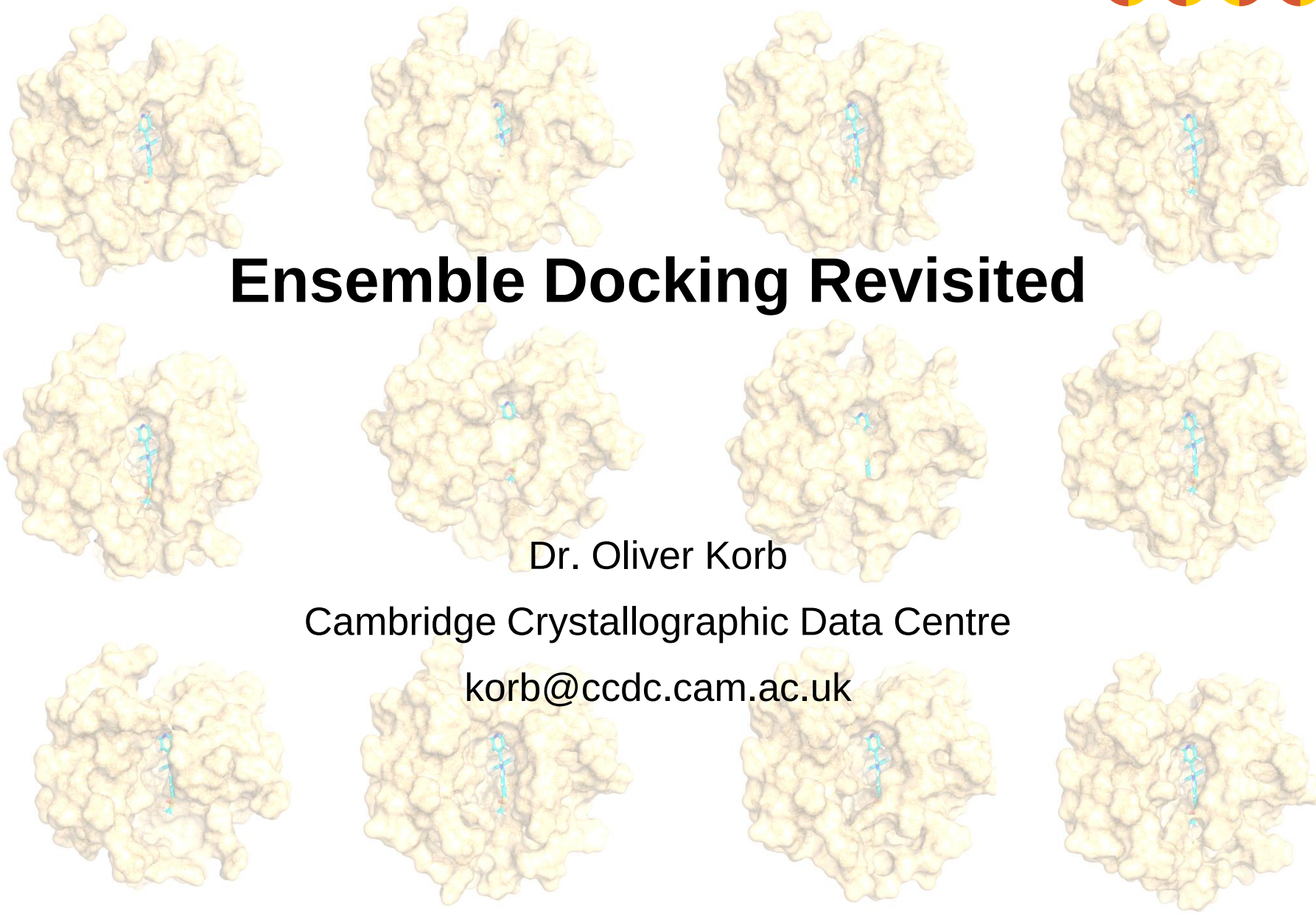


Ensemble Docking Revisited

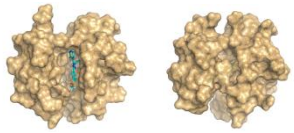
Dr. Oliver Korb

Cambridge Crystallographic Data Centre

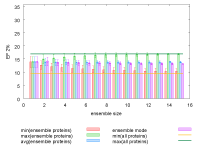
korb@ccdc.cam.ac.uk

Outline

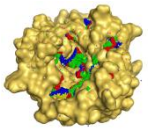
Introduction



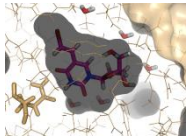
Simulated Ensemble Docking / Screening



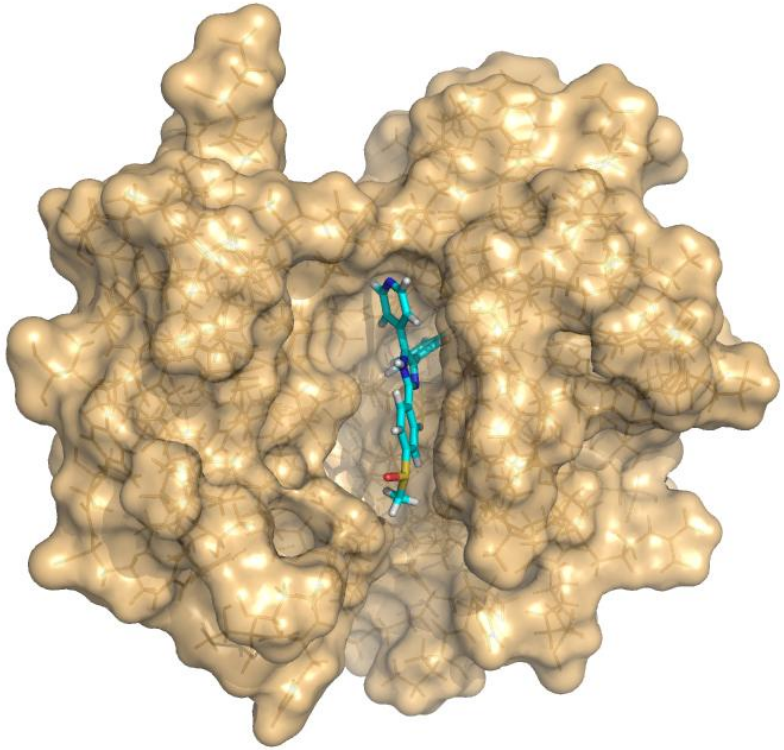
GOLD Ensemble Docking



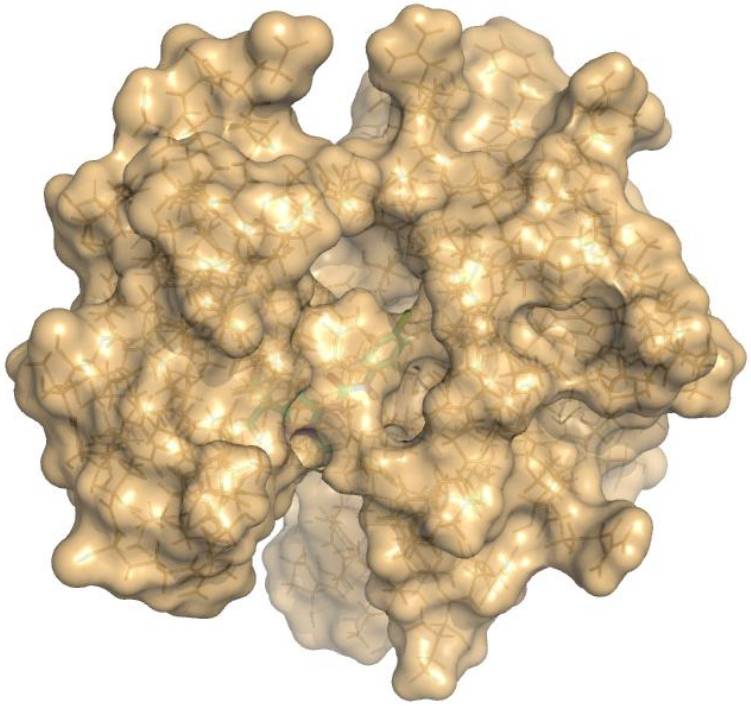
Future Work



Introduction



1a9u
DFG in



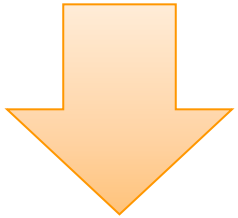
1kv1
DFG out

***induced fit* effect in p38 MAP kinase**

Introduction

- generating **meaningful** protein conformations during docking is a difficult task

- large-scale protein rearrangements can only hardly be modelled

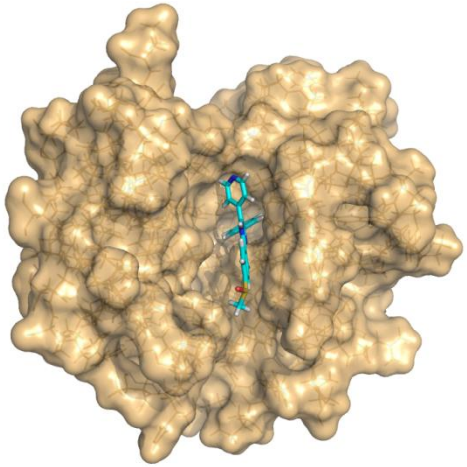


- *ensemble*-based approaches only consider a set of discrete protein conformations

Introduction – Ensemble Docking Approaches

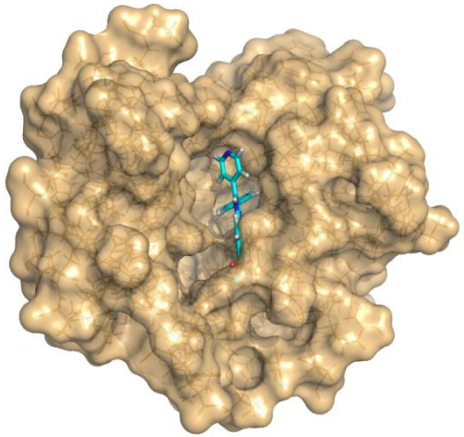
- Claussen et al. (FlexE) JMolBiol 308(2), 2001, 377-395
- Huang et al. (DOCK) Proteins 66(2), 2006, 399-421
- Rao et al. (Glide) JCAMD 22(9), 2008, pp 621-627
- Bottegoni et al. (ICM) JMedChem 52(2), 2009, 397-406

Multiple Protein Structure Docking



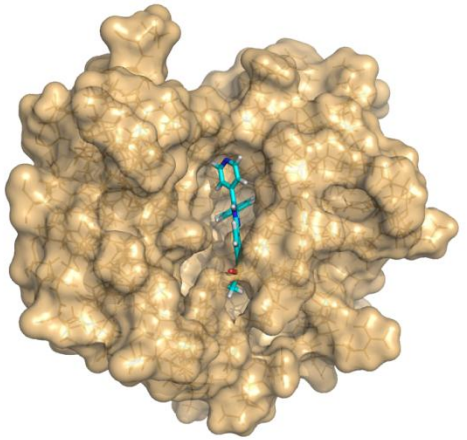
1a9u

score **74**



1bl6

60



1bl7

69

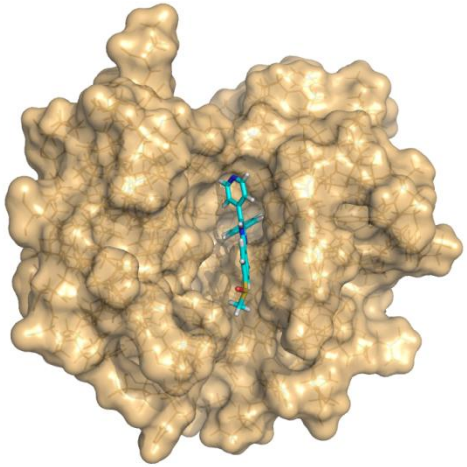
- ligands get different scores in different protein structures
- scores determine ranking performance in *virtual screening*

⇨ **which protein structure(s) to use for *virtual screening*?**

Sensitivity of Virtual Screening Results

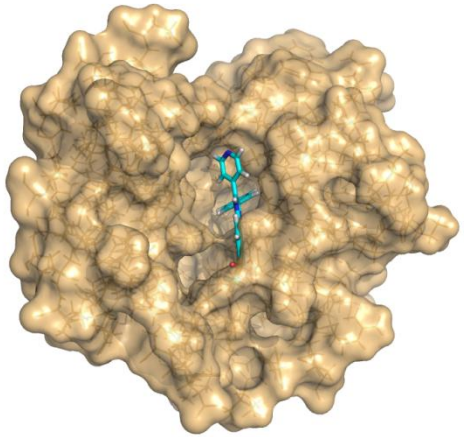
		AUC			EF 2%			EF 10%		
target	**# proteins**	**min**	**max**	**delta**	**min**	**max**	**delta**	**min**	**max**	**delta**
acetylcholine esterase	21	0.44	0.76	0.32	0.0	14.7	14.7	0.3	6.5	6.2
aldose reductase	32	0.38	0.60	0.22	9.4	16.9	7.5	2.3	5.0	2.7
cyclin-dependent kinase 2	72	0.41	0.70	0.29	0.0	17.6	17.6	0.6	5.2	4.6
dihydrofolate reductase	9	0.55	0.79	0.24	2.2	8.2	6.0	1.3	4.0	2.7
factor Xa	34	0.69	0.89	0.20	3.2	19.8	16.6	2.6	7.4	4.8
heat shock protein 90	30	0.68	0.88	0.20	0.0	14.5	14.5	1.7	6.2	4.6
neuraminidase	13	0.77	0.87	0.10	2.0	15.1	13.1	2.4	5.5	3.1
p38 MAP kinase	31	0.44	0.76	0.32	1.3	12.5	11.2	0.7	4.4	3.7
phosphodiesterase 5A	5	0.67	0.73	0.06	9.6	11.5	1.9	3.5	4.7	1.2

Simulated Ensemble Docking



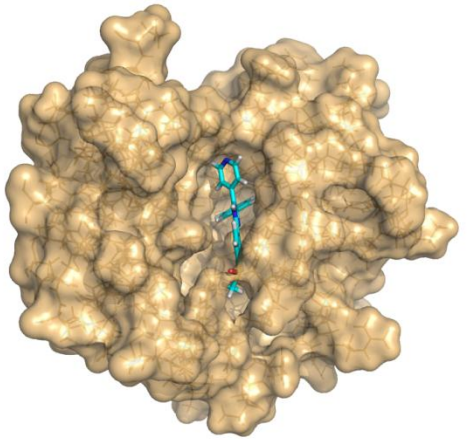
1a9u

score **74**



1bl6

60



1bl7

69

- for each ligand pick the best-scoring protein structure

⇨ simulates a **perfect** ensemble docking approach

Simulated Ensemble Docking

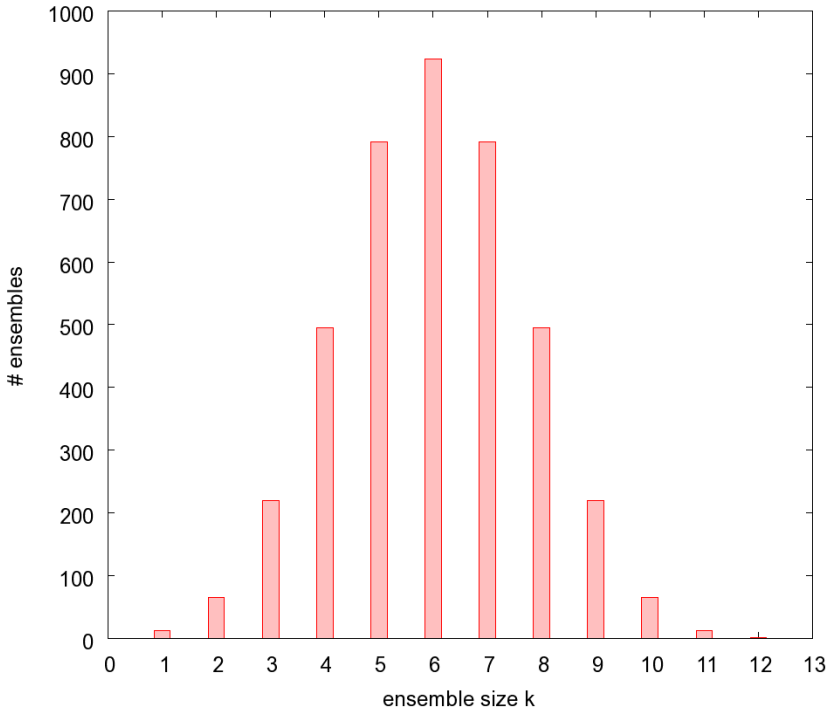
- perform docking / screening for ***n*** protein structures

  $2^n - 1$ different ensembles (size 1 or greater)

  $\binom{n}{k}$ ensembles of size *k*

- example ***n*** = 12
  - 4095 different ensembles

- simulate docking into all $2^n - 1$ ensembles by post-processing ***n*** docking results



Targets

target	PDB	# holo proteins[a]	# actives	# inactives
acetylcholine esterase	1gpk	21	105	3623
aldose reductase	1t40	32	26	902
cyclin dependent kinase 2	1ke5	72	50	1661
dihydrofolate reductase	1s3v	9	201	6496
factor Xa	1lpz	34	141	4535
heat shock protein 90	2bsm	30	24	823
neuraminidase	1l7f	13	49	1726
p38 MAP kinase	1ywr	31	240	8203
phosphodiesterase 5A	1xoz	5	51	1808

curated DUD[b] set

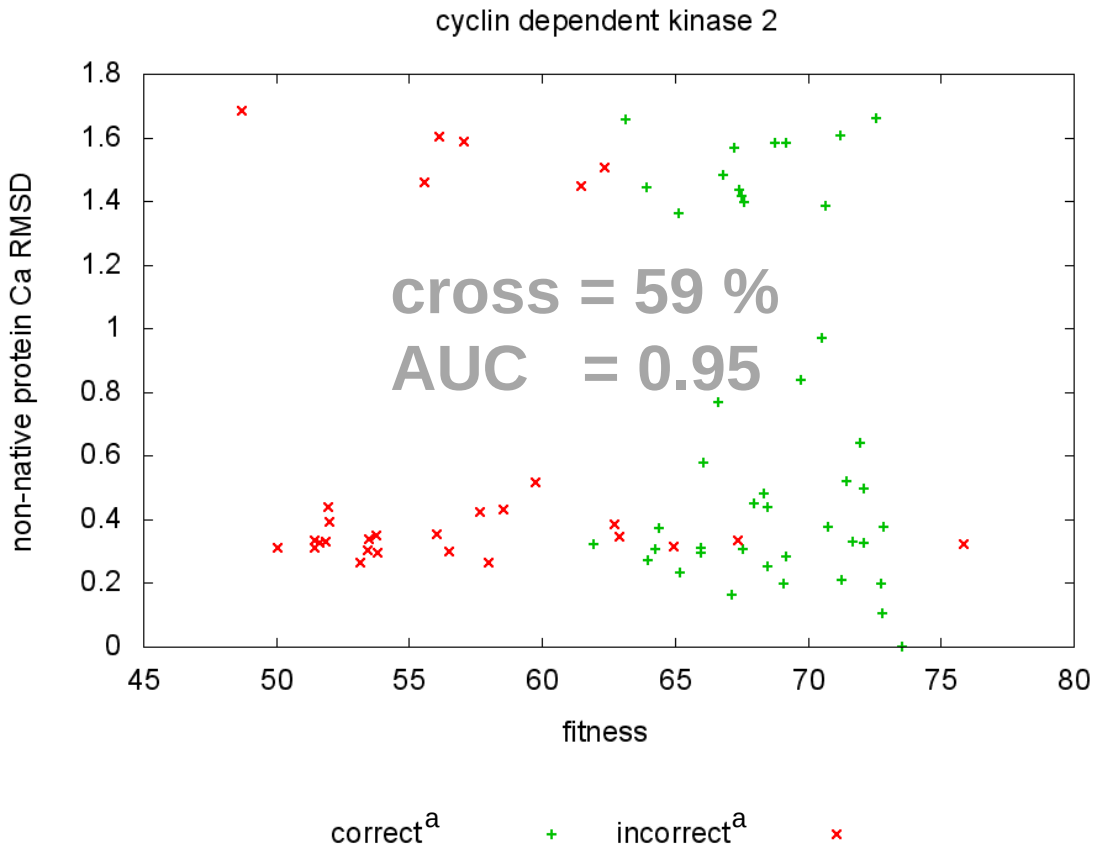
- pose prediction results averaged over 20 independent runs
- virtual screening: single run with *autoscale = 1.0*

[a] Verdonk et al. *JCIM*, 48, 2214-2225 (2008)
[b] Huang et al. *JMedChem*, 49, 6789-6801 (2006)

Assessing Ensemble Docking Performance

- a good ensemble scoring function should
  - exhibit a good cross-docking performance
  - discriminate well between correctly and incorrectly docked solutions
- *cross-docking performance*: number of correctly predicted poses in non-native protein structures
- *discrimination performance*: calculate AUC for discrimination between correctly and incorrectly docked solutions (ranked by fitness)

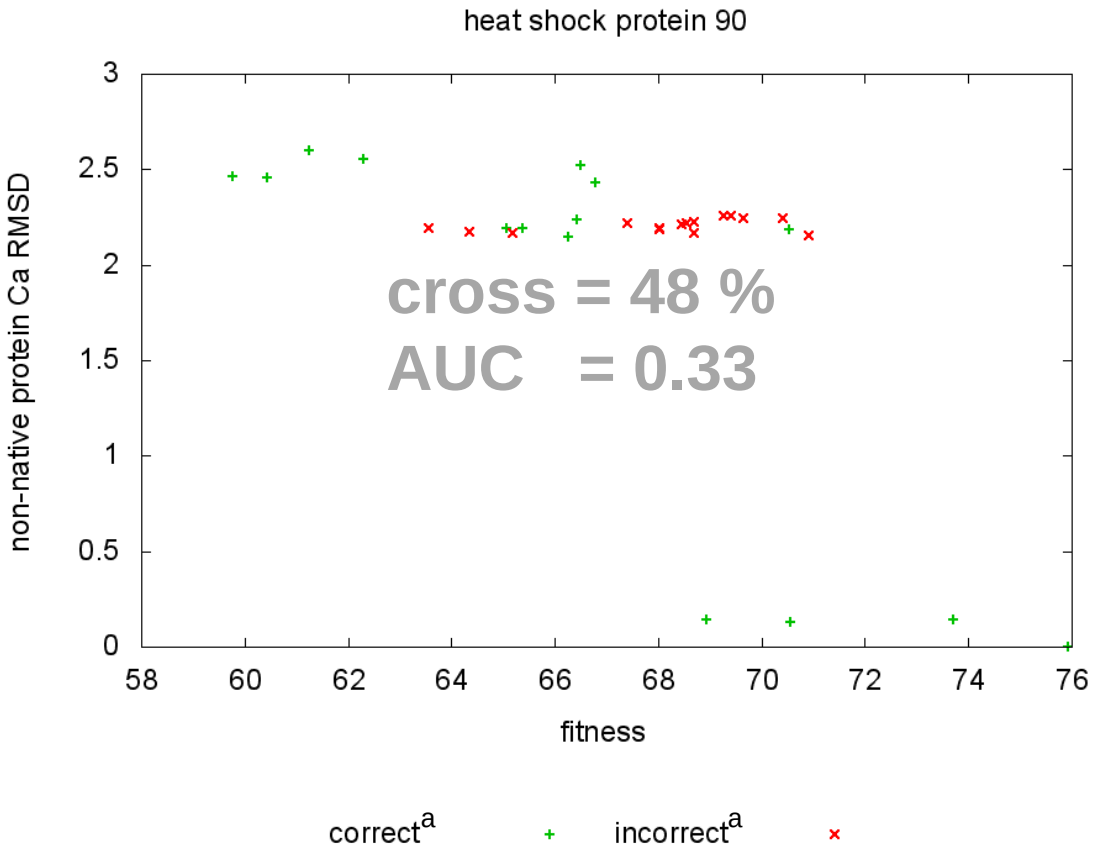
Assessing Ensemble Docking Performance



each data point represents the docking result for one protein structure (72 for CDK2)

[a] correct if top-ranked solution rmsd < 2 Å, incorrect otherwise

Assessing Ensemble Docking Performance



each data point represents the docking result for one protein structure (30 for HSP90)

[a] correct if top-ranked solution rmsd < 2 Å, incorrect otherwise

Ensemble Docking – Pose Prediction

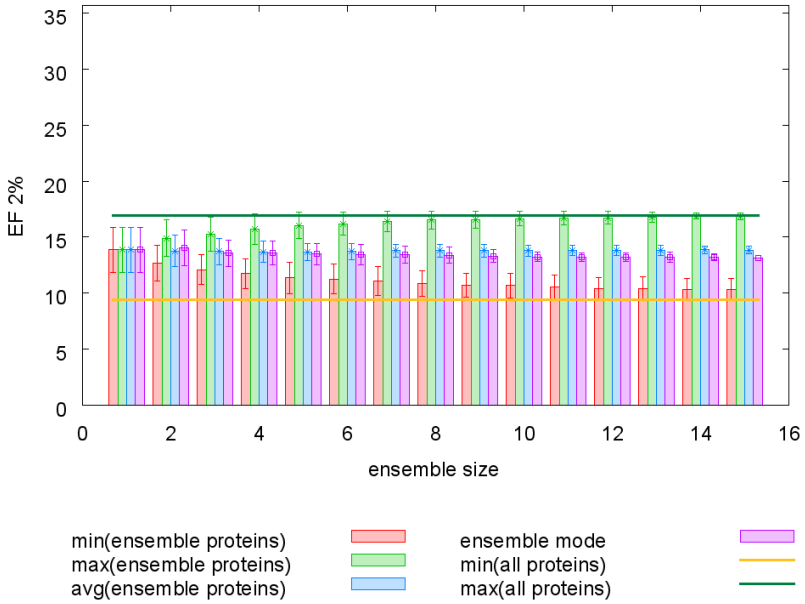
	AUC[a]	# correct	# proteins	% correct	rank[b]	improvement[c]
CHEMPLP						
acetylcholine esterase	0.55	10	20	50	1	
aldose reductase	0.83	15	31	48	1	
cyclin dependent kinase 2	0.95	42	71	59	2	
dihydrofolate reductase	1.00	7	8	88	1	
factor Xa	0.61	16	33	48	1	
heat shock protein 90	0.33	14	29	48	1	
neuraminidase	1.00	12	12	100	1	
p38 MAP kinase	0.65	3	30	10	5	
phosphodiesterase 5	1.00	2	4	50	1	
avg.	0.77			56		
GOLDSCORE						
acetylcholine esterase	0.22	2	20	10	15	
aldose reductase	0.89	11	31	35	2	
cyclin dependent kinase 2	0.75	36	71	51	1	
dihydrofolate reductase	0.58	6	8	75	1	
factor Xa	0.66	26	33	79	1	
heat shock protein 90	0.77	26	29	90	1	
neuraminidase	1.00	12	12	100	1	
p38 MAP kinase	0.51	3	30	10	2	
phosphodiesterase 5	1.00	1	4	25	1	
avg.	0.71			53		

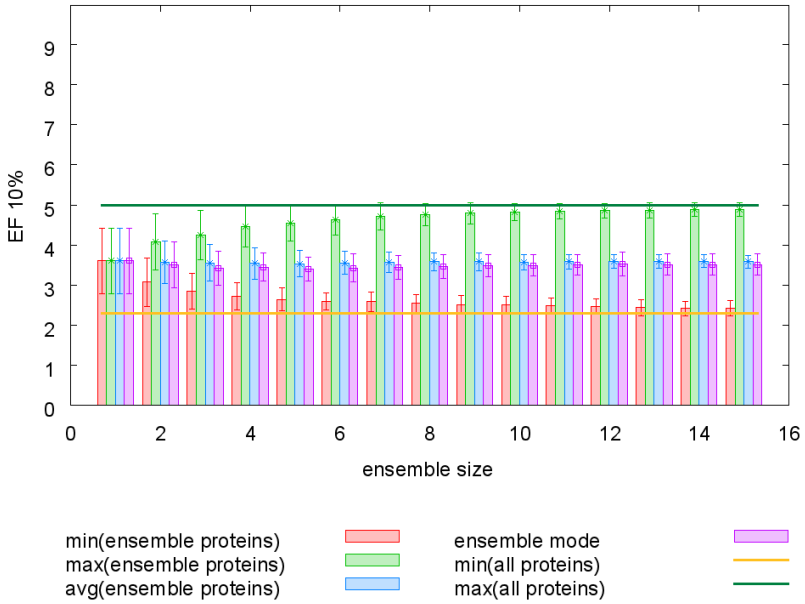
[a] discrimination between correctly and incorrectly predicted solutions

[b] rank of first correctly docked solution

[c] if ensemble docking performs better than the average single protein structure, otherwise

Virtual Screening – Heat Shock Protein 90

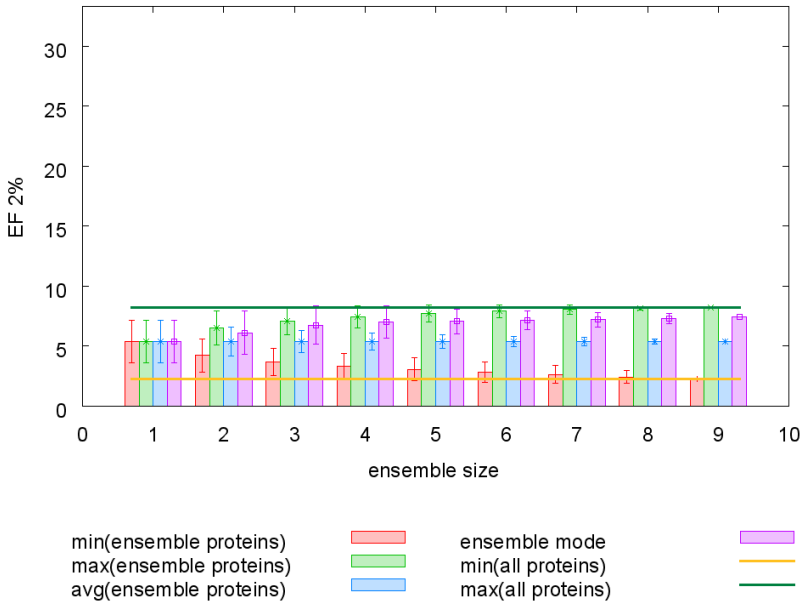


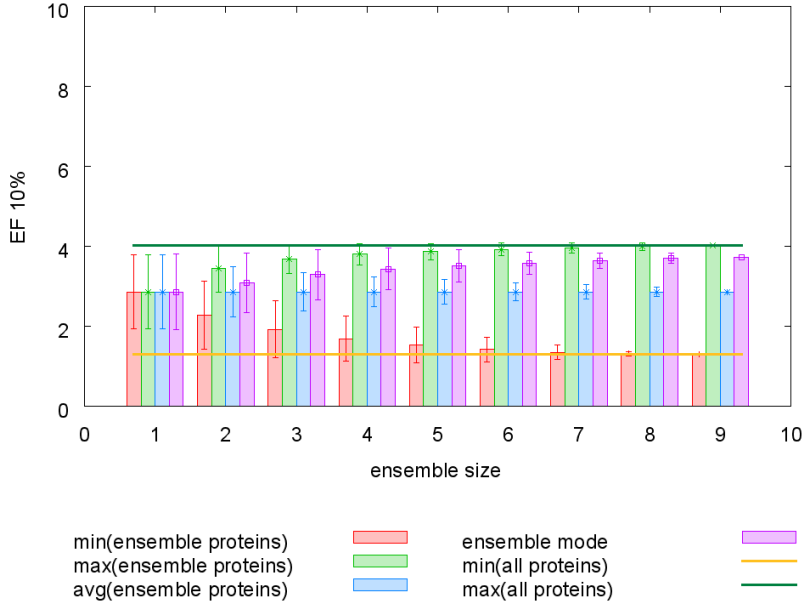


no improvement

no improvement

Virtual Screening – Dihydrofolate Reductase

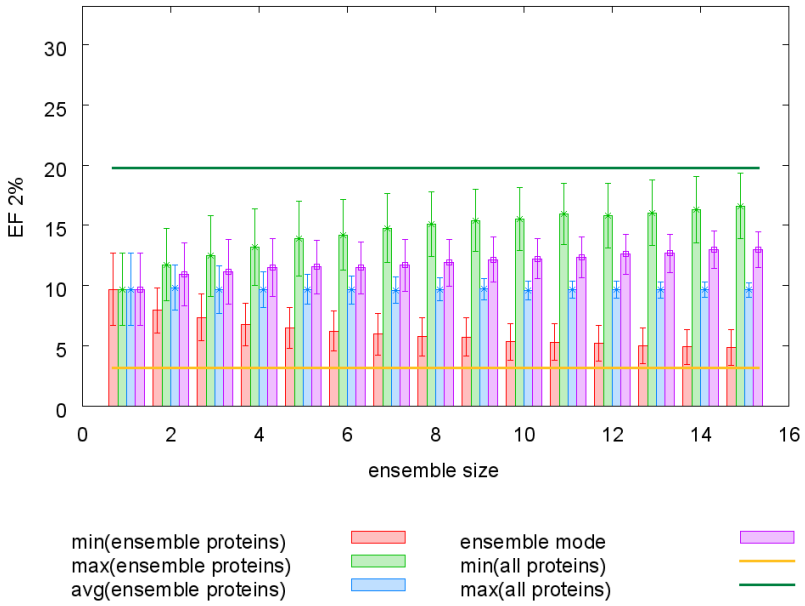


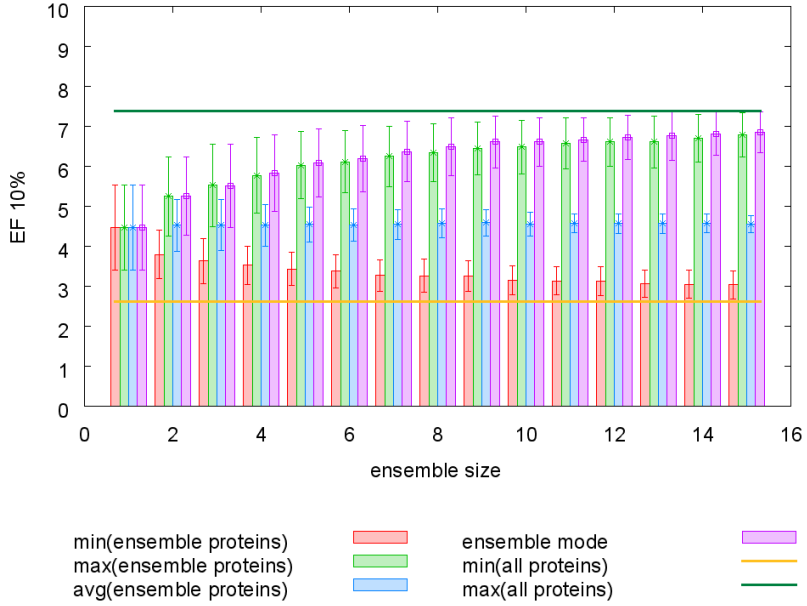


medium improvement

medium improvement

Virtual Screening – Factor Xa

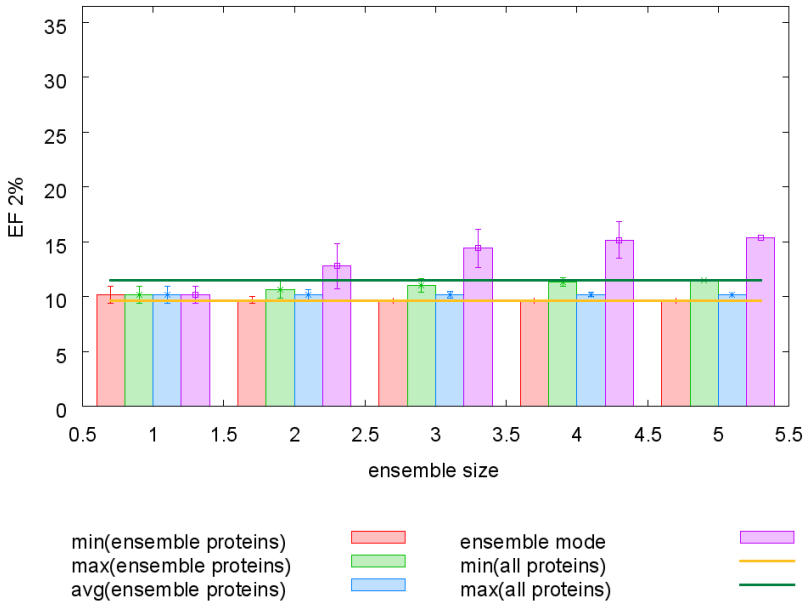


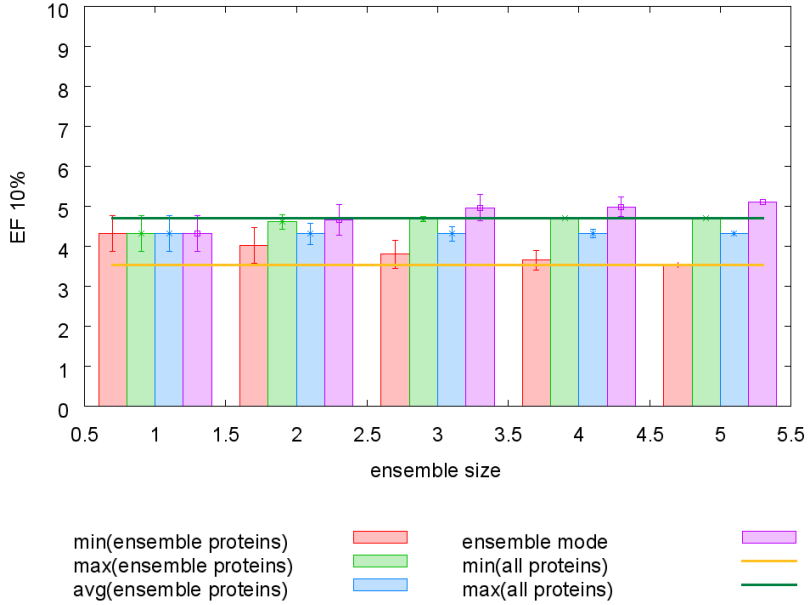


medium improvement

major improvement

Virtual Screening – Phosphodiesterase 5A





major improvement

medium improvement

Improving Upon the Best Single Protein Structure

protein 1		protein 2		ensemble	
L1	*70*	L2	*60*	L1	*70*
D	*50*	D	*45*	L2	*60*
L2	*40*	L1	*30*	D	*50*

Virtual Screening Results

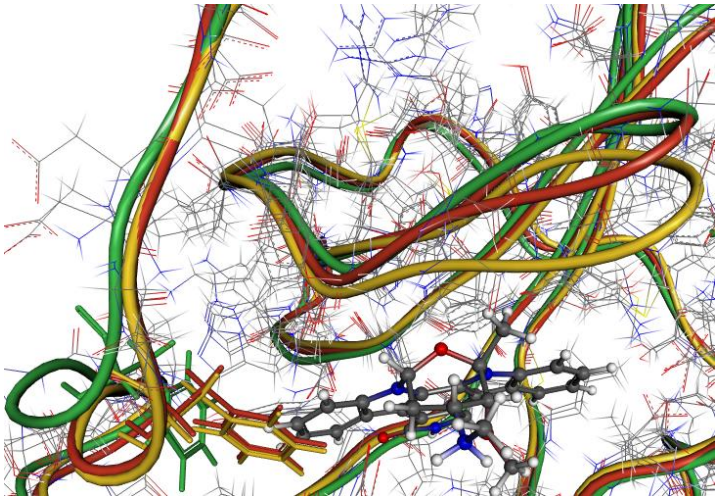
target	AUC	EF 2%	EF 10%
acetylcholine esterase			
aldose reductase			
cyclin dependent kinase 2			
dihydrofolate reductase			
factor Xa			
heat shock protein 90			
neuraminidase			
p38 MAP kinase			
phosphodiesterase 5A			

ensemble performance compared to average performance of single protein structures

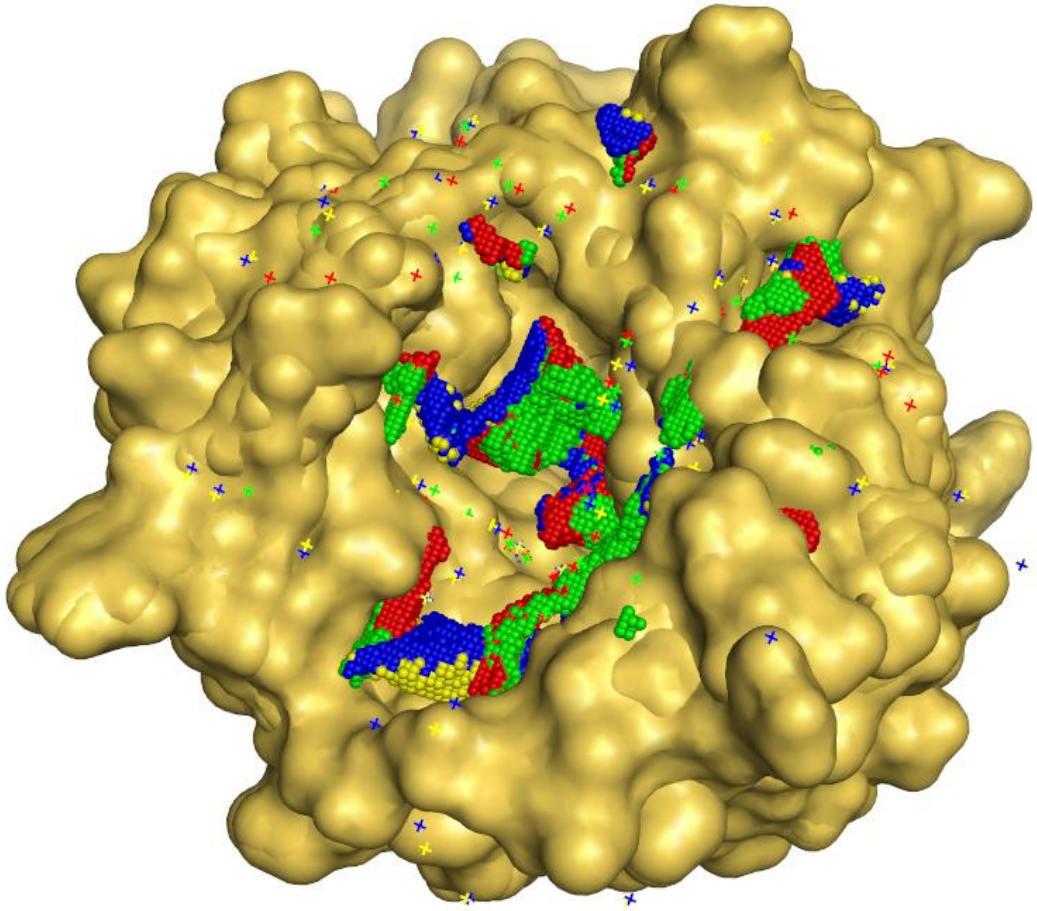
no improvement
medium improvement
major improvement

GOLD ensemble

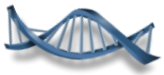
- results so far based on sequential docking
- modified *genetic algorithm* to treat protein ensembles
- requires a superimposed set of protein structures
- searches all protein conformations **concurrently**



GOLD ensemble - Fitting Points



GOLD ensemble – Genetic Algorithm



mapping degrees of freedom	protein degrees of freedom	ligand degrees of freedom	protein ID

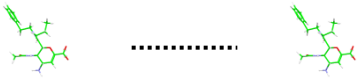
for *n* protein structures

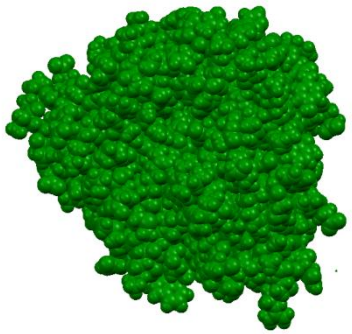
selects protein structure for scoring

- **ID mode**: change the protein during the GA-search by *mutation*
- **island mode**: search all protein structures concurrently

GOLD ensemble – Island Mode

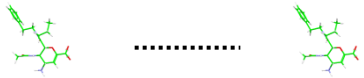
protein ID: 1

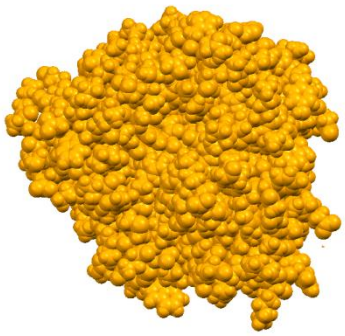




island 1

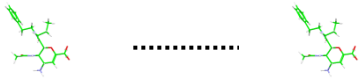
protein ID: 2

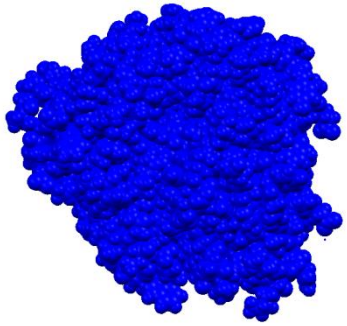




island 2

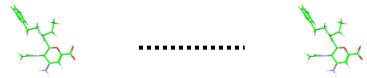
protein ID: 3

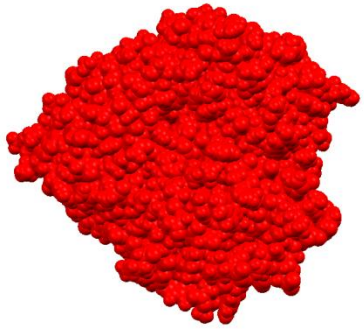




island 3

protein ID: 4





island 4

GOLD ensemble - Timing results (Island mode)

	sequential			ensemble			
target	**succ [%]**	**max fitness**	**time[a] [s]**	**succ [%]**	**max fitness**	**time[a] [s]**	**speedup**
acetylcholine esterase	100	73.55	69.1	100	74.1	61.7	1.1
aldose reductase	100	122.81	405.8	65	119.8	118.3	3.4
cyclin-dependent kinase 2	95	74.69	210.8	85	70.8	79.4	2.7
dihydrofolate reductase	100	96.03	273.0	100	95.9	93.7	2.9
factor Xa	95	108.88	566.8	90	108.5	142.8	4.0
heat shock protein 90	45	71.92	432.0	100	71.0	156.6	2.8
neuraminidase	100	92.45	227.9	100	91.6	111.5	2.0
p38 MAP kinase	0	98.05	594.3	15	90.4	180.7	3.3
phosphodiesterase 5A	100	83.57	19.8	95	83.8	22.3	0.9
avg.	81.7		311.1	83.3		107.4	

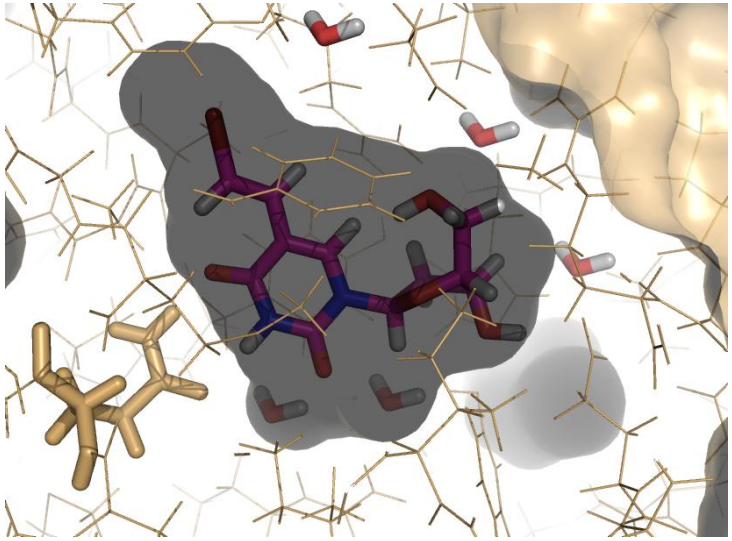
[a] Intel Xeon CPU 5130, 2GHz

Conclusions

- ensemble docking can improve hit rates
  - increases worst and average case performance in many cases
  - performs sometimes as good as the best single protein structures
- trends suggest to use multiple protein structures in an ensemble protocol
- GOLD has been extended to search ensembles time-efficiently

Future Work

- analysis of *chemotype* enrichment
- investigation of protein energies
- combine ensemble docking with flexible side-chains and switching of explicit water molecules



Acknowledgments

Jason Cole
Simon Bowden
Tjelvar Olsson
David Frenkel
John Liebeschuetz
Colin Groom

DAAD

UNIVERSITY OF CAMBRIDGE
Darwin Supercomputing Facilities

astex therapeutics

Thank you for your attention!