

# What is it with all this shape nonsense anyway?

Paul Hawkins  
Senior Applications Scientist



OpenEye Scientific Software

# A word from the president..

- “Damn it feels good to be a gangster, gettin’ voted into the White House. Everythin’ lookin’ good to the people of the world...
  - But I’m soon going to put a stop to that.



# Virtual screening

- Finding signal in the noise.
  - Using shape similarity
- How to measure signal:noise?
  - In aggregate
  - Individually
- How does sampling of conformer space affect signal to noise ratio?
  - Need a dataset



# What dataset to use?

- Make up our own
  - Results have no baseline
  - Ridiculously easy or hard dataset
    - No dynamic range
    - 100's to 1,000's of targets for robust results
- Use one from the literature
  - Baseline comparison
  - Dataset may very hard or very easy
    - Guidance from the published results



# DUD – “Not just crap sucked out of the ACD.”

- Huang *et al.*, *J. Med. Chem.*, **2006**, *49*, 6789.
  - [dud.docking.org](http://dud.docking.org)
- **Forty targets; actives and decoys.**
  - 38 exptl. co-crystal structures
  - Large number of test cases reduces variance
- **Property matched, diverse decoys.**
  - HBA, HBD, MW, logP, RB



# Virtual screening

- Run ROCS over each of the 38 db's
  - Xtal ligand as query
- Two query conformations of xtal ligand
  - X-ray conformation
  - Lowest energy OMEGA conformation
- Use shape + chemistry as ranking metric

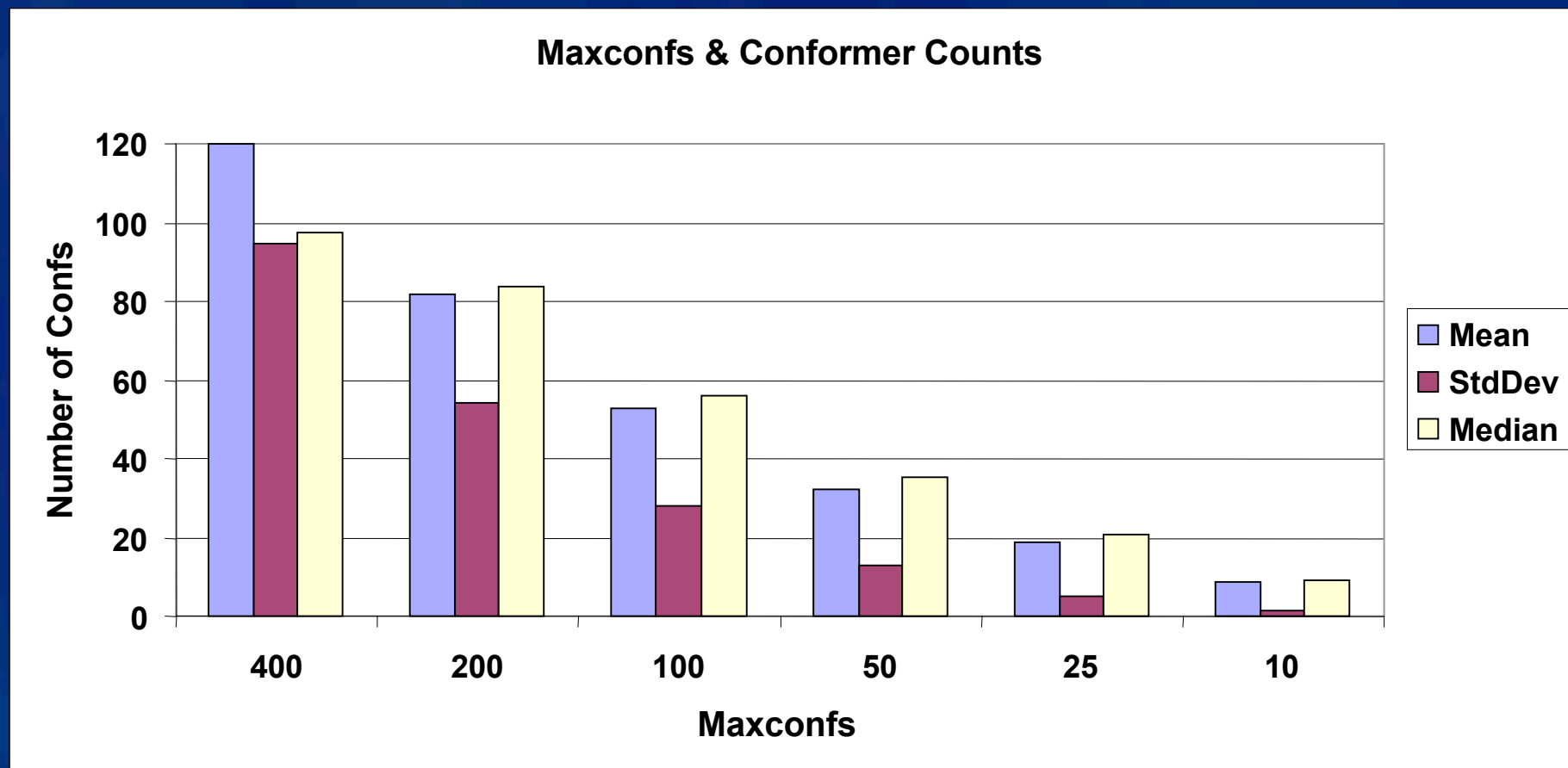


# OMEGA2 Parameters

- Maxconfs
  - 400, 200, 100, 50, 25, 10, 1
- Ewindow
- RMSD

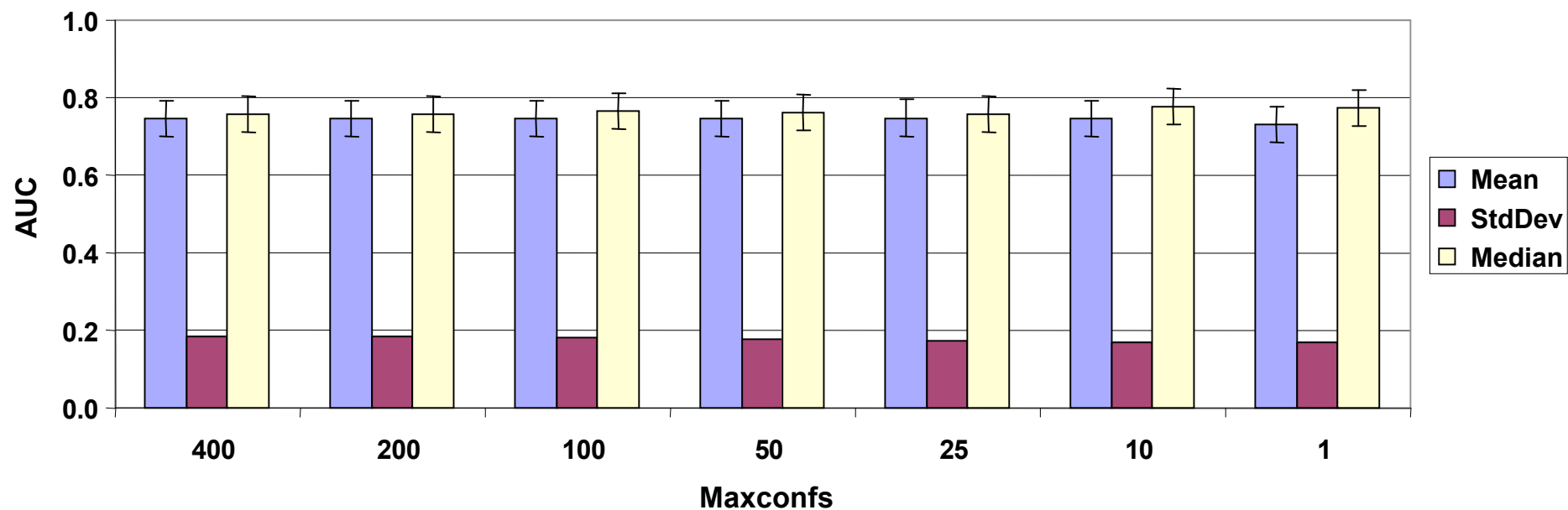


# Effect of Maxconfs on conformer #

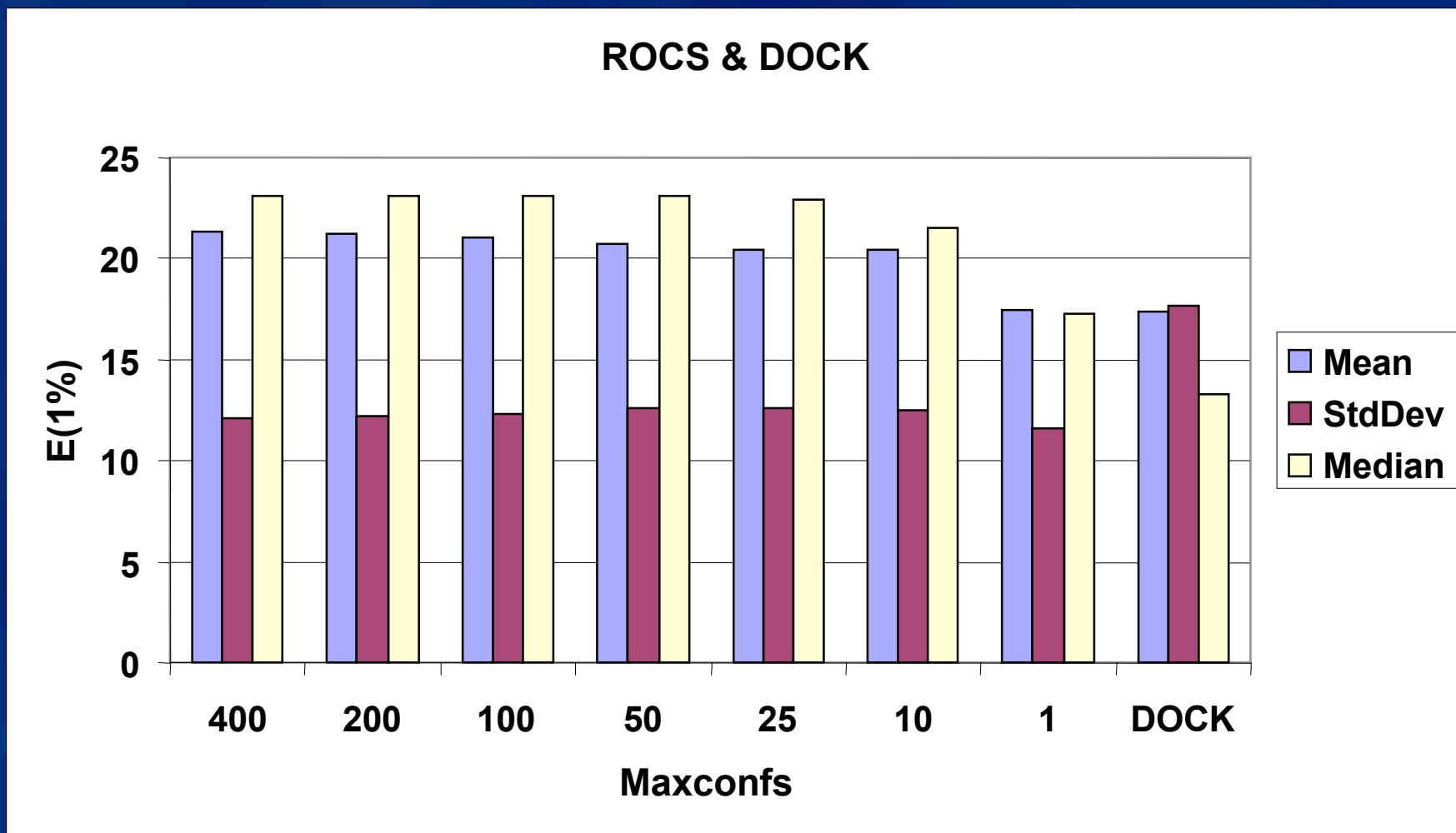


# AUC: X-ray query

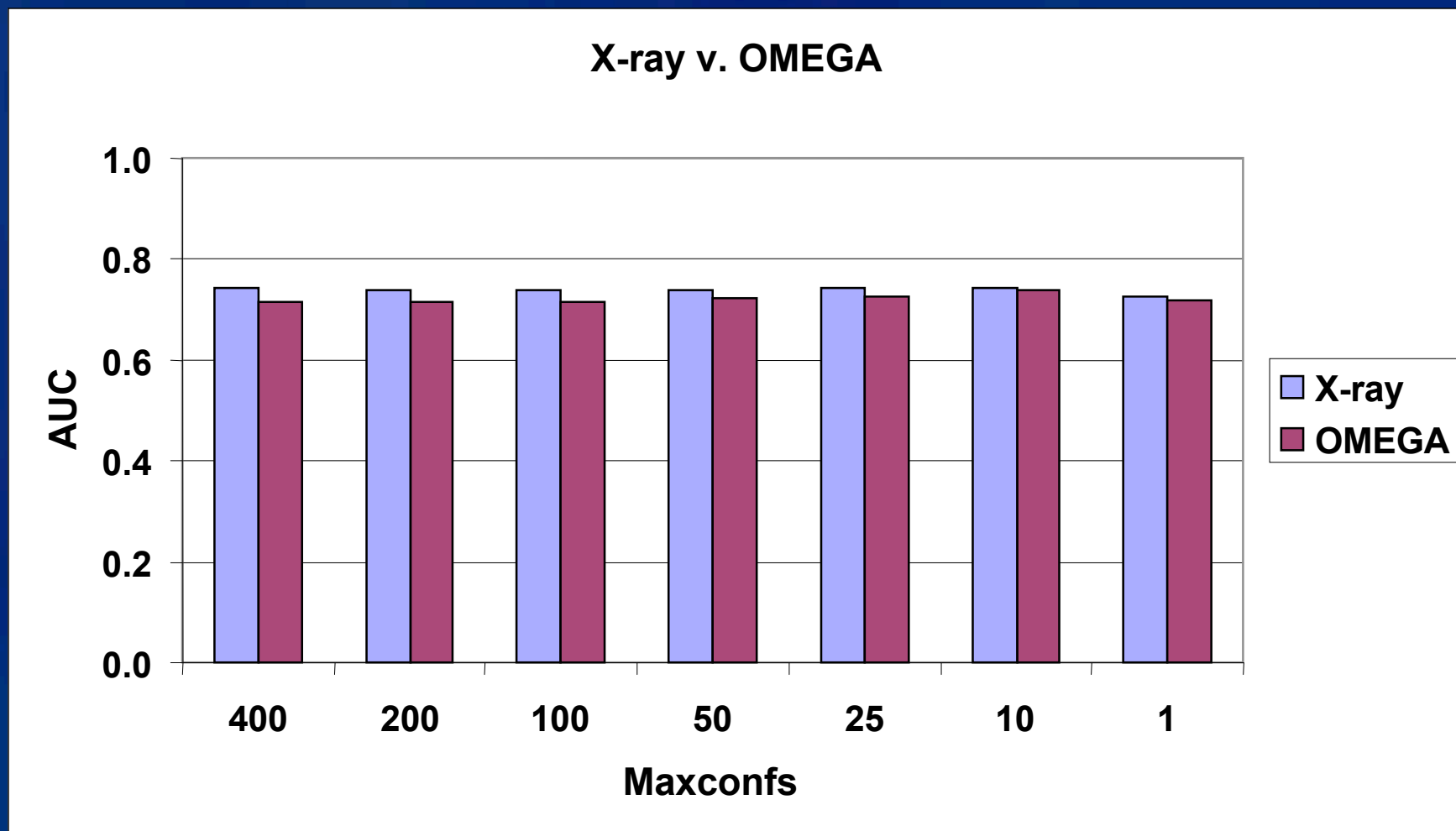
X-ray Conformer



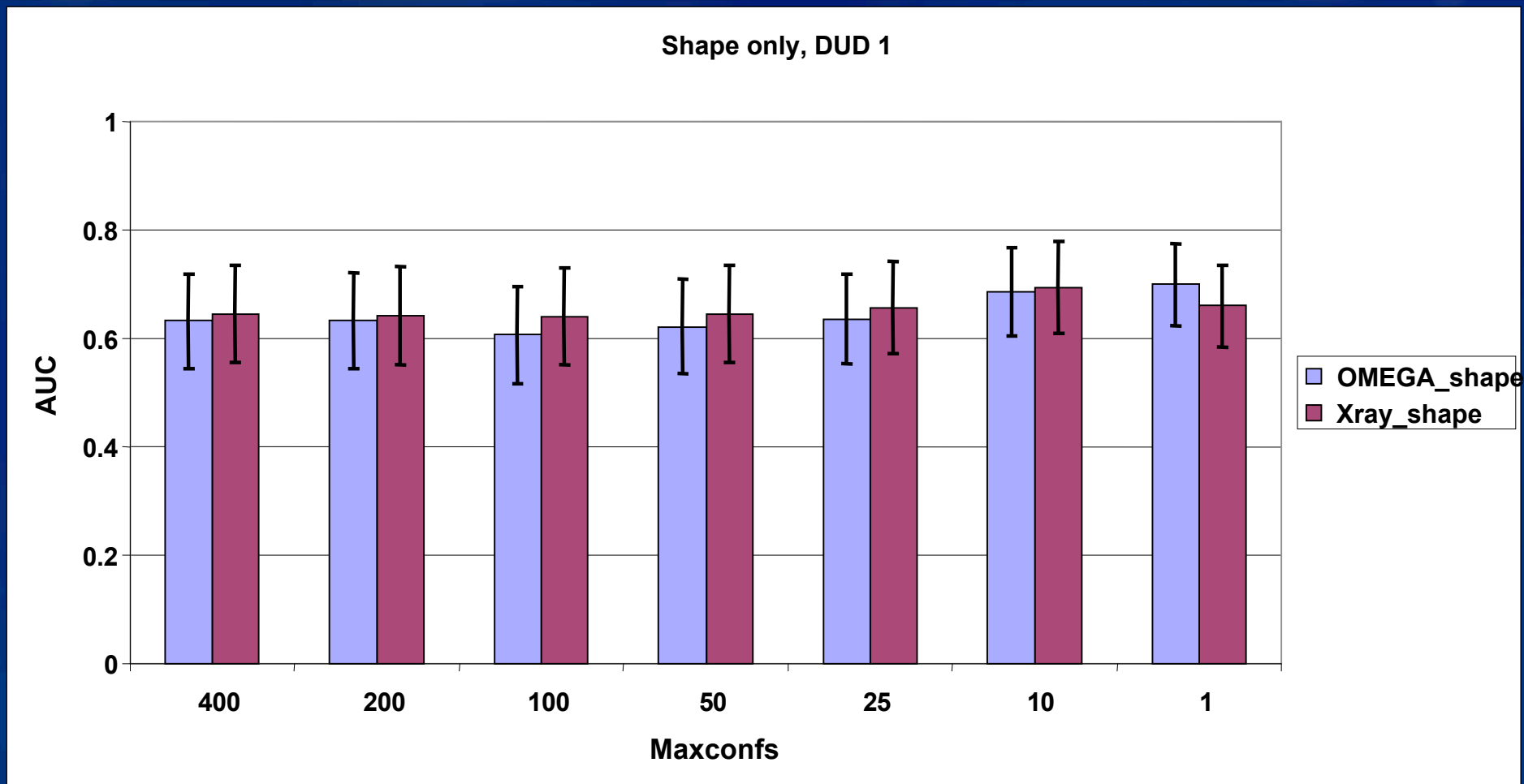
# Enrichment: X-ray query



# Is "the" bioactive conformer special?

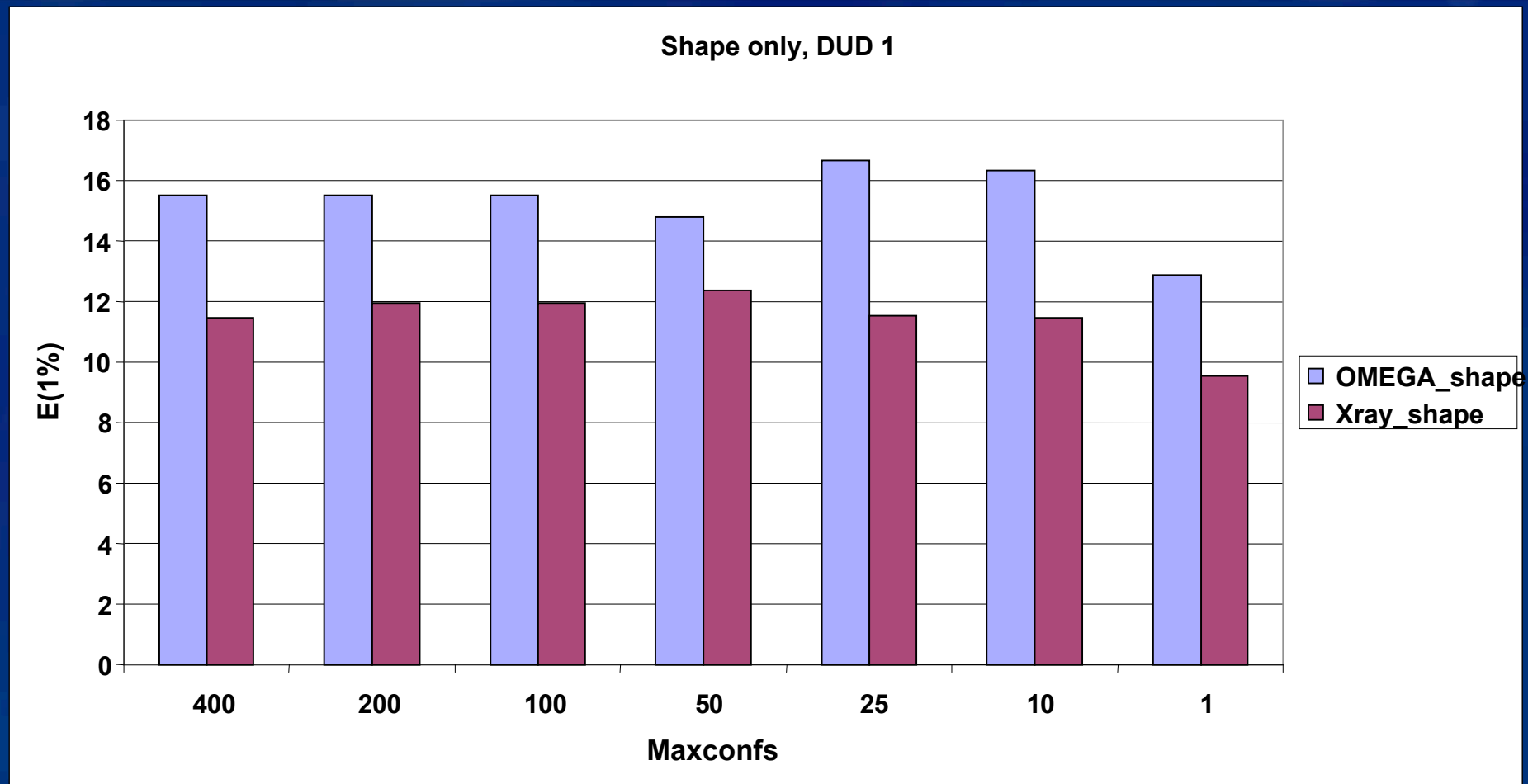


# "Simplify, simplify."



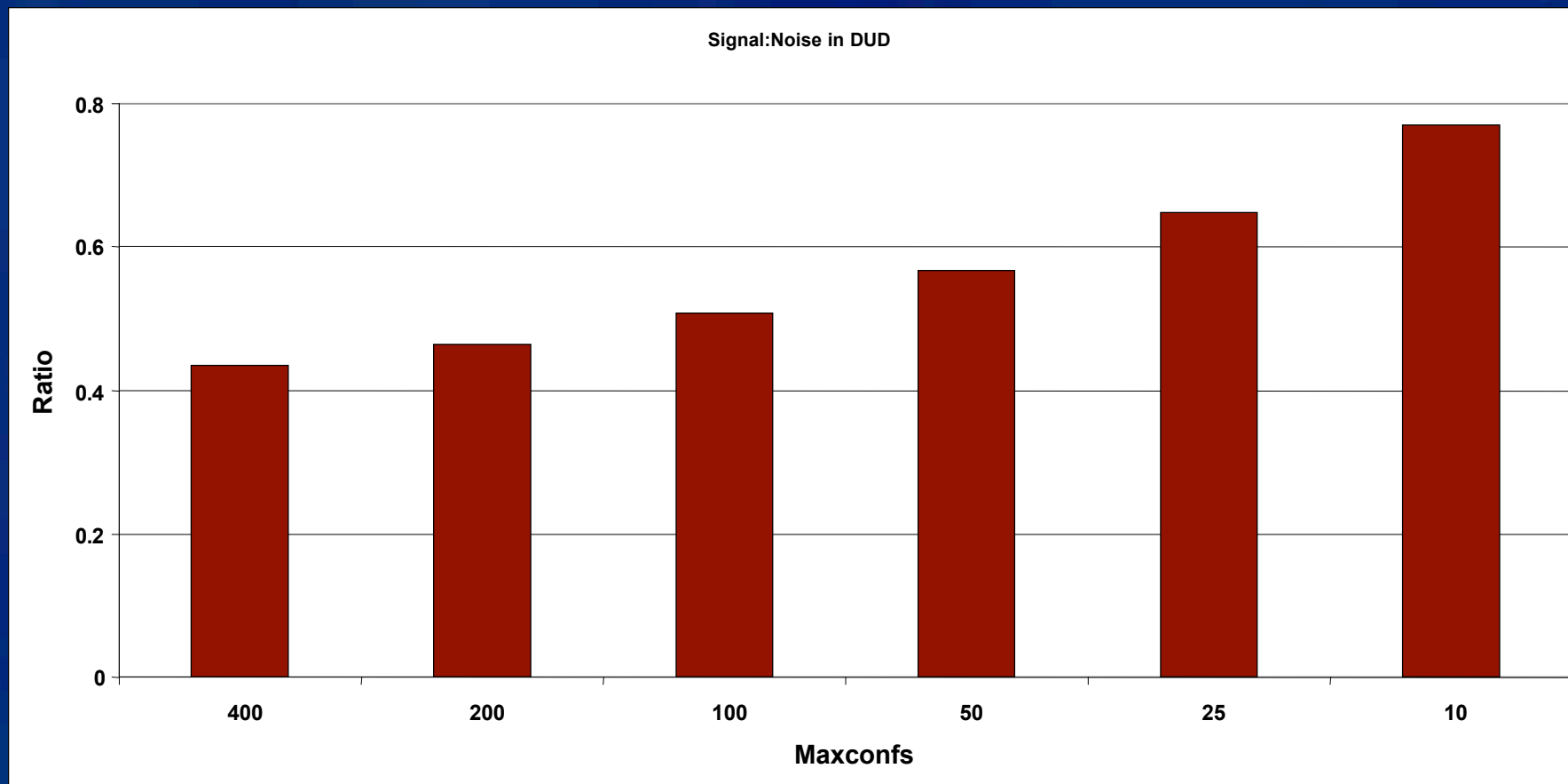
Median AUC when ranking by ST only

# Not simple enough?



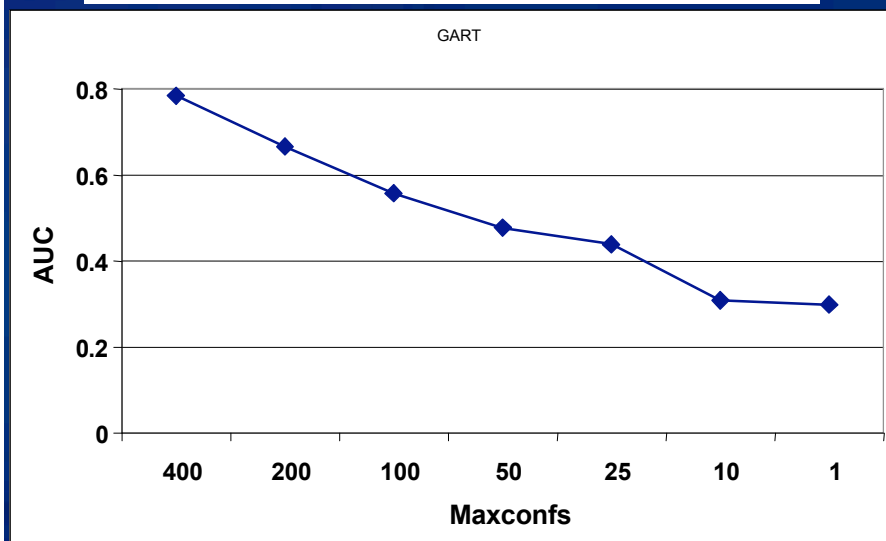
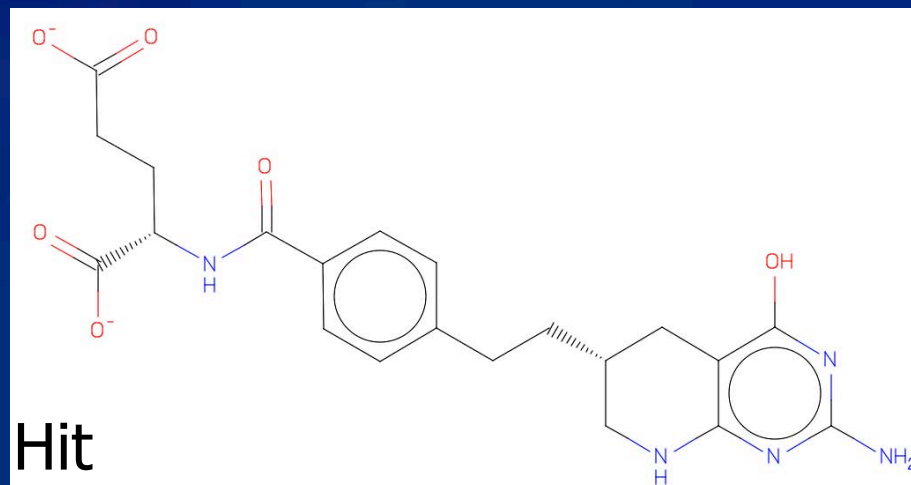
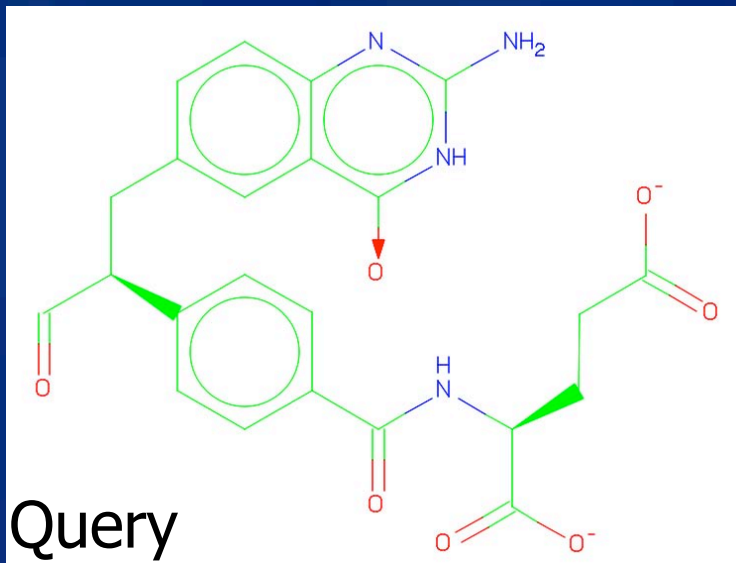
Enrichment at 1% when ranking by ST only.

# Is there more active "signal"?



Ratio of conformers per active to conformers per decoy

# GART & shape



**Maxconfs**

400

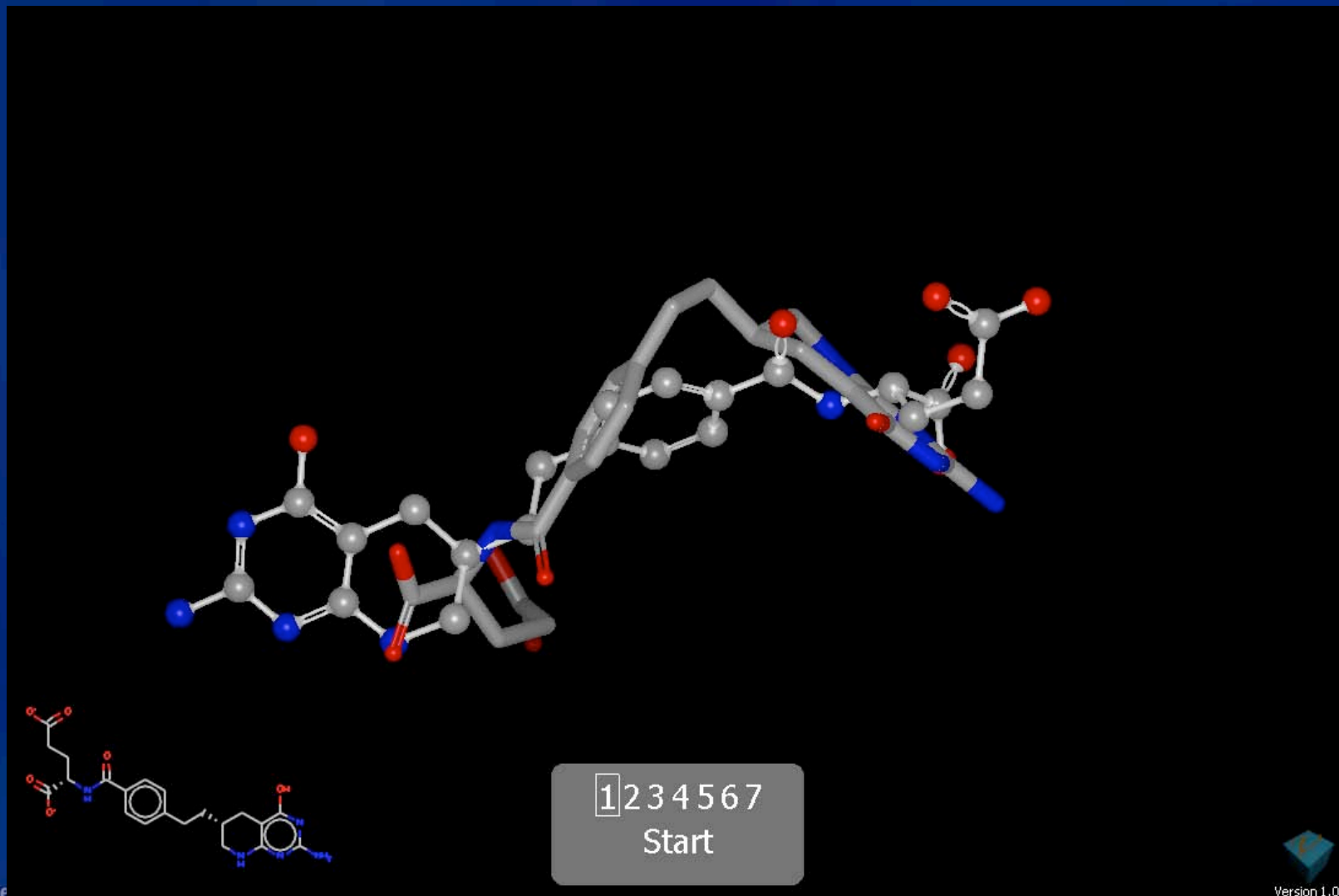
1

**Rank**

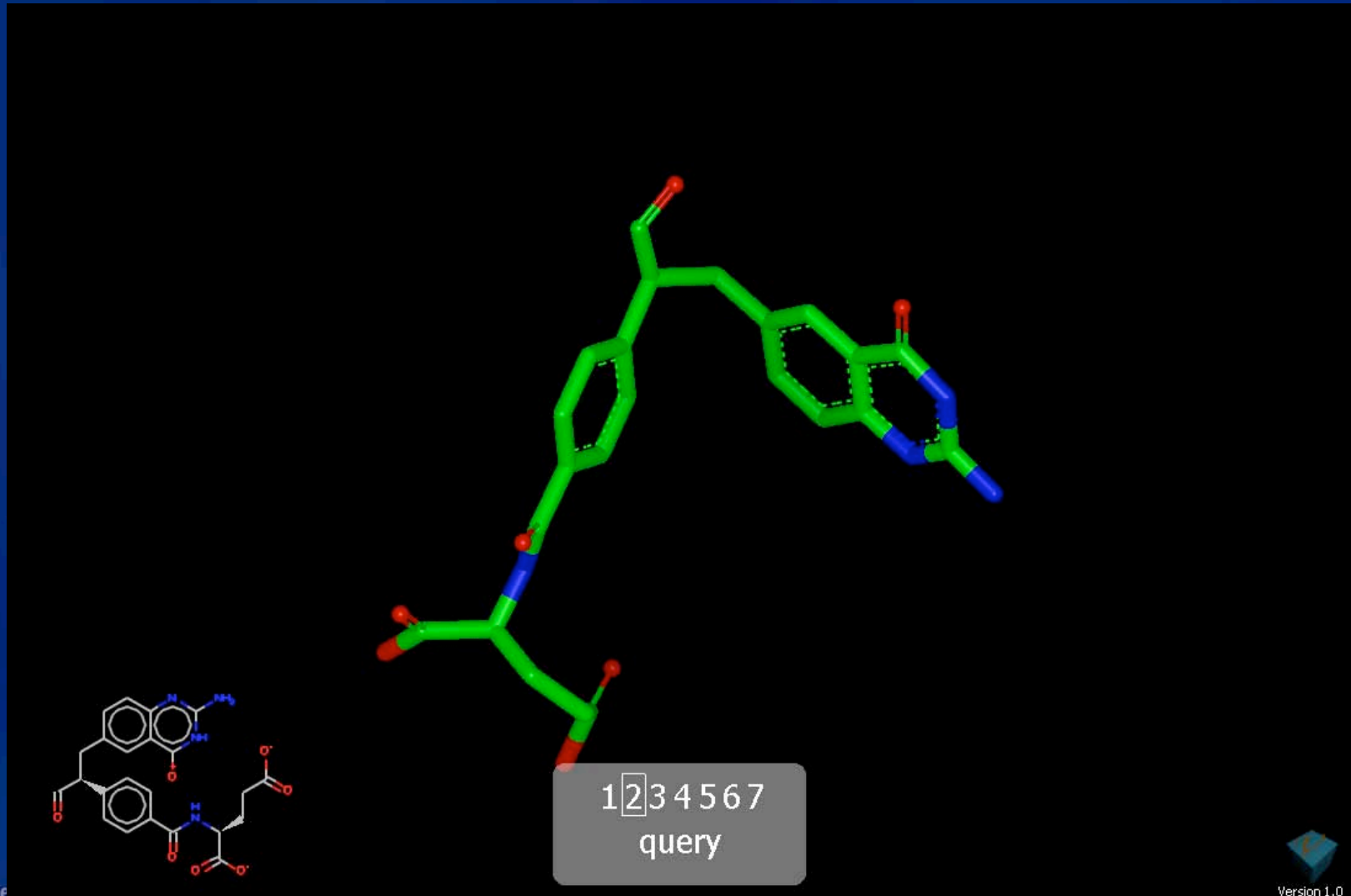
23

720

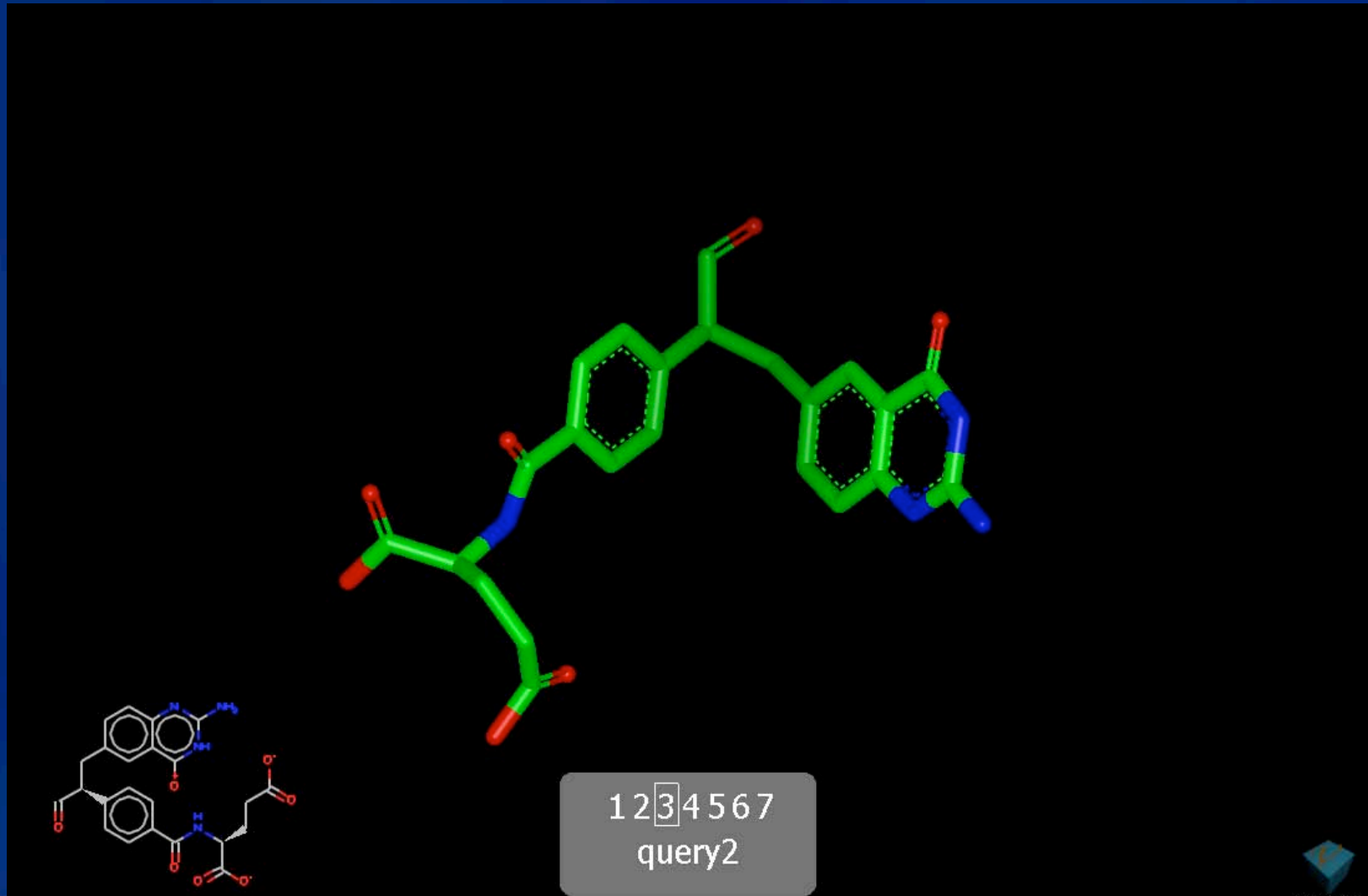
# What we see



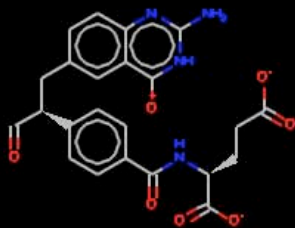
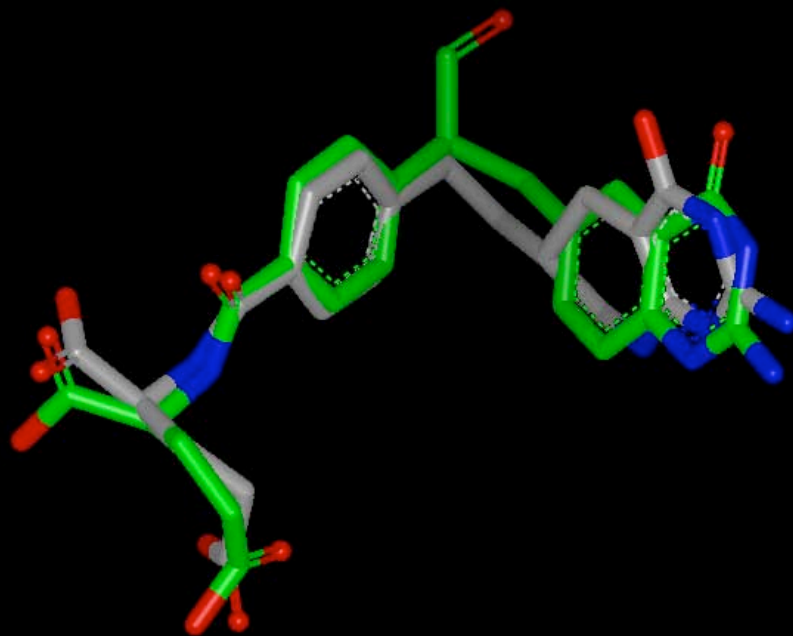
# What we see



# What we see



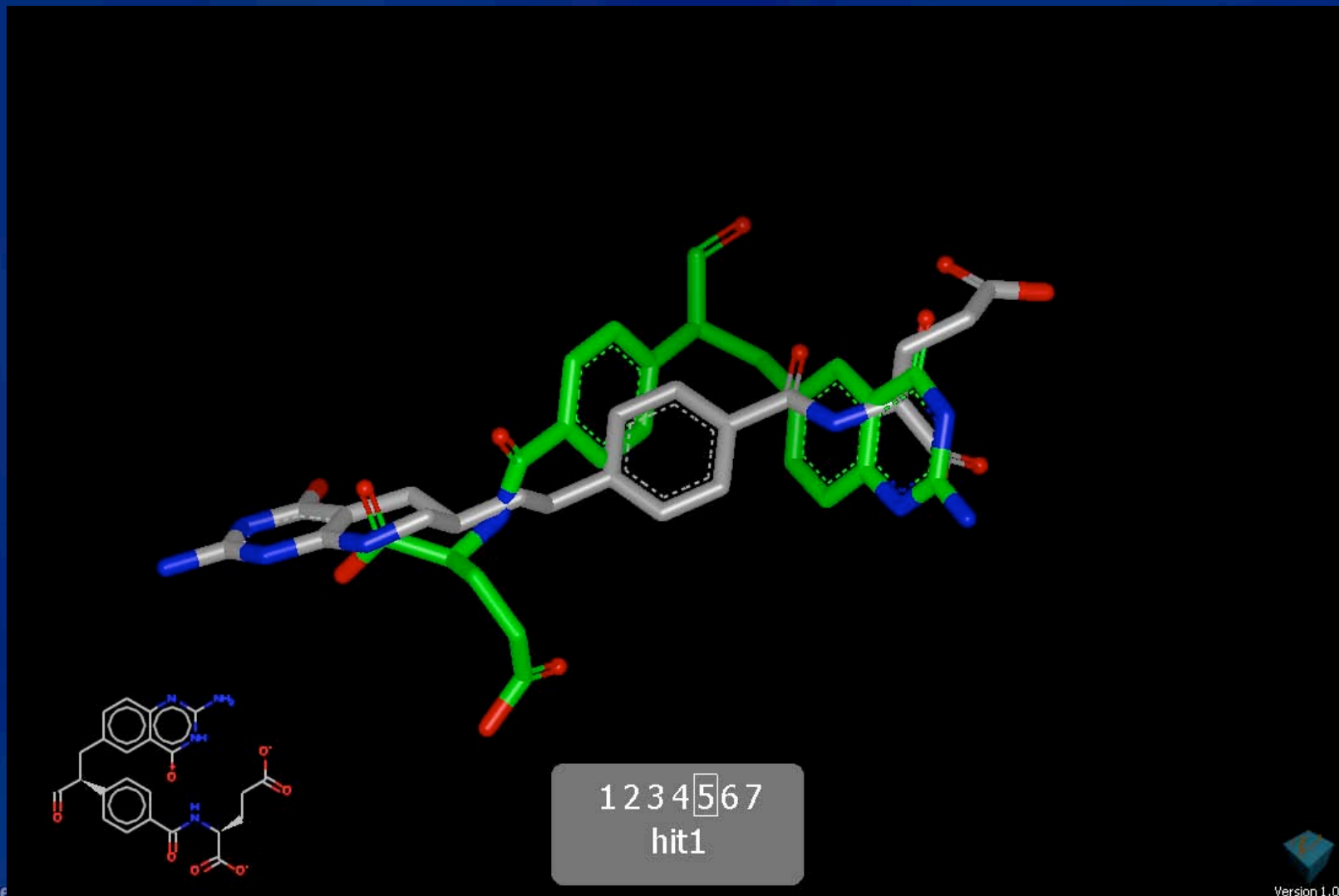
# What we see



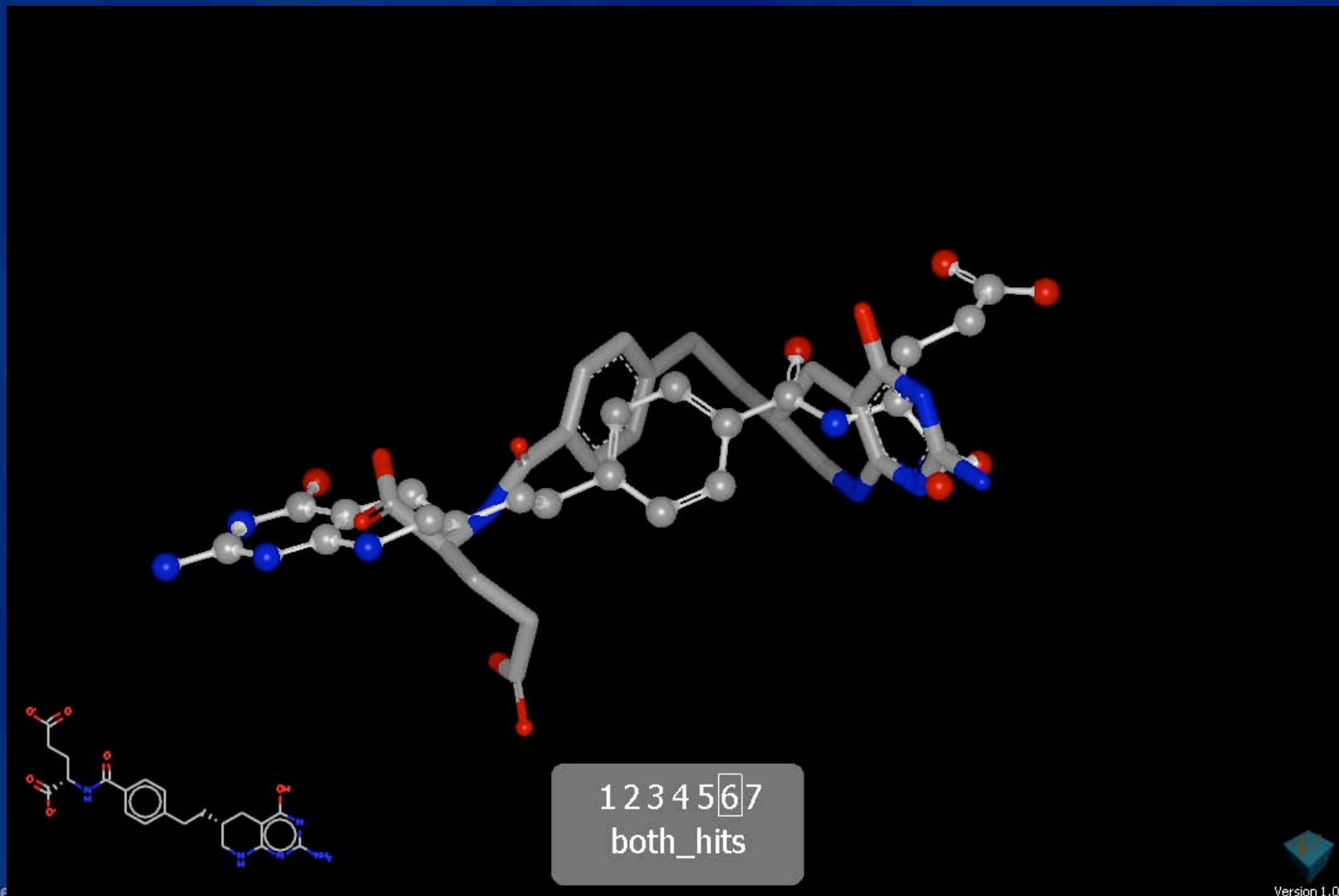
1234567  
hit400



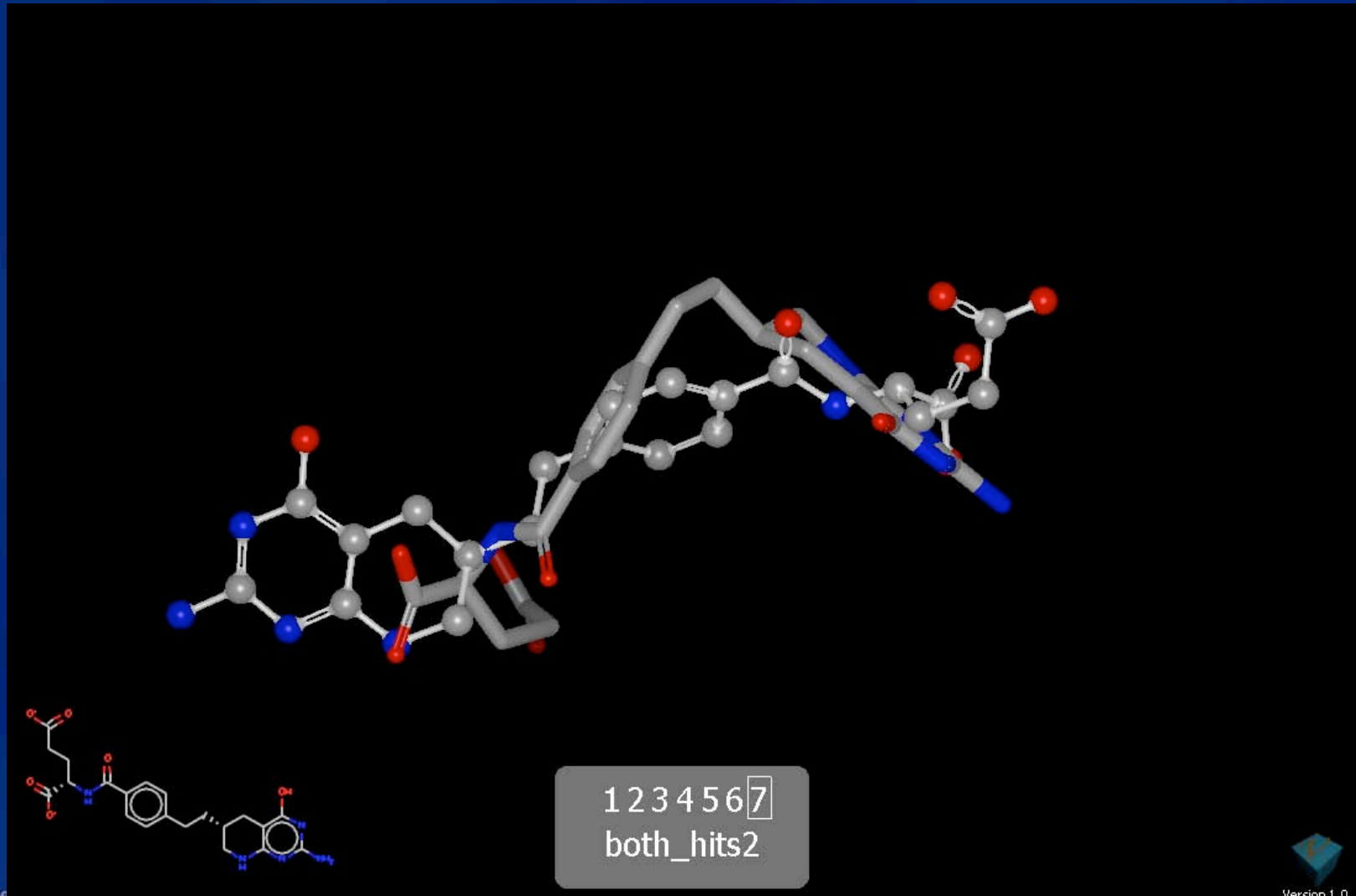
# What we see



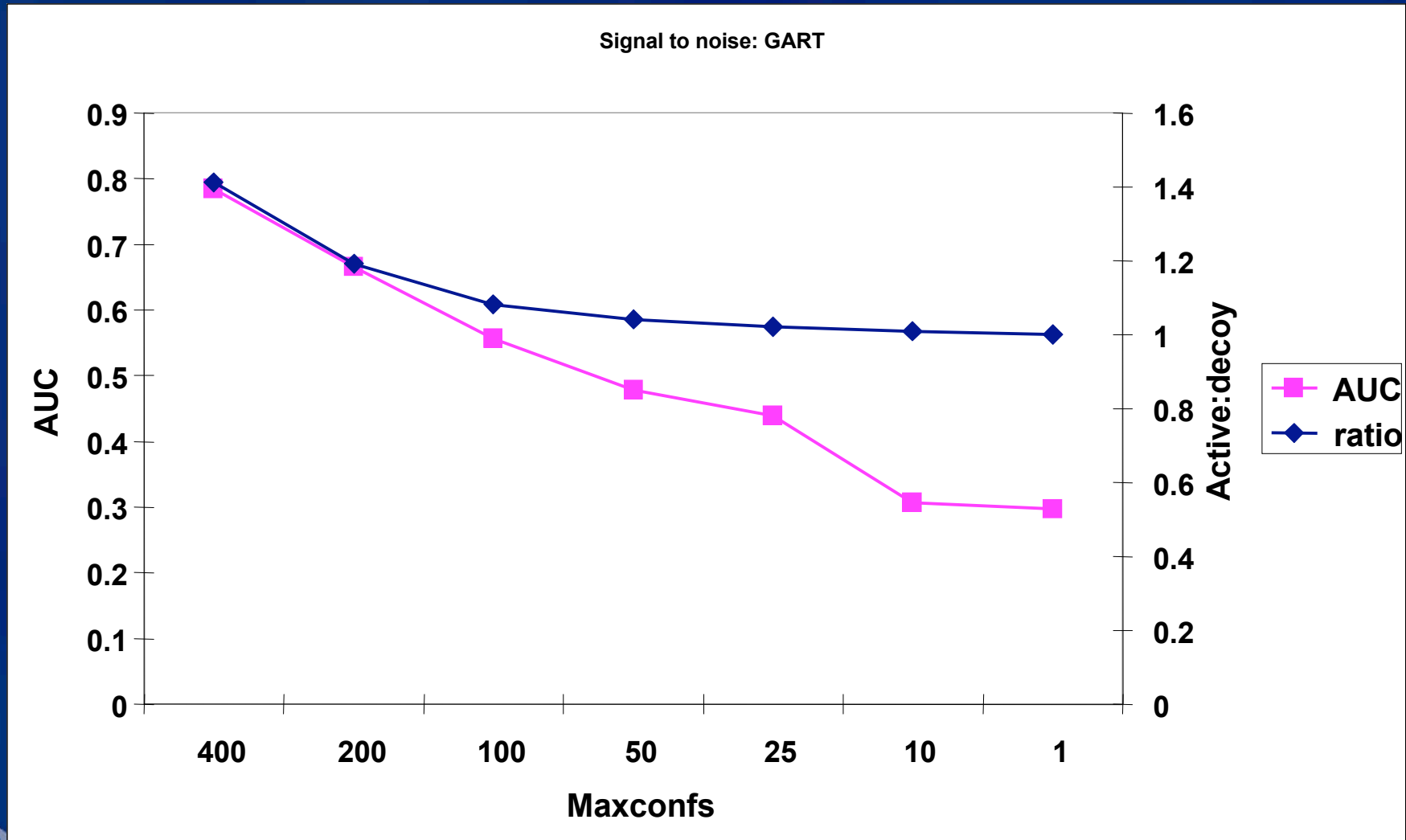
# What we see



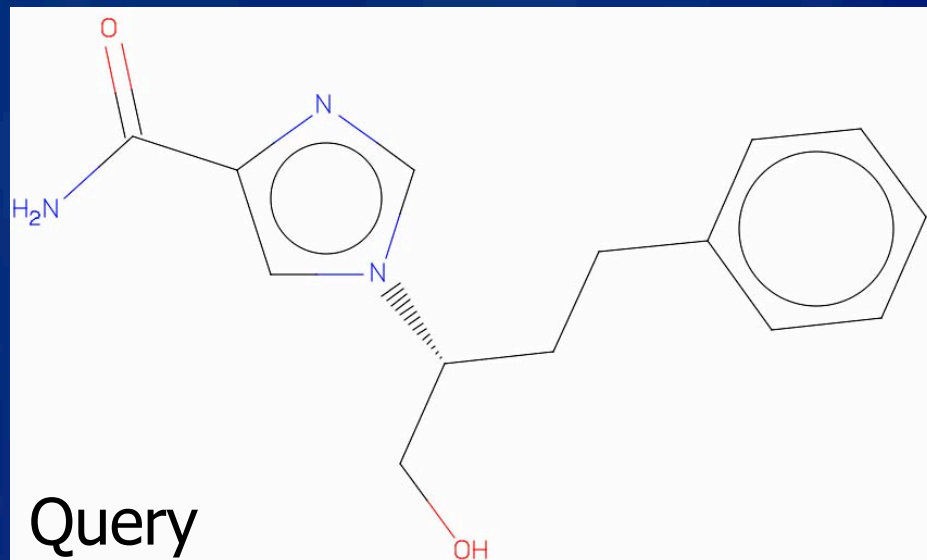
# What we see



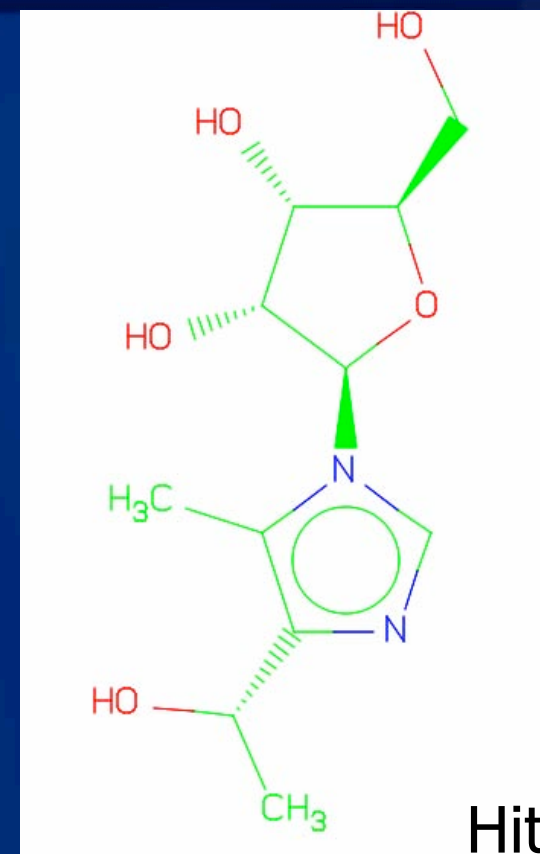
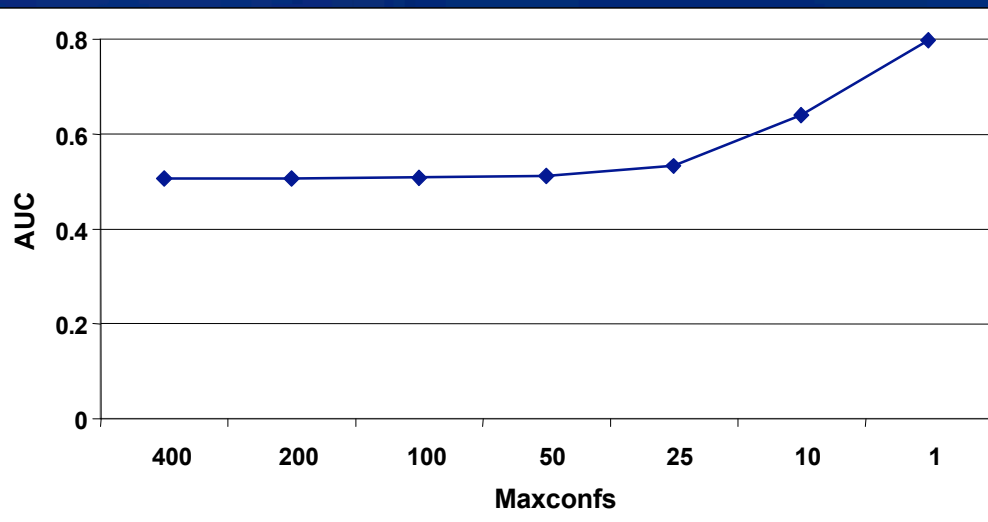
# Conformer sampling



# ADA & shape



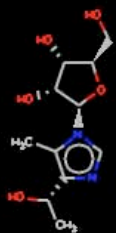
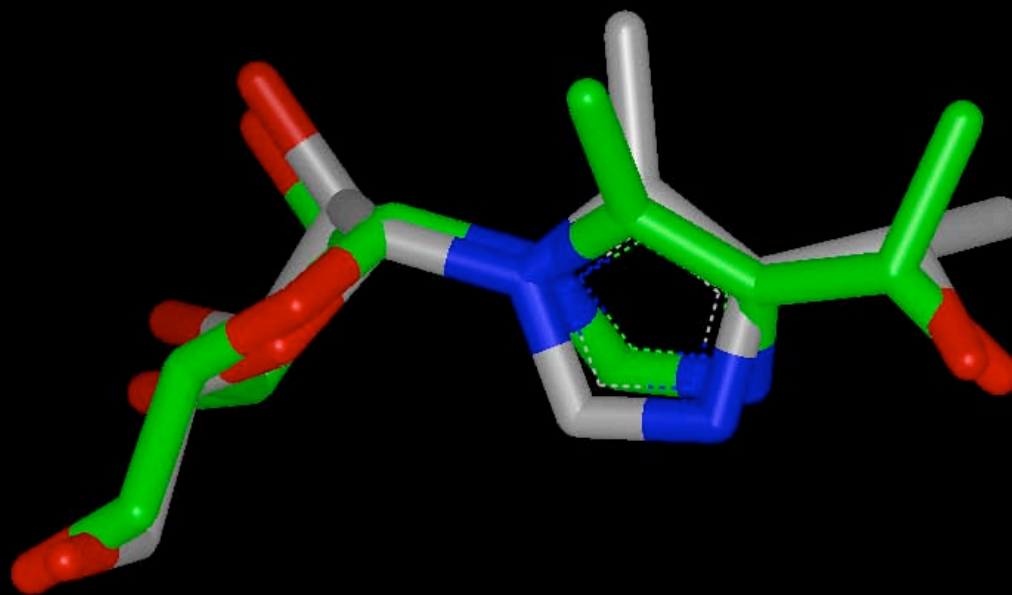
Query



Hit

Maxconfs	Rank
400	440
1	92

# What we see

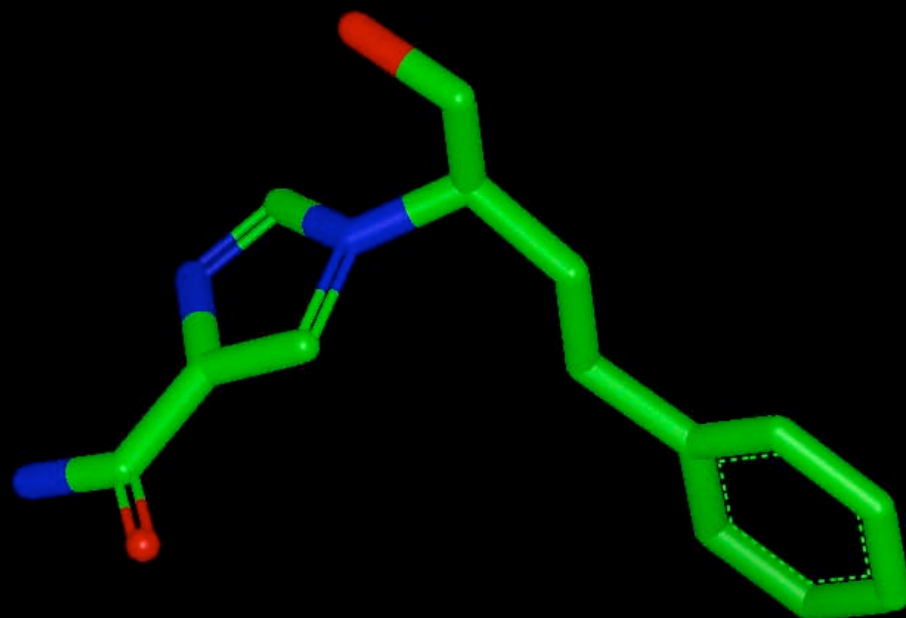


1 2 3 4 5 6  
Start



Version 1.0

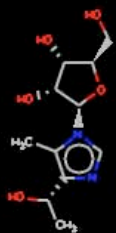
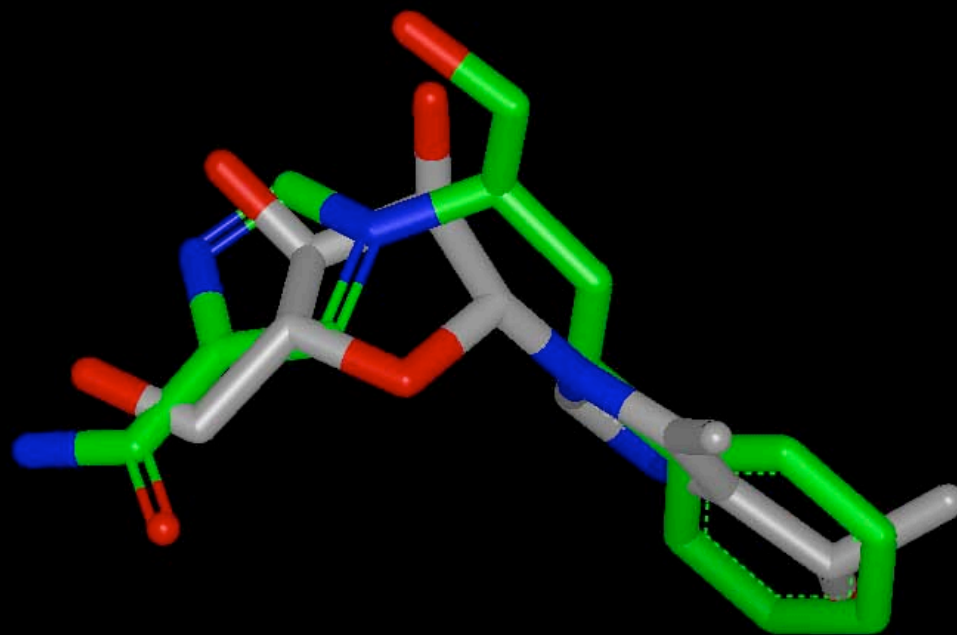
# What we see



1 2 3 4 5 6  
query



# What we see

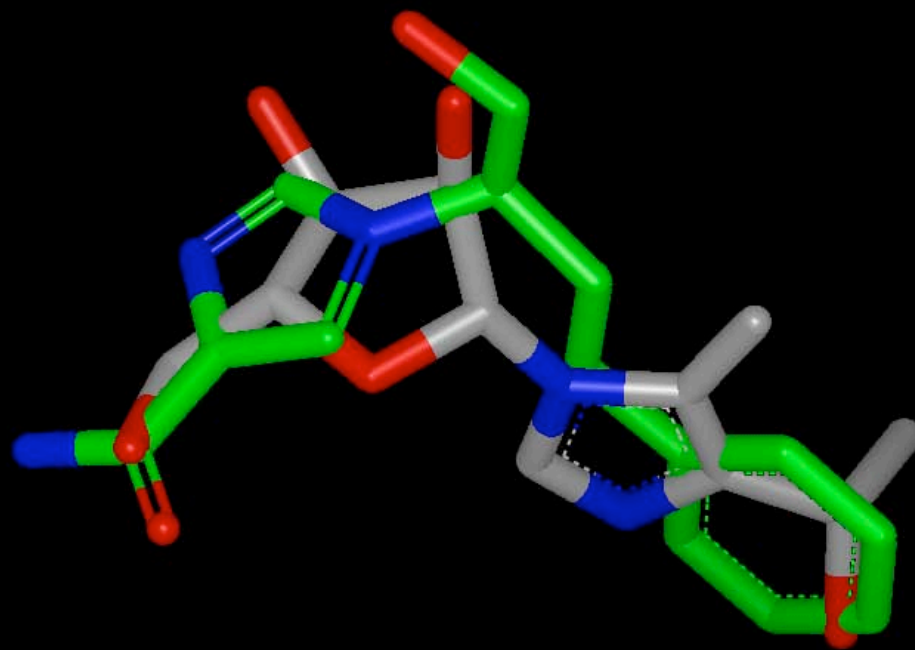


123456  
overlay\_400



Version 1.0

# What we see



1 2 3 4 5 6  
overlay\_1

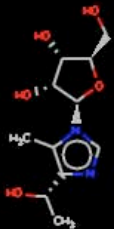
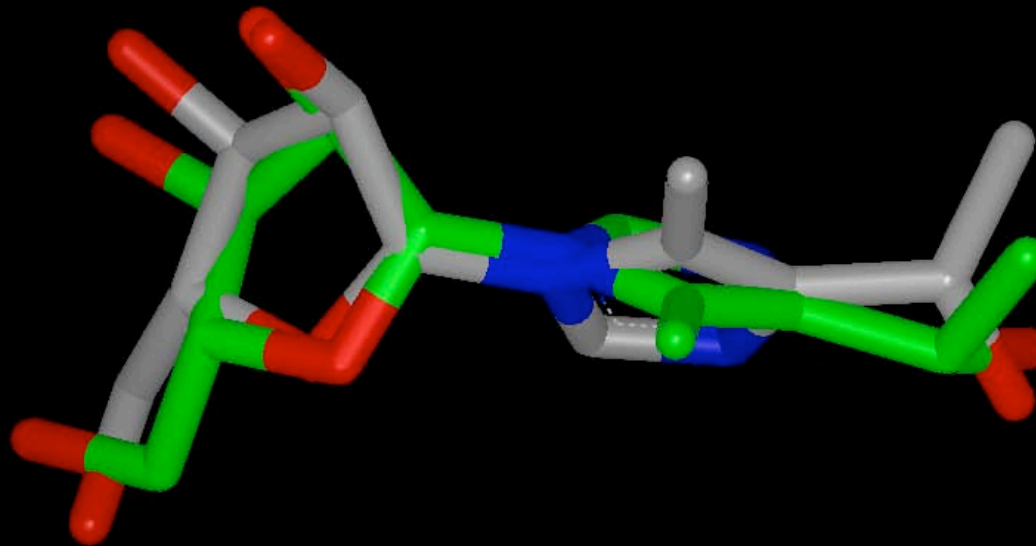


Version 1.0

# What we see

ZINC03814312ada\_act\_8

Green = maxconfs 400

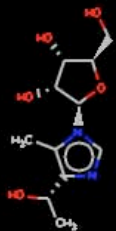
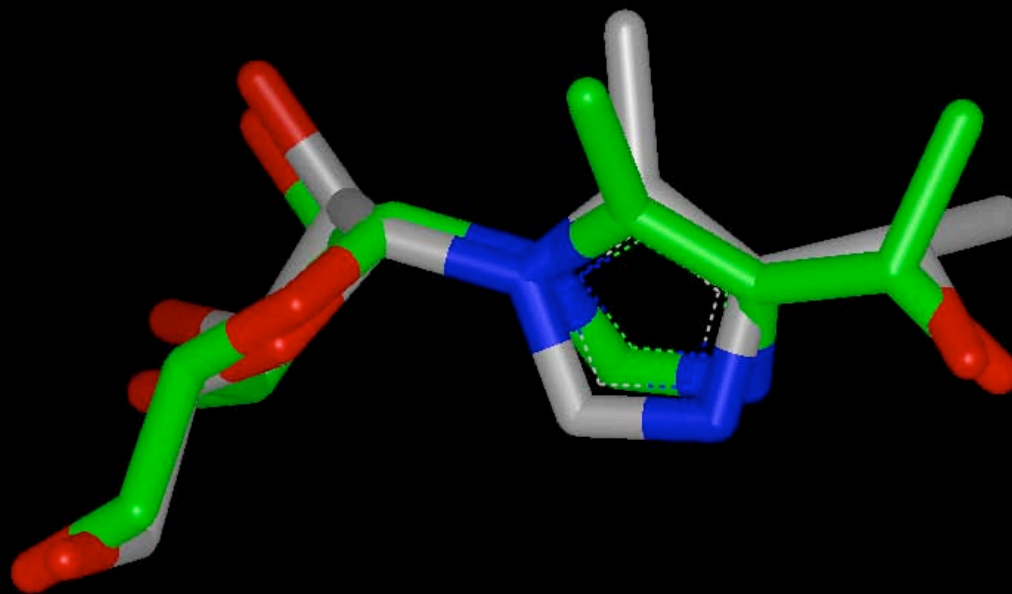


1 2 3 4 5 6  
hits



Version 1.0

# What we see

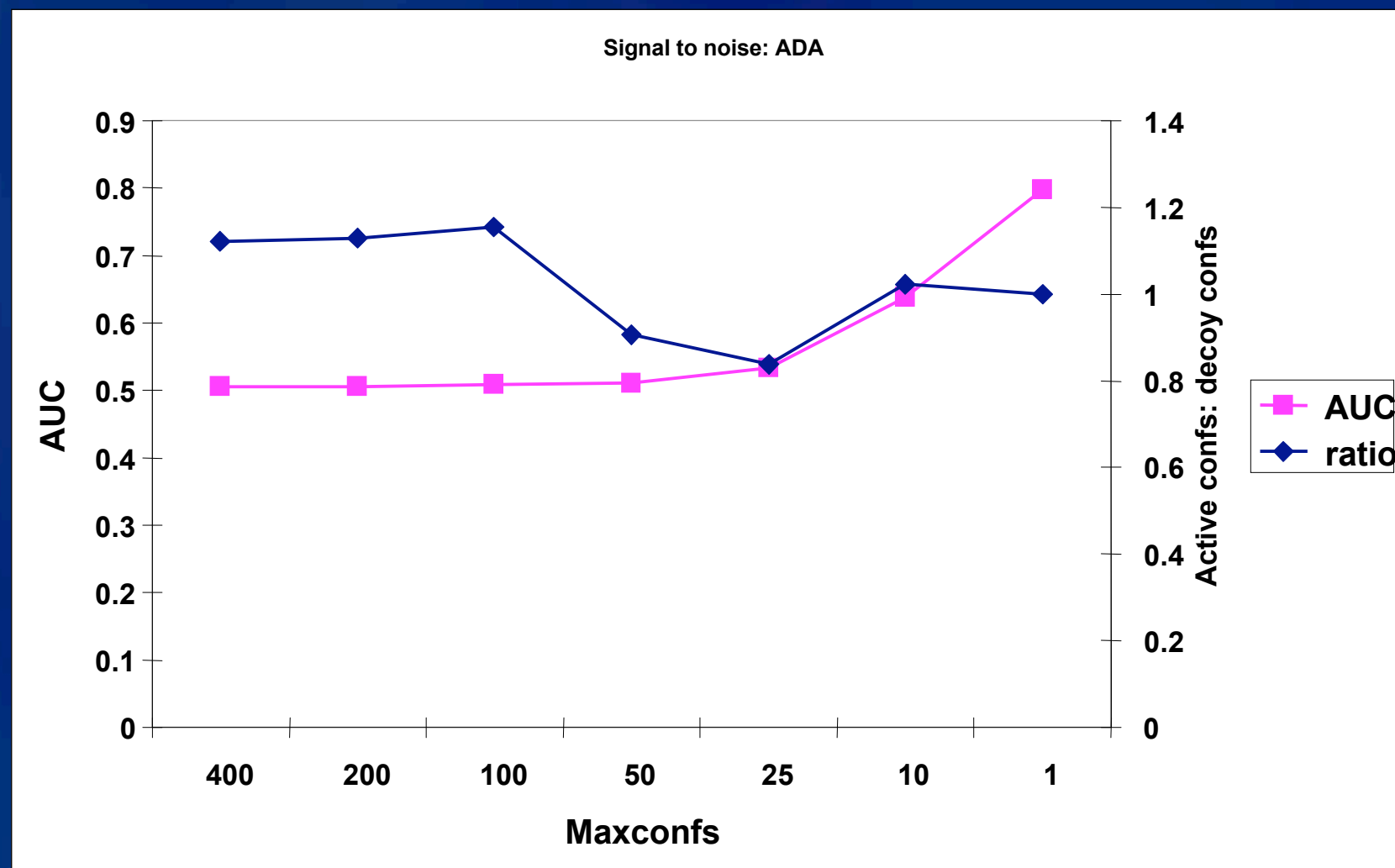


1 2 3 4 5 6  
hits1

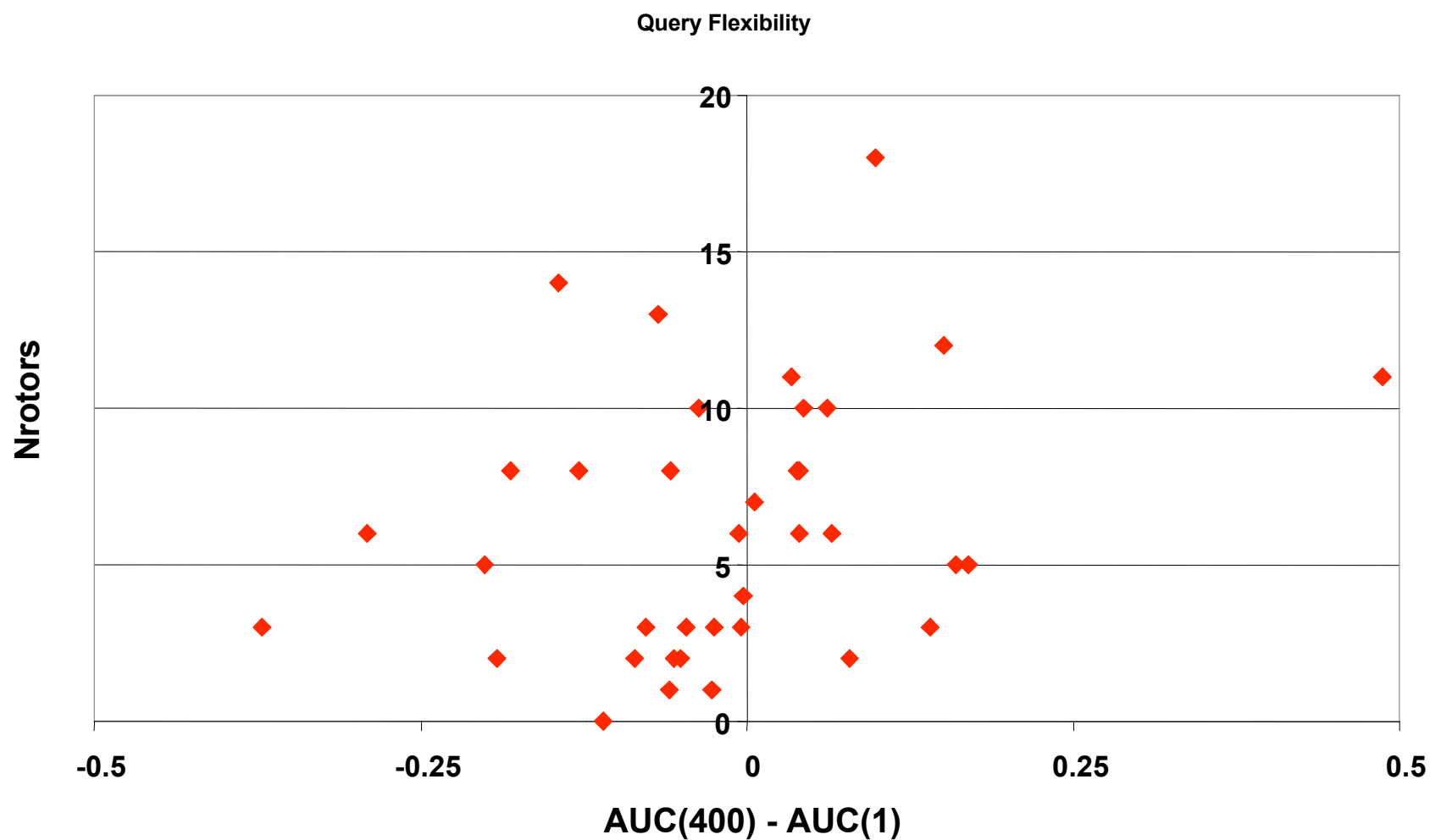


Version 1.0

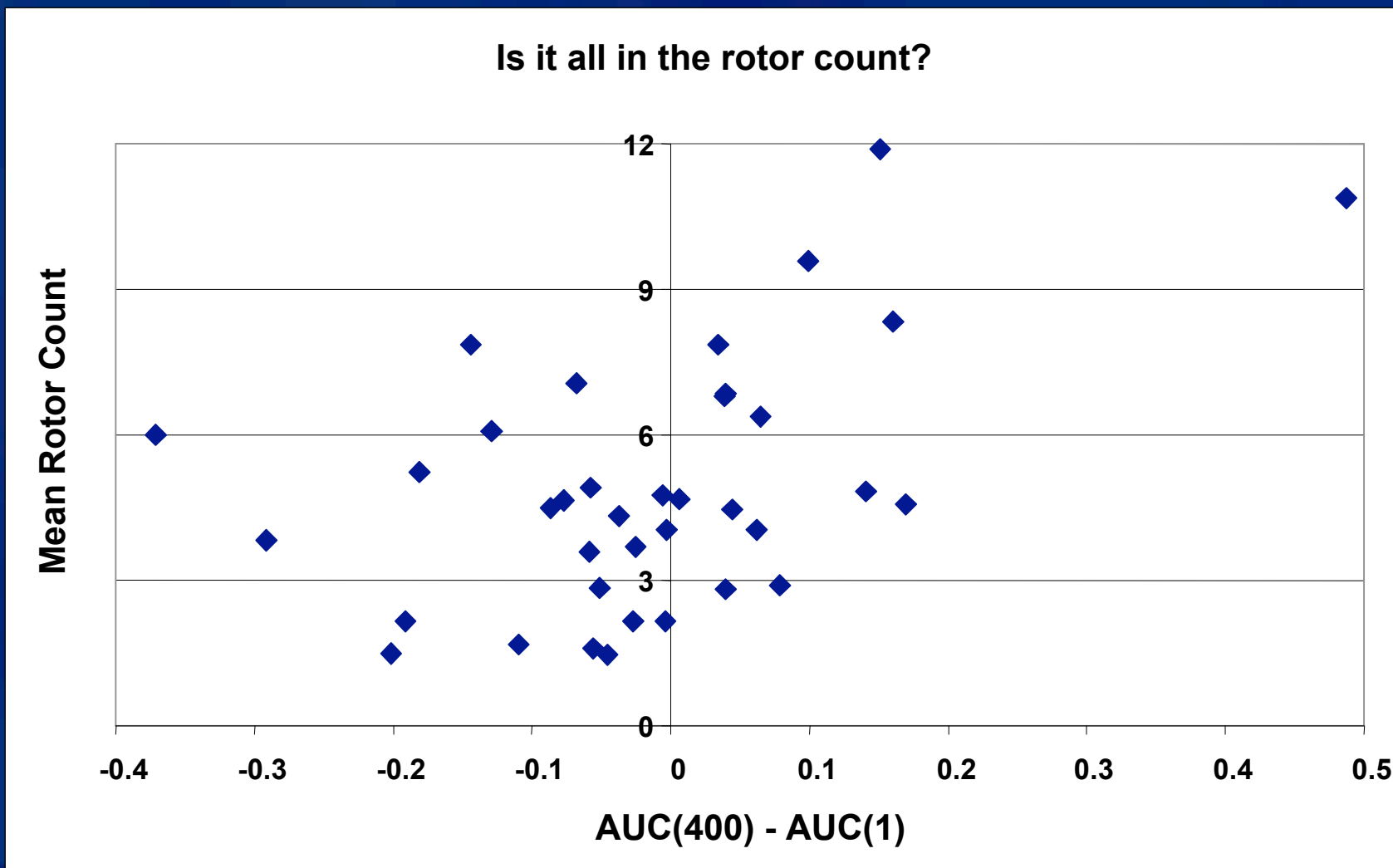
# Conformer sampling



# Rotor count for the query



# Mean Rotors per active

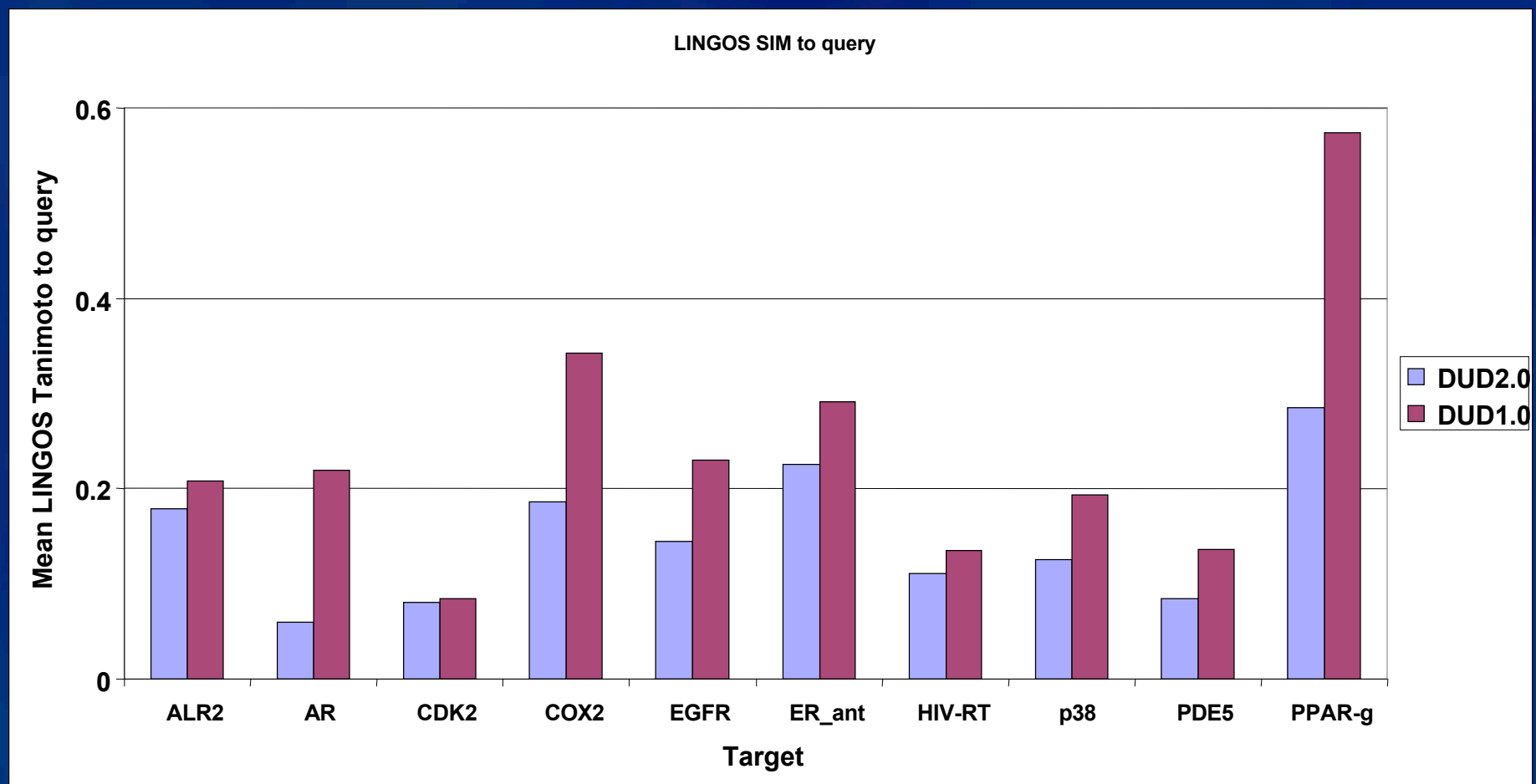


# Trouble at t'mill?

- DUD 1.0 has many close analogues in the active set
- Andrew Good developed DUD 2.0 to address this “analogue bias”
  - Filter for lead-like compounds
  - Augment with actives from WOMBAT
  - Cluster based on reduced graphs
    - Only 1 cluster representative
  - Only 11 targets overlap with DUD 1.0

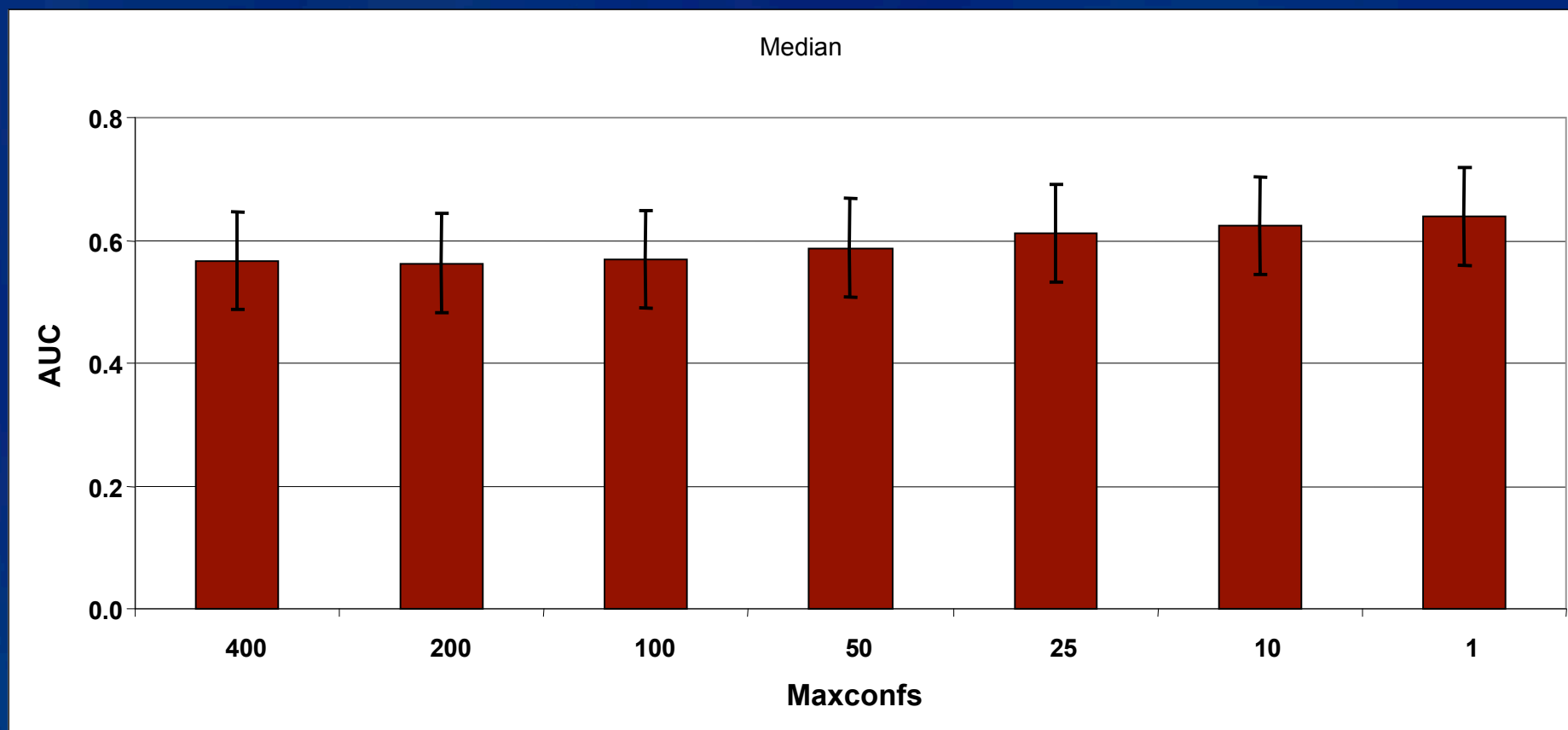


# No baseline for DUD 2.0



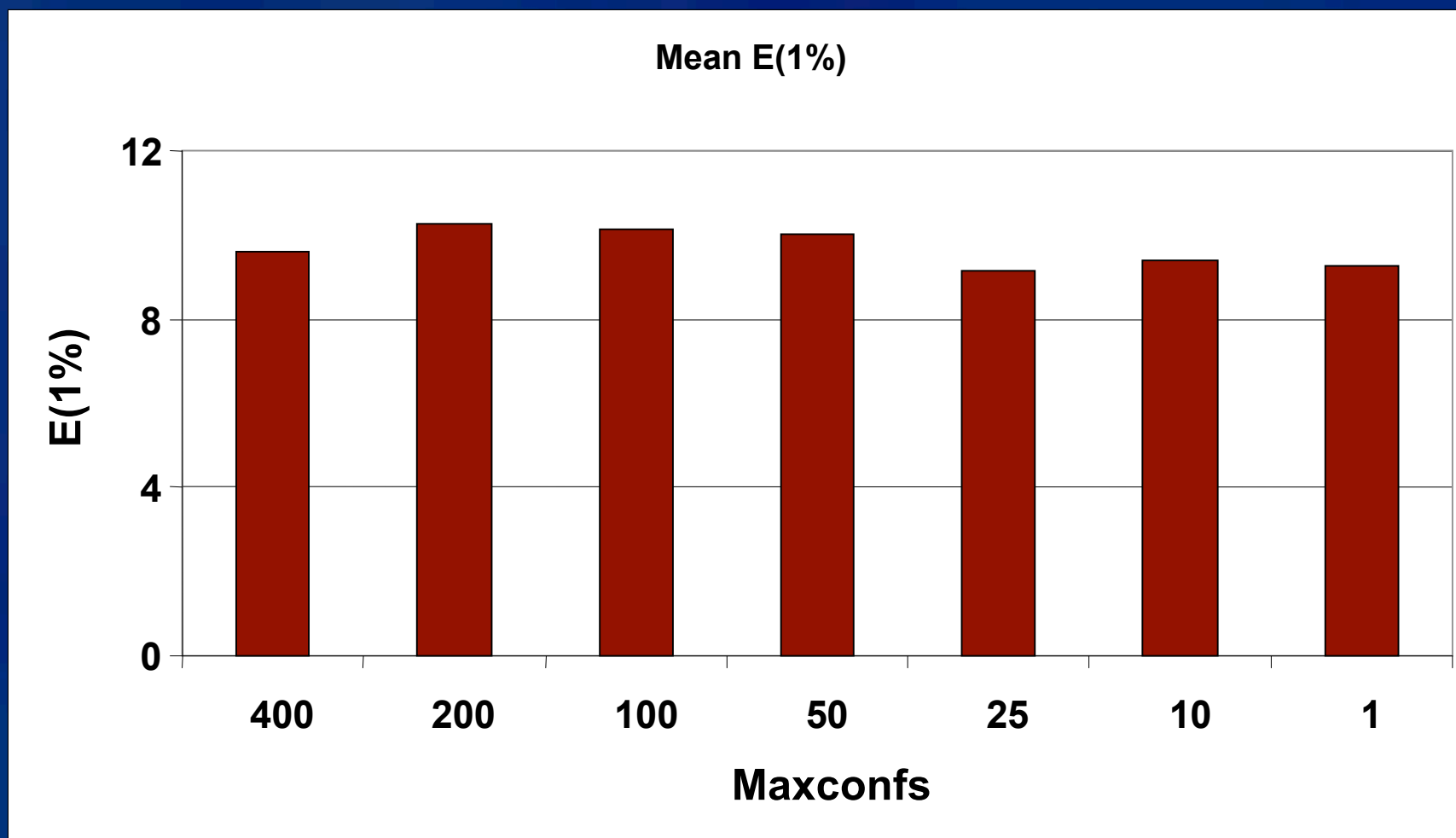
DUD 2.0 is no easier than DUD 1.0,  $p < 0.005$

# DUD 2.0/WOMBAT: AUC

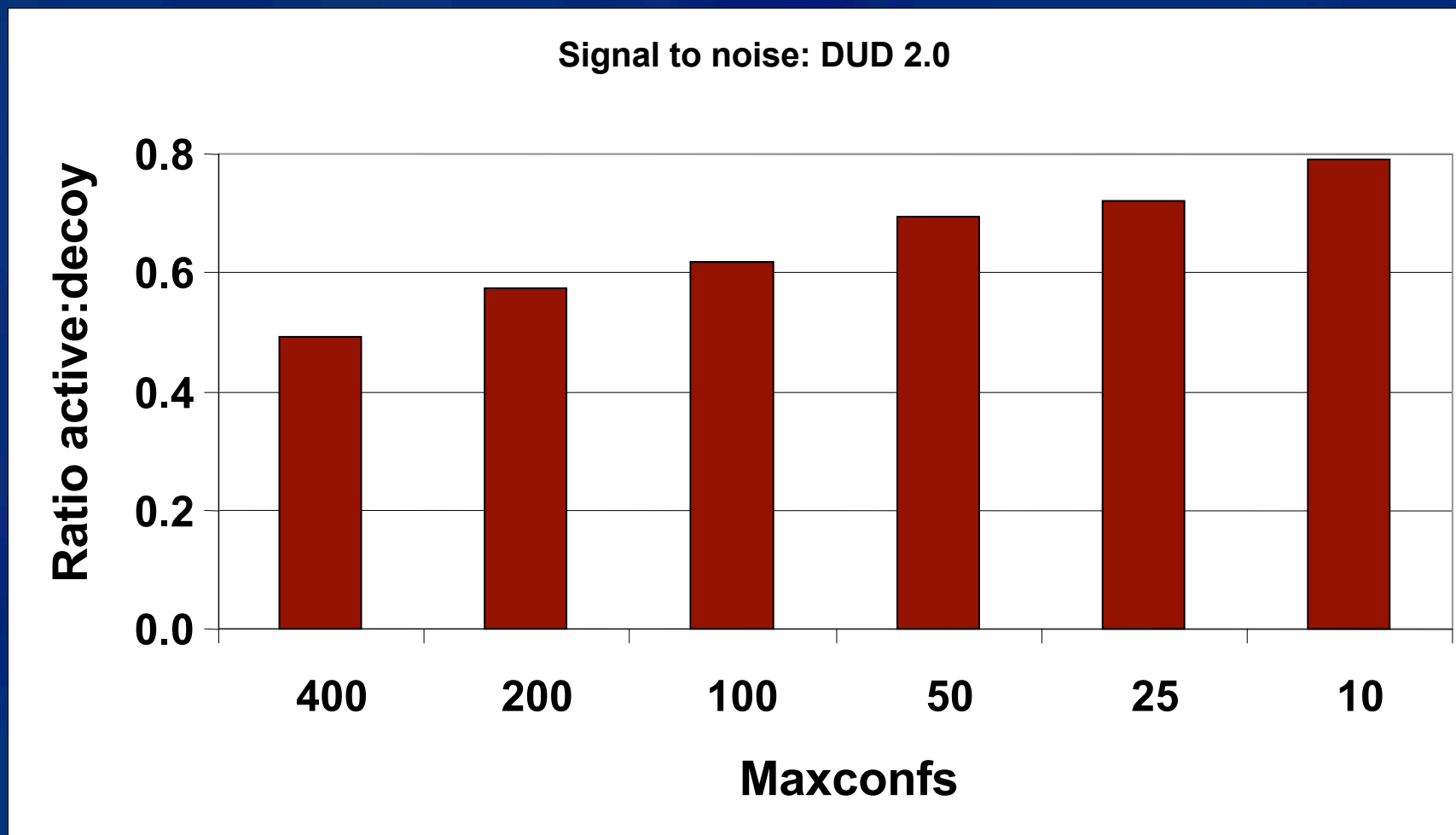


Median across 11 targets

# DUD 2.0: Enrichment



# Is there more active "signal"



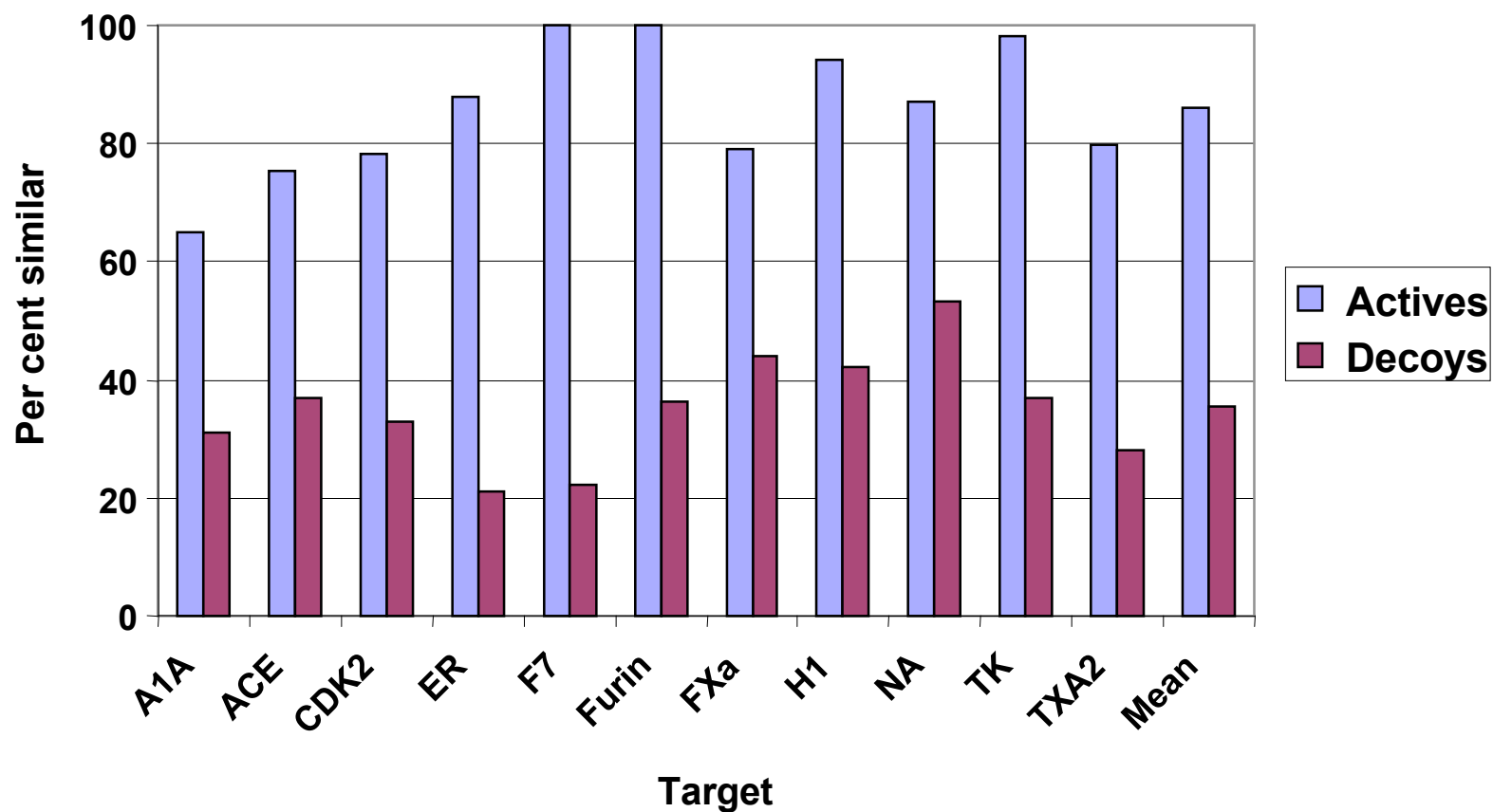
Ratio of conformers per active to conformers per decoy

# Just suppose....

- Molecules that share biological activity share shapes
  - Molecules that do not share activity share fewer shapes
- Select actives and decoys for same target and compare % of like shapes
  - Shape tanimoto  $> 0.7$
  - Combo  $> 1.0$

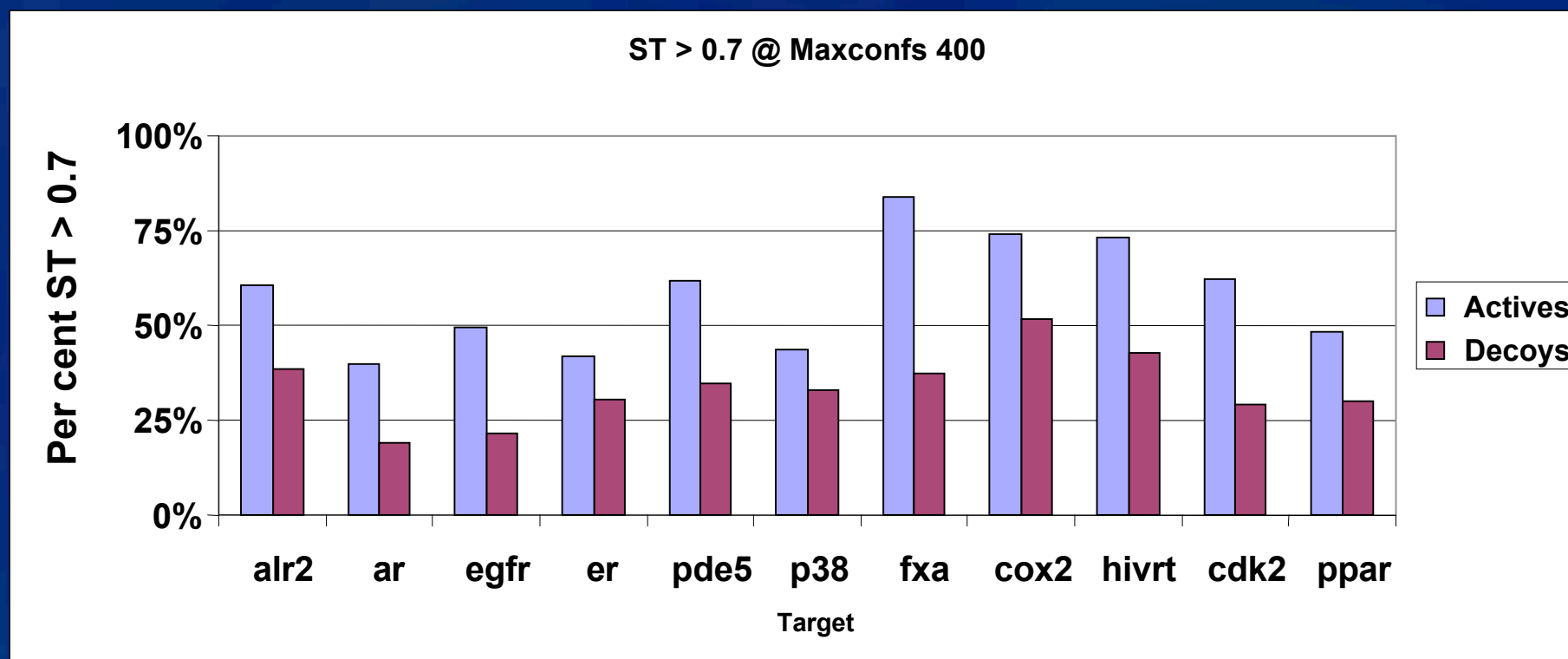


# Combo Similarity



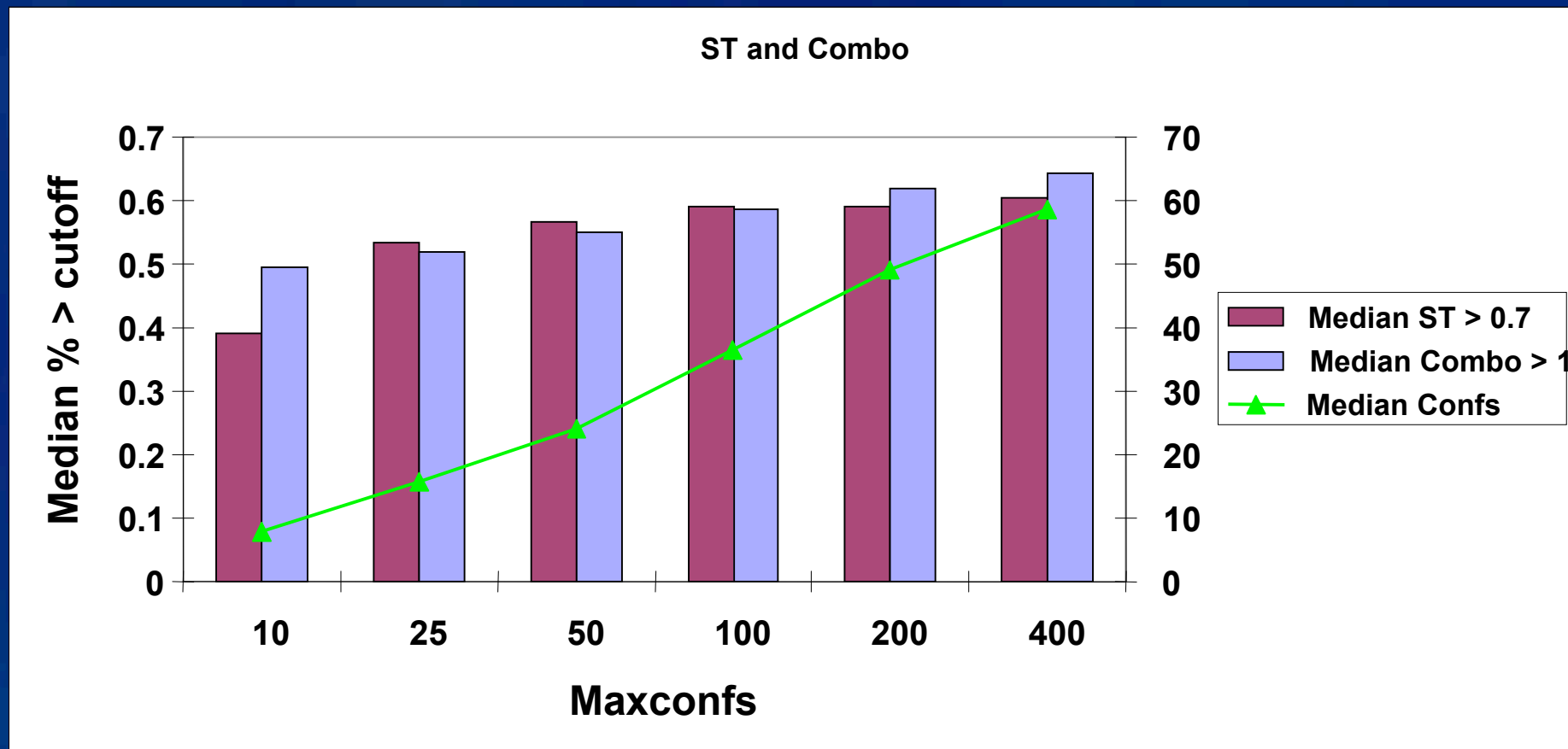
10 actives, 10 property matched decoys

# DUD 2.0 actives and decoys



15 actives/decoys, maxconfs 400  
Per cent having at least one pair of conformers  $ST > 0.7$ .

# DUD 2.0 actives share chemistry too



# But...

- This is a snapshot
  - Different actives, different results
  - Different decoys, different results
  - Bootstrapping
- No confidence intervals



# Conclusion

- Effective searching in shape space does not require an exhaustive sampling of conformer space.
  - Query or database molecule
- Molecules with shared biological activity share more shape space than molecules that do not share biological activity.
- More work required to provide robust conclusions.



# Acknowledgements

- In the “real” world
  - Niu Huang, John Irwin, Andrew Good
- The Hive
  - Geoff “Skillmaldo” Skillman
  - Bob “Train Man” Tolbert
  - Anthony “People Person” Nicholls



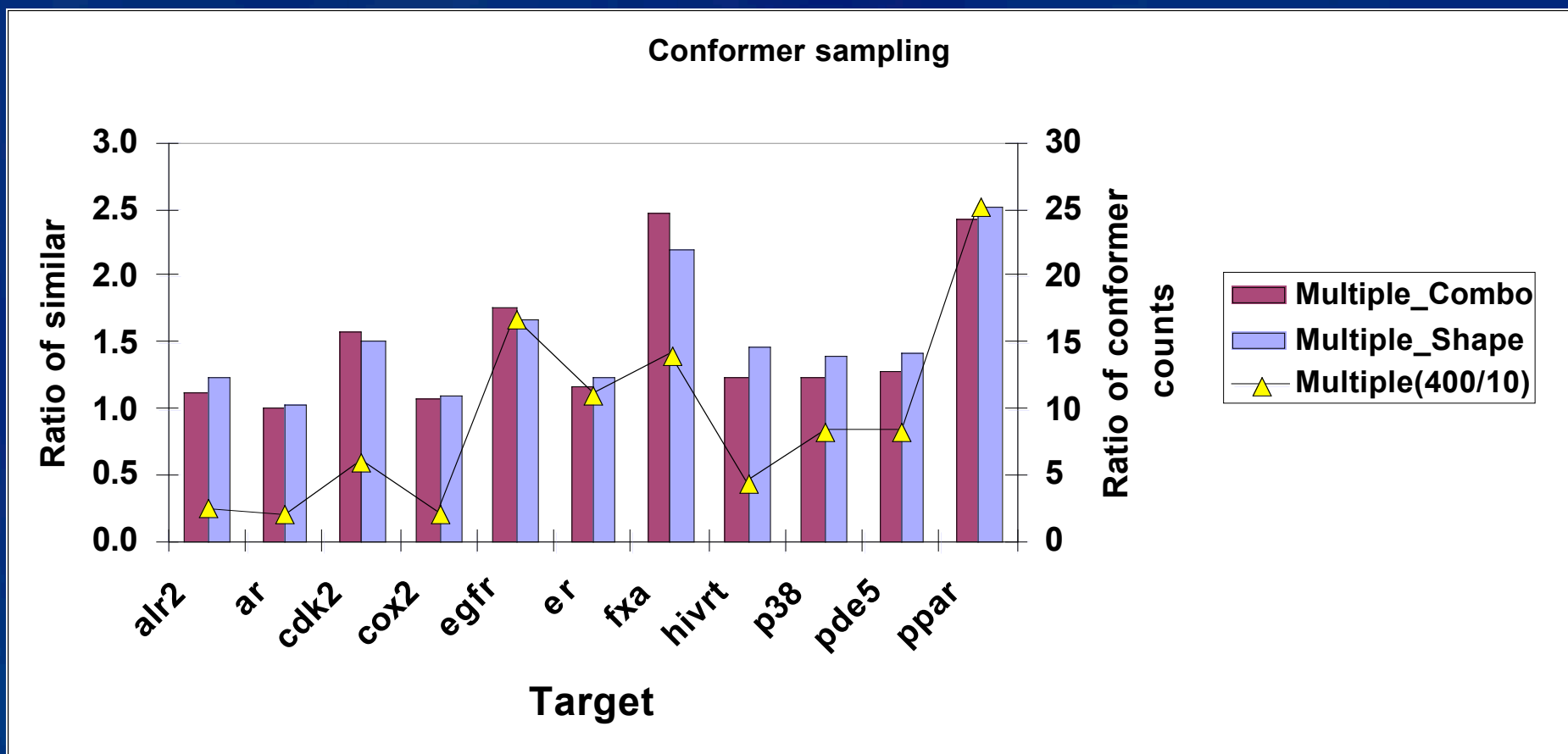
# Almost forgot....

- “To one who has faith, no explanation is necessary. To one without faith, no explanation is possible”
  - St. Thomas Aquinas, *Summa Theologica*





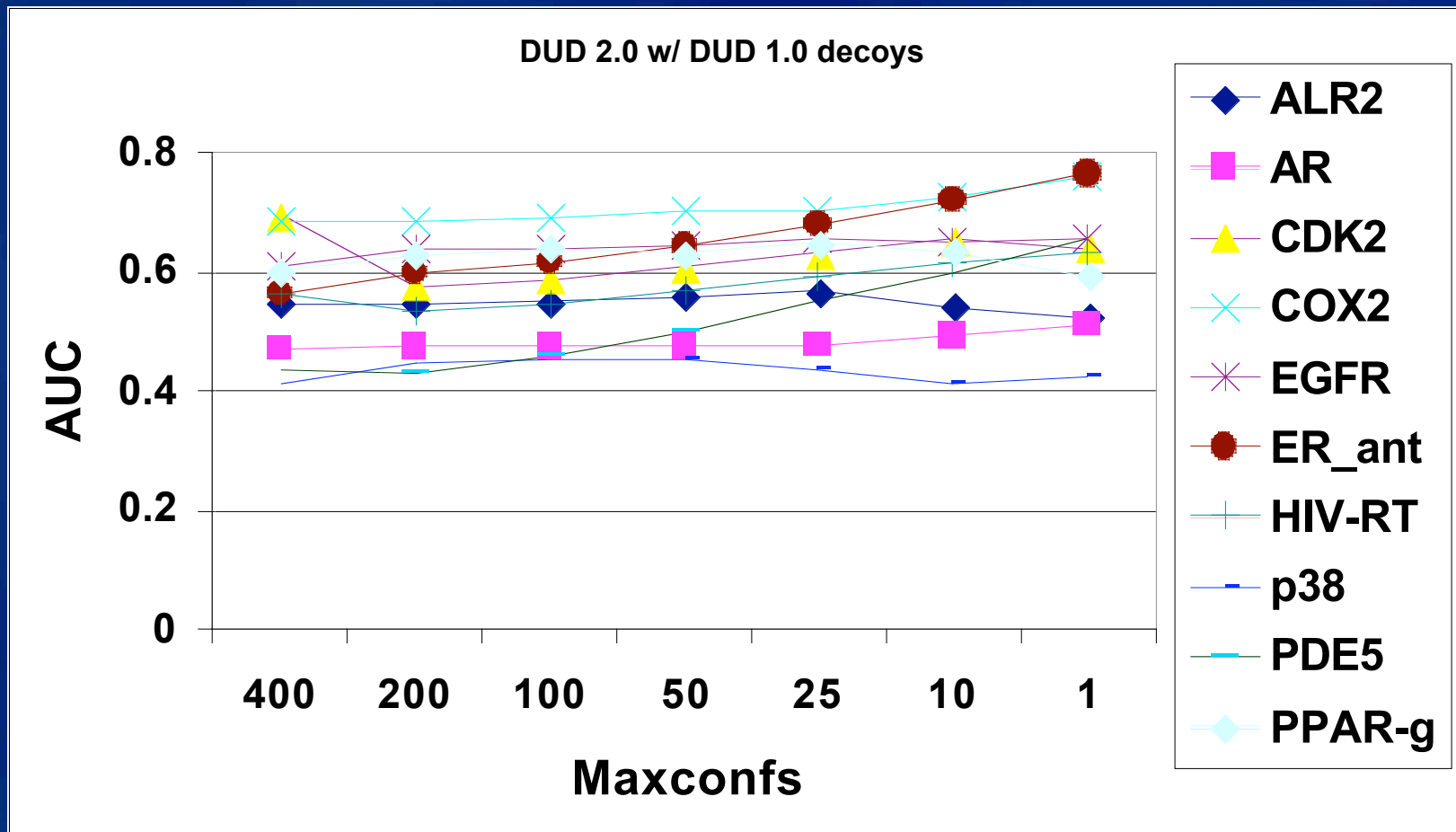
# The obligatory indecipherable graph



Ratios on going from maxconfs 400 to maxconfs 10



# Target by target



1 case random -> signal

# Multi-conformer queries give no extra signal

