Refinement Strategies for Single Particle Structure Determination

N. Grigorieff
Goals

• Higher resolution

NSF, Fürst et al. (2003)

• Sorting of structural heterogeneity
The Prophecy

King Richard hath decreed... (QRB, 1995)

• Use 5 e⁻ per Å²
• Demand a signal-to-noise ratio of 9 or better
• Aim for 3 Å resolution
  - Thou shall need to image 13,000 molecules
  - For 6 Å, thou shall need only 7,000 images
Resolving Power

InGaAs

[101]

1.3 Å
Protein Crystals

Bacteriorhodopsin

8.3 Å
Purple Membrane

2.6 Å resolution
The Puzzle

5 parameters to determine

Additional parameters:
- CTF (3 parameters)
- Magnification
- Beam Tilt (2 parameters)

200Å
A Crazy Idea

• Assume reliable resolution measure
• Search entire parameter space for highest resolution
• Given enough images, atomic resolution is reached
• Example:
  3 angles, 1 deg step; two coordinates, 1 pixel step:
  $360 \times 360 \times 360 \times 100 \times 100 = 5 \times 10^{11}$
  13000 particles: $(5 \times 10^{11})^{13000}$ structures to search
• This is a big number!
Strategy 1: Projection Matching
Strategy 2: Alignment in Reciprocal Space
Strategy 3: MRA and Classification
Strategy 4: Maximum Likelihood

Structure for $n+1$ iteration

$$A^{(n+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\int X_i(\phi)p_i(\phi, \Theta^{(n)})d\phi}{\int p_i(\phi, \Theta^{(n)})d\phi}$$

Probability function

$$p_i(\phi, \Theta) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^M \exp\left[-\frac{|X_i(\phi)-A|^2}{2\sigma^2}\right]f(\phi | \Theta)$$

$x_i$: $i$th image

$\Theta$: model parameters

$\sigma$: noise in images

$\phi$: alignment parameters

$N$: # of images

$f$: positional probab.

N = 4000
SNR = 1/200
Maximum likelihood alignment

ML processing of 2D crystals

Crystallography

Alignment of individual unit cells using ML approach
Defocus/Astigmatism and Magnification

CTFFIND3

CTFTILT
Problem 1: Local Optima

Particle  |  Reference  |  Correlation map
Problem 2: Missing Views

> 60°
Problem 3: Heterogeneity

- Misalignment of particles
- Lower resolution in disordered regions
- Loss of features
Classification Using ML

Structure for \( n+1 \) iteration

\[
A_k^{(n+1)} = \frac{1}{\sum_i q_i^k(\Theta)} \sum_{i=1}^N \int X_i(\phi) p_i^k(\phi, \Theta^{(n)}) d\phi
\]

\[
\sum_k \int p_i^k(\phi, \Theta^{(n)}) d\phi
\]

Probability function

\[
p_i^k(\phi, \Theta) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^M \exp \left[ - \frac{|X_i(\phi) - A_k|^2}{2\sigma^2} \right] f(\phi | \Theta)
\]

Probability for class \( k \)

\[
q_i^k(\Theta) = \int p_i^k(\phi, \Theta) d\phi
\]

- \( X_i \): \( i \)th image
- \( N \): # of images
- \( \Theta \): model parameters
- \( \sigma \): noise in images
- \( \phi \): alignment parameters
- \( f \): positional probab.
Classification Using ML

SNR = 1/50
N = 2000

Correlation alignment

Difference map
Problem 4: Processing Artifacts

Aligned Particles ➔ Reference

- Low-resolution structure
- High-resolution structure

N = 1000, SNR = 1/20

- Interpolation errors
- Masking
- Negative B-factor
- …
Problem 5: Noise Bias

\[ \otimes = 0 \text{ on average} \]

> 0

for 64x64 image:
average correlation = 0.064
Seeing is NOT Always Believing

100 Images 1000 Images Reference
Resolution Measurement

Images of Particles

Alignment

Averaging

FSC

Resolution
Swiss Cheese

Dangerous:
Boosting of high-resolution terms
(application of a negative B-factor)
Gedanken Experiments

$N = 30000$
$\text{SNR} = 1/50$
Weighted Correlation

Estimated resolution

True resolution

\[
PRES(X, Y) = \frac{\sum_{k \in [0, 0.5]} \Delta \Phi_{X,Y}(k) |F_X(k)|}{\sum_{k \in [0, 0.5]} |F_X(k)|}
\]

\[
CC(X, Y) = \frac{\sum_{k \in [0, 0.5]} F_X(k) F_Y^*(k)}{\sqrt{\sum_{k \in [0, 0.5]} |F_X(k)|^2 \sum_{k \in [0, 0.5]} |F_Y(k)|^2}}
\]

\[
CC_W(X, Y) = \sum_{k \in [0, 0.5]} W_X(k) F_X(k) F_Y^*(k)
\]
Coherence Constraint

$$CC_W = \sum_i |CC|_i^3$$
Acknowledgements

• Ca Channel
  Matthias Wolf
  (Glossmann/Striessnig)

• NSF/20S
  Johannes Fürst
  (Axel Brünger)

• Noisy Face
  David DeRosier

• Financial Support:
  HHMI, NIH, NSF