BIOINFORMATICS

Vol. 1 no. 1 2003
Pages 1–1

Fast and accurate pseudo-polar multi-resolution docking

Yosi Keller[1],....., Pablo Chacón[2]

[1]Math department. Yale University, 06520 Connecticut, USA. [2]Centro de Investigaciones Biológicas. CSIC. Ramiro de Maeztu 9. 28040 Madrid. Spain

ABSTRACT

Motivation: The interpretation of the electron microscope macromolecular 3D reconstructions in terms of available atomic structures of their components by docking techniques is a critical step for the understanding of the functioning of cellular machines. This paper introduces a novel formulation for the exhaustive and fast registration of density volumes. The proposed method accurately estimates arbitrary large rotations and translations without requiring an optimization scheme. By using a 3-D pseudo-polar Fourier transform, rotations can be reduced to translations which can be estimated with only 1D operations. Translational displacement is also efficiently recovered by means of well-know phase correlation technique.

Results: An early-stage implementation is able to dock with high precision an atomic structure into an electron density low resolution map from one to few minutes. The experimental results demonstrate that the algorithm is very accurate and robust to noise. Since the proposed algorithm is a general 3D rigid-body registration framework, we also expect a high impact to other computational biology fields.

1 INTRODUCTION

Despite of the explosive growth of research in structural biology in last decades, the atomic resolution access to large macromolecular complexes implicated in the main cellular functions still is rather limited. However, with a more limited resolution, electron microscopy (EM) techniques are able to capture such large macromolecules in diverse near-physiological conditions (Frank, 1996). EM can provide medium (10Δ) to low (30Δ) resolution 3D structures of large complexes even if only small amounts of material are available and can tolerate some sample heterogeneity. When combined with X-ray structures of components parts 3D EM reconstructions can yield a detail description of the structure and action of an entire macromolecular cellular machine. With current structural genomics efforts and the near perspective of 3D imaging macromolecules in their native context, the development of bioinformatics tools for merging 3D structural information constitutes the key to decipher the macromolecular inner workings.

There are impressive examples that already illustrates the strength of this integrative structural information approach, termed multi-resolution docking, to deliver relevant structural insights into macromolecule function: RNAP polymerases (Asturias, 2004; Darst, et al., 2002; Opalka, et al., 2003)}, actin-myosin complex(Volkmann, et al., 2000) , Ribosome (Agrawal, et al., 2004; Halic, et al., 2004; Valle, et al., 2003), bacteriophage T4 (Leiman, et al., 2004; Leiman, et al., 2003), IGG(Sandin, et al., 2004). These and many others examples, used recently developed bioinformatics tools that allow automatic building atomic models of large complexes imaged by EM using the atomic structures of their components [for a review see (Volkmann and Hanein, 2003; Wriggers and Chacon, 2001)]. The multiresolution docking programs diverged mainly on the fit function used. The match criteria is typically defined as the simple scalar product of the electron density, termed cross-correlation (Rossmann, 2000; Rossmann, et al., 2001; Volkmann and Hanein, 1999), but more elaborated functions that includes surface terms (Ceulemans and Russell, 2004; Chacon and Wriggers, 2002), or functions that emphasized the local correlation (Roseman, 2000; Wu, et al., 2003) can be used to enhance the fitting contrast. Even though these methods have been successfully employed to study molecular assemblies, additional computational development to improve their accuracy, high throughput coverage and efficiency is a strong requirement in current structural biology. (Baker and Johnson, 1996; Baumeister and Steven, 2000; Russell, et al., 2004; Sali, et al., 2003)

To bridge the resolution gap between a low-resolution EM map and an atomic structure requires that both data sets must be compared at like resolution. Density maps can be easily derived at any resolution from atomic resolution structure using several approaches (Belnap, et al., 1999; Wriggers, et al., 1999). In essence, the multi-resolution docking can be reduced to geometrically register two 3D electron density maps: the experimental EM map with a simulated map obtained by lowering the resolution of the atomic structure. This problem is equivalent to a general rigid body 3D registration problem, and it can be found in a

* To whom correspondence should be addressed.

© Oxford University Press 2003

diverse range of fields and applications, such as molecular replacement and molecular docking in structural biology (Rossmann and Arnold, 2001) (docking??), or 3-D assembly modelling (Papaioannou, et al., 2002; Wyngaerd and Van Gool, 2002) and medical imaging (Pluim, et al., 2003; Zhu, 2002) in image processing.

Let be $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$, $\mathbf{x}=(x,y,z)$ two maps related by a 3-D rigid transformation:

$$I_1(\mathbf{x}) = I_2(R\mathbf{x} + \Delta\mathbf{x}) \tag{1}$$

The multi-resolution docking process consists in estimate the 3D rotation matrix R and the translational parameter $\Delta\mathbf{x}=(\Delta x,\Delta y,\Delta z)$ that maximizes the density overlap, *i.e.* maximizing the correlation function. To this end, a full 6D rigid-body search to explore all possible docking solutions must be performed. The exhaustive exploration is needed to avoid any missing valid registration. Note that we are confronting a non trivial problem and several docking alternatives can be obtained. Even when taking into account the resolution differences and the EM low signal to noise ratio, there is not often an exact correspondence between crystal and EM data structures. EM densities may be missing small regions that are accounted for in the atomic structure (or vice versa) due to disorder or simply because for each technique we have not exactly the same biological construction. Unfortunately, the exhaustive exploration is highly computational demanding. Thus, several methods have been elaborated to speed up the fitting process. Traditionally, the computation of the correlation as a function of translational displacements can be efficiently computed as a product of Fast Fourier Transforms (FFT) (Bracewell, 1986). In the simple case, the registration combines a systematic rotational scan of a probe object relative to a fixed reference with a FFT translation accelerated search. This well-known FFT scheme was original used in ligand-protein docking programs (Katchalski-Katzir, et al., 1992) and it is the basis of many of the current docking programs (Gabb, et al., 1997; Mandell, et al., 2001; Vakser, et al., 1999). Apart from of the speeding up of the translational search, the computational cost of the systematic rotational scan makes this simple approach inefficient especially for relatively large data sets. To avoid the search sampling limitations, the application of this approach is followed by an off-lattice optimization, adding an extra computational cost.

The Fourier representation of the density maps is preferred if an extra acceleration is needed. Working on Fourier domain, i.e. considering only the signal amplitudes, is natural in crystallography field, where numerous tools related with rigid body fitting performed in reciprocal space can be found (Rossmann and Arnold, 2001). The Fourier domain allows the computation of the rotation and translation parameters separately, reducing a problem with six degrees of freedom to two problems of three degrees of freedom each. Let $\hat{I}_1(\mathbf{x})$ and $\hat{I}_2(\mathbf{x})$ he corresponding Fourier transform of the maps to be register; Eq. 1 can be expressed as:

$$\hat{I}_1(\mathbf{w}) = \hat{I}_2(R\mathbf{w})e^{-i\cdot\Delta\mathbf{w}} \tag{2}$$

By considering only the amplitudes, $A(\mathbf{w}) = |\hat{I}|$, we get:

$$A_1(\mathbf{w}) = A_2(R\mathbf{w}) \tag{3}$$

Since A_1 and A_2 are related only by the rotation R with no translation, it is possible to recover the relative rotation by registering the amplitudes with a least square minimization:

$$\min\left(\left|A_1(\mathbf{w}) - A_2(R\mathbf{w})\right|^2\right) \tag{4}$$

Note that minimize the quadratic misfit of the amplitudes is equivalent to maximizing the cross-correlation coefficient in direct space (Bracewell, 1986; Wriggers and Chacon, 2001).

Once the rotation is estimated, it is possible to provide straightforward estimation of rigid translational motion between two images using the phase correlation method (Foroosh, et al., 2002; Reddy and Chatterji, 1996). This method exploits the Fourier shift theorem which states that shifts in the spatial domain correspond to linear phase changes in the Fourier domain. Reordering and taking the inverse Fourier transform in both sides of Eq. 2, it is easy to obtain:

$$\Im^{-1}\left(\hat{I}_1 / \hat{I}_2\right) = \delta(\mathbf{x} - \Delta\mathbf{x}) \triangleq C(\mathbf{x}) \tag{5}$$

Therefore, we can recover the translation $\Delta\mathbf{x}$ by maximizing the phase correlation coefficient $C(\mathbf{x})$.

State of the art 3D object registration algorithms in image processing estimates both translation and rotation in Fourier domain (Hoge, 2003; Lucchese, et al., 2002)**.** To recover the rotation parameters, these algorithms normalize the Fourier transform of the input objects, and integrate it in the radial direction. The direction in which this integral is minimal gives the direction of the rotation axis. Once it recovers the rotation parameters, the algorithm recovers the translation parameters by using phase correlation. The integration in the radial direction suffers from inaccuracies caused by discretization that affect the docking accuracy.

More recently an extra acceleration has been obtained in muliresolution docking by accelerating the rotational search part (Kovacs, et al., 2003). This formulation can be considered as a 6D extension of Crowther method widely used for molecular replacement in X-ray crystallography (Navaza, 2001; Rossmann and Arnold, 2001; Vagin and Teplyakov, 1997). The density objects are expressed in spherical harmonics, which requires a radial integration that again may introduce distortions that could affect the accuracy. This very efficient method also suffers from strong memory limitations if high precision is needed. Finally, without the warrantee of exhaustive search, faster alternatives based on the

reduction of complexity of the EM maps using neural networks (Wriggers, et al., 1999)or considering only the density surface (Ceulemans and Russell, 2004)can be used. In summary, although there are many approximations to perform multi-resolution docking, efficient and robust implementation of a 6D search algorithm is still an open problem.

In this paper we present a new Fourier domain based approach, which does not require any interpolation to recover the rotation. It is based on pseudo-polar FFT (Averbuch, et al., 2004; Averbuch and Shkolnisky, 2003; Keller and Averbuch, 2004; Keller, et al., 2004), which compute the Discrete Fourier Transform (DFT) on an over-sampled set of angularly non-equispaced frequencies, termed pseudo polar grid. The transformation of the 3D electron density maps into this uniform polar representation, where rotations are reduced to translations, allows the implementation of a fast and algebraically accurate registration algorithm. The proposed new methodology follows the Euler's theorem for the estimation of the 3-D registration parameters in three steps. First, the 3D pseudo-polar transform (PPFT3D) is used to recover the rotation axis. Then, the rotation around the axis is estimated using a pseudo-cylindrical representation computed with the 2D pseudo-polar transform (PPFT2D). Finally, the translation is computed by using phase correlation. The algorithm accurately estimates arbitrary large rotations and translations without requiring an optimization scheme. Moreover, the implementation requires only 1-D operations and is appropriate for real-time implementations. The experimental results show that the algorithm is accurate and robust to noise.

2 SYSTEMS AND METHODS

The quantitative multiresolution docking is reduced to find the pose which maximizes the density overlap between a target experimental EM map and a probe simulated map obtained from the atomic structure available components. To achieve maximal efficiency we define the 6D exhaustive search in Fourier domain, where rotational and translational parts can be conveniently separated. In particular, our translational search implementation is build from the state of the art phase correlation 2D registration algorithms that can achieve sub-lattice estimation accuracy (Hoge, 2003; Keller and Averbuch, 2004; Reddy and Chatterji, 1996). The problem of rotations is by far more complex than the translational shifts. Here we propose an original rotation estimation procedure which accomplishes the transformation of rotations in translations on a spherical grid. This innovative approach yields a complexity reduction without precedents.

The rotation estimation algorithm is based on Euler's theorem, which allows representing a rotation using only three angular parameters. An arbitrary 3D rotation can be expressed as θ angle rotation around an axis given by a unit vector $\mathbf{n} = (n_x, n_y, n_z)$ whose direction is given by the angles α and β. The rotation matrix can be expressed as (Trucco and Verri, 1998):

$$R = I\cos\theta + (1-\cos\theta)\begin{bmatrix} n_x^2 & n_x n_y & n_x n_z \\ n_y n_x & n_y^2 & n_y n_z \\ n_z n_x & n_z n_y & n_z^2 \end{bmatrix} + \sin\theta \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ n_y & n_x & 0 \end{bmatrix} \quad (6)$$

where I is the identity matrix. Both θ and $\mathbf{n}$ can be easily recovered from the rotation matrix. Since $\mathbf{n}$ is invariant under the transformation of R, i.e. $R\mathbf{n} = \lambda\mathbf{n}$, $\mathbf{n}$ is an eigenvector of R. The three eigenvalues of R are λ_1=1 and $\lambda_{2,3} = e^{\pm j\theta}$. The rotation axis $\mathbf{n}$ is the eigenvector corresponding to λ_1, and θ can be recovered from $\lambda_{2,3}$. This allows to split the 3D rotation evaluation in two steps: the estimations of the rotation axis $\mathbf{n}$ and the rotation angle θ.

The rotation axis $\mathbf{n}$ can be recovered by finding the vector along which the electronic density difference between the target map, I_1, and probe map, I_2, is minimal. At this point it is convenient to work in Fourier domain and compute the difference between A_1 and A_2, the amplitudes of I_1 and I_2. We define now the 3-D angular difference function (ADF) as:

$$\Delta A(\alpha,\beta) = \int_0^\infty \left| A_1(\alpha,\beta,r) - A_2(\alpha,\beta,r) \right|^2 dr \quad (7)$$

where (α,β,r) are spherical coordinates and

$$(\alpha_0,\beta_0) = \arg\min \Delta A(\alpha,\beta) \quad (8)$$

are the parameters of the rotation axis $\mathbf{n}$. Ideally $\Delta A(\alpha_0,\beta_0)$=0, but due to noise and partial overlapping, we can find rotation axis by simply scanning over ΔA for its minima (Eq. 8). Unfortunately, there is no fast algorithm for computing the Fourier transform of the 3D density data in spherical coordinates. But, here we devise the use of a transformation of the 3D electron density maps into a uniform pseudo polar representation (Averbuch, et al., 2004; Averbuch and Shkolnisky, 2003), where rotations can be reduced to translations which can be FFT accelerated. Based on this novel transformation we propose a fast and algebraically accurate scheme for the computation of the difference angular function ΔA.

The rotation θ can be estimated by taking into account that I_1 and I_2 are related by a pure rotation:

$$I_1(R_n\mathbf{x}) = I_2(R_z(\theta)R_n\mathbf{x}) \quad (9)$$

where R_n is a 3-D rotation which aligns $\mathbf{n}$ with the Z axis, and $R_z(\theta)$ is a rotation of angle θ about the Z axis. If the volumes are rotated such that the rotation axis $\mathbf{n}$ is parallel to the Z axis, the resulting volumes will be related only by a planar rotation of angle θ. To estimate this angle we use

again the pseudo polar representation, but now in 2D. This procedure is a modification of the 2D rotation registration algorithm given in (Keller, et al., 2004).

In summary, the proposed multi-resolution, first estimates the rotation axis, second computes the planar rotation relative to the rotation axis, and finally recovers the translational displacement by using the phase correlation technique. Since the key to achieve a further acceleration in the rotational search lies in the employ of a pseudo polar transform, it is first introduced before each docking step is described in detail.

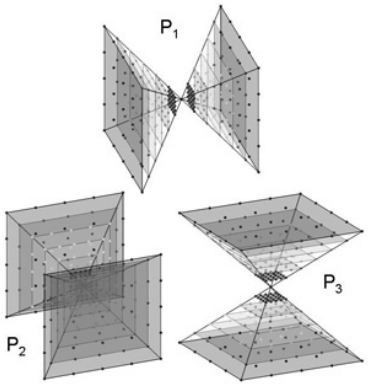

Fig. 1. The 3-D pseudo-polar sectors. The 3D pseudo polar grid P is formed by the intersection of three showed pseudo polar sectors; mathematically $P = P_1 \cup P_2 \cup P_3$.

2.1 The 3-D pseudo-polar transform

Given a 3-D density map I of size $N\times N\times N$, its 3-D discrete Fourier transform (DFT) is:

$$\hat{I}^{Cart}_{m,k,l} = \sum_{u,v,w=-N/2}^{N/2-1} I_{u,v,w}\, e^{-\frac{2\pi j}{M}(um+vk+wl)} \tag{10}$$

with $m,k,l = -\frac{M}{2},...,\frac{M}{2}-1$ where parameter M ($M\geq N$) sets the frequency resolution. Instead sampled on the Cartesian grid for some applications it is desirable to compute the Fourier transform of I on the spherical grid defined by:

$$\begin{aligned} &u = r_m \cos\theta_k \sin\phi_l, v = r_m \sin\theta_k \sin\phi_l, w = r_m \cos\phi_l, \\ &r_m = m, \theta_k = 2\pi k/K, \phi_l = \pi l/L, \\ &m = 0,...,M-1, k = 0,...,K-1, l = 0,...,L-1 \end{aligned} \tag{11}$$

Hence, the DFT in spherical coordinates has the form:

$$\hat{I}^{Sph}_{m,k,l} = \sum_{-N/2}^{N/2-1} I_{u,v,w}\, e^{-\frac{2\pi j\, r_m}{M}(u\cos\theta_k \sin\phi_l + v\sin\theta_k \sin\phi_l + w\cos\phi_l)} \tag{12}$$

Note that the spherical grid (Eq. 11) is equally spaced both in the radial and angular directions

$$\begin{aligned} \Delta r = r_{m+1} - r_m = 1, \qquad \Delta\theta = \theta_{k+1} - \theta_k = \frac{2\pi}{K}, \\ \Delta\phi = \phi_{l+1} - \phi_l = \frac{\pi}{L}, \end{aligned} \tag{13}$$

Even though there is no fast algorithm for computing the Fourier transform of the map in spherical coordinates we use a clever alternative: the 3-D pseudo-polar Fourier transform**.** PPFT3D samples the 3-D Fourier transform (Eq. 12) of the volume on the 3-D pseudo-polar grid (Eq. 15) defined by the set of samples: $P = P_1 \cup P_2 \cup P_3$ (see Figure 1):

$$\begin{aligned} P_1 &\triangleq \{(m, -\frac{2k}{N}m, -\frac{2l}{N}m)\}, \\ P_2 &\triangleq \{(-\frac{2k}{N}m, m, -\frac{2l}{N}m)\} \\ P_3 &\triangleq \{(-\frac{2k}{N}m, -\frac{2l}{N}m, m)\} \end{aligned} \tag{14}$$

where $k,l = -\frac{N}{2},...,\frac{N}{2}$, and $m = -\frac{3N}{2},...,\frac{3N}{2}$. Formally, the PPFT3D, denoted by $\hat{I}^{PP}_s$ (s=1,2,3), is the linear transformation defined by:

$$\begin{aligned} \hat{I}^{PP}_1(m,k,l) &= \sum_{u,v,w=-N/2}^{N/2-1} I(u,v,w) e^{-\frac{2\pi j}{M}(mu-\frac{2k}{N}mv-\frac{2l}{N}mw)} \\ \hat{I}^{PP}_2(m,k,l) &= \sum_{u,v,w=-N/2}^{N/2-1} I(u,v,w) e^{-\frac{2\pi j}{M}(-\frac{2k}{N}mu+mv-\frac{2l}{N}mw)} \\ \hat{I}^{PP}_3(m,k,l) &= \sum_{u,v,w=-N/2}^{N/2-1} I(u,v,w) e^{-\frac{2\pi j}{M}(-\frac{2k}{N}mu-\frac{2l}{N}mv+mw)} \end{aligned} \tag{15}$$

As we can see from Figure 1, for fixed angles k and l, the samples of the 3-D pseudo-polar grid are equally spaced in the radial direction. However, this spacing is different for each angle. Also, the grid is not equally spaced in the angular direction, but has equally spaced slopes. Two important properties of the PPFT3D are that it is invertible and that both the forward and inverse pseudo-polar Fourier transforms can be implemented using fast algorithms. Moreover, the implementations require only 1-D equispaced FFT's. In particular, the algorithms do not require re-gridding or interpolation. A detailed description of how compute the PPFT3D is shown in Appendix A. The overall complexity to compute the PPFT3D is $O(M^3 \log M)$.

2.2 Rotation axis estimation

An important property of $\Delta A(\alpha,\beta)$, given by Eq. 7, is that it can be discretized using very general sampling grids. Specifically, ΔA does not require an exact uniform spherical representation of the Fourier transforms of I_1 and I_2. We use a discretization of ΔA based on the PPFT3D, denote by ΔA^{PP}, that can be calculate by:

(1) Compute A_1^{PP} and A_2^{PP}, the magnitudes of the PPFT3D of the 3D density volumes.

(2) Evaluate the discrete ADF Eq. 7 as:

$$\Delta A^{PP} = \sum_{0\le r_k\le\pi}^{N} \left| A_1^{PP}\left(r_k,\alpha_i,\beta_j\right) - A_2^{PP}\left(r_k,\alpha_i,\beta_j\right)\right| \Delta r_{ij} \quad (16)$$

where Δr_{ij} is the radial sampling interval. Note that the integration is computed over rays of the same length ($0\le r_k\le\pi$) within the sphere bounded in a N×N×N cube in the Fourier domain.

The PPFT3D was shown to be algebraically accurate (Averbuch, et al., 2004; Averbuch and Shkolnisky, 2003), and hence, the approximation error in Eq. 16 results from the integration being computed over $0 \le r_k \le r_{ij}^{\max}$ instead of $0 \le r_k \le \pi$, where for each ray $r_{ij}^{\max}$ is the sample nearest to $r=\pi$ such that $r_{ij}^{\max} \le \pi$. The integration interval error for each ray is:

$$\left|\pi - r_{ij}^{\max}\right| < \pi\sqrt{2}/N$$

Furthermore, the interval $[r_{ij}^{\max}, \pi]$ is located at the high frequency range, where the magnitude of the PPFT3D is negligible. As given by Eq. 8, the angles of rotation axis can be estimated by finding the minima of ΔA^{PP}, that is:

$$(\alpha_0,\beta_0) = \arg\min \Delta A^{\mathrm{PP}}(\alpha,\beta) \quad (17)$$

The rotation can be directly calculated from:

$$\boldsymbol{n} = \left[\cos\alpha_o \sin\beta_o, \sin\alpha_o \sin\beta_o, \cos\alpha_o\right]^T \quad (18)$$

2.3 Planar Rotation Estimation

After recovering the rotation axis **n**, the density volumes are related by a rotation which aligns **n** to Z axis and by the rotation angle θ about the Z axis (see Eq. 9). To estimate the planar rotation, first, **n** is aligned to Z axis and, subsequently θ is estimated by using a variant of a 2-D rotation estimation algorithm presented in (Keller, et al., 2004). Quaternions are employed to compute the rotation that aligns **n** with the Z axis (Hearn and Baker, 1995). Specifically, given vectors $\mathbf{n}_1,\mathbf{n}_2 \in R^3$, the rotation that transform $\mathbf{n}_1$ to $\mathbf{n}_2$ is given by the rotation axis **n** and the angle ψ:

$$\psi = \arccos\left(\frac{\mathbf{n}_1\cdot\mathbf{n}_2}{|\mathbf{n}_1||\mathbf{n}_2|}\right) \qquad \mathbf{n} = \mathbf{n}_1 \times \mathbf{n}_2$$

Using these equations, setting $\mathbf{n}_1$=(0,0,1) and computing $\mathbf{n}_2$ using Eq. 3, the alignment parameters can be directly calculated. Then, the input density volumes I_1 and I_2 are rotated such the axis **n** is made parallel to Z axis. The resulting rotate replicas, $\tilde{I}_1$ and $\tilde{I}_2$, are now only related by a 3D rotation and a 2D translation of the XY planes about the Z axis. The Z axis alignment and the resulting geometric relationship between the rotate replicas are illustrated in

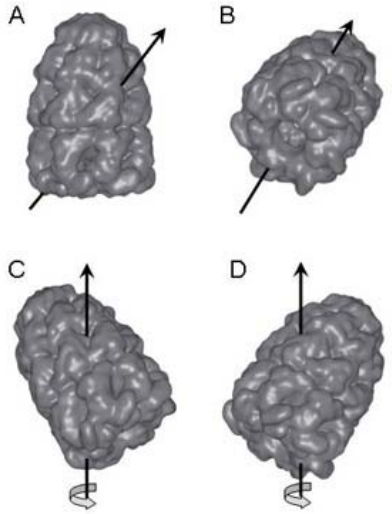

Figure 2.

Fig. 2. Alignment of rotation axis. The input volumes A and B are related by a rotation about the rotation axis. After recovering the rotation axis, the volumes in C and D are related by a translation and a planar rotation. T

To recover the relative planar rotation of $\tilde{I}_1$ and $\tilde{I}_2$ is convenient use a cylindrical grid in Fourier domain, where considering only the amplitude, the rotation can be estimated regardless the translation by using any corresponding pair of XY planes. We use a pseudo-cylindrical representation computed with the 2D pseudo-polar transform (PPFT2D) to recover the relative cylindrical motion. Using the separability of the 3D Fourier transform first we can compute the 1-D FFT transform in the *Z* axis:

$$\hat{I}_Z(\theta_i,r_i,\omega_k) = FFT_Z^{1D}\ \mathrm{I}(\theta_i,r_i,\omega_k)$$

And then we can compute the 2-D Polar Fourier transform of each *XY* plane in $I_z(\theta_i,r_i,\omega_k)$:

$$I^C(\theta_i,r_i,\omega_k) = PPFT2D\left\{\hat{I}_Z(\theta_i,r_i,\omega_k)\right\}$$

Denote

$$\tilde{A}_1 = \left|\tilde{I}_1^C\right|, \quad \tilde{A}_2 = \left|\tilde{I}_2^C\right|,$$

the amplitudes of the pseudo polar cylindrical FFT. These amplitudes are related by a planar rotation around the *Z* axis, with no relative translation. In other words, each *XY* plane in $\tilde{A}_1$ is a rotated replica of the corresponding plane in $\tilde{A}_2$:

$$\tilde{A}_1(\theta,r,\omega_z) = \tilde{A}_2(\theta+\Delta\theta,r,\omega_z) \quad (19)$$

Next, we recover the rotation $\Delta\theta$ by computing the difference function $\Delta\tilde{A}(\theta)$:

$$\tilde{A}(\theta) = \int_0^{\pi}\int_0^{\pi} \left|\tilde{A}_1(\theta, r, \omega_z) - \tilde{A}_2(-\theta, r, \omega_z)\right| dr\, d\omega_z \qquad (20)$$

Where $\tilde{A}_2(-\theta, r, \omega_z)$ corresponds to a volume where each *XY* plane is left-right flipped.

Because of the Hermitian symmetry of the Fourier Transform the rotation angle can be either $\Delta\theta$ or $\Delta\theta+\pi$. This ambiguity can be resolved by rotating A_1 using the rotations $(\alpha_0,\beta_0,\Delta\theta)$ and $(\alpha_0,\beta_0,\Delta\theta+\pi)$, and recovering the translation by choosing the rotation angle that yields the highest correlation peak.

2.4 3D translation estimation

Given the rotation parameters $(\alpha,\beta,\Delta\theta)$, the 3-D rotation matrix *R* can be calculated by Eq. 6. The resulting rotated replicas obtained by apply *R* to the input volumes are only related by a 3-D translation, which is recovered using the phase-correlation algorithm (Hoge, 2003; Reddy and Chatterji, 1996). The phase correlation method is based on the well-known Fourier shift property, and is widely used because estimation of motion in the phase domain is exceptionally robust in the presence of noise (Foroosh, et al., 2002). However, the accuracy of the phase correlation scheme is limited by the lattice translational sampling (i.e. grid size) of the density maps. For sub-lattice translations will cause the peak of the phase correlation function C(**x**) (Eq. 4) that measures the quality of the alignment to spread to neighboring voxels. Fortunately, sub-lattice accuracy and improved robustness to noise can be achieved by applying recent algorithm developments (Hoge, 2003; Keller and Averbuch, 2004). Please explain with a sentence or two the sub-lattice algorithm.

3 ALGORITHM

3.1 Method improvements

Multiresolution docking may lead to ambiguous matches or false positives. This can critical when the resolution is sufficient low that the internal structure enables the recognition or to localize small components in large low resolution. Several alternatives can be adopted to improve fitting contrast most of them compatibles with the general docking framework proposed. For example, the fit can be changed by a local correlation criterion (Roseman, 2000), or maps can be pre-filter with a Laplacian kernel (Chacon and Wriggers, 2002). The filtering enhances the numerical contrast among potential solution by maximizing both density and contour overlap. Since its implementation does not require any change in the registration scheme, here we adopt the Laplacian Filter approach. Thus, the docking is preformed with Laplacian filter maps instead the original density maps.

3.2 Implementation details

The proposed multiresolution algorithm has been implemented in C++. Although is a proof of concept implementation, it permits to perform an exhaustive docking search with high precision with times compatibles with the faster multiresolution docking tools. A docking case takes approximately one minute (70sec) for a 64^3 map, and <20 minutes for a 128^3 on a 2.8Ghz PC computer. Note that the timing is only a function of the size of the registered maps and not to their content. We anticipate at least a 3 fold speed up can be obtained by code optimization. Since in rotational and translational estimation parts the kernel calculations are independent 1D DFT, the method is a perfect candidate to be parallelized. Future method developments will be pursuing real time implementation with both code improvements and parallelization.

It is important to mention that the time consuming part of the algorithm (>99%) is the PPFT3D calculation. The precomputation of such transforms reduces the angular registration to simple differences operations (Eq 16 & 20). This procedure will permit to dock two volumes almost interactively (~2 sec, is this too optimistic?). Thereby, the proposed methodology will be extremely important to perform fast and accurate docking queries in a database framework-such the recently released EM macromolecular structure Database(http://www.ebi.ac.uk/msd-srv/emdep). The near necessity to combine the information stored in this database with other structural databases (e.g. PDB) could be carried out by our methodology.

Talk about angular spacing of the 3-D pseudo-polar grid v.s. accuracy

4 DISCUSSION

The novel docking algorithm was evaluated using ten simulated EM density maps at various resolution levels according to experimental limits. The systematic test consists in recover the original positions and orientations of several rotated and translated replicas of the atomic structure used for generating the simulated maps. To have statistical significance, the registration process has been repeated ten times for each macromolecule at five different resolutions. We also evaluated the method performance on the challenging cases when the structure to be docked represents only a portion of the EM density. We repeat the test procedure to dock several rotated and translated monomeric atomic structure inside the corresponding maps of the hetero-oligomeric structures included in the test dataset. In figure 3A the absolute error of both translation and rotation parameters are shown as a function of the resolution. The accuracy of the proposed algorithm is in the range of ~1.2º for rotation es-

timation, which corresponds to the average angular spacing of the 3-D pseudo-polar grid which is?. The sub-lattice translation is clearly achieved, being error range in the estimation of the position ~0.5Å, which is always below to the lattice size (2-3Å). These values correspond to rmsd of 1-2Å between the docked and the original atomic structures. Note that the resolution dependence is marginal. The algorithm is able to find the correct pose with high precision even in the low resolution limit. In figure 4A, a typical docking result is shown…..

As in the previous case, the obtained results demonstrate the accuracy and robustness of the proposed approach. The accuracy registration values….
In figure 4B, the ….

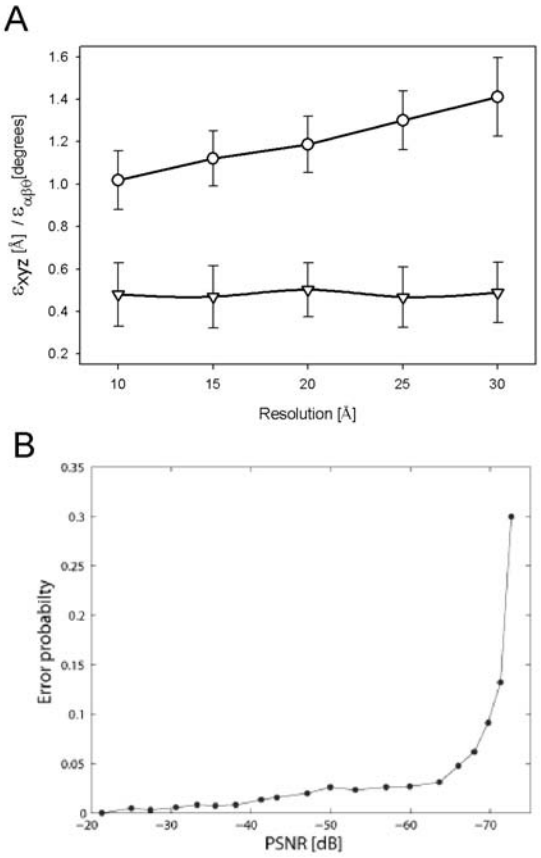

Fig. 3. The docking tests were performed on simulated EM maps calculated from atomic structure by lowering its resolution. The chosen resolutions (10Å, 15Å, 20Å, 25Å and 35Å) correspond to EM measurements experimental range. The test structures include prefoldine (??), R factor II (1gqe), ribosome (1ffk and 1fjf), RNA-pol II *E. coli* (), RNA-pol II yeast(1g03, 1l6h) and the hetero oligomers of pilin (pentamer, 2pil), nitrite reductase (trimer, 1nic), catalase (tetramer,7cat), dehidrogenase (octamer, 1aw5), groel (dodecamer); thermosome (hexadecameric, ther); The grid size of the maps was chosen at the end N=64, 128, or varies as a function of the map size? The tests replicas have been obtained by randomly rotate and translate the original structure in the range of ±10Å and ±180º in each axis, respectively. The angular difference (circles) between the original and the docked volumes with highest correlation value was measured by: $\varepsilon_{\alpha\beta\theta} = arcCos(0.5(Tr(R_0 \times R_D)^T - 1))$; where the R stands for rotation matrix, Tr for the trace, and T for the transpose, and the subscripts 0 and R are referred to the original and docked positions, respectively. The translation error (triangles) was measured with standard Euclidian distance defined by: $\varepsilon_{xyz} = \left((x_0 - x_D)^2 + (y_0 - y_D)^2 + (z_0 - z_D)^2\right)^{1/2}$. Panel B) Noised cases. Definition error probability, PSNR.

In order to verify the algorithm robustness it was applied to noisy EM maps. The results were computed by adding a range of white Gaussian noise signals to the test macromolecular EM maps. Figure 3B presents the docking accuracy as a function of the signal to noise ratio SNR. Up to SNR=?? the scheme is not affected by the noise, and from that point on, its accuracy degrades as a function of the noise. Please explain the noised experiment, how many runs, the way you add noise … etc.
In figure 4B, a noised docking result is shown…..

To effectively test the method we used with experimental EM datasets. The test includes prefoldine, Groel, RNA polymerase II E.Coli, RNA polymerase ??. The results in all cases were found in agreement with those obtained in earlier and much more expensive evaluation (Colores tool of Situs package ref). In figure 4C and 4D….

This work presents an original 3D data registration technique which has several points of interest. On the conceptual side, it is the first multi-resolution docking tool which accelerated both rotational and translational search in the frequency domain achieving high accuracy. By using Euler's theorem, the original problem, which involves six parameters, is decoupled into three sub-problems: estimating the rotation axis, estimating the planar rotation, and computing the translation. A significant characteristic is that the estimation of the rotation axis and the planar rotation is reduced to the computation of an angular difference function. Their computation is based on the PPFT2D and PPFT3D and is algebraically accurate and only needs 1D operations. Thus, on the practical side, the proposed scheme can considerably cut down the working time for aligning with precision large sets of 3D data objects, as it is typically the case of the electronic density maps in multi-resolution docking.
Besides its efficiency, the most significant advantage by this approach is the precision obtained without any post refinement scheme. The algorithm yields very accurate (~1.2º-0.5Å) registrations for a variety of simulated multi-resolution docking test cases. The results obtained noised

and experimental maps confirm and validate the method robustness for the quantitative docking of atomic structures of components into low-resolution maps of macromolecular complexes. The level of performance in both speed and registration precision open up a new application window, where fast and robust 6D exhaustive search are required. Even at this early stage of development, the docking tool may be rather helpful in practical situations where the 3D objects to be aligned are of the order of a few thousands, because it can cut the computational time needed for pairwise alignment. For example, it is frequent to explore for a given low resolution structure different docking candidates. This can be even more extreme if the docking search is performed in a database framework. In this context, by precalculating the PPF transforms the rotational search can be reduced to simple angular difference functions that take only a couple seconds. In this context, the proponed optimized methodology will be extremely important to perform fast and accurate docking queries, something faraway from the application range of current multi-resolution docking tools.

We believe that this tool and their future improvements will effectively contribute to making as automatic as possible 3D macromolecular modeling, even in a challenging database context. The proposed docking tool will have a wide impact since it anticipates the near future strong demand of tools for merging structural information at different levels of resolution. This demand is currently effective with the massive atomic resolution data available, and the different outgoing projects dedicated to obtain high throughput of experimental data from EM. Since the method constitutes a general 3D registration algorithm, the application range could be extended to other fields. Future work includes the application of the proposed algorithm to protein-ligand and protein-protein docking.

ACKNOWLEDGEMENTS

PC was supported by funds from MEC BFU2004-01282/BMC and Fundación BBVA. We would like to thanks Prof. JM Valpuesta (CNB(CSIC), Spain) for Prefoldin and F. Asturias (The Scripps Institute, USA) for RNA polymerase 3D reconstruction.

APPENDIX A

Computation details of 3D pseudo-polar transform.

The algorithm for computing the PPFT3D is based on the fractional Fourier transform. The fractional Fourier transform (Bailey and Swarztrauber, 1991), with its generalization given by the chirp z-transform (Rabiner, et al., 1969), is an algorithm that evaluates the Fourier transform of a sequence X on any equally spaced set of N points on the unit circle. Specifically, given a vector X of length N, $X = (X(u), u = -N/2, ..., N/2-1)$, and an arbitrary $\alpha \in R$ the fractional Fourier transform is defined as

$$(F^{\alpha} X)(k) = \sum_{u=-N/2}^{N/2-1} X(u) e^{-2\pi j \alpha k u / N}, k = -N/2, ..., N/2.$$

The fractional Fourier transform samples the spectrum of X at the frequencies: $\omega_k = \alpha k$, and its complexity for a given vector X of length N and an arbitrary $\alpha \in R$ is $O(N\log N)$ operations.

The algorithm for computing the 3-D pseudo-polar Fourier transform samples the Fourier transform of an image I on the pseudo-polar grid, with arbitrary frequency resolution in the radial and angular directions. The algorithm we present uses frequency resolution of $3N+1$ in the radial direction and $N+1$ in the angular directions.

The algorithm to compute, $\hat{I}_1^{PP}$ (Eq. 16), starts by zero padding the image to size $(3N+1)\times N\times N$ and applying the 1D DFT along the z direction, denoted by $\hat{I}(m,k,\cdot)$. Then for each m and k we compute:

$$T_1(m,k,\cdot) = G_{m,n}(\hat{I}(m,k,\cdot)) \quad \text{being} \quad G_{m,n} \triangleq F^{2k/n} \circ F^{-1}$$

Where F^{α} denoted the 1D Fractional Fourier transform with factor α. The operator F^{α} accepts a sequence of length N, pads it symmetrically to length $3N+1$, applies to it the fractional Fourier transform with factor α, and returns the $N+1$ central elements. Subsequently, for each m and l (along the y direction) we compute:

$$T_1'(m,\cdot,l) = G_{m,n}(T_1(m,\cdot,l)).$$

And finally for each m,k,l we get:

$$\hat{I}_1^{PP}(m,k,l) = T_1'(m,-k,-l).$$

The complexity of the algorithm for computing $\hat{I}_1^{PP}$ is $O(N^3\log N)$, i.e. N (along z) by N (along y) 1D fractional Fourier transforms. Since, the algorithm for computing $\hat{I}_2^{PP}$ and $\hat{I}_3^{PP}$ is equivalent to $\hat{I}_1^{PP}$, the total complexity of computing the full PPF3D remains $O(N^3\log N)$.

REFERENCES

Agrawal, R.K., Sharma, M.R., Kiel, M.C., Hirokawa, G., Booth, T.M., Spahn, C.M., Grassucci, R.A., Kaji, A. and Frank, J. (2004) Visualization of ribosome-recycling factor on the Escherichia coli 70S ribosome: functional implications, *Proc Natl Acad Sci U S A*, **101**, 8900-8905.

Asturias, F.J. (2004) RNA polymerase II structure, and organization of the preinitiation complex, *Curr Opin Struct Biol*, **14**, 121-129.

Averbuch, A., Donoho, D., Coifman, R., Israeli, M. and Shkolnisky, Y. (2004) Fast slant stack: A notion of Radon transform for data in cartesian grid which is rapidly computable, algebri-

cally exact, geometrically faithful and invertible, *SIAM Scientific Computing*.

Averbuch, A. and Shkolnisky, Y. (2003) 3D Fourier based discrete Radon transform, *Applied and Computational Harmonic Analysis*, **15**, 33-69.

Bailey, D.H. and Swarztrauber, P. (1991) The Fractional Fourier Transform and Applications, *SIAM Review*, **33**, 389-404.

Baker, T.S. and Johnson, J.E. (1996) Low resolution meets high: towards a resolution continuum from cells to atoms, *Curr Opin Struct Biol*, **6**, 585-594.

Baumeister, W. and Steven, A.C. (2000) Macromolecular electron microscopy in the era of structural genomics, *Trends Biochem Sci*, **25**, 624-631.

Belnap, D.M., Kumar, A., Folk, J.T., Smith, T.J. and Baker, T.S. (1999) Low-resolution density maps from atomic models: how stepping "back" can be a step "forward", *J Struct Biol*, **125**, 166-175.

Bracewell, R.N. (1986) *The Fourier transforms and its applications*. McGraw-Hill, New York.

Ceulemans, H. and Russell, R.B. (2004) Fast fitting of atomic structures to low-resolution electron density maps by surface overlap maximization, *J Mol Biol*, **338**, 783-793.

Chacon, P. and Wriggers, W. (2002) Multi-resolution contour-based fitting of macromolecular structures, *J Mol Biol*, **317**, 375-384.

Darst, S.A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G. and Wriggers, W. (2002) Conformational flexibility of bacterial RNA polymerase, *Proc Natl Acad Sci U S A*, **99**, 4296-4301.

Foroosh, H., Zerubia, J.B. and Berthod, M. (2002) Extension of phase correlation to subpixel registration, *IEEE Trans. Image Processing*, **11**, 188-200.

Frank, J. (1996) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Academic Press, San Diego.

Gabb, H.A., Jackson, R.M. and Sternberg, M.J. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information, *J Mol Biol*, **272**, 106-120.

Halic, M., Becker, T., Pool, M.R., Spahn, C.M., Grassucci, R.A., Frank, J. and Beckmann, R. (2004) Structure of the signal recognition particle interacting with the elongation-arrested ribosome, *Nature*, **427**, 808-814.

Hearn, D. and Baker, M.P. (1995) *Computer Graphics, C version*. Prentice-Hall, New Jersey.

Hoge, W. (2003) Subspace identification extension to the phase correlation method, *IEEE Transactions on Medical Imaging*, **22**, 277-280.

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C. and Vakser, I.A. (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, *Proc Natl Acad Sci U S A*, **89**, 2195-2199.

Keller, Y. and Averbuch, A. (2004) A projection-based extension of the phase correlation method, *IEEE Transactions on Signal Processing*, Submitted.

Keller, Y., Shkolnisky, Y. and Averbuch, A. (2004) The angular difference function and its application to image registration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, To appear.

Kovacs, J.A., Chacon, P., Cong, Y., Metwally, E. and Wriggers, W. (2003) Fast rotational matching of rigid bodies by fast Fourier transform acceleration of five degrees of freedom, *Acta Crystallogr D Biol Crystallogr*, **59**, 1371-1376.

Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V. and Rossmann, M.G. (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host, *Cell*, **118**, 419-429.

Leiman, P.G., Shneider, M.M., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V. and Rossmann, M.G. (2003) Structure and location of gene product 8 in the bacteriophage T4 baseplate, *J Mol Biol*, **328**, 821-833.

Lucchese, L., Doretto, G. and Cortelazzo, G. (2002) A frequency domain technique for range data registration, *IEEE transactions on Pattern Analysis and Machine Intelligence*, **24**, 1468-1484.

Mandell, J.G., Roberts, V.A., Pique, M.E., Kotlovyi, V., Mitchell, J.C., Nelson, E., Tsigelny, I. and Ten Eyck, L.F. (2001) Protein docking using continuum electrostatics and geometric fit, *Protein Eng*, **14**, 105-113.

Navaza, J. (2001) Implementation of molecular replacement in AMoRe, *Acta Crystallogr D Biol Crystallogr*, **57**, 1367-1372.

Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W. and Darst, S.A. (2003) Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase, *Cell*, **114**, 335-345.

Papaioannou, G., Karabassi, E. and Theoharis, T. (2002) Reconstruction of three-dimensional objects through matching of their parts, *IEEE transactions on Pattern Analysis and Machine Intelligence*, **24**, 114-124.

Pluim, J.P., Maintz, J.B. and Viergever, M.A. (2003) Mutual-information-based registration of medical images: a survey, *IEEE Trans Med Imaging*, **22**, 986-1004.

Rabiner, L., Schafer, R. and Rader, C. (1969) The chirp z-transform algorithm, *IEEE Transactions on Audio Electroacoustics*, **AU-17**, 86-92.

Reddy, B.S. and Chatterji, B.N. (1996) An FFT-based technique for translation, rotation, and scale-invariant image registration, *IEEE Transactions on Image Processing*, **3**, 1266-1270.

Roseman, A.M. (2000) Docking structures of domains into maps from cryo-electron microscopy using local correlation, *Acta Crystallogr. D*, **56**, 1332-1340.

Rossmann, M.G. (2000) Fitting atomic models into electron-microscopy maps, *Acta Crystallogr. D*, **56**, 1341-1349.

Rossmann, M.G. and Arnold, E. (2001) *Crystallography of Biological Macromolecules*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Rossmann, M.G., Bernal, R. and Pletnev, S.V. (2001) Combining electron microscopic with x-ray crystallographic structures, *J Struct Biol*, **136**, 190-200.

Russell, R.B., Alber, F., Aloy, P., Davis, F.P., Korkin, D., Pichaud, M., Topf, M. and Sali, A. (2004) A structural perspective on protein-protein interactions, *Curr Opin Struct Biol*, **14**, 313-324.

Sali, A., Glaeser, R., Earnest, T. and Baumeister, W. (2003) From words to literature in structural proteomics, *Nature*, **422**, 216-225.

Sandin, S., Ofverstedt, L.G., Wikstrom, A.C., Wrange, O. and Skoglund, U. (2004) Structure and flexibility of individual im-

munoglobulin G molecules in solution, *Structure (Camb)*, **12**, 409-415.

Trucco, E. and Verri, A. (1998) *Introductory Techniques for 3-D Computer Vision.* Prentice-Hall, New Jersey.

Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement, *J. Appl. Cryst.*, **30**, 1022-1025.

Vakser, I.A., Matar, O.G. and Lam, C.F. (1999) A systematic study of low-resolution recognition in protein--protein complexes, *Proc Natl Acad Sci U S A*, **96**, 8477-8482.

Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M. and Frank, J. (2003) Locking and unlocking of ribosomal motions, *Cell*, **114**, 123-134.

Volkmann, N. and Hanein, D. (1999) Quantitative fitting of atomic models into observed densities derived by electron microscopy, *J Struct Biol*, **125**, 176-184.

Volkmann, N. and Hanein, D. (2003) Docking of atomic models into reconstructions from electron microscopy, *Methods Enzymol*, **374**, 204-225.

Volkmann, N., Hanein, D., Ouyang, G., Trybus, K.M., DeRosier, D.J. and Lowey, S. (2000) Evidence for cleft closure in actomyosin upon ADP release, *Nat Struct Biol*, **7**, 1147-1155.

Wriggers, W. and Chacon, P. (2001) Modeling tricks and fitting techniques for multiresolution structures, *Structure (Camb)*, **9**, 779-788.

Wriggers, W., Milligan, R.A. and McCammon, J.A. (1999) Situs: A package for docking crystal structures into low-resolution maps from electron microscopy, *J Struct Biol*, **125**, 185-195.

Wu, X., Milne, J.L., Borgnia, M.J., Rostapshov, A.V., Subramaniam, S. and Brooks, B.R. (2003) A core-weighted fitting method for docking atomic structures into low-resolution maps: application to cryo-electron microscopy, *J Struct Biol*, **141**, 63-76.

Wyngaerd, J. and Van Gool, L. (2002) Automatic crude patch registration: Toward automatic 3D model building, **87**, 8-26.

Zhu, Y.M. (2002) Volume image registration by cross-entropy optimization, *IEEE Trans Med Imaging*, **21**, 174-180.