THE
SCRIPPS
RESEARCH
INSTITUTE

NCRR

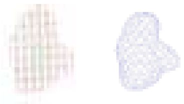
Correlation Based Search

Pablo Chacón, Ph.D.
Department of Molecular Biology
The Scripps Research Institute
10550 N. Torrey Pines Road, Mail TPC6
La Jolla, California, 92037

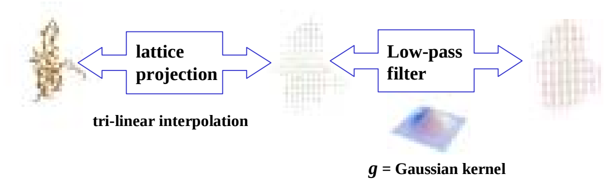
Situs Modeling Workshop, San Diego, CA, Feb. 3-5, 2003

Correlation-Based ‘Interior’ Docking

$\rho_{\mathrm{em}}(\mathbf{r})$ target density on lattice



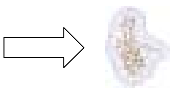
$\rho_{\mathrm{atomic}}(\mathbf{rR})$ rotated probe molecule density projected to the lattice:



$g \otimes \rho_{\mathrm{atomic}}(\mathbf{rR})$

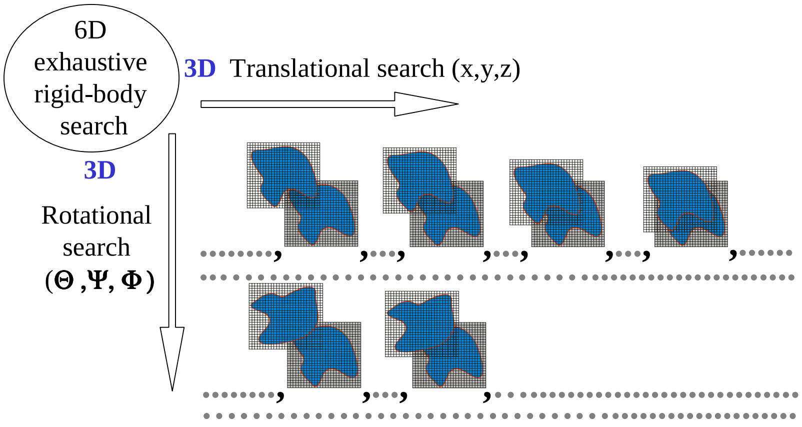
$\equiv \rho_{\mathrm{calc}}(\mathbf{r})$

$$\boxed{C(\mathbf{T}) = \int \rho_{\mathbf{em}}(\mathbf{r}) \times \rho_{\mathbf{calc}}(\mathbf{r}+\mathbf{T})\ \mathbf{d}^3\mathbf{r}}$$



Fitting criterion: e.g. linear cross-correlation,
evaluate for every rotation **R** and translation **T**

How can we perform the docking search?



Compute one correlation coefficient cost N multiplications **X** **N possible translational shifts** **X** **Number of rotations**

Where N = number of voxels

How can we reduce the computational cost?

1) Accelerate the translational search

FTM (Fast Translational Matching)

$$C(\mathbf{T}) = \int \rho_{\text{em}}(\mathbf{r}) \times \rho_{\text{calc}}(\mathbf{r}+\mathbf{T})\ d^3r = f^{-1}\left[f\left(\rho_{\text{em}}\right)^* \times f\left(\rho_{\text{calc}}\right) \right]$$

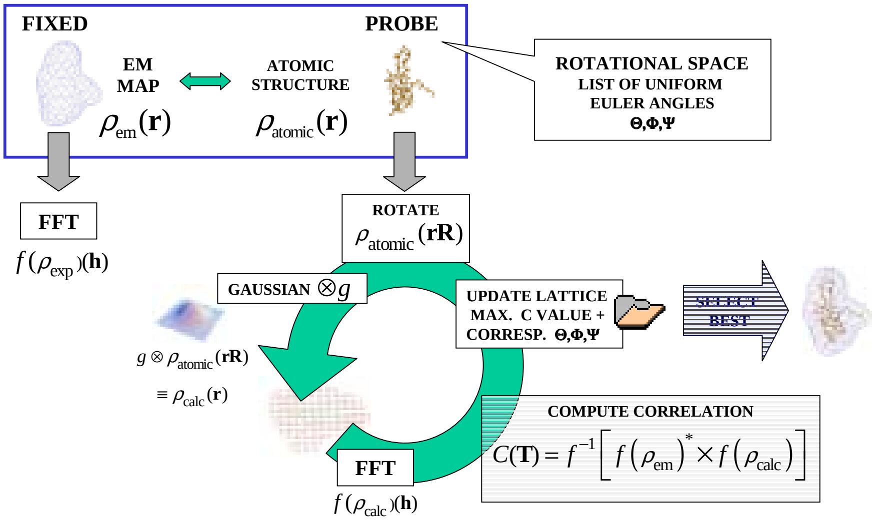
Fourier Convolution Theorem:

Direct Approach: N^2 multiplications
FFT Approach: N log N multiplications (x2)

Where N = number of voxels

2) Accelerate the rotational search →
Fast Rotational Matching, FRM
(J. Kovacs)

How can we implement the FTM search?



FIXED

EM MAP

$\rho_{\text{em}}(\mathbf{r})$

ATOMIC STRUCTURE

$\rho_{\text{atomic}}(\mathbf{r})$

PROBE

ROTATIONAL SPACE
LIST OF UNIFORM EULER ANGLES
Θ,Φ,Ψ

FFT

$f(\rho_{\text{exp}})(\mathbf{h})$

ROTATE

$\rho_{\text{atomic}}(\mathbf{rR})$

GAUSSIAN $\otimes g$

$g \otimes \rho_{\text{atomic}}(\mathbf{rR})$

$\equiv \rho_{\text{calc}}(\mathbf{r})$

UPDATE LATTICE
MAX. C VALUE +
CORRESP. **Θ,Φ,Ψ**

SELECT BEST

COMPUTE CORRELATION

$C(\mathbf{T}) = f^{-1}\left[f\left(\rho_{\text{em}}\right)^* \times f\left(\rho_{\text{calc}}\right) \right]$

FFT

$f(\rho_{\text{calc}})(\mathbf{h})$

This is an adaptation of the protein-protein shape complementarity algorithm of Katchalski-Katzir et al. (1992) [FTDOCK (Gabb 1997), DOT (Roberts 1999) and GRAMM (Vakser 1999)]

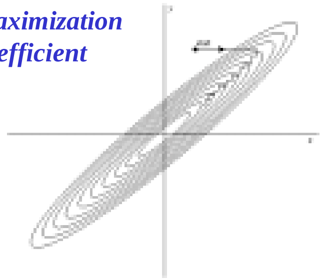
How we can improve the results of FTM?

6D exhaustive search is limited:

- Rotational search → Angular sampling
- Translational search → Grid size

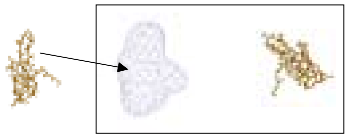
Improve the accuracy

Off-lattice (6D) local maximization of the correlation coefficient

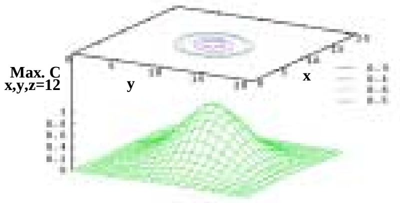


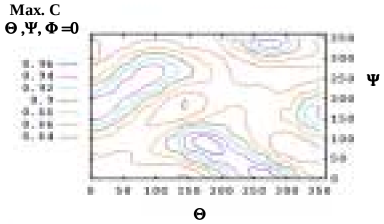
Powell's quadratically convergent maximization method can be used to perform a 6D search around the best fits found on the grid.

Example: Transcription Elongation FSII

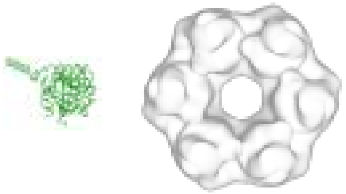


Grid size 6Å
Resolution 20 Å
ΔAngle 9º
(30481 rotations)





Multiple molecule docking
Example: 2recA



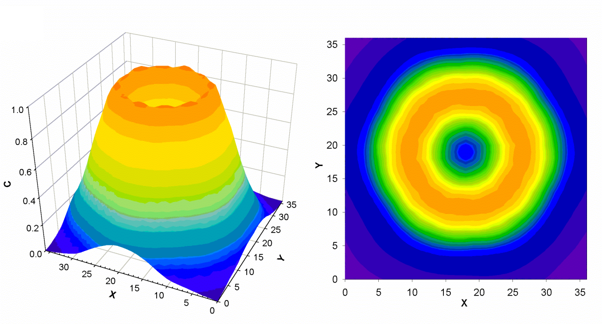
6D exhaustive search
(3D translations+3D rotations)

FTM

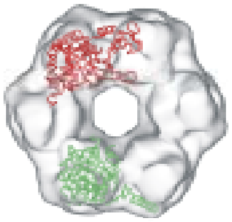
ΔAngle 9º
(30481 rotations)

Resolution 15 Å
Grid size 4Å

Translational Space (x,y,z)



With density cross-correlation we can not distinguish between correct and spurious fit

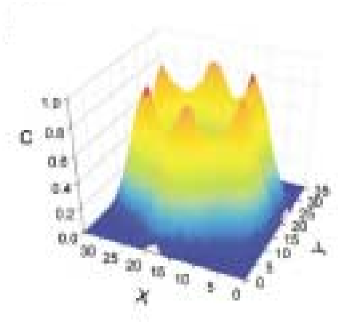


How to enhance the fitting contrast (I)

1) Renormalize (mask) the correlation locally

$$C(\mathbf{T}) = \frac{\int_{mask} \rho_{\mathrm{em}}(\mathbf{r}) \times \rho_{\mathrm{calc}}(\mathbf{r}+\mathbf{T})\ \mathrm{d}^3\mathrm{r}}{\sqrt{\int_{mask} \rho_{\mathrm{em}}^2(\mathbf{r})\mathrm{d}^3\mathrm{r}}\sqrt{\int_{mask} \rho_{\mathrm{calc}}^2(\mathbf{r})\ \mathrm{d}^3\mathrm{r}}}$$

$$\text{mask} \rightarrow \rho_{\mathrm{calc}_{l,m,n}} > 0$$



→extends the liability of correlation based docking (<15Å)

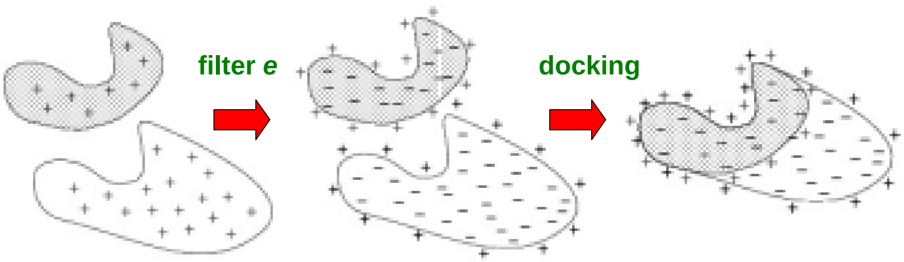
→Can not be FFT accelerated

DOCKEM
A.M. Roseman

How to enhance the fitting contrast (II)

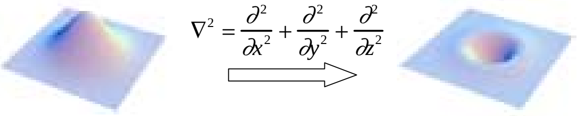
2) By adding surface/contour information

A suitable filter would assign negative values to the interior, positive values to the molecular contour. Both volume and contour matches would provide positive contributions to the correlation criterion:

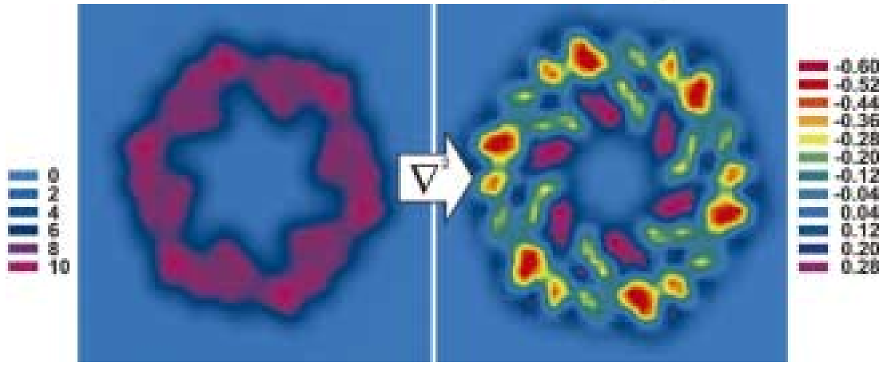


How to characterize the contour/molecular surface in a low-resolution map

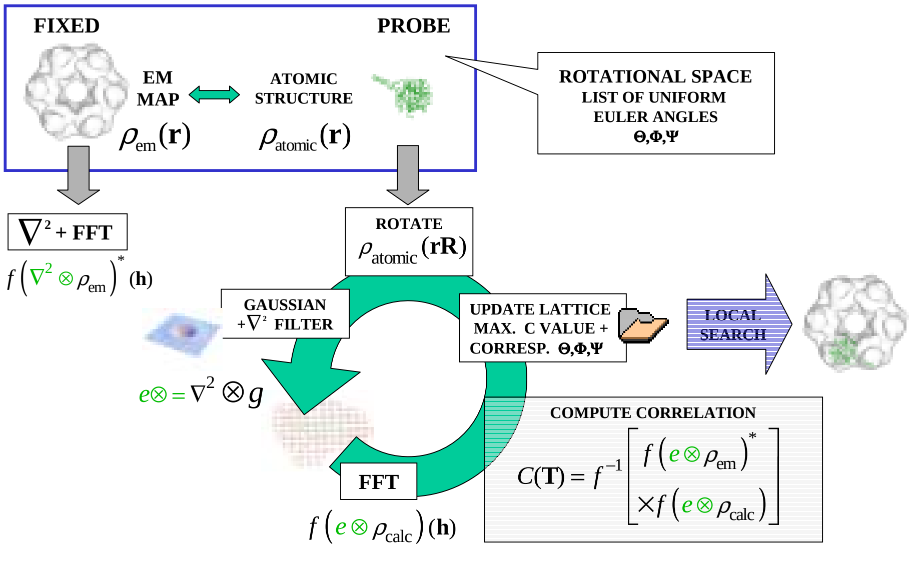
Laplacian filter



$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$



FTM + Filtering



FIXED

PROBE

EM MAP ⟺ ATOMIC STRUCTURE

$\rho_{\mathrm{em}}(\mathbf{r})$ $\rho_{\mathrm{atomic}}(\mathbf{r})$

ROTATIONAL SPACE
LIST OF UNIFORM
EULER ANGLES
Θ,Φ,Ψ

∇^2 + FFT

$f\left(\nabla^2 \otimes \rho_{\mathrm{em}}\right)^*(\mathbf{h})$

ROTATE
$\rho_{\mathrm{atomic}}(\mathbf{rR})$

GAUSSIAN
$+\nabla^2$ FILTER

$e\otimes = \nabla^2 \otimes g$

UPDATE LATTICE
MAX. C VALUE +
CORRESP. Θ,Φ,Ψ

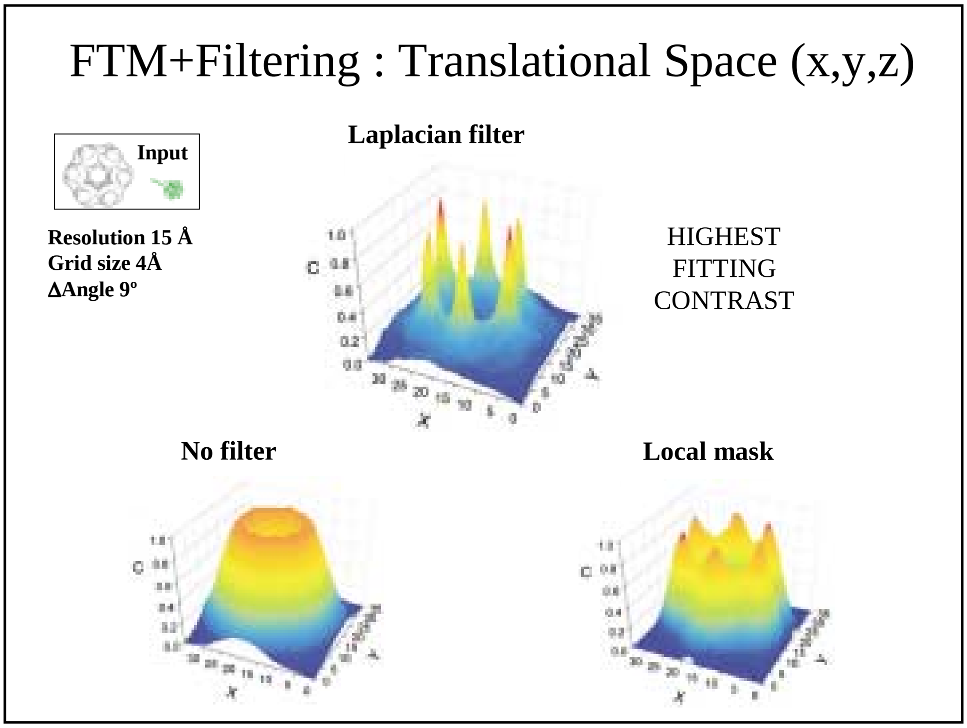
LOCAL
SEARCH

COMPUTE CORRELATION

$$C(\mathbf{T}) = f^{-1}\left[f\left(e\otimes\rho_{\mathrm{em}}\right)^* \times f\left(e\otimes\rho_{\mathrm{calc}}\right) \right]$$

FFT

$f\left(e\otimes\rho_{\mathrm{calc}}\right)(\mathbf{h})$



FTM+Filtering : Translational Space (x,y,z)

Input

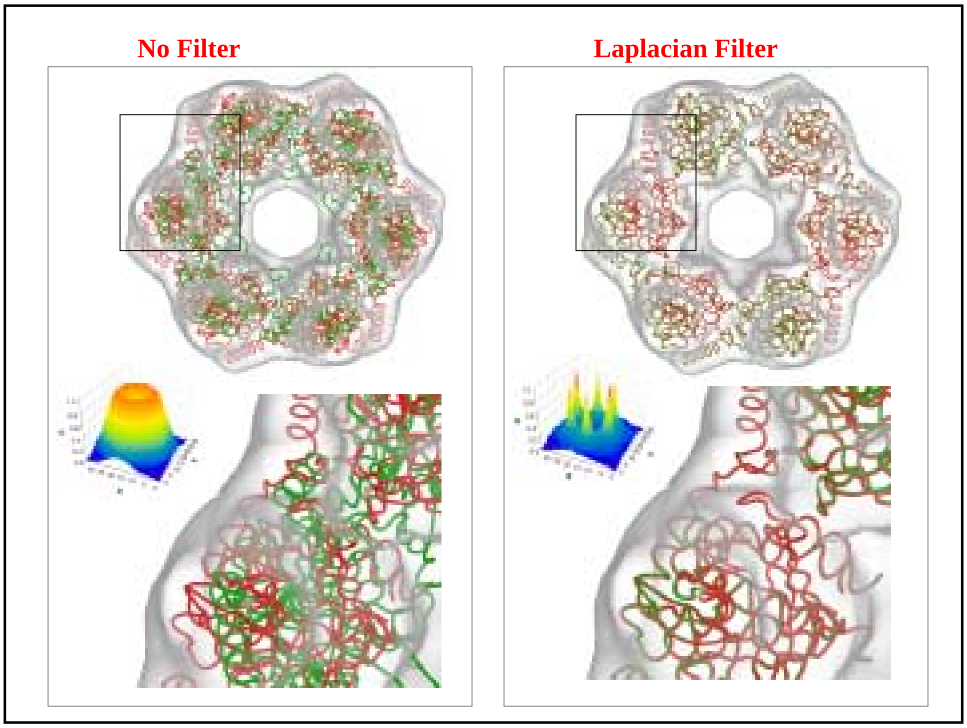
Resolution 15 Å
Grid size 4Å
ΔAngle 9°

Laplacian filter

HIGHEST
FITTING
CONTRAST

No filter

Local mask



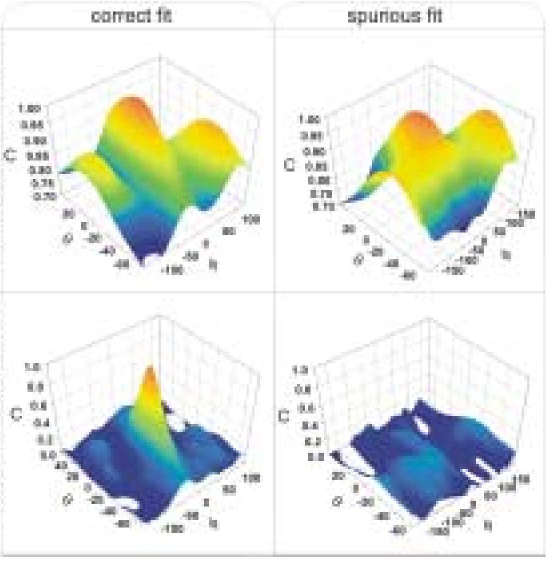
No Filter

Laplacian Filter

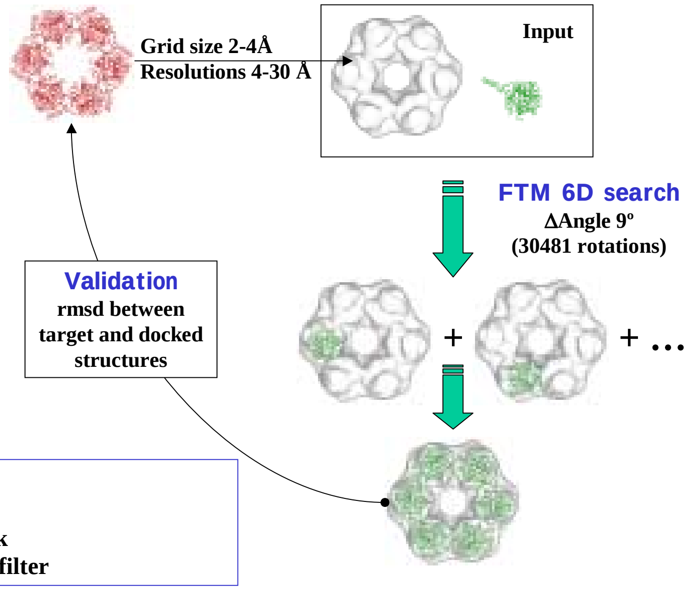
Rotational Space (θ,ψ,ϕ)

No Filter

Laplacian Filter



Restoration Tests with Simulated Data



Grid size 2-4Å
Resolutions 4-30 Å

Input

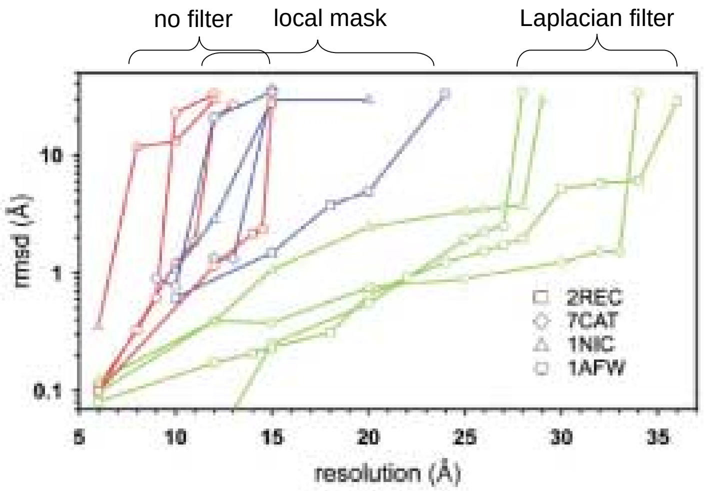
FTM 6D search
ΔAngle 9º
(30481 rotations)

Validation
rmsd between target and docked structures

\+ \+ …

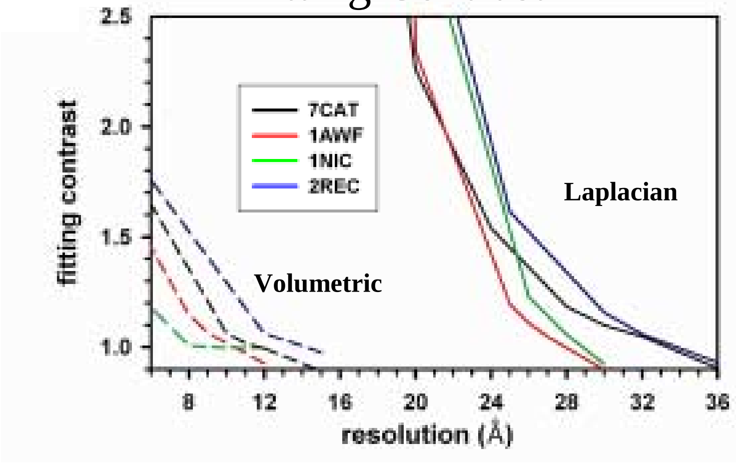
USING:
- **No filter**
- **Local mask**
- **Laplacian filter**

Restoring Various Oligomers



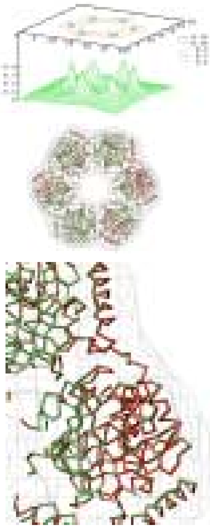
RecA (2REC), thiolase (1AFW), catalase (7CAT), and oxidoreductase (1NIC).

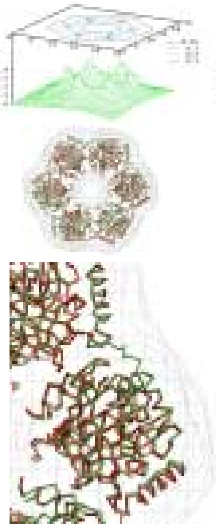
Fitting Contrast

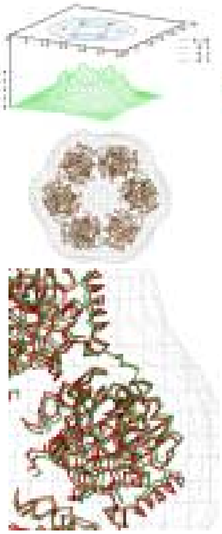


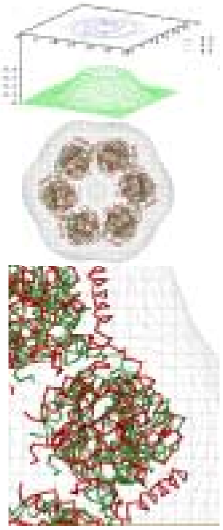
The fitting contrast is defined as the ratio between the correlation coefficients of the highest (correct fit) and the second highest (spurious fit) peak found in the 6D searches

Resolution Dependence


Resolution 15 Å
rmsd < 1 Å

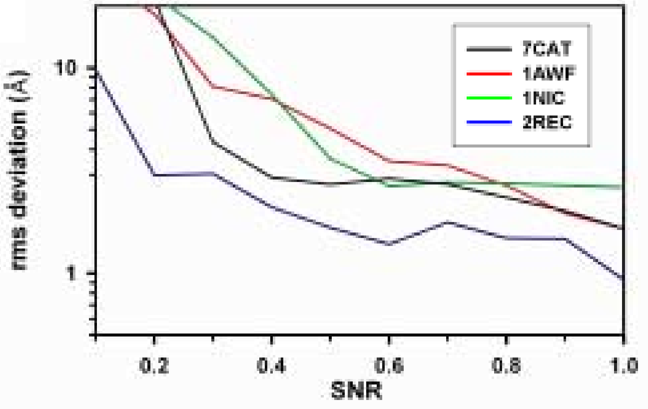

Resolution 20 Å
rmsd <1 Å


Resolution 25 Å
rmsd < 2 Å


Resolution 35 Å
rmsd>>

Noise Dependence

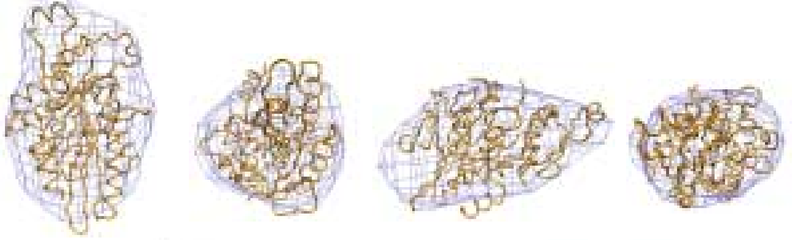
adding Gaussian noise to 20Å simulated maps

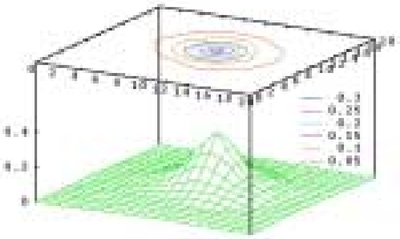


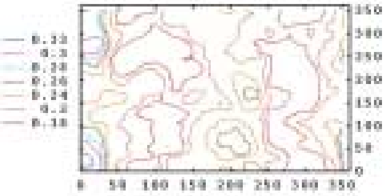
$$SNR = \mathrm{var}(o)/\mathrm{var}(n)$$

Example: ncd motor

Grid size 4Å
Resolution 20 Å
ΔAngle 9 º(30481 rotations)



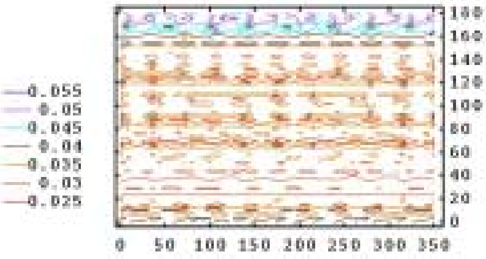


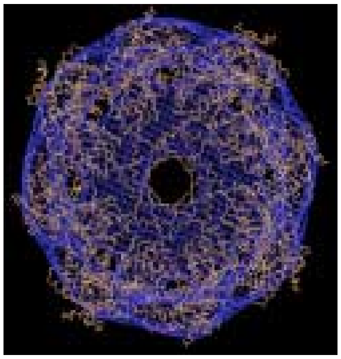


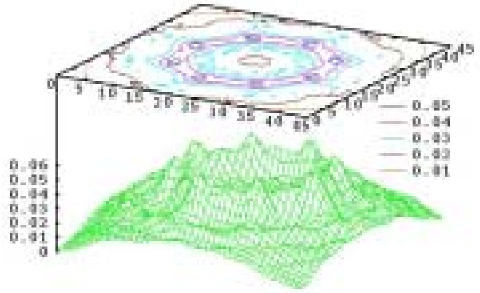
It works for single molecule….but is more expensive

Example: CCT chaperonin

Grid size 4Å
Resolution 27 Å
ΔAngle 9Å (30481 rotations)





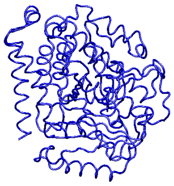


Example: Microtubule

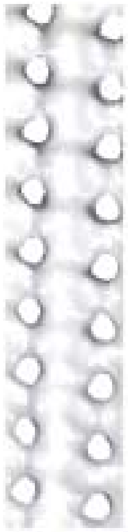
independent subunit docking

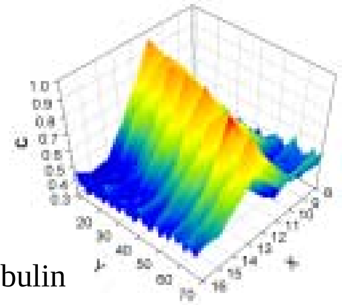
Resolution 20Å, Grid size 5Å
Δangle 9º (30481 rotations)

β-tubulin

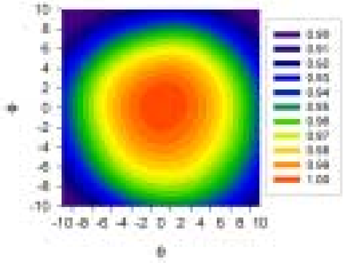


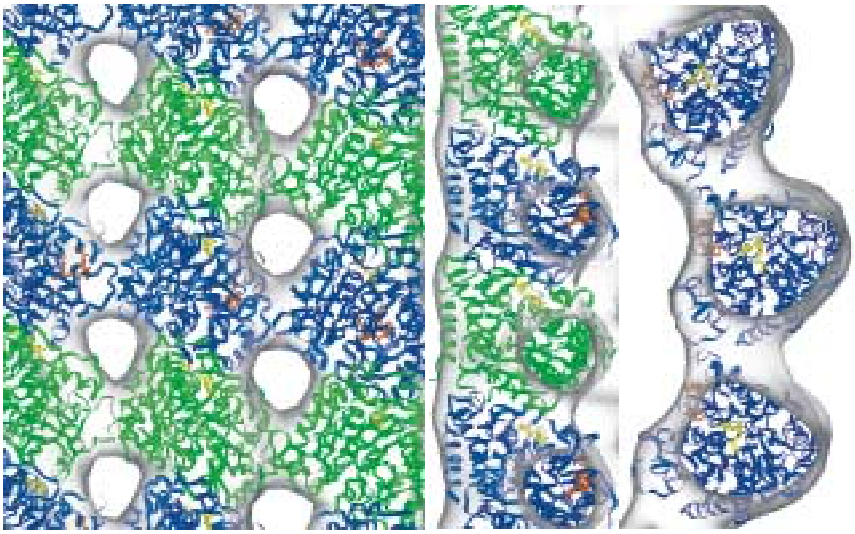
α-tubulin

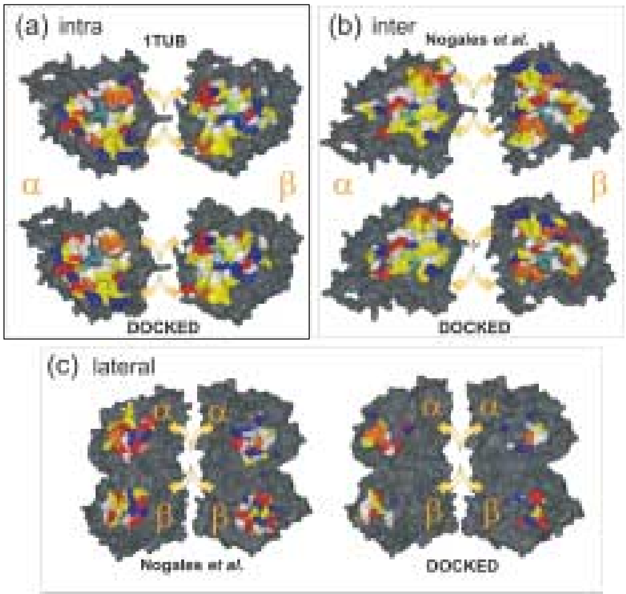




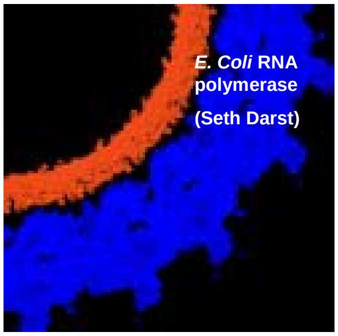
β-tubulin

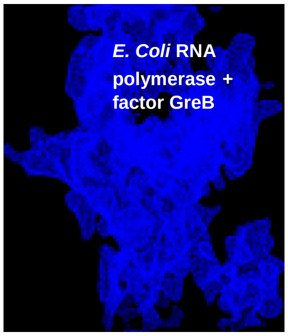






Outlook: Registration of two EM maps

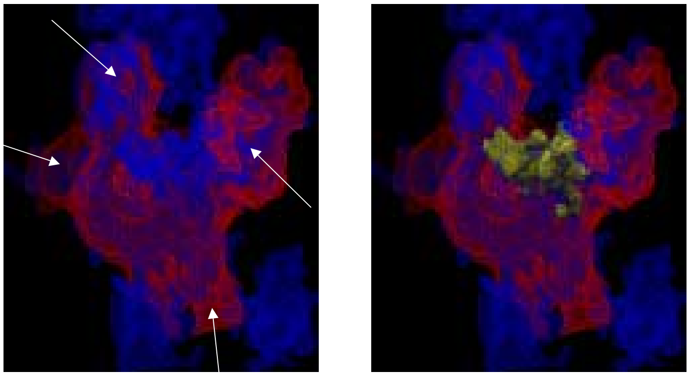




Problem: different helical arrays

Need to perform difference mapping to localize GreB (difficult at variable helical symmetry)

Registration of Maps



Rigid-body docking: The RNAP “jaws” are open in presence of GrepB factor, need to perform flexible map fitting (in progress).

Map fitting will be available in Situs 2.2.

Conclusions

We have implemented FTM in Situs 2.x with filtering. Laplacian filtering enhances the fitting contrast and offers:

(1) <u>A discriminative scoring function</u>: The non-specific, cross-correlation by itself is not a stringent criterion for the docking. A Laplacian filter includes the specific contour features as a fitting criterion.

(2) <u>Exhaustive search</u>: FTM performs a full six-dimensional rigid-body search, eliminating the risk of missing viable docking conformations.

(3) <u>Precision:</u> With Powell off-lattice search we can achieve sub-lattice precision

(4) <u>Scope of applicability</u>: The proposed can be applied directly to dock subunits to macromolecular complexes, or to localize small components in macromolecular low-resolution structures.

Possible Limitations of Laplacian Filtering:

- When the target map lacks a sufficiently large surface to define a reliable contour
- When the target map exhibits localized high-frequency noise.

Availability (with and without Laplacian filter):

- CoLoRes (COntour LOw RESolution) tool in Situs 2.x
- URL http://situs.biomachina.org/fguide.html#colores

CoLoRes Tutorial

YAO CONG (LAB 1)

http://situs.biomachina.org/tutorial_colores.html

> **colores** <*Situs density map*> <PDB structure> **-res** <number> **-cutoff** <number> **-deg** <number >

EM data

PDB

Resolution

Density cutoff

Angular sampling

Correlation flag:

-corr 0 → No filter
-corr 1 → Laplacian filter

Acknowledgements:

The Scripps Research Institute

Willy Wriggers
- Julio Kovacs
- Yao Cong

Ron Milligan

References:

Willy Wriggers and Pablo Chacón. Modeling Tricks and Fitting Techniques for Multi-Resolution Structures. *Structure*, 2001, 9:779-788.

Pablo Chacón and Willy Wriggers. Multi-Resolution Contour-Based Fitting of Macromolecular Structures. *J. Mol. Biol.*, 2002, 317:375-384.

Funding:

National Institutes of Health (R01 GM62968 and P41 RR 12255)

LJIS Interdisciplinary Training Program / Burroughs Wellcome Fund