SuMo: structure comparison of proteins focused on functional properties of ligand binding sites

Martin Jambon

mjambon@burnham.org

IBCP

The Burnham Institute

Research performed at: IBCP, Lyon, France
MEDIT, Palaiseau, France

Presented to: Automated Function Prediction SIG,
ISMB 2005, June 24th, Detroit, USA

Problems and limits

Ligand binding elements

Comparison algorithm

Examples

Conclusions

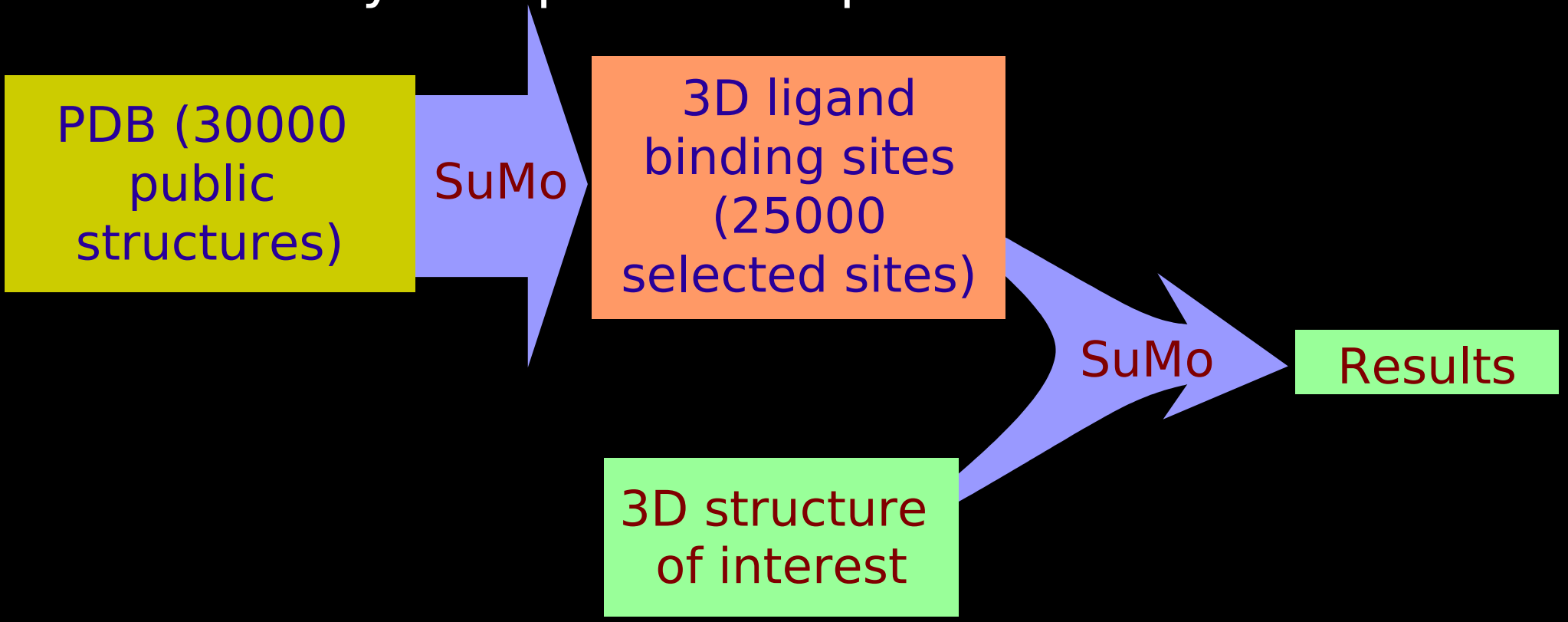
- problem
- SuMo
- limits

Problem: what kind of ligands fit my protein structure?

- Kinds of interactions:
  - hydrogen bonds
  - shape complementarity
  - hydrophobic effect
- Functional groups
  - binding only
  - reactive

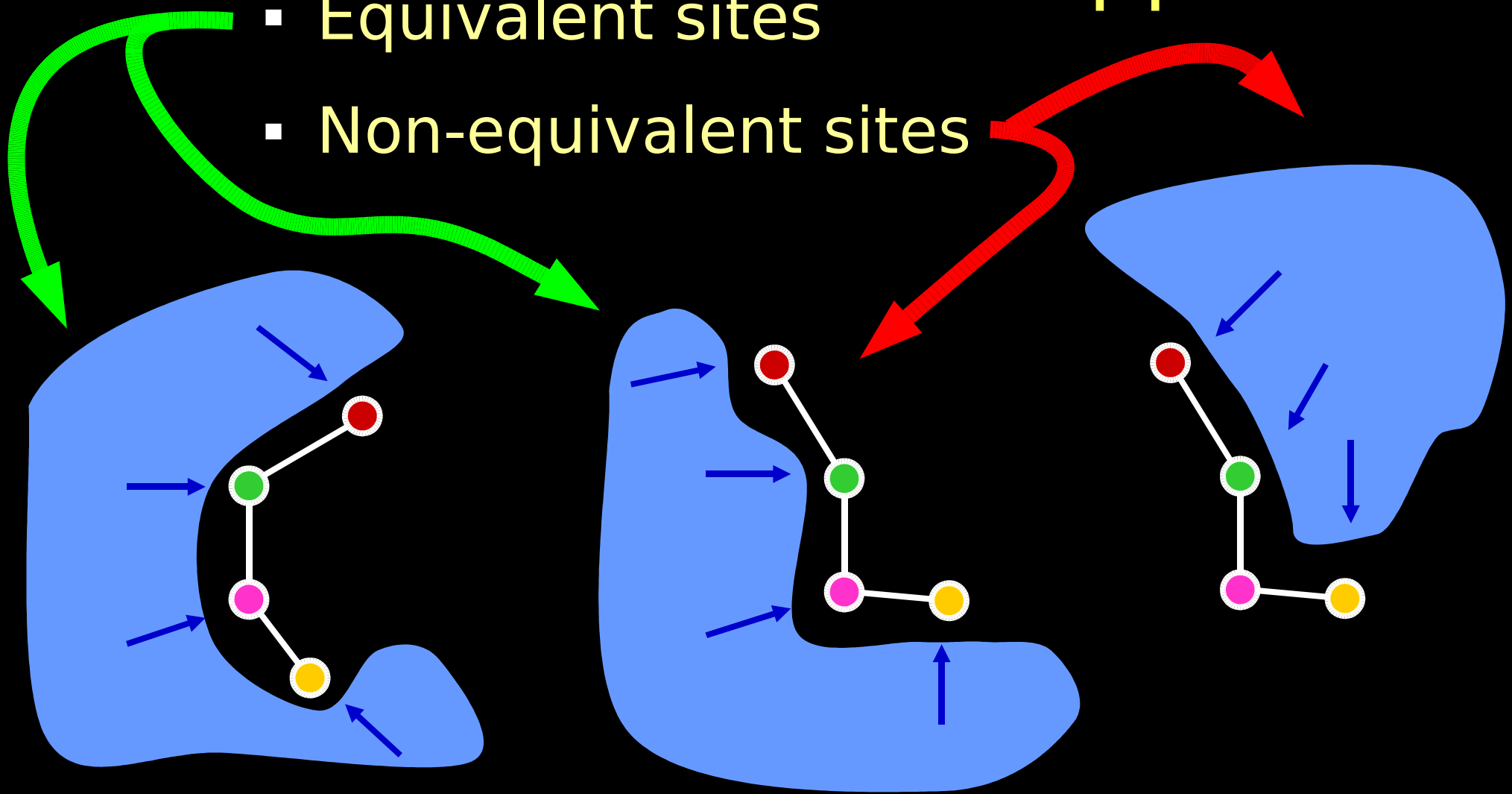
The SuMo system

- Searches for functional sites
  - structure-based, fold and sequence independent
  - works by comparison of protein structures or sites



Limits of comparison-based approaches

- Sequence motifs
- Main chain fold
- Superposable sites
- Equivalent sites
- Non-equivalent sites



Problems and limits

Ligand binding elements

Comparison algorithm

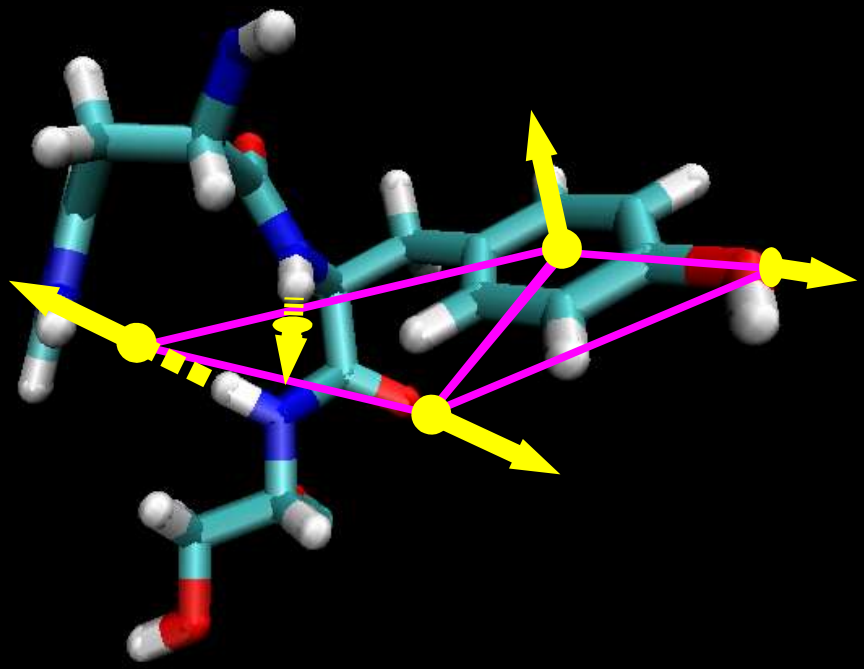
Examples

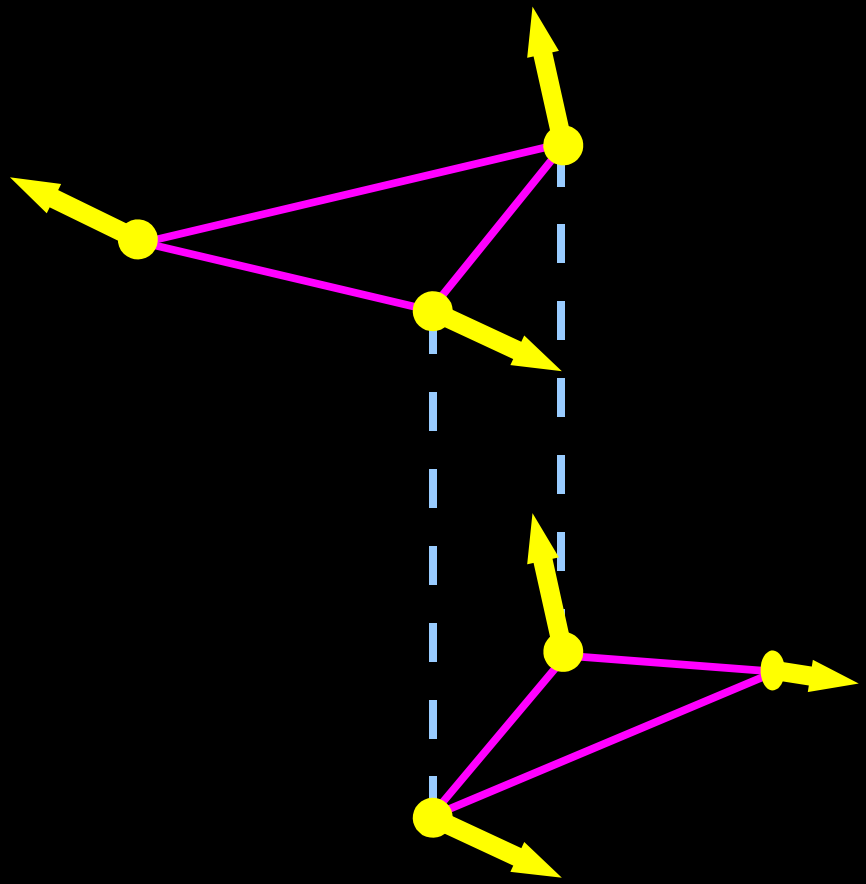
Conclusions

- microsites
- larger sites

Microsites in SuMo

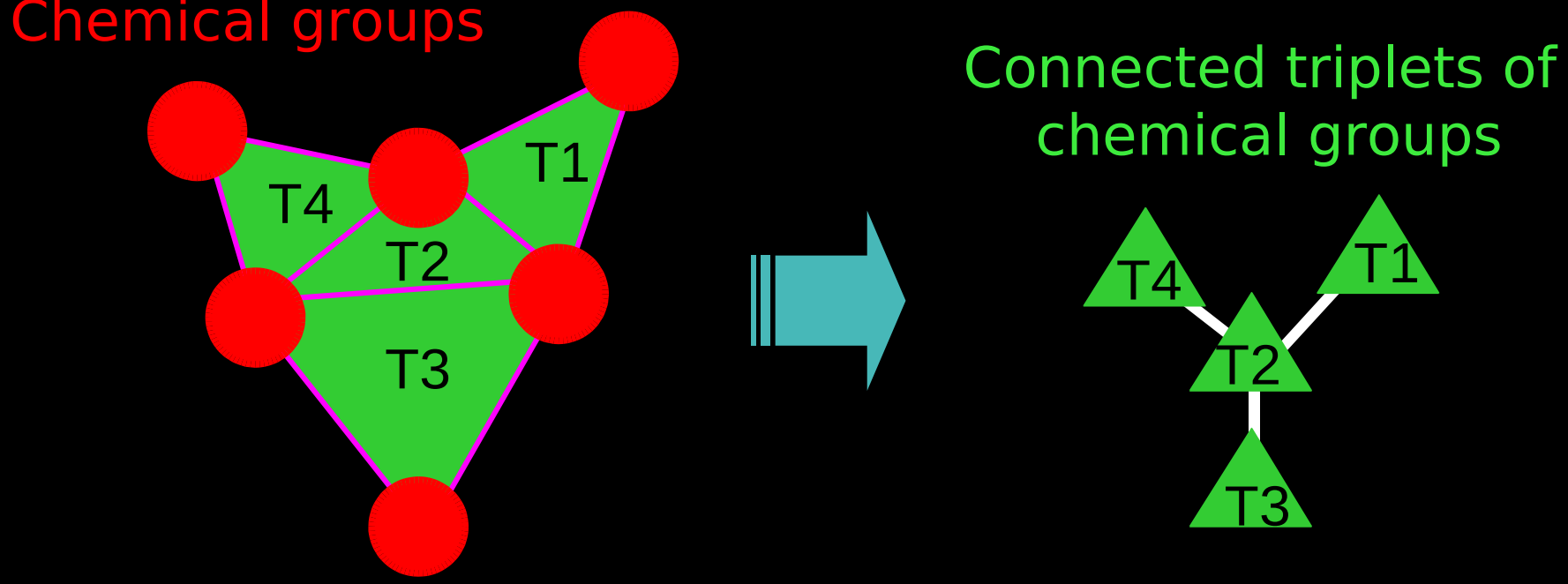
- Microsite =
  - 3 anchors forming a triplet of bounded size
  - atomic shape





Sites of arbitrary size

- Connected triplets
  - virtual hinges
  - angles



Problems and limits

Ligand binding elements

Comparison algorithm

Examples

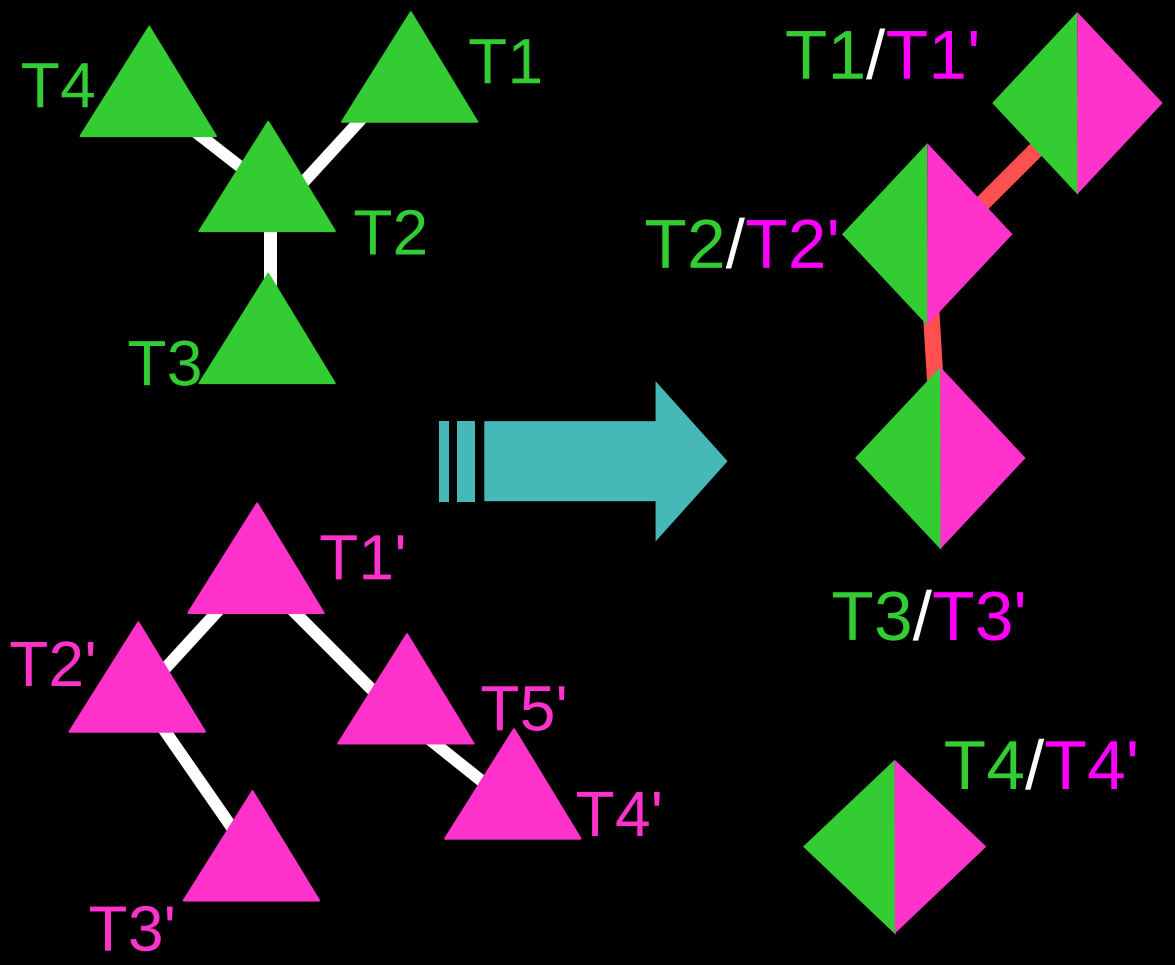
Conclusions

- pairs of microsites
- extraction of matched sites

Comparison (I) matching microsites

- Matching the triangles
  - chemical groups of same type and
  - same orientation
  - edges of similar length
- Matching the shape in the region of the triangle
  - weighted shape
  - comparison of weighted volumes

Comparison (II)
extended matched sites



- pairs of similar triplets
- double connection rule
- extraction of independent subgraphs

Problems and limits

Ligand binding elements

Comparison algorithm

Examples

Conclusions

- NTP binding
- unknown function

When the lateral chains don't help much: P-loop NTP-binding sites

- Poor Prosite motif [AG] - x(4) - G - K - [ST]
- Phosphate-main-chain H-bonds

1B0U - TRANSPORT PROTEIN - ATP-BINDING SUBUNIT OF THE HISTIDINE PERMEASE FROM SALMONELLA TYPHIMURIUM

Matching sites: 3468 (13.2%)
Analyzed sites: 26207
Patches: 5006

See also:
- Ligands sorted by best score
- Map

Pages: First | **Previous** | All | Next | Last

	PDB structure sort	Number of pairs sort	Number of amino acids sort	Volume sort	Relative volume sort	Ligand code sort	DESCRIPTION
1 (1:1) [more] [less]	1B0U(5)	11	5	**4.8**	100%	ATP	TRANSPORT PROTEIN - ATP-BINDING SUBUNIT OF THE HISTIDINE PERMEASE FROM SALMONELLA TYPHIMURIUM
2 (1:1) [more] [less]	1JI0(20)	11	6	**4.5**	70%	ATP	TRANSPORT PROTEIN - CRYSTAL STRUCTURE ANALYSIS OF THE ABC TRANSPORTER FROM THERMOTOGA MARITIMA
3 (1:2) [more] [less]	1G3R(3)	10	6	**4.5**	66%	ACP	CELL CYCLE, HYDROLASE - CRYSTAL STRUCTURE ANALYSIS OF PYROCOCCUS FURIOSUS CELL DIVISION ATPASE MIND
4 (1:2) [more] [less]	1FP6(25)	9	5	**4.4**	46%	ADP	OXIDOREDUCTASE - THE NITROGENASE FE PROTEIN FROM AZOTOBACTER VINELANDII COMPLEXED WITH MGADP
5 (1:1) [more] [less]	1CJK(8)	9	5	**4.2**	32%	GSP-MG	LYASE/LYASE/SIGNALING PROTEIN - COMPLEX OF GS-ALPHA WITH THE CATALYTIC DOMAINS OF MAMMALIAN ADENYLYL CYCLASE: COMPLEX WITH ADENOSINE 5'-(ALPHA THIO)- TRIPHOSPHATE (RP), MG, AND MN
6 (1:2) [more] [less]	1JJ7(4)	10	5	**4.2**	60%	ADP	PROTEIN TRANSPORT - CRYSTAL STRUCTURE OF THE C-TERMINAL ATPASE DOMAIN OF HUMAN TAP1
7 (1:1) [more] [less]	1AS0(2)	9	5	**4.1**	23%	GSP-MG	SIGNAL TRANSDUCTION - GTP-GAMMA-S BOUND G42V GIA1
8 (1:1) [more] [less]	1CUL(8)	9	5	**4**	30%	GSP-MG	LYASE/LYASE/SIGNALING PROTEIN - COMPLEX OF GS-ALPHA WITH THE CATALYTIC DOMAINS OF MAMMALIAN ADENYLYL CYCLASE: COMPLEX WITH 2',5'-DIDEOXY-ADENOSINE 3'- TRIPHOSPHATE AND MG
9 (1:2) [more] [less]	1N6R(4)	9	5	**4**	33%	GNP	PROTEIN TRANSPORT - CRYSTAL STRUCTURE OF HUMAN RAB5A A30L MUTANT COMPLEX WITH GPPNHP
10 (1:1) [more] [less]	1CJT(6)	9	5	**3.9**	28%	GSP-MG	LYASE/LYASE/SIGNALING PROTEIN - COMPLEX OF GS-ALPHA WITH THE CATALYTIC DOMAINS OF MAMMALIAN ADENYLYL CYCLASE: COMPLEX WITH BETA-L-2',3'-DIDEOXYATP, MN, AND MG

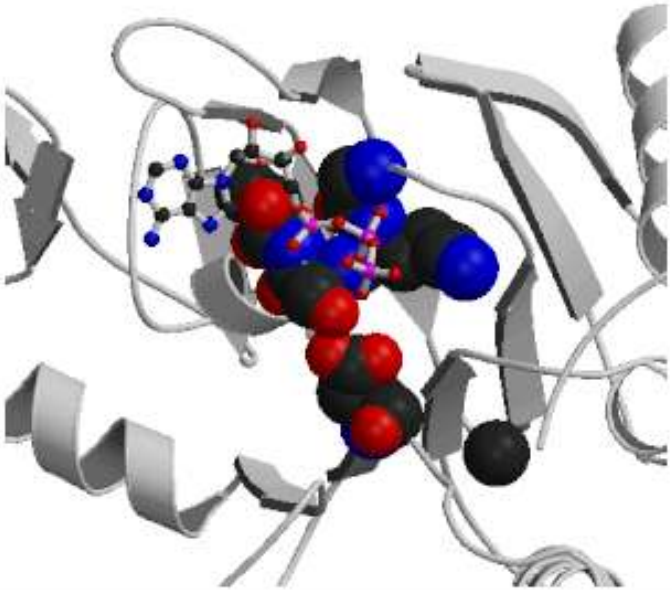
	1B0U				1AS0 - GSP-MG				Deformation (+ coef.)	Deviation (Å)	Depth difference	Weight
delta_plus	backbone	GLY	44	A	backbone	GLY	45		5.3% (5.62)	0.536	0.087	0.6
delta_plus	backbone	LYS	45	A	backbone	LYS	46		3.6% (5.67)	0.442	0.056	0.6
delta_minus		SER	46	A		SER	47		2.6% (6.33)	0.354	0.065	0.6
delta_plus	backbone	SER	46	A	backbone	SER	47		5.9% (6.90)	0.353	0.046	0.6
hydroxyl		SER	46	A		SER	47		2.7% (6.84)	0.354	0.065	0.65
delta_minus		THR	47	A		THR	48		1.8% (5.42)	0.151	0.078	0.6
delta_plus	backbone	THR	47	A	backbone	THR	48		2.9% (6.85)	0.443	0.042	0.6
hydroxyl		THR	47	A		THR	48		1.8% (5.86)	0.151	0.078	0.65
acyl		ASP	178	A		ASP	200		6.5% (8.91)	0.840	0.015	0.75

[PDB file] | [RasMol] | [PDB] | [PDBsum]

1B0U

TRANSPORT PROTEIN - ATP-BINDING SUBUNIT OF THE HISTIDINE PERMEASE FROM SALMONELLA TYPHIMURIUM

Matched annotations	Weighted number of SuMo groups	Number of SuMo groups	Volume
ATP binding site [automatic] [details]	42% (4.9 / 11.75)	44% (8 / 18)	37% (3.25 / 8.76)

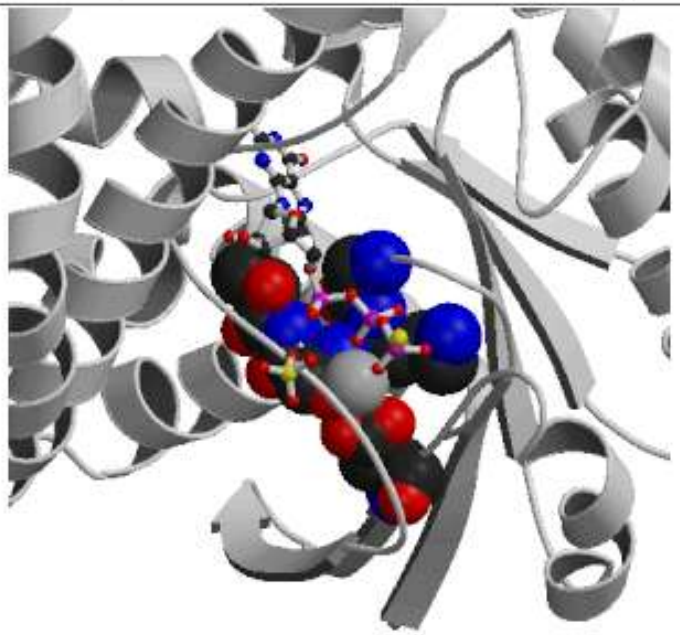


[PDB file] | [RasMol] | [PDB] | [PDBsum]

1AS0 - GSP-MG

SIGNAL TRANSDUCTION - GTP-GAMMA-S BOUND G42V GIA1

Matched annotations	Weighted number of SuMo groups	Number of SuMo groups	Volume
GSP-MG binding site [automatic] [details]	22% (5.65 / 25.65)	22% (9 / 41)	21% (4.12 / 19.77)



Images are dynamically generated using MolScript.

Usefulness

- Different folds, similar sites
- Known ligand & site?
  - estimation of cross-reactivity
  - search for functional families
- Unkown function?
  - prediction of ligand binding sites on drug targets
  - structural genomics: search for protein activity

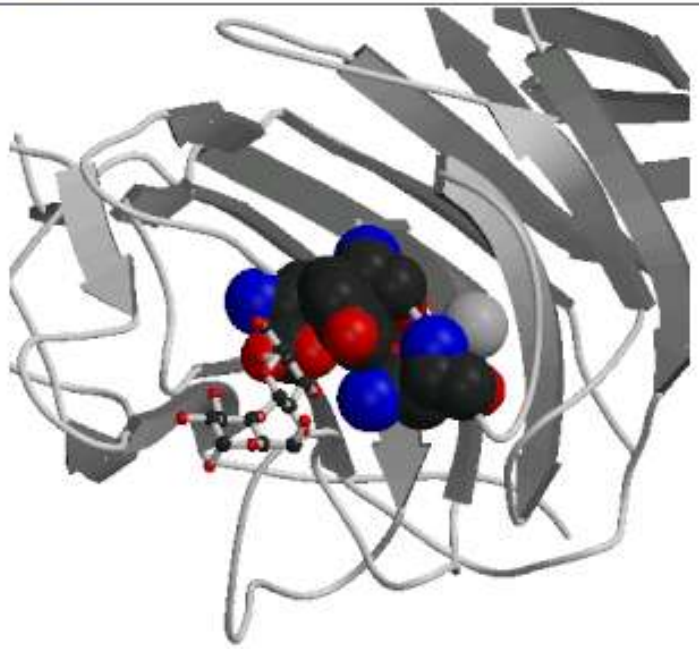
	2PEL				1V6V - XYS-XYS-XYS				Deformation (+ coef.)	Deviation (Å)	Depth difference	Weight
delta_minus	d1	ASP	83	A	d2	ASP	325	A	3.8% (2.52)	0.320	0.004	0.6
aromatic	ǂ1	TYR	125	A	ǂ1	HIS	343	A	10% (3.43)	0.329	0.022	0.9
amide		ASN	127	A		ASN	347	A	5.3% (4.68)	0.332	0.015	0.75

[PDB file] | [RasMol] | [PDB] | [PDBsum]

2PEL

LECTIN (AGGLUTININ) - PEANUT LECTIN

Matched annotations	Weighted number of SuMo groups	Number of SuMo groups	Volume
CA 1/4 binding site [automatic] [details]	100% (0.75 / 0.75)	100% (1 / 1)	100% (0.75 / 0.75)
LAT 1/4 binding site [automatic] [details]	27% (2.25 / 8.45)	23% (3 / 13)	34% (2.25 / 6.53)

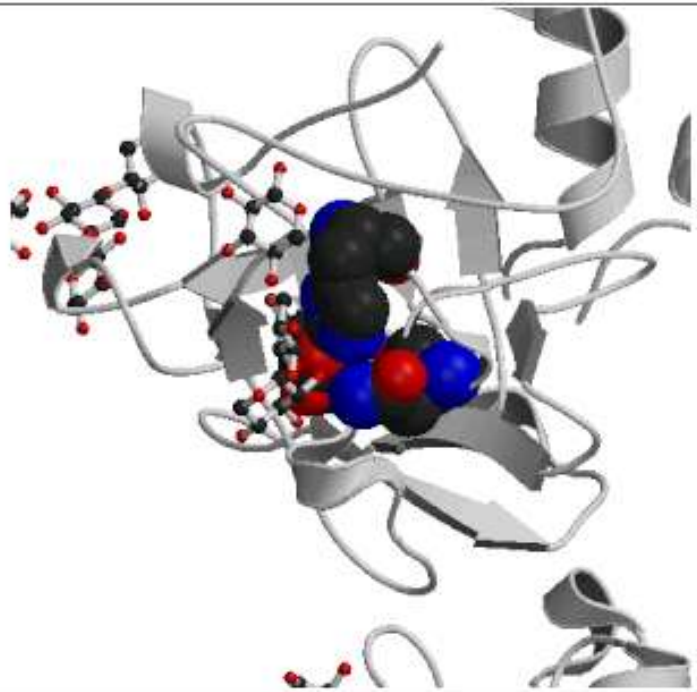


[PDB file] | [RasMol] | [PDB] | [PDBsum]

1V6V - XYS-XYS-XYS

HYDROLASE - CRYSTAL STRUCTURE OF XYLANASE FROM STREPTOMYCES OLIVACEOVIRIDIS E-86 COMPLEXED WITH 3(2)-ALPHA-L- ARABINOFURANOSYL-XYLOTRIOSE

Matched annotations	Weighted number of SuMo groups	Number of SuMo groups	Volume
XYS-XYS-XYS binding site [automatic] [details]	20% (2.25 / 11)	19% (3 / 16)	24% (2.25 / 9.34)



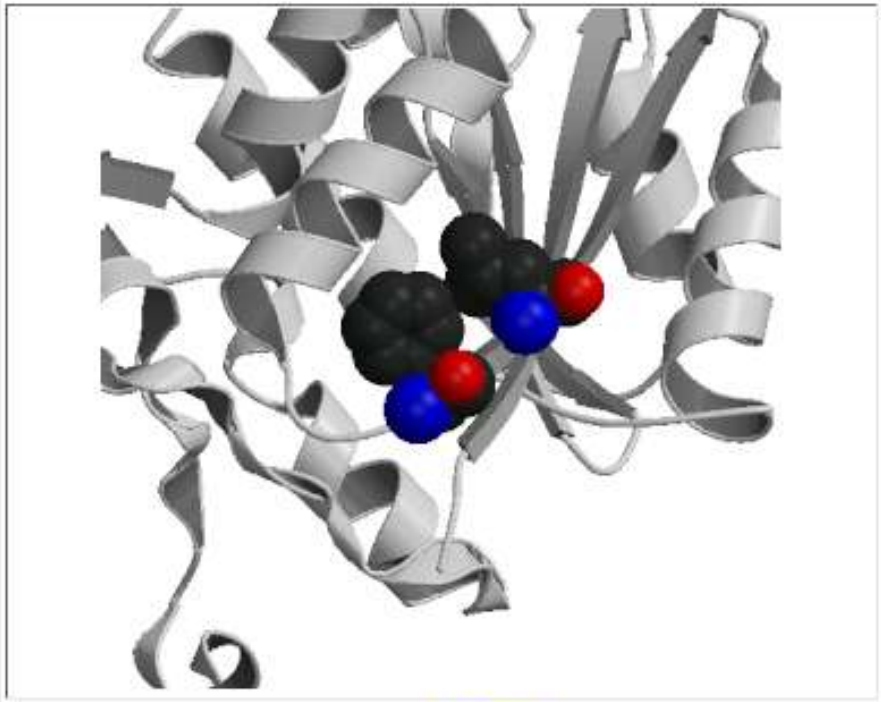
Images are dynamically generated using MolScript.

	1DI6				1PTO - SIA-GAL				Deformation (+ coef.)	Deviation (Å)	Depth difference	Weight
aromatic	♯1	PHE	40	A	♯2	TYR	102	C	4.9% (4.72)	0.152	0.063	0.9
delta_minus	backbone	PHE	40	A	backbone	TYR	102	C	3.1% (3.42)	0.223	0.081	0.6
delta_plus	backbone	PHE	40	A	backbone	TYR	102	C	1.4% (3.10)	0.032	0.070	0.6
delta_plus	backbone	LEU	42	A	backbone	SER	104	C	0.88% (2.64)	0.056	0.097	0.6

[PDB file] | [RasMol] | [PDB] | [PDBsum]

1DI6

UNKNOWN FUNCTION - 1.45 A CRYSTAL STRUCTURE OF THE MOLYBDENUMM COFACTOR BIOSYNTHESIS PROTEIN MOGA FROM ESCHERICHIA COLI

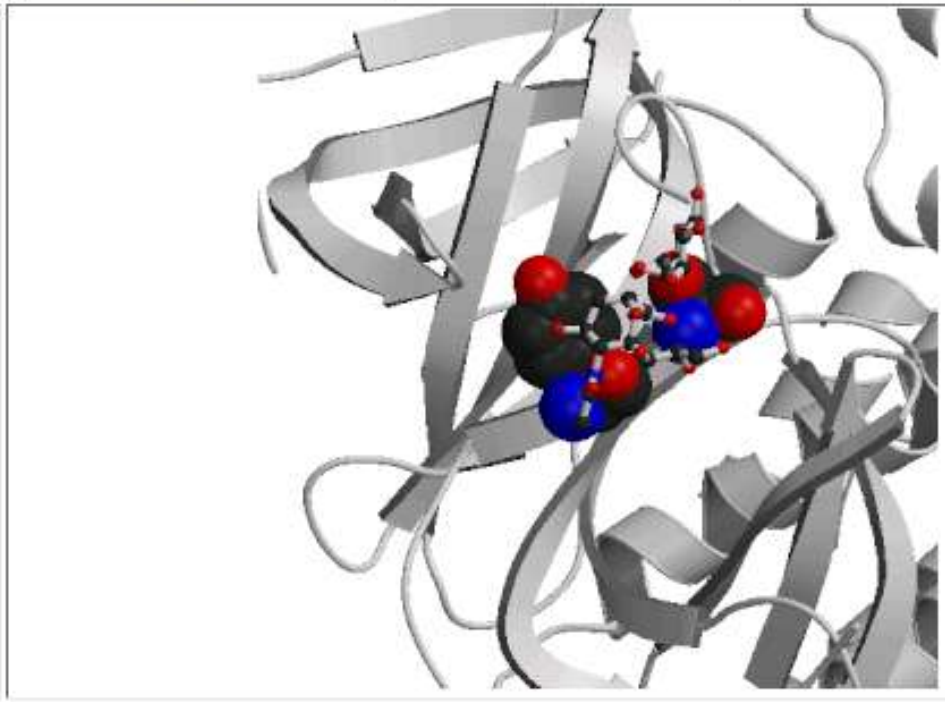


[PDB file] | [RasMol] | [PDB] | [PDBsum]

1PTO - SIA-GAL

TOXIN - MOL_ID: 1; MOLECULE: PERTUSSIS TOXIN; CHAIN: A, B, C, D, E, F, G, H, I, J, K, L; HETEROGEN: SIALIC ACID; CHAIN: B, C, I; HETEROGEN: D-GALACTOSE; CHAIN: B, C, I; OTHER_DETAILS: SACCHARIDE CONTAINS TERMINAL N-ACETYLNEURAMINIC ACID (ALPHA 2,6) GALACTOSE

Matched annotations	Weighted number of SuMo groups	Number of SuMo groups	Volume
SIA-GAL 2/3 binding site [automatic] [details]	25% (2.7 / 10.6)	25% (4 / 16)	30% (2.70 / 9.15)



Images are dynamically generated using MolScript.

Problems and limits

Ligand binding elements

Comparison algorithm

Examples

Conclusions

Practical information & conclusions

- Implementation
  - in a concise, safe and fast programming language: Objective Caml (OCaml)
- Availability
  - public: http://sumo-pbil.ibcp.fr
  - private: http://www.medit.fr
- References in the field
  - http://martin.jambon.free.fr/search-protein-3D-sites.html
- See my ISMB poster #A-82