Assignment 10

for lecture "Bioinformatics III" WS 08/09

Return before lecture on Nov. 20, 2008 or by email to p.walter@bioinformatik.uni-saarland.de until Feb. 15. This assignment will be discussed in the tutorial on Feb. 16, 2009, room15, building E1 3

ZBI ZENTRUM FÜR BIOINFORMATIK

Mapping of Crystal Structures into EM Maps

For some protein complexes both a low resolution image like, e.g., an AFM image or an EM density map of the whole complex and the atomic structure of the individual constituents is available. Then, one is interested in where these constituents are located in the cluster. The task is consequently to fit the position and orientation of a given structure with atomic resolution into a blurred density map such that the correlation is maximized.

To achieve a maximal overlap, the high resolution structure has to be blurred, i.e., convoluted with the experimental resolution. An efficient way to calculate the convolution of the atomic structure data with the experimental resolution is via the Fourier theorem. Therefore you will look at various properties of the Fourier transform in the first exercise. However, you will not perform a full three dimensional reconstruction of multiple fragments into a blurred complex but try to fit a two dimensional structure into a smeared image of itself.

(1) Properties of the Fourier transform (35 pts)

The (complex) Fourier transform of a function $f(x)$ is defined as:

$$FT[f(x)] = F(k) = \int dx\, \mathrm{e}^{-ikx} f(x)$$

Note the change of the variable from x to its conjugate variable k. Its inverse is, consequently,

$$f(x) = FT^{-1}[F(k)] = \frac{1}{2\pi} \int dk\, \mathrm{e}^{ikx} F(k)$$

Remember that the complex exponential is defined as $\exp[\pm ix] := \cos(x) \pm i \sin(x)$ and that the complex integral can be split up into real and imaginary parts, which can be evaluated independently.

(a) The Fourier transform and it's inverse

With a definition of the delta distribution as

$$\delta(x_1 - x_2) = \frac{1}{2\pi} \int dk\, \exp[ik(x_1 - x_2)]$$

show that $FT^{-1}[FT[f(x)]] = f(x)$, i.e., that the definitions of the Fourier transform and its inverse given above do match.

Hint: Be careful not to mix up the different (integration) variables of the Fourier transform and its inverse. Use, e.g., x1, x2, …

(b) Shift of the argument

Show that a shift of the argument of $f(x) \Rightarrow f(x + \Delta x)$ shows up as a phase factor $\exp[ik\Delta x]$ in the Fourier transform of f.

(c) Linearity
Show that $FT[f(x) + g(x)] = F(k) + G(k)$.

(d) The Convolution Theorem
The convolution of two functions $f(x)$ and $g(x)$ is defined as

$$(f * g)(x) = \int dy\, g(y)\, f(x - y)$$

Show that $FT[(f * g)](k) = F(k)\, G(k)$, i.e., that the Fourier transform of the convolution of f and g equals the product of the individual Fourier transforms of f and g.

(2) Blurring the structure (15 pts)
Calculate the convolution of a model molecule with an experimental uncertainty, which is described by a Gaussian distribution

$$g(x; x_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - x_0)^2}{2\sigma^2}\right]$$

of width σ, centered around $x0$. The density $\rho(x)$ of the model molecule is given by a sum of delta peaks with masses mi at the atom positions

$$x_i: \ \rho(x) = \sum m_i\, \delta(x - x_i)$$

Hint: Note that

$$\int dx\, \delta(x - x_0)\, f(x) = f(x_0)$$

you don't need the Fourier transform to evaluate g * ρ.

Hint: The result should be a sum of displaced Gaussians…

(3) Reconstruction of low resolution images (50 pts)
For this 2D fit you are given a file hello.dat with the atomic "structure" of the hypothetical HLO (nicknamed "hello") protein and various smeared images, where this structure was shifted, rotated, scaled, and blurred. Implement a reconstruction program with which you perform the tasks given below.

The objective is to minimize the difference between the given "experimental" maps and the blurred map from the structure. For this, use the sum of the squared differences between the two maps at each grid point.

As the center of mass is the same for the original structure and for the resulting blurred image (why?), you deal with the displacement by first determining the center of mass of the blurred map and then shifting the (rotated) structure to have the same center of mass. This will give you different offsets for different rotation angles.

Hint: for creating the blurred map, see exercise (2) above.

Hint: in the structure file each line holds, in this order, the x– and y–positions of an atom and its "mass". This mass determines, how much a given atom contributes to the image, i.e., how visible this atom is to the imaging.

Hint: You can start from the supplied Python script construct_example.py*, which was used to generate the blurred maps. This should explain you how to read and interpret the structure file and how to create a blurred image on a grid. Note that the shifting and rotation parameters saved in this script are not the ones used to generate the given density maps! This script omits the scaling.*

(a) Resolution Calibration

To calibrate the resolution to be used for the reconstruction, minimize the difference between the given map hello_shift.dat and the map generated from the “atomic” structure of hello.dat by varying the width σ for the Gaussian used to smear the high resolution structure. To generate hello_shift.dat, the structure was only displaced. Thus, the only parameter you have to change is σ, once you corrected for the offset. Give the offset and plot the sum of the squared differences against the width σ.

Keep this optimal σ for the subsequent reconstructions.

Hint: the best σ is somewhere in the range 0.1 to 0.2. Choose enough values from this interval to create the plot of the σ-dependent difference. Do not forget to give the optimal value of σ that you find.

Create a 2D plot of the smoothed image with the optimal σ. Try to include the atom positions, too.

Hint: Most spreadsheet programs give you an option to plot a coloured “height map” or a contour plot of the grided 2D data.

(b) Angular Correlation

In the next “experimental” map, hello_rotshift.dat, the HLO–protein is rotated and displaced. Calculate the difference between the given map and the blurred known structure for rotation angles between 0 and 2π in at least 100 angular steps. Plot this difference vs. the rotation angle and determine the best fit rotation angle. Plot the reconstructed image. In a second plot show the angle-dependent x– and y–offsets (two curves).

Hint: perform the rotation around the center of mass of the "atomic" structure — the given structure is not centered!

(c) Displacement, rotation and scaling

For hello_rotshiftscale.dat, the structure was scaled, rotated, and displaced. For every angular step, first determine the center of mass for the displacement and then use the radius of gyration to scale the structure. Give the sum of the squared differences for every angle and, in a second plot, the x- and y-offsets and the scaling factor. What are the best fit values? Plot this configuration.

Hint: The radius of gyration is the sum of the mass-weighted squared distances from the center of mass.

Contact: Peter Walter, phone -3613, email: p.walter@bioinformatik.uni-saarland.de