All-atom Molecular Dynamics Simulations of EGFR with Prediction of Inhibitors Fold Resistance

Trent E. Balius

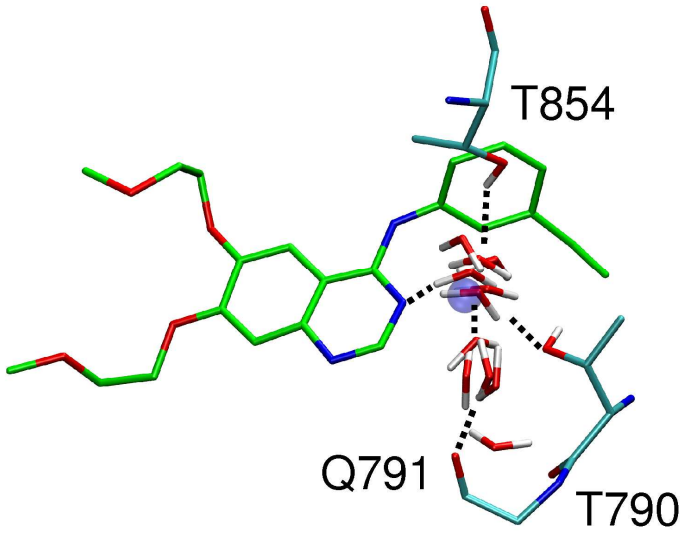
Advisor: Robert C. Rizzo

AMS 535

11-18-2009

Overview

- Background on EGFR
  - Motivations
  - Literature binding values
- Molecular Dynamics Simulations
- Post-processing Methods
  - Background on MM-GBSA
  - Molecular Footprint
- Results
- Conclusions



Balius, T. E.; Rizzo, R. C., Quantitative prediction of fold resistance for inhibitors of EGFR. *Biochemistry* **2009,** 48, (35), 8435-48.

EGFR Background

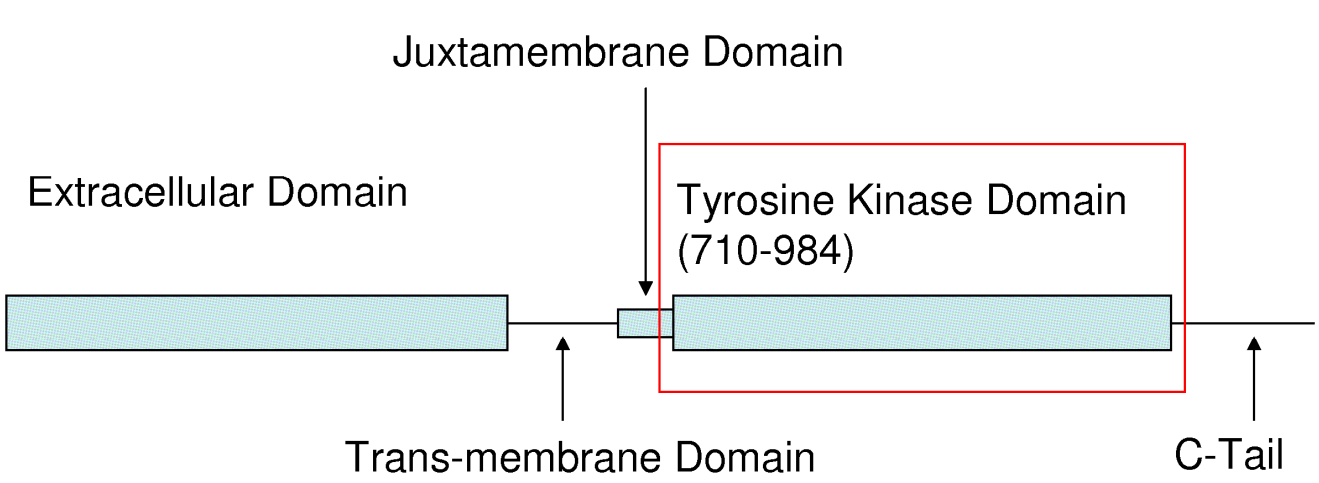
Cancer Background

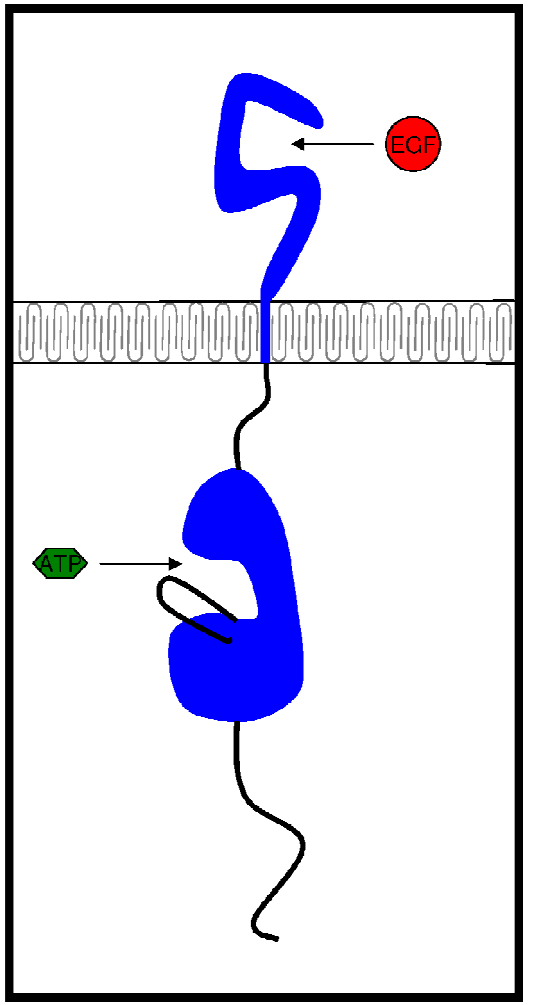
- leading cause of death in US under the age of 85
- 2nd highest cause of death in US
- lung & bronchial cancer is leading cause of death of cancer
- non-small cell lung cancer (NSCLC) largest subset of lung cancer
- EGFR is a target for NSCLC
- member of the ErbB family: EGFR, ErbB2, ErbB3, & ErbB4

Jemal, A., et al. (2008) Cancer statistics, 2008, CA Cancer J Clin 58, 71-96.

Hynes, N. E., and Lane, H. A. (2005) ERBB receptors and cancer: the complexity of targeted inhibitors, Nat Rev Cancer 5, 341-354.

EGFR: A Chemotherapeutic Target

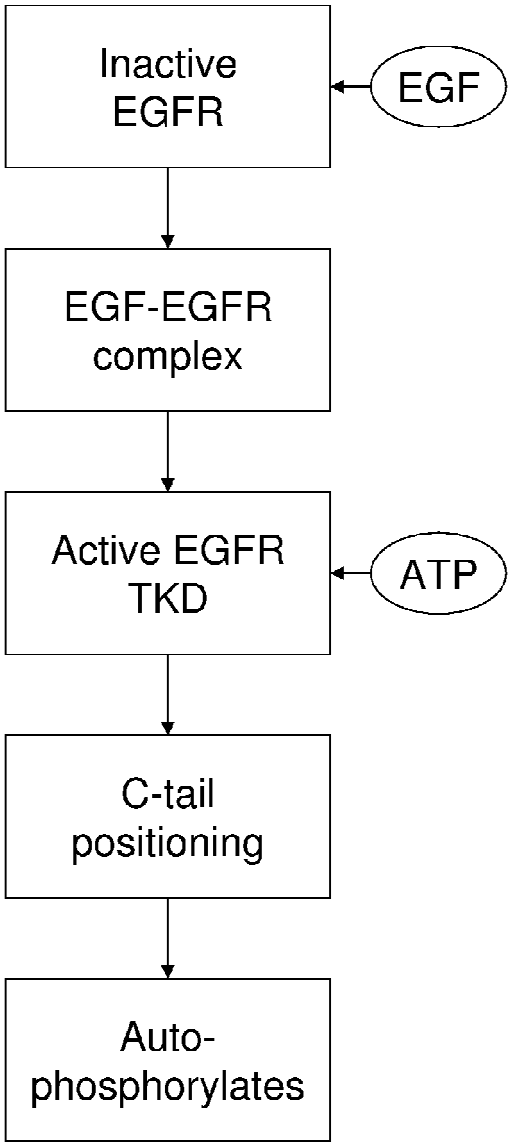


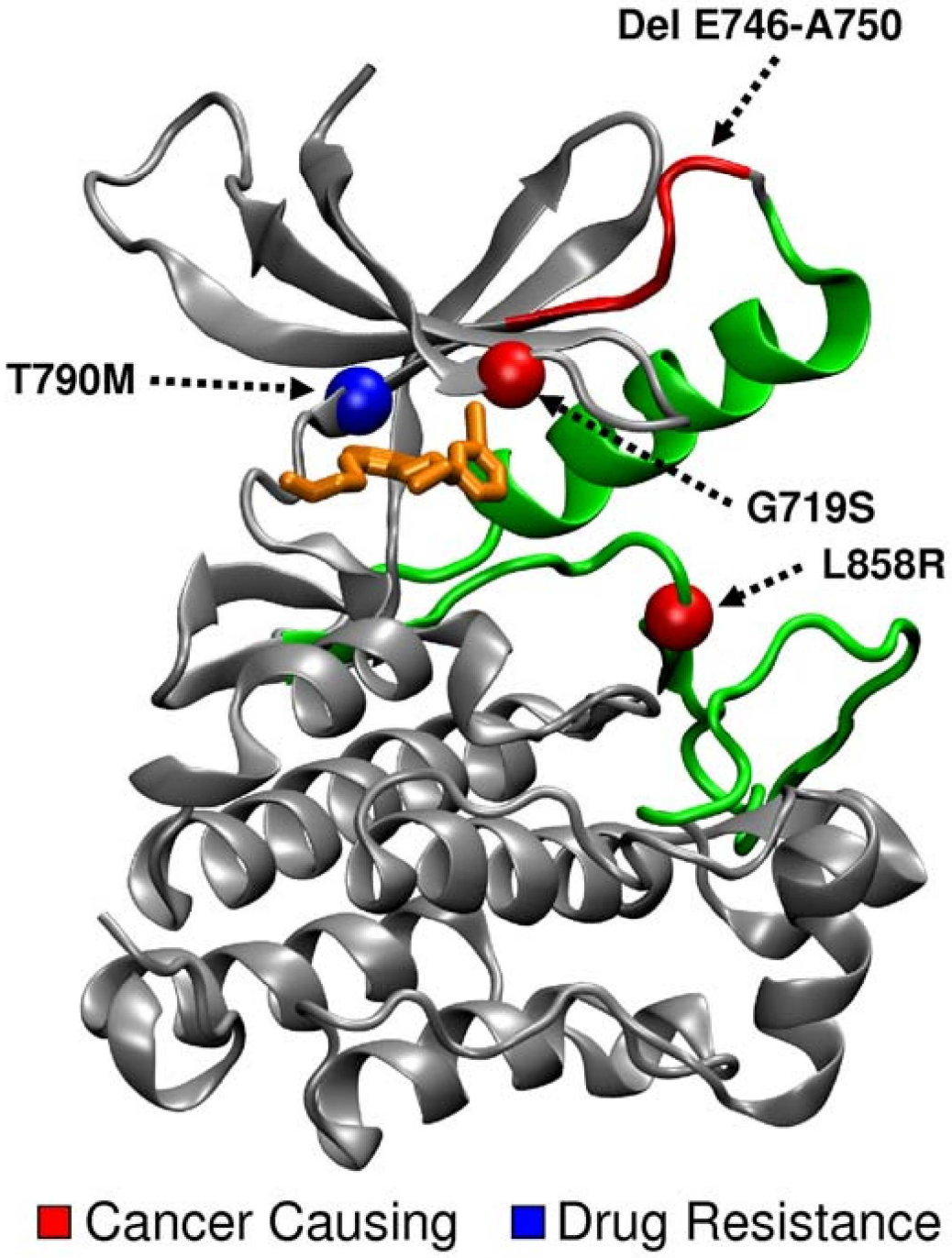


- two classes of anticancer agents
  - monoclonal antibodies (extracellular)
  - small molecule inhibitors (intracellular)

EGFR Pathway

- ligand-free EGFR is a monomer
- EGFR is inactive as a monomer
- EGF binds, the EGFR homo- or hetero dimerizes
- EGFR is active as a dimer
- ATP binds Tyrosine Kinase Domain (TKD)
  - EGFR auto-phosphorylates C-tail
- leads to signal transduction
  - pathways involved in cellular proliferation





Key EGFR Mutations

- Cancer Causing
  - L858R

    increase binding to erlotinib & gefitinib
  - Del E746-A750

    increase binding to erlotinib & gefitinib
  - G719S

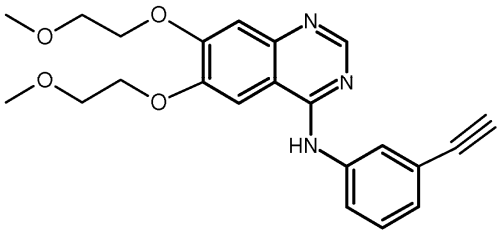
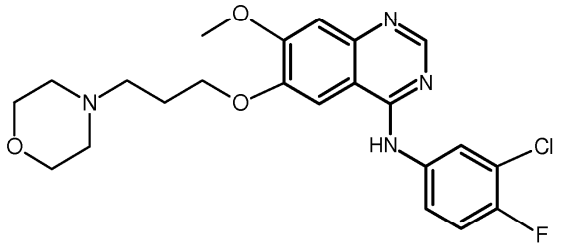
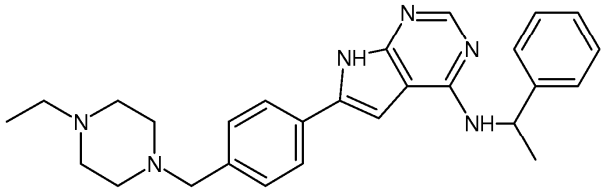
    decrease binding to gefitinib

- Drug Resistance
  - T790M

    decrease binding to all

Ligands

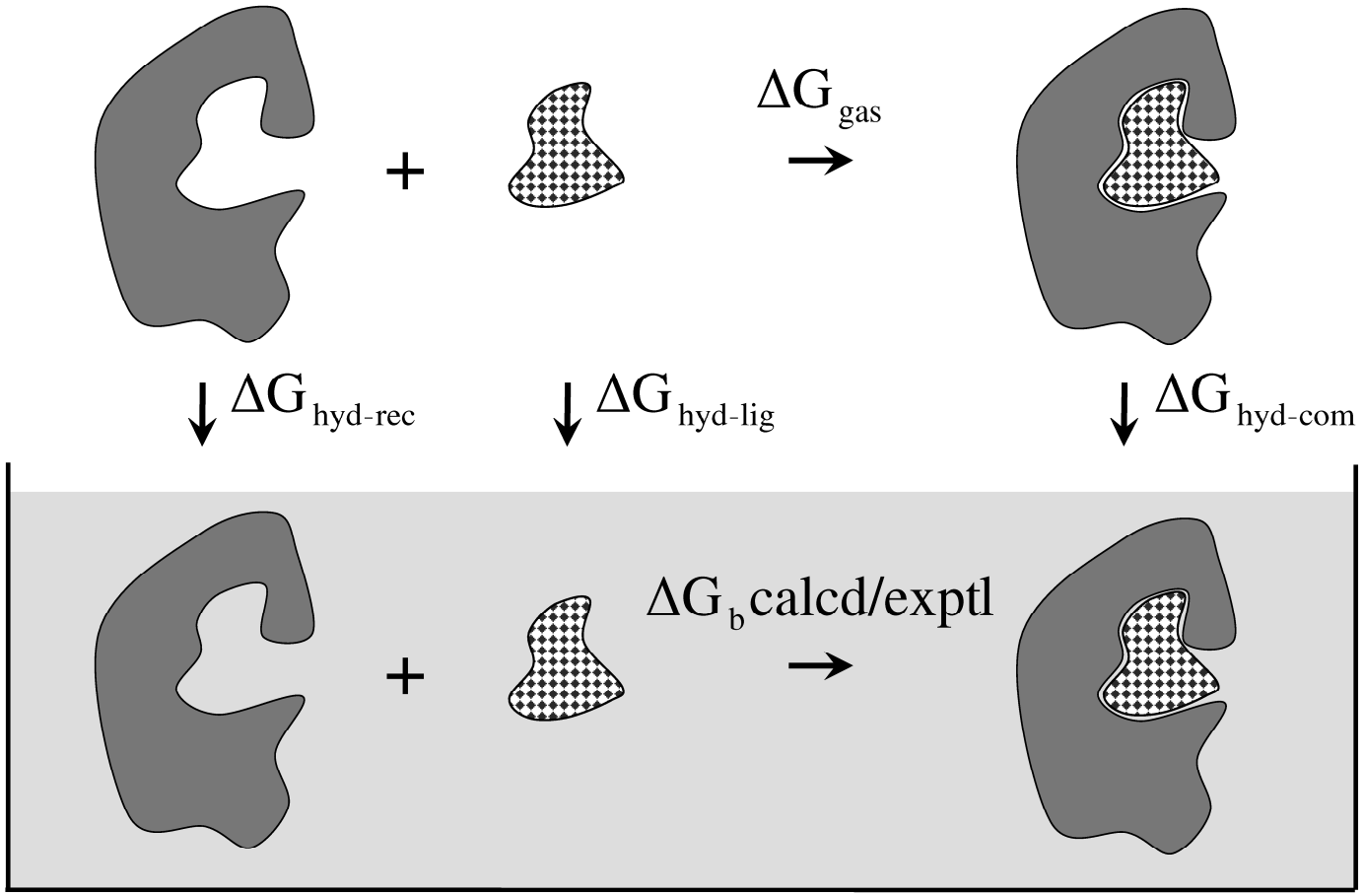
Table 1. Experimental Fold Resistance (FR) values for ATP-competitive inhibitors with EGFR.

Inhibitor	Structure	Experimental Fold Resistance[a]		
		L858R / WT	**L858R&T790M / L858R**	**G719S / WT**
erlotinib		6.25 / 17.5 nM[b] 0.36 FR −0.61 $\Delta\Delta G_{FR}$	>10000 / 12.5 nM[c] >800 FR >3.96 $\Delta\Delta G_{FR}$	–
gefitinib		2.4 / 35.3 nM[d] 0.068 FR −1.59 $\Delta\Delta G_{FR}$	10.9 / 2.4 nM[d] 4.54 FR 0.90 $\Delta\Delta G_{FR}$	123.6 / 53.5 nM[e] 2.31 FR 0.50 $\Delta\Delta G_{FR}$
AEE788		1.1 / 5.3 nM[d] 0.21 FR −0.92 $\Delta\Delta G_{FR}$	18.6 / 1.1 nM[d] 16.9 FR 1.68 $\Delta\Delta G_{FR}$	11.3 / 10.9 nM[e] 1.04 FR 0.02 $\Delta\Delta G_{FR}$

[a]Fold Resistance (FR) = ratio of experimental activities. $\Delta\Delta G_{FR}$ exptl ≈ RTln(FR) at 298.15 K in kcal/mol. [b]Ki values (nM) from Carey, K. D., et al. , Cancer Res 66, 8163-8171. (2006). [c] IC_{50} values (nM) from Ji, H., et al., Proc Natl Acad Sci U S A 103, 7817-7822. (2006). [d] Kd values (nM) from Yun, C. H., et al., Proc Natl Acad Sci U S A 105, 2070-2075. (2008). [e]Kd values (nM) from Yun, C. H., et al., Cancer Cell 11, 217-227. (2007).

Using Molecular Dynamics generated ensembles

Thermodynamic Cycle

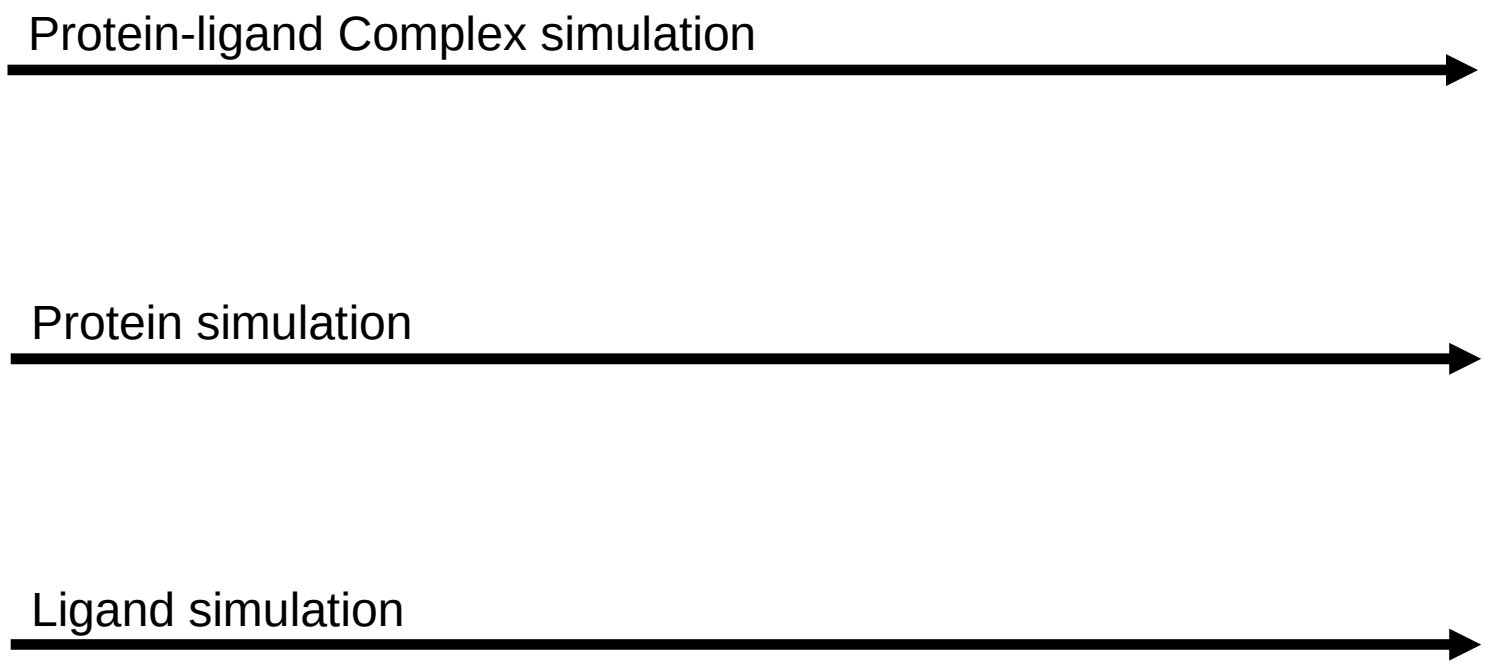


$$\Delta G_b \text{ exptl} \approx \Delta G_b \text{ calcd} = \Delta G_{gas} + \Delta G_{hyd\text{-}com} - (\Delta G_{hyd\text{-}rec} + \Delta G_{hyd\text{-}lig})$$

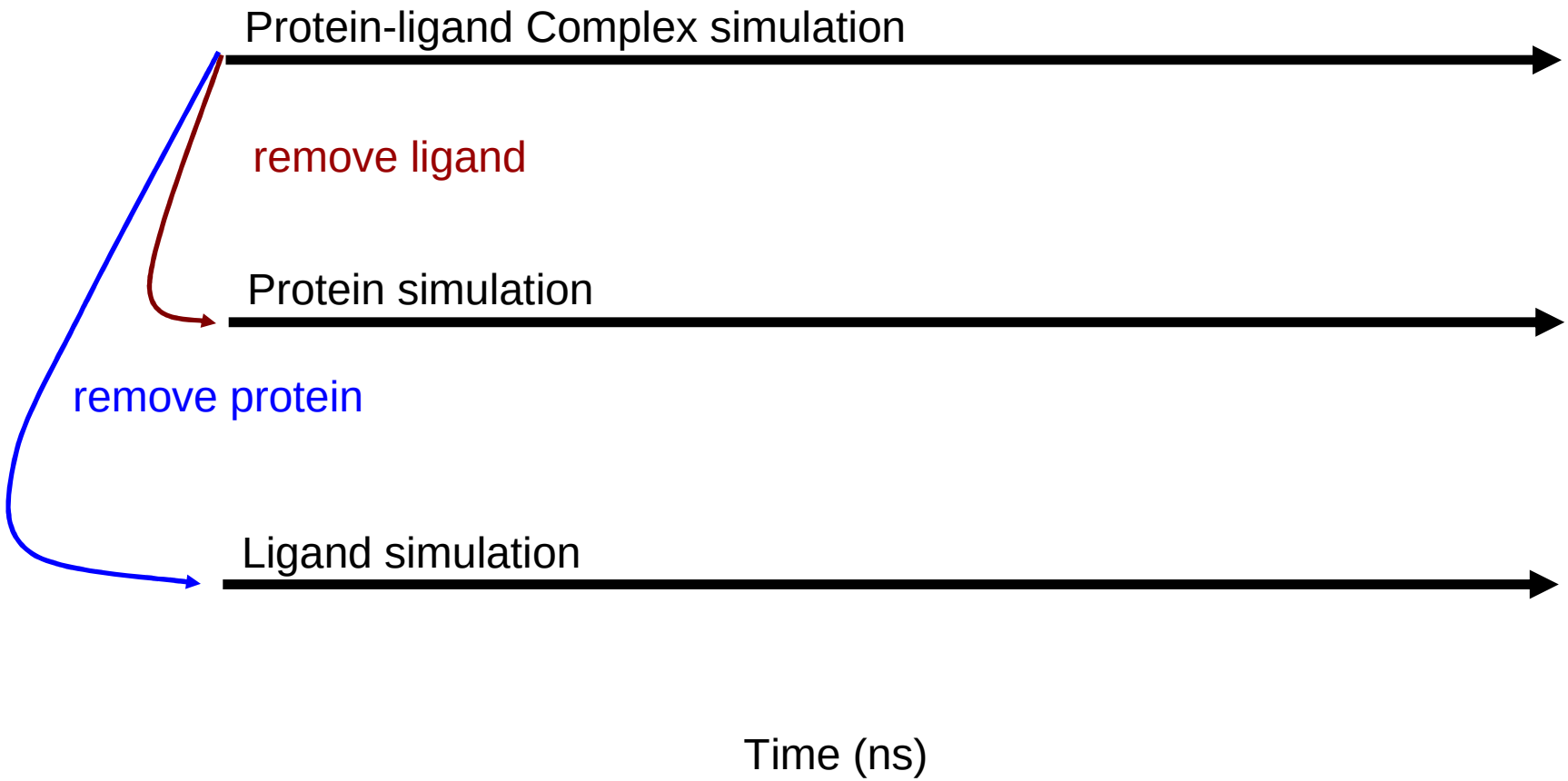
$$\Delta G_{gas} = \Delta E_{vdw} + \Delta E_{coul}$$

$$\Delta G_{hyd\text{-}species} = \Delta G_{polar} + \Delta G_{nonpolar}$$

Run 3 independent Simulations



Run 1 Simulations



Simulation Methods

- force field: FF99SB (protein); GAFF (ligand) augmented with CHELPG charges (6-31G* basis set);TIP3P (water)

- 9 step equilibration (minimization + MD) with decreasing restraints (amber 8 package)

- final production run for 5 ns @ 298.15 K
  - constant T & P with periodic boundary conditions

- post processing
  - root mean-squared deviations (RMSD)
  - ΔG_{bind} estimation via MM-GBSA method
  - energy component decomposition
  - per-residue H-bond and energetic footprints

MM-GBSA Background

MM-GBSA Equations

$$G = E_{MM} + \Delta G_{hyd} - TS$$

such that

$$E_{MM} = E_{bond} + E_{angle} + E_{tors} + E_{vdw} + E_{es}$$

TS is calculated using quasi harmonic analysis normal mode analysis

$$\Delta G_{hyd} = \Delta G_{polar} + \Delta G_{nonpolar}$$

where the polar and the nonpolar terms are defined in the following way

ΔG_{polar} - is defined by solving the PB set of differential equations

or by using the GB equation.

$$\Delta G_{nonpolar} = \alpha \cdot SA + \beta$$

$$\Delta G = G_{complex} - G_{protein} - G_{ligand}$$

MM-GBSA Equations

$$G = E_{MM} + \Delta G_{hyd} - TS$$

such that

$$E_{MM} = E_{bond} + E_{angle} + E_{tors} + E_{vdw} + E_{es}$$

TS is calculated using quasi harmonic analysis normal mode analysis

$$\Delta G_{hyd} = \Delta G_{polar} + \Delta G_{nonpolar}$$

where the polar and the nonpolar terms are defined in the following way

ΔG_{polar} - is defined by solving the PB set of differential equations

or by using the GB equation.

$$\Delta G_{nonpolar} = \alpha \cdot SA + \beta$$

$$\Delta G = G_{complex} - G_{protein} - G_{ligand}$$

Generalized Born Equations

$$G_{GB} = -166\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{q_i q_j}{f_{GB}\left(r_{i,j},\alpha_{i,j}\right)} \qquad \alpha_{i,j} = \sqrt{\alpha_i \alpha_j}$$

$$f_{GB}(r,\alpha) = \sqrt{r^2 + \alpha^2 \exp\left(-D(r,\alpha)\right)} \qquad D(r,\alpha) = \frac{r^2}{4\alpha^2}$$

- The trick to GB is calculating the Born Radii
- Born Radii are dependent on conformation

Still, W. C.; et al., J. Am. Chem. Soc 1990, 112, 6127-6129

Hawkins Born Radii Model

$$\hat{\alpha}_i = \frac{1}{\frac{1}{\rho_i} - \frac{1}{2}\sum_{j=1}^{n}\left[\left(\frac{1}{L_{i,j}} - \frac{1}{U_{i,j}}\right) + \frac{r_{i,j}}{4}\left(\frac{1}{L_{i,j}^{\,2}} - \frac{1}{U_{i,j}^{\,2}}\right) + \cdots \left(\frac{1}{2r_{i,j}}\ln\frac{L_{i,j}}{U_{i,j}}\right) + \left(\frac{1}{L_{i,j}} - \frac{1}{U_{i,j}}\right)\right]}$$

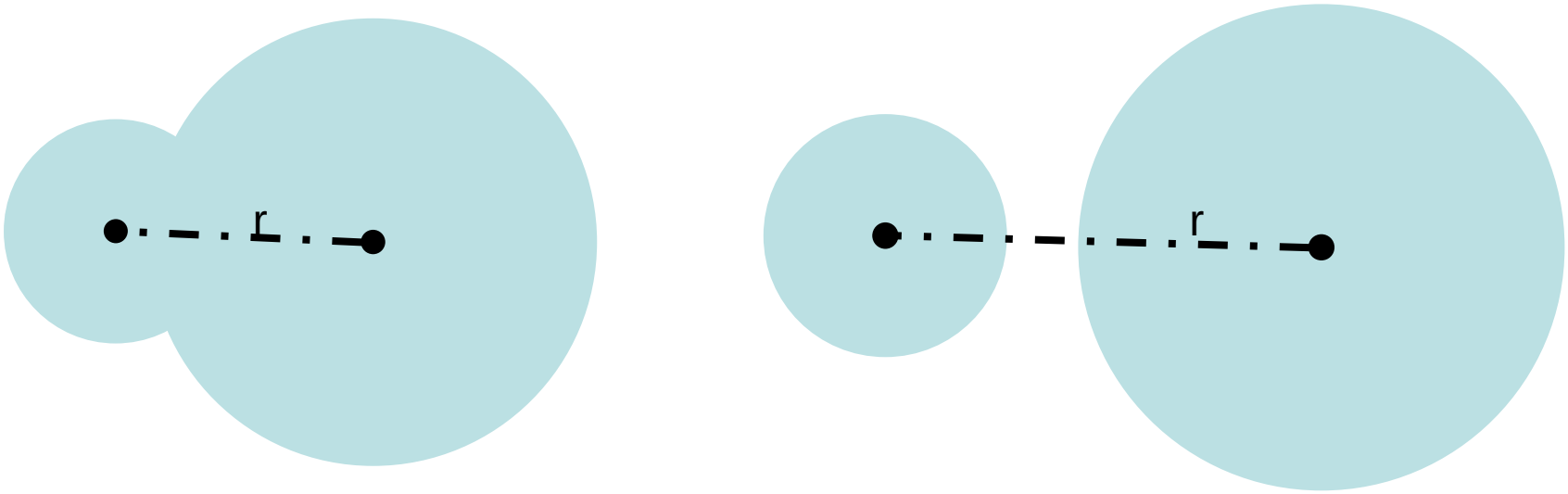
$$\alpha_i = \max\left(0, \hat{\alpha}_i\right)$$

$$L_{i,j} = \begin{cases} 1 & \text{if } r_{i,j} + s_j\rho_j \le \rho_i \\ \rho_i & \text{if } r_{i,j} - s_j\rho_j \le \rho_i < r_{i,j} + s_j\rho_j \\ r_{i,j}\text{-}s_j\rho_j & \text{if } \rho_i \le r_{i,j}\text{-}s_j\rho_j \end{cases}$$

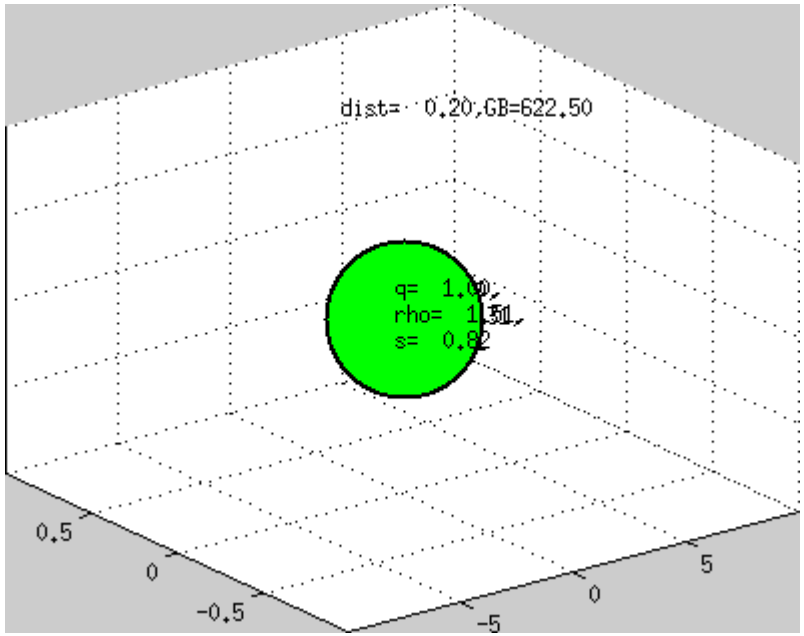
$$U_{i,j} = \begin{cases} 1 & \text{if } r_{i,j} + s_j\rho_j \le \rho_i \\ r_{i,j}\text{-}s_j\rho_j & \text{if } \rho_i < r_{i,j} + s_j\rho_j \end{cases}$$

Hawkins, G. D. et al. J. Phys. Chem., Vol. 100, No. 51, 1996 19824-19839

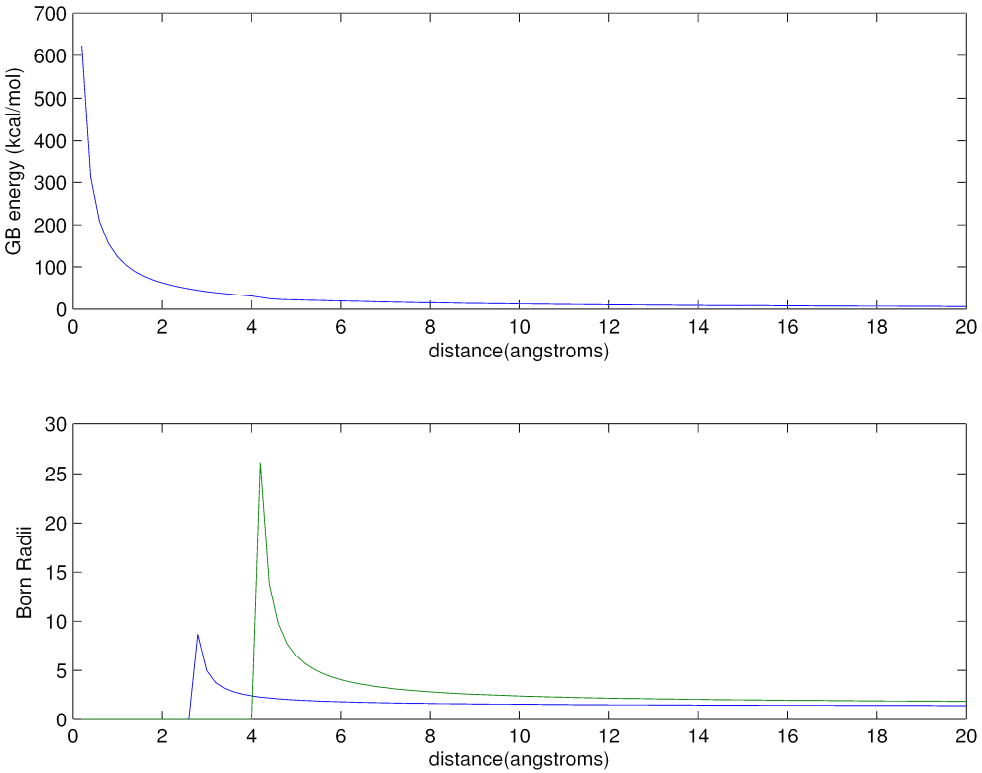
Generalized Born as a Function of atom position



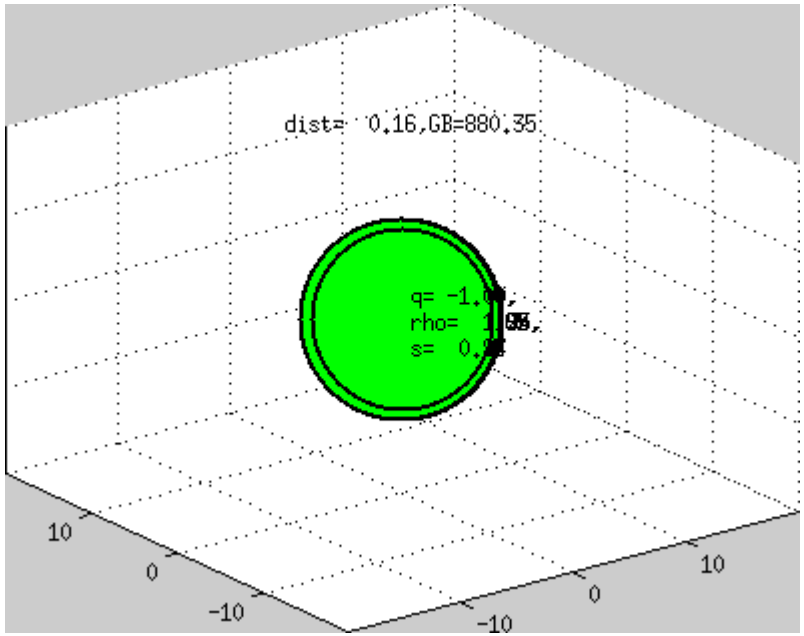
Generalized Born as a Function of atom position



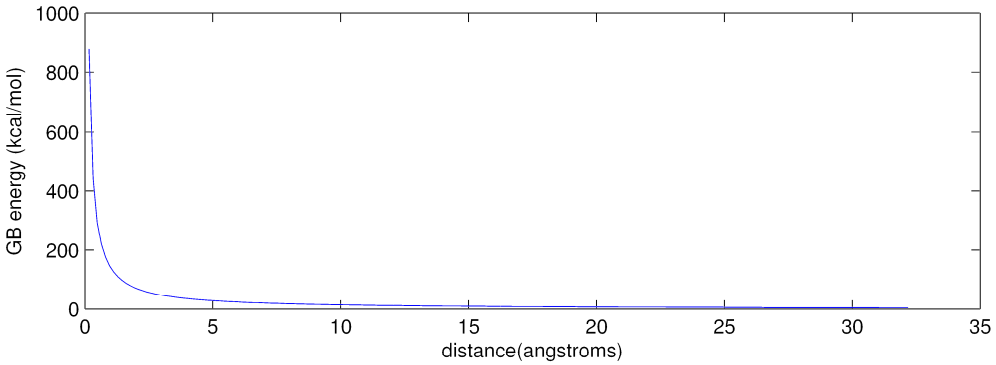
2 atoms movie

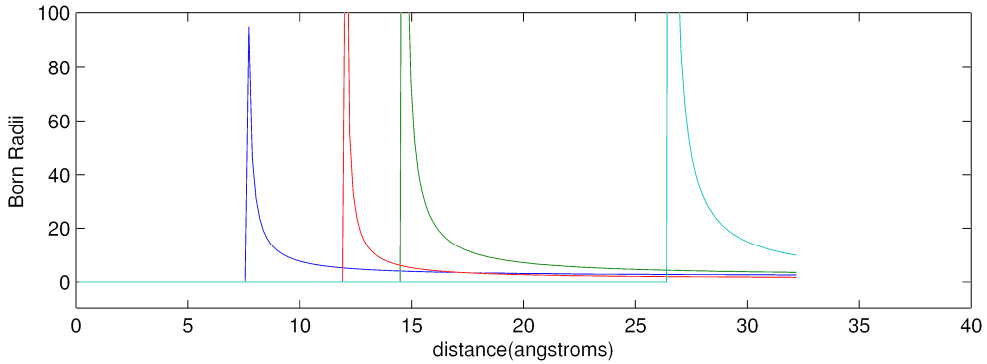


Generalized Born as a Function of atom position



4 atoms movie



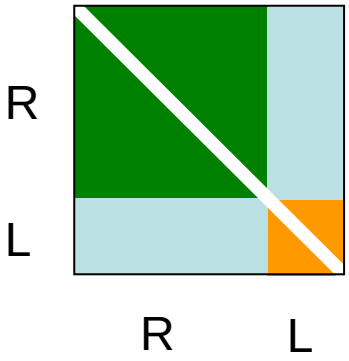


Molecular Footprints -- Per-residue Decomposition of Interactions

Footprint Introduction

$$E_{\text{vdw}}\text{comp} = \frac{1}{2}\sum_{a\in\text{comp}}\sum_{\substack{b\in\text{comp}\\ \text{s.t. } b\neq a}} E_{\text{vdw}}(a,b)$$

pair wise interaction



$$= \frac{1}{2}\sum_{a\in\text{rec}}\sum_{b\in\text{rec}} E_{\text{vdw}}(a,b) + \frac{1}{2}\sum_{a\in\text{lig}}\sum_{b\in\text{lig}} E_{\text{vdw}}(a,b) + \sum_{a\in\text{lig}}\sum_{b\in\text{rec}} E_{\text{vdw}}(a,b)$$

$$= E_{\text{vdw}}\text{rec} + E_{\text{vdw}}\text{lig} + \sum_{a\in\text{lig}}\sum_{b\in\text{rec}} E_{\text{vdw}}(a,b)$$

intermolecular component

$$\Delta E_{\text{vdw}} = E_{\text{vdw}}\text{comp} - (E_{\text{vdw}}\text{rec} + E_{\text{vdw}}\text{lig})$$

vdw component of binding energy

$$= \sum_{a\in\text{lig}}\sum_{b\in\text{rec}} E_{\text{vdw}}(a,b)$$

This same analysis can be done for other through space interactions

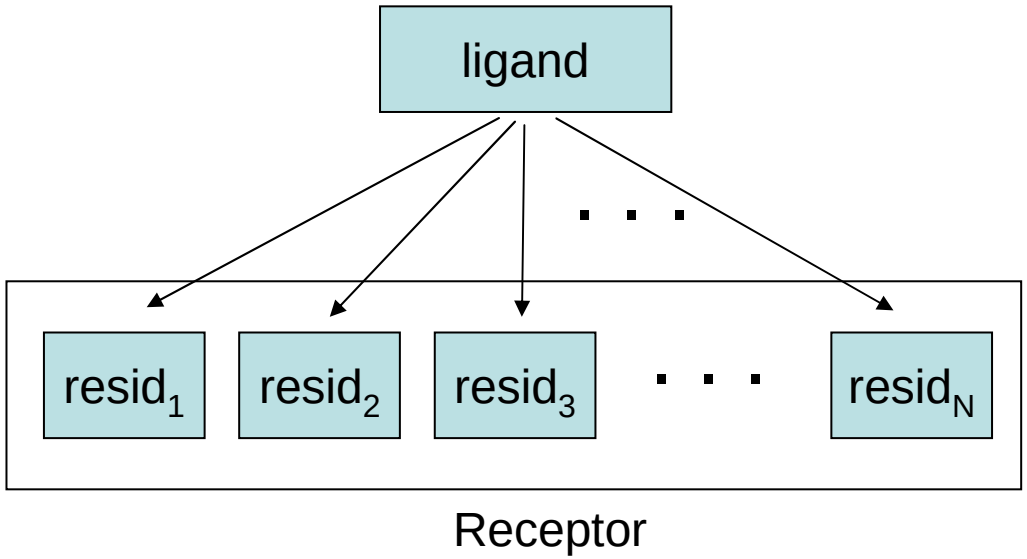
Coulombic, GB, SASA, . . .

Footprint Introduction

$$\Delta E_{\text{vdw}} = \sum_{a \in \text{lig}} \sum_{b \in \text{rec}} E_{\text{vdw}}(a,b)$$

$$= \sum_{a \in \text{lig}} \left[\sum_{i=1}^{N} \left(\sum_{b \in \text{resid}(i)} E_{\text{vdw}}(a,b) \right) \right]$$

$$= \sum_{i=1}^{N} \left[\sum_{a \in \text{lig}} \left(\sum_{b \in \text{resid}(i)} E_{\text{vdw}}(a,b) \right) \right]$$



$$\Delta E_{vdw} = \text{sum}\left(\vec{E}_{\text{vdw,fp}}\right)$$

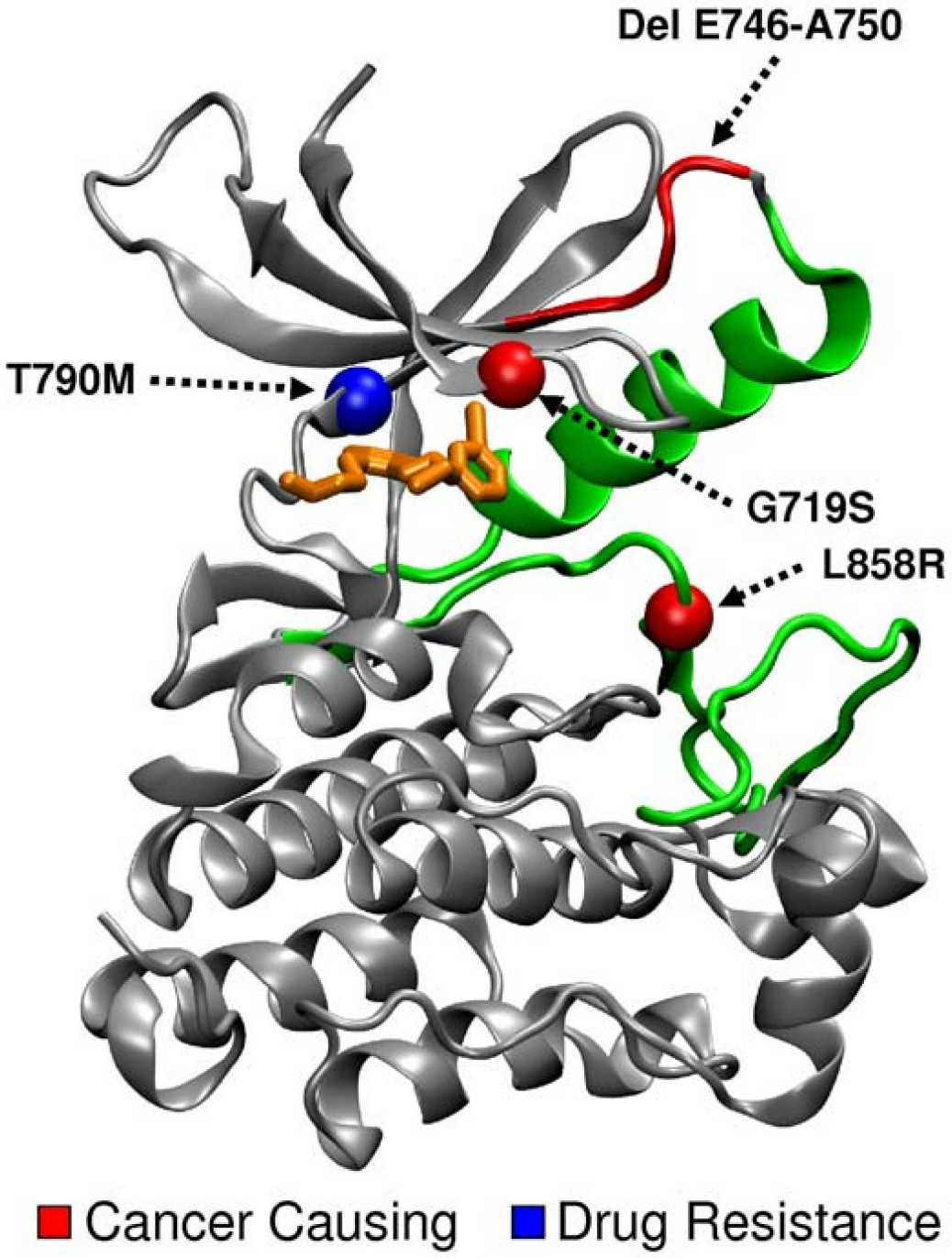
$$\vec{E}_{\text{vdw, fp}} = \left[\sum_{a \in \text{lig}} \left(\sum_{b \in \text{resid}(i)} E_{vdw}(a,b) \right) \right]$$

footprint vector each element corresponds to a residue

This same analysis can be done for other through space interactions

Coulombic, GB, SASA, . . .

Results



Key EGFR Mutations

- Cancer Causing
  - L858R

    increase binding to erlotinib & gefitinib
  - Del E746-A750

    increase binding to erlotinib & gefitinib
  - G719S

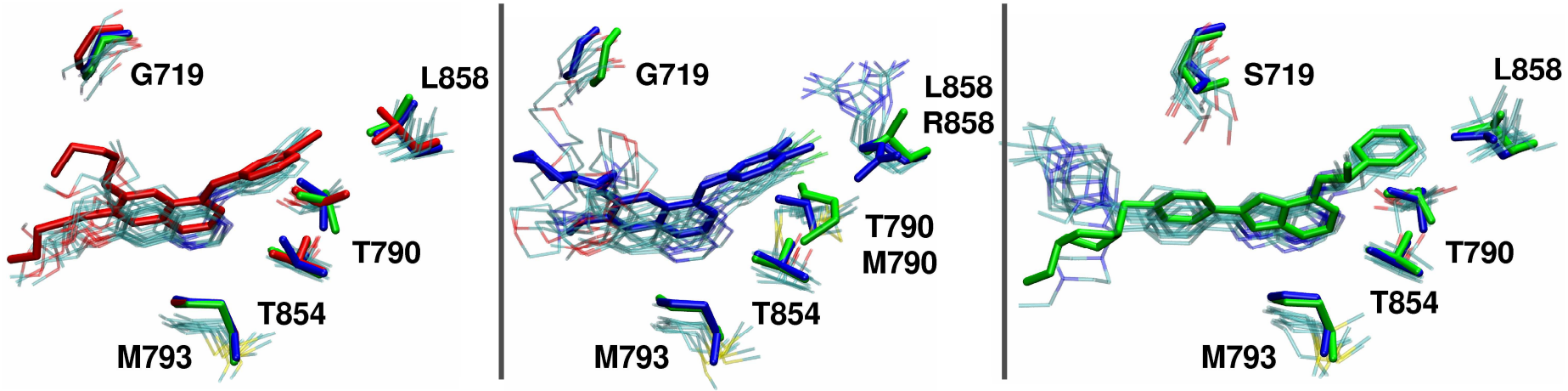
    decrease binding to gefitinib

- Drug Resistance
  - T790M

    decrease binding to all

Simulations vs. Crystallographic Structures

(a) erlotinib **(b) gefitinib** **(c) AEE788**



red=1M17 (erlotinib, WT)

blue=2ITY (gefitinib,WT)

green=2J6M (AEE788, WT)

blue=2ITZ (gefitinib,L858R)

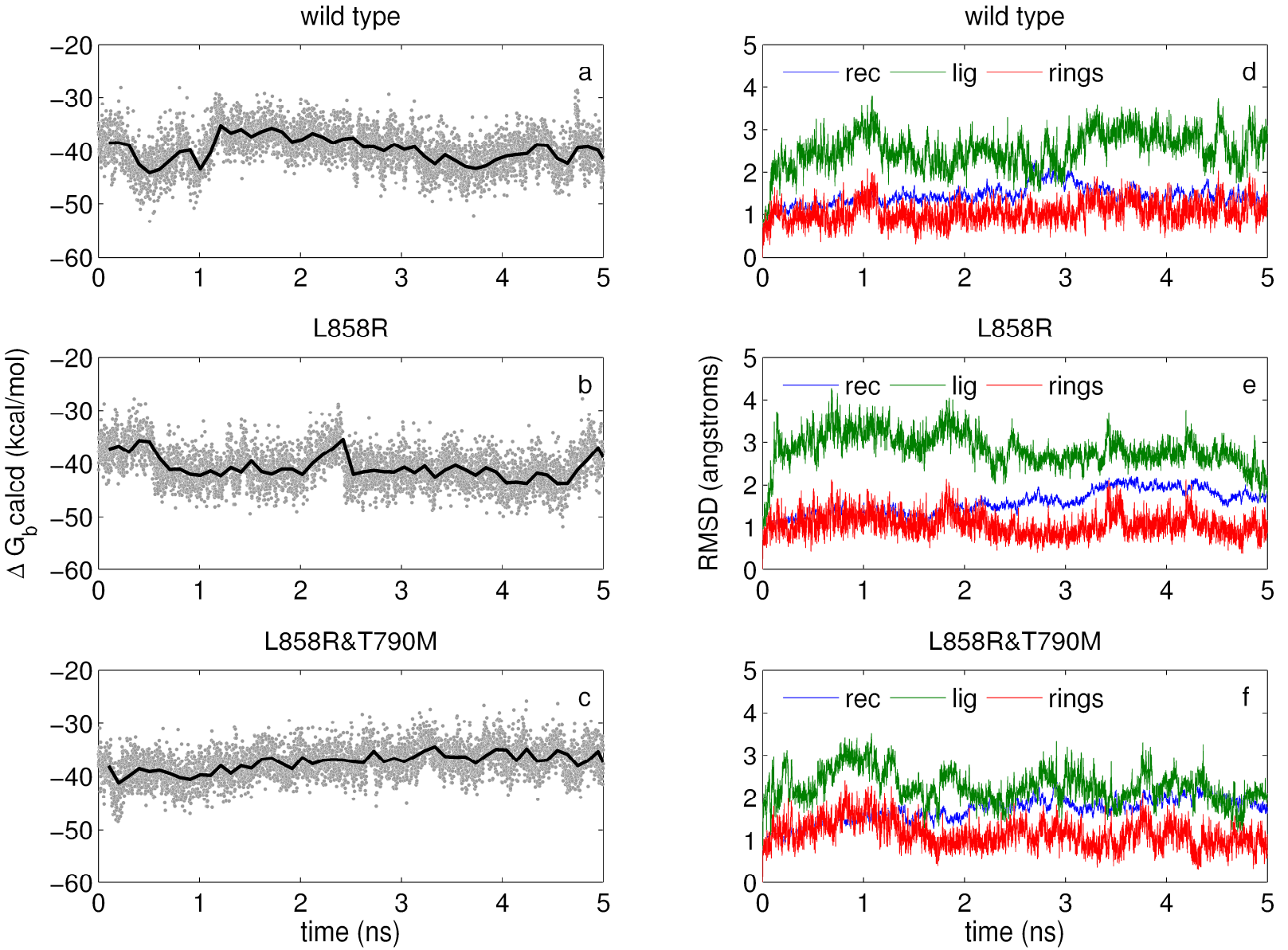
green=2JIU (AEE788,T790M)

blue=2ITO (gefitinib, G719S)

green=2ITP (AEE788, G719S)

Overlaid MD snapshots (thin lines N=10) with available crystal structure complexes of EGFR (bold lines)

Simulation and System Stability: Erlotinib



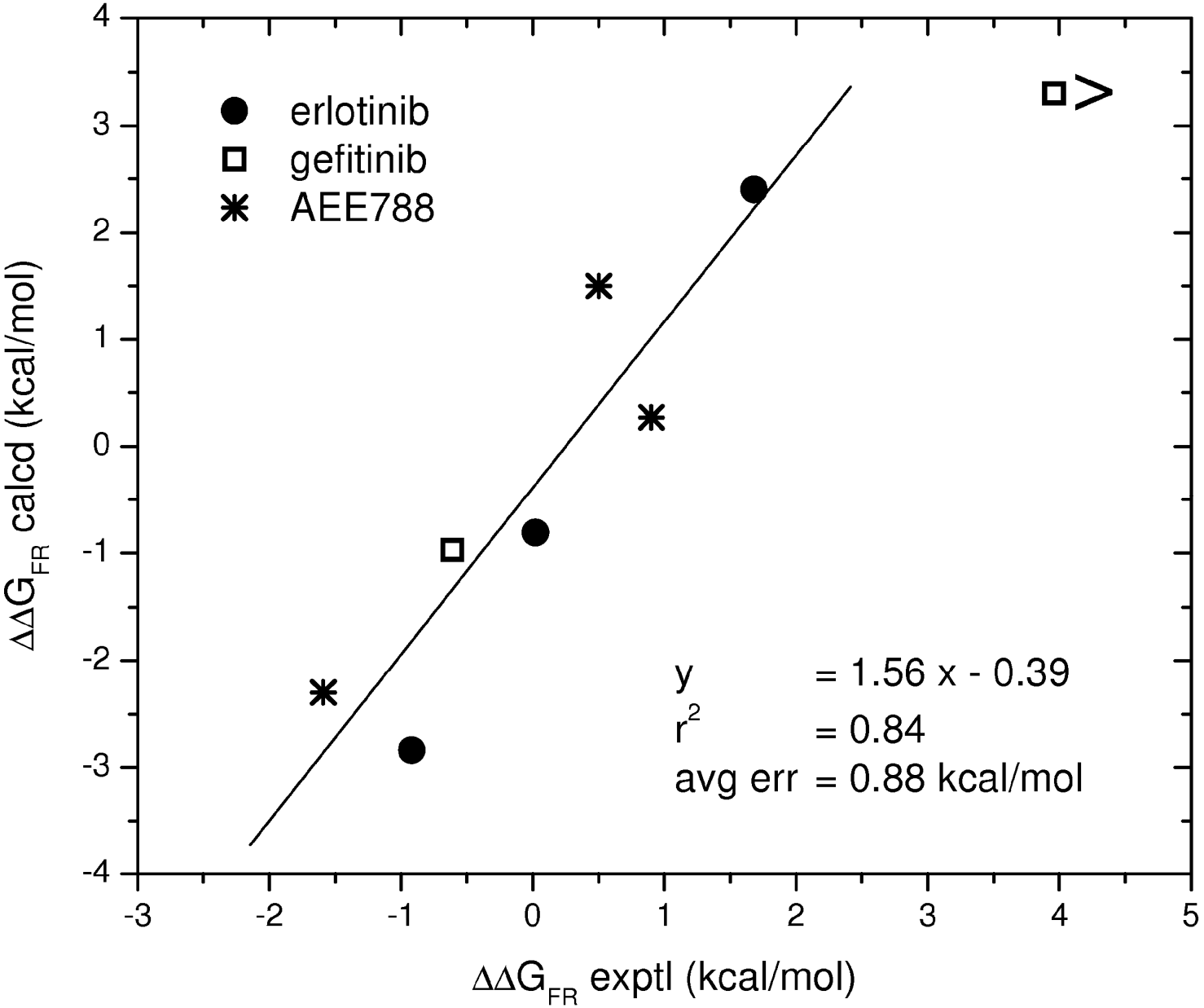
Relative Free Energies and Components

Table 2. Experimental versus calculated Fold Resistance (FR) energies ($\Delta\Delta G_{FR}$) and energy components for ligands with EGFR.

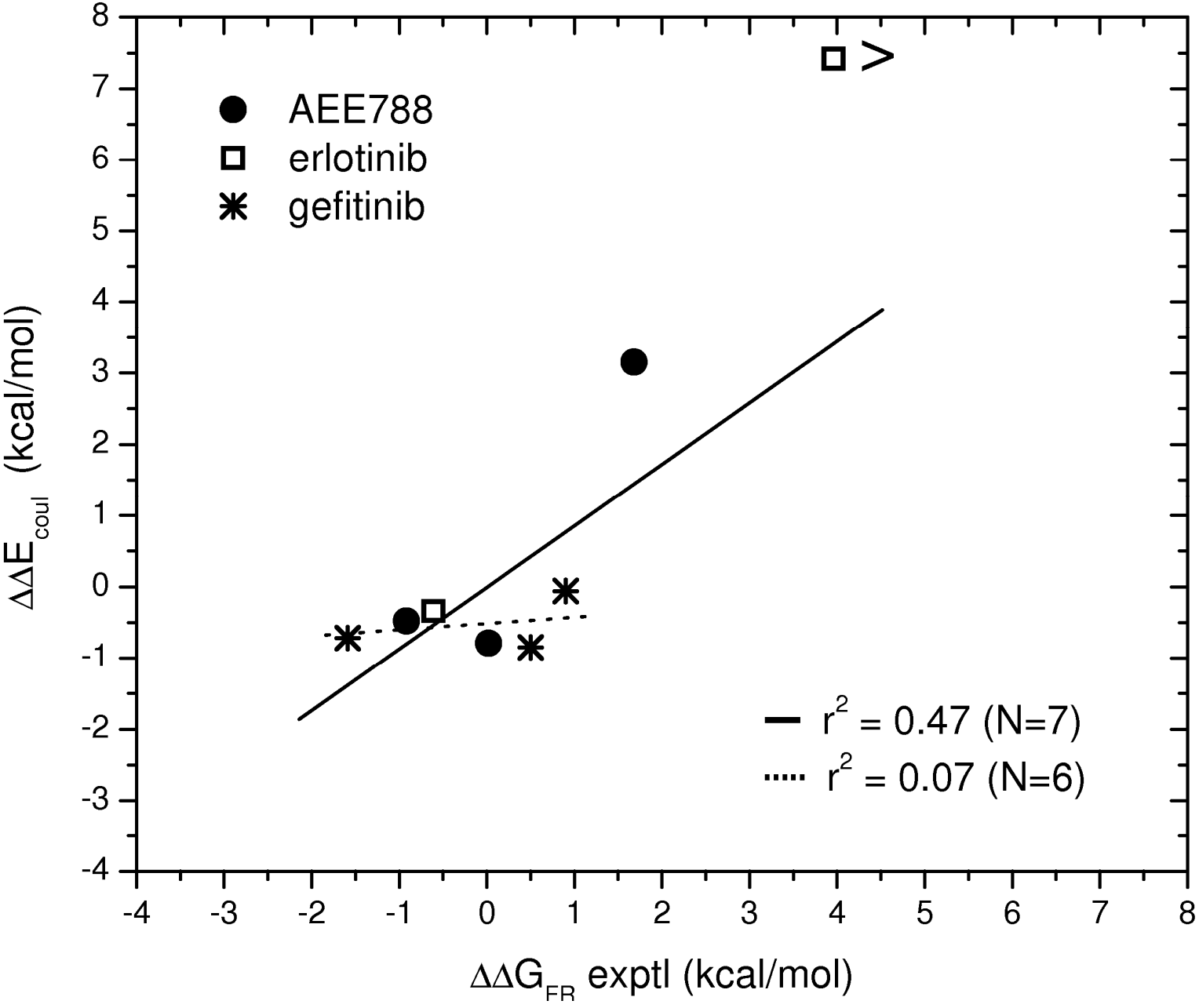
inhibitor	$\Delta\Delta E_{vdw}$ A	$\Delta\Delta E_{coul}$ B	$\Delta\Delta G_{polar}$ C	$\Delta\Delta G_{nonpolar}$ D	$\Delta\Delta G_{FR}$ calcd E =(A+B+C+D)	$\Delta\Delta G_{FR}$ exptl F
			L858R – WT			
erlotinib	–0.86 ± 0.06	–0.34 ± 0.13	0.21 ± 0.11	0.06 ± 0.003	–0.97 ± 0.07	–0.61
gefitinib	–0.99 ± 0.06	–0.72 ± 0.07	–0.58 ± 0.06	–0.01 ± 0.004	–2.30 ± 0.07	–1.59
AEE788	–2.41 ± 0.06	–0.48 ± 0.07	0.36 ± 0.06	–0.30 ± 0.005	–2.84 ± 0.06	–0.92
			L858R&T790M – L858R			
erlotinib	2.30 ± 0.06	7.42 ± 0.11	–6.56 ± 0.10	0.09 ± 0.003	3.30 ± 0.06	>3.96
gefitinib	–0.10 ± 0.05	–0.06 ± 0.07	0.49 ± 0.06	–0.06 ± 0.004	0.27 ± 0.06	0.90
AEE788	3.39 ± 0.07	3.15 ± 0.09	–4.33 ± 0.07	0.20 ± 0.004	2.40 ± 0.08	1.68
			G719S – WT			
erlotinib	–2.08 ± 0.06	–0.05 ± 0.12	–0.24 ± 0.11	0.04 ± 0.003	–2.38 ± 0.07	not reported
gefitinib	0.74 ± 0.07	–0.85 ± 0.07	1.59 ± 0.07	0.04 ± 0.004	1.50 ± 0.08	0.50
AEE788	–0.65 ± 0.06	–0.78 ± 0.06	0.55 ± 0.05	0.08 ± 0.005	–0.81 ± 0.07	0.02
r^2 =	0.70	0.47	0.19	0.30	0.84	7 data points[c]

[a]$\Delta\Delta G_{FR}$ calcd derived from the difference of two independent simulations (eg L858R – WT) computed using eqs 1-3. [b]$\Delta\Delta G_{FR}$ exptl values from Table 1. Correlations coefficients (r^2 values) obtained from fitting the change in each energy component to $\Delta\Delta G_{FR}$ exptl. All energies in kcal/mol ± standard errors of the mean from 5000 MD snapshots. [c]Data point for erlotinib with double mutant (>3.96) excluded from r^2 calculations given ambiguity in the experimental $\Delta\Delta G_{FR}$ measurement.

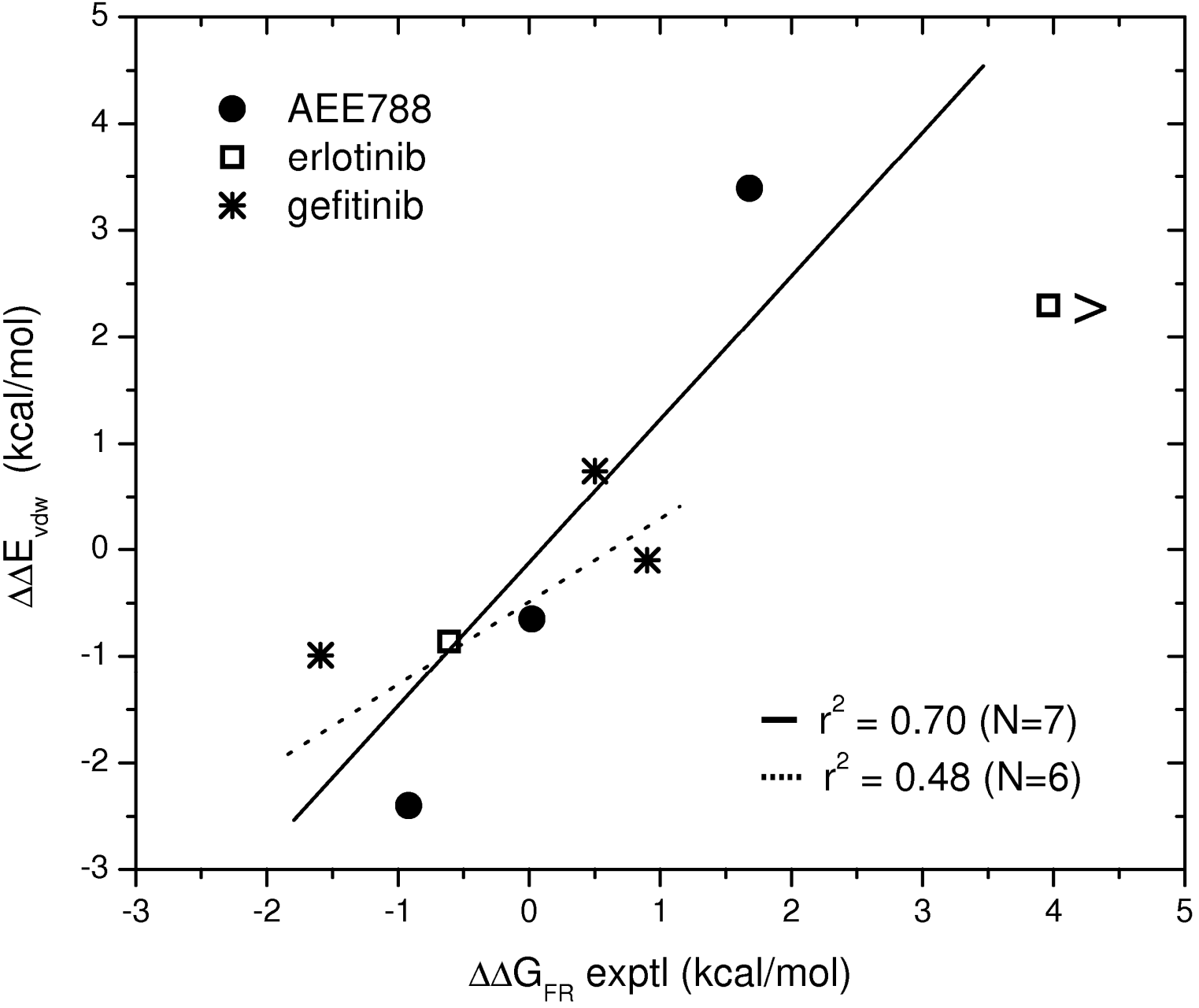
Correlation With Experimental Fold Resistance



Component Correlations



Component Correlations



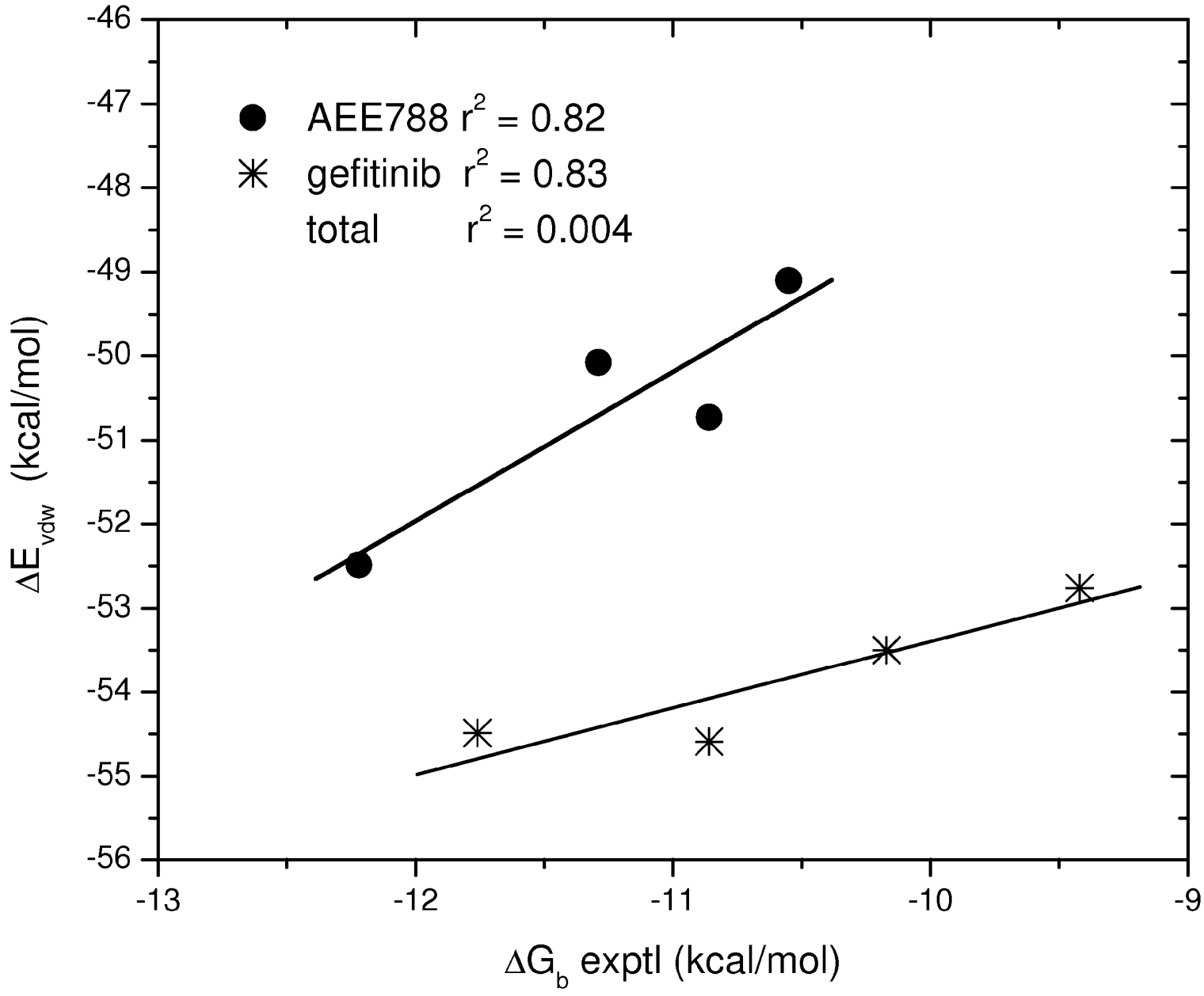
Absolute Free Energies and Components

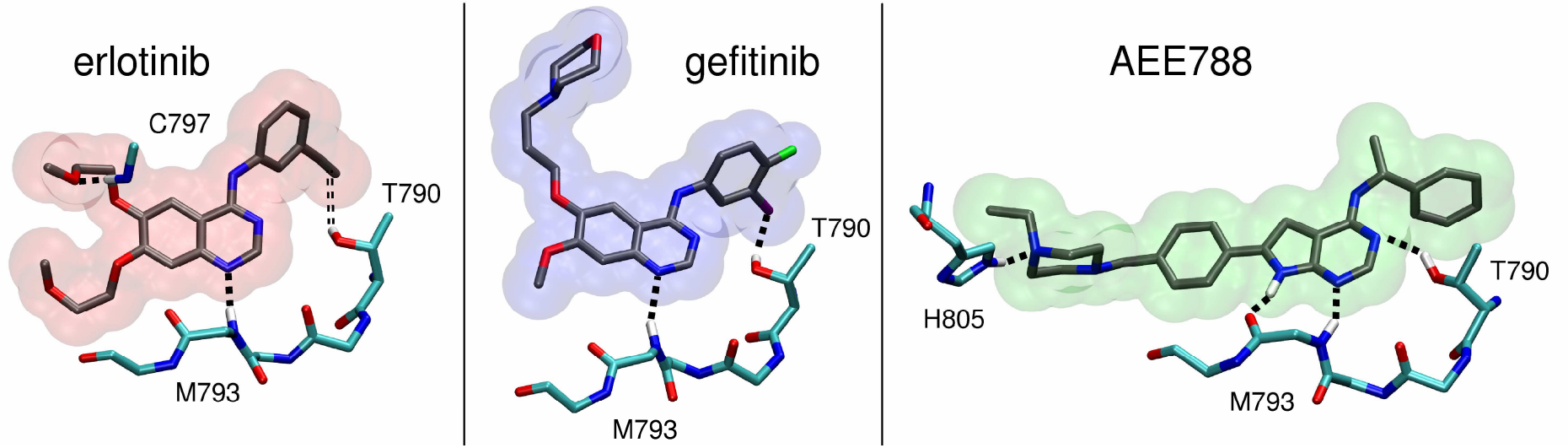
Table 3. Absolute free energies and component decomposition for inhibitors with EGFR.

system	ΔE_{vdw} A	ΔE_{coul} B	ΔG_{polar} C	$\Delta G_{nonpolar}$ D	ΔG_b calcd E=A+B+C+D	ΔG_b exptl[a] F	Hbond G
			erlotinib				
wildtype	−49.01 ± 0.04	−24.71 ± 0.09	39.73 ± 0.08	−6.05 ± 0.002	−39.69 ± 0.05	−10.58[b]	1.82
L858R	−49.86 ± 0.04	−25.04 ± 0.09	39.94 ± 0.07	−5.99 ± 0.002	−40.66 ± 0.05	−11.19[b]	2.17
L858R&T790M	−47.57 ± 0.05	−17.62 ± 0.07	33.38 ± 0.06	−5.89 ± 0.002	−37.36 ± 0.04	> −6.82[c]	0.99
G719S	−51.09 ± 0.04	−24.76 ± 0.08	39.49 ± 0.07	−6.01 ± 0.003	−42.07 ± 0.05	not reported	1.95
			gefitinib				
wildtype	−53.50 ± 0.05	−14.02 ± 0.05	28.80 ± 0.04	−6.30 ± 0.003	−45.01 ± 0.06	−10.17[d]	1.16
L858R	−54.49 ± 0.04	−14.74 ± 0.04	28.22 ± 0.04	−6.31 ± 0.003	−47.32 ± 0.05	−11.76[d]	1.24
L858R&T790M	−54.59 ± 0.04	−14.80 ± 0.05	28.71 ± 0.05	−6.37 ± 0.003	−47.05 ± 0.05	−10.86[d]	1.05
G719S	−52.76 ± 0.04	−14.87 ± 0.06	30.39 ± 0.05	−6.26 ± 0.002	−43.51 ± 0.05	−9.42[e]	1.08
			AEE788				
wildtype	−50.08 ± 0.05	−21.77 ± 0.04	31.97 ± 0.03	−5.93 ± 0.004	−45.81 ± 0.05	−11.29[d]	2.02
L858R	−52.49 ± 0.04	−22.26 ± 0.06	32.33 ± 0.05	−6.24 ± 0.003	−48.65 ± 0.04	−12.22[d]	2.19
L858R&T790M	−49.10 ± 0.06	−19.11 ± 0.07	28.00 ± 0.05	−6.03 ± 0.003	−46.25 ± 0.07	−10.55[d]	2.48
G719S	−50.73 ± 0.04	−22.56 ± 0.04	32.52 ± 0.03	−5.85 ± 0.003	−46.62 ± 0.04	−10.86[e]	1.99

[a]ΔG_b exptl ≈ RTln(activities) at 298.15 K in kcal/mol. [b]Ki values (nM) from Carey et al. [c]IC_{50} values (nM) from Ji et al. [d]Kd values (nM) from Yun et al (2008). [e]Kd values (nM) from Yun et al (2007).

Component Correlations

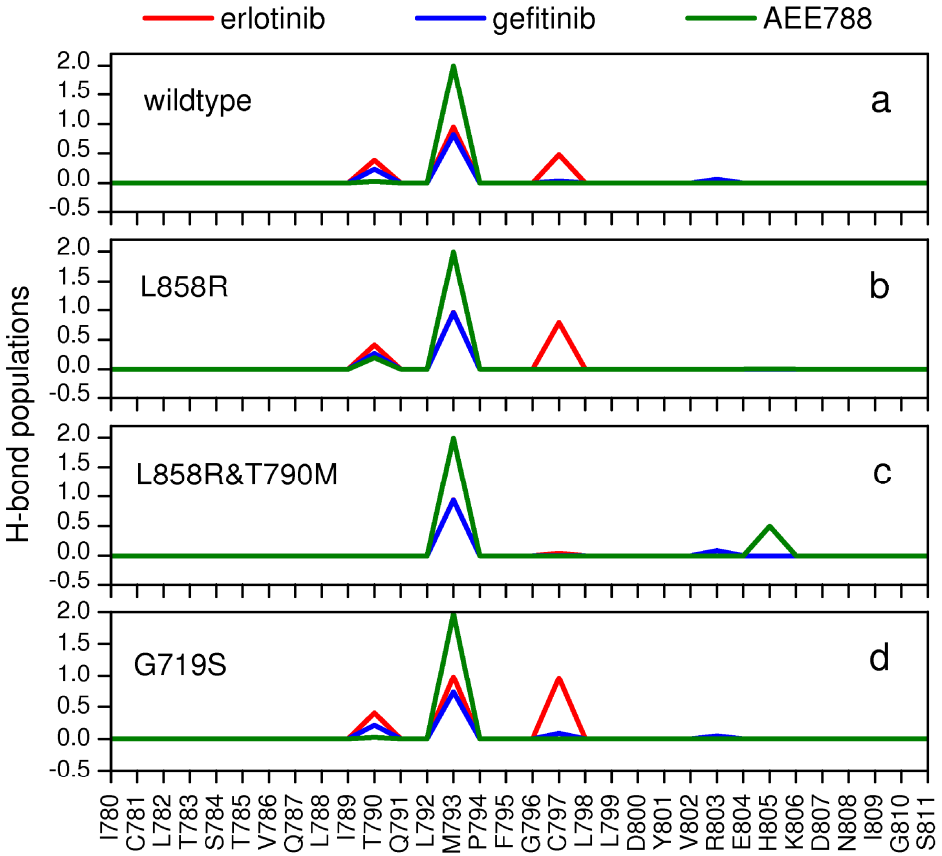




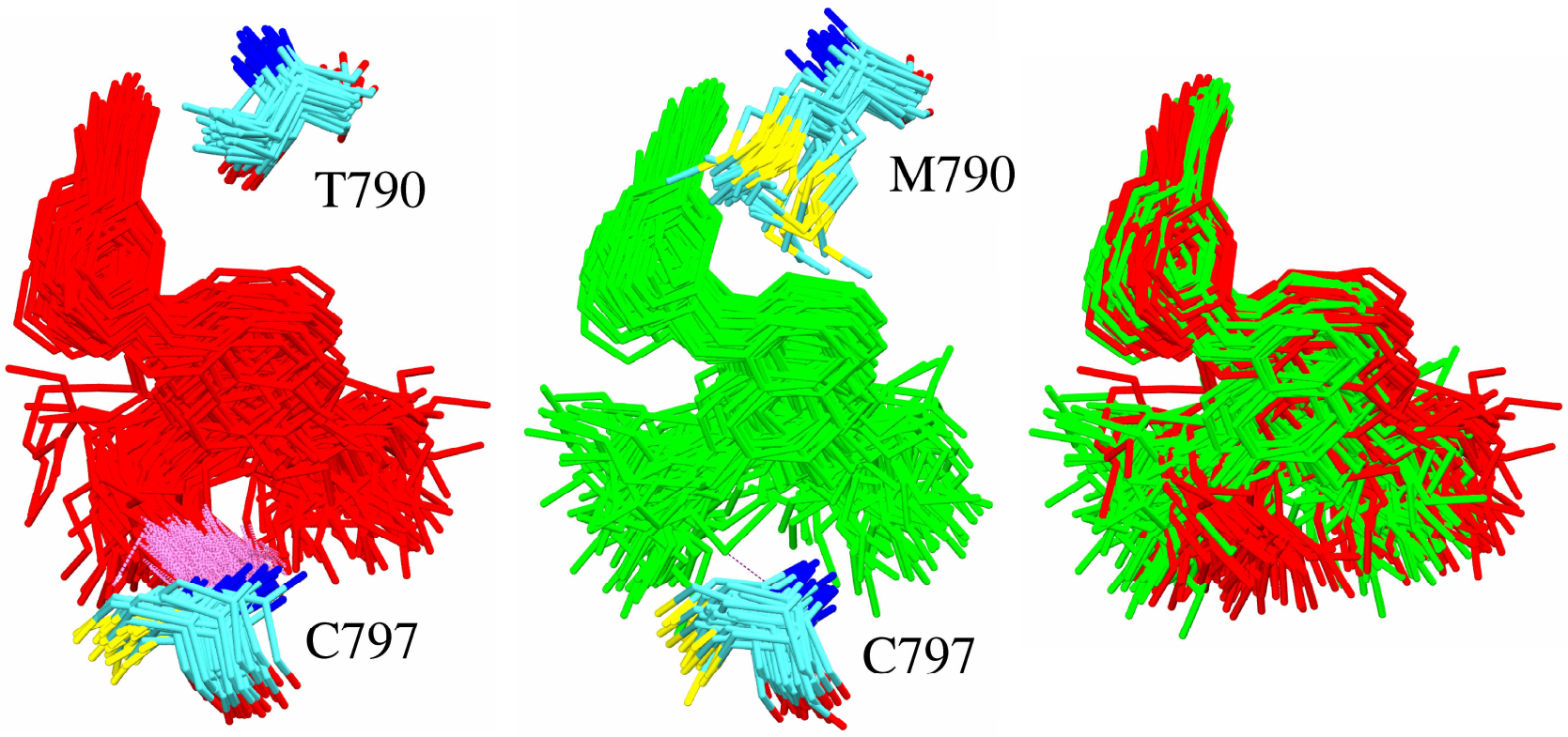
Key Inhibitor H-bonding

- M793 at backbone
- interactions with T790
  - erlotinib Pi-type h-bond
  - gefitinib Cl-type h-bond
  - AEE788 scaffold

- erlotinib at C797
- AEE788 at 805



Erlotinib Binding Comparison



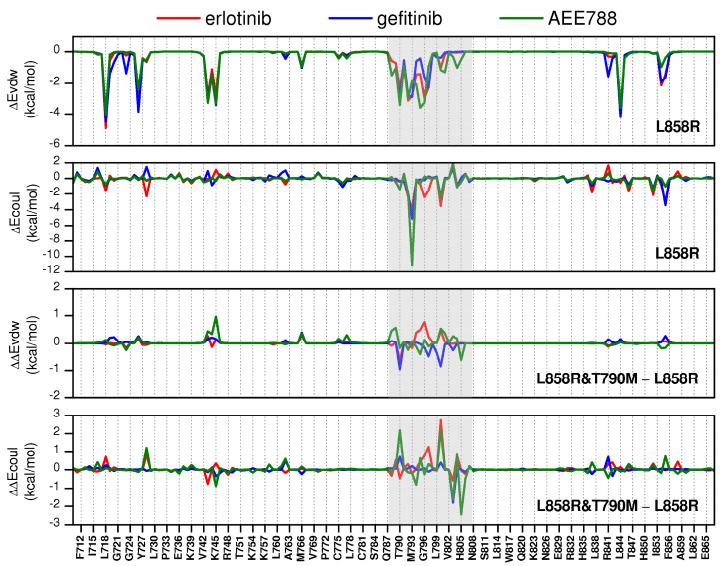
L858R simulations

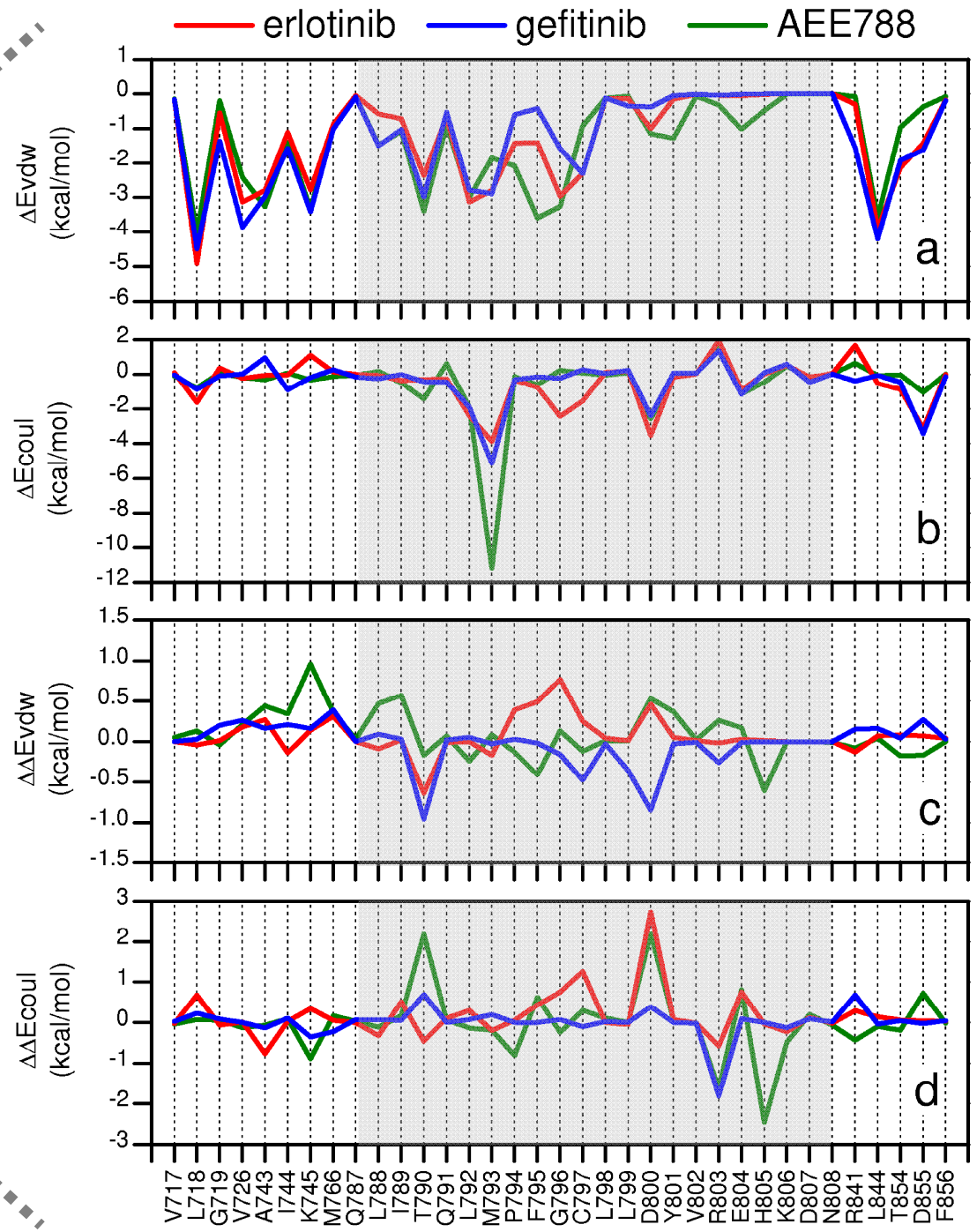
L858R&T790M simulations

overlay

The T790M mutation does not lead to a steric clash with erlotinib however there is change in H-bonding at position C797

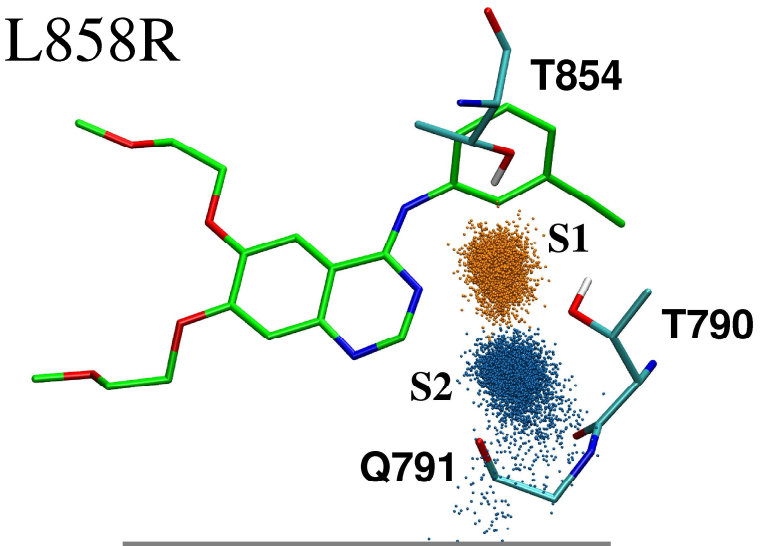
Footprints: Energy

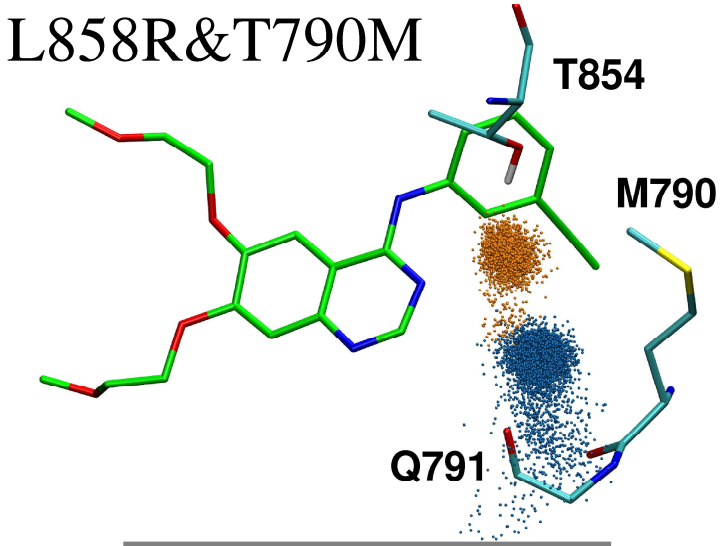




Water Mediated Interactions

a



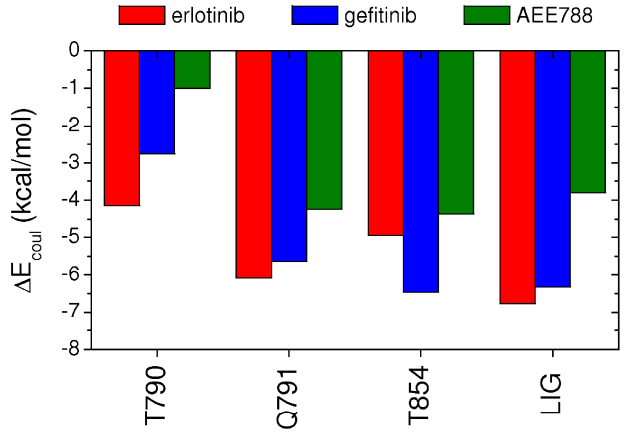


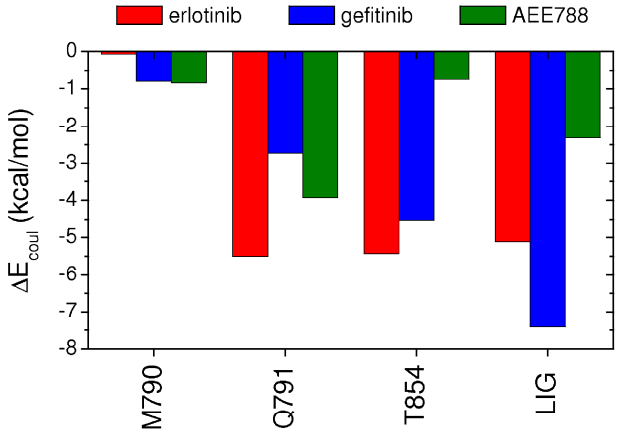
b

	S1 count	S2 count
erlotinib	4628	4168
gefitinib	4544	4385
AEE788	2489	5076

	S1 count	S2 count
erlotinib	3279	4111
gefitinib	3126	2536
AEE788	172	3563

c





Conclusions

- Good agreement ($\Delta\Delta$Gb calcd vs. exptl correlation, $r^2 = 0.84$, N=7)
- VDW is the most correlated term
- Coul. is important for orienting the ligand in the pocket
- FP regions with similar and dissimilar energies suggest good convergence/reproducibility
- Increased VDW interactions at M790 suggest this is not a steric clash mechanism
- Coulombic energies mirror H-bond trends (AEE788 shows largest interactions at M793)
- Flatter difference FP profiles for gefitinib shows agreement with exptl FR trend
- Water-mediated interaction is meaningful
- Mutants effect on affinity for ATP is not the sole reason for modulated affinity for the three inhibitors
- Differences of direct and water-mediated interactions contribute to changes in energies

Acknowledgments

- Funding: Stony Brook University Office of the Vice President for Research and School of Medicine (Carol M. Baldwin Breast Cancer Research Award), and NIH grants F31CA134201 (to TEB) and R01GM083669 (to RCR).

- Todd Miller (helpful discussion)

- Computational Science Center at Brookhaven National Laboratory (computational support)

- Rizzo Research Group: Robert C. Rizzo; Sudipto Mukherjee; Rashi Goyal; Brian McGillick, Yulin Huang

Questions??