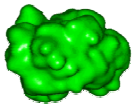
Protein-Protein Docking

Thomas Funkhouser
Princeton University
CS597A, Fall 2005

Introduction

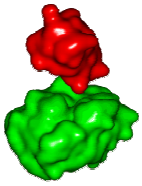
Goal:
- Given two protein structures, predict how they form a complex

Applications:
- Quaternary structure prediction
- Protein interaction prediction
- etc.



Introduction

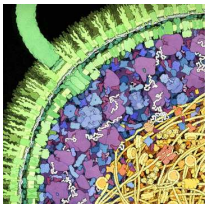
Goal:
- Given two protein structures, predict how they form a complex



Applications:
- Quaternary structure prediction
- Protein interaction prediction
- etc.

Introduction

Proteins are densely packed inside cell
- 20-30% of total volume inside cell

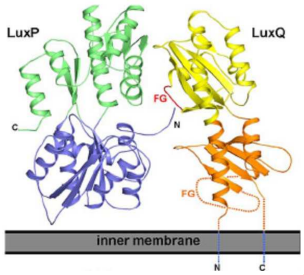


Representation of the approximate numbers, shapes and density of packing of macromolecules inside a cell of *Escherichia coli*. (Illustration by David S Goodsell)

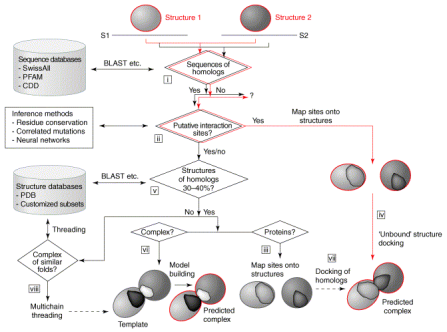
[Szilágyi05]

Introduction

Many biological processes are controlled by protein-protein interactions
- Signal transduction
- Transport
- Cellular motion



Protein Interaction Prediction



Outline

Introduction

Binding analysis

Docking methods

Evaluation

Discussion

Outline

Introduction

Binding analysis ←

Docking methods

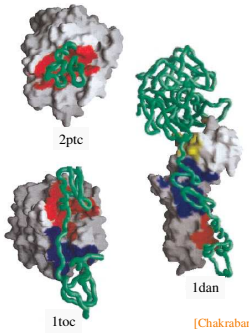
Evaluation

Discussion

Binding Site Analysis

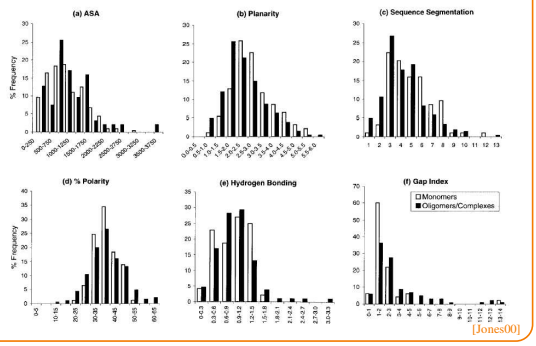
Proteins sometimes contact each other in more than one distinct patch

- One patch (46/70)
- Two patches (18/70)
- More patches (6/70)



[Chakrabarti02]

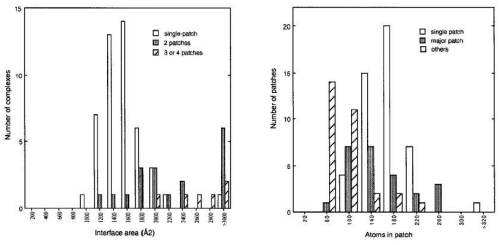
Binding Site Analysis



[Jones00]

Binding Site Analysis

Protein interfaces tend to bury 1320 ± 520 Å^2



[Chakrabarti02]

Binding Site Analysis

Some residues have higher propensity to be in site

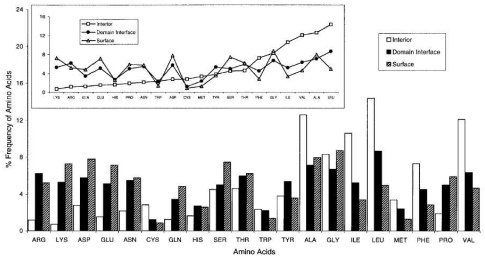
TABLE II. Amino Acid Composition of Protein-Protein Interfaces

	Number (a)			Area (b)			Propensities (c)			
Residue	Interface	Core	Rim	Interface	Core	Rim	Core	Rim	Lo Conte et al. (d)	Jones and Thornton (e)
All	100.0	100.0	99.9	99.9	100.0	100.0				
Ala	3.9	4.0	3.8	2.8	2.7	3.1	−0.40	−0.26	−0.43	−0.17
Arg	6.4	5.9	7.0	10.1	10.1	9.9	0.13	0.11	0.13	0.27
Asn	5.9	5.4	6.4	5.7	5.4	6.4	−0.14	0.03	−0.12	0.12
Asp	6.6	5.4	8.0	5.1	4.5	6.6	−0.46	−0.07	−0.31	−0.38
Cys	3.5	4.7	2.1	1.7	1.9	1.3	1.00	0.62	0.76	0.43
Gln	3.7	3.7	3.8	4.3	4.3	4.2	−0.34	−0.36	−0.36	−0.11
Glu	6.5	4.6	8.6	6.0	4.4	10.0	−0.80	0.02	−0.47	−0.13
Gly	8.1	7.5	8.7	4.8	4.2	6.4	−0.08	0.35	0.02	−0.07
His	3.4	4.4	2.3	3.8	4.4	2.4	0.84	0.23	0.64	0.41
Ile	3.6	4.1	3.1	4.6	4.9	3.5	0.71	0.38	0.56	0.44
Leu	5.0	5.5	4.5	5.7	5.8	5.3	0.34	0.25	0.29	0.40
Lys	5.7	3.7	8.0	6.5	5.2	9.7	−0.82	−0.20	−0.57	−0.36
Met	2.0	2.6	1.4	3.2	3.7	2.0	1.13	0.51	0.98	0.66
Phe	3.5	5.1	1.7	4.1	5.5	1.1	1.01	−0.60	0.79	0.82
Pro	3.8	3.4	4.2	3.6	3.5	4.1	−0.38	−0.22	−0.25	−0.25
Ser	7.9	7.8	8.1	5.4	4.8	7.3	−0.56	−0.14	−0.42	−0.33
Thr	6.2	5.7	6.8	5.0	4.7	5.9	−0.44	−0.21	−0.35	−0.18
Trp	2.8	4.1	1.3	4.2	5.3	1.6	1.41	0.21	1.25	0.83
Tyr	6.8	8.1	5.4	9.4	10.9	5.3	1.22	0.50	1.04	0.66
Val	4.5	4.3	4.7	3.8	3.8	3.9	0.08	0.11	0.09	0.27

[Chakrabarti02]

Binding Site Analysis

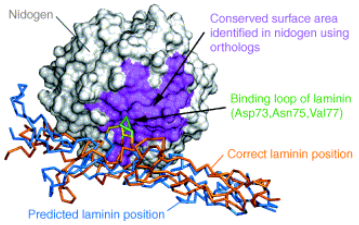
Some residues have higher propensity to be in site



[Jones00]

Binding Site Analysis

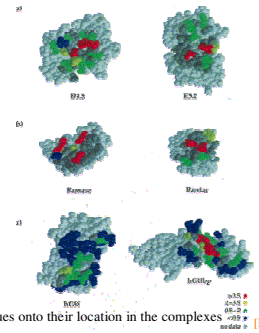
Residues in protein-protein interfaces are often better conserved than others



[Wodak04]

Binding Site Analysis

Many residues often contribute to binding energetics



Mapping of $\Delta\Delta G$ of individual residues onto their location in the complexes [Bogan98]

Outline

Introduction

Binding analysis

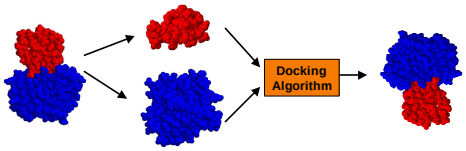
Docking methods ←

Evaluation

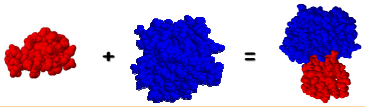
Discussion

Protein-Protein Docking

Bound docking:



Unbound docking:



[Gidalevitz]

Protein-Protein Docking

Similar to protein-ligand docking

- Search of conformations
- Scoring of energetics

Protein-Protein Docking

Main differences:

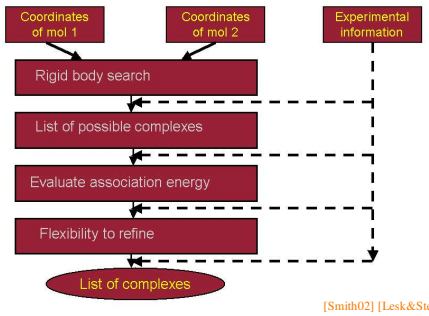
- Sites have …
  - § Large, flat surfaces
  - § Conservation, maybe
  - § Hydrophobic core
- Binding energetics are usually dominated by …
  - § Geometry
  - § Hydrophobicity
- Protein flexibility is important
  - § Side-chains
  - § Backbone

Protein-Protein Docking

Programs:

- 3D-Dock
- HEX
- GRAMM
- PPD
- DOT
- BIGGER
- DOCK
- AutoDock
- FlexX
- Darwin
- ZDOCK

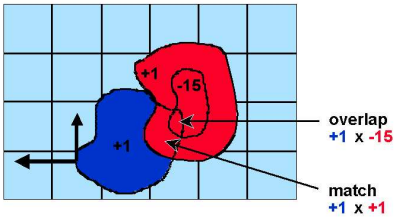
Protein-Protein Docking Pipeline



[Smith02] [Lesk&Sternberg]

Rigid Docking

Shape complementarity:



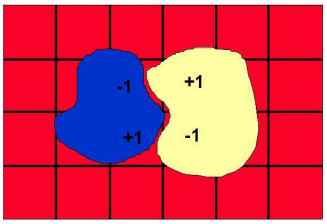
A(i,j,k) B(l,m,n)

$C = \sum\sum\sum A(i,j,k) \times B(l,m,n)$

[Lesk&Sternberg]

Rigid Docking

Electrostatic complementarity:



Charge in 1 = Q(i,j,k) Potential outside 2 V(l,m,n)

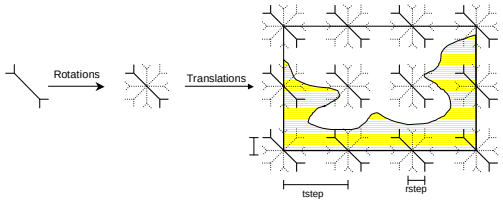
$E = \sum\sum\sum Q(i,j,k) \times V(l,m,n)$

[Lesk&Sternberg]

Rigid Docking

Search methods:

- Exhaustive
- FFT

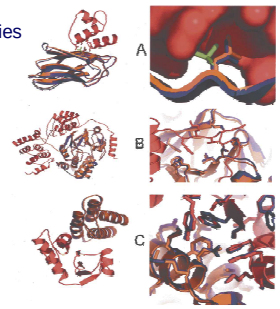


FRED [Yang04]

Flexible Docking

Search methods:

- Side-chain rotamer libraries
- Monte Carlo algorithms
- Genetic algorithms



[Wang05]

Outline

Introduction

Binding analysis

Docking methods

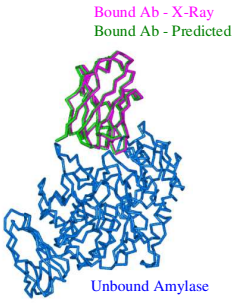
Evaluation ←

Discussion

Evaluation Methods

Metrics:

- RMSD (usually Cα)
- % of contacts predicted

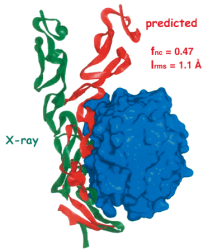


[Lesk&Sternberg]

Evaluation Methods

Metrics:

- RMSD (usually Cα)
- % of contacts predicted



[Janin05]

Evaluation Methods

Benchmarks:

- CAPRI

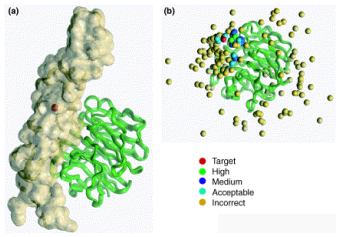
Target	Reference	Submitting groups	Model quality[a]	
			High/good	Acceptable
T01	HPr kinase/HPr	Fieulaine et al. 2002	16	0	8
T02[b]	Rotavirus VP6/Fab	Thouvenin et al. 2001	15	1	6
T03	Flu hemagglutinin/Fab	Barbey-Martin et al. 2002	13	2	0
T04	Amylase/camel V_{HH}	Desmyter et al. 2002	13	0	0
T05	Amylase/camel V_{HH}	Desmyter et al. 2002	13	0	0
T06	Amylase/camel V_{HH}	Desmyter et al. 2002	13	8	0
T07[c]	Superantigen/TCR[d]	Sundberg et al. 2002	14	12	8
T08	Ndc80/tubulin	Takagi et al. 2003	18	11	18
T09	LicT dimer	H. van Tilbeurgh and M. Graille, in prep.	17	0	0
T10	TBE virus E trimer	Bressanelli et al. 2004	20	1	3
T11	Cohesin/dockerin (unbound)	Carvalho et al. 2004	19	0	13
T12	Cohesin/dockerin (bound)	Carvalho et al. 2004	22	20	10
T13	SAG1/Fab	M. Graille and F. Ducancel, in prep.	21	5	8
T14	Phosphatase 1#/MYPT1	Terrak et al. 2004	25	38	32
T15[d]	Colicin D/Imm D	Graille et al. 2004	10	9	6
T16[e]	Xylanase/TAXI	Sainsen et al. 2004	26	5	4
T19	Orist prion/Fab	Eghiaian et al. 2004	24	11	9

[Janin05]

Evaluation Methods

Benchmarks:

- CAPRI



X-Ray Structure for Capri Target 08

Distribution of Centers of Mass for predicted Complexes

[Wodak04]

Discussion

?

References

[Bogan98] A.A. Bogan, K.S. Thorn, "Anatomy of hot spots in protein interfaces," J. Mol. Biol., 280, 1998, pp. 1-9.

[Chakrabarti02] P. Chakrabarti, J. Janin, "Dissecting protein-protein recognition sites," Proteins: Structure, Function, and Genetics, 47, 3, 2002, pp. 334-343.

[Gidalevitz] Gidalevitz T, Biswas C, Ding H, Schneidman D, Wolfson HJ, Stevens F, Radford S, Argon Y. "Guiding "in vitro" experiments with "in silico" predictions", http://bioinfo3d.cs.tau.ac.il/Education/BioInfo04/LastLect/Docking-grp94.ppt.

[Janin05] J. Janin, "Assessing predictions of protein-protein interaction: The CAPRI experiment," Protein Science, 14, 2005, pp. 278-283.

[Jones00] S. Jones, A. Marin, J.M. Thornton, "Protein domain interfaces: characterization and comparison with oligomeric protein interfaces," Protein Engineering, 13, 2, 2000, pp. 77-82.

[Smith02] G.R. Smith, M.J.E. Sternberg, "Prediction of protein-protein interactions by docking methods," Current Opinion in Structural Biology, 12, 2002, pp. 28-35.

[Szilagyi05] A. Szilagyi, V. Grimm, A.K. Arakaki, J. Skolnick, "Prediction of physical protein-protein interactions," Phys. Biol., 2, 2005, pp. S1-S16.

[Wang05] C. Wang, O. Schueler-Furman, and D. Baker, "Improved side-chain modeling for protein-protein docking", Protein Science, 14, 2005, pp. 1328-1339.

[Wodak04] S.J. Wodak, R. Mendez, "Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications," Current Opinion in Structural Biology, 14, 2004, pp. 242-249.