

Markov Random Field Modeling of the Spatial Distribution of Proteins on Cell Membranes

Jun Zhang^a, Stanly L. Steinberg^b, Bridget S. Wilson^c, Janet M. Oliver^c,
Lance R. Williams^{a,*}

^aDepartment of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA

^bDepartment of Mathematics and Statistics, University of New Mexico, Albuquerque,
NM 87131, USA

^cDepartment of Pathology, University of New Mexico, Albuquerque, NM 87131, USA

Received: 16 November 2006 / Accepted: 22 June 2007 / Published online: 29 September 2007
© Society for Mathematical Biology 2007

Abstract Cell membranes display a range of receptors that bind ligands and activate signaling pathways. Signaling is characterized by dramatic changes in membrane molecular topography, including the co-clustering of receptors with signaling molecules and the segregation of other signaling molecules away from receptors. Electron microscopy of immunogold-labeled membranes is a critical technique to generate topographical information at the 5–10 nm resolution needed to understand how signaling complexes assemble and function. However, due to experimental limitations, only two molecular species can usually be labeled at a time. A formidable challenge is to integrate experimental data across multiple experiments where there are from 10 to 100 different proteins and lipids of interest and only the positions of two species can be observed simultaneously. As a solution, we propose the use of Markov random field (MRF) modeling to reconstruct the distribution of multiple cell membrane constituents from pair-wise data sets. MRFs are a powerful mathematical formalism for modeling correlations between states associated with neighboring sites in spatial lattices. The presence or absence of a protein of a specific type at a point on the cell membrane is a state. Since only two protein types can be observed, i.e., those bound to particles, and the rest cannot be observed, the problem is one of deducing the conditional distribution of a MRF with unobservable (hidden) states. Here, we develop a multiscale MRF model and use mathematical programming techniques to infer the conditional distribution of a MRF for proteins of three types from observations showing the spatial relationships between only two types. Application to synthesized data shows that the spatial distributions of three proteins can be reliably estimated. Application to experimental data provides the first maps of the spatial relationship between groups of three different signaling molecules. The work is an important step toward a more complete understanding of membrane spatial organization and dynamics during signaling.

Keywords Markov random field · Spatial distribution · Cell membrane · Gold label · Parameter estimation

*Corresponding author.

E-mail address: williams@cs.unm.edu (Lance R. Williams).

1. Introduction

Cell membranes display a range of receptors that bind signaling molecules and initiate transmembrane responses. Receptors and the signaling proteins and lipids they activate are distributed nonrandomly in membranes; in addition, receptor activation is accompanied by dramatic reorganization of membrane components as well as by recruitment of new signaling proteins from the cytoplasm to the membrane (Wilson et al. 2000, 2001). Strict regulation of signal transduction from the outer cell surface to the cytoplasm and nucleus is crucial for cell survival, differentiation, proliferation and other activities. Unregulated signaling is an important component in the pathogenesis of many diseases, including cancer. Nevertheless, many aspects of how the cell maintains spatio-temporal control of signaling pathways remain unclear. Correlating the activities of receptors and signaling proteins and lipids with their spatial distribution and dynamics is essential to better understand the regulation of cell signaling.

To observe the topographical events associated with cell signaling, several groups have generated high resolution spatial maps of colloidal gold particles marking receptors and signaling proteins and lipids in native membranes (Wilson et al. 2000, 2001, 2002, 2004; Prior et al., 2003; Volna et al., 2004; Kim et al., 2005; Plowman et al., 2005; Xue et al., 2007). The technique involves labeling membrane sheets stripped from the dorsal cell surface with ligand- or antibody-coated electron-dense 2–10 nm gold nanoprobe, imaging the labeled sheets by transmission electron microscopy (TEM), and extracting probe coordinates from the digitized images. The spatial distributions of the probes with respect to each other and with respect to membrane features such as clathrin-coated pits and caveolae are subsequently analyzed (Oliver et al., 2004; Hancock and Prior, 2005; Zhang et al., 2006; Nicolau et al., 2006). In general, due to limitations of applicable gold particle size, only two different protein species can be labeled with confidence in the same experiment. Experimentalists are exploring the use of new metal, semi-conductor, and ceramic electron-dense nanoprobe with different shapes to expand the number of probes than can be discriminated (Hernandez-Sanchez et al., 2006; Andrews et al., 2007). However, even with these new tools, the limited availability of antibodies raised in different species to label signal pathway components makes it difficult to substantially expand the number of probes that can be used in a single experiment. Consequently, there is a need to integrate experimental data across multiple experiments. In this paper, we use Markov random field (MRF) modeling to address this problem.

MRFs provide a powerful and robust framework for modeling correlations between states associated with neighboring sites in spatial lattices. Cell membranes can be modeled as a 2D lattice, and the presence or absence of a protein of a specific type at a point on the cell membrane is a state. Since only two protein types can be observed, i.e., those bound to particles, and the rest cannot be observed, the problem is one of deducing the conditional distribution of a MRF with unobservable (hidden) states. A question of significant importance is: What fraction of the conditional distribution of the MRF modeling spatial relationships between proteins of all types can be inferred from particle preparations showing the spatial relationships between only two types? If the conditional distribution can be reliably estimated, then the Gibbs sampler (Geman and Geman, 1984) can be used to synthesize sample MRFs allowing the complete set of protein spatial distributions

to be visualized. In the subsequent sections, we briefly review the background literature on MRF (Section 2), show how the conditional distribution of a MRF for proteins of three types can be deduced from data documenting the spatial relationships between only two types (Section 3) and describe a multiscale MRF model, with results on synthesized and experimental data (Section 4). Finally, we propose applications of this technology for reconstructing topographical distributions of membrane constituents during signal transduction.

2. Background and related work

2.1. Background on Markov random field

MRFs have several components: a lattice S with m sites s ; a neighborhood system $\mathcal{N} = \{\mathcal{N}_s | s \in S\}$, where \mathcal{N}_s is the subset of sites in S which are the neighbors of s ; a field of random variables $X = \{X_s | s \in S\}$, and a conditional probability mass function (p.m.f.), $P(X_s = x_s | X_t = x_t, t \in \mathcal{N}_s)$. Each random variable X_s takes a value in a finite set $Q = \{l_1, \dots, l_q\}$ of the possible states. $X_s = x_s$ denotes the event that X_s takes the value x_s and the notation $(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m)$ denotes a joint event. The joint event is abbreviated as $X = x$ in which $x = \{x_1, x_2, \dots, x_m\}$ is a realization of X . Therefore, there is also a joint p.m.f., $P(X = x)$. Either the conditional p.m.f. or the joint p.m.f. can be used to specify a MRF. The Markov property means that the state at a site is dependent only on those at its neighboring sites:

$$P(X_s = x_s | X_t = x_t, t \neq s, t \in S) = P(X_s = x_s | X_t = x_t, t \in \mathcal{N}_s). \tag{1}$$

For example, the neighbor set of $s = (i, j)$ for a regular lattice S is commonly defined as

$$\mathcal{N}_s = \{r = (k, l) \in S : 0 < (k - i)^2 + (l - j)^2 \leq o\}, \tag{2}$$

where o is the order of the neighborhood system. Figure 1a–c show the neighborhood systems for $o = 1, 2$ and 8 .

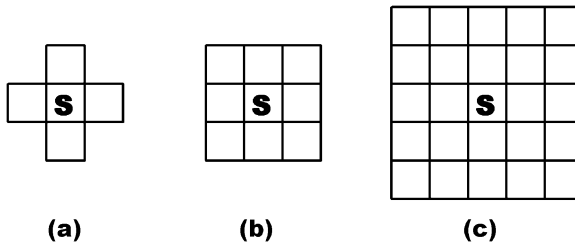


Fig. 1 MRF neighborhood systems for site $s \in S$: (a) first-order; (b) second-order; and (c) eighth-order (Li, 1995).

2.2. Related work

MRFs have been used in computer vision and image processing for texture synthesis (Cross and Jain, 1983; Efros and Leung, 1999; Paget and Longstaff, 1998), image segmentation (Derin and Elliott, 1987), and image restoration (Geman and Geman, 1984; Li, 1995). The states of MRF texture models are all possible gray levels and directly observable. A sample texture is regarded as a realization of the MRF model and is used to estimate the conditional distribution of the model through either parametric (Cross and Jain, 1983) or nonparametric methods (Efros and Leung, 1999; Paget and Longstaff, 1998). Texture can be then synthesized by sampling from the conditional distribution. In addition to MRFs where all states are observable, there are also hidden Markov models which contain states that are not directly observable. These hidden models are flexible and powerful when used in applications such as image restoration and segmentation (Geman and Geman, 1984; Besag, 1986; Derin and Elliott, 1987; Tonazzini et al., 2006), because many kinds of prior knowledge can be modeled. *Maximum a posteriori* (MAP) estimation is commonly used to restore or segment images. For the nonhierarchical MRFs discussed thus far, exact MAP estimation is computationally expensive. Consequently, approximation methods such as iterated conditional modes (ICM) (Besag, 1986), dynamic programming (Derin and Elliott, 1987), or simulated annealing with Gibbs sampling (Geman and Geman, 1984) are used. Another difficulty with nonhierarchical MRFs is that the neighborhood systems need to be small because larger neighborhood systems dramatically increase the number of parameters and the running time of the MAP algorithms.

To address the problems associated with nonhierarchical models, multiscale MRF models were formulated and have been extensively discussed in the image processing literature (Bouman and Shapiro, 1994; Kato et al. 1996, 1999; Laferté et al., 2000; Liang and Tjahjedi, 2006; Mignotte et al., 2000; Wilson and Li, 2003). In those hierarchical MRF models, there is a series of random fields at a range of scales or resolutions, and the random field at each scale depends only on the next coarser random field above it. The MAP algorithms for the hierarchical MRF models have performance comparable to or better than that of MAP estimation by simulated annealing for a nonhierarchical MRF, but require less computation than simulated annealing or ICM.

Parameter estimation is important for Bayesian image analysis because model parameters are required for MAP estimation. Maximum likelihood (ML) estimation is often used for MRF parameter estimation (Besag, 1974). When part of the data is hidden, the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) is commonly used for ML estimation. A major difficulty in applying the EM algorithm to MRFs is in the calculation of the conditional expectation, which is generally intractable because it requires summing over all possible configurations. Therefore, approximation techniques such as the mean field approximation (Celeux et al., 2003; Tonazzini et al., 2006; Zhang, 1992) and pseudo-likelihood method (Chalmond, 1989; Zhang et al., 1994) are used. The EM algorithm has been extended for parameter estimation on a quadtree (Bouman and Shapiro, 1994; Laferté et al., 2000).

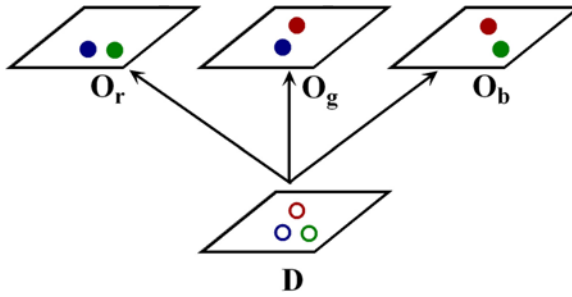


Fig. 2 Structure of a quadruply stochastic random field used in modeling of protein spatial distributions. Red, green and blue open circles represent R , G , and B , and red, green and blue closed circles represent R' , G' , and B' .

3. Modeling protein spatial distributions

In a specific signaling pathway, there are typically from 10 to 100 proteins of interest. Let us consider a simple, idealized case, where there are three proteins of interest which we call $\{R, G, B\}$, and where only two proteins can be observed in any single sample. For this scenario, we use a hidden process D with four states $\{R, G, B, X\}$ to characterize the distribution of proteins on the cell membrane. The additional state X corresponds to background. D is called a *distribution process*. In addition, there are three observable processes, O_r , O_g and O_b , to model observations where only two kinds of proteins can be observed at a time. These processes are called *observation processes* and have four observable states $\{R', G', B', X'\}$. Following Rabiner (1989), we call these observable states *observation symbols*, corresponding to the presence or absence of a gold particle bound to a protein visualized by TEM. The observation symbol probability distribution (Rabiner, 1989) is used to describe the probability that hidden state j will be observed as symbol i .

Our approach is based on the quadruply stochastic model shown in Fig. 2. This model assumes that the three observation processes depend on the same distribution process. The behavior of the observation processes given the distribution process is defined by the observation symbol probability distributions. There are three observation symbol probability distributions, one for each observation process. We assume that the state in a given observation process depends only on the corresponding state in the distribution process, and that three *observation matrices*, \mathbf{Q}_r , \mathbf{Q}_g and \mathbf{Q}_b , can be used to represent the conditional probability mass functions, $P((O_r)_s = i | D_s = j)$, $P((O_g)_s = i | D_s = j)$ and $P((O_b)_s = i | D_s = j)$:

$$\mathbf{Q}_r = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q}_g = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q}_b = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

where $(\mathbf{Q}_k)_{ij}$ is the probability that the hidden state j will be observed as the symbol i in O_k . The problem is to infer the conditional p.m.f. of D given the observation processes O_r , O_g and O_b .

In the following sections, we demonstrate that the conditional p.m.f of a hidden MRF with a first-order neighborhood system can be reliably estimated given the visible MRFs.

3.1. A Markov random field model

MRFs with a first-order neighborhood system (see Fig. 1a) are used for both the distribution and observation processes. A Markov random field can be specified by a conditional probability mass function, $P(X_s = x_s | X_t = x_t, t \in \mathcal{N}_s)$. We use \mathbf{P} to denote a matrix (we use this term even though \mathbf{P} is five dimensional, i.e., it is a *tensor*) that contains these conditional probabilities. For MRFs with a first-order neighborhood system, there are $4^5 = 1024$ elements in \mathbf{P} . The first step is to generate samples of visible Markov random fields. To generate these visible samples, we must generate a realization of the distribution process, d . Given \mathbf{P} , d is generated using the Gibbs sampler (Geman and Geman, 1984):

1. Initialize d randomly.
2. Choose $s \in S$ randomly and replace D_s with d_s drawn from $P(D_s = d_s | D_t = d_t, t \in \mathcal{N}_s)$.
3. Repeat Step 2 many times.

After the realization is generated, samples of three visible fields, o_r , o_g , and o_b , are obtained by mapping the states in d to symbols of the observation process using the observation matrices. For example, we replace R, G, B, X with X', G', B', X' respectively to generate o_r . For exposition purposes, we number the sites in the first-order neighborhood system as in Fig. 3a. A five-tuple $\langle m, l, k, j, i \rangle$ and a four-tuple $\langle l, k, j, i \rangle$ can then be used to denote the joint events $(D_4 = m, D_3 = l, D_2 = k, D_1 = j, D_0 = i)$ and $(D_3 = l, D_2 = k, D_1 = j, D_0 = i)$, respectively. The conditional probability $P(D_4 = m | D_3 = l, D_2 = k, D_1 = j, D_0 = i)$ is estimated using histogramming (Gurelli and Onural, 1994):

$$P(D_4 = m | D_3 = l, D_2 = k, D_1 = j, D_0 = i) = p_{mlkji} \approx \frac{H(\langle m, l, k, j, i \rangle)}{H(\langle l, k, j, i \rangle)}, \quad (3)$$

where H are frequencies of the tuples in the samples of D . In order to get good estimates for the conditional probabilities, the sample size need to be quite large. The realization of the distribution process, d , is generated on a lattice of size 600×740 . Corresponding to each d , there is a set of three visible samples. Our experiments showed that 100 sets of visible samples results in good estimates.

Unfortunately, the histograms for all states in the hidden field are not directly observable. However, we can deduce them from the histograms of symbols in the visible

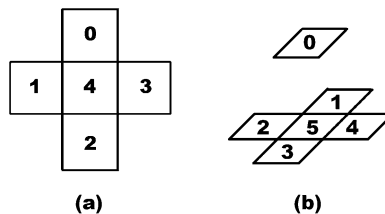


Fig. 3 Numbering of the sites in a first-order (a) and 5-neighbor (b) neighborhood systems.

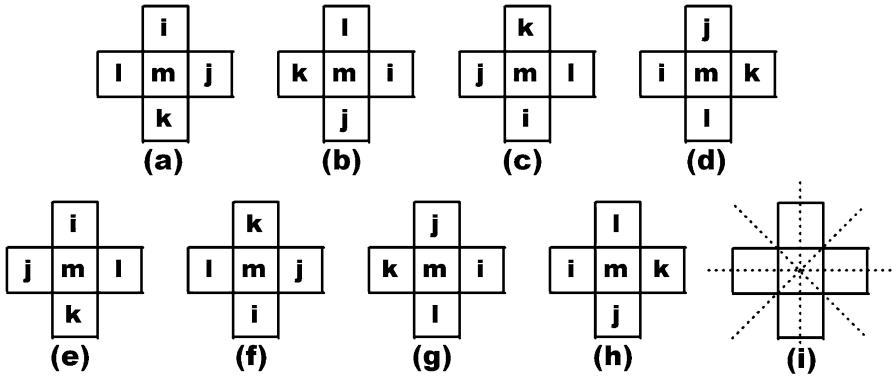


Fig. 4 Isotropy in a first-order neighborhood system. (b), (c) and (d) are rotations of (a) by multiples of 90°. (e)–(h) are mirror images of (a) with respect to four axes of reflective symmetry. All eight conditional probabilities, e.g., $P(m|i, l, k, j)$ and $P(m|l, k, j, i)$, are equal under the isotropy assumption.

samples. By performing a raster scan of the window shown in Fig. 3a over the visible samples, we obtain the frequencies of four-tuples and five-tuples of observation symbols. Because of the one-to-one relationships between G in d and G' in o_r and between B in d and B' in o_r , we observe that

$$H(\langle G, G, B, G \rangle) = H_r(\langle G', G', B', G' \rangle), \tag{4}$$

$$H(\langle B, G, G, B, G \rangle) = H_r(\langle B', G', G', B', G' \rangle), \tag{5}$$

where H_r are histograms of four-tuple's and five-tuple's of symbols in o_r . It follows that

$$p_{bggbg} \approx \frac{H_r(\langle B', G', G', B', G' \rangle)}{H_r(\langle G', G', B', G' \rangle)}. \tag{6}$$

Using analogous methods, we can directly estimate 108 of the conditional probabilities (i.e., those involving at most two of the nonbackground states) from these observed frequencies. There are still 916 unknown conditional probabilities. To reduce the number of unknown variables, we exploit isotropy in the matrix of conditional probabilities as shown in Fig. 4. The isotropy assumption is reasonable because the protein distributions are independent of orientation. The number of unknown variables is reduced to 181 under this assumption. We adopted ML estimation to infer these conditional probabilities. Although the EM algorithm is the standard method for ML estimation when part of the data is hidden, it converges very slowly. By exploiting a property of the observation matrix (i.e., the probabilities in the observation matrix are either 0 or 1), we developed a noniterative method using mathematical programming techniques to solve the ML estimation problem.

3.2. Parameter estimation for MRFs with a first-order neighborhood system

Given a realization d of the distribution process, the ML estimate maximizes the conditional probability, $P(d|\mathbf{P})$, where \mathbf{P} is the matrix of conditional probabilities. We approx-

imate the probability of d given \mathbf{P} by the pseudo-likelihood (Li, 1995) which is simply a product of the conditional probabilities:

$$P(d|\mathbf{P}) \approx \prod_{s \in S} P(d_s | d_t, t \in \mathcal{N}_s). \quad (7)$$

However, the distribution process is not directly observable. Only visible samples are available. Consequently, we need to compute the probability of a visible sample o_k given \mathbf{P} and \mathbf{Q}_k . We also approximate the conditional probability of o_k given \mathbf{P} and \mathbf{Q}_k by the pseudo-likelihood:

$$P(o_k | \mathbf{P}, \mathbf{Q}_k) \approx \prod_{s \in S} P((o_k)_s | (o_k)_t, t \in \mathcal{N}_s), \quad (8)$$

where (by Bayes' rule):

$$\begin{aligned} & P((o_k)_s | (o_k)_t, t \in \mathcal{N}_s) \\ &= \sum_{d_s, d_t, t \in \mathcal{N}_s} P((o_k)_s | d_s) P(d_s | d_t, t \in \mathcal{N}_s) P(d_t, t \in \mathcal{N}_s | (o_k)_t, t \in \mathcal{N}_s) \\ &= \sum_{d_s, d_t, t \in \mathcal{N}_s} \frac{P((o_k)_s | d_s) P(d_s | d_t, t \in \mathcal{N}_s) P((o_k)_t, t \in \mathcal{N}_s | d_t, t \in \mathcal{N}_s) P(d_t, t \in \mathcal{N}_s)}{P((o_k)_t, t \in \mathcal{N}_s)}. \end{aligned} \quad (9)$$

Because we assume that a symbol at a site in an observation process depends only on the corresponding state in the distribution process, it follows that

$$P((o_k)_t, t \in \mathcal{N}_s | d_t, t \in \mathcal{N}_s) = \prod_{t \in \mathcal{N}_s} P((o_k)_t | d_t). \quad (11)$$

Finally, we observe that

$$\begin{aligned} & P((o_k)_s | (o_k)_t, t \in \mathcal{N}_s) \\ &= \sum_{d_s, d_t, t \in \mathcal{N}_s} \frac{P(d_s | d_t, t \in \mathcal{N}_s) P(d_t, t \in \mathcal{N}_s) \prod_{t \in \{s\} \cup \mathcal{N}_s} P((o_k)_t | d_t)}{P((o_k)_t, t \in \mathcal{N}_s)} \end{aligned} \quad (12)$$

and where $k \in \{r, g, b\}$.

In Eq. (12), $P((o_k)_s | d_s)$ is known and defined in \mathbf{Q}_k . $P((o_k)_t, t \in \mathcal{N}_s)$ are the probabilities of four-tuples of symbols in the visible fields, and can be estimated from samples of the visible fields. $P(d_t, t \in \mathcal{N}_s)$ are the probabilities of four-tuples of states in the hidden field (determined by \mathbf{P}). The unknown variables are $P(d_s | d_t, t \in \mathcal{N}_s)$ and $P(d_t, t \in \mathcal{N}_s)$. The maximum likelihood estimation of \mathbf{P} is achieved by maximizing the logarithm of the probability that the process generated the visible samples as a function of $P(d_s | d_t, t \in \mathcal{N}_s)$ and $P(d_t, t \in \mathcal{N}_s)$:

$$\hat{\mathbf{P}} = \arg \max_{\mathbf{P}} \sum_{k \in \{r, g, b\}} \log P(o_k | \mathbf{P}, \mathbf{Q}_k). \quad (13)$$

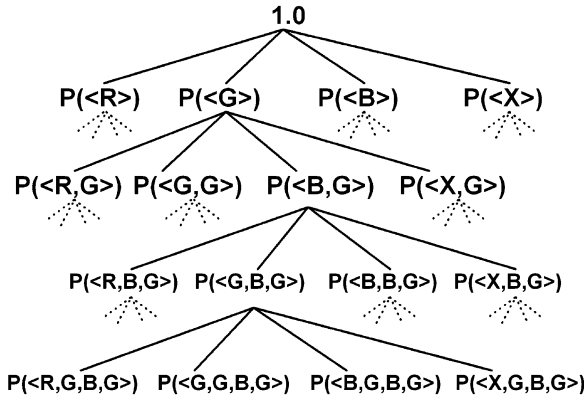


Fig. 5 Quadtree of tuple probabilities. The probability of each vertex equals the sum of the probabilities of its children.

Under the isotropy assumption, there are 181 unknown conditional probabilities. In addition, 37 of the joint probabilities of state four-tuples are also unknown. The conditional probabilities are constrained by

$$\forall i, j, k, l \quad \sum_m P_{mlkji} = 1, \tag{14}$$

where $i, j, k, l, m \in \{R, G, B, X\}$. There are also constraints on the joint probabilities of state four-tuples. To illustrate these constraints, we can build a quadtree of the four-tuple probabilities (see Fig. 5). The constraint is

$$\sum_{\langle l, k, j, i \rangle \in \mathcal{C}(a)} P(\langle l, k, j, i \rangle) = P(a), \tag{15}$$

where a is the nearest known ancestor of the four-tuple and $\mathcal{C}(a)$ is the set of leaves of the quadtree which are on the branch which starts at a . We know that at least one of the ancestors of the four-tuple is known because the probabilities of all one-tuple's are known:

$$P(\langle R \rangle) = P_g(\langle R' \rangle), \tag{16}$$

$$P(\langle G \rangle) = P_r(\langle G' \rangle), \tag{17}$$

$$P(\langle B \rangle) = P_g(\langle B' \rangle), \tag{18}$$

$$P(\langle X \rangle) = 1 - P(\langle R \rangle) - P(\langle G \rangle) - P(\langle B \rangle). \tag{19}$$

Therefore, there are 218 unknown variables with constraints defined by Eqs. (14) and (15). The objective function is

$$\begin{aligned} f &= \sum_k \log P(o_k | \mathbf{P}, \mathbf{Q}_k) \quad (\text{by Eq. (8)}) \\ &= \sum_k \log \left[\prod_{s \in S} P((o_k)_s | (o_k)_t, t \in \mathcal{N}_s) \right] \end{aligned}$$

$$= \sum_k \sum_{(o_k)_t, t \in \mathcal{N}_s} H_k((o_k)_t, t \in \mathcal{N}_s) \log P((o_k)_s | (o_k)_t, t \in \mathcal{N}_s), \quad (20)$$

where $k \in \{r, g, b\}$ and $P((o_k)_s | (o_k)_t, t \in \mathcal{N}_s)$ is defined in Eq. (12). $H_k((o_k)_t, t \in \mathcal{N}_s)$ is the observed frequency of the four-tuple of symbols. The problem we need to solve is

- maximize: f
 subject to:
- (i) $0 \leq P(d_s | d_t, t \in \mathcal{N}_s) \leq 1$;
 - (ii) $0 \leq P(d_t, t \in \mathcal{N}_s) \leq 1$;
 - (iii) the constraint defined by Eq. (14);
 - (iv) the constraint defined by Eq. (15).

A software package, SNOPT (Gill et al., 2002), is used to solve the above problem.

3.3. Results for the Markov random field model

To test our method for ML estimation, we first specified a conditional distribution for a MRF with a first-order neighborhood system. 100 realizations of the MRF was generated on a lattice of size 600×740 by sampling from the conditional distribution using the Gibbs sampler. For each realization, three visible samples were obtained by mapping the hidden states in the realization to the observation symbols using the observation matrices. By means of a raster scan with a window (see Fig. 3a) over the visible samples, the frequencies of the four-tuples and five-tuples of observation symbols were computed. We estimated 108 conditional probabilities directly from those frequencies. To estimate the remaining conditional probabilities, Eq. (20) was used to create an objective function and maximization subject to the constraints described in Section 3.1 was solved by SNOPT. This resulted in good estimates which generate realizations that are visually very similar to those generated using the specified conditional distributions (see Fig. 6). We computed the relative errors of estimates in terms of the matrix 2-norm, $\|\hat{\mathbf{P}} - \mathbf{P}\|_2 / \|\mathbf{P}\|_2$ (see Table 1). We note that the estimates are not numerically good for the case shown in Fig. 6b and c. One factor that may contribute to the error in the estimates is the lack of sufficient data for accurate estimation of conditional probabilities. Another factor may be the fact that the visible fields may be consistent with many different hidden distributions, and the ML estimation process can yield any of these distributions (Derin and Elliott, 1987).

4. A multiscale modeling approach

Unfortunately, the first-order neighborhood system is too small to get optimal results, especially when there are long range correlations between states. In addition, because

Table 1 Relative errors of estimates for the MRF modeling

	1	2	3	4
$\frac{\ \hat{\mathbf{P}} - \mathbf{P}\ _2}{\ \mathbf{P}\ _2} (\times 10^2)$	2.98	14.3	37.4	6.20
Figure	6a	6b	6c	6d

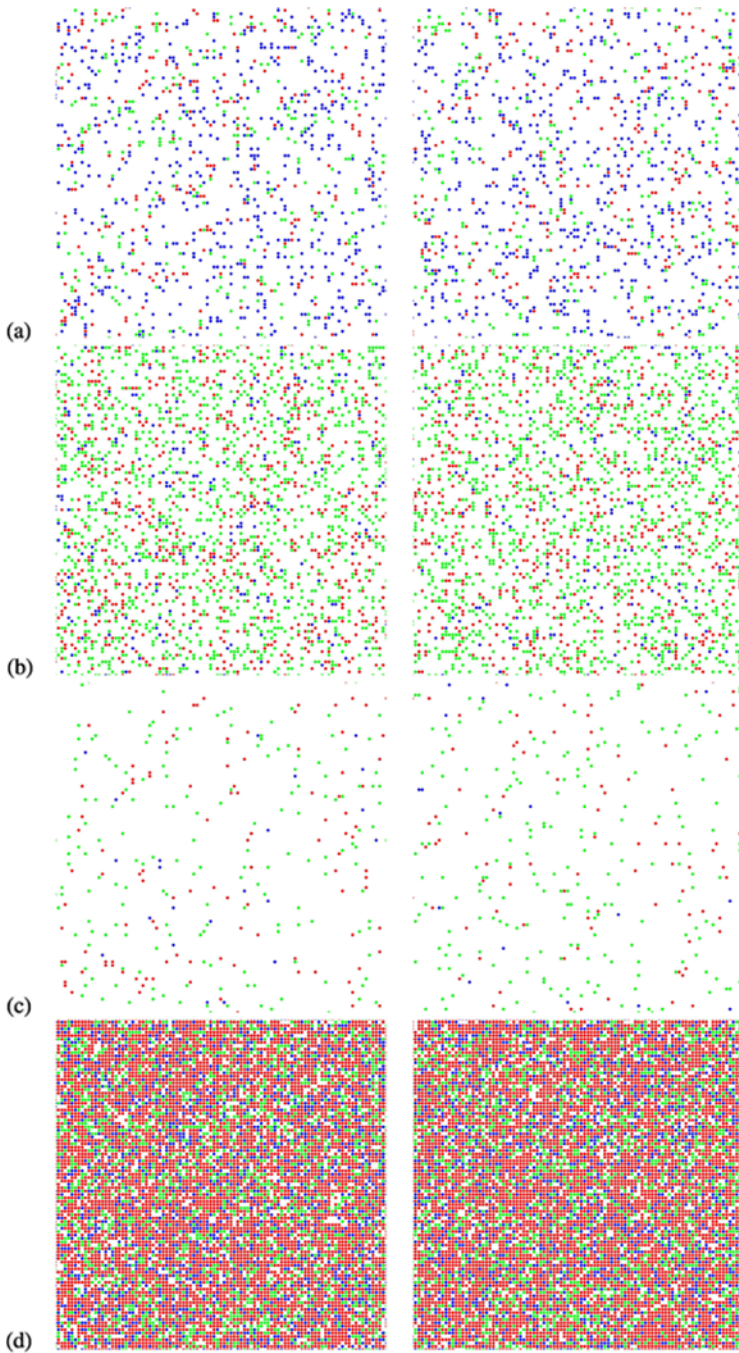


Fig. 6 Non-hierarchical MRF modeling on synthesized data. Each row shows the result of one experiment: samples generated with specified (right) and estimated (left) conditional distributions.

10 nm particles are larger than 5 nm particles, the interactions between 10 nm particles are at distances longer than those between 5 nm particles in experimental data. Figure 7 illustrates the problem: The clustering of 10 nm particles is optimally characterized at the scale of 80 pixels, but clustering of 5 nm particles is optimally characterized at the scale of 40 pixels. We also note that artifacts due to the small neighborhood system (i.e., particle positions aliasing with the lattice) become worse with coarser grids. Although we can use larger neighborhood systems following Tjelmeland and Besag (1998), the number of unknown variables dramatically increases. For example, there are 19,400 unknown variables for a second-order neighborhood system, which is too large for SNOPT to handle. Even if it were possible to find another package able to solve a system with such a large number of unknown variables, we do not have large enough datasets to reliably estimate the conditional distributions, because there are only 10 to 20 images captured in a typical experiment using nanoprobe.

4.1. A multiscale MRF model

To solve those problems, we have developed a multiscale MRF model. Following Bouman and Shapiro (1994), we build a pyramid with three layers (Fig. 8a). Each site in a coarse layer is a parent of four sites in the layer beneath it. Let us denote the three layers from coarse-to-fine, $L^{(2)}$, $L^{(1)}$, and $L^{(0)}$. For each layer, there is a corresponding distribution process, $D^{(i)}$ where $i \in \{2, 1, 0\}$. Each of these is modeled as a MRF with a first-order neighborhood system. Furthermore, two additional distribution processes are used to model the correlation between the values of sites in one layer and those in the layer beneath it: $D^{(2,1)}$ between $L^{(2)}$ and $L^{(1)}$, and $D^{(1,0)}$ between $L^{(1)}$ and $L^{(0)}$. Both of these are modeled as MRFs with a 5-neighbor neighborhood system, where four neighbors are in the same layer and the fifth neighbor is their parent in the layer above (Fig. 8b). There are also three corresponding observation processes for each of the distribution processes: $O_k^{(2)}$, $O_k^{(1)}$, $O_k^{(0)}$, $O_k^{(2,1)}$, and $O_k^{(1,0)}$ where $k \in \{r, g, b\}$. In our multiscale MRF model, the value of a site at a given scale depends not only on its parent in the layer above but also on its neighbors at the same scale. In this respect, our model is closely related to the models presented in (Kato et al. 1996, 1999; Mignotte et al., 2000; Wilson and Li, 2003). However, unlike the models described by these authors, we solve the statistical inference problem by means of a sequence of related multi-resolution problems rather than as a single problem representing the entire quadtree. Multi-resolution representations of the observed data at three scales are realizations of $O^{(i)}$, $i \in \{2, 1, 0\}$, and data between two scales are realizations of $O^{(2,1)}$ and $O^{(1,0)}$. We can infer the conditional p.m.f.'s of $D^{(i)}$ from $O_k^{(i)}$ as described in Section 3.1. Furthermore, if we can estimate the conditional p.m.f.'s of $D^{(2,1)}$ from $O_k^{(2,1)}$, and $D^{(1,0)}$ from $O_k^{(1,0)}$, then the Gibbs sampler can be used to generate samples from the conditional p.m.f.'s in a coarse-to-fine manner.

We tested the multiscale MRF model on the data used in Fig. 7. The data were processed at three different scales, with grid sizes of 20, 40 and 80 pixels, respectively. The finest scale was chosen to be 20 pixels because that is the most frequent distance of the nearest neighbor for every particle in the data set. When there are multiple proteins present at a single site, simple majority is used to determine the state of the site. Because there are only two proteins, the frequencies of the four-tuples and five-tuples of states are

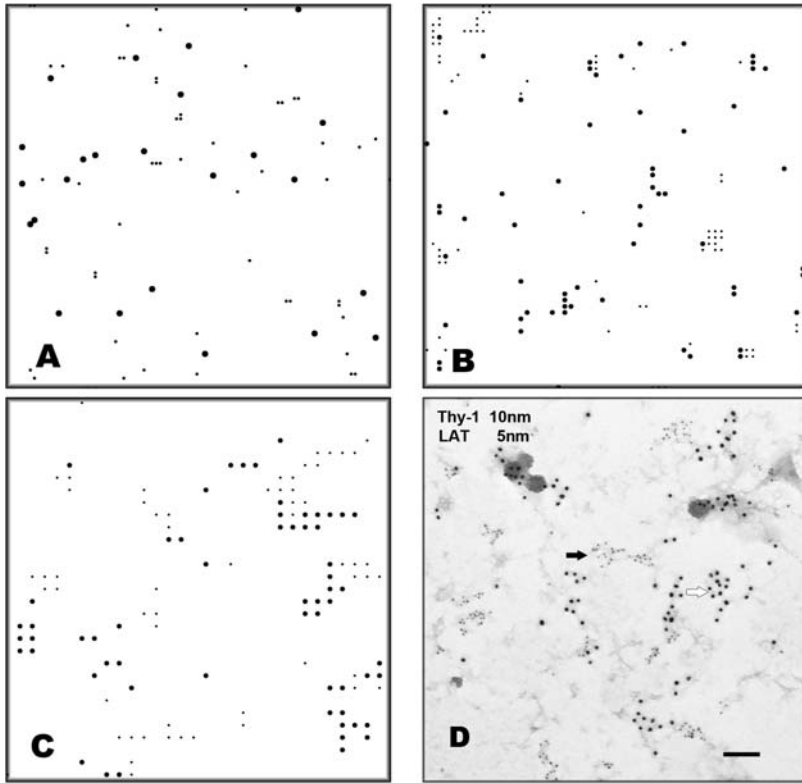


Fig. 7 Non-hierarchical MRF modeling on experimental data with two proteins. The first protein is a glycosylphosphatidylinositol-linked protein, Thy-1 (labeled with 10 nm particles), and the second is a linker for activation of T cells, LAT (labeled with 5 nm particles). The MRF model has a first-order neighborhood system. Because only two proteins are involved, the frequencies of the four-tuple's and five-tuple's of states can be computed directly from the observed data. The conditional probabilities of the MRF are computed using Eq. (3). A Gibbs sampler is used to generate samples from the conditional distribution of the MRF. The reconstruction depends on the grid size used to process the data. Grid sizes of 20, 40 and 80 pixels are used to process and reconstruct samples in (a), (b) and (c). A real TEM image is shown in (d). The black arrow points to 5 nm particles and the white arrow points to 10 nm particles. The bar equals 0.1 micron.

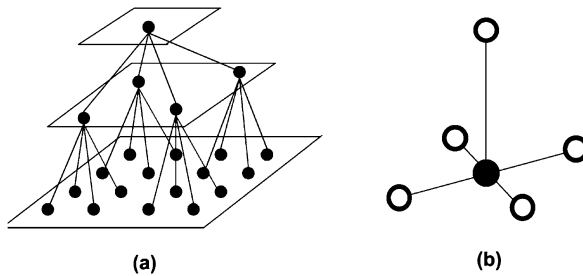


Fig. 8 Multiscale MRF model and 5-neighbor neighborhood system.

computed by means of a raster scan with a 80×80 pixel window (see Fig. 3a). The conditional probabilities of $D^{(2)}$ are then computed using Eq. (3). In an analogous way, the conditional probabilities of $D^{(1)}$ and $D^{(0)}$ are computed using 40×40 and 20×20 pixel windows respectively. By means of a raster scan using 80×80 and 40×40 pixel windows, the frequencies of the five-tuple's and six-tuple's of states (see Section 4.2) can also be computed. The conditional probabilities of $D^{(2,1)}$ are then computed using Eq. (21). In an analogous manner, the conditional probabilities of $D^{(1,0)}$ are computed using windows of size 40×40 and 20×20 pixels. The Gibbs sampler is then used to generate samples from the conditional p.m.f.'s:

1. Initialize $D^{(2)}$ randomly.
2. Sample from the conditional p.m.f. of $D^{(2)}$.
3. Initialize $D^{(1)}$ randomly.
4. While keeping $D^{(2)}$ unchanged, sample from the conditional p.m.f. of $D^{(2,1)}$.
5. Initialize $D^{(0)}$ randomly.
6. While keeping $D^{(1)}$ unchanged, sample from the conditional p.m.f. of $D^{(1,0)}$.

Keeping the upper layer unchanged while sampling from MRFs modeling adjacent layers allows long range properties to propagate from coarse scales to fine scales. The multiscale MRF model is very good at characterizing both short and long range interactions between states as shown in Fig. 9. Compared to Fig. 7a, b and c, particle distributions in Fig. 9c are visually very similar to those in the real sample images and the artifacts due to the use of the relatively small neighborhood system are minimal. There are 10 images in our sample dataset, which is adequate to obtain good estimates of the conditional probabilities as shown in Fig. 9.

4.2. Parameter estimation for MRFs with a 5-neighbor neighborhood system

In order to apply the multiscale MRF model to data where there are three proteins, the parameters of the hidden MRF with a 5-neighbor neighborhood system first need to be estimated from the visible fields. We adopted a strategy similar to that described in Section 3.1 to solve the ML estimation problem. As shown in Fig. 3b, we number the sites in a 5-neighbor neighborhood system. Using this numbering scheme, a six-tuple $\langle n, m, l, k, j, i \rangle$ and a five-tuple $\langle m, l, k, j, i \rangle$ can be used to denote the joint events

$$(D_5^{(u,w)} = n, D_4^{(u,w)} = m, D_3^{(u,w)} = l, D_2^{(u,w)} = k, D_1^{(u,w)} = j, D_0^{(u,w)} = i)$$

and

$$(D_4^{(u,w)} = m, D_3^{(u,w)} = l, D_2^{(u,w)} = k, D_1^{(u,w)} = j, D_0^{(u,w)} = i)$$

respectively, where u denotes an upper layer and w denotes a lower layer. For the sake of notation, we use the prime symbol to denote (u, w) . The conditional probability

$$P(D'_5 = n | D'_4 = m, D'_3 = l, D'_2 = k, D'_1 = j, D'_0 = i)$$

can then be estimated:

$$\begin{aligned} P(D'_5 = n | D'_4 = m, D'_3 = l, D'_2 = k, D'_1 = j, D'_0 = i) \\ = P_{nmlkji} \approx \frac{H(\langle n, m, l, k, j, i \rangle)}{H(\langle m, l, k, j, i \rangle)}, \end{aligned} \quad (21)$$

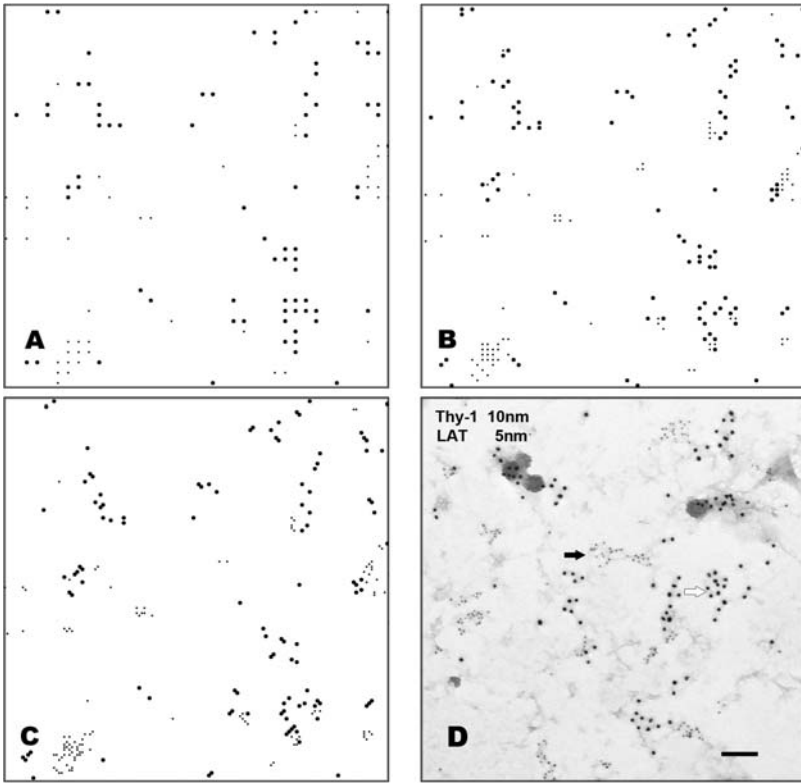


Fig. 9 Multiscale MRF modeling on experimental data with two proteins. The first protein is a glycosylphosphatidylinositol-linked protein, Thy-1 (labeled with 10 nm particles), and the second is a linker for activation of T cells, LAT (labeled with 5 nm particles). (a) A sample of $D^{(2)}$ generated from the conditional p.m.f. of $D^{(2)}$; (b) a sample of $D^{(1)}$ after sampling from the conditional p.m.f. of $D^{(2,1)}$; (c) a sample of $D^{(0)}$ after sampling from the conditional p.m.f. of $D^{(1,0)}$; and (d) a real TEM image. The black arrow points to 5 nm particles and the white arrow points to 10 nm particles. The bar equals 0.1 micron.

where H are histograms of the tuples. The probability of a realization of the distribution process, d' , given \mathbf{P} is also approximated by a product of the conditional probabilities:

$$P(d'|\mathbf{P}) \approx \prod_{s \in S^{(w)}} P(d'_s | d'_t, t \in \mathcal{N}_s), \quad (22)$$

where $S^{(w)}$ is the set of sites in the lower layer. The probability of a realization of the observation process, o'_k , given \mathbf{P} and \mathbf{Q}_k is approximated as:

$$P(o'_k | \mathbf{P}, \mathbf{Q}_k) \approx \prod_{s \in S^{(w)}} P((o'_k)_s | (o'_k)_t, t \in \mathcal{N}_s), \quad (23)$$

where

$$P((o'_k)_s | (o'_k)_t, t \in \mathcal{N}_s)$$

$$= \sum_{d'_s, d'_t, t \in \mathcal{N}_s} \frac{P(d'_s | d'_t, t \in \mathcal{N}_s) P(d'_t, t \in \mathcal{N}_s) \prod_{t \in \{s\} \cup \mathcal{N}_s} P((o'_k)_t | d'_t)}{P((o'_k)_t, t \in \mathcal{N}_s)} \quad (24)$$

and where $k \in \{r, g, b\}$. $P((o'_k)_s | d'_s)$ are known and defined in \mathbf{Q}_k . $P((o'_k)_t, t \in \mathcal{N}_s)$ are the probabilities of symbol five-tuples in the visible fields and can be estimated from the visible samples. $P(d'_t, t \in \mathcal{N}_s)$ are the probabilities of state five-tuples in the hidden field. In Eq. (24), $P(d'_s | d'_t, t \in \mathcal{N}_s)$ and $P(d'_t, t \in \mathcal{N}_s)$ are unknown variables. The maximum likelihood estimation of \mathbf{P} is achieved by maximizing the logarithm of the probability that the process generated the visible samples as a function of $P(d'_s | d'_t, t \in \mathcal{N}_s)$ and $P(d'_t, t \in \mathcal{N}_s)$:

$$\hat{\mathbf{P}} = \arg \max_{\mathbf{P}} \sum_{k \in \{r, g, b\}} \log P(o'_k | \mathbf{P}, \mathbf{Q}_k). \quad (25)$$

Isotropy is assumed only between neighbors that are in the lower layer. The conditional probabilities vary if we swap between one neighbor at the lower layer and the neighbor at the upper layer. The number of unknown conditional probabilities in \mathbf{P} is 802. In addition, there are 181 unknown probabilities of state five-tuples. The total number of unknown variables is 983. As in Section 3.1, we build a quadtree of state tuples where the five-tuples are leaves. The probabilities of the five-tuples are constrained by

$$\sum_{(m, l, k, j, i) \in \mathcal{C}(a)} P((m, l, k, j, i)) = P(a), \quad (26)$$

where a is the nearest known ancestor of the five-tuples and $\mathcal{C}(a)$ is the set of leaves of the quadtree which are on the branch which starts at a . The problem we need to solve is

maximize:

$$\begin{aligned} f &= \sum_k \log P(o'_k | \mathbf{P}, \mathbf{Q}_k) \\ &= \sum_k \sum_{(o'_k)_t, t \in \mathcal{N}_s} H_k((o'_k)_t, t \in \mathcal{N}_s) \log P((o'_k)_s | (o'_k)_t, t \in \mathcal{N}_s) \end{aligned} \quad (27)$$

subject to:

- (i) $0 \leq P(d'_s | d'_t, t \in \mathcal{N}_s) \leq 1$;
- (ii) $0 \leq P(d'_t, t \in \mathcal{N}_s) \leq 1$;
- (iii) $\forall d'_t, t \in \mathcal{N}_s \quad \sum_{d'_s} P(d'_s | d'_t, t \in \mathcal{N}_s) = 1$;
- (iv) the constraint defined by Eq. (26).

The above problem is also solved using SNOPT.

We did seven experiments to examine the quality of estimation. A conditional distribution for a MRF with a 5-neighbor neighborhood system was first specified. 100 realizations of the MRF was generated on a double lattice (a lattice of size 300×370 above a lattice of size 600×740) by sampling from the conditional distribution using the Gibbs sampler. For each realization, three visible samples were obtained by mapping the hidden

Table 2 Relative errors of estimates for the MRF modeling with a 5-neighbor neighborhood system

	1	2	3	4	5	6	7
$\frac{\ \hat{\mathbf{P}}-\mathbf{P}\ _2}{\ \mathbf{P}\ _2} (\times 10^2)$	44.1	33.7	29.9	35.6	35.9	45.3	33.3

states in the realization to the observation symbols using the observation matrices. By means of a raster scan with a window (see Fig. 3b) over the visible samples, the frequencies of the five-tuples and six-tuples of observation symbols were computed. To estimate the conditional probabilities, Eq. (27) was used to create an objective function and maximization subject to the constraints described above was solved by SNOPT.

Table 2 shows relative errors of estimates in terms of matrix 2-norm for the seven experiments. Like the MRF model with a first-order neighborhood system, there are errors in the estimates. When the number of samples are increased by a factor of 10, similar errors were also observed. It is very likely that the errors are due to the fact that the visible fields are consistent with many different hidden distributions. As demonstrated in Sections 4.3 and 4.4, although there were errors in estimates of the conditional probabilities of the MRF with a 5-neighbor neighborhood system, the multiscale MRF model worked well when applied to synthesized and experimental data.

4.3. Results on synthesized data

We applied the multiscale MRF model to synthesized data where we know the ground truth. The conditional probabilities of $D^{(2)}$, $D^{(2,1)}$, and $D^{(1,0)}$ were first specified. A Gibbs sampler was then used to generate samples from the conditional distributions through a coarse-to-fine pass described in Section 4.1. The three scales used to generate the samples were 80, 40 and 20 pixels. The coordinates of all three proteins at the finest layer in the samples were exported as data. We eliminated the coordinates of R to generate a set of data where R is missing. In an analogous way, a set of data where G is missing and a set of data where B is missing were obtained. We processed the data sets at the same three scales used to generate them. The data sets at scale of 80 pixels are regarded as samples of $O_k^{(2)}$ and were used to estimate the conditional p.m.f. of $D^{(2)}$. The data sets at scales of 80 and 40 pixels were used to estimate the conditional p.m.f. of $D^{(2,1)}$. The data sets at scales of 40 and 20 pixels were used to estimate the conditional p.m.f. of $D^{(1,0)}$. Finally, the estimates of the conditional probabilities of $D^{(2)}$, $D^{(2,1)}$, and $D^{(1,0)}$ were used to generate samples by means of Gibbs sampling in a coarse-to-fine manner as before. This produced satisfactory results. As shown in Fig. 10, the samples of the finest layer generated with specified and estimated conditional probabilities are visually very similar.

4.4. Results on experimental data

As a critical test, we applied the multiscale MRF model to two sets of experimental data obtained by immunogold labeling of membrane sheets. They were prepared from rat basophilic leukemia 2H3 (RBL-2H3) mast cells, that express the high affinity IgE receptor, FcεRI. Crosslinking this receptor with multivalent ligand activates a complex, multicomponent tyrosine kinase-dependent signaling pathway leading to the release of histamine

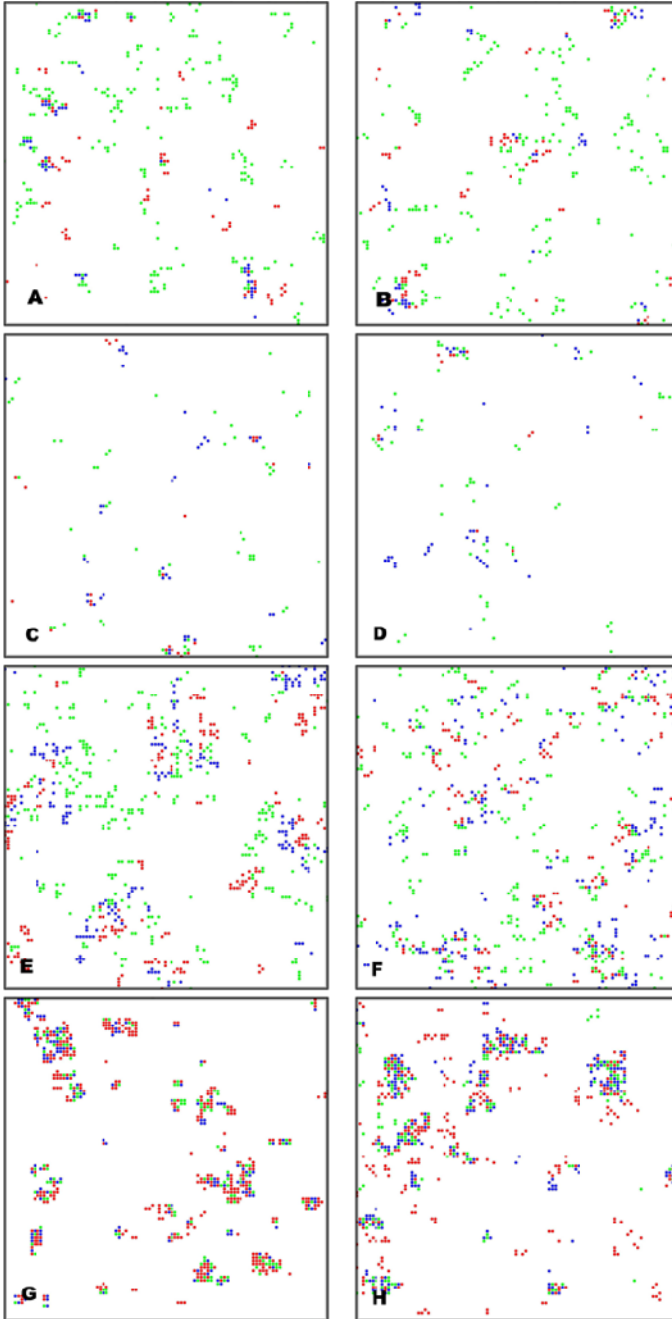


Fig. 10 Multiscale MRF modeling on synthesized data. Each row shows the result of one experiment: samples of the finest layer of the multiscale MRF model generated with specified (left) and estimated (right) conditional distributions.

and other mediators of allergic and asthmatic responses. Early events in the Fc ϵ RI signaling cascade include receptor redistribution into large clusters, the recruitment of both membrane-bound and cytoplasmic signaling proteins to the receptor-rich domains and also the segregation of certain tyrosine phosphorylated scaffolding and signaling proteins away from the receptor-rich domains (Wilson et al. 2000, 2001, 2002).

The first experiment involved three signaling species: the β subunit of the Fc ϵ RI itself (β), linker for activation of T cells (LAT), and phospholipase C γ isoform 1 (PLC γ 1). LAT is a transmembrane scaffolding protein that helps to recruit signaling proteins to the membrane. PLC γ 1 is an enzyme that cleaves phosphatidylinositol lipid, generating second messengers such as inositol triphosphate and diacylglycerol. Previous work from this group implicated LAT in the organization of “secondary signaling domains,” since LAT clusters segregate from IgE receptors and these clusters colocalize with PLC γ 1 (Wilson et al., 2001). To reconstruct the distributions all three proteins in the same membrane, we analyzed three sets of double-label data: (1) PLC γ 1 labeled with 5 nm particles and LAT labeled with 10 nm particles (Fig. 11a); (2) LAT labeled with 5 nm particles and β labeled with 10 nm particles (Fig. 11b); and (3) PLC γ 1 labeled with 5 nm particles and β labeled with 10 nm particles (Fig. 11c). There were 10 images in each data set. Spatial point patterns for the two particle sizes were extracted from digitized images as described in (Zhang et al., 2006). We computed histograms of the distances of the nearest neighbor for every particle in the data sets and found that the most frequent distance is 16 pixels. Consequently, windows of size 16×16 , 32×32 , and 64×64 pixels were used to estimate the histograms of symbols at the three scales. LAT, PLC γ 1, and β were assigned to R , G , and B , respectively. In Fig. 11d, the sample reconstructed from the modeling demonstrates both the colocalization of PLC γ 1 (green) with LAT (red) and the segregation of LAT/PLC γ 1 from β (blue) in a single integrated image.

In the second experiment, we again chose the Fc ϵ RI β subunit but added two different signaling species, Grb2-binding protein 2 (Gab2) and the p85 subunit of phosphatidylinositol 3-kinase (p85 of PI 3-kinase). Gab2 is an adaptor protein and PI 3-kinase is an enzyme that phosphorylates phosphatidylinositol lipids in the 3 position on the inositol ring. Previous double-label studies indicated that Gab2 and p85 are both recruited to receptor-rich domains in activated RBL-2H3 cells (Wilson et al., 2001). This dataset was also composed of three double-label protocols, with 10 images each: (1) Gab2 was labeled with 5 nm particles and β was labeled with 10 nm particles (Fig. 12a); (2) Gab2 was labeled with 5 nm particles and p85 was labeled with 10 nm particles (Fig. 12b); and (3) p85 was labeled with 5 nm particles and β was labeled with 10 nm particles (Fig. 12c). The data are processed at scales of 20, 40, and 80 pixels. In Fig. 12d, the sample reconstructed from the modeling confirms the colocalization of p85 (red), Gab2 (green), and β (blue) in a single integrated image.

5. Discussion

The heterogeneous organization of plasma membrane components and functions has been recognized for decades (Berlin et al., 1974) and there is growing consensus that plasma membrane heterogeneity is a critical regulator of signaling pathway activity. Thus, correlating the activities of receptors and signaling proteins and lipids with their spatial distribution and dynamics is essential to understand the regulation of cell signaling.

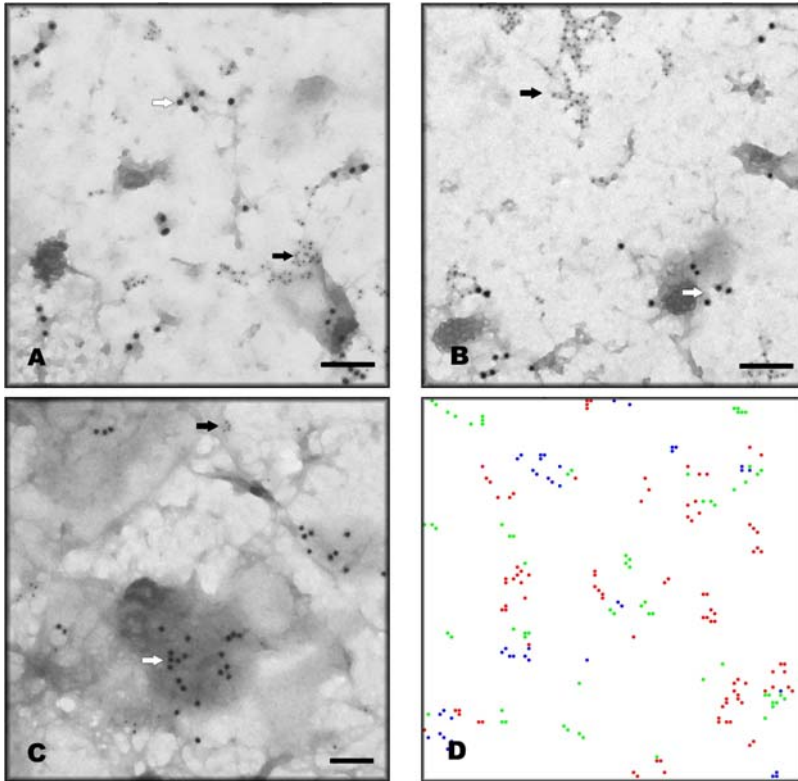


Fig. 11 Multiscale MRF modeling for the first experiment involving three proteins where only two are observable in any single sample. (a) TEM image where PLC γ 1 (5 nm) and LAT (10 nm) are observed; (b) TEM image where LAT (5 nm) and β (10 nm) are observed; (c) TEM image where PLC γ 1 (5 nm) and β (10 nm) are observed; and (d) a reconstruction computed using Gibbs sampling that shows the spatial distributions of all three proteins, LAT (red), PLC γ 1 (green), and β (blue). The black arrows point to 5 nm particles and the white arrows point to 10 nm particles. The bar equals 0.1 micron.

A range of static and dynamic methods is available to investigate receptor distributions on membranes. In the mast cell system that our biology team studies, the high affinity IgE receptor, Fc ϵ RI, has been tracked by scanning electron microscopy with immunogold probes (Seagrave et al., 1991), fluorescence recovery after photobleaching (FRAP) with fluorescent probes (Thomas et al., 1992), time-resolved phosphorescence anisotropy (TPA) with phosphorescent probes (Rahman et al., 1992) and single particle tracking using either gold nanoprobles (Baris et al., 2007) or quantum dot (QD) probes (Andrews et al., 2007; Lidke et al., 2007). Some of these methods can be used to localize more than one molecular species, for example, the analysis of QD-Fc ϵ RI dynamics in the context of GFP-actin distribution (Lidke et al., 2007). Correlated movement of receptors and signaling proteins can in principle be studied by fluorescence correlation spectroscopy (FCS) and protein-protein interactions can be studied by fluorescence resonance energy transfer (FRET). However, these dual-label techniques require transfection with fluorescent analogs of signaling proteins and, in the case of FRET, favorable molecular geometry to

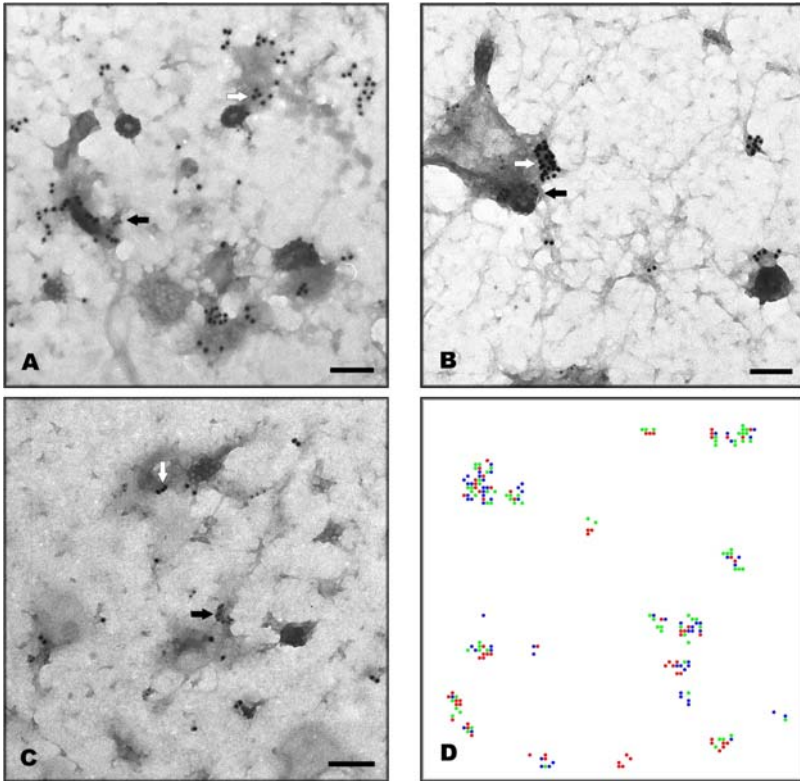


Fig. 12 Multiscale MRF modeling for the second experiment involving three proteins where only two are observable in any single sample. (a) TEM image where Gab2 (5 nm) and β (10 nm) are observed; (b) TEM image where Gab2 (5 nm) and p85 (10 nm) are observed; (c) TEM image where p85 (5 nm) and β (10 nm) are observed; and (d) a reconstruction computed using Gibbs sampling that shows the spatial distributions of all three proteins, p85 (red), Gab2 (green), and β (blue). The black arrows point to 5 nm particles and the white arrows point to 10 nm particles. The bar equals 0.1 micron.

enable FRET to occur. Although protein-protein colocalization can be measured dynamically, relationship to membrane architecture is not revealed and extension to more than two labels would be very difficult.

To date, immuno-electron microscopy on membrane sheets provides the highest resolution view of membrane topography. It also enables observations of spatial relationships between several different molecular species at high resolution and without the need for transfection. Importantly, it also reveals details of membrane architecture, such as the proximity or residency of receptors within coated pits or caveolae. Thus, its chief limitation is the small number of different molecular species that can be imaged on the same membrane sheet. As a first step toward overcoming this limitation, we have developed a multiscale hidden MRF model that can be used reliably to estimate the spatial relationships among three proteins from experiments showing spatial relationships between only two of these proteins. The code is available on our web-site cellpath.health.unm.edu and we are optimistic that it will be adopted by other groups.

New applications of the MRF will enhance the utility of the current model. For example, in the experiments above, it is assumed that all proteins of interest on cell membranes are labeled by gold particles. We use *labeling efficiency* to denote the percentage of a protein labeled by gold particles in a biological experiment. However, in our experience, labeling efficiencies in experiments using nanoprobe are typically less than 100%. These values must be determined individually for each molecular species probed, using quantitative methods such as flow cytometry or western blotting that can be fitted to reliable standard curves, and may range from 50–70% (data not shown). Once these quantitative methods have been used to estimate labeling efficiency in a given biological experiment, MRF modeling can be used to correct for underlabeling. The strategy is to include the labeling efficiency in the observation matrices as follows:

$$\mathbf{Q}_r = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & p_r^g & 0 & 0 \\ 0 & 0 & p_r^b & 0 \\ 1 & 1 - p_r^g & 1 - p_r^b & 1 \end{bmatrix}, \quad \mathbf{Q}_g = \begin{bmatrix} p_g^r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & p_g^b & 0 \\ 1 - p_g^r & 1 & 1 - p_g^b & 1 \end{bmatrix},$$

$$\mathbf{Q}_b = \begin{bmatrix} p_b^r & 0 & 0 & 0 \\ 0 & p_b^g & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 - p_b^r & 1 - p_b^g & 1 & 1 \end{bmatrix},$$

where p_i^j is the labeling efficiency of protein j in experiments where protein i is missing. The method has recently been used to correct for 40% underlabeling of the epidermal growth factor receptor (EGFR) on breast cancer cell membranes, permitting more accurate estimates of the numbers of receptors per cluster (Yang et al., 2007).

We currently use MRFs with a first-order neighborhood system. In the next phase of work, we predict that a larger neighborhood system can be implemented, improving the quality of the reconstruction computed using Gibbs sampling. However, as the size of the neighborhood system increases, the number of unknown variables becomes too large to be solved using SNOPT. By exploiting Markov–Gibbs equivalence (Besag, 1974; Li, 1995), we can write the conditional probability in terms of clique potentials. We believe that by using the Gibbs formulation and by exploiting isotropy within cliques the number of unknown variables can be further reduced, so that we can utilize a MRF model based on a second-order neighborhood system (i.e., an 8-neighborhood). In addition, it is easier to obtain biological insights when the results are interpreted in terms of clique potentials. For example, if the potential of RB 2-clique is larger than that of RG 2-clique, R is more likely to colocalize with G than with B . Using clique potentials, combined with MRF modeling’s ability to derive confidence intervals, we can provide quantitative measurements of co-clustering for molecular species.

These new approaches will allow the modeling to keep pace with technological advancements that permit the experimentalists to routinely label more than two molecular species at a time. These advancements will rely on successful applications of new shape-defined 2–10 nm nanoparticles (Hernandez-Sanchez et al., 2006). The upper limit for simultaneous labeling is likely to be at most 4 or 5 molecular species at a time, for at least

as long as antibodies remain the principal reagents for probe targeting. Our proposed extensions to the code will permit combinations of these larger data sets to be analyzed, permitting more complete maps of molecular topography during signal transduction.

6. Conclusion

We have described a novel approach for reconstructing spatial relationships between three proteins on cell membranes from samples showing relationships between only two proteins. This approach utilizes a multiscale hidden Markov random field model where mathematical programming techniques are used to deduce the conditional distributions. To our knowledge, we are the first to use Markov random fields to model the spatial distribution of proteins on cell membranes. The application of our approach to synthesized data has demonstrated that the multiscale MRF model is good at characterizing both short and long range statistical properties and that the spatial relationships among three proteins can be reliably estimated. The application to experimental data has provided the first maps of the spatial relationship between groups of three different signaling molecules. The ability to analyze the spatial organization and dynamics of multiple membrane proteins during signaling is a critical step towards a more complete understanding of the spatial and temporal regulation of signal transduction pathways.

Acknowledgements

We thank Janet R. Pfeiffer for technical assistance. This work was supported in part by NIH grants R01 GM49814, R01 CA119232, R01 AI051575 and P20 GM065794 and by a graduate fellowship to JZ from the Gies Foundation of the UNM Cancer Research and Treatment Center. Use of the electron microscopy facility at the UNM SOM, and NIH support for this facility, is gratefully acknowledged.

References

- Andrews, N.L. et al., 2007. Dynamics, topography, and microdomains in FcεRI signaling. *Biophys. J.*, submitted.
- Barisas, B.G. et al., 2007. Compartmentalization of the Type I Fc epsilon receptor and MAFA on mast cell membranes. *Biophys. Chem.* 126, 209–217.
- Berlin, R.D. et al., 1974. Control of cell surface topography. *Nature* 247, 45–46.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Stat. Soc. Ser. B* 36(2), 192–236.
- Besag, J., 1986. On the statistical analysis of dirty pictures. *J. Roy. Stat. Soc. Ser. B* 48(3), 259–302.
- Bouman, C.A., Shapiro, M., 1994. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. Image Process.* 3(2), 162–177.
- Celeux, G., Forbes, F., Peyrard, N., 2003. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recognit.* 36(1), 131–144.
- Chalmond, B., 1989. An iterative Gibbsian technique for reconstruction of M-ary images. *Pattern Recognit.* 22(6), 747–762.
- Cross, G.R., Jain, A.K., 1983. Markov random field texture models. *IEEE Trans. Pattern Anal. Mach. Intell.* 5(1), 25–39.
- Dempster, A.P., Laird, N., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. Ser. B* 39(1), 1–38.

- Derin, H., Elliott, H., 1987. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 9(1), 39–55.
- Efros, A.A., Leung, T.K., 1999. Texture synthesis by non-parametric sampling. In: *ICCV '99: Proceedings of the International Conference on Computer Vision*, vol. 2, p. 1033, Washington, DC, USA. IEEE Computer Society.
- Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 6(6), 721–741.
- Gill, P.E., Murray, W., Saunders, M.A., 2002. SNOPT: an SQP algorithm for large-scale constrained optimization. *SIAM J. Optim.* 12(4), 979–1006.
- Gurelli, M.I., Onural, L., 1994. On a parameter estimation method for Gibbs–Markov random fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 16(4), 424–430.
- Hancock, J.F., Prior, I.A., 2005. Electron microscopic imaging of ras signaling domains. *Methods* 37(2), 165–712.
- Hernandez-Sanchez, B.A. et al., 2006. Synthesizing biofunctionalized nanoparticles to image cell signaling pathways. *IEEE Trans. NanoBiosc.* 5, 222–230.
- Kato, Z., Berthod, M., Zerubia, J., 1996. A hierarchical Markov random field model and multitemperature annealing for parallel image classification. *CVGIP: Graph. Model Image Process.* 58(1), 18–37.
- Kato, Z., Zerubia, J., Berthod, M., 1999. Unsupervised parallel image classification using Markovian models. *Pattern Recognit.* 32(4), 591–604.
- Kim, J.H. et al., 2005. Independent trafficking of Ig- α /Ig- β and μ -heavy chain is facilitated by dissociation of the B cell antigen receptor complex. *J. Immunol.* 175, 147–154.
- Laferté, J.M., Pérez, P., Heitz, F., 2000. Discrete Markov image modeling and inference on the quadtree. *IEEE Trans. Image Process.* 9(3), 390–404.
- Li, S.Z., 1995. *Markov Random Field Modeling in Computer Vision*. Springer, London.
- Liang, K.H., Tjahjadi, T., 2006. Adaptive scale fixing for multiscale texture segmentation. *IEEE Trans. Image Process.* 15(1), 249–256.
- Lidke, K.A. et al., 2007. Direct observation of membrane proteins confined by actin corrals. *J. Cell Biol.*, submitted.
- Mignotte, M. et al., 2000. Sonar image segmentation using an unsupervised hierarchical MRF model. *IEEE Trans. Image Process.* 9(7), 1216–1231.
- Nicolau, D.V. et al., 2006. Identifying optimal lipid raft characteristics required to promote nanoscale protein-protein interactions on the plasma membrane. *Mol. Cell Biol.* 26, 313–323.
- Oliver, J.M. et al., 2004. Membrane receptor mapping: the membrane topography of Fc ϵ RI signaling. In: P. Quinn (Ed.), *Membrane Dynamics and Domains, Subcellular Biochemistry*, vol. 37. Kluwer Academic/Plenum, Dordrecht/New York, pp. 3–34.
- Paget, R., Longstaff, I.D., 1998. Texture synthesis via a noncausal nonparametric multiscale Markov random field. *IEEE Trans. Image Process.* 7(6), 925–931.
- Plowman, S.J. et al., 2005. H-ras, k-ras and inner plasma membrane raft proteins operate in nanoclusters with differential dependence on the actin cytoskeleton. *Proc. Natl. Acad. Sci.* 102, 15500–15505.
- Prior, I.A. et al., 2003. Direct visualization of ras proteins in spatially distinct cell surface microdomains. *J. Cell Biol.* 160(2), 165–170.
- Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 257–286.
- Rahman, N.A. et al., 1992. Rotational dynamics of type I Fc epsilon receptors on individually-selected rat mast cells studied by polarized fluorescence depletion. *Biophys. J.* 61, 334–346.
- Seagrave, J.C. et al., 1991. Relationship of IgE receptor topography to secretion in RBL-2H3 mast cells. *J. Cell Physiol.* 148, 139–151.
- Thomas, J.L., Feder, T.J., Webb, W.W., 1992. Effects of protein concentration on IgE receptor mobility in rat basophilic leukemia cell plasma membranes. *Biophys. J.* 61, 1402–1412.
- Tjelmeland, H., Besag, J., 1998. Markov random fields with higher order interactions. *Scand. J. Stat.* 25, 415–433.
- Tonazzini, A., Bedini, L., Salerno, E., 2006. A Markov model for blind image separation by a mean-field EM algorithm. *IEEE Trans. Image Process.* 15(2), 473–482.
- Volna, P. et al., 2004. Negative regulation of mast cell signaling and function by the adaptor lab/ntal. *J. Exp. Med.* 200(8), 1001–1013.
- Wilson, R., Li, C.T., 2003. A class of discrete multiresolution random fields and its application to image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(1), 42–56.

- Wilson, B.S., Pfeiffer, J.R., Oliver, J.M., 2000. Observing Fc ϵ RI signaling from the inside of the mast cell membrane. *J. Cell Biol.* 149(5), 1131–1142.
- Wilson, B.S. et al., 2001. High resolution mapping of mast cell membranes reveals primary and secondary domains of Fc ϵ RI and LAT. *J. Cell Biol.* 154(3), 645–658.
- Wilson, B.S., Pfeiffer, J.R., Oliver, J.M., 2002. Fc ϵ RI signaling observed from the inside of the mast cell membrane. *Mol. Immun.* 38, 1259–1268.
- Wilson, B.S. et al., 2004. Markers for detergent-resistant lipid rafts occupy distinct and dynamic domains in native membranes. *Mol. Biol. Cell* 15(6), 2580–2592.
- Xue, M. et al., 2007. FPR and Fc ϵ RI occupy common signaling domains for localized crosstalk. *Mol. Biol. Cell* 18, 1410–1420.
- Yang, S. et al., 2007. Mapping ErbB receptors on breast cancer cell membranes during signal transduction. *J. Cell Sci.* 120, 2763–2773.
- Zhang, J., 1992. The mean field theory in EM procedures for Markov random fields. *IEEE Trans. Image Process.* 40(10), 2570–2583.
- Zhang, J., Modestino, J.W., Langan, D.A., 1994. Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. *IEEE Trans. Image Process.* 3(4), 404–420.
- Zhang, J. et al., 2006. Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale electron-dense probes and electron microscopy. *Micron* 37(1), 14–34.