Article No. mb982232 *J. Mol. Biol.* (1998) **284**, 1247–1254

JMB

COMMUNICATION

Self-organizing Neural Networks Bridge the Biomolecular Resolution Gap

Willy Wriggers[1,2]*, Ronald A. Milligan[2], Klaus Schulten[3] and J. Andrew McCammon[1]

[1]*Department of Chemistry and Biochemistry, Department of Pharmacology, University of California, San Diego 9500 Gilman Drive, La Jolla CA 92093-0365, USA*

[2]*Department of Cell Biology The Scripps Research Institute 10666 N. Torrey Pines Road La Jolla, CA 92037, USA*

[3]*Department of Physics and Beckman Institute, University of Illinois at Urbana-Champaign, 405 N. Mathews Ave, Urbana, IL 61801, USA*

**Corresponding author*

Topology-representing neural networks are employed to generate pseudo-atomic structures of large-scale protein assemblies by combining high-resolution data with volumetric data at lower resolution. As an application example, actin monomers and structural subdomains are located in a three-dimensional (3D) image reconstruction from electron micrographs. To test the reliability of the method, the resolution of the atomic model of an actin polymer is lowered to a level typically encountered in electron microscopic reconstructions. The atomic model is restored with a precision nine times the nominal resolution of the corresponding low-resolution density. The presented self-organizing computing method may be used as an information-processing tool for the synthesis of structural data from a variety of biophysical sources.

Keywords: actin; topology representing networks; electron microscopy; multi-resolution; macromolecular assemblies

Advances in modern biology and medicine depend on an understanding of fundamental cellular processes, most of which involve the actions and interactions of large biomolecular aggregates. Three-dimensional (3D) image reconstructions of aggregates, involving hundreds of thousands to millions of atoms, are now routinely determined (DeRosier & Harrison, 1997) by electron microscopy (EM). To bridge the resolution gap from individual atoms to the basic functional units in biological cells, attempts have been made to combine EM data with high-resolution structures determined by NMR spectroscopy or X-ray crystallography (Baker & Johnson, 1996). The most common of these hybrid strategies is the visual docking of crystallographic structures into envelopes derived from low-resolution data to obtain models of aggregates at pseudo-atomic resolution. The successful construction of such hybrid models for viruses and cytoskeletal motor-filament complexes provides a clear indication of the value of this approach (Chiu *et al.*, 1997; Schröder *et al.*, 1993; Milligan, 1996). The reliability of cryo EM data provided, in some cases, a basis for docking atomic structures into assemblies with a precision of four to five times the resolution of the experimental data (Baker & Johnson, 1996).

Despite the success of the visual docking method, there are serious limitations that call for the development of more advanced techniques. Apart from the qualitative character of docking by eye, crystal structures and EM envelopes are, in general, unreliable determinants of the architecture of macromolecular assemblies. Often, proteins change conformations by ''induced fit'' in the aggregate (Rayment *et al.*, 1993a; Mendelson & Morris, 1997) or exhibit spatial variability not present in the crystal (Sosa *et al.*, 1997). Also, visual docking requires the choice of a molecular boundary in the reconstruction volume. Usually, a cutoff density value is chosen for the surface representation, although it is clear that EM images of molecules lack hard boundaries (Frank, 1996). A faithful representation of the molecular surface may even become impossible if static disorder in averaged EM data alters the shape of the surface contour significantly.

Abbreviations used: 3D, three-dimensional; EM, electron microscopy; TRN, topology-representing network; rms, root-mean-square.

E-mail address of the corresponding author: wriggers@ucsd.edu

0022–2836/98/501247–08 $30.00/0

We describe a self-organizing algorithm to represent both crystal structures and low-resolution volumetric data by a small number of neural pointers. The quantization of the volumetric data is independent of the protein surface and will be particularly useful in practical applications where spatial disorder renders the low-resolution data unsuitable for a definition of molecular boundaries. This is the first time neural computing has been applied to characterize the shape and density distribution of the occupied volume of large-scale protein assemblies.

Neural network quantization of 3D biological data

Self-organizing feature maps organize the connectivity of neurons in a cortical layer to optimize the spatial distribution of the neural responses to a training sequence of input signals (Willshaw & von der Malsburg, 1976; Kohonen, 1982; Ritter *et al*., 1992). The purpose of the optimization is to convert the similarity of signals into proximity relationships among the neurons in the cortical layer. Self-organizing neural networks are typically applied to solve important information-processing tasks in artificial intelligence, including the formation of topographic sensory maps and visuomotor control of robots (Ritter *et al*., 1992; Zeller *et al*., 1997). A consequence of the map formation is a quantization that approximates the probability density function $p(\mathbf{v})$ of input signals $\mathbf{v} \in \Re^D$, using a finite number of neural pointers $\mathbf{w}_i \in \Re^D$, $i = 1, \ldots, N$. Here, we train neural networks on 3D biological data sets, using the topology-representing network (TRN) algorithm developed by Martinetz & Schulten (1994)

We assume that each neural unit $\mathbf{w}_i$ receives input stimuli $\mathbf{v}(t)$, $t = 1, \ldots, t_{\max}$, on a manifold $M \subset \Re^D$. The sequence of stimuli will be randomly selected on the manifold according to $p(\mathbf{v})$. In practical applications to biomolecular datasets, M is the discrete set of atom positions or density voxels, and $p(\mathbf{v})$ is proportional to the atom masses or to the density values, respectively. The number $t_{\max}$ of stimuli necessary for a reliable representation of a discrete manifold M is of the order of its number of data points. Furthermore, we assume that the set of neural units $\mathbf{w}_i$ can develop lateral connections (synaptic links) with each other. The lateral connections are described by a connection matrix $\mathbf{C}$ with elements $C_{ij} \in \{0, 1\}$. Initially, the C_{ij} are set to zero and the weight vectors $\mathbf{w}_i$ are randomly distributed near M.

The adaptation of the neural pointers $\mathbf{w}_i$ at a given time step t is given by:

$$\mathbf{w}_i(t) = \mathbf{w}_i(t-1) + \varepsilon(t) e^{-k_i(t)/\lambda(t)} [\mathbf{v}(t) - \mathbf{w}_i(t-1)]$$
$$i = 1, \cdots, N; t = 2, \cdots, t_{\max} \quad (1)$$

The parameter $\lambda(t)$ and the step size $\varepsilon(t)$ are calculated dynamically (see equation (2), opposite). Information about the arrangement of the receptive fields is given by the closeness rank k_i (t) of each neuron i depending on $\mathbf{v}(t)$, i.e. the number of neurons $\mathbf{w}_j$ with $\| \mathbf{v}(t) - \mathbf{w}_j(t) \| < \| \mathbf{v}(t) - \mathbf{w}_i(t) \|$ ($\| \cdot \|$ is the Euclidean distance). The ranking ensures that the adaptation of the $\mathbf{w}_i(t)$ is controlled by the topological arrangement of neighboring neurons, avoiding confinement of neurons to local minima of the quantization error (Martinetz *et al*., 1993). Simultaneously, at a given time step t, neural units can develop or refresh a connection with another unit by setting the corresponding matrix elements of $\mathbf{C}(t)$ from zero to 1. Unrefreshed connections are removed after they exceed a maximum age T (t). The time-dependent parameters λ, ε and T, are computed according to:

$$\lambda = \lambda_i (\lambda_f/\lambda_i)^{t/t_{\max}}$$
$$\varepsilon = e_i (\varepsilon_f/\varepsilon_i)^{t/t_{\max}} \quad (2)$$
$$T = T_i (T_f/T_i)^{t/t_{\max}}$$

The particular choice of λ_i, λ_f, ε_i, ε_f, T_i and T_f is not very critical. The parameters given in the Figure legends were optimized by trial and error.

Figure 1 presents TRN tessellations of the two-dimensional surface of a torus and of the occupied volume of the crystal structure of the myosin motor domain. The TRN algorithm distributes the weight vectors over the manifold M according to the probability density of the stimuli (Figure 1). In general, the neural pointers $\mathbf{w}_i$ and the corresponding connectivities C_{ij} define a perfect topology-preserving map and form a discrete representation of the manifold M, even in cases where M has an intricate topology (Martinetz *et al*., 1993). Figure 1(b) illustrates the fact that a manifold with periodic boundaries, such as the two-dimensional surface of the torus, can be embedded in a space of higher dimension (here $D = 3$) to represent its topology.

The described quantization technique can be applied to both high and low-resolution datasets. For polymeric structures, N should be a multiple of the number of identical subunits in the dataset. Let K neural pointers $\mathbf{w}_i^{\mathrm{h}}$ and $\mathbf{w}_j^{\mathrm{l}}$ ($i, j = 1, \ldots, K$) discretize the occupied volume of a given subunit of the aggregate at high (h) and low (l) resolution. As described below, the K neurons may be derived from more than one network to identify features at variable resolution. The neurons divide both high-resolution and low-resolution data sets into a number of subregions:

$$V_i^{\mathrm{l,h}} = \{\mathbf{q} \in \Re^D | \, ||\mathbf{q} - \mathbf{w}_i^{\mathrm{l,h}}|| \leqslant ||\mathbf{q} - \mathbf{w}_j^{\mathrm{l,h}}|| \; j = 1, \cdots, K\}$$
$$i = 1, \cdots, K \quad (3)$$

known in the literature as receptive fields, Voronoi cells, or Dirichlet regions.

We now define the index map:

$$\Im : m \to n \; (m, n = 1, \cdots, K) \quad (4)$$

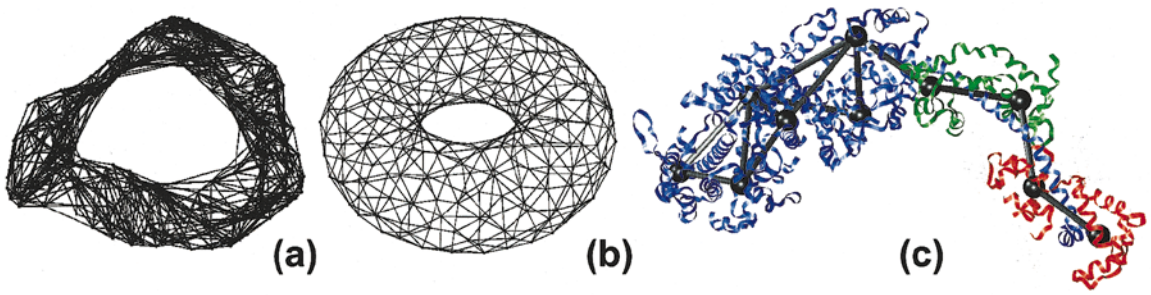

Figure 1. Topology-preserving tessellation with the TRN: surface of a torus ($N = 250$ neurons) after (a) 10,000 and (b) 100,000 time-steps, and (c) volumetric tessellation ($N = 10$ neurons, black) of the crystal structure (Rayment *et al*., 1993b) of myosin subfragment S1 after 200,000 time-steps (the myosin α-helical tail is decorated by the essential light chain, green, and the regulatory light chain, red). Lines represent lateral connections $C_{ij} = 1$. Parameters (equations (1) and (2)): $\varepsilon_i = 0.3$; $\varepsilon_f = 0.02$; $T_i = 0.1\ N$; $T_f = 0.8\ N$; $\lambda_i = 0.2\ N$; $\lambda_f = 0.02$. The computations required five minutes computing time on a Silicon Graphics Indigo 2 workstation. Atomic models of biopolymers (Figures 1–3) were rendered with the molecular graphics program VMD (Humphrey *et al*., 1996). Other software, including the TRN code, was developed by W.W. (unpublished results).

which identifies pairs of corresponding pointers in both high-resolution and low-resolution data. The resulting topographic map:

$$f : V_i^{\mathrm{h}} \to V_{\Im(i)}^{\mathrm{l}} \quad (i = 1, \cdots, K) \tag{5}$$

then correlates regions of the high-resolution structure with regions of the corresponding low-resolution density. To obtain a pseudo-atomic resolution model, the pointers $\mathbf{w}_i^{\mathrm{h}}$ and $\mathbf{w}_{\Im(i)}^{\mathrm{l}}$ are superimposed by least-squares fitting (Kabsch, 1976, 1978).

We note that in practical applications the index map $\Im$ equation (4) is not known *a priori*. The distance matrices:

$$D_{ij}^{\mathrm{h,l}} = ||\mathbf{w}_i^{\mathrm{h,l}} - \mathbf{w}_j^{\mathrm{h,l}}|| \quad (i, j = 1, \cdots, K) \tag{6}$$

determine the geometric arrangement of the K neurons in a given subunit. $\Im$ can be optimized with respect to the distance matrix mean-square deviation of the high-resolution and low-resolution pointers:

$$\Delta_{\Im} = \frac{2}{K^2} \sum_{i>j} (D_{ij}^{\mathrm{h}} - D_{\Im(i)\Im(j)}^{\mathrm{l}})^2 \tag{7}$$

The minimum of $\Delta_{\Im}$ can be found by direct enumeration of the $K!$ possible index maps.

Reconstruction of blurred data

We applied the method to the experimentally well-studied (Holmes *et al*., 1990; Milligan *et al*., 1990; Orlova & Egelman, 1993; Bremer *et al*., 1994) actin filament (Figure 2). Actin comprises four structural subdomains (Kabsch *et al*., 1990), identifiable by quantization with four neurons (Figure 2(a)). A single neuron from a separate computation identifies the centroid of the receptive field correponding to a single actin monomer. The atomic structure of the filament is not known, but crystal structures of actin subunits have been refined against X-ray fiber diffraction data (Holmes *et al*., 1990; Lorenz *et al*., 1993; Tirion *et al*., 1995). The Holmes *et al*. (1990) model of the actin filament is based on 13 monomers, whose cumulative helical rotation yields a 180° twist of the 13-mer (cf. Figure 2(b)). To lower the resolution, the atomic model was convoluted with a Gaussian filter with standard deviation $\sigma = 10$ Å. A discrete volumetric density map was computed on a lattice with voxel spacing 3 Å by a weighted sum of the Gaussian contributions over all atoms (Figure 2(c)). The value $\sigma = 10$ Å was chosen to imitate the appearance of experimental (Orlova & Egelman, 1993; Bremer *et al*., 1994) EM images of actin. We note that there is only an approximate relation between σ and the nominal resolution of EM image reconstructions, as the latter value is estimated from the limiting layer-lines of experimental diffraction patterns (Hawkes & Valdrè, 1990).

Using periodic boundaries (Figure 2(c)) in the z-direction, 13 neurons and 52 neurons were distributed in two separate computations to identify the centroid and the adjacent subdomains of each actin monomer. The two resulting networks are superimposed with the filament density in Figure 2(d). The five neurons defining the position and orientation of a monomer (Figure 2(a)) were least-squares fit (Kabsch, 1976, 1978) to the corresponding neurons in the 13-mer with a rms deviation of 3.4 Å. Monomer orientations in the reconstructed filament, shown in Figure 2(e), were correctly reproduced according to the distance matrix criterion (equation (7)). The reconstructed filament (Figure 2(e)) exhibits an (atom-) rms deviation of only 1.1 Å from the original model (Figure 2(b)).

Application to electron microscopy

The identification of features in volumetric data from EM is complicated by the fact that the

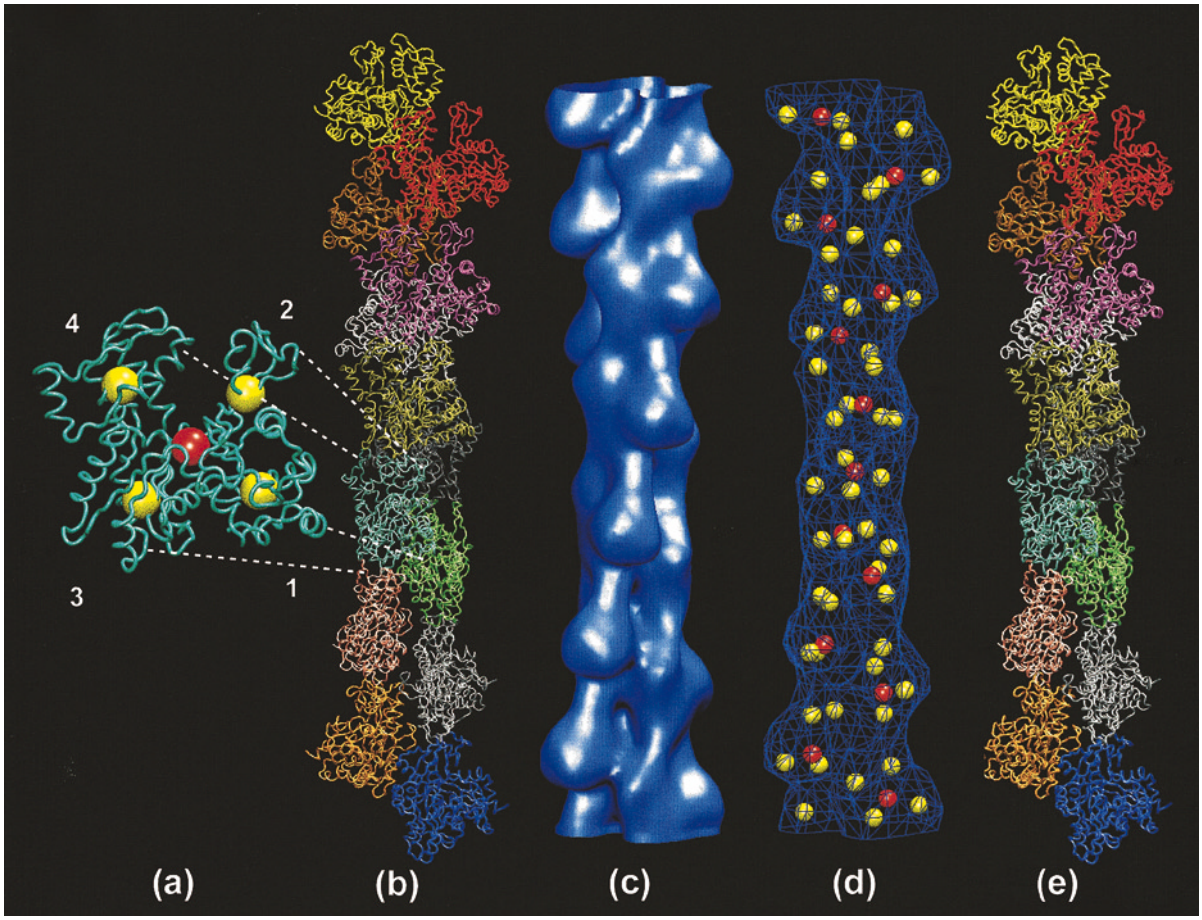

Figure 2. Reconstruction of an actin filament from low-resolution data. (a) Quantization of the actin monomer structure. Actin was modeled (Wriggers & Schulten, 1998) based on the actin-gelsolin segment-1 complex (McLaughlin *et al.*, 1993). One neuron (red) was placed to identify the centroid of the atomic structure. Four neurons (yellow), corresponding to the actin structural subdomains 1-4 (Kabsch *et al.*, 1990), were placed in a second computation. The neuron positions are averages from ten statistically independent computations. Parameters (equations (1) and (2)): $t_{\max} = 1{,}000{,}000$; $\varepsilon_i = 0.03$; $\varepsilon_f = 0.001$; $\lambda_i = 0.2\ N$; $\lambda_f = 0.02$. (b) Model of the actin filament (Holmes *et al.*, 1990) built from 13 subunits (rise per monomer 27 Å; rotation per monomer −116.154°). (c) Isosurface of the actin filament at low resolution rendered at half-maximal contour level. The contribution of a single atom to the density value of a voxel on a 3D lattice is given by the Gaussian:

$$G(\Delta x, \Delta y, \Delta z) = \exp\left(\frac{-3(\Delta x)^2}{2\sigma^2}\right)\exp\left(\frac{-3(\Delta y)^2}{2\sigma^2}\right)\exp\left(\frac{-3(\Delta z)^2}{2\sigma^2}\right) \tag{8}$$

where $\sigma = 10$ Å is the 3D standard deviation of G, and $(\Delta x, \Delta y, \Delta z)$ is the difference vector between an atom and a voxel on the lattice. Periodic boundaries were realized by embedding the 3D unit cell (comprising a 180° turn of the filament) in $\Re^4$ according to the transformation:

$$\begin{aligned} u_1 &= x \\ u_2 &= y \\ u_3 &= (R/2\pi)\cos(2\pi z/R) \\ u_4 &= (R/2\pi)\sin(2\pi z/R) \end{aligned} \tag{9}$$

where $R = 351$ Å is the unit cell repeat (comprising 13 monomers) in the z-direction, $(x, y, z) \in \Re^3$ is the position of an atom or low-resolution voxel, and $\mathbf{u} = (u_1, u_2, u_3, u_4) \in \Re^4$ is the corresponding position in the periodic space defined by $\sqrt{(u_3^2 + u_4^2)} = R/2\pi$. (d) Quantization of the actin filament density (wire mesh). Two neural networks are superimposed: 13 neurons (red) identify the actin monomers, and 52 neurons (yellow) identify individual actin subdomains. The shown positions were averaged over the 13 subunits by helical projection. Parameters (equations (1) and (2)): $t_{\max} = 10{,}000{,}000$; $\varepsilon_i = 0.03$; $\varepsilon_f = 0.001$; $\lambda_i = 0.2\ N$; $\lambda_f = 0.02$. (e) Atomic model of the filament, reconstructed with VMD (Humphrey *et al.*, 1996) by least-squares fitting of corresponding neurons from high-resolution and low-resolution data. The isosurfaces were generated with an improved version (Heiden *et al.*, 1993) of the ''marching cubes'' algorithm (Lorensen & Cline, 1987), adapted from the program VIEWMOL (Bleiber & Hill, 1993). The computations (Figures 2 and 3) required six hours computing time on a Silicon Graphics Indigo 2 workstation.

experimental density does not strictly correspond to the object density. The voxel histogram (Frank *et al.*, 1991; Frank, 1996) of an image reconstruction of actin from micrographs (Orlova & Egelman, 1993) in Figure 3(a) illustrates two general properties of experimental density distributions. First, a pronounced peak at low densities (here shifted to the origin) is due to background scattering. The protein density corresponds to a second, broader peak at higher densities. When integrating the histogram ''from the top down'', the known molecular volume of actin corresponds to a (dimensionless) boundary density value of 30 (E. H. Egelman, *personal communication*). Figure 3(b) shows the experimental surface (Orlova & Egelman, 1993), rendered at the slightly smaller threshold value of 24. At this density, the isosurface resembles the surface of the control data in Figure 2(c). For the quantization with the TRN (Figure 3(c)), experimental densities below a value of ten were assumed to be caused by background scattering (cf. Figure 3(a)) and were set to zero. Stimuli $\mathbf{v}$ were then randomly selected on the 3D lattice according to the density distribution above the cutoff. The atomic structure of each actin monomer was docked into the low-resolution data by least-squares superposition (Kabsch, 1976, 1978) of the five neural pointers (Figure 2(a)) with a rms deviation of 3.9 Å. The correct orientation of the monomer was determined from the minimum of $\Delta_{\Im}$ (equation (7)). Figure 3(d) shows the resulting atomic resolution model. The hybrid model of the polymer obtained by combining crystal and EM data closely resembles the Holmes (1990) model of actin. The (atom-) rms deviation of the hybrid model from the Holmes model (Figure 2(b)) is 2.7 Å.

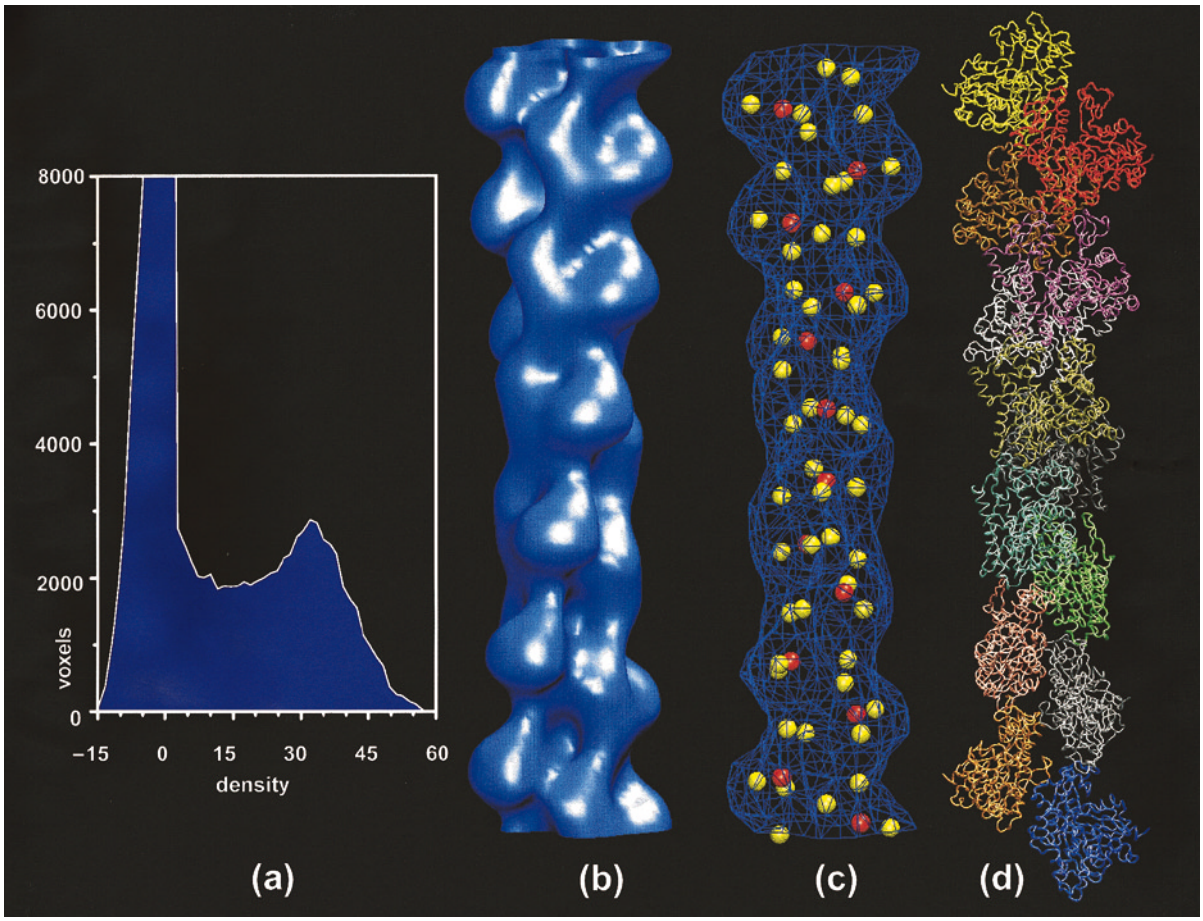

Figure 3. Construction of the actin filament structure from experimental low-resolution data. (a) Histogram of the density distribution in the image reconstruction from negatively stained specimens (Orlova & Egelman, 1993) ($51 \times 51 \times 117$ voxels; 50 histogram bins in the density interval $[-15, 57]$). The sharp peak at lower densities is due to the background scattering density, the broader peak due to the protein. (b) Isosurface of the actin density map (Orlova & Egelman, 1993) with periodic boundaries (Figure 2), rendered at a threshold value of 24 (helical parameters same as in Figure 2(b)). (c) Quantization of the actin filament density (wire mesh) with periodic boundaries (Figure 2). Two neural networks are superimposed: 13 neurons (red) identify the actin monomers and 52 neurons (yellow) identify individual actin subdomains. The neuron positions were averaged over the 13 subunits by helical projection. Parameters (equations (1) and (2)): $t_{\max} = 10{,}000{,}000$; $\varepsilon_i = 0.03$; $\varepsilon_f = 0.001$; $\lambda_i = 0.2\ N$; $\lambda_f = 0.02$. (d) Atomic model of the filament, constructed with VMD (Humphrey *et al.*, 1996) by least-squares superposition of corresponding neurons from high-resolution and low-resolution data.

Table 1. Spatial variabilities (rms fluctuations) of neuron positions

Structure	Figure	Resolution	COD	rms fluctuation (Å) SD1	SD2	SD3	SD4
Monomer	2(a)	Atomic	0.53	0.35	0.36	0.30	0.36
13-mer	2(c)	10 Å	0.48	0.40	0.58	0.54	0.33
13-mer	3(b)	20 Å[a]	0.91	3.70	1.40	2.89	1.50

COD denotes the centroid neuron; SD1, SD2, SD3 and SD4 denote neurons corresponding to the four subdomains of actin (Figure 2(a)). The monomer data were obtained from ten statistically independent calculations. The variabilities of neuron positions in the simulated (Figure 2(c)) and in the experimental (Figure 3(b)) low-resolution filament data sets were computed by helical projection.

[a] Resolution estimated based on limiting layer-lines in reciprocal space (E. H. Egelman, personal communication).

Table 1 summarizes the spatial variability of the neuron positions in the monomer and filament computations. The neuron positions in the ''simulated'' volumetric density are well defined and their precision (<1 Å) is comparable to that of neurons placed into the high-resolution structure. The low rms deviations from helical symmetry indicate that the placed neurons are not in register with the lattice, i.e. the positions are not dependent on the discretization of the low-resolution data. Neuron positions in the experimental volumetric data are less precise. Neurons corresponding to actin subdomains 1 and 3 exhibit deviations from the helical symmetry by 3-4 Å. Hence, the neuron positions were averaged over the 13 subunits by helical projection in the construction of the hybrid model (Figure 3(d)).

Discussion

The TRN identifies features in low-resolution images and relates these features with available atomic structures. So far, researchers were able to achieve this synthesis of data at various resolutions intuitively by inspection, but depended on the accuracy of the visual representation of the data (Frank, 1996). The presented computational strategy bypasses the need to visualize the unreliable surface contour of biomolecules, and accurately and reproducibly characterizes features within the volumetric data set.

It is assumed that the electron density of 3D image reconstructions from EM is proportional to the object density and, thereby, can be related to the distribution of atoms in a high-resolution structure. This relationship can be achieved by correction of the contrast transfer function (Hawkes & aldrè, 1990) in cryo EM, but is not strictly valid for image reconstructions from negatively stained specimens (Frank, 1996). The constructed pseudo-atomic model of the actin filament (Figure 3(d)) suggests that images from negatively stained micrographs are suitable in practical applications of the self-organizing method.

The hybrid model derived from the experimental image reconstruction (Figure 3(d)) corroborates the Holmes (1990) model of the actin filament. This well-known model of filamentous actin was refined against X-ray fiber diffraction data at 8 Å resolution. The 2.7 Å rms deviation of the hybrid model from the Holmes model is within the experimental uncertainty. The atomic model of actin filament reconstructed from ''simulated'' low-resolution data suggests that the self-organizing method restores the atomic structure with a precision nine times the nominal resolution of the volumetric data.

In applications to macromolecular assemblies comprising identical subunits (such as actin filaments), the centroid of individual subunits is identifiable by a single neuron per subunit, and the subunit orientation is identifiable by a larger number (here four) neurons per subunit. The ''optimum'' number of neurons is determined by the shape of the protein and may vary among applications. As a rule of thumb, there should be at least three neurons per subunit to determine the orientation, but no more than ten, to avoid combinatorial complexity (equation (7)). Alternatively, an optimum representation may be defined as the number of neurons that minimize a normalized quantization error (Martinetz *et al*., 1993) of a given data set (W.W., unpublished results).

In realistic situations, one can expect certain limitations that require extensions to the method described here. First, the resolution of the EM data may be too low for an identification of features that would unambiguously determine the orientation of a protein within the EM data set (McGouch *et al*., 1997). In such situations, a small number N of neurons (e.g. $N = 2$ for an ellipsoidal density distribution) may be useful to reduce the degrees of freedom to be explored in an exhaustive conformational search of possible docking orientations (Strynadka *et al*., 1996). Second, additional neurons may be necessary to account for small regions of the low-resolution density that are not completely represented by the crystal structure of the subunits. In this case, the size of the receptive fields of these neurons can be adjusted (Ritter *et al*., 1992) by assigning individual enlargement factors $\varepsilon = \varepsilon_i$ (equation (1)). Third, subunits may change their conformation in the aggregate by ''induced fit''. To determine these conformational changes at atomic resolution, the topographic map f (equation (5)) might be useful as a constraint in steered molecular dynamics simulations (Collins *et al*., 1995; Grubmüller *et al*., 1996; Izrailev *et al*., 1997, 1998).

In conclusion, research in artificial intelligence has produced powerful computing techniques that

imitate the information processing employed by biological neural systems (Amit, 1989; Hertz *et al*., 1991; Ritter *et al*., 1992). Neural network algorithms are of increasing importance in biological applications, where they help to integrate a multitude of experimental data (Wade *et al*., 1992; Andrade *et al*., 1993; van Osdol *et al*., 1994; Schuchhardt *et al*., 1996; Fernandez & Carazo, 1996; Schneider *et al*., 1998). The presented topology-representing method for the quantization of macromolecular aggregates will be valuable in many areas of structural biology where datasets are studied at varied resolution.

Acknowledgments

Edward H. Egelman provided the 3D image reconstruction of the actin filament. We thank Michael Zeller and Amy McGough for discussions. This work was supported, in part, by grants from NIH and the NRAC program of the NSF Supercomputer Centers (J.A.M.). W.W. acknowledges the LJIS Interdisciplinary Training Program and The Burroughs Wellcome Fund for fellowship support.

References

Amit, D. (1989). *Modeling Brain Function*, Cambridge University Press, Cambridge.

Andrade, M., Chacon, P., Merelo, J. & Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.* **6**, 383–390.

Baker, T. & Johnson, J. (1996). Low resolution meets high: towards a resolution continuum from cells to atoms. *Curr. Opinion Struct. Biol.* **6**, 585–594.

Bleiber, A. & Hill, J. R. (1993). VIEWMOL, version 1.4 Arbeitsgruppe Quantenchemie an der Humboldt-Universität zu Berlin. Max-Planck-Gesellschaft.

Bremer, A., Henn, C., Goldie, K., Engel, A., Smith, P. & Aebi, U. (1994). Towards atomic interpretation of F-actin filament three-dimensional reconstructions. *J. Mol. Biol.* **242**, 683–700.

Chiu, W., Burnett, R. & Garcea, R. (1997). *Structural Biology of Viruses*, Oxford University Press, New York.

Collins, J., Burt, S. & Erickson, J. (1995). Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics. *Nature Struct. Biol.* **2**, 334–338.

DeRosier, D. & Harrison, S. (1997). Macromolecular assemblages: sizing things up. *Curr. Opin. Struct. Biol.* **7**, 237–238.

Fernandez, J. & Carazo, J. (1996). Analysis of structural variability within two-dimensional biological crystals by a combination of patch averaging techniques and self organizing maps. *Ultramicroscopy*, **65**, 81–93.

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

Frank, J., Penczek, P., Grassucci, R. & Srivastava, S. (1991). Three-dimensional reconstructuion of the 70 S *E. coli* ribosome in ice: the distribution of ribosomal RNA. *J. Cell Biol.* **115**, 597–605.

Grubmüller, H., Heymann, B. & Tavan, P. (1996). Ligand binding and molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, **271**, 997–999.

Hawkes, P. & Valdrè, U. (1990). *Biophysical Electron Microscopy*, Academic Press, London.

Heiden, W., Goetze, T. & Brickmann, J. (1993). Fast generation of molecular surfaces from 3D data fields with an enhanced ''marching cube'' algorithm. *J. Comp. Chem.* **14**, 246–250.

Hertz, J., Krogh, A. & Palmer, R. (1991). *Introduction to the Theory of Neural Computation*, Addison-Wesley, New York.

Holmes, K., Popp, D., Gebhard, W. & Kabsch, W. (1990). Atomic model of the actin filament. *Nature*, **347**, 44–49.

Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD–visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.

Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. (1997). Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72**, 1568–1581.

Izrailev, S., Stepaniants, S., Isralewitz, B., Kosztin, D., Lu, H., Molnar, F., Wriggers, W. & Schulten, K. (1998). Steered molecular dynamics. In *Algorithms for Macromolecular Modelling, Lecture Notes in Computational Science and Engineering*, Springer Verlag.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallog. sect. A*, **32**, 922–923.

Kabsch, W. (1978). A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallog. sect. A*, **34**, 827–828.

Kabsch, W., Mannherz, H., Suck, D., Pai, E. & Holmes, K. (1990). Atomic structure of the actin:DNase I complex. *Nature*, **347**, 37–44.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **44**, 59–69.

Lorensen, W. & Cline, H. (1987). Marching cubes: a high resolution 3D surface construction algorithm. *Comput. Graph.* **21**, 163–169.

Lorenz, M., Popp, D. & Holmes, K. (1993). Refinement of the F-actin model against X-ray fiber diffraction data by the use of a directed mutation algorithm. *J. Mol. Biol.* **234**, 826–836.

Martinetz, T. & Schulten, K. (1994). Topology representing networks. *Neural Networks*, **7**, 507–522.

Martinetz, T. M., Berkovich, S. G. & Schulten, K. (1993). ''Neural gas'' for vector quantization and its application to time-series prediction. *IEEE Trans. Neural Networks*, **4**, 558–569.

McGough, A., Pope, B., Chiu, W. & Weeds, A. (1997). Cofilin changes the twist of F-actin: implications for actin filament dynamics and cellular function. *J. Cell Biol.* **138**, 771–781.

McLaughlin, P., Gooch, J., Mannherz, H.-G. & Weeds, A. (1993). Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. *Nature*, **364**, 685–692.

Mendelson, R. & Morris, E. (1997). The structure of actomyosin subfragment 1 complex: results of searches using data from electron microscopy and X-ray crystallography. *Proc. Natl. Acad. Sci. USA*, **94**, 8533–8538.

Milligan, R. (1996). Protein-protein interactions in the rigor actomyosin complex. *Proc. Natl. Acad. Sci. USA*, **93**, 21–26.

Milligan, R., Whittaker, M. & Safer, D. (1990). Molecular structure of F-actin and location of surface binding sites. *Nature*, **348**, 217–221.

Orlova, A. & Egelman, E. (1993). A conformational change in the actin subunit can change the flexibility of the actin filament. *J. Mol. Biol.* **232**, 334–341.

Rayment, I., Holden, H., Whittaker, M., Yohn, C., Lorenz, M., Holmes, K. & Milligan, R. (1993a). Structure of the actin-myosin complex and its implications for muscle contraction. *Science*, **261**, 58–65.

Rayment, I., Rypniewski, W., Schmidt-Bäse, K., Smith, R., Tomchick, D., Benning, M., Winkelmann, D., Wesenberg, G. & Holden, H. (1993b). Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science*, **261**, 50–58.

Ritter, H., Martinetz, T. & Schulten, K. (1992). *Neural Computation and Self-organizing Maps: An Introduction*, Addison-Wesley, New York.

Schneider, G., Sjoling, S., Wallin, E., Wrede, P., Glaser, E. & von Heijne, G. (1998). Feature-extraction from endopeptidase cleavage sites in mitochondrial targeting peptides. *Proteins: Struct. Funct. Genet.* **30**, 49–60.

Schröder, R., Manstein, D., Jahn, W., Holden, H., Rayment, I., Holmes, K. & Spudich, J. (1993). Three-dimensional atomic model of F-actin decorated with *Dictyostelium* myosin S1. *Nature*, **364**, 171–174.

Schuchhardt, J., Schneider, G., Reichelt, J., Schomburg, D. & Wrede, P. (1996). Local structural motifs of protein backbones are classified by self-organizing neural networks. *Protein Eng.* **9**, 833–842.

Sosa, H., Dias, P., Hoenger, A., Whittaker, M., Wilson-Kubalek, E., Sablin, E., Fletterick, R., Vale, R. & Milligan, R. (1997). A model for the microtubule-ncd motor protein complex obtained by cryo-electron microscopy and image analysis. *Cell*, **90**, 217–224.

Strynadka, N., Eisenstein, M., Katchalski-Katzir, E., Shoichet, B., Kuntz, I., Abagyan, R., Totrov, M., Janin, J., Cherfils, J., Zimmerman, F., Olson, A., Duncan, B., Rao, M., Jackson, R., Sternberg, M. & James, M. (1996). Molecular docking programs successfully predict the binding of a β-lactamase inhibitory protein to TEM-1 β-lactamase. *Nature Struct. Biol.* **3**, 233–239.

Tirion, M., ben-Avraham, D., Lorenz, M. & Holmes, K. (1995). Normal modes as refinement parameters for the F-actin model. *Biophys. J.* **68**, 5–12.

van Osdol, W., Myers, T., Paull, K., Kohn, K. & Weinstein, J. (1994). Use of the Kohonen self-organizing map to study the mechanisms of action of chemotherapeutic agents. *J. Natl Cancer Inst.* **86**, 1853–1859.

Wade, R., Bohr, H. & Wolynes, P. (1992). Prediction of water binding sites on proteins by neural networks. *J. Am. Chem. Soc.* **114**, 8284–8285.

Willshaw, D. & von der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. *Proc. Roy. Soc. ser. B*, **194**, 431–445.

Wriggers, W. & Schulten, K. (1998). Investigating a back door mechanism of actin phosphate release by steered molecular dynamics. *Proteins: Struct. Funct. Genet.* In the press.

Zeller, M., Sharma, R. & Schulten, K. (1997). Motion planning of a pneumatic robot using a neural network. *IEEE Control Systems Magazine*, **17**, 89–98.

Edited by W. Baumeister

(Received 1 June 1998; received in revised form 13 August 1998; accepted 13 September 1998)