two-dimensional structures for the 60 cell line tested compounds are available in MDL's SDFile format. In order to utilise the structural data, users will need to obtain a chemical software package. These datasets may be downloaded from http://dtp.nci.nih.gov/docs/cancer/cancer_data.html.

The Molecular Target data for the 60 cell line panel may be downloaded at http://dtp.nci.nih.gov/mtargets/download.html. There are currently three datasets available. There are two files with microarray data. The first presents data from a cDNA microarray experiment [36,38], analysing baseline mRNA expression in each of the 60 cell lines, relative to that in a pool of 12 of these cell lines. This dataset contains 9706 individual measurements for each of the cell lines. The second microarray

dataset contains data provided by Millennium Pharmaceuticals using Affymetrix oligonucleotide microarrays. This dataset consists of 5365 feature sets for which baseline mRNA expression was assayed in each of the 60 cell lines. For both of these microarray datasets, the gene assignments associated with each measurement are updated periodically utilising data from the human UniGene database of the National Center for Biotechnology Information. For the cDNA array data, the assignments are derived using the GenBank accession number for the sequence derived from the 3′ end of the cDNA. For the oligonucleotide array data, gene assignments are made using the GenBank accession number of the sequence used to design the oligonucleotides. A third dataset contains data determined from individual experiments (non-microarray data) and currently includes data from 255 measurements, including DNA mutation status, RNA expression, protein data and enzyme activity measurements.

Files containing structural information on several hundred thousand compounds may be downloaded at http://dtp.nci.nih.gov/docs/3d_database/structural_information/structural_data.html. Three-dimensional coordinates and Simplified Molecular Input Line Entry Specification (SMILES) strings are provided for several hundred thousand compounds. For those that are interested in particular subsets of compounds, separate files are available with structural data for compounds in the Diversity Set, the Mechanistic Set, the complete set of 140 000 plated compounds, as well as for compounds that were tested in the 60 cell line screen or in the AIDS screen. Chemical software is needed to utilise the information contained in these files.

Yeast screen data is provided in comma-delimited files for each of the 3 stages of the screen (http://dtp.nci.nih.gov/yacds/download.html). These files contain data for 87 264 compounds tested in the initial Stage 0 phase of the screen, 14 837 compounds that underwent further testing in Stage 1, and 2189 compounds that were selected for multi-dose testing in Stage 2. AIDS Antiviral Screen data for 43 905 compounds tested for ability to protect human CEM cells from infection by human immunodeficiency viris-1 (HIV-1) is available for download at http://dtp.nci.nih.gov/docs/aids/aids_screen.html [50].

Recently, DTP has begun to assemble these diverse datasets into a unified data warehouse. In this model, all assay results and chemical data are being merged into a single set of relational database tables. These tables are in a public Oracle instance, with a library of Java classes to facilitate data retrieval. Researchers will be able to write their own code to access the DTP data, and to seamlessly integrate it with their own data. Currently, the data warehouse contains GI_{50}, TGI and LC_{50} data from the 60 cell line screen (both the standard 2 day screen, as well as a non-routine 6 day assay), yeast screen data, data from the molecularly-targeted screens, chemical data, AIDS screen data and data from NCI's historical *in vivo* drug screen. The presentation of these data on the DTP web site is currently utilising the data warehouse architecture and the library of Java classes. Researchers interested in querying the data warehouse directly can find information at http://dtp.nci.nih.gov/dw/dw_main.html.

8. Conclusions

The DTP databases represent a tremendous resource available to the public, particularly to those interested in the drug discovery process. Suppliers submit their compounds to the 60 cell line screen for a variety of reasons. Some submit a series of compounds and utilise the resulting data to aid in selection of a lead compound for further development. Others with an interest in a particular type of cancer wish to determine whether a particular class of molecules has activity against human tumour cell lines derived from that tissue type. Some researchers have already done *in vitro* testing and seek to use the 60 cell line data for a COMPARE analysis, to help develop hypotheses about how the compound might function.

It is important to emphasise that while COMPARE has proven valuable in the generation of hypotheses, that any correlations need to be verified experimentally. There are some issues that arise frequently in doing a COMPARE analysis. Many of the compounds tested in the 60 cell line screen showed minimal activity, and were not tested further. The data for singly tested compounds should be scrutinised before investing resources in following up on such a correlation. Often compounds tested just once show little variability in the response of the cell lines. Other times, the pattern might appear to be dominated by one or a few cell lines, yielding correlation coefficients that may be quite high. While such results may be real, if the compound was only tested once, one does not know whether these results would be reproducible.

When running COMPARE with Molecular Target data, it is important to consider how a particular gene is regulated. This is particularly true in interpreting data from microarray experiments. RNA levels are not always well correlated with expression data for the protein it encodes [49]. Even with protein measurements, the relevant measure for interaction with a particular compound might be a modified form of the protein (i.e. phosphorylated, acetylated, acylated, etc.). A protein might localise to different locations within the cell, or might need to associate with other cellular components. These caveats make it particularly important to experimentally verify any correlations between compounds and Molecular Targets.

Since different researchers have differing goals for utilising the DTP databases, the web site provides multiple ways of accessing the datasets. Users may explore the data using the tools that DTP provides to retrieve and analyse particular pieces of data. For those who wish to develop their own analysis tools and techniques, the datasets can be downloaded and subjected to these protocols.

Acknowledgements

The author wishes to credit those who have performed the screening and developed many of the analysis tools described in this article. 60 cell line screening was performed in DTP's Screening Technology Branch, directed by Robert Shoemaker and Dominic Scudiero. Richard Klausner, Leland Hartwell, Stephen Friend, Julian Simon, Michele Cronk, Heather Dunstan, David Evans, John Lamb, and Philippe Szankasi contributed to the Yeast screen, which was performed at the Fred Hutchinson Cancer Research Center in Seattle. Data analysis tools and the DTP web site are developed by members of DTP's Information Technology Branch, including Daniel Zaharevitz, David Segal, Marie Hose, Mark Kunkel, Timothy Myers, Ken Paull and Penny Svetlik. Molecular Target data in the 60 cell line panel was generated by numerous researchers. Credit for each of these measurements is indicated on the DTP web site.

References

1. Alvarez M, Paull K, Monks A, *et al*. 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* 1995, **95**, 2205–2214 PMID: 7738186.
2. Alvarez M, Robey R, Sandor V, *et al*. Using the national cancer institute anticancer drug screen to assess the effect of MRP expression on drug sensitivity profiles. *Mol Pharmacol* 1998, **54**, 802–814.
3. Amundson SA, Myers TG, Scudiero D, Kitada S, Reed JC, Fornace Jr AJ. An informatics approach identifying markers of chemosensitivity in human cancer cell lines. *Cancer Res* 2000, **60**, 6101–6110 PMID: 11085534.
4. Blaskovich MA, Sun J, Cantor A, Turkson J, Jove R, Sebti SM. Discovery of JSI-124 (cucurbitacin I), a selective Janus kinase/signal transducer and activator of transcription 3 signaling pathway inhibitor with potent antitumor activity against human and murine cancer cells in mice. *Cancer Res* 2003, **62**, 1270–1279 PMID: 12649187.
5. Bode CJ, Gupta ML, Reiff EA, Suprenant KA, Georg GI, Himes RH. Epothilone and paclitaxel: unexpected differences in promoting the assembly and stabilization of yeast microtubules. *Biochemistry* 2002, **41**, 3870–3874.
6. Boyd MR, Farina C, Belfiore P, *et al*. Discovery of a novel antitumor benzolactone enamide class that selectively inhibits mammalian vacuolar-type (H+) ATPases. *J Pharmacol Exp Ther* 2001, **297**, 114–120 PMID: 11259534.
7. Budde RJ, Ke S, Levin VA. Activity of pp60c-src in 60 different cell lines derived from human tumors. *Cancer Biochem Biophys* 1994, **14**, 171–175 PMID: 7537173.
8. Bykov VJ, Issaeva N, Shilov A, *et al*. Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. *Nat Med* 2002, **8**, 282–288 PMID: 11875500.
9. Chandler LA, Sosnowski BA, Greenlees L, Aukerman SL, Baird A, Pierce GF. Prevalent expression of fibroblast growth factor (FGF) receptors and FGF2 in human tumor cell lines. *Int J Cancer* 1999, **81**, 451–458 PMID: 10209961.
10. Cleaveland ES, Monks A, Vaigro-Wolff A, *et al*. Site of action of two novel pyrimidine biosynthesis inhibitors accurately predicted by the compare program. *Biochem Pharmacol* 1995, **49**, 947–954 PMID: 7741767.
11. Duan Z, Feller AJ, Toh HC, Makastorsis T, Seiden MV. TRAG-3, a novel gene, isolated from a taxol-resistant ovarian carcinoma cell line. *Gene* 1999, **229**, 75–81 PMID: 10095106.
12. Dunstan HM, Ludlow C, Goehle S, *et al*. Cell-based assays for identification of novel double-strand break-inducing agents. *J Natl Cancer Inst* 2002, **94**, 88–94 PMID: 11792746.
13. Efferth T, Sauerbrey A, Halatsch ME, Ross DD, Gebhart E. Molecular modes of action of cephalotaxine and homoharringtonine from the coniferous tree Cephalotaxus hainanensis in human tumor cell lines. *Naunyn Schmiedebergs Arch Pharmacol* 2003, **367**, 56–67 PMID: 12616342.
14. Fitzsimmons SA, Workman P, Grever M, Paull K, Camalier R, Lewis AD. Reductase enzyme expression across the National Cancer Institute Tumor cell line panel: correlation with sensitivity to mitomycin C and EO9. *J Natl Cancer Inst* 1996, **88**, 259–269.
15. Grem JL, Danenberg KD, Behan K, *et al*. Thymidine kinase, thymidylate synthase, and dihydropyrimidine dehydrogenase profiles of cell lines of the National Cancer Institute's Anticancer Drug Screen. *Clin Cancer Res* 2001, **7**, 999–1009 PMID: 11309351.
16. Holbeck SL, Simon J. The FHCRC/NCI yeast anticancer drug screen. In Nitiss JL, Heitman J, eds. *Yeast as a Tool in Cancer Research*, in press.
17. Imai Y, Nakane M, Kage K, *et al*. C421A polymorphism in the human breast cancer resistance protein gene is associated with low expression of Q141K protein and low-level drug resistance. *Mol Cancer Ther* 2002, **1**, 611–616 PMID: 12479221.
18. Izquierdo MA, Shoemaker RH, Flens MJ, *et al*. Overlapping phenotypes of multidrug resistance among panels of human cancer-cell lines. *Int J Cancer* 1996, **65**, 230–237 PMID: 8567122.
19. Kitada S, Krajewska M, Zhang X, *et al*. Expression and location of pro-apoptotic Bcl-2 family protein BAD in normal human tissues and tumor cell lines. *Am J Pathol* 1998, **152**, 51–61 PMID: 9422523.
20. Koo HM, Monks A, Mikheev A, *et al*. Enhanced sensitivity to 1-beta-D-arabinofuranosylcytosine and topoisomerase II inhibitors in tumor cell lines harboring activated ras oncogenes. *Cancer Res* 1996, **56**, 5211–5216 PMID: 8912859.
21. Krajewska M, Zapata JM, Meinhold-Heerlein I, *et al*. Expression of Bcl-2 family member Bid in normal and malignant tissues. *Neoplasia* 2002, **4**, 129–140 PMID: 11896568.
22. Kubo A, Nakagawa K, Varma RK, *et al*. The p16 status of tumor cell lines identifies small molecule inhibitors specific for cyclin-dependent kinase 4. *Clin Cancer Res* 1999, **5**, 4279–4286.
23. Lazo JS, Aslan DC, Southwick EC, *et al*. Discovery and biological evaluation of a new family of potent inhibitors of the dual specificity protein phosphatase Cdc25. *J Med Chem* 2001, **44**, 4042–4049 PMID: 11708908.
24. Lee JS, Paull K, Alvarez M, *et al*. Rhodamine efflux patterns predict P-glycoprotein substrates in the National Cancer Institute drug screen. *Mol Pharmacol* 1994, **46**, 627–638.
25. Leteurtre F, Sackett DL, Madalengoitia J, *et al*. Azatoxin derivatives with potent and selective action on topoisomerase II. *Biochem Pharmacol* 1995, **49**, 1283–1290 PMID: 7763310.
26. Meinhold-Heerlein I, Stenner-Liewen F, Liewen H, *et al*. Expression and potential role of Fas-associated phosphatase-1 in ovarian cancer. *Am J Pathol* 2001, **158**, 1335–1344 PMID: 11290551.

27. Monks A, Scudiero D, Skehan P, *et al*. Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. *J Natl Cancer Inst* 1991, **83**, 757–766.
28. Moscow JA, Connolly T, Myers TG, Cheng CC, Paull K, Cowan KH. Reduced folate carrier gene (RFC1) expression and antifolate resistance in transfected and non-selected cell lines. *Int J Cancer* 1997, **72**, 184–190 PMID: 9212241.
29. Mu F, Coffing SL, Riese 2nd DJ, *et al*. Design, synthesis, and biological evaluation of a series of lavendustin A analogues that inhibit EGFR and Syk tyrosine kinases, as well as tubulin polymerisation. *J Med Chem* 2001, **44**, 441–452 PMID: 11462983.
30. Mulsch A, Schray-Utz B, Mordvintcev PI, Hauschildt S, Busse R. Diethyldithiocarbamate inhibits induction of macrophage NO synthase. *FEBS Lett* 1993, **321**, 215–218.
31. Myers TG, Anderson NL, Waltham M, *et al*. A protein expression database for the molecular pharmacology of cancer. *Electrophoresis* 1997, **18**, 647–653 PMID: 9150955.
32. O'Connor PM, Jackman J, Bae I, *et al*. Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents. *Cancer Res* 1997, **57**, 4285–4300 PMID: 9331090.
33. Pasquale GD, Davidson BL, Stein CS, *et al*. Identification of PDGFR as a receptor for AAV-5 transduction. *Nat Med* in press. PMID: 14502277.
34. Paull KD, Shoemaker RH, Hodes L, *et al*. 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* 1989, **81**, 1088–1092.
35. Rapisarda A, Uranchimeg B, Scudiero DA, *et al*. Identification of small molecule inhibitors of hypoxia-inducible factor 1 transcriptional activation pathway. *Cancer Res* 2002, **62**, 4316–4324 PMID: 12154035.
36. Ross DT, Scherf U, Eisen MB, *et al*. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000, **24**, 227–235.
37. Scheffer GL, de Jong MC, Monks A, *et al*. Increased expression of beta 2-microglobulin in multidrug-resistant tumour cells. *Br J Cancer* 2002, **86**, 1943–1950 PMID: 12085191.
38. Scherf U, Ross DT, Waltham M, *et al*. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000, **24**, 236–244.
39. Schnier JB, Nishi K, Monks A, Gorin FA, Bradbury EM. Inhibition of glycogen phosphorylase (GP) by CP-91,149 induces growth inhibition correlating with brain GP expression. *Biochem Biophys Res Commun* 2003, **309**, 126–134 PMID: 12943673.
40. Solary E, Leteurtre F, Paull KD, Scudiero D, Hamel E, Pommier Y. Dual inhibition of topoisomerase II and tubulin polymerization by azatoxin, a novel cytotoxic agent. *Biochem Pharmacol* 1993, **45**, 2449–2456 PMID: 8392342.
41. Sreerama L, Sladek NE. Class 1 and class 3 aldehyde dehydrogenase levels in the human tumor cell lines currently used by the National Cancer Institute to screen for potentially useful antitumor agents. *Adv Exp Med Biol* 1997, **414**, 81–94 PMID: 9059610.
42. Stinson SF, Alley MC, Kopp WC, *et al*. Morphological and immunocytochemical characteristics of human tumor cell lines for use in a disease-oriented anticancer drug screen. *Anticancer Res* 1992, **12**, 1035–1053 PMID: 1503399.
43. Takayama S, Krajewski S, Krajewska M, *et al*. Expression and location of Hsp70/Hsc-binding anti-apoptotic protein BAG-1 and its variants in normal tissues and tumor cell lines. *Cancer Res* 1998, **58**, 3116–3131 PMID: 9679980.
44. Tamm I, Kornblau SM, Segall H, *et al*. Expression and prognostic significance of IAP-family genes in human cancers and myeloid leukemias. *Clin Cancer Res* 2000, **6**, 1796–1803 PMID: 10815900.
45. Tamm I, Wang Y, Sausville E, Scudiero DA, Vigna N, Oltersdorf T, Reed JC. IAP-family protein survivin inhibits caspase activity and apoptosis induced by Fas (CD95), Bax, caspases, and anticancer drugs. *Cancer Res* 1998, **58**, 5315–5320 PMID: 9850056.
46. Taverna P, Liu L, Hanson AJ, Monks A, Gerson SL. Characterization of MLH1 and MSH2 DNA mismatch repair proteins in cell lines of the NCI anticancer drug screen. *Cancer Chemother Pharmacol* 2000, **46**, 507–516 PMID: 11138465.
47. Tew KD, Monks A, Barone L, *et al*. Glutathione-associated enzymes in the human cell lines of the National Cancer Institute Drug Screening Program. *Mol Pharmacol* 1996, **50**, 149–159 PMID: 8700107.
48. Vulevic B, Chen Z, Boyd JT, *et al*. Cloning and characterization of human adenosine 5′-triphosphate-binding cassette, sub-family A, transporter 2 (ABCA2). *Cancer Res* 2001, **61**, 3339–3347 PMID: 11309290.
49. Washburn MP, Koller A, Oshiro G, *et al*. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* 2003, **100**, 3107–3112.
50. Weislow OS, Kiser R, Fine DL, Bader J, Shoemaker RH, Boyd MR. New soluble-formazan assay for HIV-1 cytopathic effects: application to high-flux screening of synthetic and natural products for AIDS-antiviral activity. *J Natl Cancer Inst* 1989, **81**, 577–586.
51. Woo ES, Monks A, Watkins SC, Wang AS, Lazo JS. Diversity of metallothionein content and subcellular localization in the National Cancer Institute tumor panel. *Cancer Chemother Pharmacol* 1997, **41**, 61–68 PMID: 9443615.
52. Wosikowski K, Schuurhuis D, Johnson K, *et al*. Identification of epidermal growth factor receptor and c-erbB2 pathway inhibitors by correlation with gene expression patterns. *J Natl Cancer Inst* 1997, **89**, 1505–1515.
53. Wu L, Smythe AM, Stinson SF, *et al*. Multidrug-resistant phenotype of disease-oriented panels of human tumor cell lines used for anticancer drug screening. *Cancer Res* 1992, **52**, 3029–3034 PMID: 1350507.
54. Xia Y, Yang ZY, Xia P, *et al*. Antitumor Agents. 211. Fluorinated 2-phenyl-4-quinolone derivatives as antimitotic antitumor agents. *J Med Chem* 2001, **44**, 3932–3936 PMID: 11689079.
55. Xu Z, Chen ZP, Malapetsa A, *et al*. DNA repair protein levels vis-a-vis anticancer drug resistance in the human tumor cell lines of the National Cancer Institute drug screening program. *Anticancer Drugs* 2002, **13**, 511–519 PMID: 12045463.
56. Yu LJ, Matias J, Scudiero DA, *et al*. P450 enzyme expression patterns in the NCI human tumor cell line panel. *Drug Metab Dispos* 2001, **29**, 304–312 PMID: 11181500.
57. Zaharevitz DW, Gussio R, Leost M, *et al*. Discovery and initial characterization of the paullones, a novel class of small-molecule inhibitors of cyclin-dependent kinases. *Cancer Res* 1999, **59**, 2566–2569.