Supporting Information

Discovery of Novel Myc-Max Heterodimer Disruptors with a 3-Dimensional Pharmacophore Model

Gabriela Mustata[1*], Ariele Viacava Follis[2], Dalia I. Hammoudeh[2], Steven J. Metallo[2], Huabo Wang[3], Edward V. Prochownik[3,4], John S. Lazo[5], and Ivet Bahar[1]

[1]*Department of Computational Biology, University of Pittsburgh, Pittsburgh, PA 15260, USA,*
[2]*Department of Chemistry, Georgetown University, Washington, DC 20057, USA*
[3]*Section of Hematology/Oncology Children's Hospital of Pittsburgh Rangos Research Center, Pittsburgh, PA 15201, USA*
[4]*Department of Microbiology and Molecular Genetics and the University of Pittsburgh Cancer Institute, Pittsburgh, PA 15260, USA*
[5]*Department of Pharmacology and Chemical Biology, University of Pittsburgh, Pittsburgh, PA 15260, USA.*

Contents

Pharmacophore Model Generation, Refinement, and Validation

The software GALAHAD[1-3] was used for the ligand-based pharmacophore model generation. The program works in two stages: first, it uses a genetic algorithm (GA) to identify a set of ligand conformations with minimal energies while maximizing pharmacophore multiplet and pharmacosteric similarities. The best fits are carried forward into the second stage, which is a rigid-body alignment that overlays the ligands in Cartesian space.[3] This approach frees the user from having to select a template molecule. The flexibility of a molecular structure is treated through the initial alignment in torsional angle space by an advanced GA. The resulting torsional angles are applied to the base structure of each ligand to generate the ligand conformation. The 3D similarities among the ligand conformations are estimated through the fast pharmacophoric and steric multiplets (Tuplets).[3] The fact that not all features are required to map contributes to the ability of the models to allow for more structural diversity. GALAHAD employs Pareto multi-objective optimization[4] to simultaneously balance steric, pharmacophoric, and energy information to calculate multiple pharmacophore models. GALAHAD was performed for 100 generations with a population size of 70 and a tournament pool size of 210. Pharmacophore and steric quartets were used to evaluate overlap during the GA run. At least four ligands had to contribute to any given consensus feature for it to be included in the model query; lower levels of stringency tended to produce partial match constraints that were too weak to be useful. Default values were used for other settings. Twenty models were produced in our study, which differed somewhat in the number and type of features, and in the conformations and overlay of the molecules. The best model in this study was selected based on the best Pareto score alignment of the pharmacophoric features.

The refinement stage was performed using the Tuplets module in SYBYL 8.0.[5] The Tuplet program allows for the decomposition of the full pharmacophoric pattern found for each inactive ligand into its constituent distance multiplets, which are encoded into a vector fingerprint. Virtual screening results from Tuplets are qualitatively similar to those obtained from "classical" flexible 3D searches. However, the application of multiplet hypothesis differs from such searches in such a way that the implicit partial match constraint in pharmacophore multiplet hypothesis applies across the groups of features exhibited by the multiplets, rather than individual features. In our study, the Tuplets hypothesis resulted from the refinement of the previously generated GALAHAD pharmacophore model. The Tuplets program uses the original hypothesis to generate a vector fingerprint that is modified using the information added from the inactive compounds in the refinement set. The Tuplet hypothesis is then generated as a bitmap consisting of the k highest scoring vector fingerprint elements. The value of k (i.e. number of bits set) determines the depth of detail included in the model. We used in our study a value of 100, which proved to be a reasonably good starting point.

Test set used for validation of the pharmacophore model:

a. Active Molecules

12RH-NCN-1   28RH-NCN-1   1RH-S-Me

10031-B8   10075-G5   10009-G9

b. Inactive Molecules

6RH   7RH   33RH   35RH
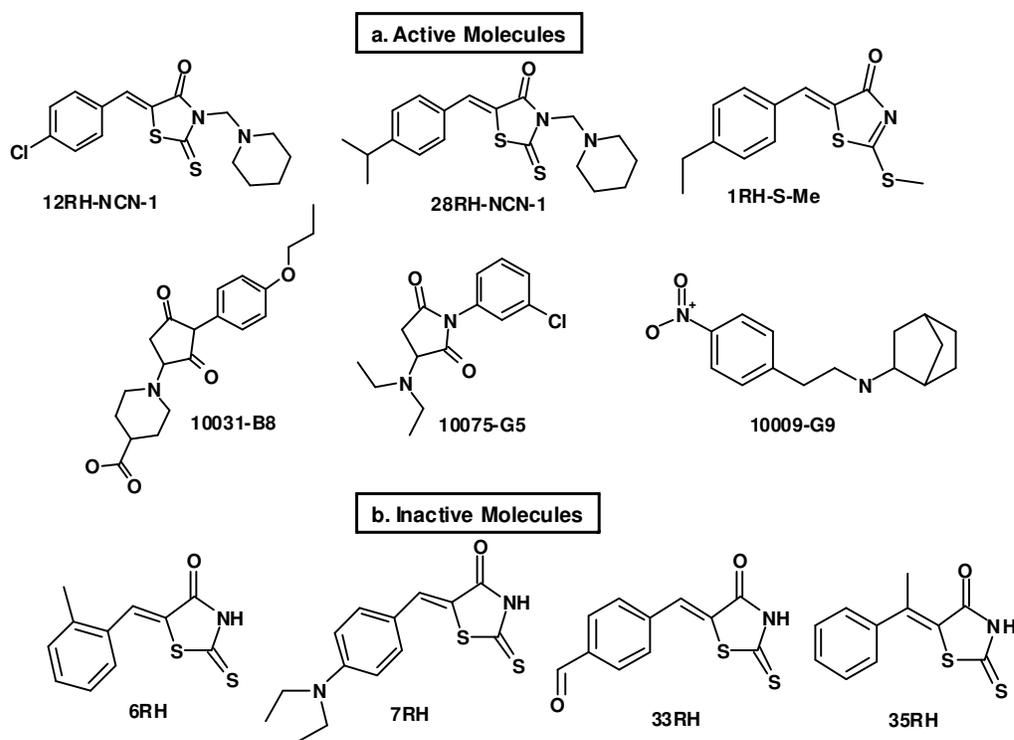
The active molecules have IC50 values similar to 10058-F4.

Summary of HPLC Purity for the tested compounds

Varian ProStar instrument, Alltech Econosphere column
C18 reversed phase 3uM beads
30' gradient 100% H20-0.1%TFA - 100% Acetonitrile -0.1%TFA
wavelength:350 nM

| Compound	Purity	Time (minutes)
5140069	96.4%	16.219
6569963	95.8%	16.904
6525237	99.8%	21.648
5360134	99.0%	24.276
6370870	96.6%	21.065
5928105	91.9%	21.613
5149518	94.9%	20.629
7116536	95.3%	19.127
5248624	98.0%	19.715

The solubility for all compounds was at least 10mM in DMSO and at least 200 µM in PBS buffer.

# Summary of NMR data for the tested compounds

b0343102

| ID | 5140069 | 286.3335 | $C_{16}H_{18}N_2O_3$ |

B0356972

ID	6525237	326.4204	$C_{18}H_{18}N_2O_2S$

B1345-91

ID	6569963	247.2965	$C_{14}H_{17}NO_3$

b0342728

| ID | 5360134 | 348.2475 | C$_{17}$H$_{18}$BrNO$_2$ |

B0078185

| ID 6370870 | 293.3688 | C$_{19}$H$_{19}$NO$_2$ |

b0357122

| ID | 5928105 | 319.4070 | C$_{21}$H$_{21}$NO$_2$ |

ID		
248624	C$_{20}$H$_{20}$O$_5$	340.3794

ID		
7116536	337.4627	C$_{16}$H$_{19}$NO$_3$S$_2$

S8
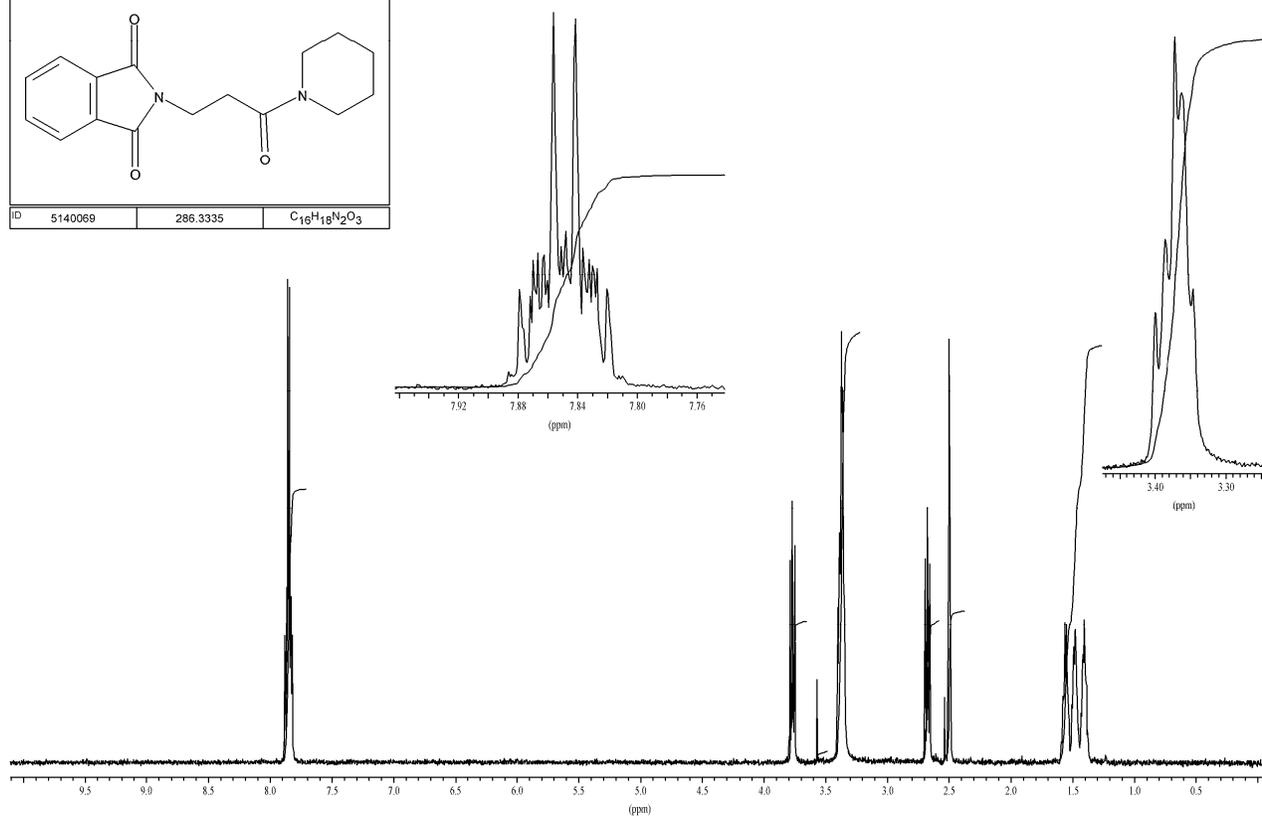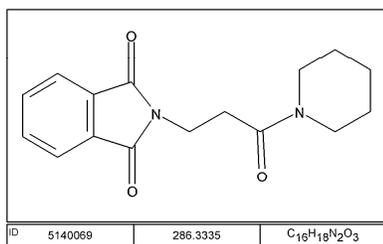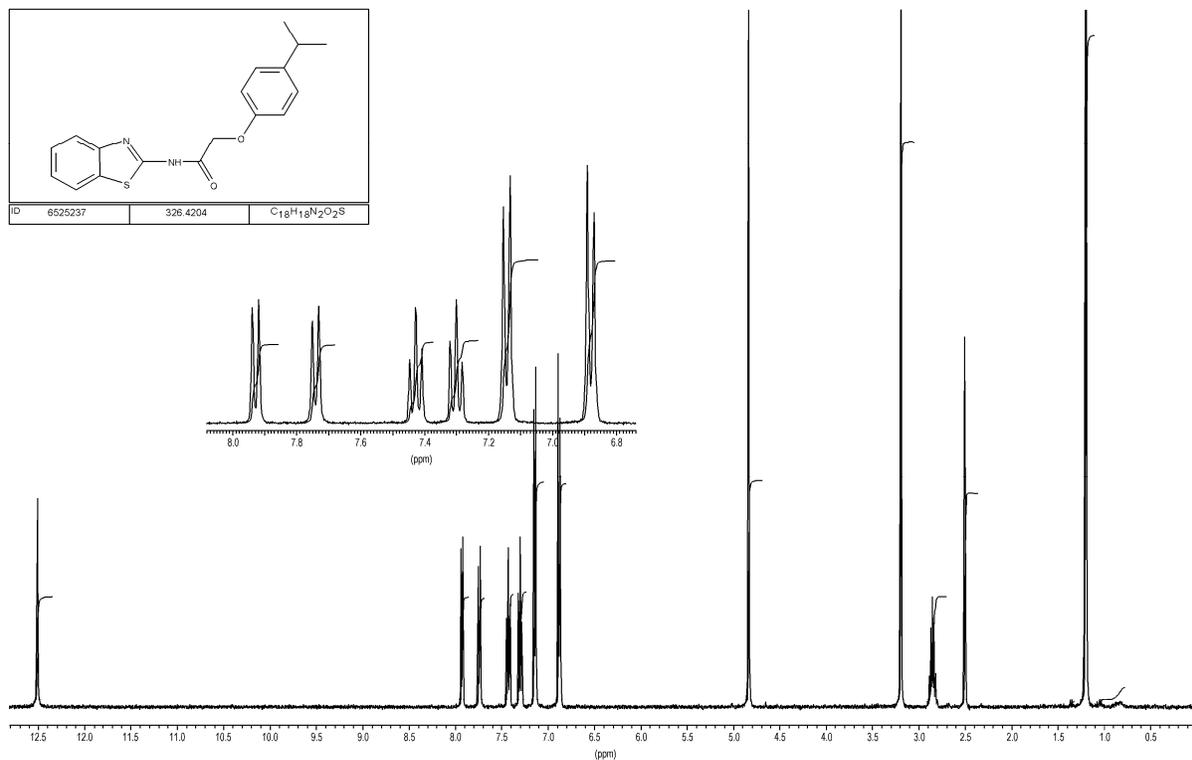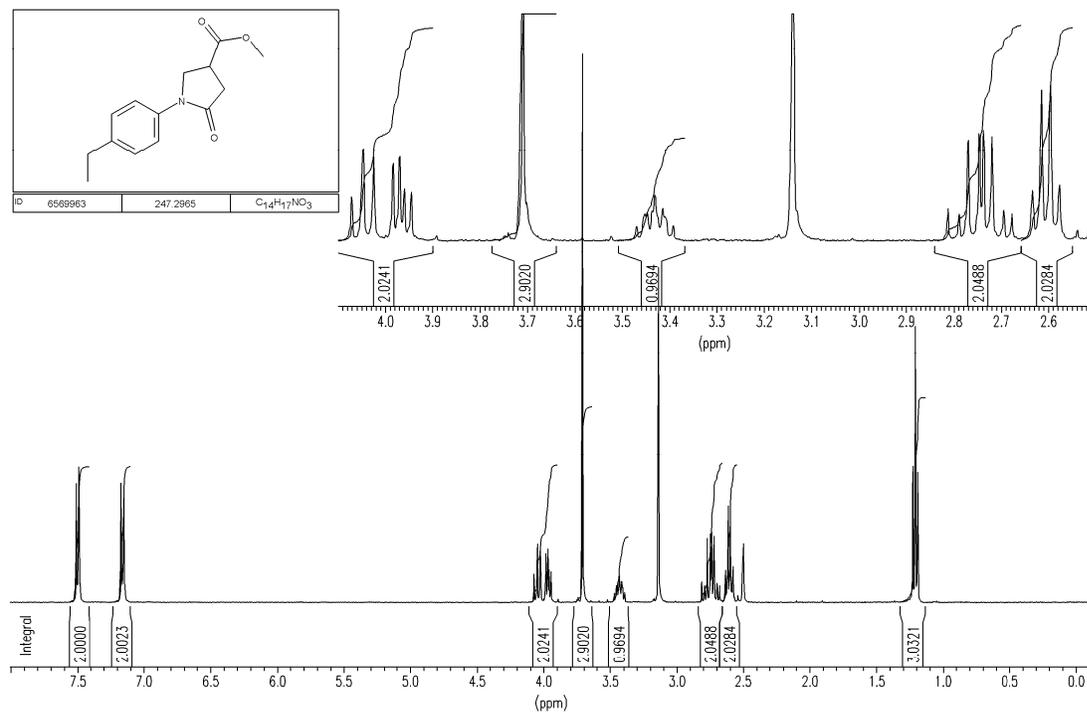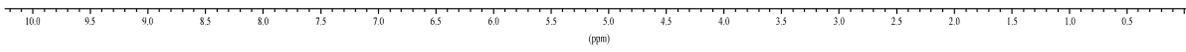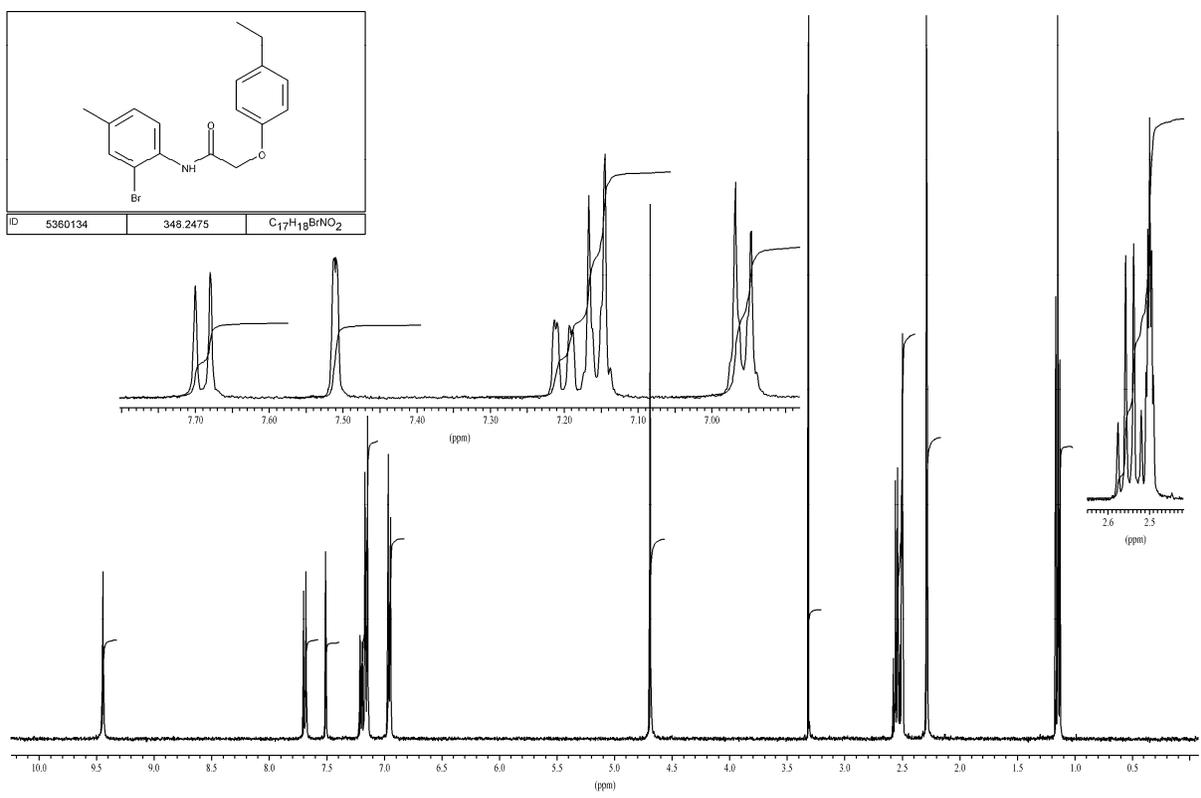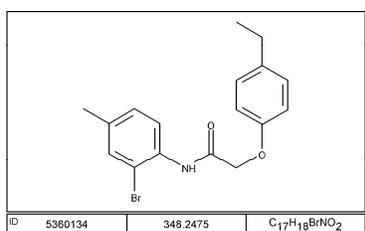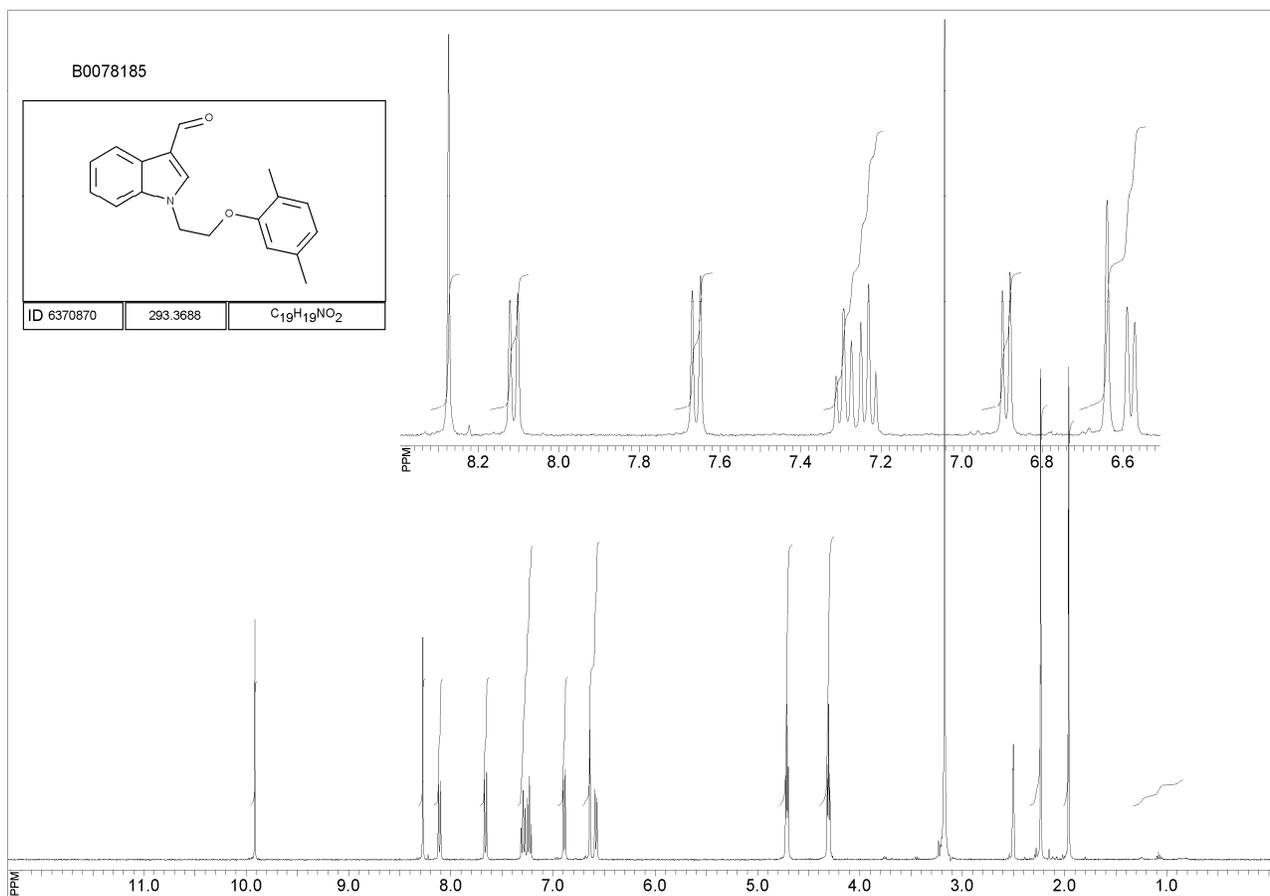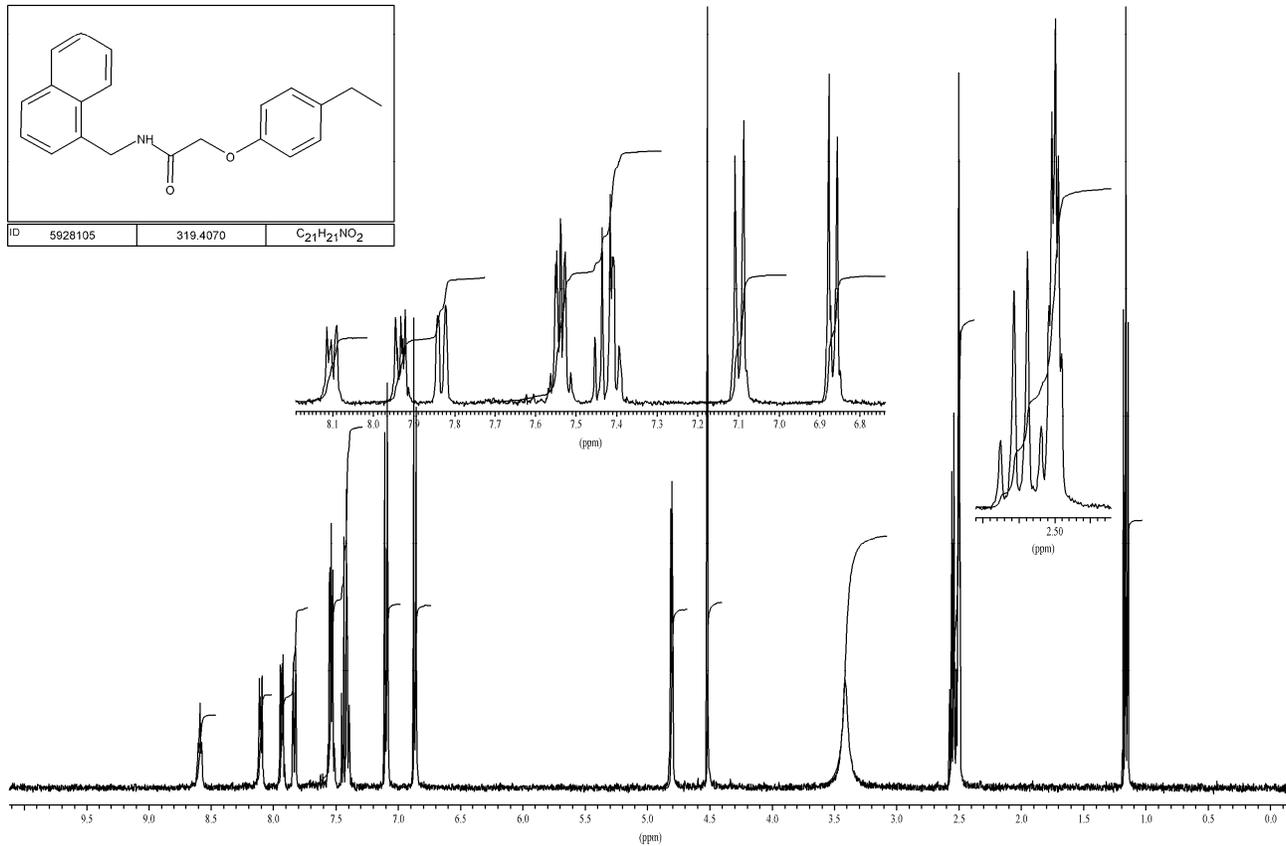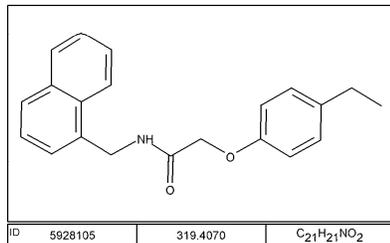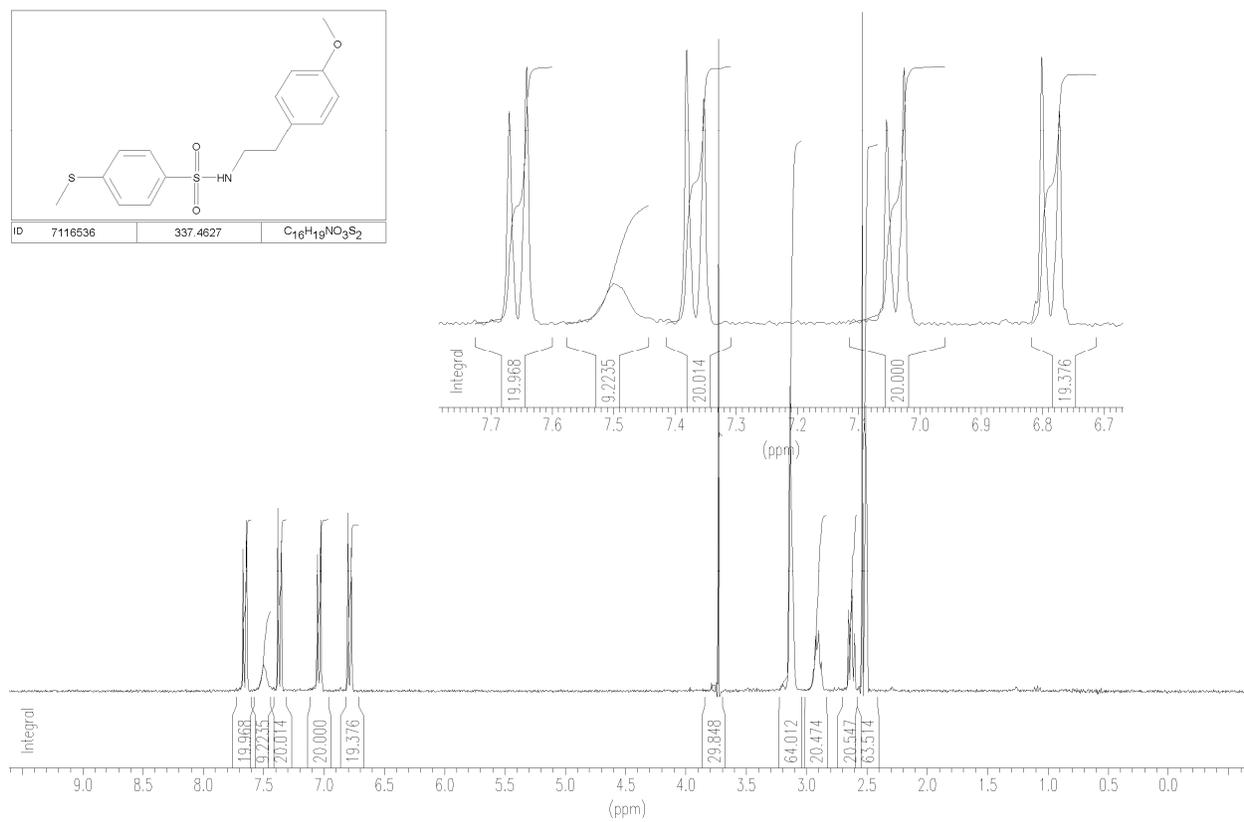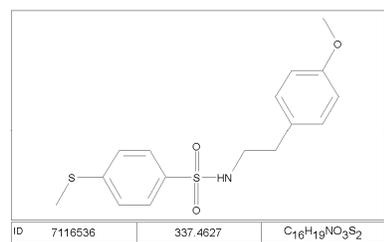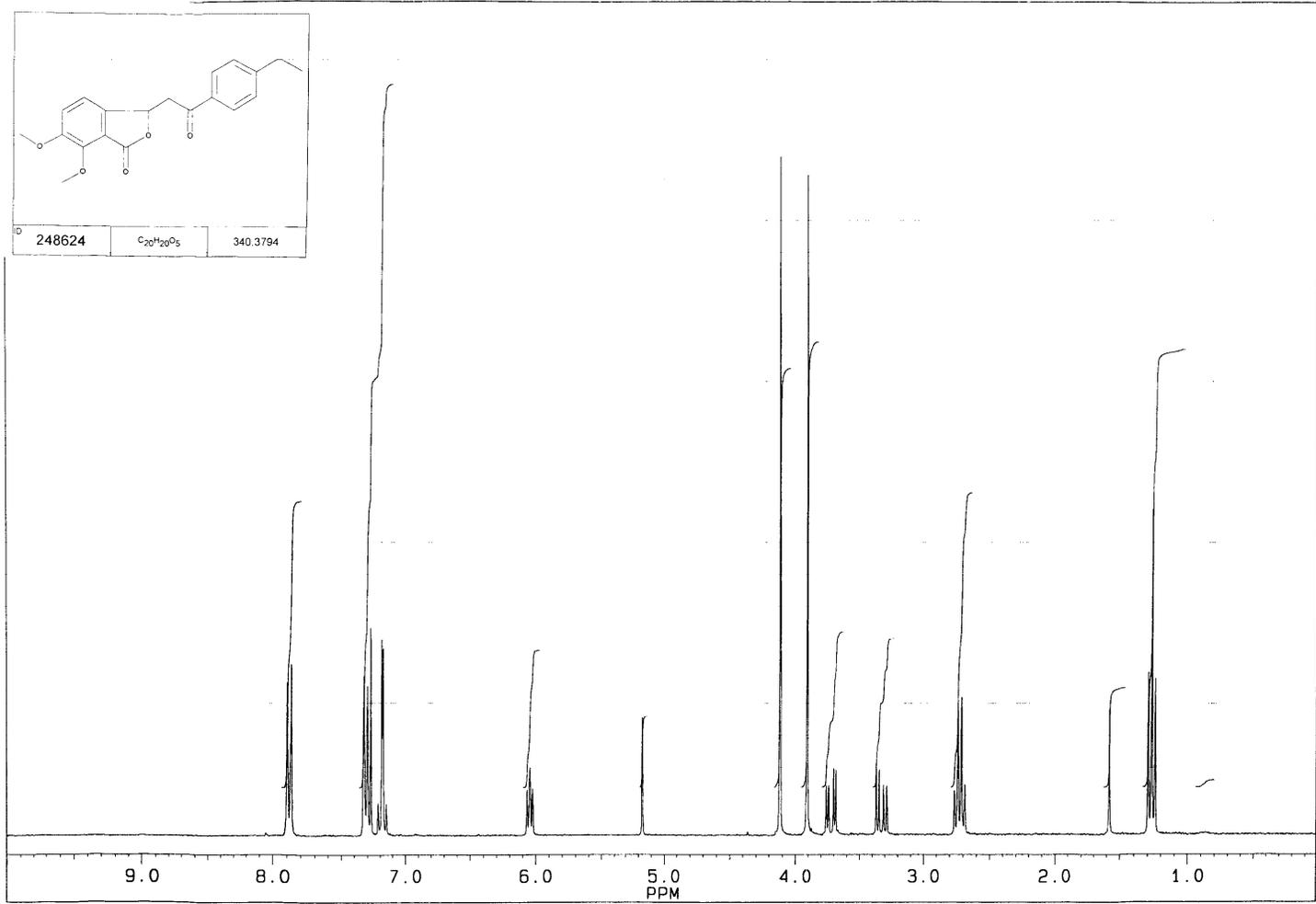
## Expression and Purification of Recombinant c-Myc353-437 and Max

The recombinant bHLHZip domain of c-Myc (c-Myc353-437) was produced from the c-Myc/pET SKB3 construct, kindly supplied by Dr. S. K. Nair (University of Illinois, Urbana-Champaign), by insertion into the pET151D vector (encoding an N-terminal hexahistidine (His 6)-tag, separated by a TEV protease digestion site) with the TOPO® ligation system. The protein was over-expressed in E.coli BL21DE3(pLysS). The 6xHis tagged human 151 amino acid "short" Max isoform [Max(S)],[6] cloned into the pQE10 vector (Qiagen, Inc.), was over-expressed in E.coli M15(pRep4). Bacterial cultures were grown at 37°C in LB medium to OD600• 0.8, then induced with 0.5 mM IPTG for 5 hours. Proteins were purified by Ni-agarose chromatography with a pH gradient elution. The 6xHis tag of c-Myc353-437 was cleaved using TEV protease (the TEV protease was expressed in a pET24 vector [courtesy of S.K. Nair] and purified by Ni-agarose chromatography under native conditions). All the proteins were further purified by HPLC and lyophilized. Protein concentrations were determined by measurement of OD280.

## Screening of c-Myc-Max dimer disruption

A preliminary screening of c-Myc-Max dimer disruption by candidate inhibitors was performed as follows. Each compound was added from a 10 mM stock solution in ethanol to reactions containing 1.5 μM c-Myc; 1X PBS buffer (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na2HPO4, 1.4 mM KH2PO4, pH=7.4) to a final concentration of 200 μM. Reactions were incubated for 20 minutes, followed by addition of 1.5 μM Max(S) – a Max isoform, which, unlike its 160 amino acid "long" isoform, is unable to homodimerize. Ellipticity at 222 nm (Θ222), indicative of α-helical content in the protein component, was then monitored for each reaction in a 1 mm path length cuvette with a JASCO J710 spectropolarimeter. Triplicate samples were tested for each compound. The dimer formation between c-Myc and Max causes an increase in α-helical content compared to the protein monomers. Samples containing active inhibitors displayed less negative Θ222 values than that observed for c-Myc-Max dimer samples, close to that observed for the weighted average of pure c-Myc and Max(S) samples. Full titrations of the inhibitors' competition for c-Myc binding against Max(S) were then performed for compounds that displayed significant disruption of c-Myc-Max dimers in the preliminary screening. Samples were similarly prepared upon serial dilutions of the inhibitors stock solutions. Three independent titrations were averaged for each compound. Experimental data were fit using the following equation, which is derived from thermodynamic considerations, relying on the presence of an equimolar ratio between c-Myc353-437 and Max(S).

$$\Theta = \Theta_0 + \Delta\Theta \bullet \left[\frac{-[I]\big/[M]-1+\sqrt{([I]/[M]+1)^2 + 4\bullet[I]/[M]\bullet(K_{comp}-1)}}{2\bullet K_{comp}-2} \right] \qquad (1)$$

Here Θ0 and ΔΘ were introduced to match the 0 to 1 scale of the thermodynamic equation (no competition to full competition) to the endpoint ellipticity values; [I] is the concentration of tested inhibitor (variable), [M] is the concentration of c-Myc and Max (fixed) and Kcomp represents the ratio between the c-Myc-inhibitor and c-Myc-Max dissociation constants (Kcomp=KDinhibitor/KDdimer). An estimate of the inhibitors affinity for c-Myc was then obtained by substituting a previously determined KDdimer value of 0.432±0.012 μM for the affinity between c-Myc and Max(p21).

## Competition assay against 10058-F4 for c-Myc353-437 binding

The same active compounds were similarly screened for competition against 10058-F4, to verify their binding to the same site as this compound on the c-Myc bHLHZip domain. 10058-F4 is one of the original small molecule c-Myc-Max heterodimer inhibitors, that binds to a short segment of the bHLH-ZIP domain between residues 402-412. The fluorescence polarization of 10058-F4 in 10 μM equimolar mixtures with c-Myc353-437, 1X PBS buffer, was monitored in the presence of varying concentrations of each screened active compound (added from 2 mM stock solutions in DMSO). A Photon Technology International Quanta Master fluorimeter equipped with polymer sheet polarizers was employed. The excitation and emission wavelengths were set at 380 nm and 468 nm respectively. Data were collected with a sample specific G-factor determination and corrected for background scattering polarization. Three independent samples were analyzed for each compound at each concentration. The 10058-F4 tumbling rate in solution decreases upon binding to c-Myc, resulting in an increase of the compound's fluorescence polarization. That is, the fluorescence polarization of 10058-F4 decreases when a compound that binds to the same region as 10058-F4 displaces it from this interaction site on c-Myc. All the screened compounds were found to displace 10058-F4 from its binding site on c-Myc. The concentration points were fit to Eq. 1 (modified to fit endpoint polarization values), and estimated KD values were obtained for each screened compound from the product of the respective Kcomp10058-F4 and KD10058-F4 (5.3±0.7 μM). These values were generally within error of the estimated KD values obtained from the c-Myc –Max dimer disruption experiment, with the exception of the values obtained for the compound 5140069, which performed slightly worse than 10058-F4 in the dimer disruption experiment, but displayed slightly tighter binding to c-Myc than 10058-F4 when directly competing against it. Control experiments were similarly performed using 10 uM c-Myc-10074-G5 mixtures and 200 uM of each studied inhibitor, and confirmed that the new inhibitors do not interfere with the binding of 10074-G5, which occurs at a different site on c-Myc.
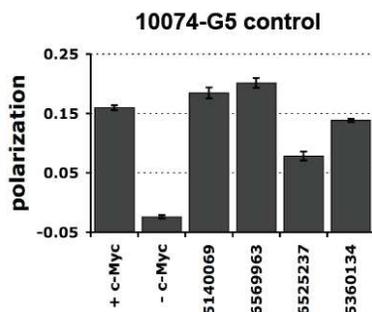
Electrophoretic Mobility Shift Assays (EMSA)

Experiments were performed as reported in our previous work.[7] Briefly, reactions containing 60 nM concentrations of c-Myc, Max(S) and varying concentrations of each tested inhibitor in a buffer containing 1X PBS, 1 mM EDTA; 0.1% NP40; 5% Glycerol; 1 mM DTT; 0.4 mg/mL BSA were incubated for 90 minutes, followed by addition of 10 nM of a synthetic double-stranded oligonucleotide containing a consensus c-Myc-binding "E-Box" element (CACGTG).[7] The binding reaction was then allowed to proceed for an additional 15 min before loading on an 8% running gel (80:1 poly acrylamide:bis-acrylamide). Gels were run at 20°C in 0.5 X TBE and scanned on a BioRad FX molecular imager. Data were analyzed with BioRad Quantity One software.[8]

Dose response experiments for hits against c-Myc-Max(S) heterodimer formation, and for hits competitively against the parent compound 10058-F4 for c-Myc monomer binding.
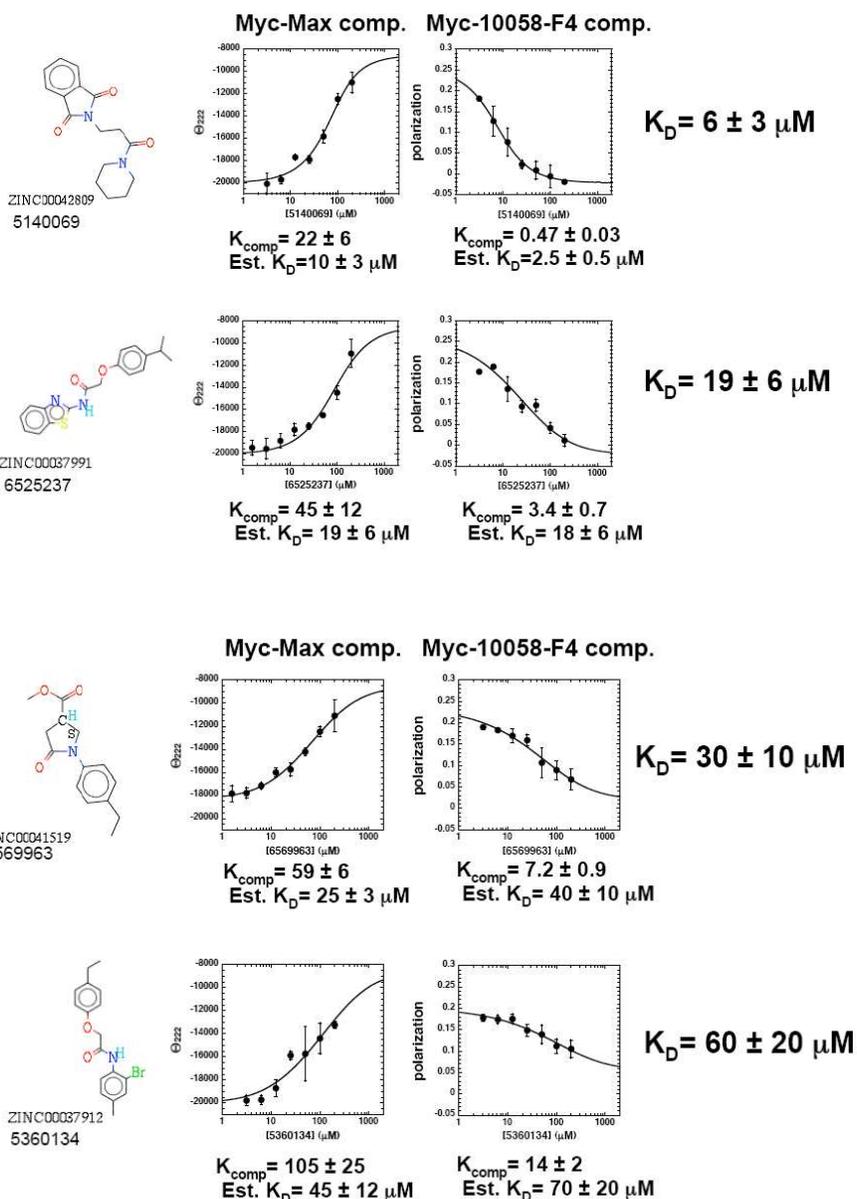

Cell-based Assay

HL60 human promyelocytic leukemia cells were grown in RPMI medium supplemented with 10% fetal calf serum, 100 U/ml penicillin G, and 100 •g/ml streptomycin (all from Mediatech, Inc., Herndon,VA). Rat fibroblast lines were grown under similar conditions in Dulbecco's modified minimal essential medium. To determine the effects of Myc-Max compounds on HL60 cell growth, logarithmically growing cells (>90% viability) were re-suspended in fresh medium. 4 ml (a total of 16,000 cells) were then seeded into 6-well plates in the presence of the indicated amount of Myc-Max compound. In all cases, 10058-F4 was included as a reference compound. Daily cell counts were performed manually in triplicate on a hemacytometer using trypan blue exclusion. Viabilities exceeded 85% throughout the course of the experiment. Each experiment was repeated at least two additional times with results similar to those depicted here obtained.

The number of viable cells was determined by using the colorimetric MTS assay (CellTiter 96® AQueous Assay). The mechanism behind this assay is that metabolically active cells will react with a tetrazolium salt in the MTS reagent to produce a soluble formazan dye that can be absorbed at 490 nm. Briefly, One hundred microliter samples of an exponentially growing cell suspension (1-2 x 10³ cells) were seeded into a 96-well microliter plate and allowed to achieve logarithmic growth for 24 hr. Fresh medium containing the indicated concentrations of ZINC compounds was then added. The number of viable cells was determined by MTS assay every day. 20 µl of CellTiter 96R AQueous One Solution were added to each well and the plates were incubated for an additional 1 h at 37°C. Absorbance was measured at 490 nm with a 96-well plate reader. Each experiment was performed in three replicate wells for each drug concentration.

Dose-response profiles of each of the compounds on HL60 cells, and with TGR1 (normal rat fibroblasts) along with TGR1 knockout cells with overexpressed HMGA1b (KO+HMG). IC50s here were calculated based on dose-response profiles on day 4 (5) following the addition of each compound. Representative experiments are shown, with each compound being assayed in separate experiments on two to four additional occasions.
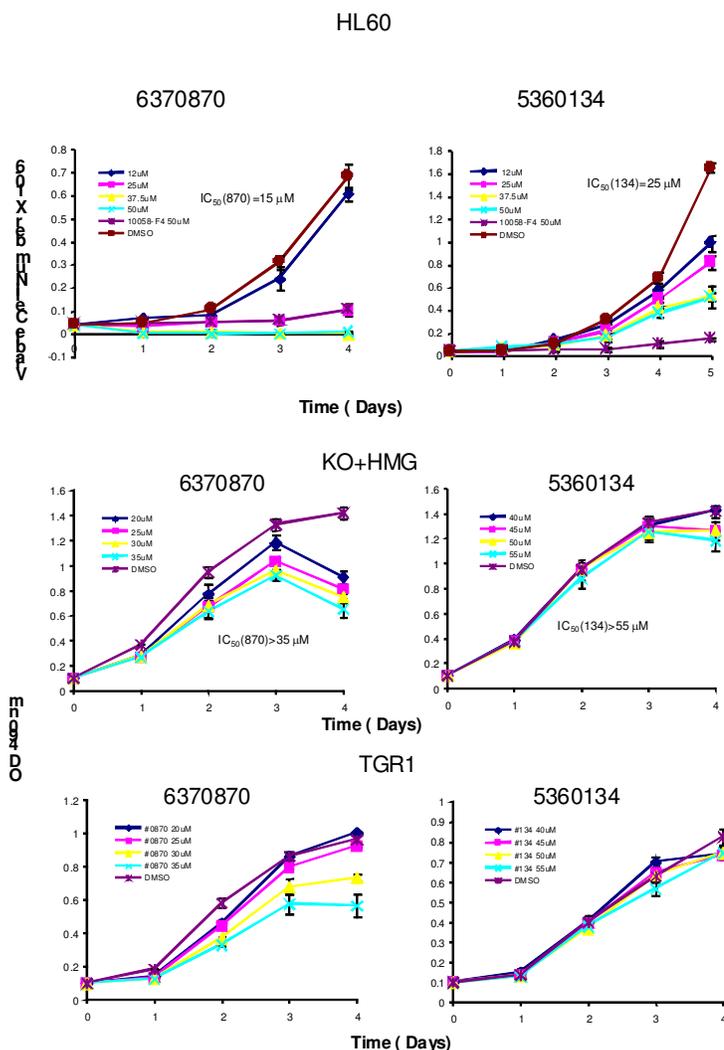
REFERENCES

(1) Clark, R. D.; Abrahamian, E. Using a staged multi-objective optimization approach to find selective pharmacophore models. *J. Comput. Aided Mol. Des* **2008**.
(2) Richmond, N. J.; Abrams, C. A.; Wolohan, P. R.; Abrahamian, E.; Willett, P.; Clark, R. D. GALAHAD: 1. pharmacophore identification by hypermolecular alignment of ligands in 3D. *J. Comput. Aided Mol. Des* **2006,** *20*, 567-587.
(3) Shepphird, J. K.; Clark, R. D. A marriage made in torsional space: using GALAHAD models to drive pharmacophore multiplet searches. *J. Comput. Aided Mol. Des* **2006,** *20*, 763-771.
(4) Cottrell, S. J.; Gillet, V. J.; Taylor, R.; Wilton, D. J. Generation of multiple pharmacophore hypotheses using multiobjective optimisation techniques. *J. Comput. Aided Mol. Des* **2004,** *18*, 665-682.
(5) SYBYL 8.0. www.tripos.com
(6) Zhang, H.; Fan, S.; Prochownik, E. V. Distinct roles for MAX protein isoforms in proliferation and apoptosis. *J. Biol. Chem* **1997,** *272*, 17416-17424.
(7) Wang, H.; Hammoudeh, D. I.; Follis, A. V.; Reese, B. E.; Lazo, J. S.; Metallo, S. J.; Prochownik, E. V. Improved low molecular weight Myc-Max inhibitors. *Mol. Cancer Ther.* **2007,** *6*, 2399-2408.
(8) BioRad Quantity One. www.biocompare.com