hTERT-Targeted Library

Medicinal and Computational Chemistry Dept., ChemDiv, Inc., 6605 Nancy Ridge Drive, San Diego, CA 92121 USA, Service: +1 877 ChemDiv, Tel: +1 858-794-4860, Fax: +1 858-794-4931, Email: ChemDiv@chemdiv.com

INTRODUCTION

Telomeres are located at the distal ends of the chromosomes and the shortening of which with successive cycles of cell division leads to cell senescence and cell death. Human telomerase, a cellular reverse transcriptase, is a ribonucleoprotein enzyme that catalyzes the synthesis and extension of telomeric DNA. It is composed of at least, a template RNA component (hTR; human Telomerase RNA) and a catalytic subunit, the telomerase reverse transcriptase (hTERT). Except germline cells, activated lymphocytes and some stem cell populations, most adult somatic cells do not express hTERT [1]. In cells where telomerase is activated, hTERT synthesizes a TTAGGG sequence from the RNA template that is then added to the end of the shortening chromosome [2], thus saving the cells from death. The above mechanism is cleverly exploited by tumor cells to maintain their immortality [3]. Together with its universal expression, hTERT represents an ideal target for cancer therapy [4,5].

Telomerase represents a prototype of a universal tumor antigen due to both its expression by the vast majority of tumors and its inherent functional involvement in oncogenic transformation. The absence of telomerase is associated with telomere shortening and aging of somatic cells, while high telomerase activity is observed in over 90% of human cancer cells, strongly indicating its key role during tumorigenesis [6]. Several details regarding telomere structure and telomerase regulation have already been elucidated, providing new targets for therapeutic exploitation. Given these attractive features, the identification of epitopes within hTERT, the catalytic subunit of telomerase, has led to the investigation of this tumor antigen as a broadly applicable immunological target [7]. Further support for anti-telomerase approaches comes from recent studies indicating that telomerase is endowed of additional functions in the control of growth and survival of tumor cells that do not depend only on the ability of this enzyme to maintain telomere length. This observation suggests that inhibiting telomerase or its synthesis may have additional anti-proliferative and apoptosis inducing effect, independently of the reduction of telomere length during cell divisions. Here we provide the basic information about the biology of telomeres and telomerase and attempt to present various approaches that are currently under investigation to inhibit its expression and its activity.

In the past decade, research in the field of telomerases has progressed tremendously, especially in relation to cellular immortality and carcinogenesis. As mentioned above, telomerase activation is observed in a vast majority of human cancers, irrespective of tumor type, while most normal tissues contain inactivated telomerase. The role and timing of telomerase activation in carcinogenesis has been revealed by telomerase-knockout mouse studies [8]. Significant telomere erosions and age- and generation-dependent increases in cytogenic abnormalities are exhibited in telomerase-knockout mice, providing evidence that telomere dysfunction with critically short telomeres causes genomic instability. This concept is further supported by studies using telomerase–/– p53–/– double-knockout mice. These mouse cells demonstrate high levels of genomic instability, exemplified by increases in both formation of dicentric chromosomes and susceptibility to oncogenic transformation. These mice exhibit significantly decreased tumor latency and overall survival. Thus, in the absence of genome checkpoint functions, telomere dysfunction accelerates genomic instability, facilitating cancer initiation [9]. According to this concept, the genomic instability caused by telomere dysfunction occurs in the early stages of carcinogenesis, before telomerase activation. Subsequently, telomeres in these initiated cells undergo further progressive shortening, generating rampant chromosomal instability and threatening cell survival. Telomerase activation necessarily occurs at this stage to stabilize the genome and confer unlimited proliferative capacity upon the emerging and evolving cancer cell. In other words, cells that have acquired telomerase activity can obtain the capacity for cancer progression. Eventually, most cancer cells exhibit telomerase activity. This cancer-specific telomerase activity provides an opportunity for us to utilize it for the design of target-specific library.

Continuous effort has been made to uncover the molecular mechanisms of telomerase activation during carcinogenesis. The hTERT gene is regulated by androgens as well as by different oncogenes including Her-2, Ras, c-Myc and Bcl-2, which seem to play an important role in cancer grow and progression. The discovery of the telomerase subunit hTERT [10], a catalytic subunit bearing the enzymatic activity of telomerase, [11] was the starting point for uncovering the cancerspecific activation of telomerase. Numerous studies have demonstrated that hTERT expression is highly specific to cancer cells and tightly associated with telomerase activity, while the other subunits are constitutively expressed both in normal and cancer cells [12]. Therefore, there is no doubt that hTERT expression plays a key role in cancer-specific telomerase activation.

1. hTERT inhibitors

Telomerase is an attractive target for anti-cancer therapeutics due to its requirement for cellular immortalization and expression in human neoplasms [13]. Because telomerase activity is essential for proliferation of most cancer cells, therapeutic strategies have been developed to inhibit its activity. These

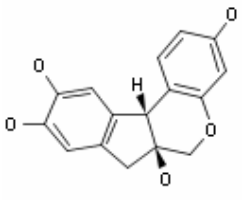
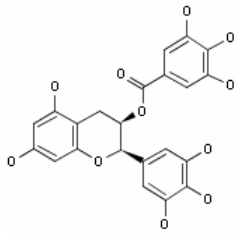
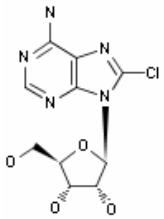
strategies centre on targeting the active site, hTERT and hTERC expression, core enzyme stability and telomeric DNA [14]. Successful approaches involve a combination of traditional drugs with telomerase inhibitors. Though initially promising, strategies that inhibit telomerase with either small molecules or antisense oligonucleotides have a major limitation, namely the lag time required for telomere shortening before cellular effects are attained. As alternative approaches, immunotherapy and gene therapy have been tailored to exploit, rather than antagonize telomerase expression and/or activity. Several Phase I studies of hTERT immunotherapy have been completed in patients with breast, prostate, lung and other cancers, and clinical and immunological results are encouraging. Immunotherapy induces functional, antitumour T cells in patients in the absence of clinical toxicity. It requires the presence of the catalytic subunit of telomerase, hTERT, to elicit an immune response directed towards hTERT peptide-presenting cells. hTERT promoter-driven gene therapy and mutant telomerase RNA (hTR) gene therapy depend on the innate telomerase activity of cancer cells to drive the expression of pro-apoptotic genes and to synthesize mutated DNA sequences onto telomeres, respectively. In addition, telomestatin, a G-quadruplex binding ligand may exert anti-proliferative effects independently of telomere shortening. Disrupting the functional expression of hTERT is particularly effective in agreement with evidence that hTERT is an antiapoptotic factor in some cancer cells. In addition, approaches that stabilise DNA secondary structures may disrupt telomere maintenance through a variety of routes making them, potentially, very potent in attacking cancer cells.

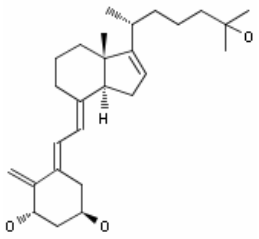
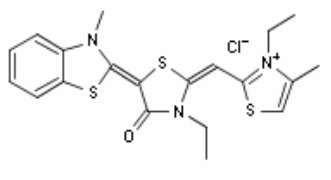
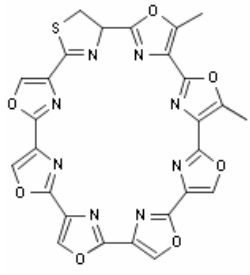
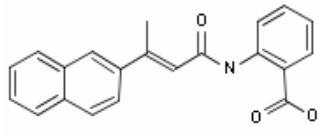
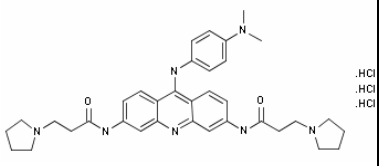
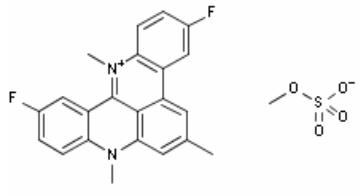
Compounds currently under development that seek to inhibit hTERT, the reverse transcriptase component of telomerase, include nucleoside analogs and the small molecule BIBR1532 [15]. Compounds inhibiting the RNA component of telomerase, hTERC, include peptide nucleic acids, 2-5A antisense oligonucleotides, and N3'-P5' thio-phosphoramidates. Recently, an oligonucleotide sharing sequence homology with terminal telomeric DNA, termed 'T-oligo', has shown cytotoxic effects in multiple cancers in culture and animal models. Independent of telomerase function, T-oligo is thought to mimic the DNA-damage response a cell normally experiences when the telomere t-loop structure becomes dysfunctional [16].

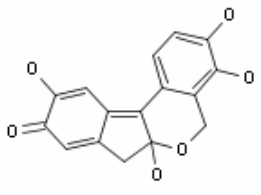
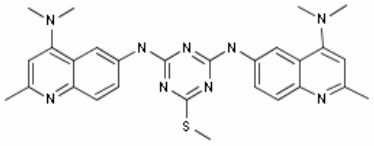
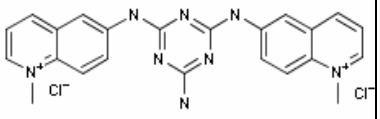
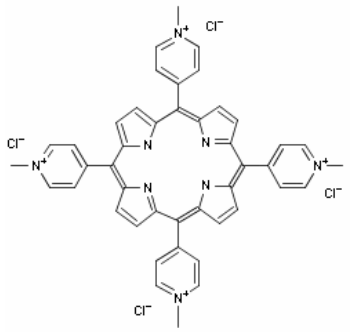
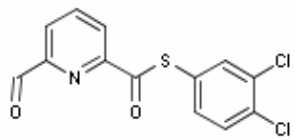
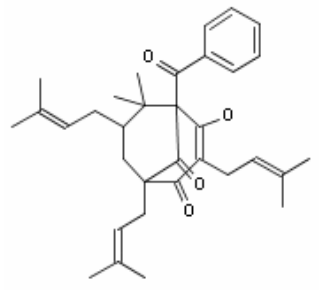
To the present day, more than 400 hTERT inhibitors have been developed including peptide-based substances and a range of various small-molecule compounds. Among them, more than 40 compounds are being evaluated in different clinical trials as well as preclinical studies (Table 1).

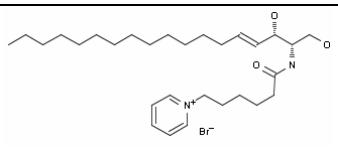
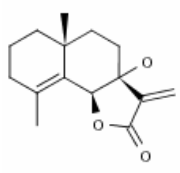
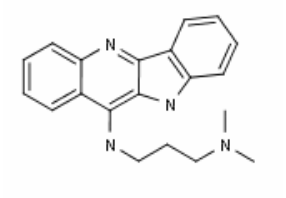
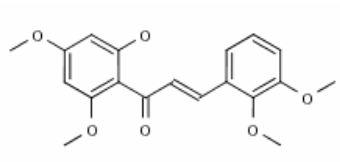
Table 1. Representative examples of small-molecule compounds evaluated in advanced biological trials

№	Structure/Name	Phase/Originator	Addition mechanism/activity	Therapeutic Group

1	 Brazilin	Preclinical/ Tsumura	Free Radical Scavengers NOS2 Expression Inhibitors Nitric Oxide Production Inhibitors	Antidiabetic Drugs Antiplatelet Therapy Immunomodulators
2	 (-)-Epigallocatechin gallate As well as curcumin was found to produce the same effect on hTERT activity	Multi Phase II/III	AP-1 Inhibitors Aromatase Inhibitors Bacterial Efflux Pump Inhibitors DNA Gyrase Inhibitors Fatty Acid Synthase Inhibitors HCV NS3 Protease Inhibitors Indoleamine 2,3-dioxygenase Inhibitors NF-kappaB (NFKB) Activation Inhibitors PDGFR Inhibitors Prolyl Endopeptidase (prolyl oligopeptidase; POP) Inhibitors Proteasome Inhibitors SGLT-1 Inhibitors Tumor NADH Oxidase (tNOX) Inhibitors VEGFR-2 (FLK-1/KDR) Inhibitors beta-Amyloid (Abeta) Aggregation Inhibitors beta-Amyloid (Abeta) Protein Neurotoxicity Inhibitors beta-Secretase (BACE) Inhibitors	Actinic Keratoses, Agents for Agents for Liver Fibrosis Alzheimer's Dementia, Treatment of Anti-Hepatitis C Virus Drugs Antineoplastic Antibiotics Antiparkinsonian Drugs Chemopreventive Agents Dermatologic Drugs Lipoprotein Disorders, Treatment of Metabolic Disorders (Not Specified) Multiple Sclerosis, Agents for Muscular Dystrophy, Agents for Ophthalmic Drugs
3	GRN-56715	Geron Preclinical	-	-
4		Phase I National Cancer Institute (NCI) (Originator) Peking University	IC50=0.750 µM Cancer, breast (adenocarcinoma) remission/reduction, IN VITRO	Lymphocytic Leukemia Therapy Oncolytic Drugs

	NSC-354258	Health Science Center (Originator)	MCF7 human breast adenocarcinoma cells (hormone-dependent)	
5	 Ro-25-4020	Preclinical Roche	IC50=0.300 nM Cancer, prostate remission/reduction, IN VITRO LNCaP human prostate carcinoma cells (androgen-dependent)	Oncolytic Drugs
6	 FJ-5002	Preclinical Dana-Farber Cancer Institute (Originator) FUJIFILM (Originator)	-	Oncolytic Drugs
7	 Telomestatin	Taiho	IC50=5.00 nM Telomerase inhibition, IN VITRO Namalva Burkitt's lymphoma cells	Oncolytic Drugs
8	 BIBR-1532	Boehringer Ingelheim	IC50=5.00 µM Telomerase inhibition, IN VITRO HeLa human cervix adenocarcinoma cells	Oncolytic Drugs
9	 BSG-01	Preclinical Cancer Research Technology	IC50=60 nM Telomerase inhibition, IN VITRO A2780 human ovary carcinoma cells (cisplatin-resistant)	Oncolytic Drugs
10		Preclinical Institute of Cancer Research (ICR) University of Nottingham	IC50=0.330 ± 0.130 µM Telomerase inhibition, IN VITRO A2780 human ovary carcinoma cells	Oncolytic Drugs

	RHPS04			
11	 Hematein	Preclinical Korea Res. Inst. Biosci. Biotechnol. (Originator) Seoul National University (SNU) (Originator)	-	Atherosclerosis Therapy
12	 Compound 115405  Compound 12459	Preclinical Sanofi	Compound 115405: IC50=72 nM Cancer, rhinopharyngeal remission/reduction, IN VITRO KB human epidermoid rhinopharyngeal carcinoma cells Compound 12459: IC50>22.7 µM	Oncolytic Drugs
13	 TMPyP4	Cylene Pharmaceuticals (Originator) University of Arizona (Originator) Preclinical	IC50=1.97 µM Telomerase inhibition, IN VITRO HeLa human cervix adenocarcinoma cells	Oncolytic Drugs
14		Preclinical Chong Kun Dang Pharm (CKD Pharm)	IC50=24 µM Telomerase inhibition, IN VITRO Telomeric repeated amplification protocol assay	Oncolytic Drugs
15	 CLU-502	Universitaetsklin., Essen Preclinical	DNA Topoisomerase I Inhibitors DNA Topoisomerase II Inhibitors Telomerase Inhibitors/ IC50=2-5 µM HL60 human acute promyelocytic leukemia cells	Oncolytic Drugs

16	 L-threo-C6-pyridinium-ceramide-bromide	MUSC Foundation for Research Development Preclinical	-	Oncolytic Drugs
17	 7-Hydroxyfrullanolide	Preclinical Himalaya Global Holdings	IL-6 Production Inhibitors TNF-alpha Production Inhibitors/ IC50=0.756 mg/l SKBr3 human breast adenocarcinoma cells (c-erbB2-overexpressing)	Inflammation, Treatment of Oncolytic Drugs
18	 SYUIQ-5	Okayama University (Originator) Okayama University of Science (Originator) Preclinical	IC50=0.440±0.030 µM Telomerase inhibition, IN VITRO K562 human myeloid leukemia cells	Oncolytic Drugs
19	 HTMC	Preclinical Chaoyang University of Technology Chung Shan Medical University	CC50=47 µM Cancer, lung (non-small cell) (NSCLC) remission/reduction, IN VITRO A549 human non-small-cell lung carcinoma cells	Oncolytic Drugs

As an example, a series of 2,7-diamidoanthraquinone were designed and synthesized [[17]]. They were evaluated for their effects on telomerase activity, hTERT expression, cell proliferations, and cytotoxicity. In the series, compounds (6, 10, 13, 16, 18, 19, 20–22, and 24) showed potent telomerase inhibitory activity, while compounds 19, 21, and 22 activated hTERT expression in normal human fibroblasts (Fig. 1). The results indicated that 2,7-diamidoanthraquinones represent an important class of compounds for telomerase-related drug developments. Compounds 8, 16, 18, 26, and 32 were also selected by the NCI for Screening Program and demonstrated high anti-proliferative activity against 60 human cancer cell lines. Structure–activity relationships (SAR) study revealed that the test compounds with side chains two carbon spacer between amido and amine are important structural moiety for telomerase inhibition. Although the exact mechanism of how this amine group contributes to its activity

is still unclear, however, the amine group in the extended arm of the bis-substituted anthraquinone might contribute to proper binding to the residues within the grove of G-quadruplex structure. These results indicated that the 2,7-disubstituted amido-anthraquinones are potent telomerase inhibitors that have the potential to be further developed into novel anticancer chemotherapeutic agents.

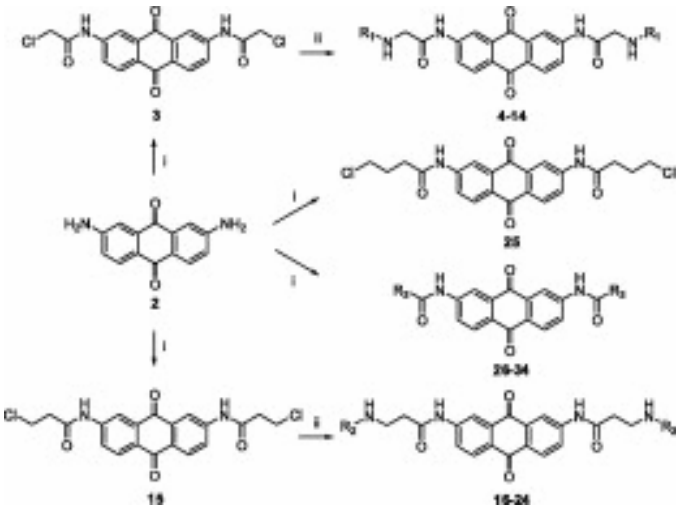


Fig. 1. 2,7-Diamidoanthraquinone as effective hTERT inhibitors

2. Concept and Applications

hTERT-targeted library design at CDL involves:

• *A combined profiling methodology that provides a consensus score and decision based on various advanced computational tools*:

1. Unique bioisosteric morphing, structure similarity and funneling procedures in designing novel potential hTERT ligands with high IP value. We apply CDL's proprietary Chemosoft™ and SmartMining™ software as well as commercially available solutions from Accelrys (MolSoft™), MOE, Daylight and other platforms.

2. Neural Network tools for target-library profiling, in particular Self-organizing Kohonen maps, performed in SmartMining Software. We have also used the Sammon mapping and Support vector machine (SVM) methodology as more accurate computational tools to create our hTERT-focused library.

3. In several cases we have used 3D-molecular docking approach to the focused library design.

4. Computational-based `*in silico*` ADME/Tox assessment for novel compounds includes prediction of human CYP P450-mediated metabolism and toxicity as well as many pharmacokinetic parameters, such as Brain-Blood Barrier (BBB) permeability, Human Intestinal Absorption (HIA), Plasma Protein binding (PPB), Plasma half-life time ($T_{1/2}$), Volume of distribution in human plasma (V_d), etc.

A general approach to limiting the space of virtual libraries of combinatorial reaction products consists of implementation of a series of special filtering procedures. The typical filtering stages are

briefly summarized in Figure 2. A variety of "Rapid Elimination of Swill" (REOS) filters is used to eliminate compounds that do not meet certain criteria [18].

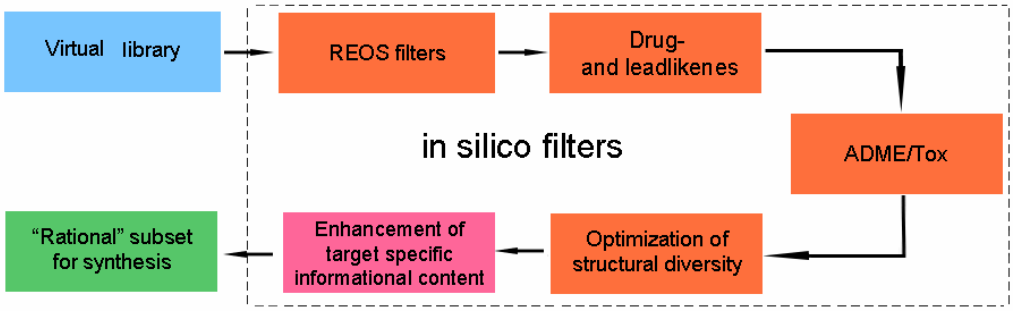


Fig. 2. General procedures of selection of a rational target-specific subset within an initial virtual combinatorial library

These criteria can include: (1) presence of certain non-desirable functional groups, such as reactive moieties and known toxicophores; (2) molecular size, lipophilicity, the number of H-bond donors/acceptors, the number of rotatable bonds. At the next stage the design focuses on "lead" and "drug-likeness" of combinatorial molecules [19]. The ADME/Tox properties of screening candidates should be taken into consideration as early as possible [20]. Additional filters are therefore used for *in silico* prediction of some crucial ADME/Tox parameters, such as solubility in water, logD at different pH values, cytochrome P450-mediated metabolism and toxicity, and fractional absorption. Optimization of structural diversity is another natural and very important way to constrain the size of combinatorial libraries (reviewed in [21]). The fundamentals for these applications are described in a series of our recent articles on the design of exploratory small molecule chemistry for bioscreening [for related data visit ChemDiv, Inc. online source: www.chemdiv.com]. Our multiple *in silico* approach to hTERT-focused library design is schematically illustrated in Fig. 3.

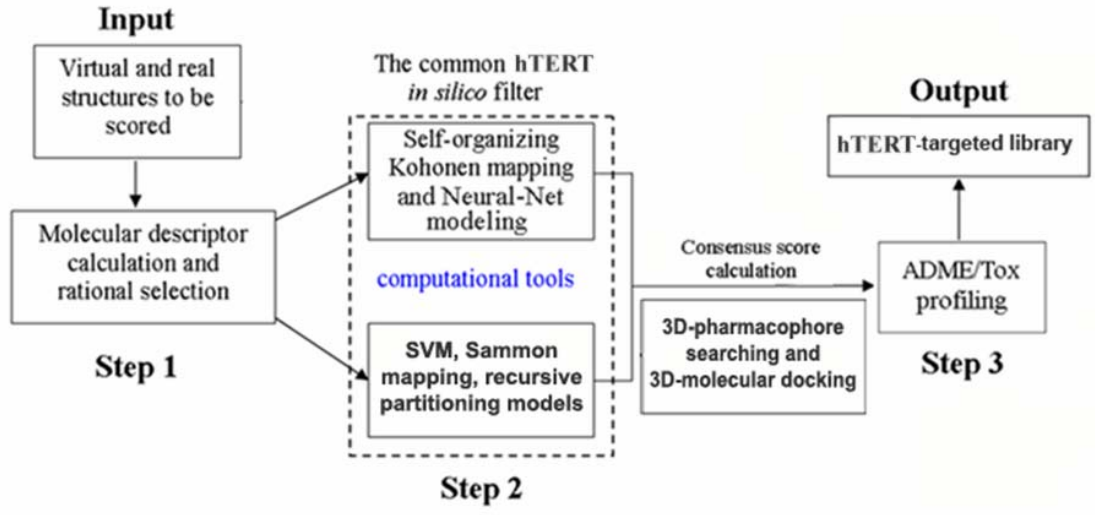


Fig. 3. Multiple computational approach to hTERT-targeted library design

• *Synthesis, biological evaluation and SAR study for the selected structures*:

1. High-throughput synthesis with multiple parallel library validation. Synthetic protocols, building blocks and chemical strategies are available.

2. Library activity validation via bioscreening; SAR is implemented in the next library generation.

2.1. Virtual Screening on hTERT-specific Activity

The common hTERT-filter

At the initial stage of our hTERT-targeted library design, we have collected a 25K-small molecule agents database of known drugs and compounds entered into preclinical or clinical trials; their structures and assignments were obtained from Integrity Database [22], scientific publications and related patents. Each compound in this database is characterized by a defined profile of target-specific activity, focused against 1 of more than 100 different protein targets. The database was filtered based on MW (not more than 800). Molecular features encoding the relevant physicochemical and topological properties of compounds were calculated from 2D molecular representations and selected by ADS and PCA (Step 1, Fig. 3, see below). These molecular descriptors encode the most significant molecular features, such as molecular size, lipophilicity, H-binding capacity, flexibility, and molecular topology. Taken in combination, they define both pharmacokinetic and pharmacodynamic behavior of compounds and are effective for property-based classification of target-specific groups of active agents. However, it should be noted that for each particular target-specific activity group, another, more optimal set of descriptors can be found, which provides better classification ability.

After the calculation a feature reduction stage has been performed. In the modeling studies described here, for reduction of the number of input variables, we have been used a unique algorithm, named Automatic Descriptors Selection (ADS), implemented in SmartMining software as well as classical Principal Component Analysis (PCA). The principles of PCA have been described many times in scientific literature and are not described here. As a result of the performed selection procedure, at the output, an experimental set consisted of 7 molecular descriptors including Zagreb index, E-state indexes for the following structural fragments: >C-, -CH2-, -CH3, the number of H-bond donors, HB2 (a structural descriptor which encodes the strength of H-bond acceptors following an empirical rule) and LogP was determined.

After all the preparatory procedures were complete, the reference database with selected molecular descriptors was used for development of *in silico* model with the most appropriate architecture and learning strategy. Key examples cited in the notes section represent real computational filtering technologies developed at ChemDiv, Inc. for enhancement of knowledge-based content of exploratory chemical libraries for biological screening at the stage of combinatorial synthesis planning. Prior to the

statistical experiments, the molecular structures should be filtered and normalized in order to fulfill certain criteria. As shown in Fig. 3, 'front-line' computational tools include Kohonen-based SOM generation as well as Neural-Net- and SVM-based modeling; these algorithms have been effectively used across the Step 2, decoded in Fig. 3.

Self-organizing Kohonen mapping

A Kohonen SOM of 25K pharmaceutical leads and drugs generated as a result of the unsupervised learning procedure is depicted in Fig. 4. It shows that the studied compounds occupy a wide area on the map, which can be characterized as the area of druglikeness. Distribution of various target-specific groups of ligands in the Kohonen map demonstrates that most of these groups have distinct locations in specific regions of the map (Fig. 5a-l). A possible explanation of these differences is in the fact that, as a rule, receptors of one type share a structurally conserved ligand-binding site. The structure of this site determines molecular properties that a receptor-selective ligand should possess to properly bind the site. These properties include specific spatial, lipophilic, and H-binding parameters, as well as other features influencing the pharmacodynamic characteristics. Therefore, every group of active ligand molecules can be characterized by a unique combination of physicochemical parameters differentiating it from other target-specific groups of ligands. Another explanation of the observed phenomenon can be related to different pharmacokinetic requirements to drugs acting on different biotargets.

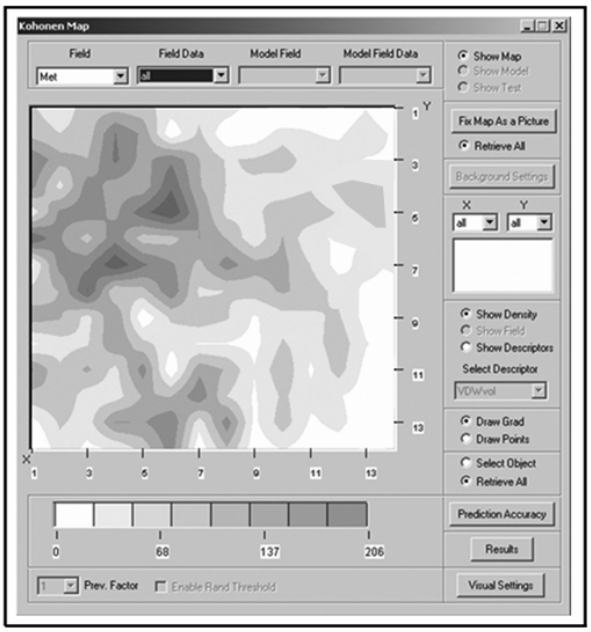


Fig. 4. Property space of 25K pharmaceutical leads and drugs visualized using the Kohonen map (*the data have been smoothed*)

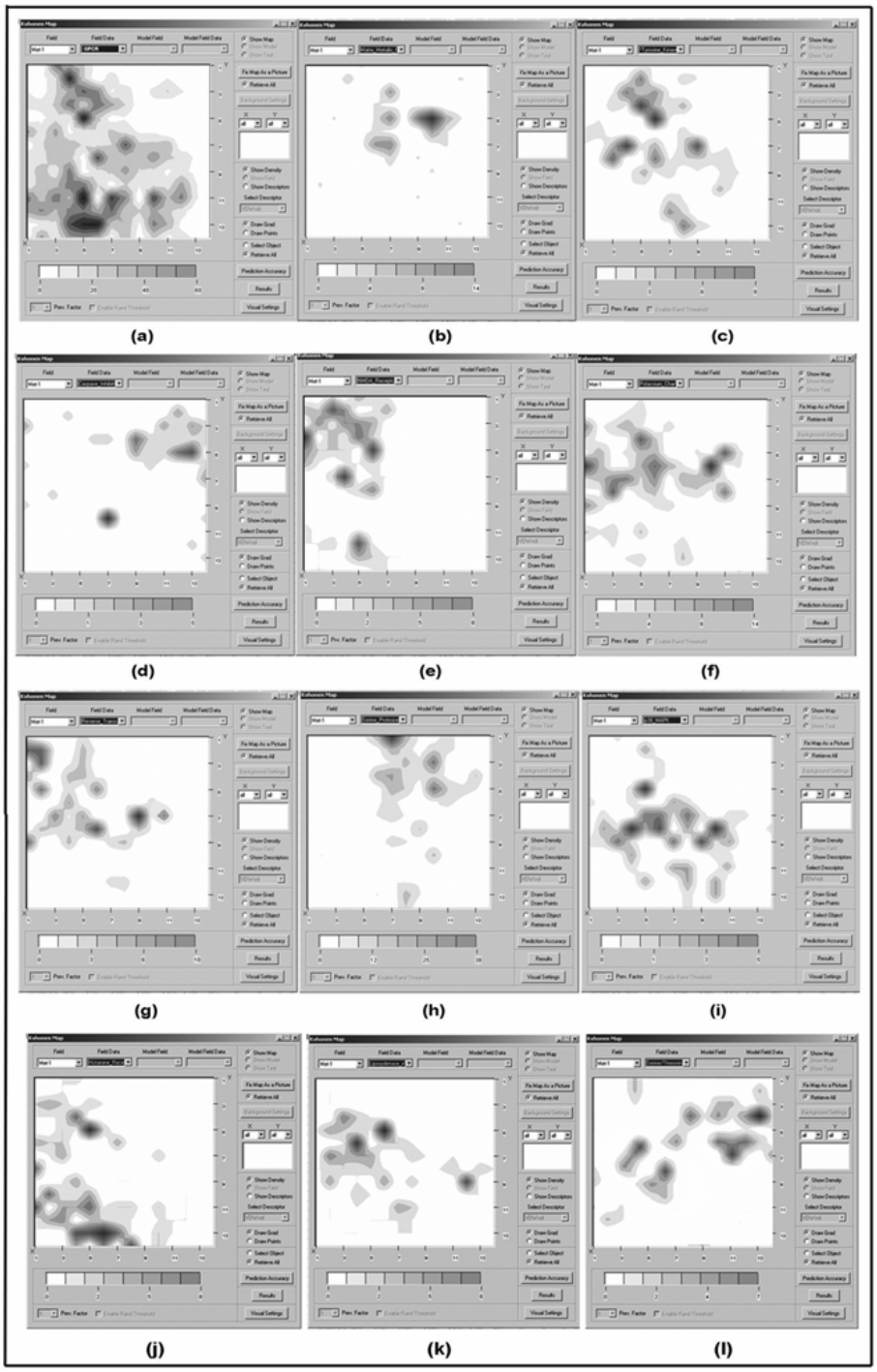


Fig. 5. Distribution of representative target-specific groups of pharmaceutical agents within the Kohonen map: (a) GPCR agonists/antagonists (5432 compounds); (b) matrix metalloproteinase inhibitors (120 compounds); (c) tyrosine kinase inhibitors (175 compounds); (d) caspase inhibitors (50 compounds); (e) NMDA receptor agonists/antagonists (150 compounds); (f) potassium channel blockers/activators (302 compounds); (g) reverse transcriptase inhibitors (160 compounds); (h) serine protease inhibitors (531 compounds); (i) p38 MAPK inhibitors (100 compounds); (j) histamine receptor antagonists (168 compounds); (k) lipoxygenase inhibitors (114 compounds); (l) hTERT inhibitors (320 compounds)

The described algorithm represents an effective procedure for selection of target-biased compound subsets compatible with high-thoughput *in silico* evaluation of large virtual chemical space. Whenever a large enough set of active ligands is available for a particular receptor, quantitative discrimination function can be generated allowing selection of a series of compounds to be assayed against the target. Once a Kohonen network is trained and specific sites of location of target-activity groups of interest are identified, the model can be used for testing any available chemical databases with the same calculated descriptors. The Kohonen mapping procedure is computationally inexpensive and permits real-time calculations with moderate hardware requirements. Thus for a training database consisting of 25K molecules with 7 descriptors using 4000 iterations, approximately 2 hours are required for a standard PC (Pentium 3-GHz processor) on a Windows XP platform to train the network. The time increases almost linearly with the size of the database. After the Kohonen network is trained, the 2D map can be created in a short time. It is important to note, that focusing on physicochemical rather than structural features makes this approach complementary to any available ligand structure similarity technique.

Our own experience and literature data demonstrate that Kohonen self-organizing maps are an efficient clustering, quantization, classification and visualization tool very useful in the design of chemical libraries. Possible limitations of this approach are related to the fact that the SOM algorithm is designed to preserve the topology between the input and grid spaces; in other words, two closely related input objects will be projected on the same or on close nodes. At the same time, the SOM algorithm does not preserve distances: there is no relation between the distance between two points in the input space and the distance between the corresponding nodes. The latter fact sometimes makes the training procedure unstable, when the minor changes in the input parameters lead to serious perturbation in the output picture. As a result, it is often difficult to find the optimal training conditions for better classification. Another potential problem is associated with the quantization of the output space. As a result, the resolution of low-sized maps can be insufficient for effective visualization of differences between the studied compound categories.

The predictive ability of the model constructed towards hTERT-active agents was approx. 75%; therefore, this model can be used for targeted-library design and rational compound selection.

Neural-Net modeling

Using the same knowledgebase we have further developed a property-based neural network (NN) algorithm for effective discrimination between hTERT inhibitors and compounds belonging to *non*-hTERT activity classes. Following the underlying strategy, 320 known hTERT ligands were used as a positive training set, hTERT (+); a subset of more than 10K compounds, representing over 200 various

non-hTERT based active ligands was used as a negative training set, hTERT(-). Using a special feature selection procedure, a 10-descriptor set was chosen for NN experiments. These descriptors encode significant molecular properties, such as lipophilicity, charge distribution, topological features, steric and surface parameters. The back-propagated feed-forward nets were constructed and trained with the molecular descriptors as input values and activity scores as output values. To assess the predictive ability of the NN models generated, we used three independent randomizations within the reference dataset which included tree groups of compounds (training, cross-validation and test group). The resulting histogram is shown in Fig. 6.

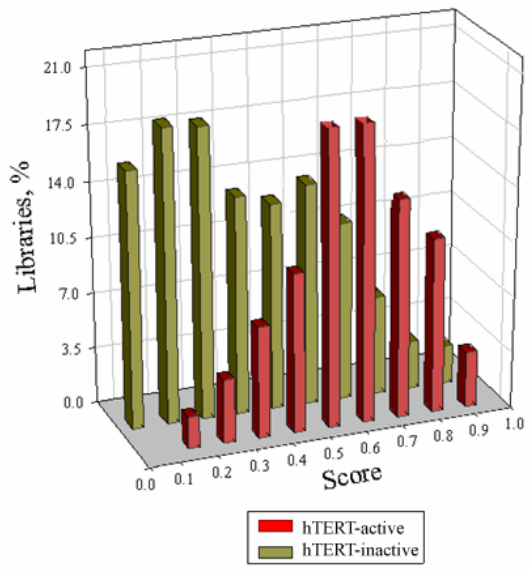


Fig. 6. Distribution of hTERT-active and hTERT-inactive compounds from the test set. An average predictive accuracy was 72%

The classification quality was approximately the same in each of these three independent cycles: up to 74% of hTERT ligands and 70% of *non*-hTERT ligands were correctly classified in the corresponding test sets. It should be noted that we carried out a wet lab experimental validation of the similar model via highthroughput screening of 5K compounds from the CDL corporate compound database against abl-kinase. The experimental activity data (hit rate) was consistent with the expected from NN calculations, which demonstrates a high utility of NNs in designing kinase-specific combinatorial libraries. The model demonstrated an enhanced level of discrimination between “active” and “inactive” libraries.

SVM-based modeling

Recently, a so-called Support Vector Machines (SVM) [[23]] method has became popular as an alternative method. At least as powerful and versatile as ANNs, SVM approach is being adjusted for

various application, from genomics to face recognition, including drug design [24]. Recently, we tested SVM as a classification tool in several drug-discovery programs and found it typically outperforming other approaches, in particular, ANNs [25]. Here, we used SVM algorithm for selection of compounds for primary and secondary screening against hTERT.

The main parameters of the SVM-based classification model are similar to that used in NN-modeling. Thus, as a training set, we used 320 known hTERT ligands from different classes (positive training set, hTERT(+)), and a set of more than 10K small molecule compounds, spraying over various *non*-hTERT active ligands (negative training set, hTERT(-)). All molecules were additionally filtered for molecular weight range (200–600) and atom type content (only C, N, O, H, S, P, F, Cl, Br, and I were permitted). For the entire database of hTERT-active and hTERT-inactive structures, we have calculated sixty five molecular descriptors encoding such molecular properties as lipophilicity, charge distribution, topological features, steric and surface parameters, using ChemoSoft™. Low-variability and highly correlated ($R > 0.9$) descriptors were removed reducing the set to 39. A sensitivity analysis [26] was applied to further reduce the number of the redundant descriptors. The resulted 8 molecular descriptors (logP, no. of H-bond acceptors, no. of H-bond donors, no. of rotatable bonds, molecular refractivity, density, Zagreb index, relative positive surface area), were used for generation of the SVM classification model [27]. Before modeling each descriptor was scaled to [0;1] range (by training set; scaled values for other subsets were derived using train set scaling factors). SVM classifiers were based on linear or nonlinear (Radial Basis Functions, RBF) kernel. In our experiments, the nonlinear RBF kernel provided the best classification ability. The goodness of the model has been evaluated using an internal validation procedure. The whole set of all compounds was divided into three parts: training set (for building the SVM model), validation set (for checking model quality while generating SVM models; this set was used to check SVM models instead of leave-one-out crossvalidation, as the latter is too slow for large data sets), and the test set (for checking prediction quality of the best models). The resulting figure 7 illustrates the distributions of calculated SVM scores for compounds in hTERT(+) and hTERT(-) test sets, correspondingly. In order to assess the classification quality of the trained SVM model, we calculated percent of correctly classified compounds in each set at different threshold scores. With the threshold score 0.4, the model correctly classified up to 70% of hTERT(+) and 78% of hTERT(-) compounds.

After models were developed and successfully validated we further classified the structures from our virtual library through this common *in silico* filter. Thus, based on the outputs outputted from these models we have calculated a consensus score for each compound tested. As a result, a large set of high-score structures (more than 60K compounds) was collected and further evaluated using specific computational models (Step 3, see above).

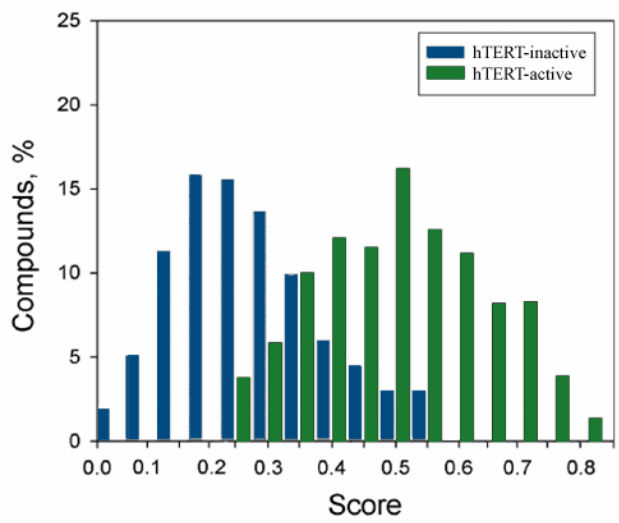


Fig. 7. SVM score distribution of the test set compounds. An average predictive accuracy was 74%

Specific in silico filters

The set of the compounds selected were further expanded and tested using specific computational approaches including bioisosteric morphing (see below), 3D-molecular docking (*not described here*), Sammon mapping (*not described here*), etc. For example, the basic concept of bioisosterism is central in drug design and development [[28]]. The term refers to the compounds or substructures that share similar shapes, volumes, electronic distributions and physicochemical properties and have similar biological activity [[29]]. Therefore, bioisosteric approach is useful for morphing the marginal chemotypes. Thus, we have carried out several bioisosteric transformations based on the structures of known hTERT-active ligands. The generated structures were further used as templates for 2D-structure similarity (*Tanimoto index*) towards various compounds from ChemDiv store. As a result, we have selected more than 85K structures in addition to "*high-score*" compounds outputted previously (see "*The common hTERT-filter"*).

It should be particularly noted that following the original concept of diversity-oriented compound library design we have effectively applied three computational methods which were based solely on physicochemical descriptors, and so they provide various structures of high diversity. In turn, bioisosteric morphing generally operates within the defined and relatively narrow scope of the core/template structure of active compound. Thus, the final set included two main groups: structures which were obtained at the output of front-line filters (60 compounds) as well as structures generated by bioisosteric transformations within hTERT-active compounds followed by similarity (85K compounds). These groups were combined and gave (145K) structures which were further evaluated using pharmacophore modeling and 3D-molecular docking approach (*not described here*). After this modeling, 52K "high-score" compounds were included in the final hTERT-targeted library. Key diversity parameters for the desired database are

listed in Table 1. As evident from the number of screens, the number of core heterocyclic fragments, and the diversity coefficients (all these parameters are calculated using the *Diversity* module [30] of the ChemoSoft™ software tool), the studied compound database has a high structural diversity and, from the "targeted diversity" point of view, it can be reasonably considered as a good hTERT-focused library.

Table 1. Diversity parameters of the hTERT-targeted dataset

Parameter	Value
Total number of compounds	52919
No. of screens[a]	7,181
Diversity coefficient[b]	0.798
No. of core heterocycles	500

[a] screens are simple structural fragments, centroids, with the topological distance equal to 1 bond length between the central atom and the atoms maximally remote from it.

[b] cosine coefficients are calculated, and the sums of non-diagonal similarity matrix elements are used in ChemoSoft™ program as a diversity measure; the diversity coefficient can possess the value from 0 to 1, which correspond to minimal and maximal possible diversity of a selection.

The representative examples of "high-score" structures entered in the final library are shown within the figure below. As a result, we have selected a set of more than 52K structures which can be regarded as potential hTERT inhibitors (see Fig. 8).

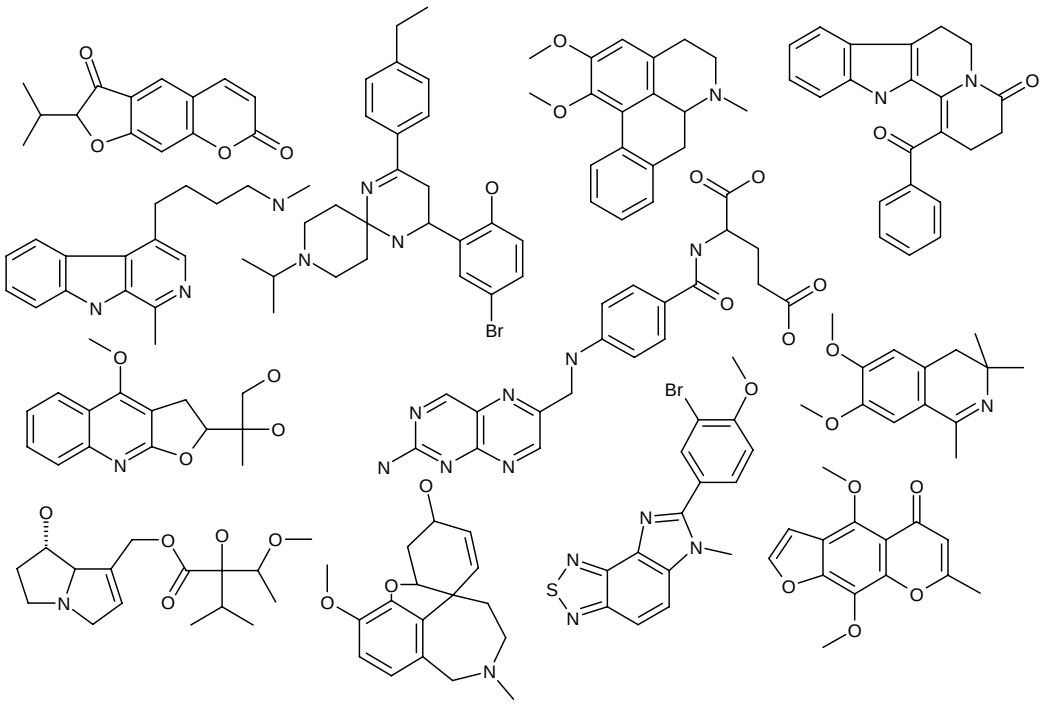


Fig. 8. Representative structures from the common hTERT-targeted library

In summary, we have developed and effectively applied a multi-step computational approach to design of our hTERT-targeted library. In particular, we have successfully validated this strategy towards a series of other biological targets, including tyrosine kinases, chemokines, caspases, etc. The related biological trials have revealed several highly potent inhibitors, and we can confidently conclude that described *in silico* pathway represents an effective method for targeted libraries design. Moreover, we provide rapid and efficient tools for follow-up chemistry on discovered hits, including single isomer chemistry, stereoselective synthesis and racemic mixture separation. The developed library is updated quarterly based on a "cache" principle. Older scaffolds/compounds are replaced by templates resulting from our in-house development (unique chemistry, literature data, computational approaches) while the overall size of the library remains the same (ca. 94K compounds). As a result, the libraries are renewed each year, proprietary compounds comprising 50-75% of the entire set. Clients are invited to participate in the template selection process prior to launch of our synthetic effort.

References

[1] Keith WN, Lilsland A, Evans JTR, Glasspool RM. Telomerase-directed molecular therapeutics. Expert reviews in molecular medicine. 2002, 22 April. http://www.expertreviews.org

[2] Morin GB. The human telomerase terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. Cell. 1989;59:521-529

[3] Shay JW, Zou Y, Hiyama E, Wright WE. Telomerase and cancer. Hum Mol Genet. 2001;10:677-685

[4] Expert Opin Ther Targets. 2005 Jun;9(3):457-69. Telomerase: a potential therapeutic target for cancer. Fletcher TM.

[5] Curr Drug Targets Immune Endocr Metabol Disord. 2004 Sep;4(3):253-6. Telomerase as drug and drug target for the treatment of thyroid cancer. Zeiger MA, Meeker AK.

[6] Curr Cancer Drug Targets. 2006 Mar;6(2):147-80. Telomeres and telomerase: Pharmacological targets for new anticancer strategies? Pendino F, Tarkanyi I, Dudognon C, Hillion J, Lanotte M, Aradi J, Ségal-Bendirdjian E.

[7] Expert Rev Vaccines. 2008 Sep;7(7):881-7. Telomerase as a universal tumor antigen for cancer vaccines. Beatty GL, Vonderheide RH.

[8] Rudolph KL, Chang S, Lee HW et al. Longevity, stress response, and cancer in aging telomerase-deficient mice. Cell 1999; 96: 701–12; Chin L, Artandi SE, Shen Q et al. p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. Cell 1999; 97: 527–38.

[9] Artandi SE, DePinho RA. Mice without telomerase: what can they teach us about human cancer? Nat Med 2000; 6: 852–5.

[10] Meyerson M, Counter CM, Eaton EN et al. hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. Cell 1997; 90: 785–95; Nakamura TM, Morin GB, Chapman KB et al. Telomerase catalytic subunit homologs from fission yeast and human. Science 1997; 277: 955–9.

[11] Bodnar AG, Ouellette M, Frolkis M et al. Extension of life-span by introduction of telomerase into normal human cells. Science 1998; 279: 349–52; Nakayama J, Tahara H, Tahara E et al. Telomerase activation by hTRT in human normal fibroblasts and hepatocellular carcinomas. Nat Genet 1998; 18: 65–8

[12] Takakura M, Kyo S, Kanaya T, Tanaka M, Inoue M. Expression of human telomerase subunits and correlation with telomerase activity in cervical cancer. Cancer Res 1998; 58: 1558–61; Kyo S, Kanaya T, Takakura M, Tanaka M, Inoue M. Human telomerase reverse transcriptase as a critical determinant of telomerase activity in normal and malignant endometrial tissues. Int J Cancer 1999; 80: 60–3; Kanaya T, Kyo S, Takakura M, Ito H, Namiki M, Inoue M. hTERT is a critical determinant of telomerase activity in renal-cell carcinoma. Int J Cancer 1998; 78: 539–43; Kyo S, Kanaya T, Takakura M et al. Expression of human telomerase subunits in ovarian malignant, borderline and benign tumors. Int J Cancer 1999; 80: 804–9.

[13] Anticancer Agents Med Chem. 2007 Jul;7(4):475-83. Harnessing telomerase in cancer therapeutics. Fakhoury J, Nimmo GA, Autexier C.

[14] J Am Geriatr Soc. 2003 Jan;51(1):116-22. Telomeres, telomerase, and telomerase inhibition: clinical implications for cancer. Ahmed A, Tollefsbol T.

[15] Curr Med Chem Anticancer Agents. 2002 Sep;2(5):613-26. How to inhibit telomerase activity for cancer therapy. Kyo S, Inoue M.

[16] Anticancer Drugs. 2008 Apr;19(4):329-38. Telomerase inhibitors and 'T-oligo' as cancer therapeutics: contrasting molecular mechanisms of cytotoxicity. Rankin AM, Faller DV, Spanjaard RA.
[17] Bioorg Med Chem. 2008 Jul 15;16(14):6976-86. Epub 2008 Jun 2. Synthesis, human telomerase inhibition and anti-proliferative studies of a series of 2,7-bis-substituted amido-anthraquinone derivatives. Huang HS, Huang KF, Li CL, Huang YY, Chiang YH, Huang FC, Lin JJ.
[18] W. P. Walters, M. T. Stahl, M. A. Murcko, *Drug Disc. Today* 1998, *3*, 160 – 178.
[19] D. E. Clark, S. D. Pickett, *Drug Disc. Today* 2000, *5*, 49 – 58.
[20] P. J. Eddershaw, A. P. Beresford, M. K. Bayliss, *Drug Disc. Today* 2000, *5*, 409 – 414.
[21] M. J. Valler, D. Green, *Drug Disc. Today* 2000, *5*, 286–293.
[22] Prous Science, URL: http://www.prous.com.
[23] Vapnik, V. Statistical Learning Theory, Whiley: New York, 1998.
[24] (a) Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Comput. Chem., 2001, 26, 5; (b) Trotter, M.; Buxton, B.; Holden, S.B. Measurement and Control, 2001, 34, 235; (c) Warmuth, M.K.; Liao, J.; Ratsch, G.; Mathieson, M.; Putta, S.; Lemmen, C. J. Chem. Inf. Comp. Sci., 2003, 43, 667.
[25] Zernov, V.V.; Balakin, K.V.; Ivaschenko, A.A.; Savchuk, N.P.; Pletnev, I.V. J. Chem. Inf. Comput. Sci., 2003, published on web, October 2003.
[26] Bigus, J.P. Data Mining with Neural Networks, McGraw-Hill, 1996.
[27] Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Comput. Chem., 2001, 26, 5.
[28] (a) Burger, A. Prog. Drug Res., 1991, 37, 287; (b) Patani, G.A.; LaVoie, E.J. Chem. Rev., 1996, 96, 3147; (c) Olesen, P.H. Curr. Opin. Drug Discov. Devel., 2001, 4, 471; (d) Chen, X.; Wang, W. Ann. Reports Med. Chem., 2003, 38, 333.
[29] (a) Patani, G.A.; LaVoie, E.J. Chem. Rev., 1996, 96, 3147; (b) Chen, X.; Wang, W. Ann. Reports Med. Chem., 2003, 38, 333.
[30] S. V. Trepalin, V. A. Gerasimenko, A. V. Kozyukov, N. Ph. Savchuk, A. A. Ivashchenko, New Diversity Calculations Algorithms Used for Compound Selection. *J. Chem. Inf. Comput. Sci.* 2002, *42*, 249 – 258.