

Dengue Drug Discovery: Using Global Computing to Combat a Global Disease

Stan Watowich
University of Texas Medical Branch



Dengue outbreak in Cambodia (C. Peskett, Doctors w/o Borders, 2007)



Child with dengue fever (Bairo Pite Clinic, Timor-Leste, Robin Taudevin)

utmb Health Biochemistry & Molecular Biology

Austin Forum: April 05, 2011

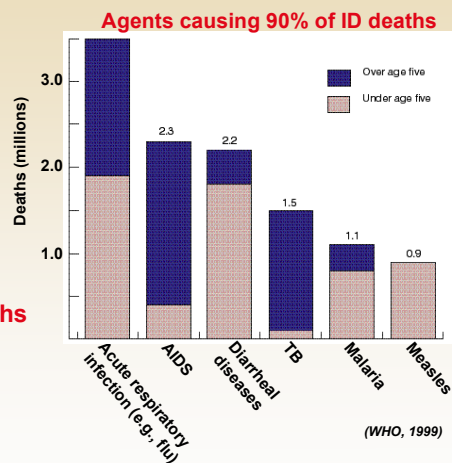
Infectious diseases take a large toll

Globally

- pathogens cause ~22% of all deaths
- **HIV**: 5% of sub-Sahara seropositive
- **HCV**: 3% of world seropositive
>3M infections/yr
- **DENV**: 50M infections/yr
>1M clinical cases/yr

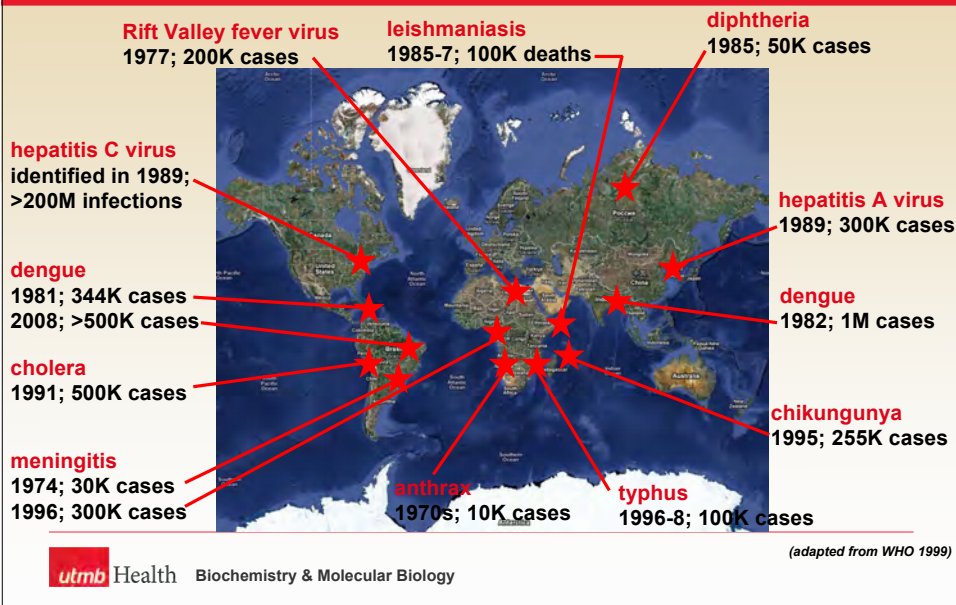
United States

- viral diseases cause ~5% of US deaths
- **HIV**: 0.3% of population infected
- **HCV**: 2% of population seropositive
- **influenza**: 30-50K deaths/yr



utmb Health Biochemistry & Molecular Biology

Sampling of infectious disease outbreaks

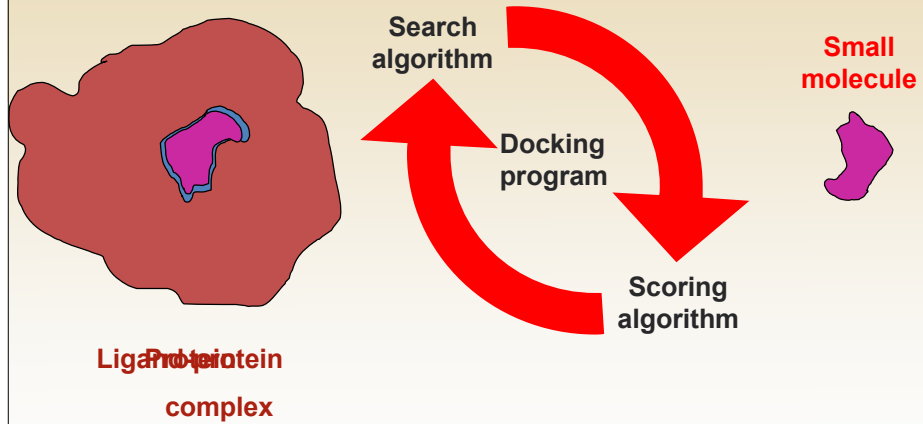


Challenge

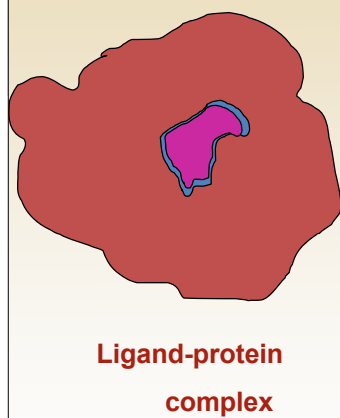
How to find drug leads for neglected emerging world diseases (e.g., dengue) given limited resources?

One approach:
develop better tools for *in silico* drug discovery

Overview: *In silico* docking



Overview: *In silico* docking



Target function to guide docking should be:

$$\Delta G = \Delta H - T \Delta S$$

Approximated as:

$$\Delta G_{binding} = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{tor} + \Delta G_{sol}$$

Typical empirical target function

- does not adequately treat solvation and entropic terms

$$\Delta G^{\text{Binding}} = \Delta G_{\text{vdw}} + \Delta G_{\text{hbond}} + \Delta G_{\text{elec}} + \Delta G_{\text{tor}} + \Delta G_{\text{sol}}$$

$$W_{\text{vdw}} \sum_{ij} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right)$$

$$W_{\text{elec}} \sum_{ij} \frac{q_i q_j}{\epsilon r_{ij}}$$

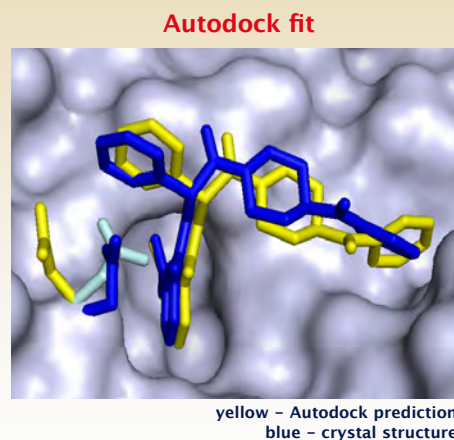
$$W_{\text{sol}} \sum_{ij} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)}$$

$$W_{\text{hbond}} \sum_{ij} E(i) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right)$$

$$W_{\text{tor}} N_{\text{tor}}$$

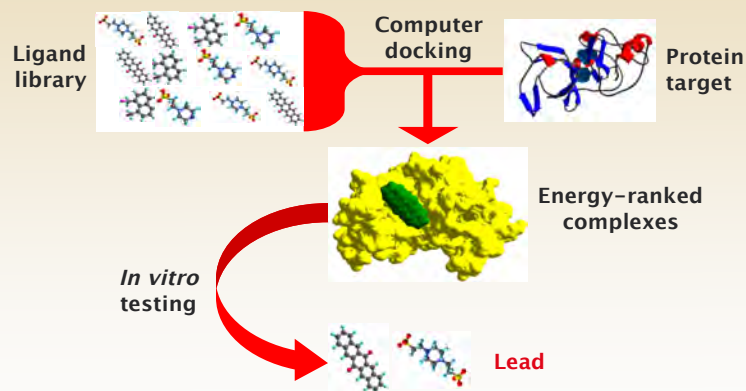
Docking can reproduce crystal structures

- flexible small molecules can be positioned correctly within a binding site
- small molecules fit within 0.5-1.2 Å of the X-ray structure conformation



Can virtual screening find drug leads?

- preliminary EUDOC screen yielded 3 dengue protease inhibitors from ~30 tested hits (Tomlinson et al. 2009)



Can virtual screening find drug leads?

- preliminary EUDOC screen yielded 3 dengue protease inhibitors from ~30 hits (Tomlinson et al. 2009)

Can we find more leads?

Can virtual screening find drug leads?

- preliminary EUDOC screen yielded 3 dengue protease inhibitors from ~30 hits (Tomlinson et al. 2009)

Can we find more leads?

- larger libraries should increase number of leads
 - 2.3M “drug-like, lead-like” molecules from ZINC (Irwin & Shoichet, 2005)

Can virtual screening find drug leads?

- preliminary EUDOC screen yielded 3 dengue protease inhibitors from ~30 hits (Tomlinson et al. 2009)

Can we find more leads?

- larger libraries should increase number of leads
 - 2.3 million “drug-like, lead-like” cmds from ZINC (Irwin & Shoichet, 2005)

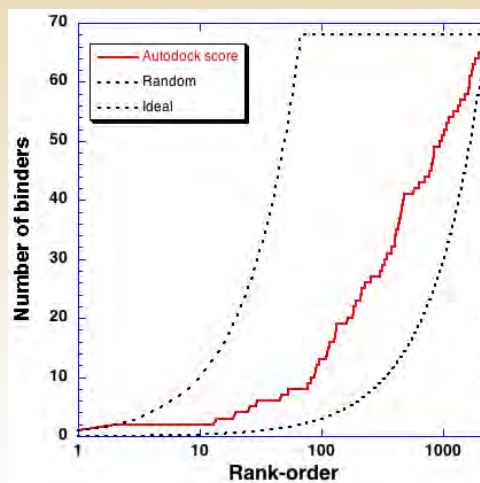
Can we find leads more effectively?

- use more accurate virtual screening programs
 - is there a “best” docking-based virtual screening program?

Virtual screening performance

Estrogen receptor - agonist

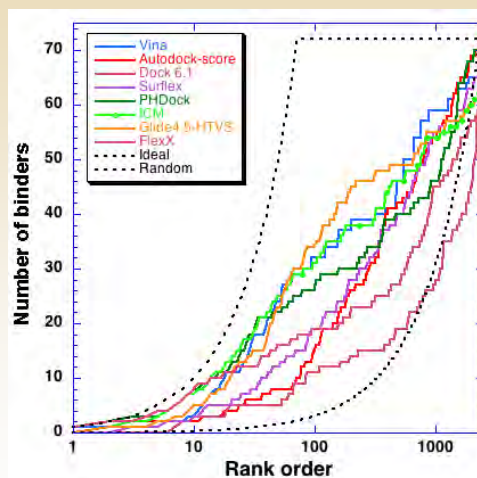
- extracted from DUD (Irwin, UCSF)
- ER agonist library
 - 68 known binders
 - ~2350 “decoys”



Comparison of virtual screening programs

Estrogen receptor - agonist

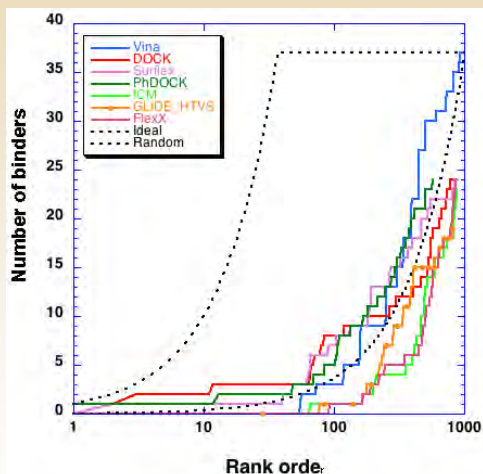
- comparison of 8 docking programs
- programs include free and commercial software
- data for several programs provided by J. Cross (J. Chem. Inf. Model. 49, 1455, 2009)
- success rates range from 10-40% (at 100 hits)



Intractable systems for screening

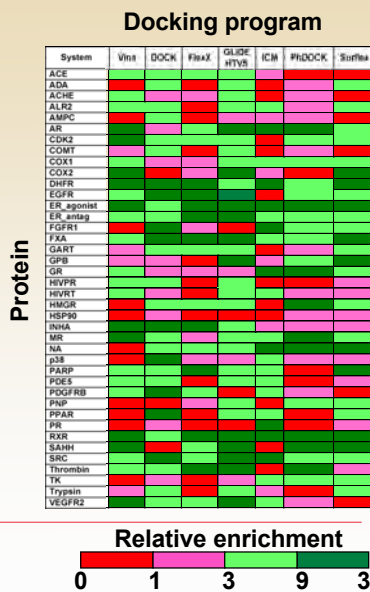
HSP90

- target & library extracted from DUD
- library
 - 30 binders
 - 1000 decoys
- success rates for almost all programs near random (or less)



Comparison of screening programs

- 40 proteins/libraries from DUD
- several 1000s molecules per library
- ~3% binders in libraries
- data for several programs provided by J. Cross (J. Chem. Inf. Model. 49, 1455, 2009)
- success rates taken @ 100 hits
- no program showed strong early enrichment for all systems
- a few proteins (FXA, ER- α , RXR) had good enrichments with all programs



Can virtual screening find drug leads?

- preliminary EUDOC screen yielded 3 dengue protease inhibitors from ~30 hits (Tomlinson et al. 2009)

Can we find more leads?

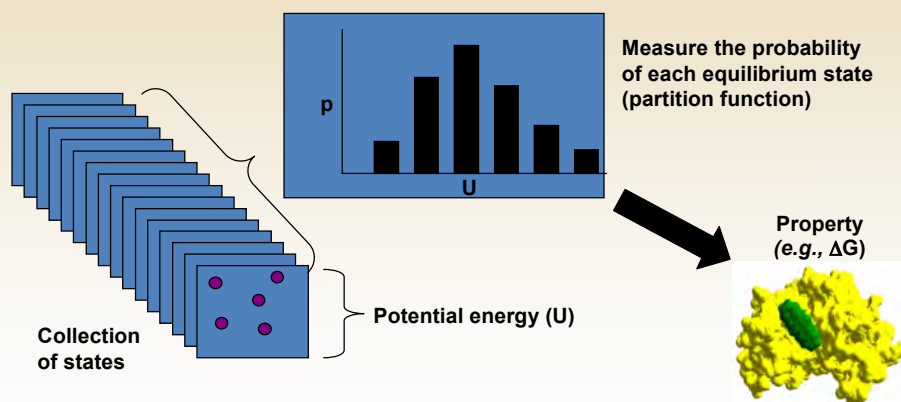
- larger libraries should increase number of leads
 - 2.3 million “drug-like, lead-like” cmds from ZINC (Irwin & Shoichet, 2005)

Can we find leads more efficiently?

- use more effective virtual screening programs
 - is there a “best” docking-based virtual screening program?
- develop more effective virtual screening approaches
 - replace empirical scoring functions with $\Delta G_{\text{binding}}$ calculations based on sound statistical thermodynamic principles

Calculating accurate $\Delta G_{\text{binding}}$

- rooted in statistical mechanical theory and partition functions
- calculated using perturbation methods



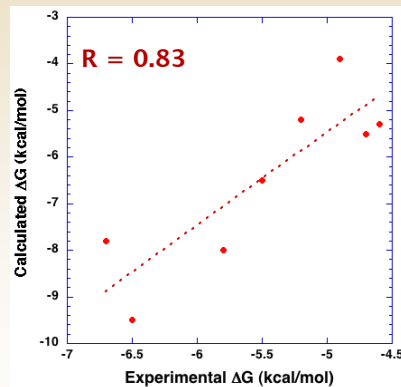
Incorporating $\Delta G_{\text{binding}}$ into docking

Two step process

- 1) docking to determine poses
- 2) rescore poses with mean-field FEB calculations to determine $\Delta G_{\text{binding}}$

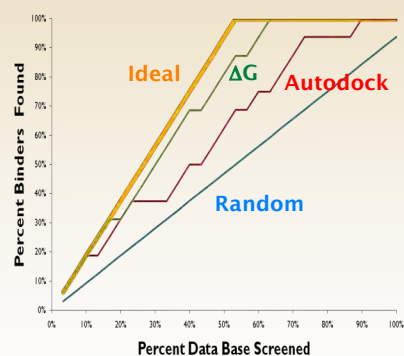
Validation: lysozyme-ligand system

- good correlation between exp. and calculated ΔG
- mean difference ~ 1.3 kcal/mol



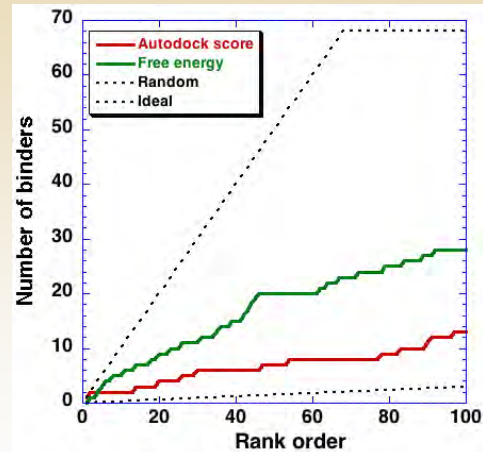
Incorporating $\Delta G_{\text{binding}}$ into screening

- target protein: L99A lysozyme
- library: 16 binders, 14 non-binders
- ΔG screening improved enrichment relative to Autodock
- ΔG screening had 83% accuracy



ΔG screening improved enrichment

- ER-agonist and library from DUD
- ~2500 ΔG calculations
- 2- to 3-fold improvement in success rate relative to Autodock
- similar improvements seen with trypsin
- continue testing all DUD systems



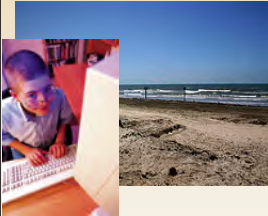
Technical challenge

- docking: ~1 minute per ligand
100,000 ligands -> 10 weeks CPU time
- ΔG calculation: ~3 weeks per ligand
100,000 ligands -> 5800 years CPU time

How to provide the 1,000s of CPU yrs required for ΔG -based drug discovery?

Distributed computing for ΔG screening

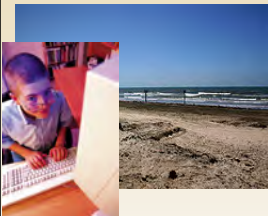
- drug discovery calculations are embarrassingly course-grained
- ideal problem for distributed computing



- project development
- target design
- hit validation

Global computing for ΔG screening

- drug discovery calculations are embarrassingly course-grained
- ideal problem for distributed computing on global computer network



- pre & post-processing
- off-site data storage & backup

Distributed computing for ΔG screening

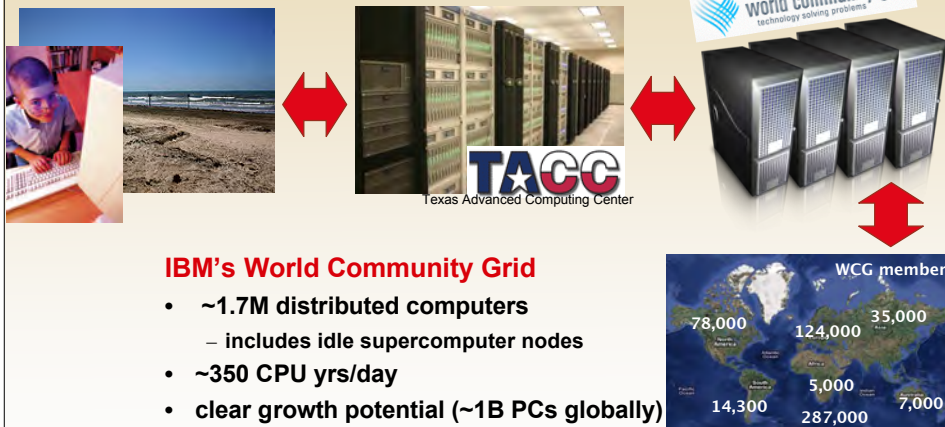
- drug discovery calculations are embarrassingly course-grained
- ideal problem for distributed computing



utmb Health Biochemistry & Molecular Biology

Distributed computing for ΔG screening

- drug discovery calculations are embarrassingly course-grained
- ideal problem for distributed computing



utmb Health Biochemistry & Molecular Biology

“Discovering Dengue Drugs-Together”

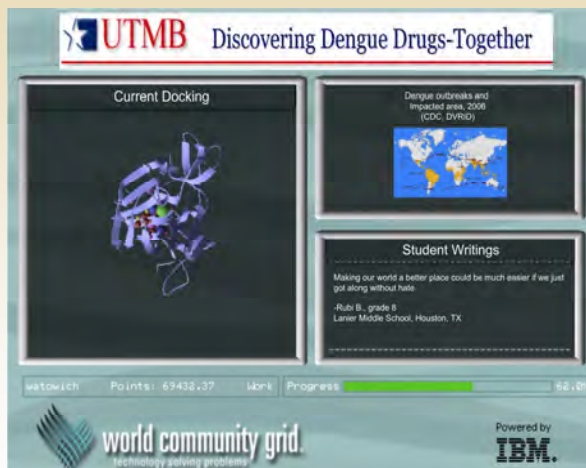
IBM World Community Grid project

Phase 1 (docking)

- completed 15 targets
- 2.3M molecule library
- 11,700 CPU yrs
- equivalent to a 13,000 node HPC

Phase 2 (ΔG calculations)

- launched Jan'11
- new leads expected Jun'11



utmb Health Biochemistry & Molecular Biology

Global access to discovery tools

World Community Grid is an invaluable resource, but ...

Can we make in silico drug discovery more accessible to resource-limited researchers?

utmb Health Biochemistry & Molecular Biology

Freely accessible drug discovery portal

- established a Web interface for Vina-based screening (docking.utmb.edu)
- designed to reduce risk, expense, & barriers to virtual drug discovery
- inputs: – PDB structure
– coordinate in the active site
- includes 3 drug-like, commercially available libraries (5K, 23K, 600K)

Drawbacks: – queue backlog
– 18–36 hr turnaround of small libraries

Job Name	Job Owner	Status	Remove
NUTAN	prakash	95% Done	Remove
Samgny	baker	Not Started	Remove
6022711	prakash	Not Started	Remove

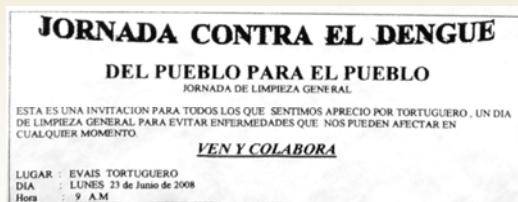
Job Name	Job Owner	Completed	Download	Remove
Khuu8bu	Khuu8bu	02/27/11	Download	Remove
Tim_sun	baker	02/27/11	Download	Remove
EXQ	udatha	02/26/11	Download	Remove
SACHINI	kaushik	02/26/11	Download	Remove
NIAD	colucci	02/23/11	Download	Remove
test	prakash	02/19/11	Download	Remove
pubs	prakash	02/19/11	Download	Remove
ALEX_1E1_AR_CB	watowich	02/18/11	Download	Remove
ALEX_1E1_AR_MB	isabell	02/15/11	Download	Remove
diabetes	praga	02/15/11	Download	Remove
ST_WN_CB	springman2	02/11/11	Download	Remove
ST_WN_MB	springman1	02/08/11	Download	Remove
SW_5002_CB2K	springman	02/07/11	Download	Remove
SW_5002_MB4K	watowich	02/04/11	Download	Remove

TACC to the rescue ...

Can we easily access TACC for drug discovery?

- moving portal to TACC Lonestar4
- easy screening setup
- rapid screening of ~100K ligands
- tools for immediate analysis
- expected to provide 500-fold (?) speedup relative to UTMB portal

Targeting dengue disease



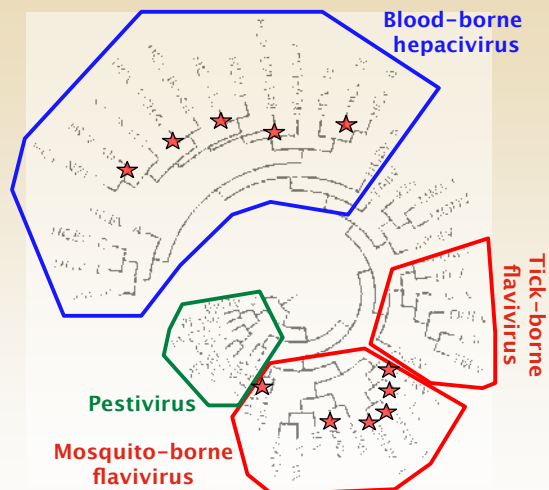
Tortuguero, Costa Rica



utmb Health Biochemistry & Molecular Biology

Dengue related to many pathogens

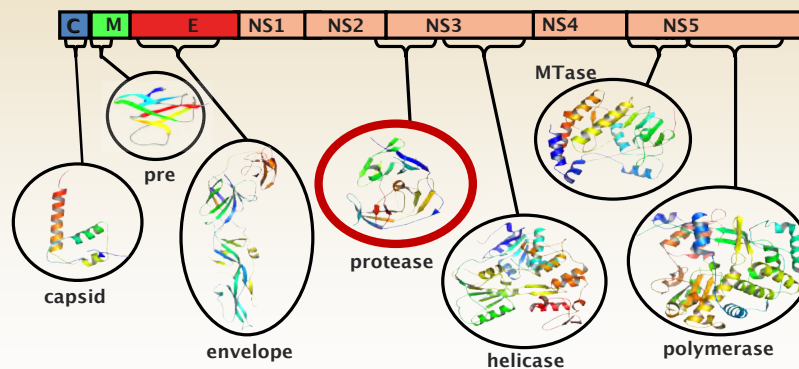
- *Flaviviridae* family
- major diseases
 - dengue
 - West Nile fever
 - Japanese enceph.
 - Yellow fever
 - hepatitis C
- no signs of remission (~1.5M clinical cases/yr)



utmb Health Biochemistry & Molecular Biology

Are there unique druggable targets?

- flavivirus genomes have similar organization & sequences
- flavivirus proteins have similar structures
- NS3 protease required for virus replication & highly conserved

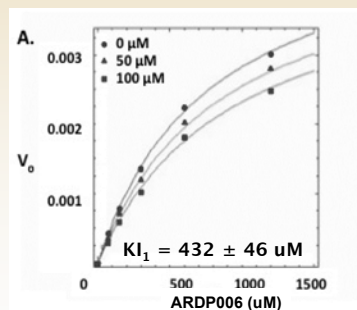
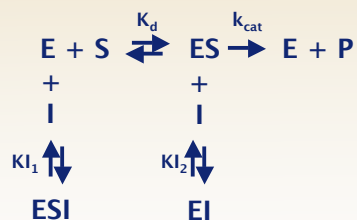


utmb Health Biochemistry & Molecular Biology

Initial *in silico* discovery

- evaluated with *in vitro* protease inhibition assays
- initial screen yielded 3 protease inhibitors from ~30 hits (Tomlinson et al. 2009)
- leads inhibited both dengue and WNV proteases
- leads were uncharged, drug-like molecules

General model for inhibition



utmb Health Biochemistry & Molecular Biology

Leads showed activity in cell culture

- phased assays used with multiple DENV, WNV strains & cell lines

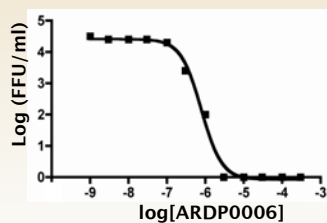
1) rapid ELISA

- EC₅₀ and CC₅₀
- 48 hr infection
- polyclonal Ab for readout



2) foci-forming assay (for DENV)

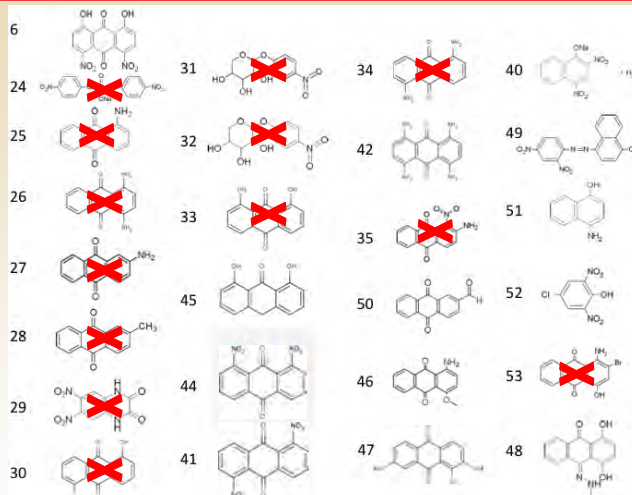
- EC₅₀ and CC₅₀
- 96 hr infection
- foci require 5 days to develop



Lead expansion from commercial sources

ARDP006 “analogues”

- several dozen ordered
- many insoluble



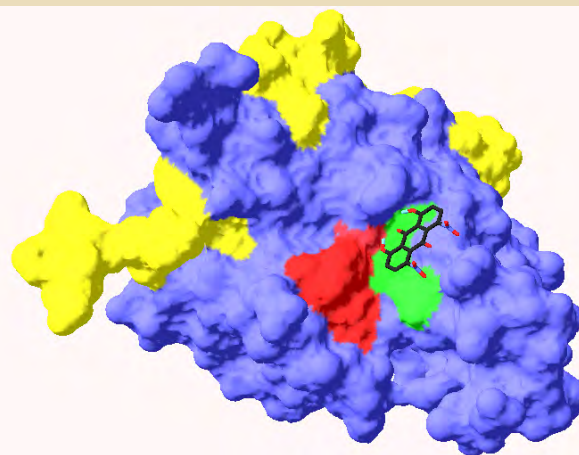
Improved activity with analogs

ARDP006 “analogs”

- DENV
~30-fold improvement
- WNV
~490-fold improvement

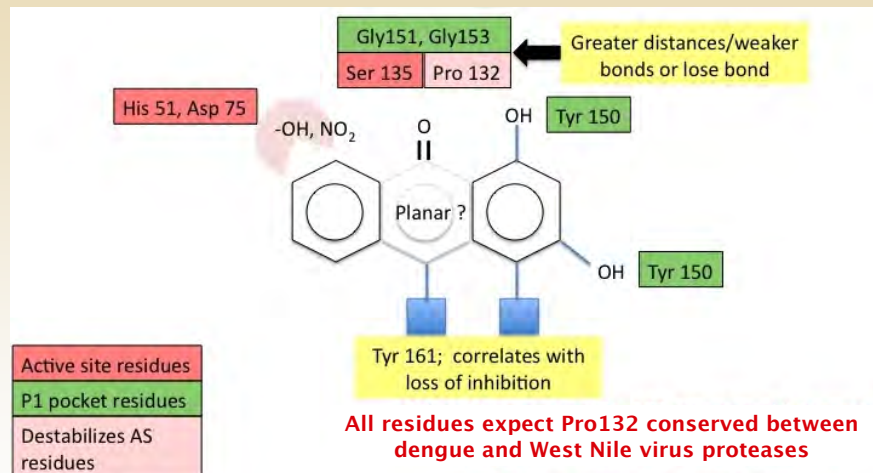
Compound	DEN2V Ki1 (uM)	WNV Ki1 (uM)
ARDP006	432 ± 46	977 ± 241
6A42	158 ± 32	-
6A45	47 ± 15	2 ± .02
6A47	215 ± 119	-
6A49	15 ± 3	34 ± 19

Binding to develop structure-activity (SAR)



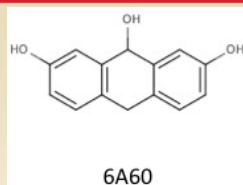
ARDP0006 docked into DENV protease

Structure-based SAR of anthraquinones



utmb Health Biochemistry & Molecular Biology

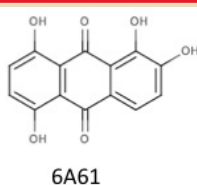
SAR-directed anthraquinone-based leads



Predicted to be a better binder

DEN2V K_{i1} = 7 ± 5

WNV K_{i1} = 11 ± 11

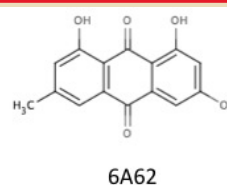


Predicted to have intermediate binding

Validated SAR

DEN2V K_{i1} = 72 ± 15

WNV K_{i1} = 31 ± 36



Predicted to be a poor binder

DEN2V K_{i1} = 508 ± 47

WNV K_{i1} = 1035 ± 122



utmb Health Biochemistry & Molecular Biology

Development of best leads

- best leads have low μM protease inhibition of DENV and WNV
- do not inhibit trypsin
- ~ 100 nM activity in cell culture against DENV, WNV
- therapeutic index ($\text{CC}_{50}/\text{EC}_{50}$) > 150

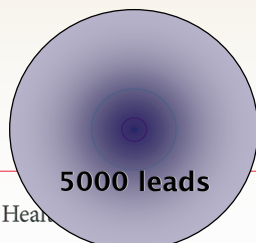
- synthetic chemical program initiated by Dr. S. Gilbertson (Univ Houston)
- *in vivo* (AG129 mouse model) studies initiated by Utah State Univ.

What next?

1 drug •

- drug development is highly inefficient
- more leads help increase odds of producing a drug

10 IND filings ●

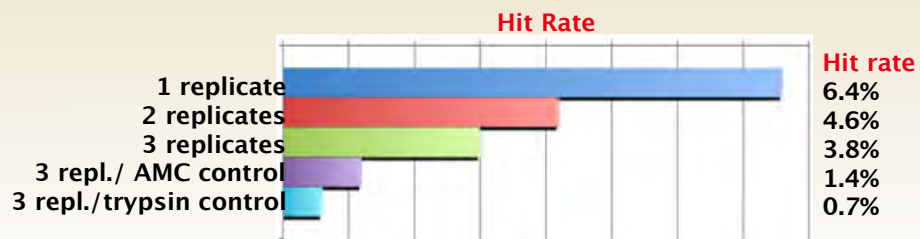


Complete WCG virtual screens, and

Use complementary (e.g., HTS) methods

How best to perform HTS?

- replicates, replicates, replicates + internal controls



Use HTS for additional leads

How best to perform HTS?

- replicates, replicates, replicates + internal controls
- **inexpensive to include replicates and controls**
 - ~\$0.053/compound (UT-HSC) + researcher
- vs
- ~\$50/compound (+researcher) to validate

In vitro HTS for additional leads

- **Microsource US drug collection**
 - 1040 compounds which have reached clinical trial stages in the USA
- **Microsource natural products collection**
 - 800 pure natural products
 - very chemically diverse
- **MayBridge HitFinder library**
 - 14,400 “drug-like” library
- **include replicates, integrated trypsin counter-screen**

HTS for additional leads

MicroSource library

Name	Known Activity	DENV Ki1 (uM)	DENV Ki2 (uM)	WNV Ki1 (uM)	WNV Ki2(uM)
M1	Transacetylase inhibitor				
M2	Anti-parasitic				
M3	Anti-infective				
M4	Topical antibacterial				
M5	Anti-parasitic				
M6	antibacterial				

MayBridge library

- 38 selective (DENV activity, no trypsin activity) molecules under study

Acknowledgements

Watowich lab

Robert Malmstrom
Suzanne Tomlinson
Muhammad Arian
Jeff Borgenson
Fan-Ping Kong
Marta Lorinczi
Andrew Russo

UTMB

Alan Barrett
Robert Tesh
Doug Watts

§ NIH, IIF, Welch, GCC

Collaborators

Jay Boisseau (TACC)
Scott Gilbertson (U of H)
Michael Gonzoles (TACC)
Justin Julander (USU)
Stephen Monk (TACC)
Benoit Roux & lab (Univ. Chicago)
Cliff Stephan (UT-HSC)

IBM

WCG team
Viktors Berstis
Bill Bovermann
Sandy Dochen
Keith Uplinger
Robin Willner