M1 ISDD (Feb 2018)

PROTEIN DOCKING

Lesson 6

Interface prediction, hot-spots and biomedical applications

Juan Fernández-Recio

CSIC, BSC

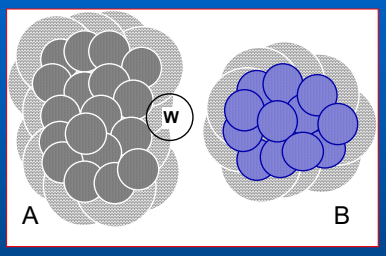
juanf@bsc.es

- Interface prediction
- Hot-spot identification
- Application to drug design targeting PPIs
- Pathological mutations affecting PPIs

Protein-protein interface size

Defining macromolecular interfaces based on solvent accessibility

ASA accessible surface area
measures molecule-solvent contacts with the rolling ball algorithm
Lee & Richards, JMB,1971



Protein-protein interface size

Defining macromolecular interfaces based on solvent accessibility

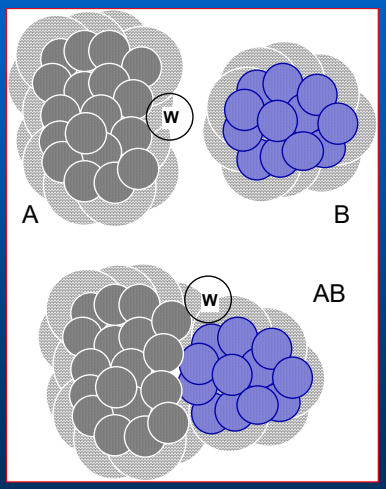
$$BSA = ASA_A + ASA_B - ASA_{AB}$$

ASA accessible surface area
measures molecule-solvent contacts with the rolling ball algorithm
Lee & Richards, JMB,1971

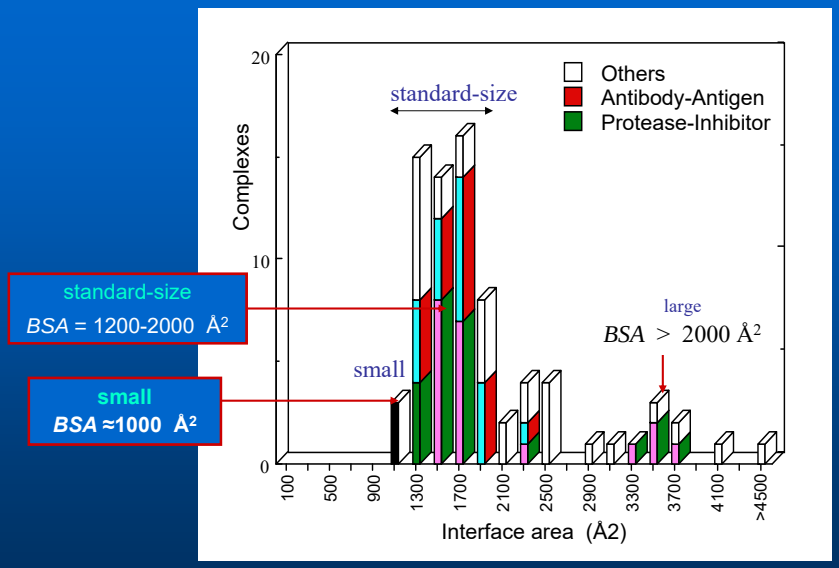
BSA buried surface area
measures molecule-molecule contacts
Chothia & Janin, Nature, 1975

Interface atoms or residues
are all atoms or residues that contribute to the *BSA*. On average, each interface atom contributes ≈ 10 Å^2

Hydrophobic effect

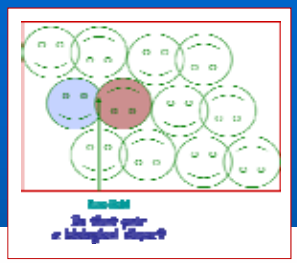


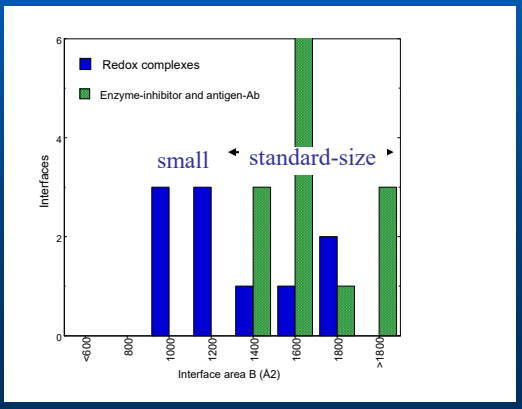
Protein-protein interface size

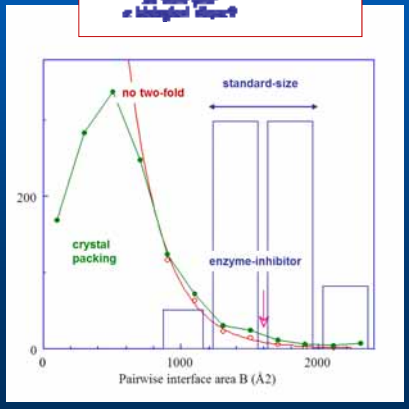


Protein-protein interface size

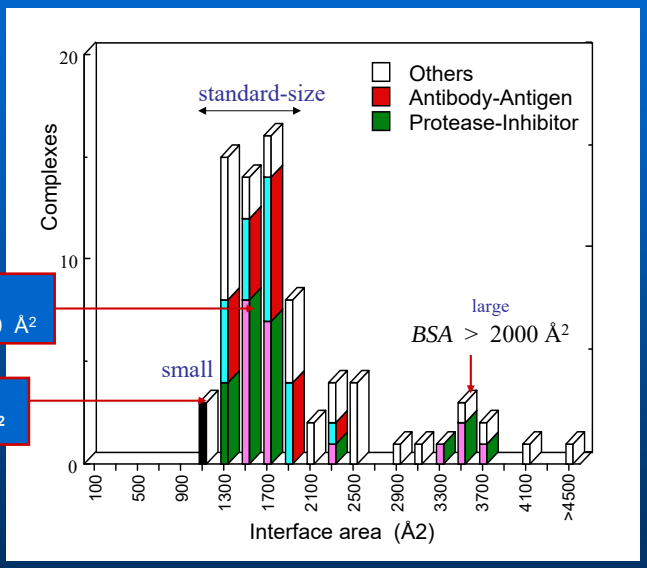
Electron transfer complexes
have small interfaces
BSA = 900 1200 Å²
0 3 H bonds
Crawley & Carrondo (2004)







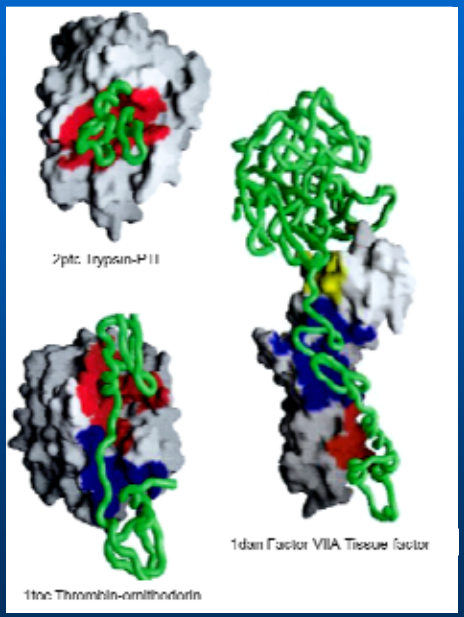
Protein-protein interface size



standard-size
BSA = 1200-2000 Å²

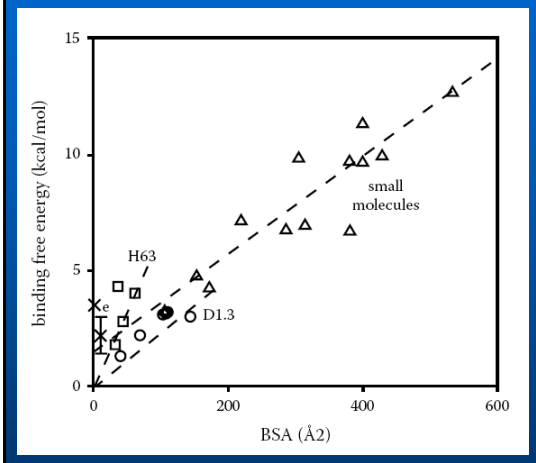
small
***BSA* ≈1000 Å²**

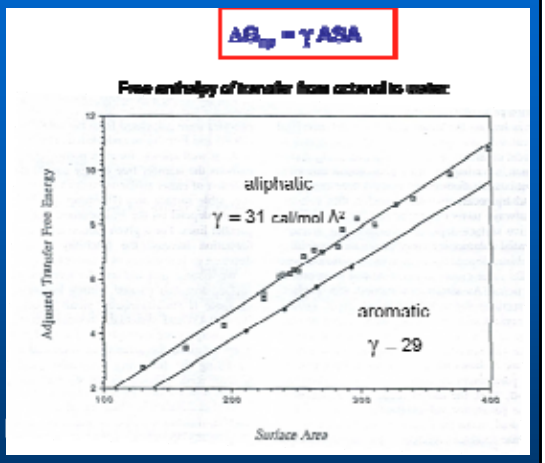
Protein-protein interface size



Large interfaces are multi-patch

Protein-protein interface size and desolvation energy





Protein-protein interface prediction

Common Descriptors Used for Prediction of Interaction Sites

Descriptor	Comments
Features Derived from 3-D Structure	
Neighbor list: Residues in spatial vicinity to the residue in question	9–20 residues
B-factor	A crystallographic measure that approximates the flexibility of a residue.
Solvent accessibility (ASA)	Measured in Å².
Relative solvent accessibility	Measured as a fraction of the overall surface of the residue that is exposed to solvent.
Shape index/curvedness	
Secondary structure	Three state (helix, strand, loop) or more.
Sequence distance	The separation in sequence between residues within the same patch. Some results indicate that structurally contiguous residues that are not adjacent in sequence are more likely to form interaction sites.
Planarity	
Predicted/Approximate Structural Features	
Predicted secondary structure	Methods relying only on sequence can use computational tools to generate predicted solvent accessibility/secondary structure. This improves performance without limiting applicability to proteins with known 3-D structure.
Predicted solvent accessibility	
Sequence neighbor list	Can be used instead of neighbor list to approximate the environment of the analyzed residue. Nine to fifteen residues around the residue in question. Four to seven on each side of the residue. Some structure-based methods use this in addition to neighbor list.

Protein-protein interface prediction

Common Descriptors Used for Prediction of Interaction Sites

Descriptor	Comments
	Evolutionary Features
Sequence profile	Extracted from a multiple sequence alignment, a profile reveals patterns of evolutionary conservation.
Conservation score	A quantification of the level of conservation of an individual position.
Conservation of physicochemical traits	If the position is not conserved, scoring conservation of traits such as charge, hydrophobicity, or size may improve prediction.
	Physicochemical Features
Hydrophobicity	Several different scales are available.
Electrostatic potential	Measured for individual residue or for a patch. Requires 3-D structure.
Atom propensities	Serves as a way to sum physicochemical properties across residues in the patch.
Desolvation energy	Used mostly in predictions for rigid-body docking.
	External Knowledge
Protein–protein interaction	Can be used to: (1) identify sequence or structural elements that are significantly overrepresented in interacting pairs, and (2) to assess coevolution of positions in interacting pairs.
Functional annotation of the protein	Enzyme–inhibitor and antigen–antibody have different types of interfaces than other complexes. Adding this information may improve prediction.

$\Delta G_{ij} = \gamma ASA$

Protein-protein interface prediction

Method name	Input data	Method	Details	Web
ISIS[58]	sequence	Neural network	Predicted structural features, evolutionary information	http://cubic.bioc.columbia.edu/services/isis/
TreeDet [71]	sequence, structure	Scoring function	sequence and structural alignments	http://treedetv2.bioinfo.cnio.es/treedet/index.html
Promate[73]	structure	Scoring function	Secondary structure, sequence conservation, residue type	http://bioinfo41.weizmann.ac.il/promate/
PINUP[76]	structure	Scoring function	side-chain energy score, propensity, sequence conservation	http://sparks.informatics.iupui.edu/PINUP/
InterProSurf[59]	structure	Scoring function	solvent accessibility, propensities	http://curie.utmb.edu/
PRISM[77]	structure	Scoring function	geometric complementarity, conservation	http://prism.ccbb.ku.edu.tr/prism/
ConSurf[68]	structure	Scoring function	conservation	http://consurf.tau.ac.il/
ET[66]	structure	Scoring function	multiple sequence alignments	http://mammoth.bcm.tmc.edu/traceview/
JET[70]	structure	Scoring function	structural and functional conservation	http://www.ihes.fr/~carbone/data.htm
WHISCY[79]	structure	Scoring function	conservation, surface properties	http://www.nmr.chem.uu.nl/Software/whiscy/startpage.htm
PIER[61]	structure	Scoring function	atomic statistical propensities	http://abagyan.ucsd.edu/PIER/

Protein-protein interface prediction

Method name	Input data	Method	Details	Web
SiteEngines[60]	structure	Hierarchical scoring function	structural matching, physico-chemical properties	http://bioinfo3d.cs.tau.ac.il/SiteEngine/
PPI-Pred[84]	structure	SVM	surface shape, electrostatic potential	http://bioinformatics.leeds.ac.uk/ppi-pred
cons-PPISP[80,81]	structure	Neural network	PSI-Blast sequence profile and solvent accessibility	http://pipe.scs.fsu.edu/ppisp.html
SPPIDER[85]	structure	Neural Network	solvent accessibility and other features	http://sppider.cchmc.org/
Patch Finder Plus[82]	structure	Neural Network	conservation, concavity, area, H-bond, residue frequency	http://pfp.technion.ac.il/
meta-PPISP[87]	structure	Meta web server	cons- PPISP, Promate and PINUP	http://pipe.scs.fsu.edu/meta-ppisp.html
PI²PE[88]	structure	Meta web server	cons-PPISP, WESA, DISPLAR	http://pipe.scs.fsu.edu/
SHARP[290]	structure	Energy-based, scoring function	Desolvation, hydrophobicity, ASA, propensity, surface shape	http://www.bioinformatics.sussex.ac.uk/SHARP2/sharp2.html
ODA[62]	structure	Energy-based	Desolvation energy	http://www.molsoft.com/oda.html
NIP[100]	structure	Energy-based	Docking simulations	https://life.bsc.es/pid/pydock/

Promate

http://bioinfo41.weizmann.ac.il/promate/

Promate

http://bioinfo41.weizmann.ac.il/promate/

Please choose the scores configuration: ◉ use default ○ let me choose

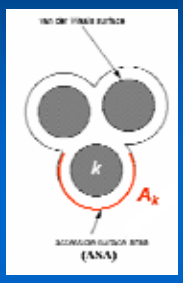
Single amino acids distribution:	○ use	◉ don't use
Atoms distribution:	◉ use	○ don't use
Chemical character:	◉ use	○ don't use
Amino acid pairs distribution:	◉ use	○ don't use
Evolutionary conserved positions:	◉ use	○ don't use
Non-regular secondary structure length:	◉ use	○ don't use
Sequence distances within a circle:	◉ use	○ don't use
Secondary structure:	◉ use	○ don't use
Domains:	○ use	◉ don't use
Hydrophobic patch rank:	◉ use	○ don't use
Hydrophobic patch size:	○ use	◉ don't use
Temperature factor (B-factor):	○ use	◉ don't use
Water molecules:	◉ use	○ don't use
Initial probabilities file:	○ use	◉ don't use

Please select the output files to be produced:

- ☐ A text file with the surface dots probabilities
- ☑ A pdb file with the AAs colored by their interface probability (full range of colors)
- ☑ A pdb file with the surface atoms colored by their interface probability (full range of colors)
- ☑ A pdb file with the predicted interface patch colored in red

Optimal Docking Areas (ODA)

Desolvation Energy for Protein-Protein Docking



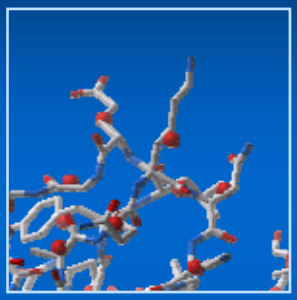
$$\Delta G_{sol} = \sum_{k=1}^{N} \sigma_k A_k$$

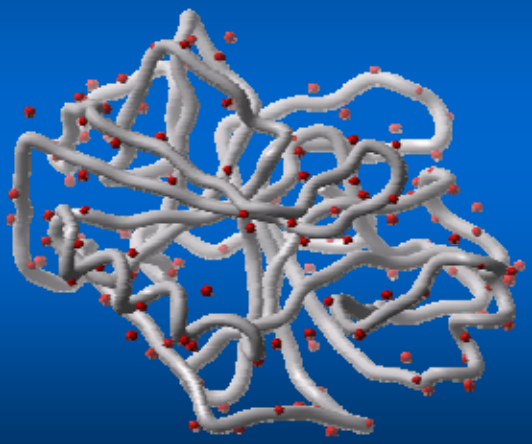
ASP are optimized for protein docking

ASPs from vacuum/water → ASPs from octanol/water → ASPs for interface/water

σ (cal/(mol Å²))	Radius (Å)	Atom type
15.1	1.95	C aliphatic
17.7	1.8	C aromatic
−17.0	1.7	N uncharged
−54.8	1.7	N^{+}, N^{ζ} in Lys^{+}
−27.3	1.7	$N^{\eta 1}$, $N^{\eta 2}$ in Arg^{+}
−18.5	1.6	O hydroxyl
−13.6	1.4	O carbonyl
−29.9	1.4	O^{-} in Glu, Asp
11.2	2.0	S in SH
2.2	1.85	S in Met or S–S

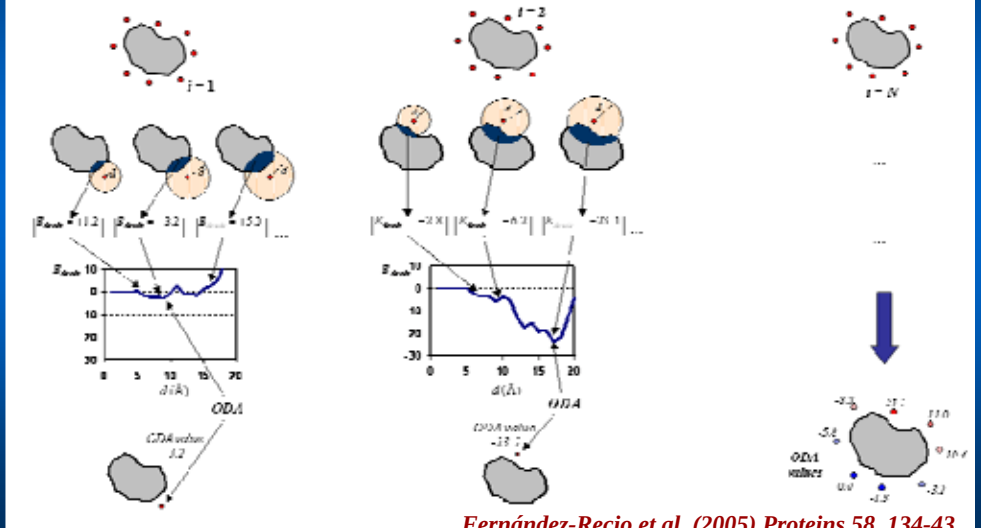
Optimal Docking Areas (ODA)





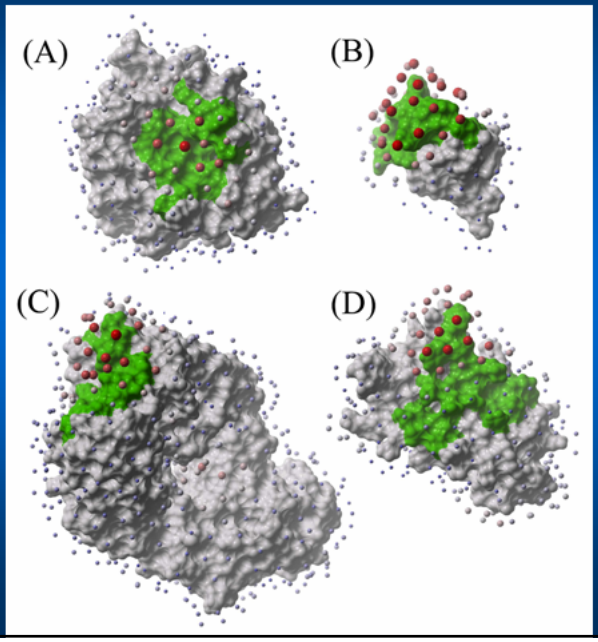
Optimal Docking Areas (ODA)

$$\Delta E_j^{desolv} \approx ASA_j \sigma_j \qquad \Delta E_i^{ODA} = Min\left(\sum_j^N \Delta E_j^{desolv}\right)$$



Fernández-Recio et al. (2005) Proteins 58, 134-43

Optimal Docking Areas (ODA)



Benchmark: 66 unbound cases

- ODAs in 50% proteins

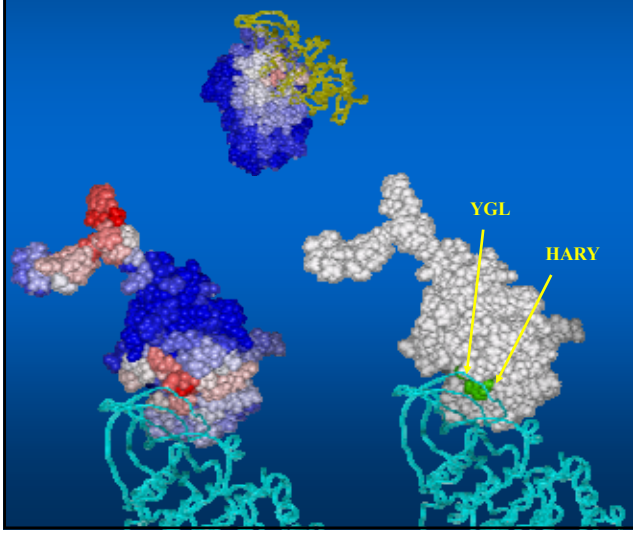
- 80% correctly located in binding sites

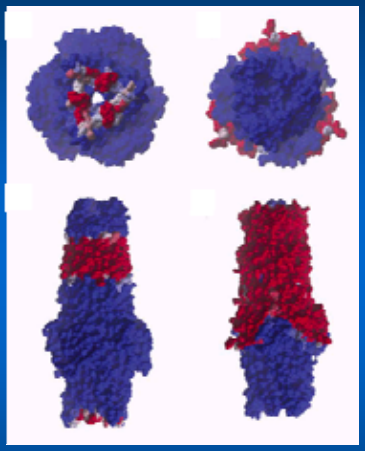
- pyDockODA (centered in residues);

Fernández-Recio et al. (2005) Proteins 58, 134-143

ODA - applications

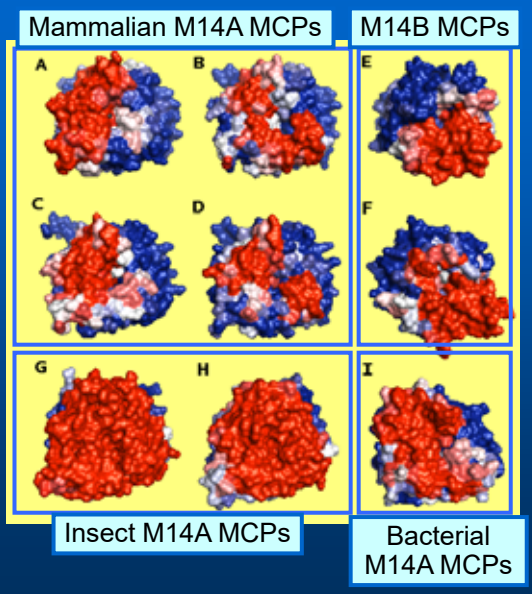
Bolanos-García, Fernández-Recio et al. TiBS (2006) 31, 654

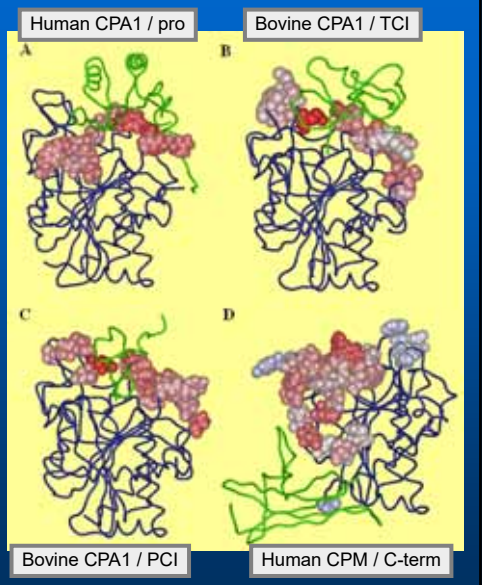




Federici, …, Fernández-Recio et al. (2005) J.Biol.Chem. 280, 15307-14

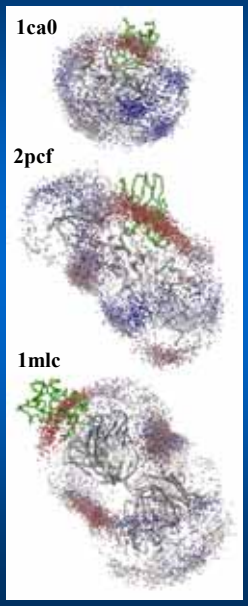
ODA - applications





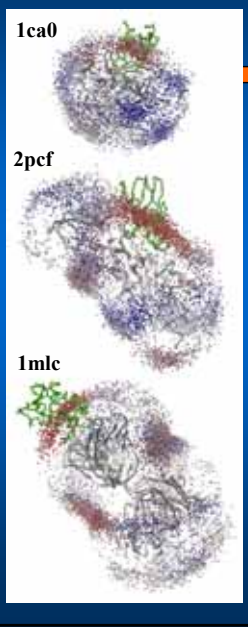
Fernández… & Fernández-Recio (2007) Proteins 68, 131

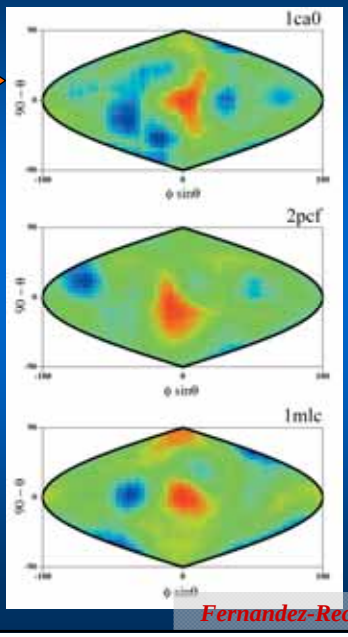
Interface predictions from docking results



Fernandez-Recio et al. (2004) JMB 335, 843-865

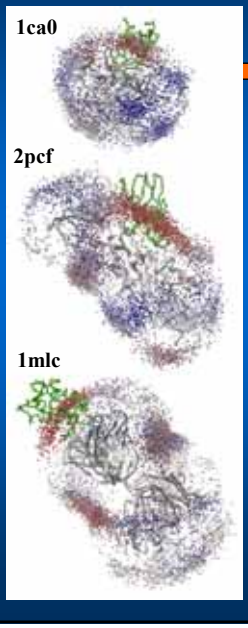
Interface predictions from docking results

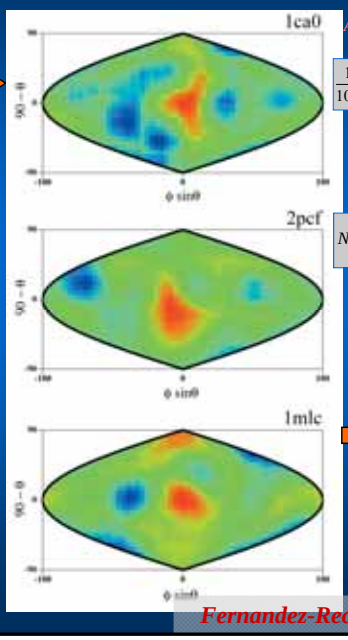




Fernandez-Recio et al. (2004) JMB 335, 843-865

Interface predictions from docking results



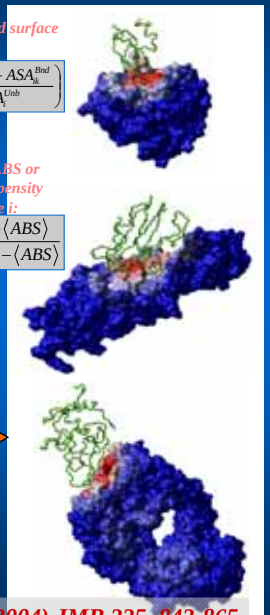


Averaged buried surface (ABS)

$$\frac{1}{100}\sum_{k=1}^{100}\left(\frac{ASA_i^{Unb} - ASA_{ik}^{Bnd}}{ASA_i^{Unb}}\right)$$

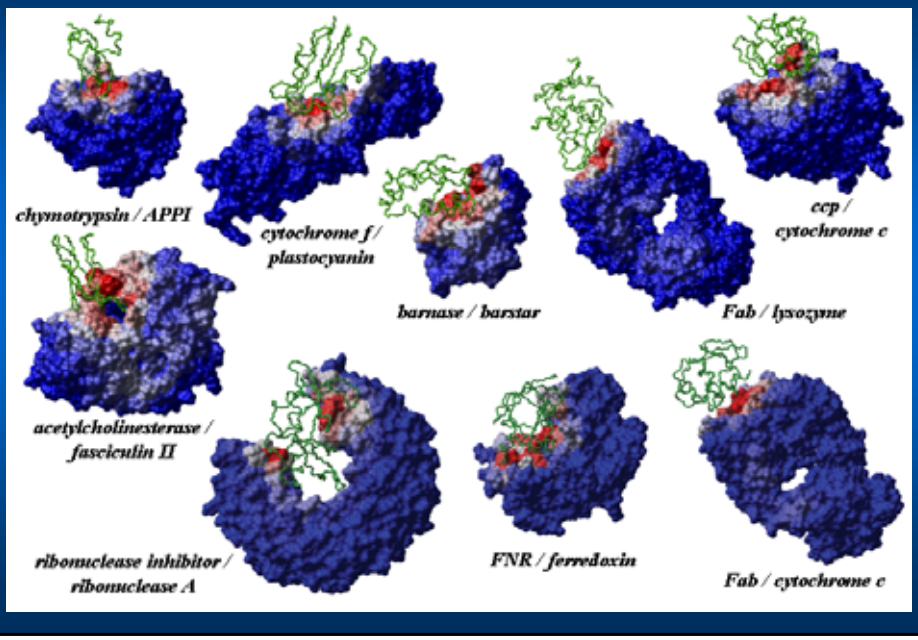
Normalized ABS or Interface Propensity for residue i:

$$NIP_i = \frac{ABS_i - \langle ABS \rangle}{ABS^{MAX} - \langle ABS \rangle}$$

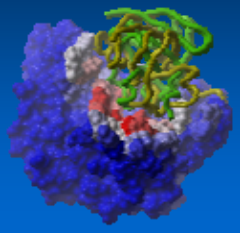


Fernandez-Recio et al. (2004) JMB 335, 843-865

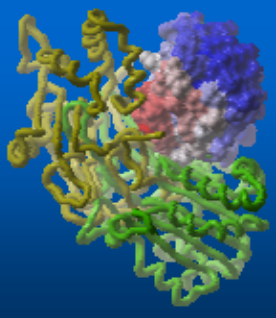
Interface predictions from docking results



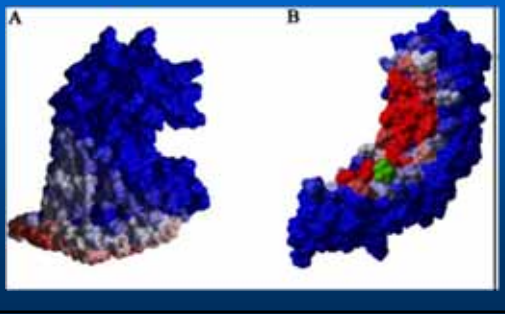
Interface predictions from docking results



Medina, Abagyan, Gómez-Moreno, Fernández-Recio (2008) Proteins 72, 848



Sicilia et al. (2005) Plant Physiology 578, 5
Federici et al., Trends in Plant Sciences (2006)



- Interface prediction
- **Hot-spot identification**
- Application to drug design targeting PPIs
- Pathological mutations affecting PPIs

Hot-Spots

The **O-ring model**: alanine-scanning **hotspots** tend to be located at the center of the interface and surrounded by energetically unimportant residues that occlude solvent from them.

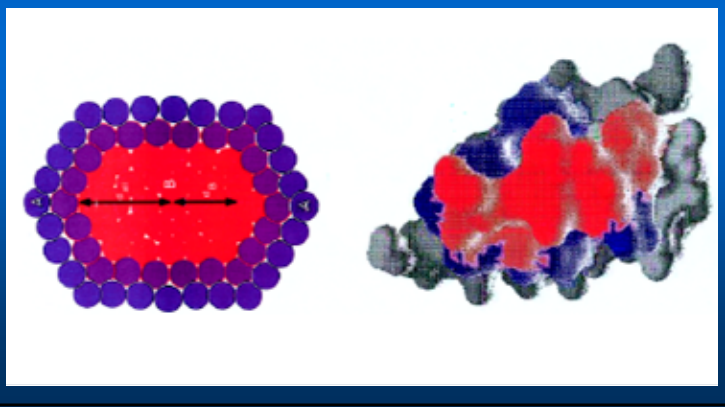
Bogan & Thorn (1998)

A sketch of the **solvent accessible** and **fully buried** atoms in a standard-size protein-protein interface. Each protein contributes about 70 atoms, and 1/3 are buried.

Lo Conte, Chothia & Janin (1999)

The **core/rim model** is a realistic implementation of that sketch. (Chakrabarti & Janin 2002)

The **core** and **rim** of the CI2 inhibitor interface with subtilisin (1sni)

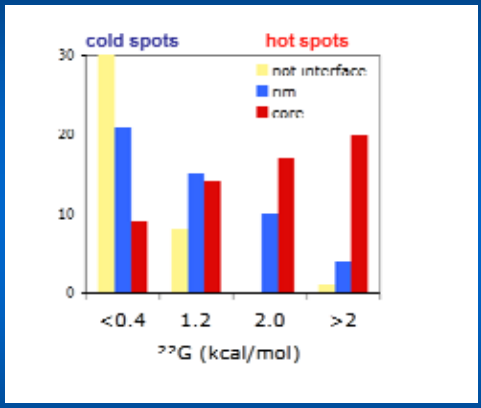


Hot-Spots

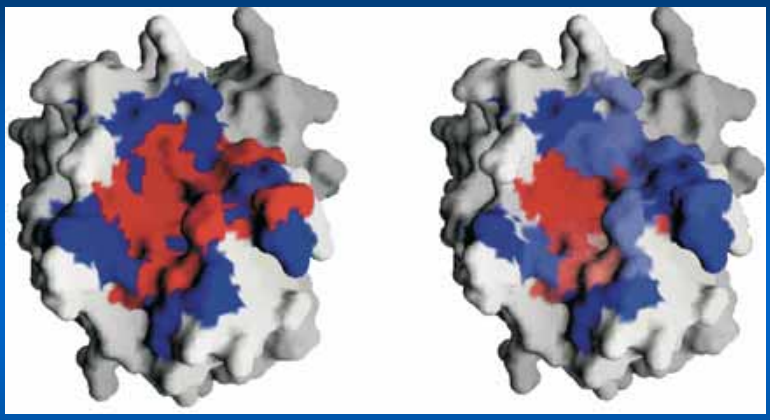
Alanine-scanning experiments

- The residues at the interface of a protein-protein complex are mutated to Ala one by one.
- The dissociation constant K'_d of the mutated complex is compared K_d of the wild type.
- The mutation changes the free enthalpy of dissociation by:

$$\Delta G_d = RT \ln K'_d/K_d$$



Hot-Spots



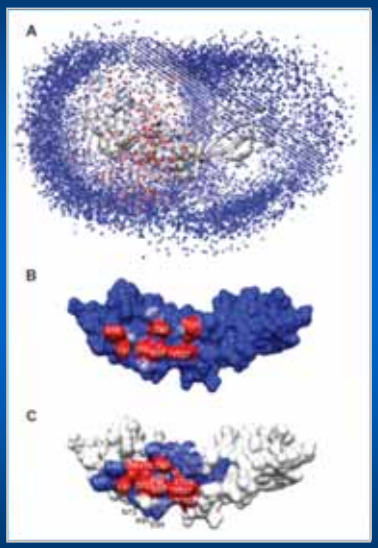
Core and **rim** in the recognition patch of the ovomucoid inhibitor (1cho)

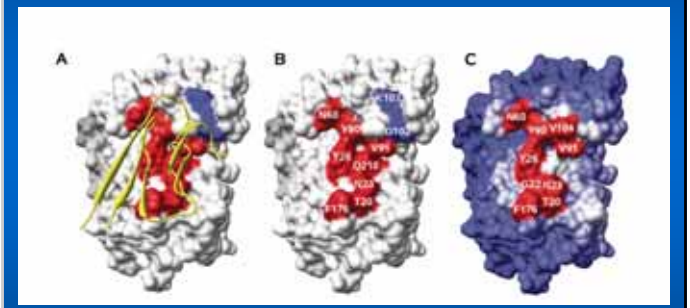
S=0 (red) fully conserved residues.
S>1 (purple) variable residues

Prediction of Binding Hot-Spots

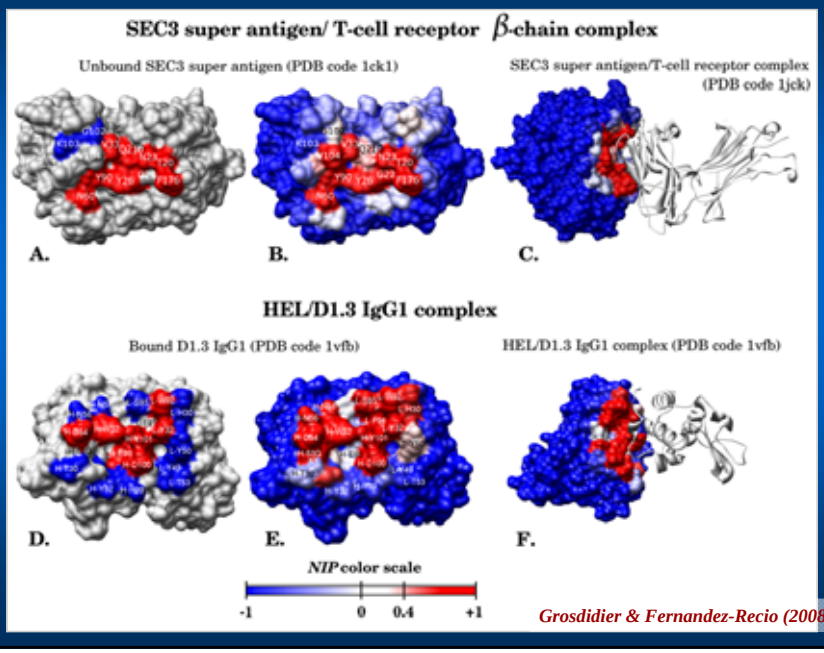
Method name	Input data	Method	Details	Sensitivity	PPV	availability
ISIS[58]	sequence	Neural network	Predicted structural features, evolutionary information	15%	89%	http://cubic.bioc.columbia.edu/services/isis/
FOLDEF[126]	complex structure	Energy-based	Alanine scanning	45-72%[a]	61-73%[a]	http://foldx.crg.es/
ROBETTA[122]	complex structure	Energy-based	Alanine scanning	28-69%[b]	60-71%[b]	http://robetta.org/submit.jsp
K-FADE[125]/ K-CON/ ROBETTA	complex structure	Machine learning algorithm	Physical-biochemical features	48%	53%	http://kfc.mitchell-lab.org
MAPPIS[119]	complex structure	Evolutionary conservation	Multiple alignments, 3D clustering	66%	63%	http://bioinfo3d.cs.tau.ac.il/MAPPIS
HotPoint[121]	complex structure	Empirical model	Accessibility, knowledge-based potentials	59%	70%	http://prism.ccbb.ku.edu.tr/hotpoint
pyDockNIP[127]	unbound protein structure	Energy-based	Docking simulations	42-43%	68-75%	http://mmb.pcb.ub.es/PyDock

Prediction of Binding Hot-Spots



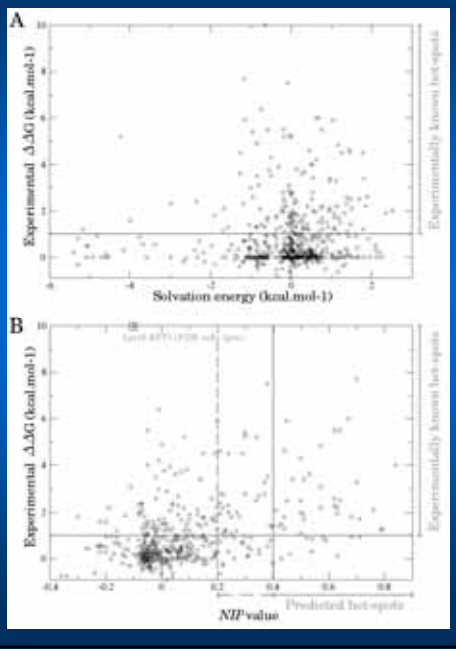


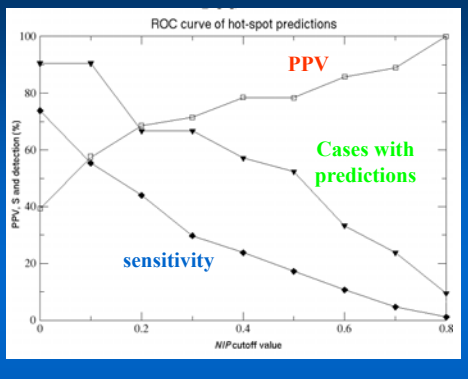
Prediction of Binding Hot-Spots



Grosdidier & Fernandez-Recio (2008) BMC Bioinf

Prediction of Binding Hot-Spots





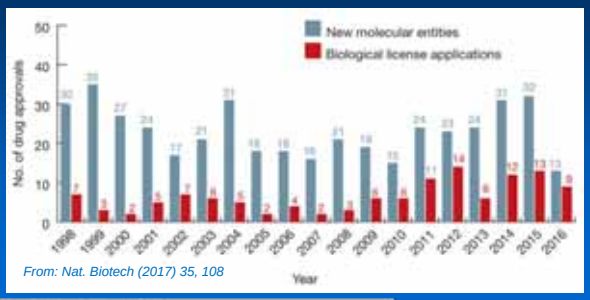
Hot-spot prediction benchmark

	PPV	*S*
Standard benchmark (21 cases)		
NIP ≥ 0.2	68%	43%
NIP ≥ 0.4	78%	24%
FOLDEF	73%	46%
ROBETTA[b]	71%	69%
Benchmark on modeled subunits (11 cases)		
NIP ≥ 0.2	59%	33%
NIP ≥ 0.4	75%	15%

Grosdidier & Fernandez-Recio (2008) BMC Bioinf

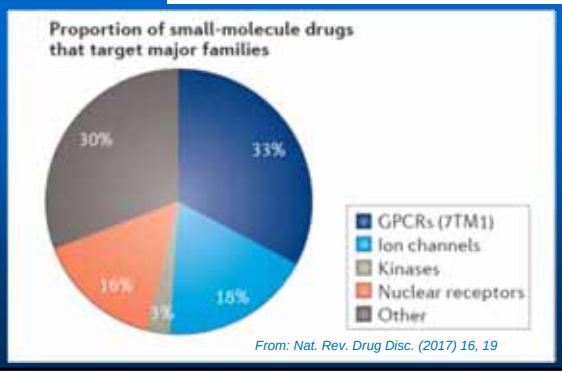
- Interface prediction
- Hot-spot identification
- **Application to drug design targeting PPIs**
- Pathological mutations affecting PPIs

Current limitations in drug discovery



From: Nat. Biotech (2017) 35, 108

Drug discovery: new targets are strongly needed!



From: Nat. Rev. Drug Disc. (2017) 16, 19

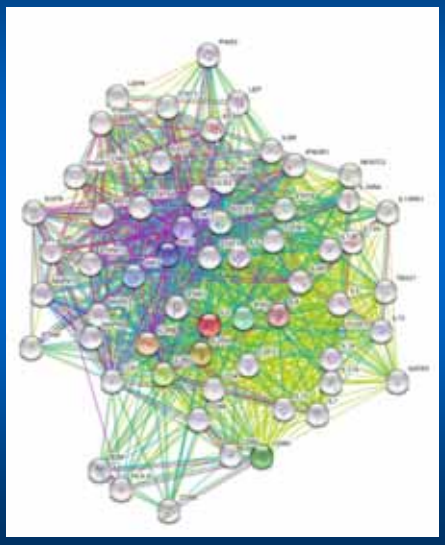
>70% FDA-approved drugs target only 4 gene families, basically cell-surface receptors and enzymes

Most of the available drugs target individual proteins

Only 10–15% of all human proteins are currently «druggable»

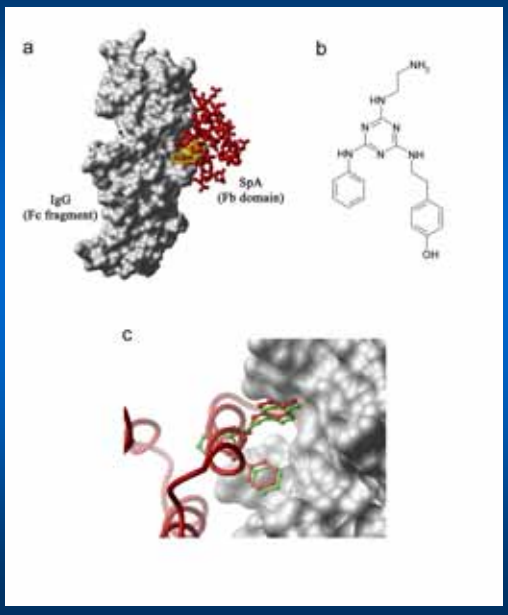
Importance of protein-protein interactions in biomedicine

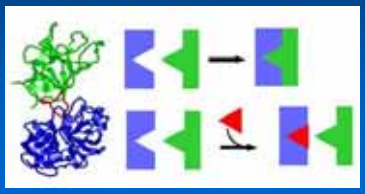
Drug discovery: new targets are strongly needed!



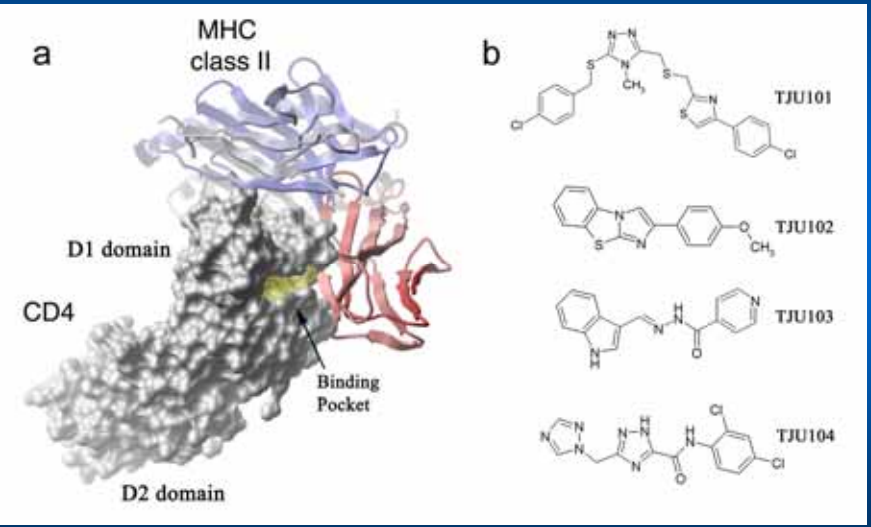
New approaches:
drugs targeting protein-protein interactions

Inhibiting protein-protein interactions

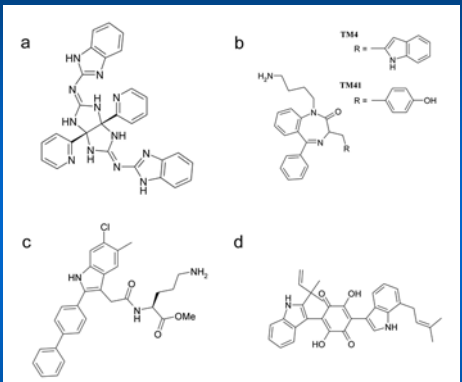




Inhibiting protein-protein interactions

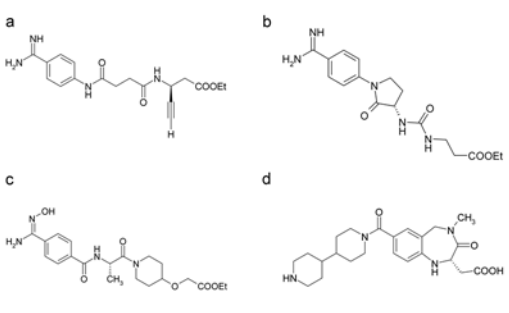


Inhibiting protein-protein interactions

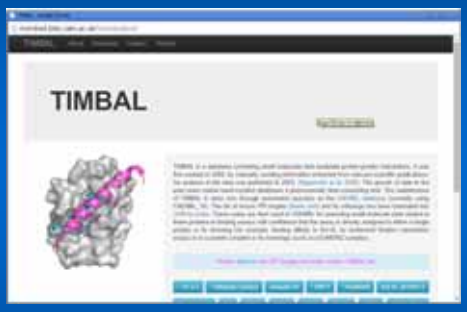


(a) Molecule SB 247464; (b) Molecules TM4 and TM41; (c) Inhibitor of EPOEPOR interaction; (d) Molecule L-783,281.

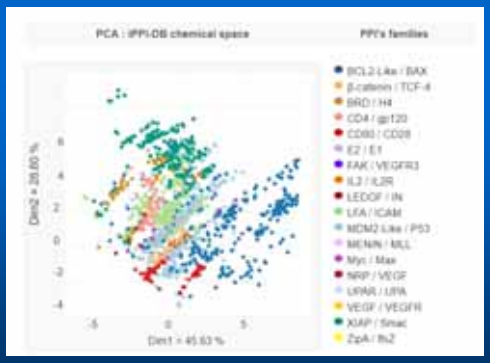
Antagonists of GPIIb-IIIa currently in clinical trials



Inhibiting protein-protein interactions



http://www.ippidb.cdithem.fr/

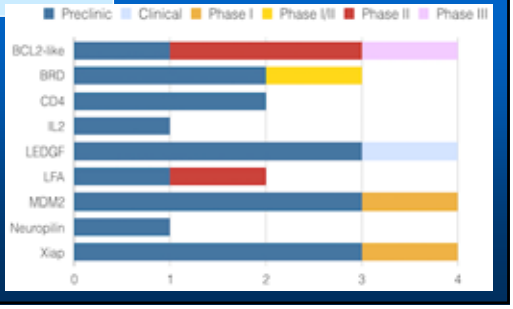


Inhibiting protein-protein interactions

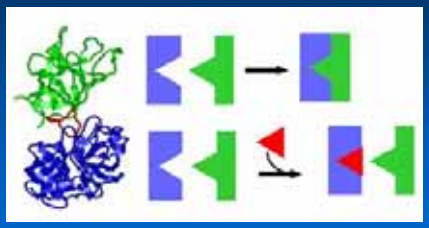
Drug discovery targeting protein interactions

PPI target	Disease	Clinical
BCL-2 family	Cancer	Phase III (disc.)
MDM2-p53	**Cancer**	**Phase III**
LFA1-ICAM1	**Dry eye**	**Pre-reg.**
αIIbβ3	**Cardiovascular**	**Approved**
αLβ2	Psoriasis	Phase II (disc.)
α4β1	**Ulcerative colitis**	**Phase III**
α5β1	Macular degeneration	Phase I (disc.)
IAP	**Cancer**	**Phase II**
Bromodomain family	**Cardiovascular, cancer**	**Phase III, II**

From: Nat. Rev. Drug Disc. (2016) 15, 533



Rational design of protein-protein inhibitors

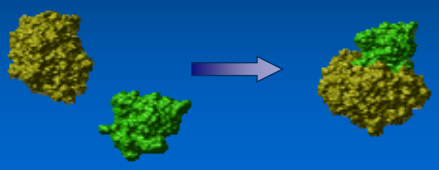


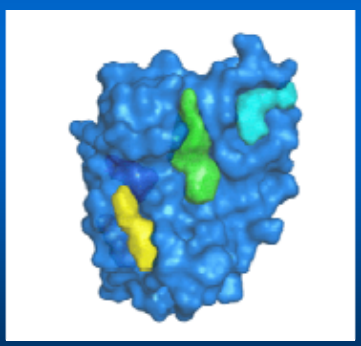
We need a suitable cavity

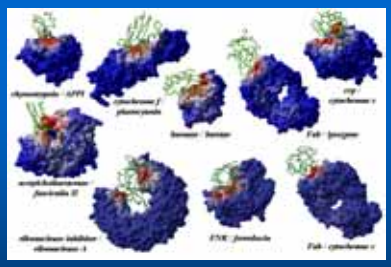
Rational design of protein-protein inhibitors

Major difficulties:

Where to search?
(In most cases protein-protein interface is not known)





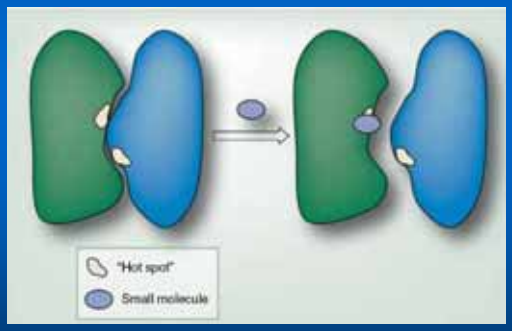


- Protein-protein docking
- Interface predictions

Rational design of protein-protein inhibitors

Major difficulties:

Absence of binding pockets in natural interfaces



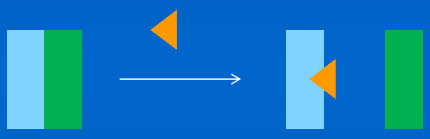
- In standard drug design:
  3D structure with bound ligands, known active site

- In PPI inhibition:
  no clear pockets (difficult to identify computationally)

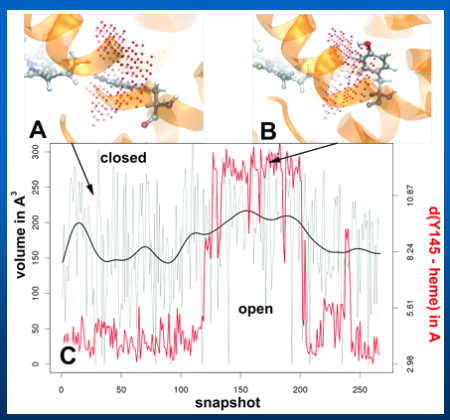
Rational design of protein-protein inhibitors

Major difficulties:

Absence of binding pockets in natural interfaces



Molecular Dynamics
+
Pocket predictors (Fpocket, CASTp,...)

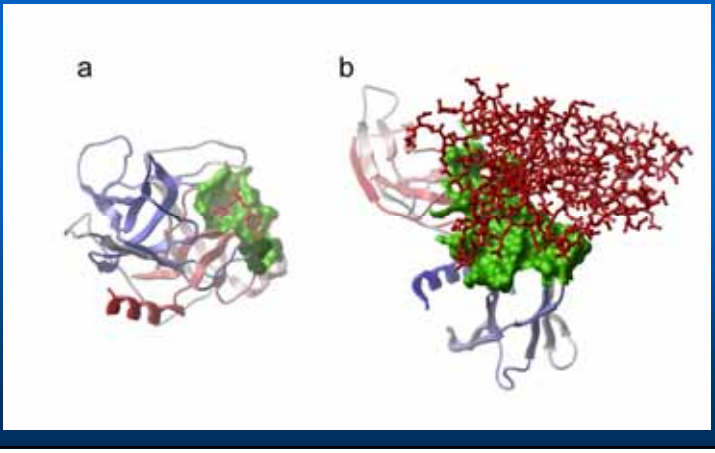


https://bioinformatictools.wordpress.com/tag/pocket-finder/

Rational design of protein-protein inhibitors

Major difficulties:

How to compete with a large interface?



Rational design of protein-protein inhibitors

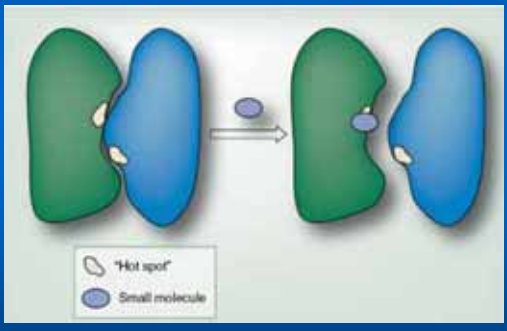
Major difficulties:

How to compete with a large interface?

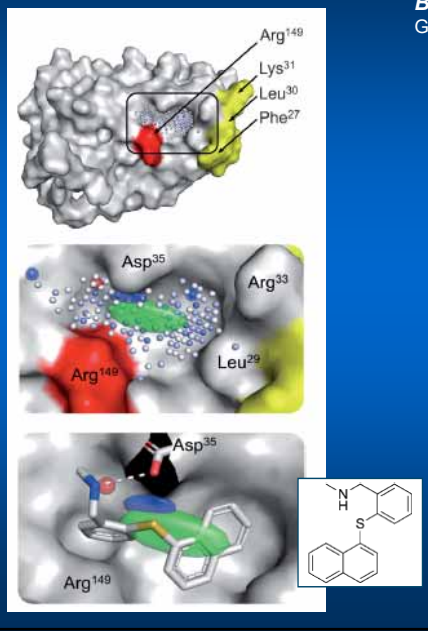
- Strategy: Targeting hot-spots.

- Hot-spot experimental identification:
  alanine-scanning (costly)

- Hot-spot computational prediction:
  ROBETTA, FOLDEF…
  (need 3D structure of complex)



Rational design of protein-protein inhibitors



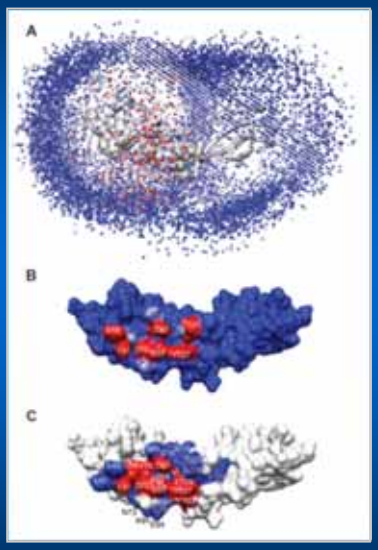
Blocking IFN-α / receptor
Geppert *et al.* (2012) *Angew.Chem*

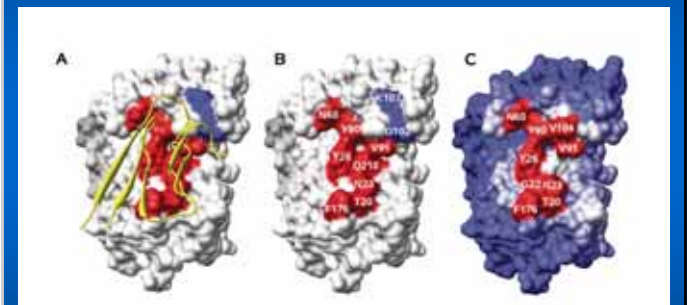
Predicted hot-spots (iPred)

Pharmacophore (VirtualLigand)

Docking pose (Gold)

Docking-based prediction of Binding Hot-Spots

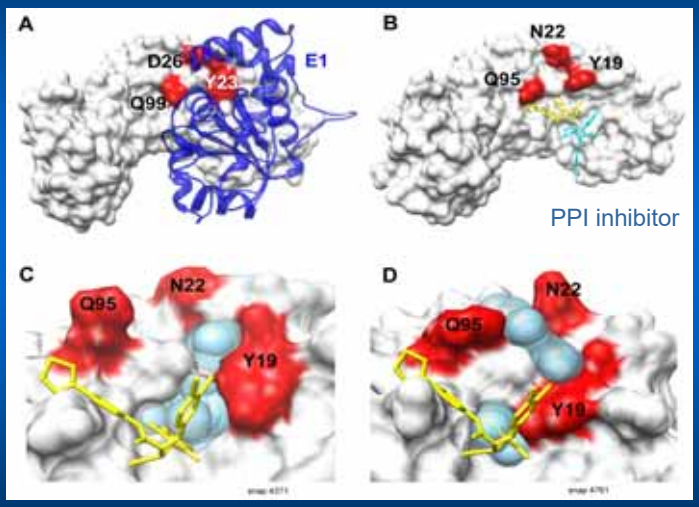




NIP **(no need complex structure)**

Transient cavities in interfaces for ligand docking

HPV E2 protein - predicted hot-pots



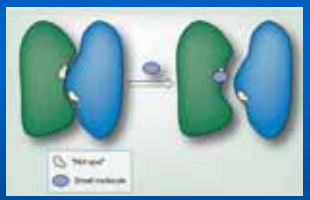
Combine MD, pocket identification and docking-based hot-spot predictions

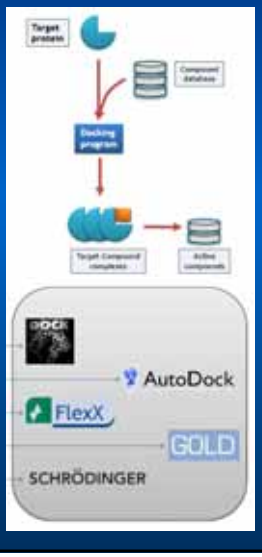
(no need complex structure)

Rational design of protein-protein inhibitors

Major difficulties:

Are current ligand libraries appropriate?





PPI-HitProfiler

http://www.cdithem.fr/getPPIHitProfiler.php

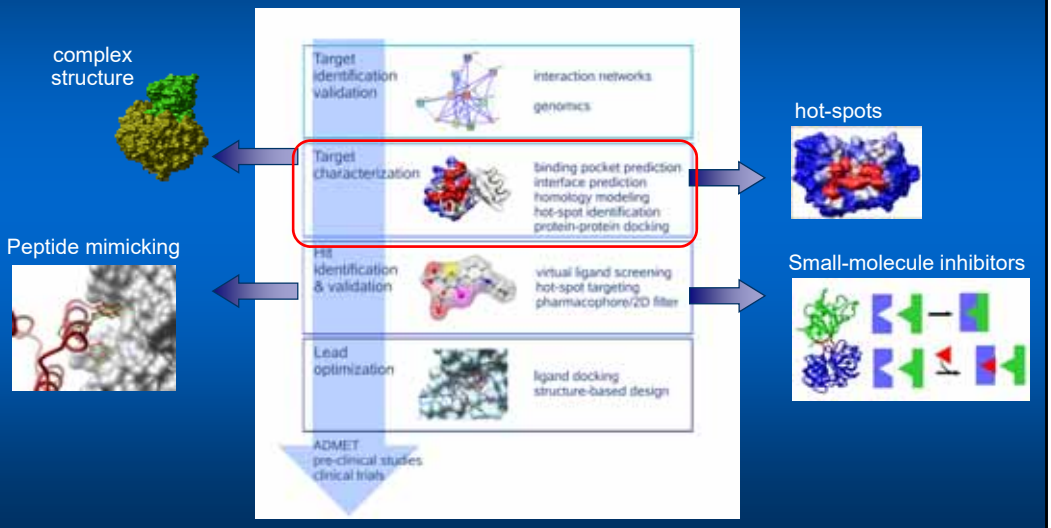
Rational design of protein-protein inhibitors

Major difficulties:

Problem	Possible solution
No natural cavities in PP interfaces	MD to generate transient cavities
Transient cavities difficult to identify (even bound-to-inhibitor cavities difficult)	Specialized pocket prediction tools
Where to search (in most cases, PP interface not known)	PP docking
Find best binding site (to compete with a large interface)	Target hot-spots
Find best PP inhibitor (chemically different from traditional ligands)	Specific libraries for PP inhibitors

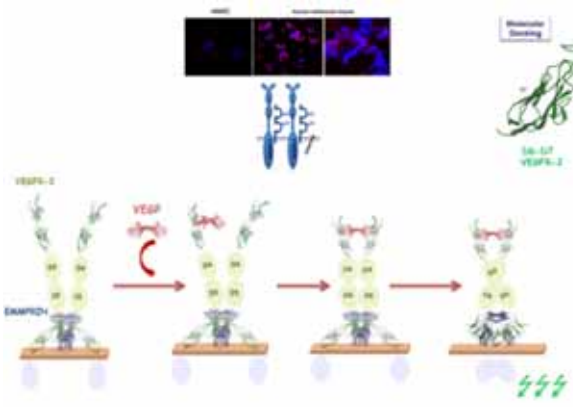
Docking applied to drug discovery targeting protein-protein interactions

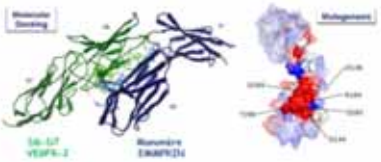
Protein-protein docking | **Interface and hot-spot predictions**



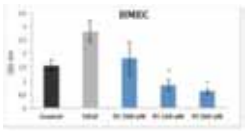
Drug discovery targeting protein interactions

- VEGF is the most relevant pro-angiongenic factor in cancer
- EMMPRIN is needed for VEFGR activation by VEGF





Inhibiting VEGFR-EMMPRIN interactions for angiogenesis treatment



Inserm

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

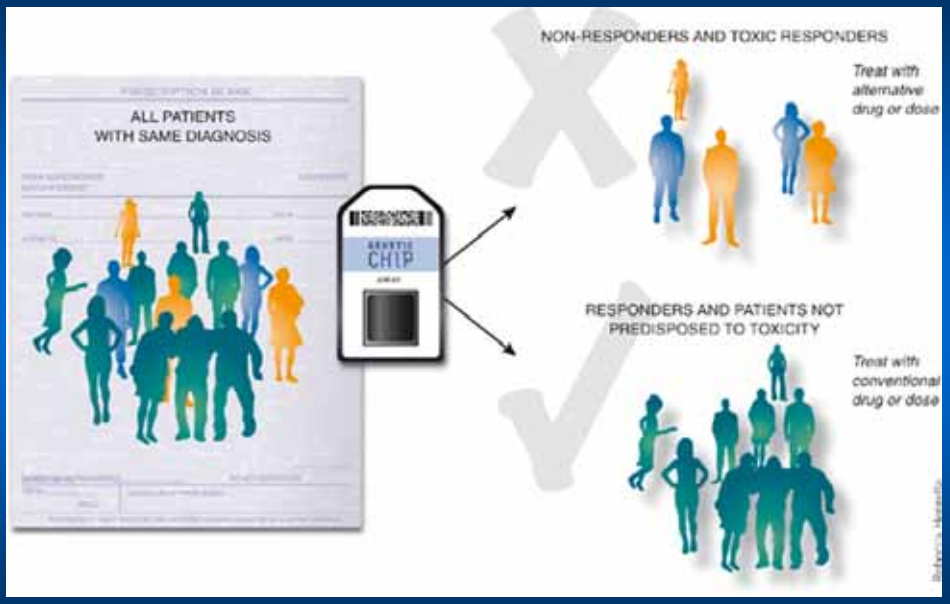
PARIS DIDEROT

Bruno Villoutreix
Samia Mourah

Khayati et al (2015) Oncotarget
Mourah et al. (2015) WO 2016/083409

- **Interface prediction**
- **Hot-spot identification**
- **Application to drug design targeting PPIs**
- **Pathological mutations affecting PPIs**

Personalized Medicine



Personalized medicine

International Cancer Genome Consortium

PCAWG

Genomics England

BabySeq

2015-2018

thousands of newly annotated mutations

BIG DATA

ANALYSIS

Population Pathology → Stratified Medicine

genome.gov
National Human Genome Research Institute

Predictive personalized medicine

ANALYSIS

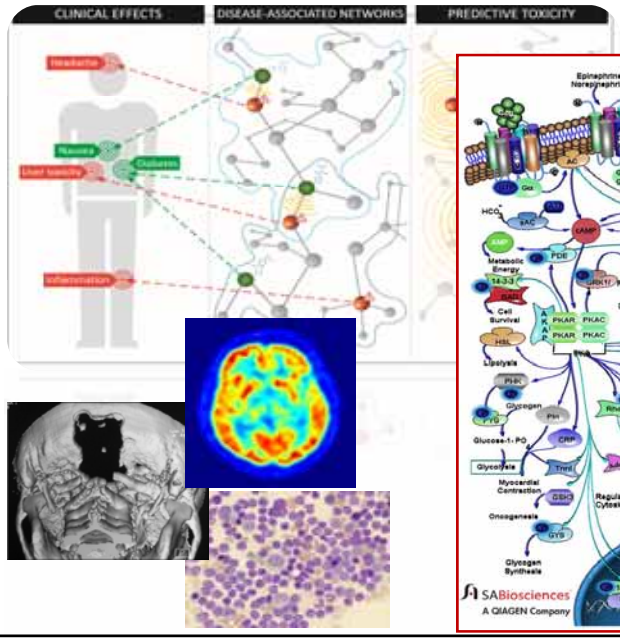
Identify Pathology *(precision diagnosis)*

Improve treatment, drug response *(personalized treatment)*

Annotated mutations → Understand cause of disease
New mutations → Predict effect on health/disease

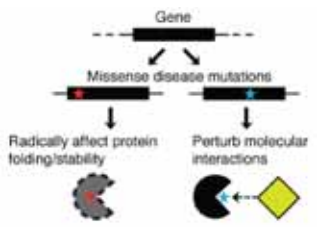
« We need to understand the functional mechanisms of pathological mutations at molecular level

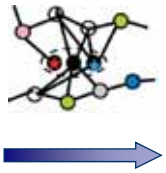
Molecular basis of disease

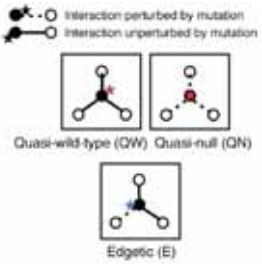


Applications in Biomedicine: Prediction and interpretation of pathological mutations

Effect of pathological mutations on PPIs and networks

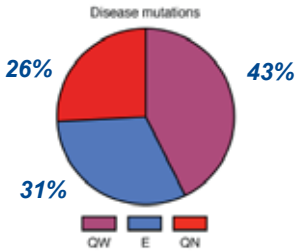






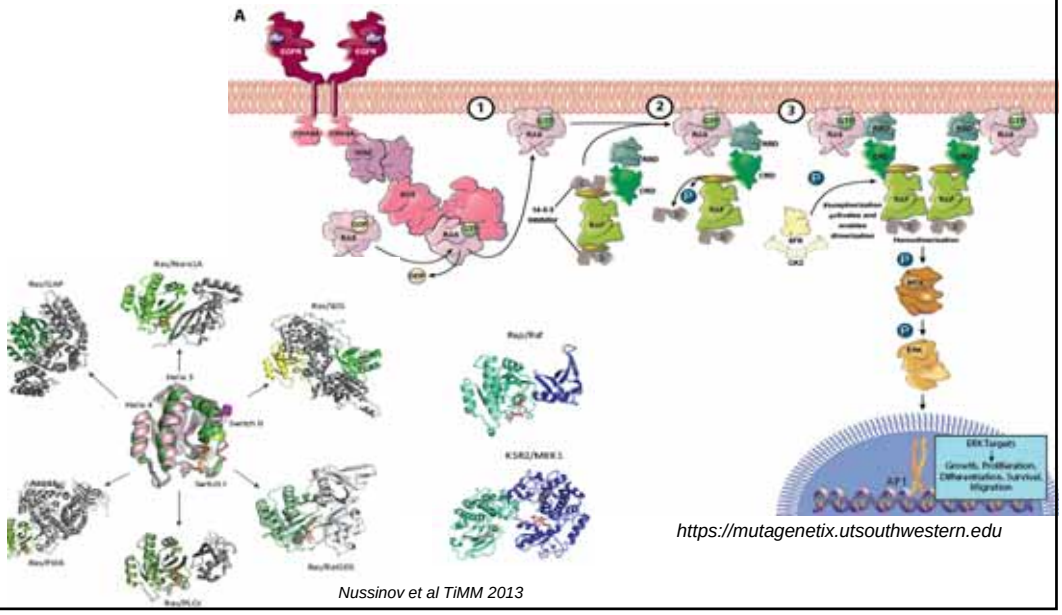
n=197 mutations

89 proteins with > 1 interaction



Sahni et al. (2015) Cell 161, 647

Characterization of pathological mutations (personalized medicine)

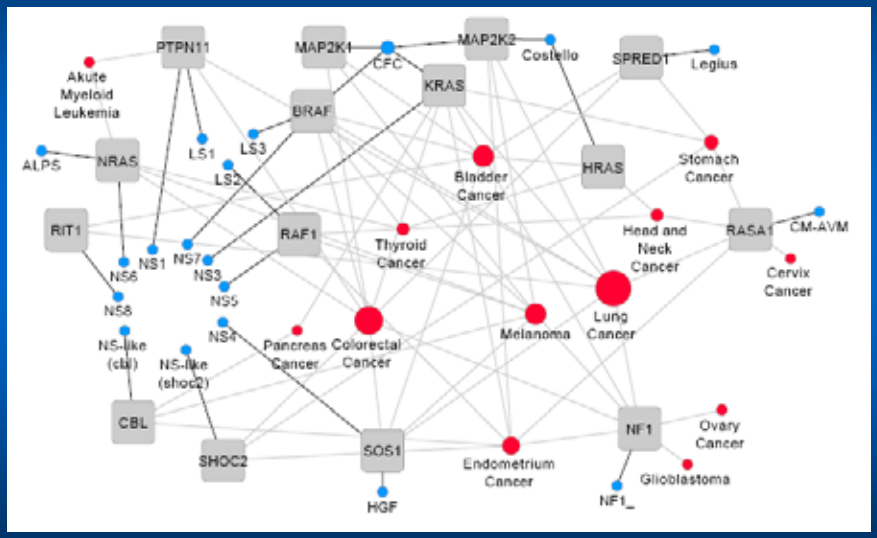


https://mutagenetix.utsouthwestern.edu

Nussinov et al TiMM 2013

Characterization of pathological mutations

Mutations in MAPK/ERK pathway cause cancer or RASopathies



Kiel & Serrano (2014) Mol.Sys.Biol.

Characterization of pathological mutations

Folding/Stability

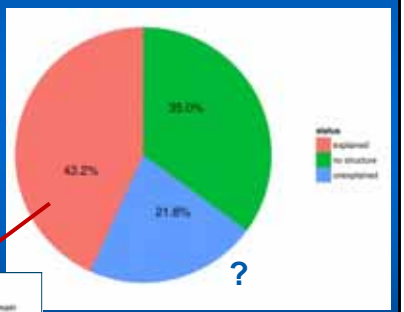
Activity / Function

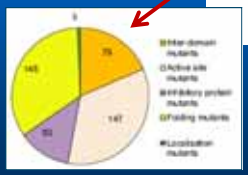
PPI

Localization

956 mutations

Germline mutations: RASopathies
Somatic mutations: cancer





We need more structural and dynamic models for PPIs

Kiel & Serrano (2014) Mol.Sys.Biol.

Pathological mutations affecting protein interactions

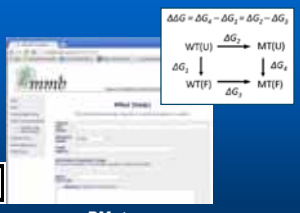
Disease nsSNPs in Protein Complexes
David *et al.* 2012 *Human Mut.* 33, 359

	Total residues	Observed	Expected	O/E percent ratio
Disease nsSNPs				
Core	42,342	781	516.1	1.5
Interface	43,381	620	528.8	1.2
Surface noninterface	112,805	1019	1375.1	0.7
Total	198,528	2420		

Affecting Protein Stability

FoldX destabilization

Kiel *& Serrano* 2014
Mol. Sys. Biol.



PMut
Ferrer-Costa *et al.* 2012 *JMB*

Pathological mutations affecting protein interactions

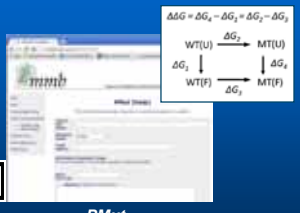
Disease nsSNPs in Protein Complexes
David *et al.* 2012 *Human Mut.* 33, 359

	Total residues	Observed	Expected	O/E percent ratio
Disease nsSNPs				
Core	42,342	781	516.1	1.5
Interface	43,381	620	528.8	1.2
Surface noninterface	112,805	1019	1375.1	0.7
Total	198,528	2420		

Affecting Protein Stability

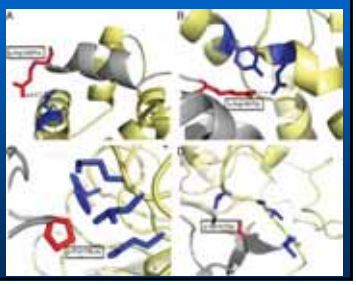
FoldX destabilization

Kiel *& Serrano* 2014
Mol. Sys. Biol.



PMut
Ferrer-Costa *et al.* 2012 *JMB*

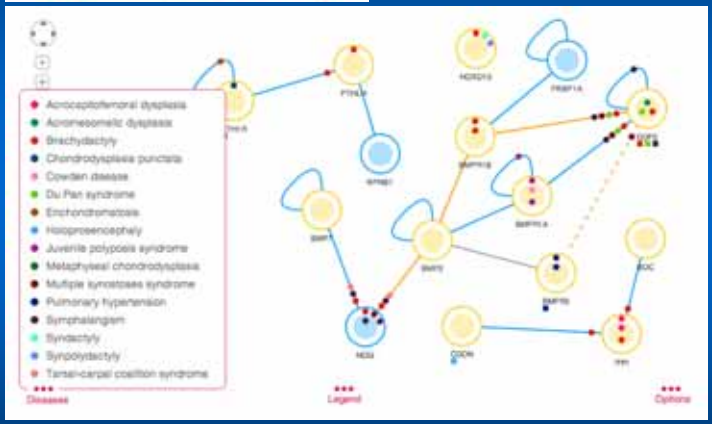
Affecting Binding Affinity



Pathological mutations affecting protein interactions

dsysmap

http://dsysmap.irbbarcelona.org/



Mosca *et al.* 2015 *Nature Methods* 12, 167

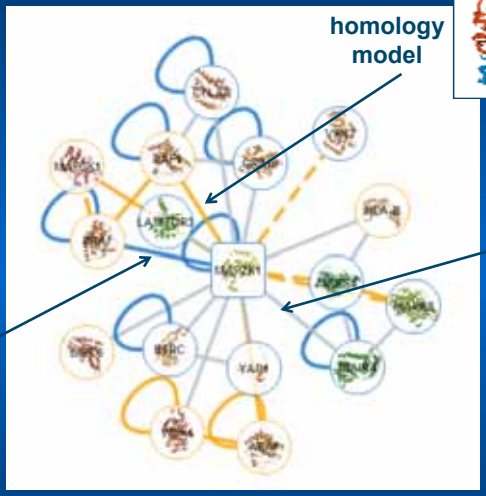
24K disease mutations
9.5K mapped on PPI structures
2.3K mapped on interfaces

In 2012:
23K disease mutations
2.4K mapped on PPI structures
0.6K mapped on interfaces

Characterization of pathological mutations in PPIs

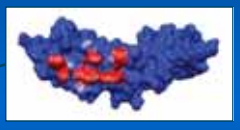
MEK1 structural interaction network

x-ray



homology model

docking binding site predictions



Applications in Biomedicine:
Prediction and interpretation of pathological mutations

Pathological mutations in PPIs, personalized diagnosis