Case Study #2

1682 J. Chem. Inf. Model. **2010,** *50,* 1682–1692

Test MM-PB/SA on True Conformational Ensembles of Protein−Ligand Complexes

Yan Li, Zhihai Liu, and Renxiao Wang*

State Key Laboratory of Bioorganic Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Science, 345 Lingling Road, Shanghai 200032, People's Republic of China

Received January 25, 2010

J. Chem. Inf. Model., 2010, 50 (9):1682–1692 doi: 10.1021/ci100036a

VERTEX

Case Study #2

1682 *J. Chem. Inf. Model.* **2010,** *50,* 1682–1692

Test MM-PB/SA on True Conformational Ensembles of Protein−Ligand Complexes

Yan Li, Zhihai Liu, and Renxiao Wang*

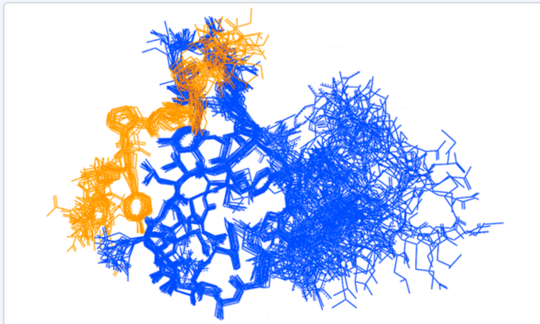
State Key Laboratory of Bioorganic Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Science, 345 Lingling Road, Shanghai 200032, People's Republic of China

Received January 25, 2010

- Binding free energies calculated with MMPBSA and MMGBSA
- Compared methods for generating protein conformations
  – Molecular dynamics
  – Averaged NMR ensembles
  – Representative from NMR ensembles
- 7 variants of charge and conformer generation

VERTEX

From the Abstract



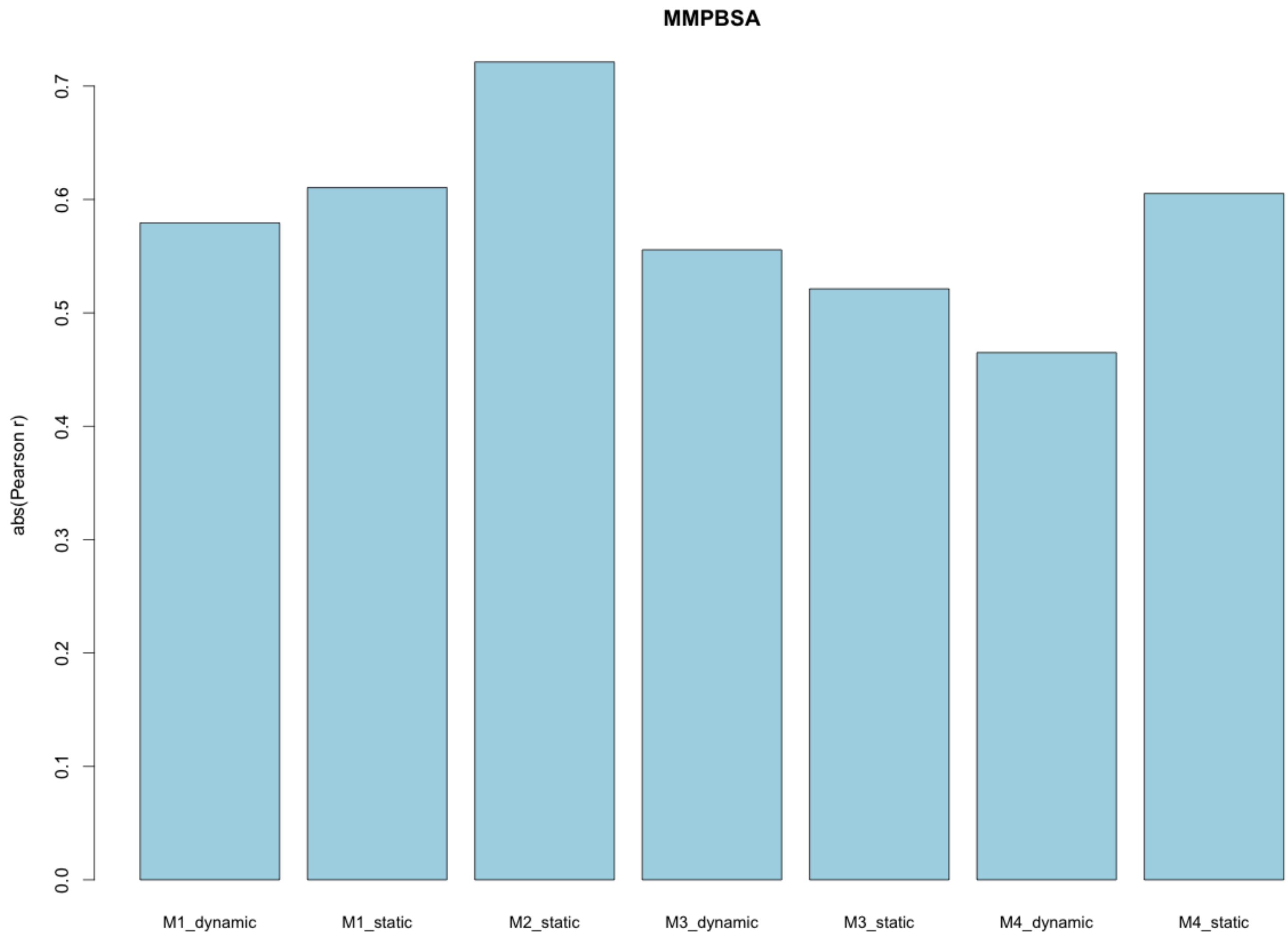
The molecular mechanics Poisson–Boltzmann surface area (MM-PB/SA) method has been popular for computing protein–ligand binding free energies in recent years. All previous evaluations of the MM-PB/SA method are based upon computer-generated conformational ensembles, which may be affected by the defective computational methods used for preparing these conformational ensembles. In an attempt to reach more convincing conclusions, we have evaluated the MM-PB/SA method on a set of 24 diverse protein–ligand complexes, each of which has a set of conformations derived from NMR spectroscopy. Our results indicate that both MM-PB/SA and molecular mechanics generalized Born surface area (MM-GB/SA) are able to produce a modest correlation between their results and the experimentally measured binding free energies on our test set. In particular, both MM-PB/SA and MM-GB/SA produced better results by using a representative structure (R = 0.72–0.79) rather than averaging over the conformational ensemble of each given complex (R = 0.61–0.74). A head-to-head comparison with four selected scoring functions (X-Score, PLP, ChemScore, and DrugScore) on the same test set reveals that MM-PB/SA and MM-GB/SA results are marginally better than those produced by scoring funcitons, supporting the value of the MM-PB/SA method. Nevertheless, scoring functions are still more cost-effective options, especially for high-throughput tasks.

In particular, both MM-PB/SA and MM-GB/SA produced better results by using a representative structure (R) 0.72-0.79) rather than averaging over the conformational ensemble of each given complex (R) 0.61-0.74

The Study

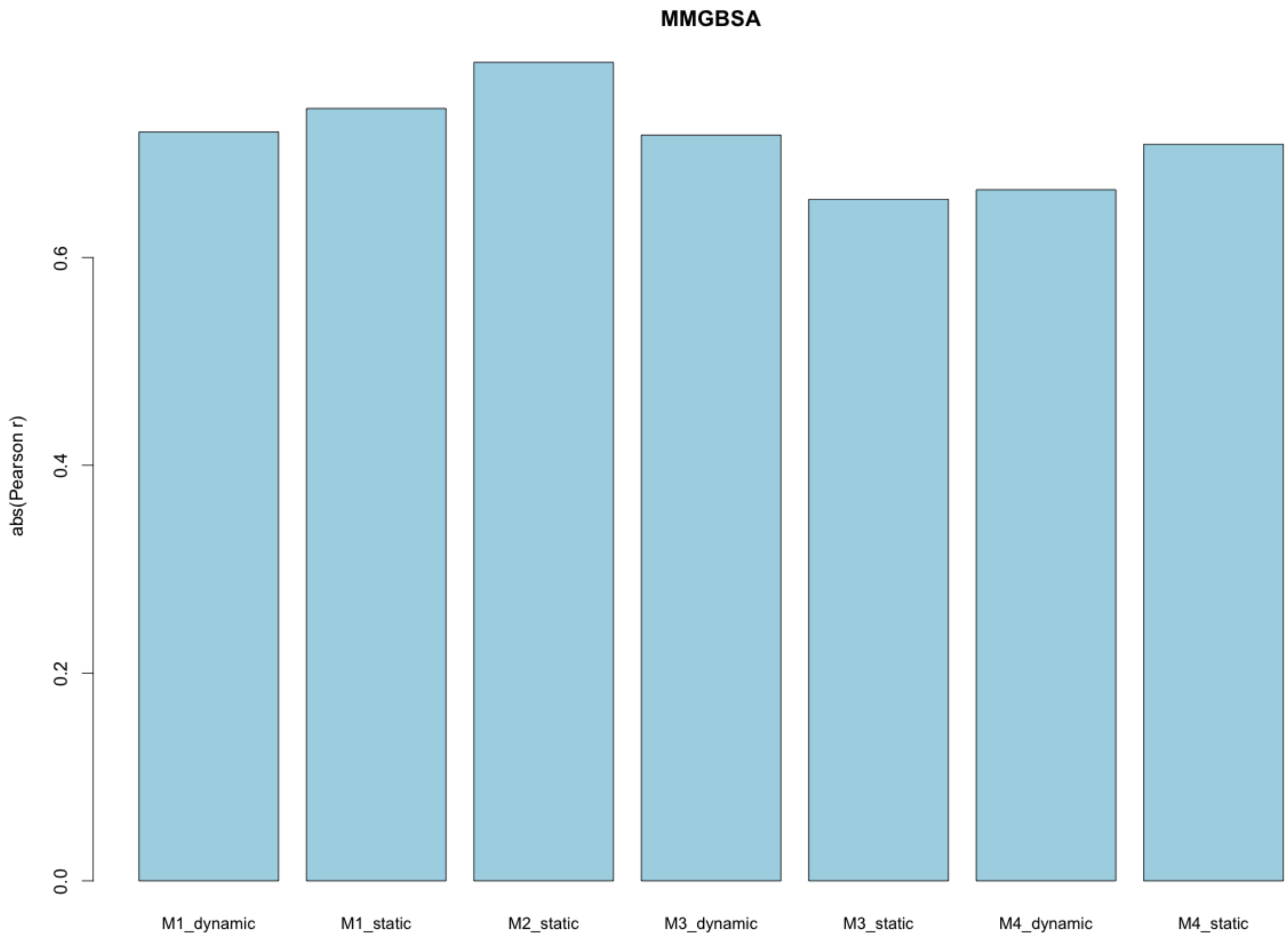
- Generate protein ensembles 7 ways
  - MD average
  - NMR average
  - NMR representatives
- Generate electrostatic potentials 2 ways
  - MMGBSA
  - MMPBSA
- Calculate Pearson r for $\log K_d$ vs calculated free energy

MMPBSA Results



VERTEX

MMGBSA Results



Let's Grab the Data

Table S1. MM-PB/SA and MM-GB/SA results on all 24 protein-ligand complexes [a]

PDB code	$-\log K_d$	ΔE_{elec}	ΔE_{vdw}	ΔG_{PB}	ΔG_{GB}	ΔG_{SA} [b]	$\Delta G_{MM\text{-}PB/SA}$ [c]	$\Delta G_{MM\text{-}GB/SA}$ [d]	$T\Delta S$	$\Delta G_{PB\text{-}total}$ [e]	$\Delta G_{GB\text{-}total}$ [f]	ΔG_{expt} [g]
1BZF	8.30	-181.76	-38.08	171.97	172.13	-5.92	-53.79	-53.63	-22.72	-31.07	-30.91	-11.32
1DDM	5.77	-153.74	-61.91	197.82	188.77	-10.50	-28.34	-37.38	-27.26	-1.08	-10.12	-7.87
1F40	6.00	-13.51	-31.97	27.37	22.47	-4.92	-23.04	-27.93	-15.52	-7.52	-12.41	-8.18
1FHR	4.00	-526.67	-16.18	501.98	518.02	-4.48	-45.35	-29.31	-46.70	1.35	17.39	-5.45
1IH0	5.00	7.07	-39.65	8.93	10.81	-5.69	-29.34	-27.46	-27.56	1.78	0.10	-6.82
1J5I	3.64	-8.70	-38.44	39.45	28.10	-5.50	-13.19	-24.54	-19.83	6.64	-4.71	-4.96
1JMQ	4.40	-33.90	-38.93	44.70	48.31	-5.60	-33.73	-30.12	-33.48	-0.25	3.36	-6.00
1K9Q	3.15	-24.81	-30.12	39.25	37.07	-4.46	-20.14	-22.32	-25.88	5.74	3.56	-4.30
1K9R	3.30	36.98	-31.64	-21.67	-25.21	-4.79	-21.12	-24.66	-26.01	-4.89	1.35	-4.50
1L2Z	3.69	-318.63	-37.07	322.67	322.43	-6.43	-39.46	-39.70	-29.09	-10.37	-10.61	-5.04
1LXF	4.10	4.00	-41.91	20.28	13.81	-5.92	-23.55	-30.02	-27.13	3.58	-2.89	-5.59
1O5P	10.00	-28.64	-71.78	76.54	58.19	-9.21	-33.09	-51.44	-19.79	-13.3	-31.65	-13.64
1P7M	4.38	-27.92	-23.01	35.74	31.93	-2.90	-18.09	-21.90	-6.50	-11.59	-15.40	-5.97
1Q5L	3.22	-81.34	-57.63	92.67	90.56	-8.63	-54.93	-57.04	-32.29	-22.64	-24.75	-4.39
1XSC	4.30	61.78	-32.41	-35.72	-25.65	-5.73	-12.08	-2.01	-18.67	6.59	16.66	-5.87
2GTV	5.15	-406.65	-15.87	338.30	356.75	-3.50	-87.72	-69.27	-11.13	-76.59	-58.14	-7.03
2JMJ	5.04	-102.06	-41.81	100.94	96.78	-8.22	-51.15	-55.31	-40.19	-10.96	-15.12	-6.87
2JOA	5.96	-56.89	-53.20	69.02	62.31	-8.24	-49.31	-56.02	-37.72	-11.59	-18.30	-8.13
2JUP	3.81	-20.47	-24.70	27.88	27.50	-4.63	-21.92	-22.30	-27.01	5.09	4.71	-5.19
2NMB	6.28	-486.39	-40.12	497.90	496.05	-8.65	-37.26	-39.11	-26.52	-10.74	-12.59	-8.56
2OQS	5.88	-270.48	-45.37	269.85	273.50	-7.07	-53.07	-49.42	-31.46	-21.61	-17.96	-8.01
2P0X	6.31	-848.12	-26.48	822.72	831.84	-4.90	-56.78	-47.66	-32.33	-24.45	-15.33	-8.60
2RLY	3.57	25.56	-28.23	-14.41	-14.05	-4.52	-21.61	-21.26	-22.55	0.94	1.29	-4.87
2RM0	4.03	28.23	-30.61	-10.55	-13.39	-4.72	-17.65	-20.49	-22.18	4.53	1.69	-5.49
Correlation Coefficient (R_p) [h]										*0.721*	*0.788*	

VERTEX

Adding a Confidence Interval to Pearson's R

$$z' = .5[\ln(1+r) - \ln(1-r)]$$

$$\sigma_{z'} = \frac{1}{\sqrt{N-3}}$$

$$z' \pm z\sigma_{z'}$$

Let's Assume R=0.7 and N=22

$$z' = .5[\ln(1+r)\ -\ \ln(1-r)]$$

$z' = 0.87$

$$\sigma_{z'} = \frac{1}{\sqrt{N-3}}$$

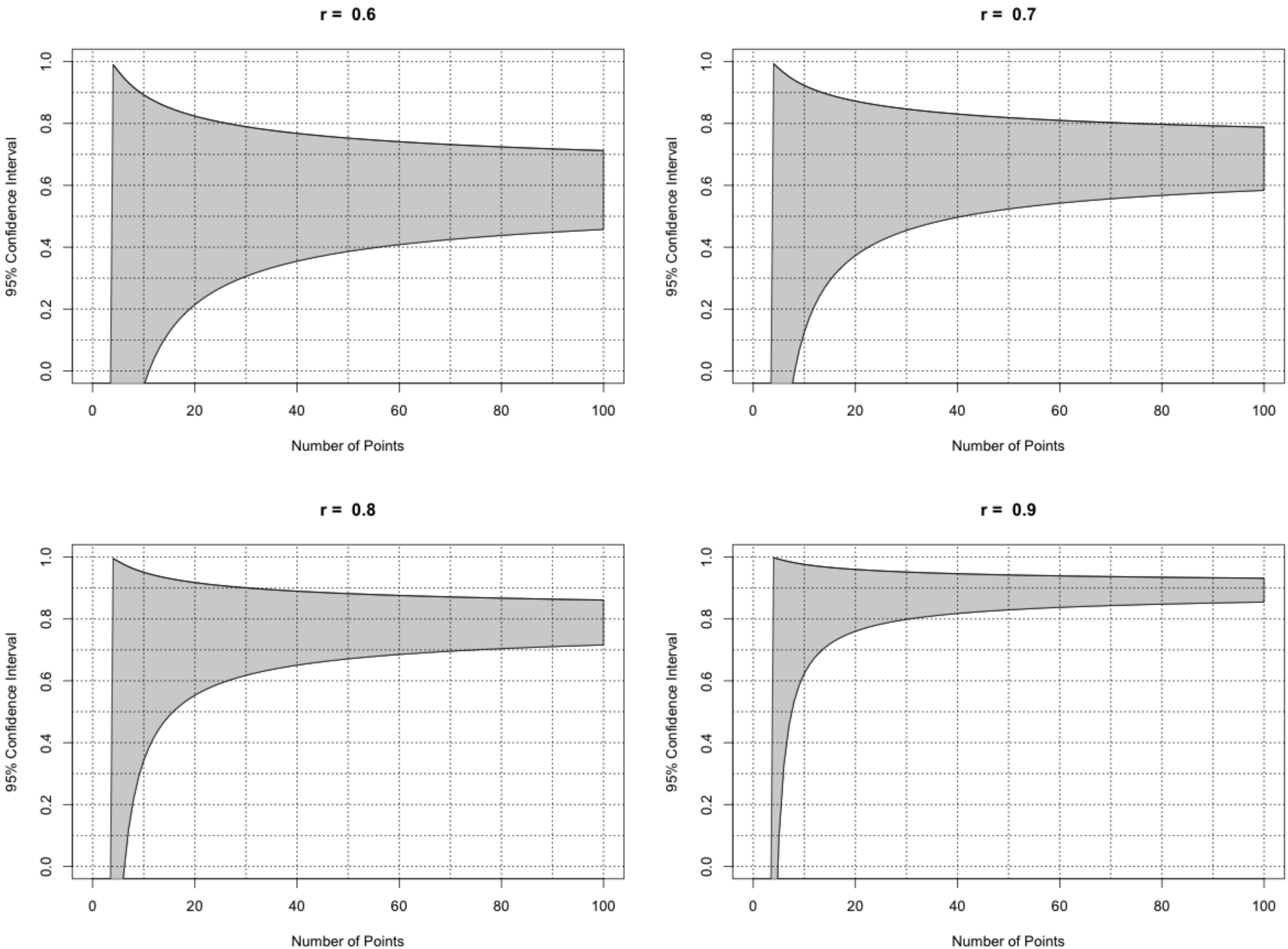
$\sigma_{z'}=0.23$

$$z' \pm z\sigma_{z'}$$

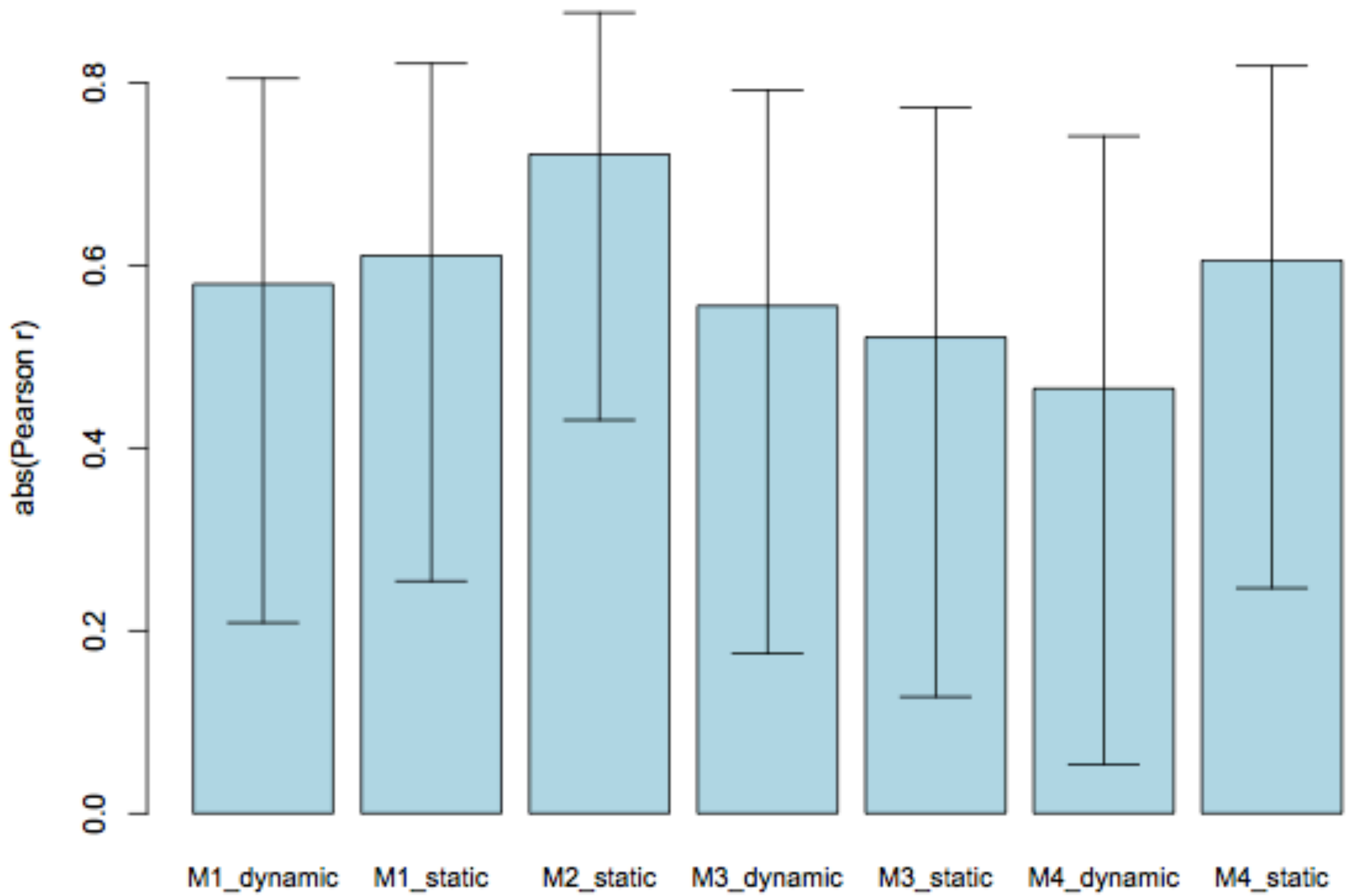
```
Lower limit = 0.87-(2.08*0.23)=0.39
Upper limit = 0.87+(2.08*0.23)=1.35
```

$$0.37 \leq r \leq 0.87$$

VERTEX


r = 0.6
r = 0.7
r = 0.8
r = 0.9
95% Confidence Interval
Number of Points
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100

MMPBSA with Error Bars



VERTEX

MMGBSA with Error Bars