Boosting Virtual Screening Enrichments Using Data Fusion

Coalescing 2D fingerprints, shape, and docking

Sastry, G. M., Inakollu, V. S. S., Sherman, W.

SCHRÖDINGER.

The Big Picture

- Ideally, we would run QM/MD/FEP for all binding energy calculations
  - Way too expensive
- Even docking with protein flexibility can be too expensive for large datasets using typical hardware
  - And virtual screening results have not been validated
- **Can we devise strategies within the current virtual screening paradigm to improve enrichment results?**

Presentation Outline

- Datasets & Metrics
- Fingerprints
- Shape
- Docking
- Data fusion

Virtual Screening Datasets

- Set 1: Glide validation set
  - 65 targets
  - ~20 actives/target
  - 1000 decoys
- Set 2: MDDR from McGaughey *et al.*
  - 11 targets
  - 8-257 actives/target
  - ~25K decoys
- Set 3: DUD
  - 40 targets
  - ~20 actives/target
  - ~2000 decoys

Mostly MDDR results are presented here, but all results are in:
Sastry M *et al. Journal of Chemical Information and Modeling* **53,** 1531–1542 (2013)

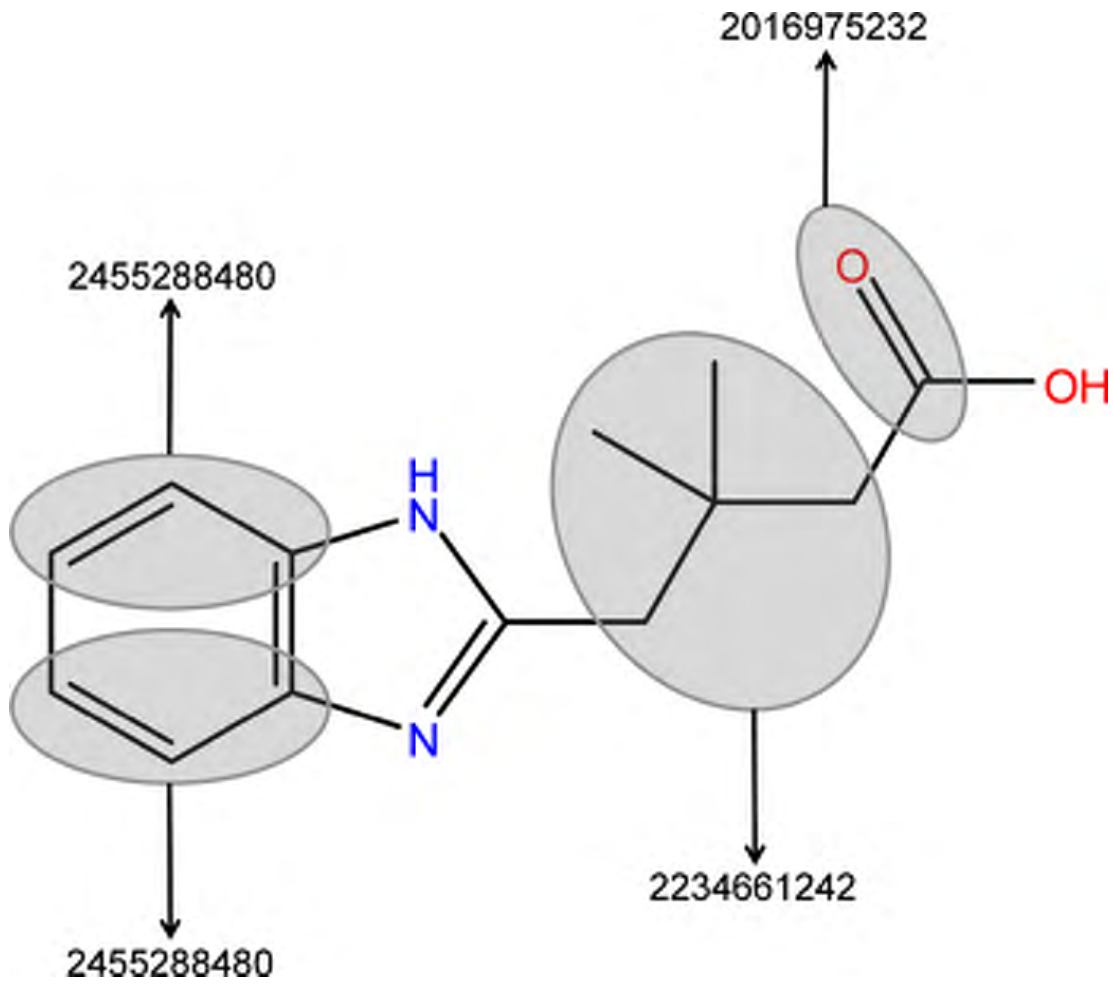
Enrichment Metrics

- BEDROC*
  - Boltzmann-enhanced discrimination of receiver-operating characteristic
  - Weights the early part of the ROC curve but accounts for the full curve
  - α allows tuning for how heavy to weight early enrichment
  - α=160.9 corresponds to 80% of the BEDROC score being accounted for in the top 1% of the database screen
  - α=20 corresponds to 80% of the BEDROC score being accounted for in the top 8% of the database screen
  - Maximum value=1.0
- EF(1%)
  - Enrichment of actives in top 1% of DB
  - Maximum value=100
- EF(10%)
  - Enrichment of actives in top 10% of DB
  - Maximum value=10

* Truchon and Bayly, *JCIM* **2007** 47 (2) 488–508

Fingerprints

- Up to 64-bit hashed fingerprints (default 32-bit = 2^{32})
- Details in 2 publications:
  - Sastry et al., *J Chem Inf Model*, **2010**, 50(5)
    - Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments
  - Duan et al., *J Mol Graph Model*, **2010**, 29
    - Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods



SCHRÖDINGER.

Effect of Address Space Size

Target	#Heavy Atoms	Query #On Bits			EF(1%)	
		2^{10}	2^{32}	2^{64}	2^{10}	2^{32}
CA	13	116	120	120	47.5	52.5
CDK2	35	953	2665	2665	7.8	11.7
COX2	26	264	303	303	10.1	18.7
DHFR	33	371	483	483	15.4	38.4
ERα	29	178	193	193	10.8	10.8
HIV Protease	45	504	694	694	5.9	28.7
HIV-RT	29	337	408	408	2.0	3.4
Neuraminidase	28	322	371	371	25.0	41.6
PTP1B	18	279	332	332	50.0	50.0
Thrombin	35	462	607	607	4.5	30.5
TS	53	439	569	569	48.4	70.9
Average	**31.3**	**384**	**613**	**613**	**20.7**	**32.5**

Linear fingerprints, Daylight atom types, no bit scaling, Tanimoto similarities

SCHRÖDINGER.

Fingerprint Methods

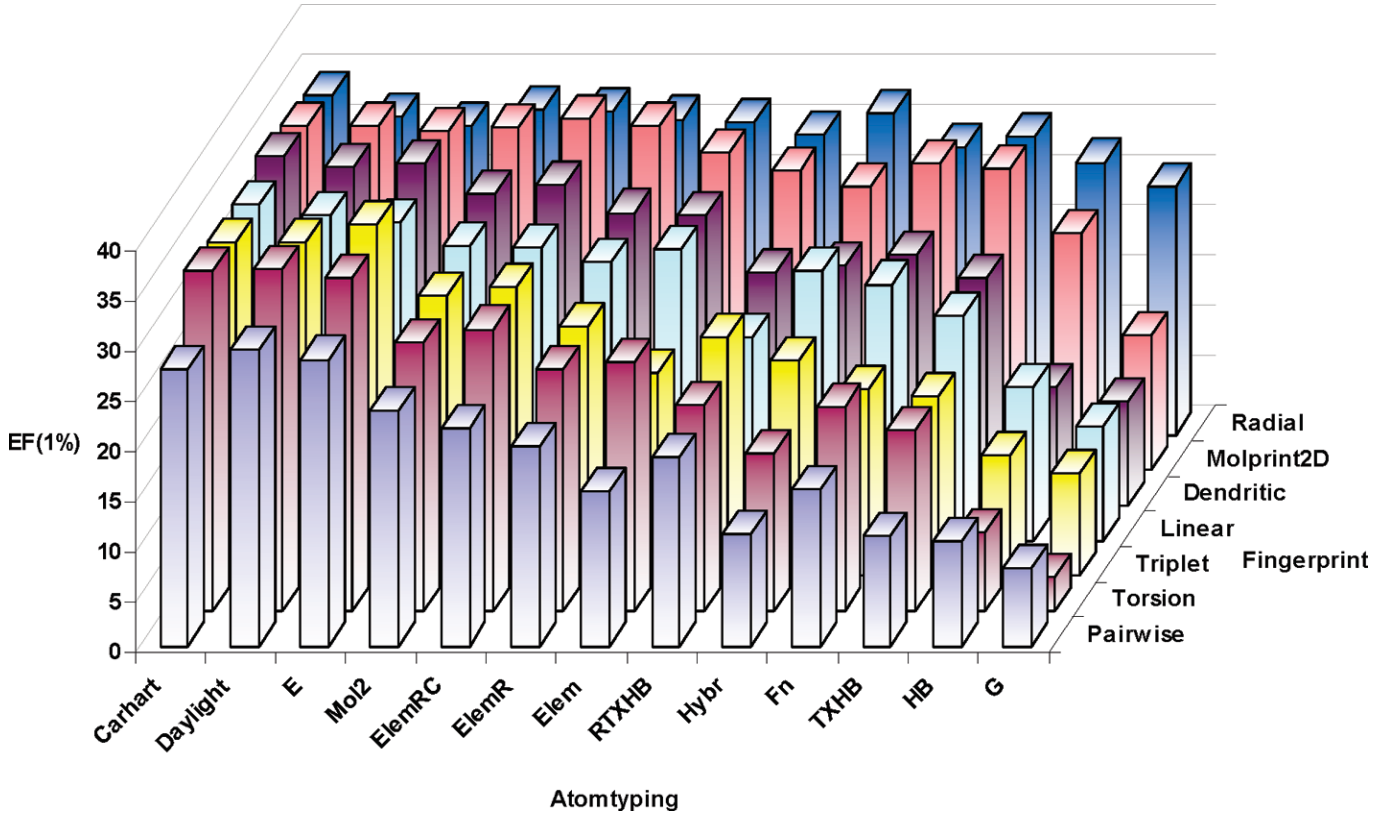
Multiple methods and options implemented in Canvas

- 13 atom types
  - Plus custom types
- 13 bit scaling rules
- 20+ metrics

FP Type	Description
Linear	Linear fragments + ring closures
Dendritic	Linear and branched fragments
Radial	Fragments that grow radially from each atom. Also known as extended connectivity fingerprints (ECFPs)[42]
Pairwise	Pairs of atoms, [44] differentiated by type and the distance separating them: $Type_i$ - $Type_j$ -d_{ij}
Triplet	Triplets of atoms, differentiated by type and the three distances separating them: $Type_i$ –d_{ij}-$Type_j$ –d_{jk}-$Type_k$ –d_{ki}
Torsion	Four consecutively bonded atoms, [45] differentiated by type: $Type_i$ - $Type_j$ - $Type_k$ -$Type_l$
MOLPRINT2D	A radial-like fingerprint that encodes atom environments using lists of atom types located at different topological distances [46,47]
MACCS	SMARTS-based implementation of the MACCS structural keys[36]

Summary of Fingerprint Screening Results

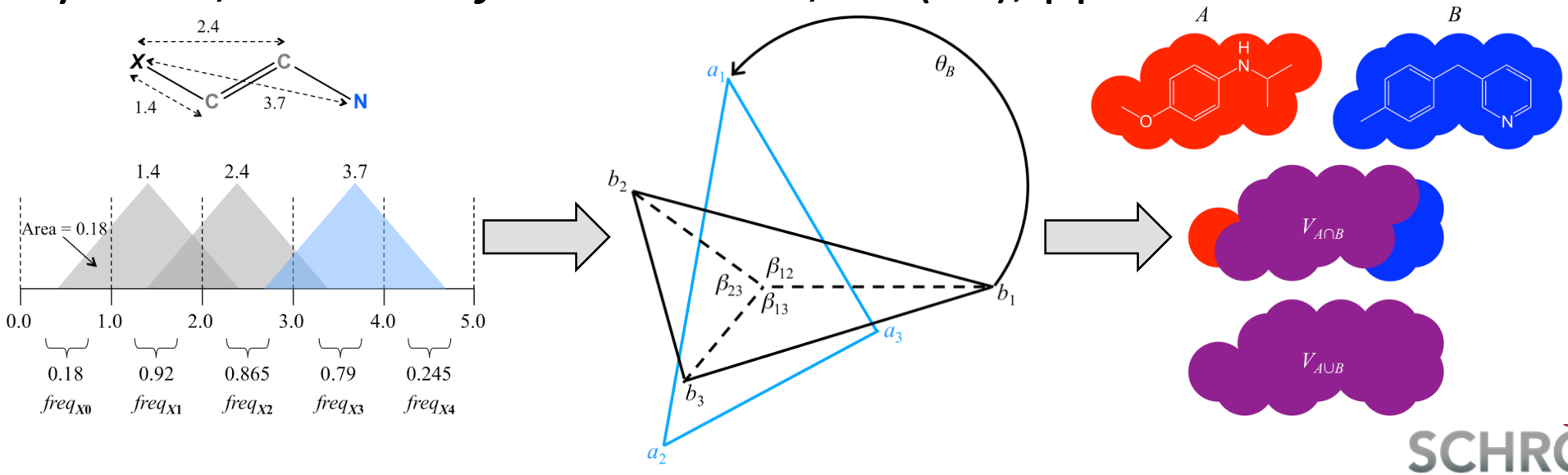
- Sastry et al., *J Chem Inf Model* **2010** *50:* 771
  - "Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments"
- Best EF(1%)=35.1 Molprint2D and element + ring/cyclic atom types
  - 33.6 with default Molprint2D settings



SCHRÖDINGER.

Phase Shape Overview

- Based on the principle of rapid initial alignments using atom triplets followed by refinement and volume overlap scoring
- Atom triplets derived from local atom environments
- Fast superposition using 2D least squares
- Hard sphere atom volume overlaps for similarity assessment
- Sastry at al., *J Chem Inf Model* **2011**, 51 (10), pp 2455–246



Virtual Screening: Effect of Atom Types

- Consistent improvement with more specific atom types
- Pharmacophore treatment outperforms all atom-based schemes

	EF(1%)				
Target	**Shape Only**	**QSAR**	**Element**	**MMod**	**Pharm**
CA	10.0	25.0	27.5	32.5	32.5
CDK2	16.9	20.8	20.8	23.4	19.5
COX2	21.4	19.1	16.7	19.5	21.0
DHFR	7.7	3.9	11.5	23.1	80.8
ER	9.5	17.6	17.6	13.5	28.4
HIVpr	13.2	17.7	19.1	14.0	16.9
HIVrt	2.7	2.0	4.7	4.7	2.0
NA	16.7	16.7	16.7	16.7	25.0
PTP1B	12.5	12.5	12.5	12.5	50.0
Throm	1.5	4.0	4.5	8.5	28.0
TS	19.4	32.3	35.5	51.7	61.3
Average	**11.9**	**15.6**	**17.0**	**20.0**	**33.2**
Median	**12.5**	**17.6**	**16.7**	**16.7**	**28.0**

Improved Enrichment

Docking

- Glide HTVS
  - ~1-2 s/cmpd
  - SP produces ~10% better enrichments at 10x computational cost
- Default Protein Preparation Wizard
  - Protein preparation paper published in JCAMD:

    "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments"

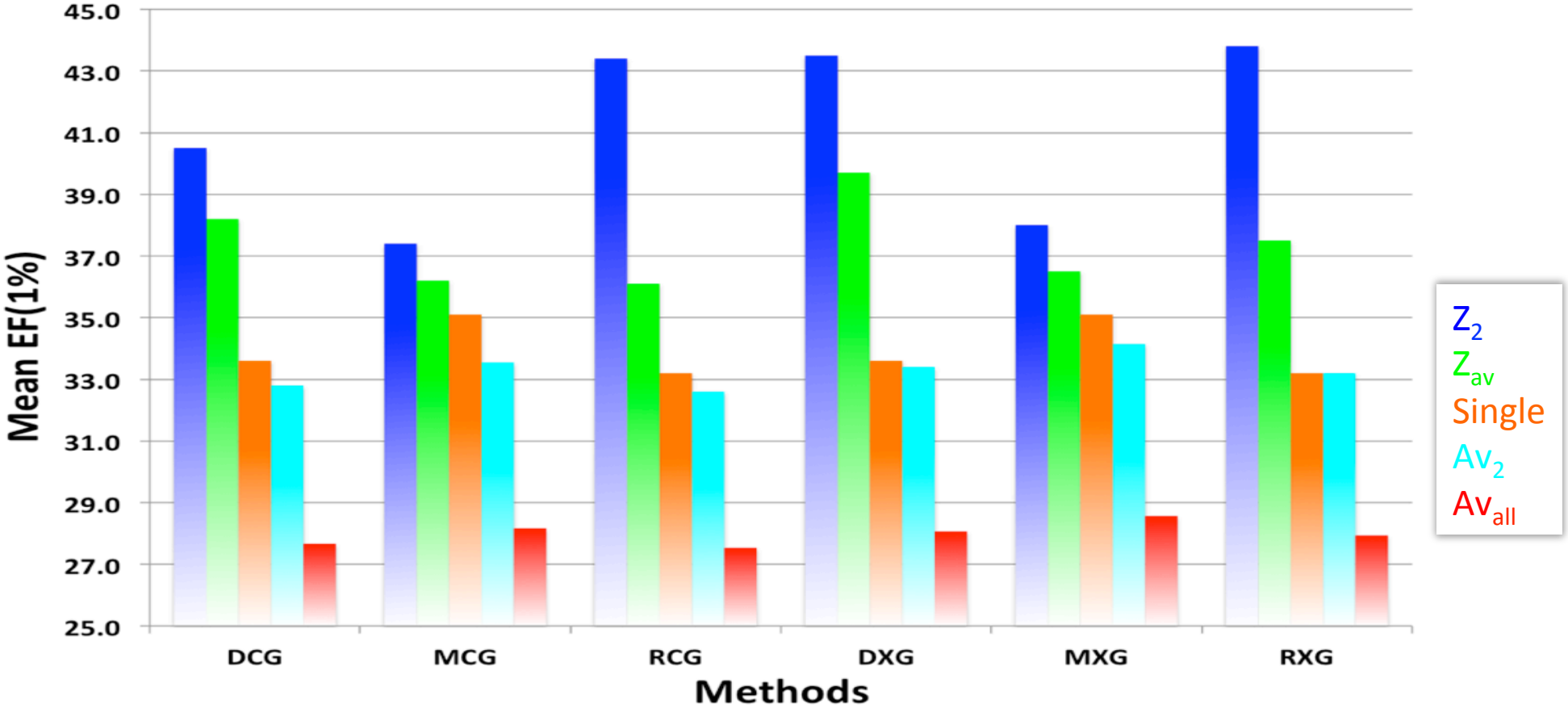
    Sastry et al., *J Comp-Aided Mol Des*, **2013**, *27(3)*, pp 221-234
- Database ligands prepared with LigPrep and Epik

Combining Multiple Scores

- Scores from fingerprints, shape, and docking cannot be directly combined
- Various options exist for combining:
  - Consensus ranking
  - Parallel selection
  - Average of normalized scores
- We like normalized scores for various reasons
  - Emphasizes underlying score, not just rank
  - Easier to gain confidence intervals
- Standard Score (aka Z-score)
  - Normalize each distribution to mean=0 and stddev=1
  - Invert sign of GlideScore so bigger is better (like FP and shape)
- Question: Combine all scores or a subset?

Comparison with Different Screening Protocols

- Z_2 scoring outperforms single methods

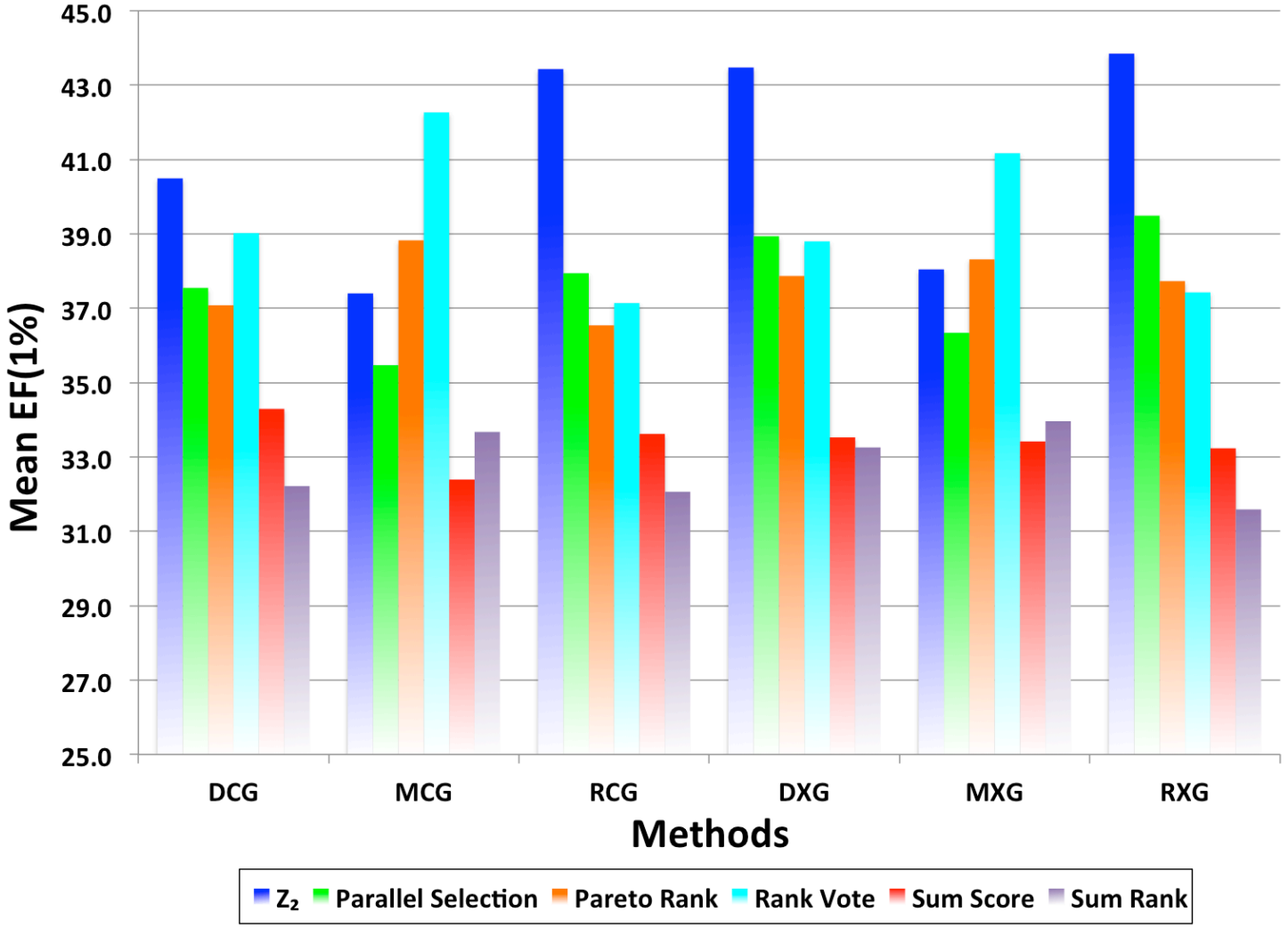


Fingerprints:	D=Dendritic M=Molprint2D
Phase Shape query:	C=ConfGen X=x-ray
Docking:	G=Glide

SCHRÖDINGER.

Comparison to Different Data Fusion Algorithms

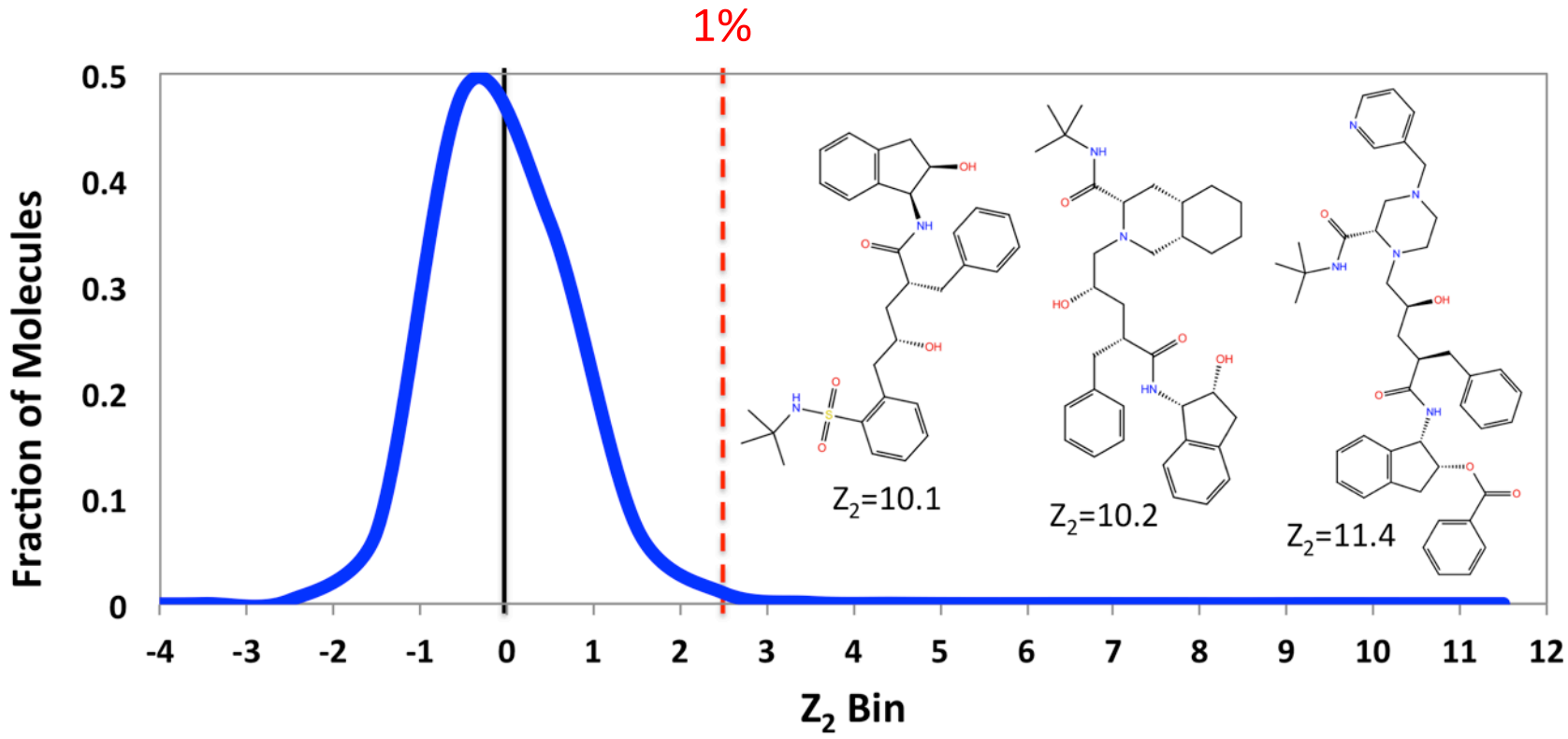
Z_2 generally outperforms other data fusion approaches



SCHRÖDINGER.

HIV Protease Example

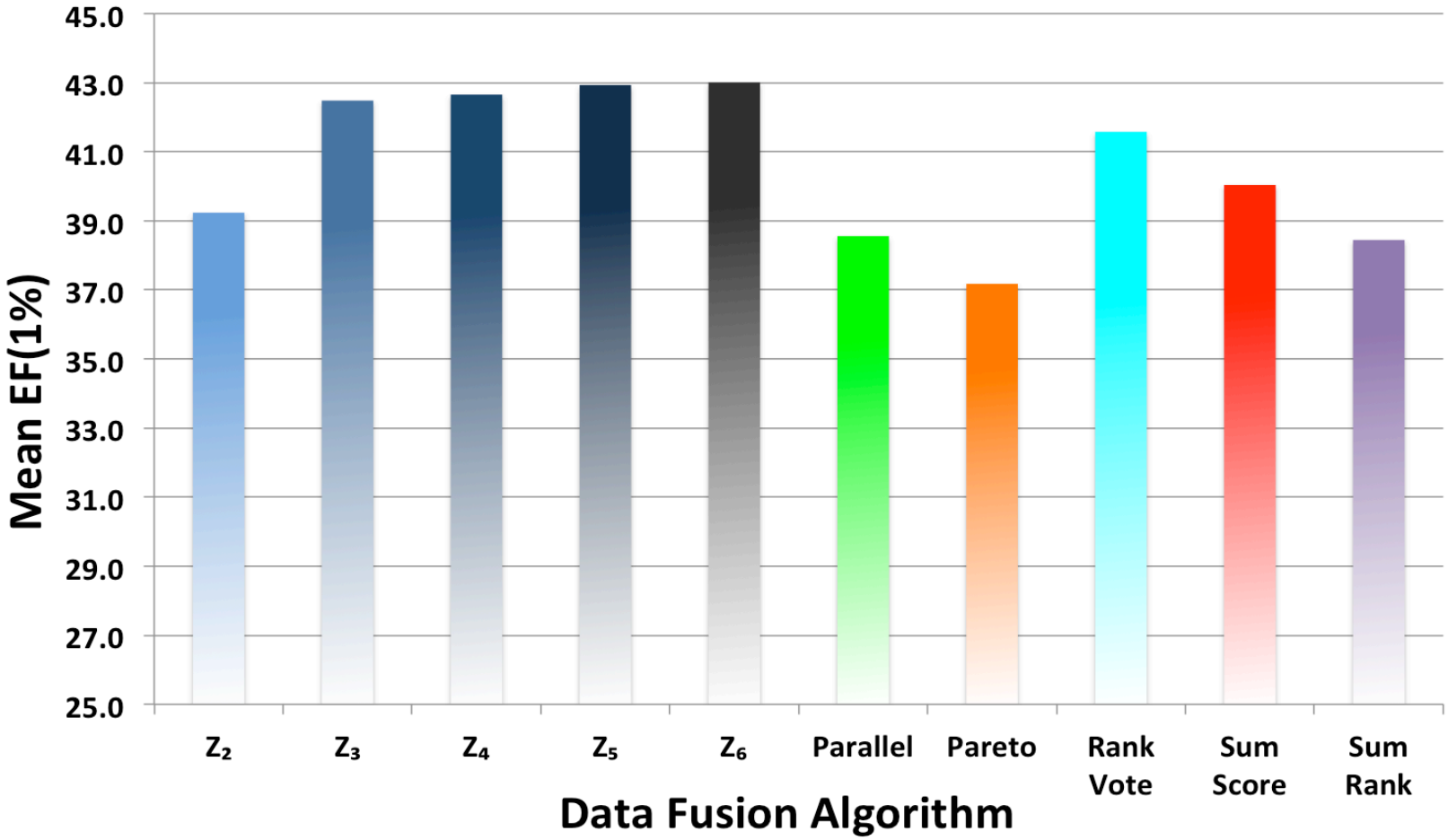
- We want a narrow peak with a fat positive tail
- Top compounds are significantly above mean
- Top compounds are active



SCHRÖDINGER.

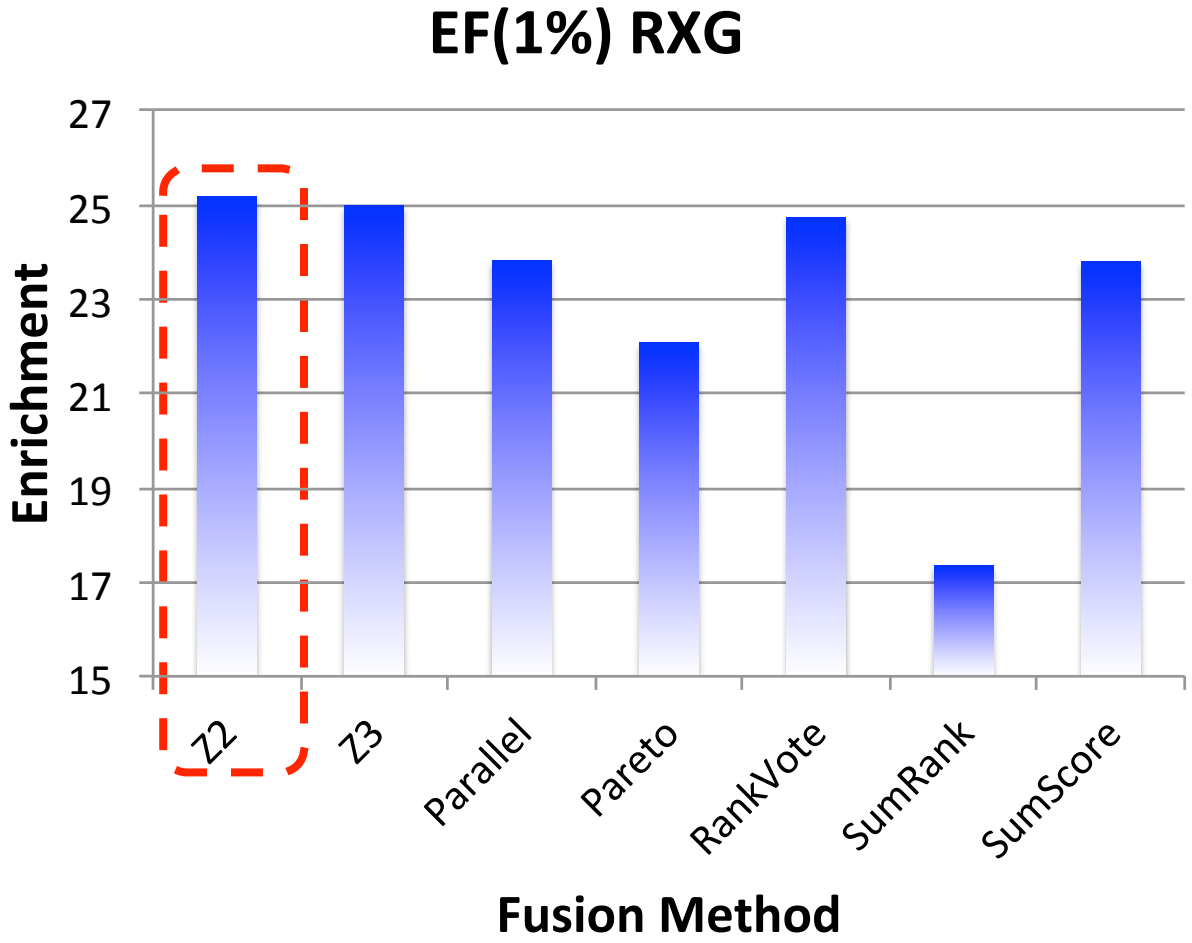
Combining More Scoring Methods

- Combined all 3 FPs, 2 shape screenings, and HTVS docking
- With more scoring methods, more Z-scores should be used



New Results on DUD

- 40 targets
- Well-selected actives and decoys



Conclusions

- Data fusion can improve virtual screening enrichments
- Z-score generally performs better than other fusion approaches
- Including more scoring methods appears to be better
  - Depends on them being “good enough”
- Results are consistent for Glide, MDDR, and DUD sets
- Fully automated workflow is available

Acknowledgements

- Data Fusion
  - Madhavi Sastry
  - V.S. Sandeep Inakollu
- Canvas Fingerprints
  - Madhavi Sastry
  - Steve Dixon
  - Jeff Lowrie
- Shape Screening
  - Madhavi Sastry
  - Steve Dixon