TIDEA™ Predicts Binding Affinity for Diverse Small Ligand/Binding Site Interactions from Ligand Structure Alone

COMP 147

Darryl Rideout, Focus Synthesis LLC, 11575 Sorrento Valley Road, Ste. 212, San Diego, CA 92121, Email: focus_synthesis@focussynthesis.com

REVISED ABSTRACT

TIDEA (Target-Independent Drug Enhancement Algorithm) is a proprietary algorithm used to estimate potency and hit rates for small molecule ligand/protein target interactions from ligand structure alone, without any knowledge of target structure or SAR. We used a Learning Set of 120 diverse bioactive ligands with known potency values for 56 distinct protein targets to develop TIDEA. The TIDEA algorithm calculates a score solely from the ligand structure, which correlates with potency (pIC50/pKi) for a wide array of ligand/target interactions. In a Test Set of 80 ligands with an average of 4 ligands/target (i.e. 20 targets), which had no overlap with the Learning Set, the percentage of subnanomolar ligands was 11-fold greater for high TIDEA scores (>9.5). with statistical significance (Chi Square p value <0.01). When applied to an Ultradiverse Set of 65 ligands with a different target for every ligand, statistical significance was observed for the percentage of potent (< 100nM) ligands (Chi Square p value < 0.03), and for the increase in average potency (T-Test p value < 0.015). TIDEA has the potential to accelerate drug discovery by addressing key limitations of structure-based and SAR-based methods, because TIDEA shows a selection bias for potent molecules while maintaining diversity.

INTRODUCTION

Traditional methods of achieving selection bias toward more potent small molecules (docking, QSAR, analog-based approaches, etc (1-7)) are limited in that they require target macromolecule structure or SAR knowledge, and because they restrict molecules diversity. TIDEA (Target Independent Drug Enhancement Algorithm), a virtual screening algorithm for bioactive small molecules that uses a proprietary algorithm developed at Focus Synthesis LLC, requires no knowledge of the target macromolecule or active site is required. Ligand potency (ClogIC50 or ClogKi) increases with increasing TIDEA scores for a wide array of ligands with scaffolds and diverse targets. This TIDEA score/potency correlation is statistically significant. TIDEA correlates with adhesiveness potential for bioactive ligands that fit some macromolecular binding site, rather than specific shape complementarity addressed by traditional approaches. TIDEA has the potential to increase the success rate in early stage drug discovery and development and decreasing late stage failures by providing a broader range of viable drug candidates in early discovery.

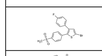
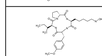
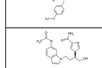
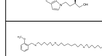
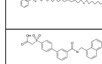
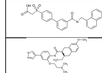
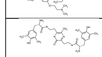
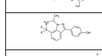
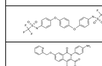
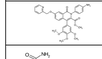
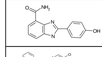
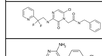
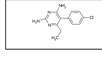
METHODS

Learning Set for Creation of the TIDEA algorithm. (See also Scheme 1). The Learning Set used to develop TIDEA was constructed from 56 distinct subsets of 2 to 4 ligands per subset reported in the literature, for a total of 120 small molecule ligands (8). Each of the 56 ligand subsets has a distinct scaffold and binds to a distinct macromolecular target, to ensure that potency/TIDEA score relationships would not be related to specific ligand/target shape complementarity.

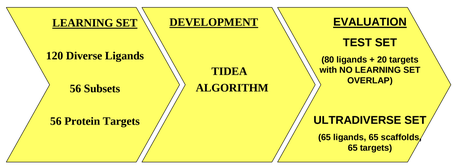
Test Set. The Test Set consists of 80 diverse small molecule ligands (FW between 300 and 500) that bind 20 distinct targets (8). We deliberately avoided overlap between the 56 Learning Set targets and the 20 Test Set targets to ensure a rigorous test of the TIDEA technology. The ligand scaffold types are also distinct for the learning set and test set.

Ultradiverse Set In order to demonstrate unequivocally that TIDEA is distinct from ligand/target complementarity and SAR-based approaches (e.g. Ensemble Docking, Inverse QSAR, CoMFA, SoMFA, AFMoC, Similarity-Potency Trees, bioisosterism (1-7)) we chose 1 molecule at random for each of 65 targets in a set of 300 small molecule ligands yielding the Ultradiverse Set of 65 molecules. Each of the 65 molecules in the Ultradiverse set represents a distinct scaffold as well as a distinct macromolecular target. To insure optimal diversity, the 65 different targets are derived from over 13 different classes (protease inhibitors, nuclear receptors, kinases etc.). Table 1 shows 13 ligands from the Ultradiverse set, each with a distinct scaffold, a distinct target, and a distinct target class.

Table 1. REPRESENTATIVE STRUCTURES AND TARGETS FOR 13 TARGET CLASSES

Structure	Target	Target Class	TIDEA Score	Potency (nM)	Potency (ClogIC50 or ClogKi)
	COX-2	Cyclooxygenases	2	10	8.00
	Histone Deacetylase type 1	Deacetylases	10	6.3	8.20
	Adenosine deaminase	Deaminases	7	1300	5.89
	Human AChE	Esterases	3	5370	5.27
	Human Papillomavirus E1 Helicase	Helicases	4	4.3	8.37
	ROCK-II	Kinases	12	9	8.05
	Delta-opioid receptor	Neurotransmitter Receptors	10	0.115	9.94
	Estrogen receptor	Nuclear Receptors	8	3	8.52
	PTP1B	Phosphatases	5	2500	5.60
	PDE5	Phosphodiesterases	9	1	9.00
	Poly (ADP-Ribose) polymerase	Polymerases	4	6	8.22
	Thrombin	Proteases	8	12	7.92
	PfDHFR	Reductases	5	0.3	9.52

SCHEME 1: DEVELOPMENT AND TESTING OF TIDEA ALGORITHM



The TIDEA Algorithm was created using the **Learning Set** only, and tested on the very distinct Test Set: It was tested using the **Test Set**, which has no overlap with the **Learning Set**.

The **Ultradiverse Set** was derived from the Learning Set, the Test Set and other targets. Every ligand has a distinct target and scaffold.

RESULTS (TEST SET)

TIDEA effectiveness for the Test Set: There is a significant correlation between TIDEA score and potency for the Test Set, as shown in Figure 1. Seven out of eight macromolecular ligands have TIDEA scores greater than 9.5 in the Test Set. Average Potency increases disproportionately with TIDEA score than with molecular weight, as shown in Figure 2. TIDEA scores above 11.5 yield a significantly larger number of potent (< 5 nM) inhibitors without compromising diversity, as shown in Figure 3. Statistical significance. The difference between the average potency (ClogP or ClogKi) for TIDEA score <9.5 (7.11) and TIDEA score >9.5 (7.89) is statistically significant by the T test ($p<0.01$). The percentage of subnanomolar inhibitors is increased more than 10 –fold when the higher (TIDEA score >9.5, 22% subnanomolar) score range is compared to the lower (TIDEA score < 9.5, 2% subnanomolar). This enrichment of subnanomolar inhibitors is statistically significant by the Chi Square method (Yates p-value < 0.01).

Score	Potent (<1nM)	Weaker (≥1nM)
< 9.5	1	52
> 9.5	6	21
Yates Chi Square: 6.9 Yates p-value = 0.009		

RESULTS (ULTRADIVERSE SET)

Figure 3 shows all of the Test Set ligands with TIDEA score > 11.5. This subset of 8 molecules, with only 1.6 molecules per target (5 different targets), is actually more diverse than the entire Test Set, with 4 molecules per target. Molecular diversity is maintained after selecting molecules with high TIDEA scores.

TIDEA effectiveness for Ultradiverse Set. Figures 4 and 5 show the trend toward higher potency with higher TIDEA score (>9.5) for the Ultradiverse Set of 65 ligands with 65 scaffolds and 65 targets.

Statistical significance. The higher average potency for ligands with TIDEA score > 9.5 is statistically significant by the T test ($p< 0.015$, Figure 4). The enrichment of more potent ligands (IC50 or Ki<100 nM) at TIDEA score > 9.5 is statistically significant by Chi Square analysis (p value<0.03).

SCORE	#> 100nM	#< 100nM
<9.5	23	25
>9.5	3	14
Chi square=6.323 p value=0.029		

RESULTS: DRUG-LIKE VS. NON DRUG-LIKE LIGANDS

Drug-like molecules have higher TIDEA scores. The TIDEA scores were calculated for a drug/non-drug test set of 140 molecules, with FW betseen 100 and 860, consisting of 70 clinical drugs and 70 non-drugs (1:1 ratio). The non-drugs were organic molecules chosen at random from the Accelrys Available Chemicals Directory®. For higher TDIEA scores (>9.5), the drug/non-drug ration increased 24-fold relative to lower TIDEA scores (<9.5). The difference is statistically significant by the Chi-Square test (p <0.0001).

SCORE	<9.5	>9.5
drugs	52	18
non-drugs	69	1
Chi Square= 17.6 *p*-value= 0.000027		

Figure 1. HIGHER TIDEA SCORES CORRELATE WITH HIGHER POTENCY IN THE TEST SET

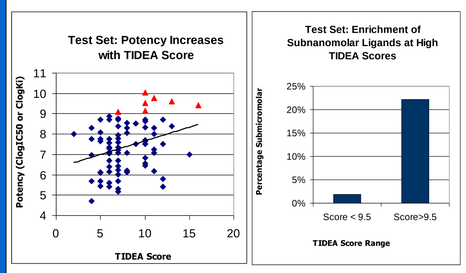


Figure 2. FOR THE TEST SET, THE AVERAGE POTENCY INCREASES 11-FOLD WITH TIDEA SCORE, BUT NOT WITH FW

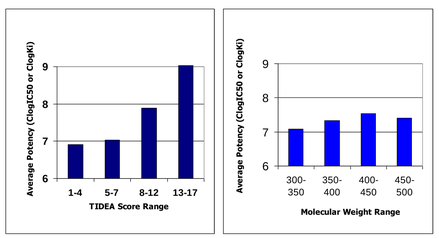


Figure 4. EVEN IN THE ULTRADIVERSE SET, WHERE EVERY MOLECULE IN THE SET IS FROM A DIFFERENT TARGET, HIGH TIDEA SCORES ARE ASSOCIATED WITH HIGHER POTENCY

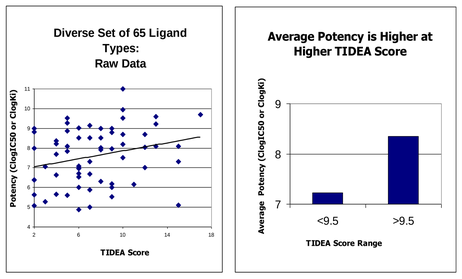


Figure 5. AVERAGE POTENCY AND % OF SUBNANOMOLAR INHIBITORS INCREASE WITH TIDEA SCORE FOR THE ULTRADIVERSE SET

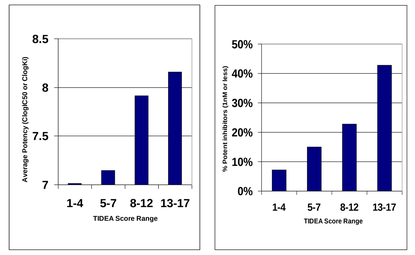
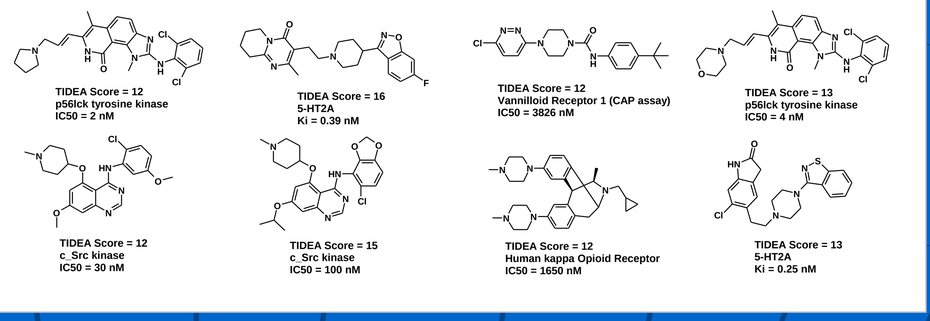


Figure 3. TEST SET MOLECULES WITH HIGHEST TIDEA SCORE MAINTAIN DIVERSITY WITH INCREASED POTENCY



DISCUSSION

TIDEA is completely target-independent and scaffold-independent, ensuring diversity at high TIDEA scores, unlike 3D protein structure-based and SAR-based approaches. Traditional approaches to enhancing hit rates in libraries all rely on knowledge of the 3D structure of the target macromolecule, ligand/target binding affinity (SAR) for related small molecules, or both (1-7). TIDEA requires no knowledge of the target or ligand/target interaction. The TIDEA score correlates with potency even when every ligand in the dataset binds to a different target (as in the Ultradiverse Set). Creating such a correlation with structure-based or SAR-based methods would be impractical, and at early stages of discovery, there is often no target structure or SAR information available. For universal libraries and corporate compound collections, TIDEA will have considerable advantages in terms of molecular diversity and utility in the absence of target information.

CONCLUSION

- Potency is significantly higher at high TIDEA scores, in terms of average potency (statistically significant by the T-test) and fraction of potent inhibitors (significant by Chi Square analysis).
- Application of TIDEA requires only the structure of each ligand, while current approaches require a 3D protein structure or structures and potencies of several similarligands with diverse bioactivities for each target type.
- TIDEA will be ideal for increasing hit rates when screening universal libraries and corporate compound collections with multiple targets because diversity is maintained at high TIDEA scores, in contrast to traditional target-structure-based tools that focus on a single ligand target interaction and restrict diversity. TIDEA can also help enhance hit rates even when there is no target structure or SAR information available.
- TIDEA has the potential to enhance and synergize the selection bias of traditional drug discovery tools without compromising diversity.

ONGOING DEVELOPMENT AND FUTURE PLANS

- To demonstrate the ability of TIDEA to increase hit rates in diverse small molecule arrangements in collaborative arrangements and independent, blinded studies (9).
- To design a diverse set of small molecules with high TIDEA scores (>9.5) for use as screening compounds (underway) (10).
- To accelerate Fragment-Based Lead Discovery by modifying the current TIDEA algorithm to select molecules with high ligand efficiency (underway)(10).

References:

1. RE Amaro and WW Li, Current Topics in Medicinal Chemistry V 10 pp 3-13 (2010)
2. M Wawer and J. Bajorath, J. Chem. Inf. Model. V50 pp 1395-1409 (2010)
3. L Tan, J Batista, J Bajorath, J. Chem. Inf. Model. V50 pp 1042-1052 (2010)
4. EM Nova LR de Pouplana X Barril and M Orozco, J. Chem. Theory Comput V6 pp 2547-2557 (2010)
5. GA Patani and EJ LaVoie, Chem. Rev. V96 pp 3147-3178 (1996)
6. J Verma, VM Khedkar, and EC Coutinho, Current Topics in Medicinal Chemistry V10 pp 95-116 (2010)
7. RE Amaro and WW Li, Current Topics in Medicinal Chemistry V10 pp 3-13 (2010)
8. The literature references and structures used in the Test Set, Learning Set, and Ultradiverse Set are too numerous to list here (over 100). A complete listing of literature references, ligand structures, targets, and potency values will be provided on request. Please email requests with your preferred format (.sdf, .pdf, or hard copy) to Darryl Rideout at focus_synthesis@focussynthesis.com.
9. Focus Synthesis is eager to establish collaborative arrangements and support independent, blinded studies of the TIDEA technology. Scientists and corporations with their own non-proprietary bioactive compound sets and compounds selected from the literature should contact Darryl Rideout at focus_ synthesis@focussynthesis.com.
10. Please visit the following page for further developments, and TIDEA-related services and products:
www.focussynthesis.com/TIDEA
www.focussynthesis.com/Products/TIDEA_screening_compounds

FOCUS SYNTHESIS

www.FocusSynthesis.com