


Modeling the Cluster Size Distribution of Vascular Endothelial Growth Factor (VEGF) Receptors

Emine Güven¹ , Michael J Wester^{2,3},
Jeremy S Edwards^{4,5,6,7} and Ádám M Halász⁸

¹Department of Biomedical Engineering, Düzce University, Düzce, Turkey. ²Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM, USA. ³Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, USA. ⁴Department of Chemistry and Chemical Biology University of New Mexico, Albuquerque, NM, USA. ⁵Department of Chemical and Biological Engineering, University of New Mexico, Albuquerque, NM, USA. ⁶Department of Molecular Genetics and Microbiology, University of New Mexico, Albuquerque, NM, USA. ⁷Comprehensive Cancer Center, University of New Mexico Health Sciences Center, Albuquerque, NM, USA. ⁸Mathematics, West Virginia University, Morgantown, WV, USA.

Bioinformatics and Biology Insights
Volume 16: 1–14
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11779322221085078


ABSTRACT: We previously developed a method of defining receptor clusters in the membrane based on mutual distance and applied it to a set of transmission microscopy images of vascular endothelial growth factor receptors. An optimal length parameter was identified, resulting in cluster identification and a procedure that assigned a geometric shape to each cluster. We showed that the observed particle distribution results were consistent with the random placement of receptors within the clusters and, to a lesser extent, the random placement of the clusters on the cell membrane. Here, we develop and validate a stochastic model of clustering, based on a hypothesis of preexisting domains that have a high affinity for receptors. The proximate objective is to clarify the mechanism behind cluster formation and to estimate the effect on signaling. Receptor-enriched domains may significantly impact signaling pathways that rely on ligand-induced dimerization of receptors. We define a simple statistical model, based on the preexisting domain hypothesis, to predict the probability distribution of cluster sizes. The process yielded sets of parameter values that can readily be used in dynamical calculations as the estimates of the quantitative characteristics of the clustering domains.

KEYWORDS: Cell signaling, receptors, receptor clustering, VEGF, membrane organization

RECEIVED: December 9, 2021. **ACCEPTED:** February 12, 2022.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors gratefully acknowledge support from US NIH grants R01GM104979 (to JSE) and K25CA131558 (to AMH). The use of the University of New Mexico (UNM) Cancer Center core facilities, as well as the NIH

P30CA118100 support for these Cores, is gratefully acknowledged. Publication support was provided by West Virginia University.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Emine Güven, Department of Biomedical Engineering, Düzce University, Düzce 81620, Turkey. Email: emine.guven@duzce.edu.tr

Introduction

Cell signaling relies on a complex network of molecular interactions between ligands, receptors, co-receptors, and associated kinases and phosphatases. The dynamics of signaling pathways within these networks is the subject of intense study and is supported by a host of computational tools.¹ Correct biochemical parameters are essential to the predictive power of the underlying mathematical models, but obtaining *in vivo* values has been a long-standing challenge. Molecular resolution imaging techniques provide a direct window into molecular processes, close to *in vivo* conditions.

Our groups^{2,3} have been pursuing a modeling approach that is in close coordination with molecular scale imaging. One challenge specific to this strategy is the “resolution gap” between molecular scale imaging and other investigative methods,^{4–6} and quantitative understanding of cell-level function.^{7,8} While some of the underlying molecular-scale factors (such as small copy numbers and spatial inhomogeneity) are understood to a large extent, it is impractical to include this level of detail in signaling pathway models, which rely on chemical reaction networks and are normally used for well-mixed chemical systems. The rate constants and other parameters relevant to cell-level dynamics correspond to emergent behaviors that are an indirect reflection of molecular level events. Receptor clustering, the tendency of

receptors to collect in groups ranging from a few up to hundreds of copies, is a molecular-level phenomenon with a significant potential impact on cell-level behavior.⁹ It can alter the dynamics of receptor-receptor interactions, with a direct impact on signal initiation.

Signal initiation by many types of membrane bound receptors (receptor tyrosine kinase families, including vascular endothelial growth factor [VEGF] receptors) involves mutual activation of receptors,^{7,10} which requires close spatial proximity. This is typically facilitated by receptor-receptor binding resulting in dimers. The key element of signaling is that receptor-receptor binding is induced or strongly enhanced by ligand binding to receptors. The dynamics of the 3-step signal initiation sequence (ligand binding, receptor oligomerization, and mutual activation of receptors) are delicately tuned to elicit the correct biological response. Accumulation of receptors in a small region increases the likelihood of collisions between unbound receptors. The effective on-rate for receptors accumulated in a fraction of the membrane surface area is higher than for the same set of receptors distributed evenly across the entire membrane.

We previously analyzed the spatial distribution of VEGF^{7,11,12} receptors in this data set.¹³ The VEGF and its receptors (VEGFR) play an important role in regulating new



blood vessel growth from extant vessels (angiogenesis), and so is a critical factor in tumor development.

The possible impact of clustering on VEGF signaling was explored in the work by Chen et al.¹⁴ We assumed that the cell membrane contains preexisting regions (domains) that have a higher affinity to VEGF receptors; the primary effect would be the accumulation of receptors in these domains. To quantify the impact on signaling, we adapted a chemical reaction network model of VEGF signal initiation⁷ to a 2-compartment system. One of the compartments (or regions) represented the high-affinity domains and the other accounted for the rest of the cell membrane. The VEGF monomers and dimers were allowed to move between the 2 regions, with an enhancement factor $\alpha \geq 1$ that corresponded to the ratio of equilibrium concentrations of monomer receptors in the 2 regions. Increasing the value of the enhancement factor α while keeping the concentration of VEGF fixed led to the increased equilibrium levels of signaling capable complexes (stably bound, liganded receptor dimers).

While the origin of the phenomenon is not completely understood, the phenomenon of receptor clustering is likely a result of the quasi-random movement of receptors in an inhomogeneous membrane landscape. In previous work using this data set,¹³ using nearest neighbor distance (NND)-based methods, we established that the receptors form clusters and that their distribution within clusters is otherwise random. We also proposed an improved method for identifying clusters, based on Espinoza et al.¹⁵ The results were consistent with the hypothesis that clusters form in preexisting regions (domains) on the cell membrane. During their random movement on the membrane, receptors are placed in these domains with a higher probability than elsewhere, leading to an “enrichment.” This hypothesis is in itself not novel and was used in spatially resolved models aimed at specific signaling systems, such as in the works by Pryor et al.¹⁶ and Kerketta et al.¹⁷ However, to our knowledge, the relation between basic characteristics of attractive domains and the induced cluster size distributions has not been explored.

In this article, we focus on the distribution of cluster sizes. Starting from previous results on cluster identification and the inferred geometric footprints, we explore whether the cluster size distributions can be explained (and thus predicted) by a simple probabilistic model based on random placement of receptors into 1 or 2 sets of preexisting domains. We first summarize our previous results on cluster identification. We then outline the model(s) proposed for cluster size distributions and obtain model parameters optimized for the observed cluster size distributions. We perform model fits for each image taken separately and also for groups of images.

Methods and Models

Experimental data

The data set analyzed here is based on the set of transmission electron microscopy (TEM) micrographs studied by Güven

et al.¹³ We briefly summarize the relevant information and refer the reader to the work by Güven et al.¹³ for more details. We investigate the spatial distribution of VEGF receptors extracted from TEM images of VEGF receptors labeled using anti-VEGF antibody conjugated with gold nanoparticles. We are interested in the clustering of membrane receptors defined as the accumulation of receptors in a fraction of the available area, as seen in Figures 1D or 2A.

TEM images were of PAE-KDR (porcine aortic endothelial) cells that artificially express VEGFR-2 (KDR) receptors. Cells were stimulated with VEGF for selected times. Samples were prepared and imaged as described previously by Wilson et al.⁵ The micrographs represent snapshots of the position of receptors at the moment when the sample was prepared.

Receptor positions were obtained by identifying the centers of the respective dark spots on the micrographs using ImageJ.¹⁸ The resulting image coordinates were inserted into text files that were the primary source of data for all analyses presented here. The images analyzed here are 2650×2650 pixels (px), with a resolution of 1.448 px/nm; the image area is a square of side:

$$a = 2650 \text{ px} \approx 1830 \text{ nm, with area} \\ A_{\text{image}} = 7.02 \times 10^6 \text{ px}^2 \approx 3.35 \mu\text{m}^2$$

Cluster analysis

The spatial distribution of receptors in our micrographs is generally not random and appears so to a casual observer. A methodical analysis of the phenomenon should address 3 aspects: assess clustering tendency following an established methodology,^{15,19} provide a definition (identification) of clusters in the sample, and finally, characterize the distribution of receptors in the micrographs in a way that quantifies the phenomenon of clustering. Below, we briefly summarize our approach to the first 2 aspects. These results were reported previously and we refer the reader to the work by Güven et al.¹³ for further details.

Clustering tendency. We tested against the hypothesis of random uniform placement of N points observed in an image in an area A (a spatial Poisson process with density $\lambda = N / A$). We computed NND distributions (see Figure 1A to C) and Hopkins and Ripley statistics (which all rely on the mutual distance between pairs of points) and compared with the corresponding distributions expected from random placement.

Cluster identification. Our approach¹³ is a version of hierarchical distance-based clustering.²⁰ It is a development of Espinoza and coworkers¹⁵ who first adapted distance-based clustering to the analysis of nanogold-labeled membrane proteins.

Hierarchical clustering relies on a single length parameter L , which induces a pattern of connections between the points of a given set. The number of clusters induced in a given image decreases as L increases. We calculate the number of clusters as a function of the length parameter, $N_C(L)$ for each image. This is a decreasing function that equals the number of points

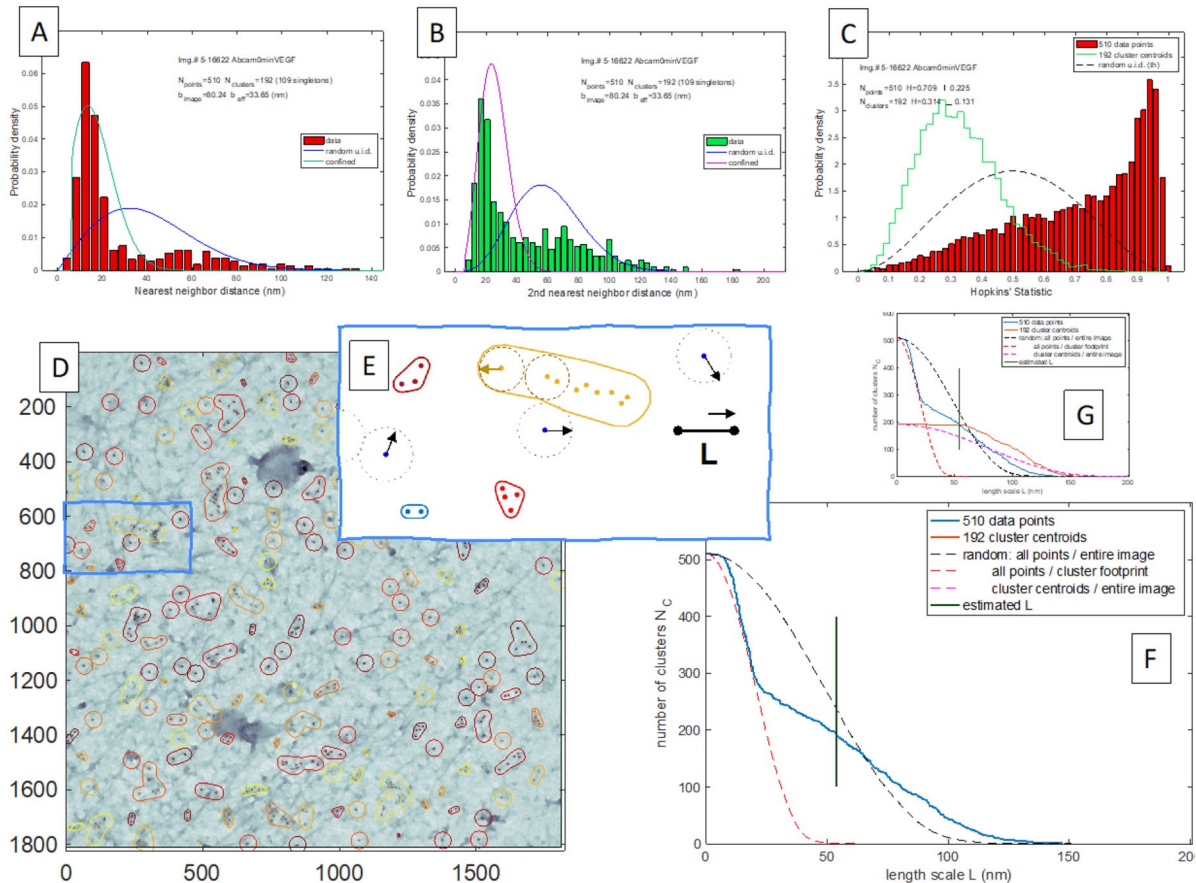


Figure 1. Measures of clustering, cluster identification, and footprint construction for one micrograph, based on¹⁰. Panel (D) shows the micrograph (5-16622, provided by courtesy of the Wilson Lab) overlaid with points identified with labeled VEGF receptors. The distribution of (A) first and (B) second NND and the Hopkins statistic (C) indicate the clustering of the points. In plot (C), bars indicate the Hopkins statistic for the set of all points, and the staircase (green) is the Hopkins statistic for cluster centroids. Hierarchical distance-based clustering: the optimal distance parameter L is identified from the curvature of the $N_C(L)$ function. (F, G) The number of induced clusters as a function of L . Cluster footprints (E) are an envelope of circles of radius $L_{\min}/2$ centered on each point in a cluster. For each cluster, we use the smallest distance parameter L_{\min} which does not split the cluster (so $L_{\min} \leq L$). Singletons (clusters of one point) are shown with a contour of diameter L .

when L is small and approaches 1 for large L . For a set of randomly distributed points, $N_C(L)$ follows a universal curve (dashed black line in Figure 1F and G). The short distance portion of $N_C(L)$ is consistent with having the same number of particles distributed randomly and uniformly in a smaller area (Figure 1F and G, red dashed line). We aim to approximate the transition between the compact distribution and the long distance behavior (consistent with the random distribution of a smaller number of points in the entire image area—compare the red staircase and purple dashed line in Figure 1G). We found it practical to identify the optimal L as 2 times the value of L corresponding to the point where the second derivative of $N_C(L)$ is zero.

Domain reconstruction algorithm for cluster footprints. For each identified cluster, we constructed a geometric footprint (shape) around the member points to provide us with an estimate of the area and perimeter for the cluster. The domain reconstruction algorithm was first described by Pryor et al.¹⁶ Points in a distance-based cluster with characteristic length L form a

connected graph whose edges have length less than or equal to L . The algorithm provides a contour that surrounds every point in the cluster, together with a disk of radius $L/2$ centered around it. Figure 1E illustrates how the contours were determined. Note that typically, clusters in a given image remain connected for length parameter values down to some $L_{\min} < L$; we used this intrinsic characteristic length to construct the footprint for each cluster.

Model for the origin of clusters

Domain hypothesis. When comparing with cluster size distributions, we assume that each receptor cluster consists of a group of receptors localized in a single microdomain. Domains of the same type have identical properties. The distribution of the observed cluster sizes should then be consistent with placing a number of receptors randomly into these preexisting domains. We allow for up to 2 types of domains, and also for free particles, which are placed randomly outside domains, into the rest of the cell membrane.

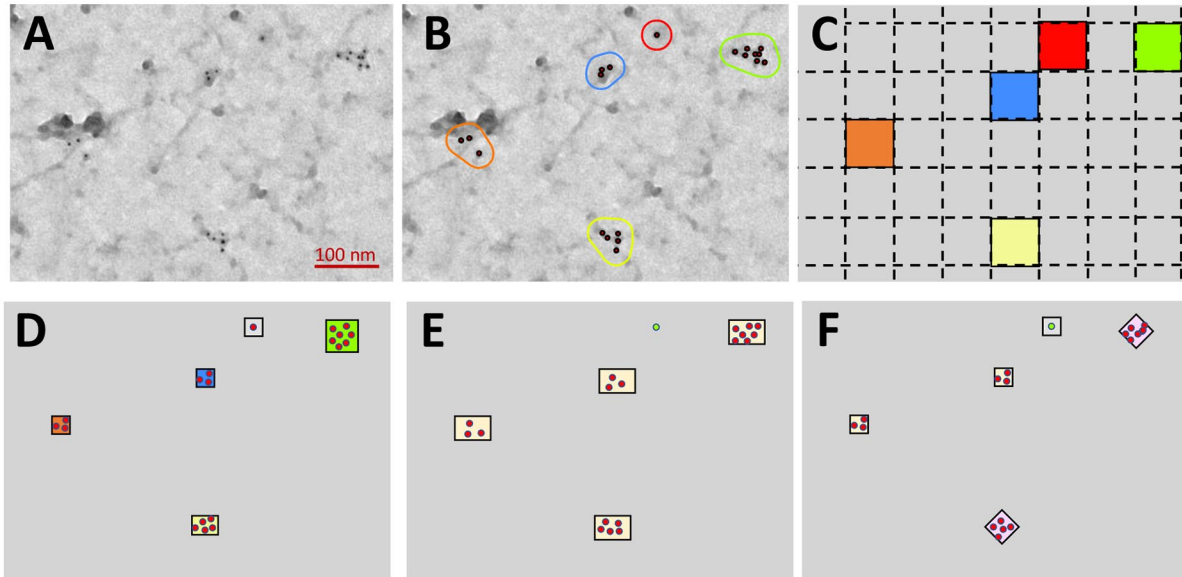


Figure 2. Model for cluster size distributions. (A) Transmission electron micrograph of a cell membrane patch (courtesy of the Wilson Lab). Dark dots correspond to receptors labeled with metal beads. Shaded areas (indicating the ability to absorb electrons) are often associated with receptors. (B) The original image is overlaid with symbols indicating the coordinates of receptors and contours of cluster “footprints” as identified by our algorithm. We hypothesize that receptors preferentially localize in a small area similar to these darker shaded regions, presumably due to their specific physical properties. (C) A simple model subdivides the cell membrane into colored “high-affinity” boxes corresponding to attractive domains and uncolored “low-affinity” boxes of the same size. Particles are placed randomly in these 2 types of boxes, and we identify the particles gathered together in a box as a “cluster.” (D) The small physical footprint of the observed clusters is consistent with a typical diameter of ≈ 20 to 60 nm; the probability of multiple particles placing randomly into a patch of area $1/1000$ to $1/6000$ of the total image area is very small. We instead defined models with one (E) or possibly 2 types of attractive domain (F), while also allowing for free particles which appear as singletons.

Domain number, area and density enrichment factors estimated from cluster footprints

In a first approximation, we identify the area occupied by the clusters of at least 2 particles (receptors) with the putative domains. Denote the corresponding footprint area $A_C^{(2+)}$, number of clusters, $N_C^{(2+)}$ and particles $N_P^{(2+)}$; the number of singleton particles is $N_S = N_P - N_P^{(2+)}$.

The empirical estimates for the area fraction $f_A^{(exp)} = A_C^{(2+)}/A_{total}$, number of domains $N_D^{(exp)} = N_C^{(2+)}$, and population fraction $f_P^{(exp)} = N_P^{(2+)}/N_P$ can be used to estimate the “attractiveness” or enrichment factor for the domains, as the ratio between the density of particles inside domains versus in the rest of the observed area:

$$\alpha^{(exp)} = \frac{N_P^{(2+)}}{A_C^{(2+)}} \cdot \left(\frac{N_S}{A_{total} - A_C^{(2+)}} \right)^{-1} = \frac{f_P^{(exp)}}{f_A^{(exp)}} \cdot \left(\frac{1 - f_P^{(exp)}}{1 - f_A^{(exp)}} \right)^{-1} \quad (1)$$

Stochastic models; distributions by particle and by domain/box. Confinement of particles is likely a stochastic process, where each of the N_D identical domains (sometimes referred to as “boxes”) receives a number of the N_P particles. Some domains will be empty, and others will have a single particle. Particles in single occupancy domains are indistinguishable from free particles.

A statistical model placing the N_{PD} confined particles into N_D boxes will predict the expected number of domains by

occupancy k based on the total number of domains N_D and confined particles N_{PD} , as $\langle n_{box}^{(k)} \rangle = N_D \cdot \text{PMF}_{box}(k)$ based on the per-box probability mass function (PMF). In addition to the occupancy k , this will depend on the number of domains N_D and the total number of confined particles, N_{PD} , and possibly additional model parameters. The corresponding per-particle PMF provides the distribution of particles by clusters of occupancy k ; the 2 PMFs are related by the expected number of particles by domain of occupancy k , $\langle n_{particle}^{(k)} \rangle = k \cdot \langle n_{box}^{(k)} \rangle$.

$$\text{PMF}_{particle}(k) = \frac{k \cdot N_D}{N_{PD}} \cdot \text{PMF}_{box}^{(k)}, \quad k = 1, 2, \dots, N_{PD} \quad (2)$$

In particular, the number of empty domains is $n_{empty} = \langle n_{box}^{(0)} \rangle = N_D \cdot \text{PMF}_{box}^{(0)}$ (no simple relation to the particle PMF) and the number of domains with a single particle is the same as the number of confined particles in single occupancy domains, $n_{singles} = \langle n_{box}^{(1)} \rangle$.

The relation to experimentally observable quantities, number of single particles (confined or not), and number of clusters of size 2 or larger and the number of particles in them is

$$N_S = N_P - N_{PD} + N_D \cdot \text{PMF}_{box}^{(1)} \quad (3)$$

Model with one type of domain. The core idea is that a number N_P of particles are placed randomly into identical N_D

domains (sometimes referred to as “boxes”). From the perspective of one domain, each particle may fall into the domain with a probability $p_{\text{bino}} = 1/N_D$. The probability that the domain receives exactly k particles is binomial PMF with N_P drawings and success probability p_{bino} :

$$\begin{aligned} P_k^{(\text{box})} &= p_{\text{bino}}^k (1 - p_{\text{bino}})^{N_P - k} \frac{N_P!}{k!(N_P - k)!} \\ &= \text{binopdf}(k; N_P, p_{\text{bino}}) \end{aligned} \quad (4)$$

As we only have access to clusters, which we identify with non-empty domains, it is practical to look at the distribution of particles by “cluster size,” i.e., the number of total particles in a domain. The probability that one specific particle falls into a box with $k-1$ other particles is the same as the PMF for boxes that contain exactly $k-1$ of the other $N_P - 1$ particles. Our main observable is the number of particles by box size $\langle n_{\text{particle}}^{(k)} \rangle$.

$$\begin{aligned} P_k^{(\text{particle})} &= \text{binopdf}(k-1; N_P - 1, p_{\text{bino}}) \\ \langle n_{\text{particle}}^{(k)} \rangle &= N_P \cdot \text{binopdf}(k-1; N_P - 1, p_{\text{bino}}) \end{aligned} \quad (5)$$

One type of domain plus singletons. A plausible explanation for the accumulation of receptors in domains is a mechanism characterized by an “enrichment factor” α that results in a proportionally higher average particle density inside domains compared with the rest of the membrane area. The particles in a portion of the membrane of area A_{total} will be divided among domains and the rest of the image. Denote the aggregate area of all domains $A_D = \alpha N_D$, where α is the average area of one domain, and the rest by $A_{\text{free}} = A_{\text{total}} - A_D$. The number of particles in each sector will be

$$\begin{aligned} N_{PD} &= N_P \frac{\alpha A_D}{A_{\text{total}} + (\alpha - 1) A_D} \equiv (1 - f_S) N_P, \\ N_{P,\text{free}} &= N_P \frac{A_{\text{free}}}{A_{\text{total}} + (\alpha - 1) A_D} \equiv f_S N_P \end{aligned} \quad (6)$$

Singletons (particles that are in “clusters” of size 1) are experimentally indistinguishable from free particles outside domains. The observed number of singles is then $N_S = N_{P,\text{free}} + n_{C,\text{singles}}$:

$$N_S = N_P \left(1 - f_P \left(1 - \left(1 - \frac{1}{N_D} \right)^{N_{PD} - 1} \right) \right) \quad (7)$$

The number of particles by cluster size for clusters of 2 or more remains as in the previous case:

$$\langle n_k^{(\text{particle})} \rangle = N_{PD} \cdot \text{binopdf} \left(k-1; N_{PD} - 1, \frac{1}{N_D} \right), k \geq 2 \quad (8)$$

Two types of domain plus singletons. Some of the distributions observed in the image set are not well fit in a model that had one type of domain. We could have multiple types of domains, resulting from different physical mechanisms. For 2 types of domain, we assume that a fraction f_S of all particles (so $f_S N_P$ “free” particles) are outside domains and (always) appear as singles outside clusters. The remaining particles are split among 2 types of domains with lower and higher affinity, labeled L and H ; denote by f the proportion of particles in the high-affinity domains. The particles in each type of domain N_{PH}, N_{PL} are distributed consistent with a binomial distribution with success probability p_H and p_L , respectively; N_{DL}, N_{DH} represent the number of each type of domains (in the image):

$$\begin{aligned} N_{PH} &= (1 - f_S) f N_P, & N_{DH} &= \frac{1}{p_H} \\ N_{PL} &= (1 - f_S) (1 - f) N_P, & N_{DL} &= \frac{1}{p_L} \end{aligned} \quad (9)$$

The number of singles is the sum of the free particles and the single particle domains of both types:

$$N_S = f_S N_P + N_{PH} (1 - p_H)^{N_{PH} - 1} + N_{PL} (1 - p_L)^{N_{PL} - 1} \quad (10)$$

The number of particles in domains with exactly k particles ($k \geq 2$) work out to

$$\begin{aligned} \langle n_{\text{particle}}^{(k)} \rangle &= N_{PH} \cdot \text{binopdf}(k-1; N_{PH} - 1, p_H) \\ &\quad + N_{PL} \cdot \text{binopdf}(k-1; N_{PL} - 1, p_L) \end{aligned} \quad (11)$$

Model fitting

For a given image and a model type, we seek the model parameters that predict the cluster size distribution most closely. We compare the observed number of particles $N_P^{(k)}$ in clusters of size $k = 1, 2, \dots$ to the corresponding expectation predicted by the model with the respective parameters, $\langle n_{\text{particle}}^{(k)} \rangle$. For larger cluster sizes, especially in images with a moderate number of particles, the model prediction may result in a significant probability for having one or a handful of large clusters. Thus, the expected number of particles in each specific cluster size between 10 and 20 might be on the order of 3, corresponding to 40 to 50 particles total in three to four clusters of size somewhere between 10 and 20. To account for this aspect, compare the number of particles in groups of cluster sizes as follows:

$$\{s_\ell\}_{\ell=1 \dots 6} = \{1, 2, (3 \dots 5), (6 \dots 10), (11 \dots 20), \geq 21\} \quad (12)$$

We construct a normalized square distance

$$\Delta_{\text{image}}^{(j)}(\vec{f}, \vec{p}) = \frac{1}{N_P} \sum_{\ell} \left(q_{\ell}^{(\text{exp})} - q_{\ell}^{(\text{model})} \right)^2$$

$$q_\ell^{(\text{exp})} \equiv \sum_{k \in s_\ell} N_p^{(k)}, q_\ell^{(\text{model})} \equiv \sum_{k \in s_\ell} \langle n_{\text{particle}}^{(k)} \rangle \quad (13)$$

where j refers to a specific image and \vec{f}, \vec{p} represent the model parameters (population fractions and success probabilities) used to compute the expected numbers of particles $\langle n_{\text{particle}}^{(k)} \rangle$ in clusters of each size. The above is different from the usual χ^2 test in that the squared differences are not individually normalized to the expectations. Our measure can be interpreted as a “mean” relative square distance where the relative differences $(O_\ell - E_\ell)^2 / E_\ell = (q_\ell^{(\text{exp})} - q_\ell^{(\text{model})})^2 / q_\ell^{(\text{model})}$ are each weighted with the relative fraction

$$p_\ell = \frac{q_\ell^{(\text{model})}}{N_p}$$

The expectation and variance of a Poisson integer random variable $p(k) = \Gamma^k e^{-\Gamma} / k!$ are equal to $\langle k \rangle = \langle (k - \Gamma)^2 \rangle = \Gamma$, therefore the standard deviation is $\sigma = \langle k \rangle^{1/2}$. If this were the case for the particle numbers by cluster size, the expectation of deviation from the model $(q_\ell^{(\text{exp})} - q_\ell^{(\text{model})})^2$ would be equal to $q_\ell^{(\text{model})}$, and therefore, we should expect $(q_\ell^{(\text{exp})} - q_\ell^{(\text{model})})^2 / q_\ell^{(\text{model})} = O(1)$. However, a significant number of images have big relative differences between the number of particles in each size group, and we used the simple sum of squares normalized to the total number of particles. The measure used here should be on the order of 1 for a good model fit.

Individual and group fit. We used a Markov chain Monte Carlo (MCMC) approach to identify best-fit model parameters for each image. The actual computation is our implementation of the Metropolis-Hastings algorithm²¹ seeking to minimize the distance $\Delta_{\text{image}}^{(j)}$ over possible values of the parameter set \vec{f}, \vec{p} . We performed individual fits for every image, using the 1-domain and 2-domain models (with singles). Group fits for multiple images were performed in a similar manner, but the optimization was aimed at minimizing the sum of square distances evaluated for each image in the group.

Results

The dataset discussed here was derived from 81 high-resolution (2500 \times) micrographs (TEM images) of nanogold-labeled VEGF receptors on PAE-KDR cells. Labeled receptors appear as points (dark spots), whose coordinates are extracted for this analysis. Details on sample preparation and imaging technique are provided in section “Methods and Models.” Results on NND distributions and cluster identification were discussed in the work by Güven et al¹³ and are the starting point for the cluster size distribution analysis we are interested in here.

Clustering

Quantifying clustering tendency. One striking feature of the spatial distribution pattern of receptors is their tendency to

accumulate in small groups or clusters. Analysis using standard measures, such as NND distributions and Ripley and Hopkins statistics revealed a compelling difference between the observed distributions and random placement.¹³ The points and their assignment to clusters are overlaid on the micrograph in Figure 1D. The NND distributions in Figure 1A and B and the Hopkins statistic (Figure 1C) are compared with the theoretical distribution expected from random placement. The situation illustrated in Figure 1A and B for one micrograph is typical. We computed these measures for the entire dataset and found that the bulk of the experimental distributions fell well below the expected distance, but closely approximated the theoretical curve corresponding to random placement of the same number of points in a smaller area.

The NND-based analysis was first discussed in the work by Güven et al.¹³ Plots similar to those in Figure 1 for the entire data set are provided as supplemental data in the accompanying Github repository.

Cluster identification. We identify clusters of points using a version of distance-based hierarchical clustering. Clusters in a set of points are induced by a parameter L , so that any 2 points whose mutual distance is below L are assigned to the same cluster. As the parameter increases from zero to the diameter of the image, the induced number of clusters $N_C(L)$ decreases from N_{particle} to 1.

The key element is identifying an optimal length parameter L^* for the points found in one image. We rely on the features of the observed $N_C(L)$ curve (Figures 1F and G) and its comparison with the random uniform placement of the same number of particles. The observed curve exhibits 2 distinct sections. Similar to the NND distribution, $N_C(L)$ for small L values is consistent with the random placement of N_p particles in an area $A' < A_{\text{image}}$; for larger values of L , the curve is consistent with random uniform placement of a smaller number of points $N'_p < N_p$ in the total image area A_{image} .

This can be understood as the separation between 2 length scales, the typical distance between neighboring points within a cluster λ_1 vs the typical distance between clusters λ_2 . When the separation is complete, $\lambda_1 < \lambda_2$, and, $L \in [\lambda_1, \lambda_2]$ will identify the same clusters and curve will exhibit a plateau. This provides a guiding principle to identify L^* from experimentally derived cluster scaling functions, as the value of L corresponding to the maximum upward curvature (second inflection point) of the $N_C(L)$ curve (illustrated in Figure 1F and G). We applied the cluster identification procedure to the entire dataset. The characteristic length (Figure 3) does not exhibit a correlation with the number of particles in an image or with the size of the resulting clusters. We will discuss the cluster size distributions in the details below. The results of cluster identification are also provided in the Github repository.

Clusters as domains. All of our results so far, and current understanding of the biochemistry and dynamics of receptor

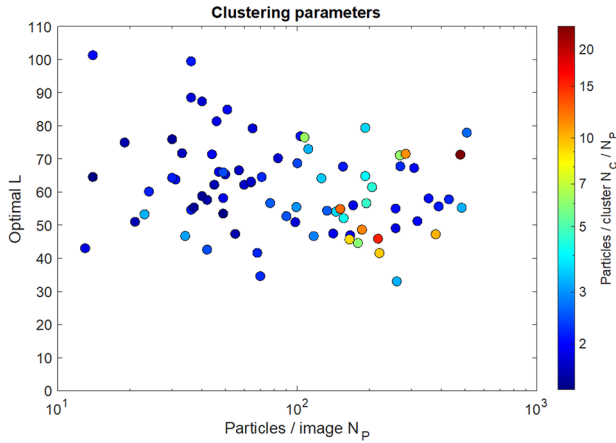


Figure 3. The characteristic length parameter L does not exhibit a clear dependence on the number of particles per image or on the average size of clusters.

movement on the cell membrane, point toward a hypothesis of clustering induced by the features of the membrane landscape. We use the term “domain” to refer to specific regions on the membrane that have an increased affinity for receptors. We assume that receptors can move in and out of these regions, but on average, the domains are “enriched” and have a consistently higher density of receptors than the rest of the membrane.

While we cannot reliably identify the underlying physical regions directly from the micrographs, we identify domains with observed clusters that have at least 2 particles. We assigned a geometric shape to each cluster, using a procedure^{12,13} that relates closely to distance-based clustering. This allows us to calculate the areas for each domain/cluster we identify.

We used the geometric measurements of cluster footprints to estimate an enrichment factor, defined as the ratio of the density corresponding to the particles in clusters of size 2 and higher over the average density of particles (Figure 4) in the entire image equation (1). The values range more than 2 orders of magnitude from 30 to 3000, with typical values concentrated between 100 and 1000. There is only a slight trend of increased enrichment with higher average cluster sizes and a larger fraction of particles in clusters.

Cluster size distributions

We would like to characterize the distribution of cluster sizes using a small number of parameters that could be used to compare the distributions between images with different particle numbers. In the same time, we explore the extent to which clustering could be explained, at least in part, by our preexisting domain hypothesis.

We devised a type of model that treats the accumulation of receptors (particles) in preexisting attractive microdomains as a random process, similar to placing items in identical “boxes.” We explored models with 1 and 2 types of domain, which are identified with observed clusters; in addition, both models

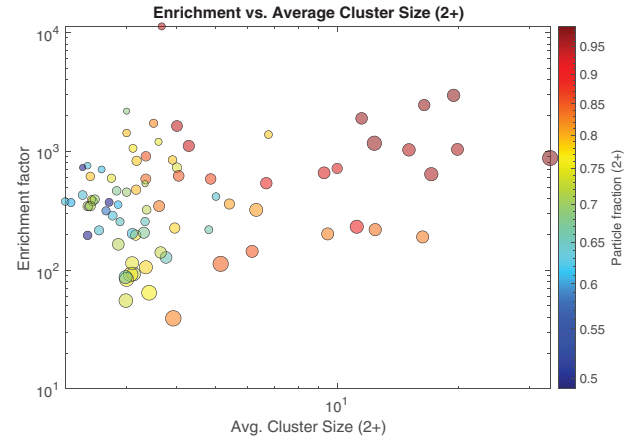


Figure 4. Enrichment factors derived directly from cluster footprints. The scatter plot illustrates the correlation between the enrichment factor, the average cluster size in each image, and the fraction of particles (in each image) that is in clusters of size 2 or higher. Circles correspond to individual images, and their size is proportional to the number of particles. Color changes by the cluster size.

allow for a fraction of free particles, which are identified with singletons (particles that are not part of a cluster). We refer to the free particles, and the domains (possibly 2 different types) as sectors. Model parameters are: fractions of the particles that fall into each sector $f_s = (N_{free} / N_P), [f = (N_{PH} / N_{PD})]$, and a binomial probability parameter for each type of domain, p_H, p_L . See section “Methods and Models” for more details.

Individual fit of cluster size distributions. We focus on the distribution of the number of particles by cluster size. Cluster sizes are identified with the number of particles in a domain. A model provides the expectation of the number of particles in clusters of size k , $\langle n_{particle}^{(k)} \rangle$.

As clusters may be relatively large, sizes are grouped into 6 ranges:

$\{s_\ell\}_{\ell=1 \dots 6} = \{1, 2, (3 \dots 5), (6 \dots 10), (11 \dots 20), \geq 21\}$ and we use the aggregate number of particles q_ℓ in each size range. For each image, we performed an MCMC fit (Metropolis-Hastings) aimed at minimizing the square distance $\Delta_{image}^{(j)}$ for each image $j = 1, 2, \dots$

$$\Delta_{image}^{(j)}(\vec{f}, \vec{p}) = \frac{1}{N_P} \sum_{\ell} \left(q_{\ell}^{(exp)} - q_{\ell}^{(model)} \right)^2$$

$$q_{\ell}^{(exp)} \equiv \sum_{k \in s_{\ell}} N_P^{(k)}, \quad q_{\ell}^{(model)} \equiv \sum_{k \in s_{\ell}} \langle n_{particle}^{(k)} \rangle \quad (14)$$

Here we report the estimated characteristics of the domains based on the results of model fitting. In general, a model is defined by a set of population fractions that determine the number of particles in each sector, free particles and particles in each type of domain. If there is only one type of domain, the N_P total particles are divided into $f_s N_P$ free particles and $N_{PD} = (1 - f_s) N_P$ particles in domains. With 2-domain types,

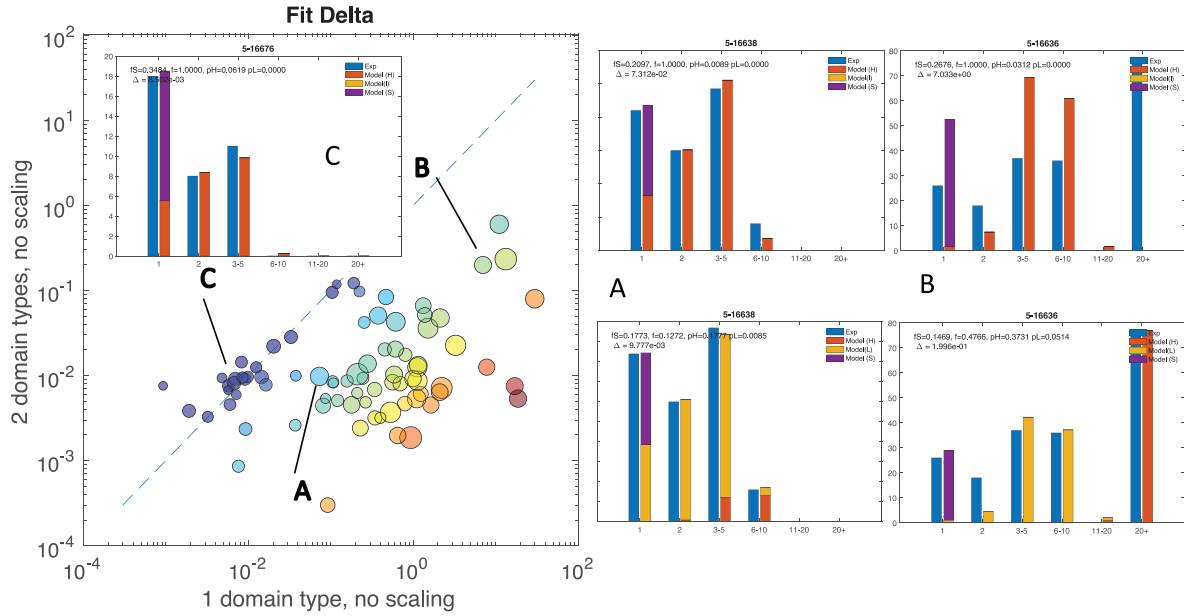


Figure 5. Square distances (equation (14)) from fitting each image to a model with one, respectively, 2-domain types (left). Circle sizes are proportional to the number of particles; shading represents the ratio of the fitted square distances obtained with 1-domain respectively 2-domain models. Some of the images are well fit by one domain (C), others are significantly better fit with 2-domain types (A or B). The histograms labeled A, B, C indicate the distribution of the number of points by cluster size for the respective images. Bars on the left represent observed counts, bars on the right are for the closest model fits. Only the 1-domain model fit is shown for image C; for images A and B, the top bar plots are with 1-domain type and the bottom ones with 2-domain types.

we have $N_{PH} = f N_{PD}$ particles in high-affinity domains and $N_{PL} = (1-f)N_{PD}$ particles in low-affinity domains. The free particles are assumed to be all singles (not part of clusters).

Consider one type of domain (there might be 1- or 2-domain types in the model). If a total number of particles N_{PD} are placed into N_D domains, the probability of one specific domain receiving exactly k particles is given by a binomial distribution

$$\text{binopdf}(k; N_P, p_{\text{bino}}) = \binom{N_{PD}}{k} (p_{\text{bino}})^k \quad (15)$$

The binomial probability p_{bino} is simply the inverse of the number of available domains N_D . We define an *overall success probability per domain* p_{succ} as the probability that one particle placed in the image will fall into one specific domain of *this type*. The binomial probabilities for each domain type and the corresponding particle fractions are our model parameters.

The relative attractiveness of the domains α is similar to the enrichment factor relevant to signaling dynamics. We define it as the ratio between the overall success probability for a domain type and what that same probability would be based only on the area corresponding to one domain

$$p_{\text{bino}} = \frac{1}{N_D}, p_{\text{succ}} = \frac{N_{PD}}{N_P} p_{\text{bino}} \quad (16)$$

$$\alpha = \frac{p_{\text{succ}}}{p_{\text{geom}}} = \left(\frac{N_{PD}}{N_P} p_{\text{bino}} \right) / \left(\frac{A_{\text{domain}}}{A_{\text{image}}} \right)$$

Here p_{geom} can be defined as the probability of geometric measurements of cluster footprints to estimate an enrichment factor.

We performed individual fits for each image, to models with 1-domain type and models with 2 different types of domains. In the latter case, we refer to the 2-domain types as lower and higher affinity, or type L , respectively H .

The resulting minimal square distances for each image are compared in Figure 5 and indicate that while some images are well fit by the 1-domain model (C), a large fraction of the images are better (A) (or significantly better—B) fit with the 2-domain model. The circles in Figure 5 (left) are shaded according to the ratio of the fitted square distance obtained with 1-domain, respectively, 2-domain types. The predicted distribution of particles by domain/cluster size is compared with the observed distribution for the 3 sample images. Similar plots are provided for all the images in the Github repository.

The relative attractiveness α (equation (16)) and equivalent average number of domains per image derived by model fitting are shown in Figure 6. For images well fit with a single domain type (left panel), the number of domains N_D per image ranges from 2 to 100 and the relative attractiveness ranges between 10 and 400. The parameters for 2-domain fits are shown in the center and right panels. The number of both high- and low-affinity domains ranges from 10 to 30 with a relative attractiveness of 10 to 800, respectively 1 to 100.

Fitting groups of images to a single model. The estimated model parameters vary over a wide range. This reflects the variability of the input data (images) but is partially due to the fact that our distance function does not always provide a sharp optimum for the model parameters, so multiple parameters sets could provide comparably good fits for the same image.

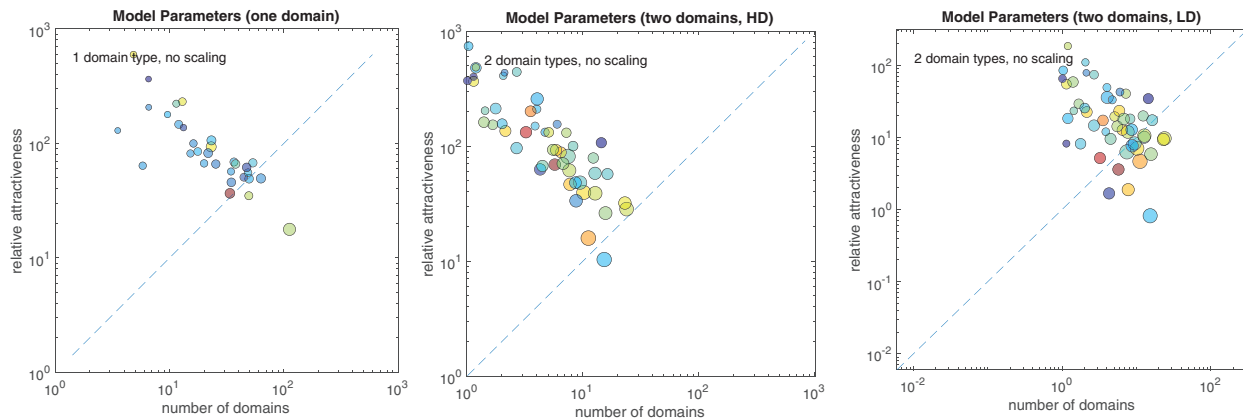


Figure 6. Individually fitted model parameters obtained using the 1-domain and 2-domain model. Each point corresponds to one image. One-domain parameters (left panel) are shown for images that are not fit significantly better by the 2-domain model. Two-domain parameters are shown (high affinity, center; low affinity, right) only for those images that are significantly better matched by the 2-domain type model.

We can leverage this property by searching for parameter sets that are optimized simultaneously for a set of images. Briefly, we perform the same MCMC fitting process, but aimed at minimizing the sum of square distances for a subset \mathcal{G} of the images,

$$\Delta_{\text{group}}^{(m)} = \sum_{j \in \mathcal{G}_m} \Delta_{\text{image}}^{(j)}, \mathcal{G}_m \subset \{1, 2, \dots, N_{\text{images}}\} \quad (17)$$

Given a set of images, the fitting process is identical to that of individual fitting, but using the group distance (equation (17)). Attempts to fit the entire set of images to a single model failed to achieve reasonably low distances. This led us to adopt a group approach where the set of 81 images was partitioned into groups (or clusters—not to be confused with clusters of receptors). The complete fitting process is similar to k -means clustering, in that we begin with a given partition of the image set into k groups and perform repeated parameter optimization passes, followed by reassignment of images to the closest group (centroid).

In Figure 7 and Table 1, we present results with $k = 6$ groups of images fitted with models with 2-domain types and parameters. The scatter plots show the overall success probabilities $\hat{p}_{\text{succ}}^{(H)}$, $\hat{p}_{\text{succ}}^{(L)}$, and expected number of domains per image N_{DH} , N_{DL} for the 2-domain types. The relation to the primary model parameters $f_s, f, \hat{p}_H, \hat{p}_L$ is as follows:

$$\begin{aligned} \hat{p}_{\text{succ},H} &= (1 - f_s) f \hat{p}_H, \hat{p}_{\text{succ},L} = (1 - f_s)(1 - f) \hat{p}_L, \\ N_{DH} &= \frac{1}{\hat{p}_H}, N_{DL} = \frac{1}{\hat{p}_L}. \end{aligned} \quad (18)$$

The initial partition was obtained by actual k -means clustering applied to feature vectors that contain various characteristics of the images.

Figures 8 to 10 illustrate the results of group fits to one image from each group. The bar plots compare the actual number of particles by cluster size with individual model fits and group fits. Similar plots are available for all images in our data set.

Discussion

Receptor clustering, the accumulation of membrane-bound receptors in groups ranging from 2 to 3 to hundreds of copies, is commonly observed for multiple receptor types. This type of clustering is not the result of a known receptor-receptor binding mechanism and is likely the result of the interaction of receptors with features of the cell membrane. A possibly related phenomenon is co-confinement of receptors, observed in single-particle tracking. One plausible explanation is the presence of small regions that have a higher affinity to the receptors, so that while receptors can move in and out of these domains, their random movement is biased in a way that results in a higher probability of placement inside the domain.

The implications of the presence of high-affinity or receptor-enriched domains on the dynamics of signaling are significant. In pathways where signal initiation relies on ligand-induced oligomerization of receptors, enrichment results in a higher rate of dimerization and more effective signaling. This is the case for VEGF signaling, which relies on the formation of receptor dimers stabilized by the bivalent VEGF ligand.

Using algorithms to identify receptor clusters and to assign a footprint area to the clusters, we aim to develop a methodology to quantify the phenomenon of receptor clustering and to investigate the hypothesis of underlying high-affinity domains. The main goal of the work reported here is to construct and parameterize a simple model that can explain the observed cluster size distributions using a small set of general parameters pertaining to the underlying domains. The data set used here, kindly provided by the Wilson Lab (University of New Mexico) consists of TEM micrographs of nanogold-labeled VEGFR-2 receptors on PAE-KDR cells that artificially express these receptors. Receptors are identified as small dark spots; their coordinates were extracted and used to compile a list of points for each image.

In previous work analyzing this data set,¹³ we used hierarchical distance-based clustering to identify clusters, supported by NND-based measures to identify the optimal length scale and establish clustering tendency. We found that the distributions

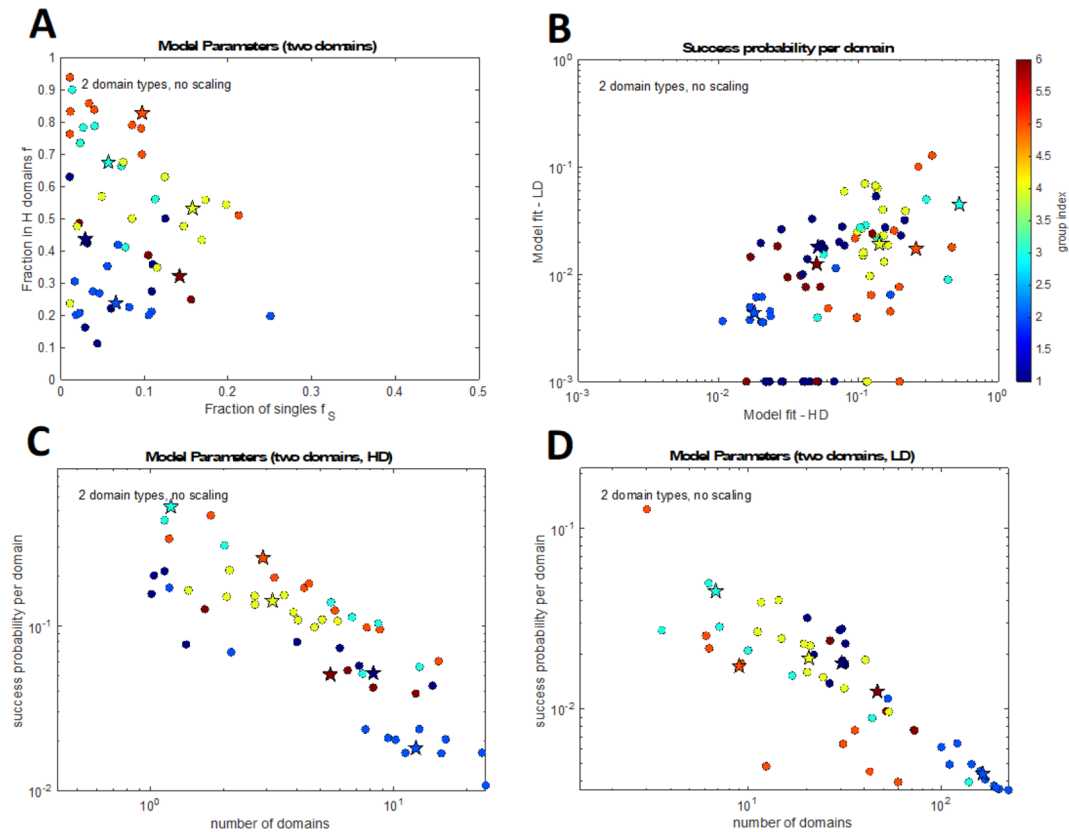


Figure 7. Summary of group fitting results. Parameters for the 2-domain model (f_S, f, p_H, p_L) were optimized for groups of images (see text for details). Stars represent group centroids and circles represent individual images; symbols are colored to reflect group assignments. The overall success probability $p_{succ,H}$, $p_{succ,L}$ for domains of type H in panel A, respectively L , are compared with each other in panel B and plotted against the number of domains of the respective type in C and D.

Table 1. Fitted model (mod.) parameters after we performed k -means clustering for the unknown model parameters for each data set.

GR.	F_S (%)		$P_{succ,H}$ (%)		$P_{succ,L}$ (%)		N_{DH}		N_{DL}	
	MOD.	OBS.	MOD.	OBS.	MOD.	OBS.	MOD.	OBS.	MOD.	OBS.
1	2.91	6.04	5.15	6.64	1.80	1.42	8.23	16.47	30.47	1210.8
2	6.60	6.71	1.80	3.26	0.44	0.47	12.31	12.43	162.60	155.4
3	5.68	4.82	52.55	19.20	4.49	2.51	1.21	6.13	6.83	14.6
4	15.74	13.67	14.13	13.09	1.91	3.00	3.17	3.44	20.63	263.8
5	9.73	8.95	25.80	21.28	1.73	3.16	2.90	4.17	9.02	139.1
6	14.21	8.93	5.04	4.47	1.25	0.93	5.48	17.95	46.53	841.91

Here, we chose 6 groups. The results for the observed (obs.) biological data are for the cluster centroids.

of points within clusters were consistent with uniform random placement at a higher density. This observation is consistent with the hypothesis of underlying high-affinity domains.

Our goal here is to take the domain hypothesis further by attempting to quantify the properties of the underlying domain structure and devise a simple model to predict/reproduce the observed distribution of particles in clusters. The main idea is that, if the observed clusters are the result of random placement of particles into a set of preexisting domains, then the observed

distribution of cluster sizes should be determined by the individual size (area), density (of domains in the cell membrane), and enrichment factor of these domains.

Small cluster footprints point to high enrichment factors

First, we used the method developed by Pryor et al¹⁶ to construct footprints associated with individual clusters. In a first

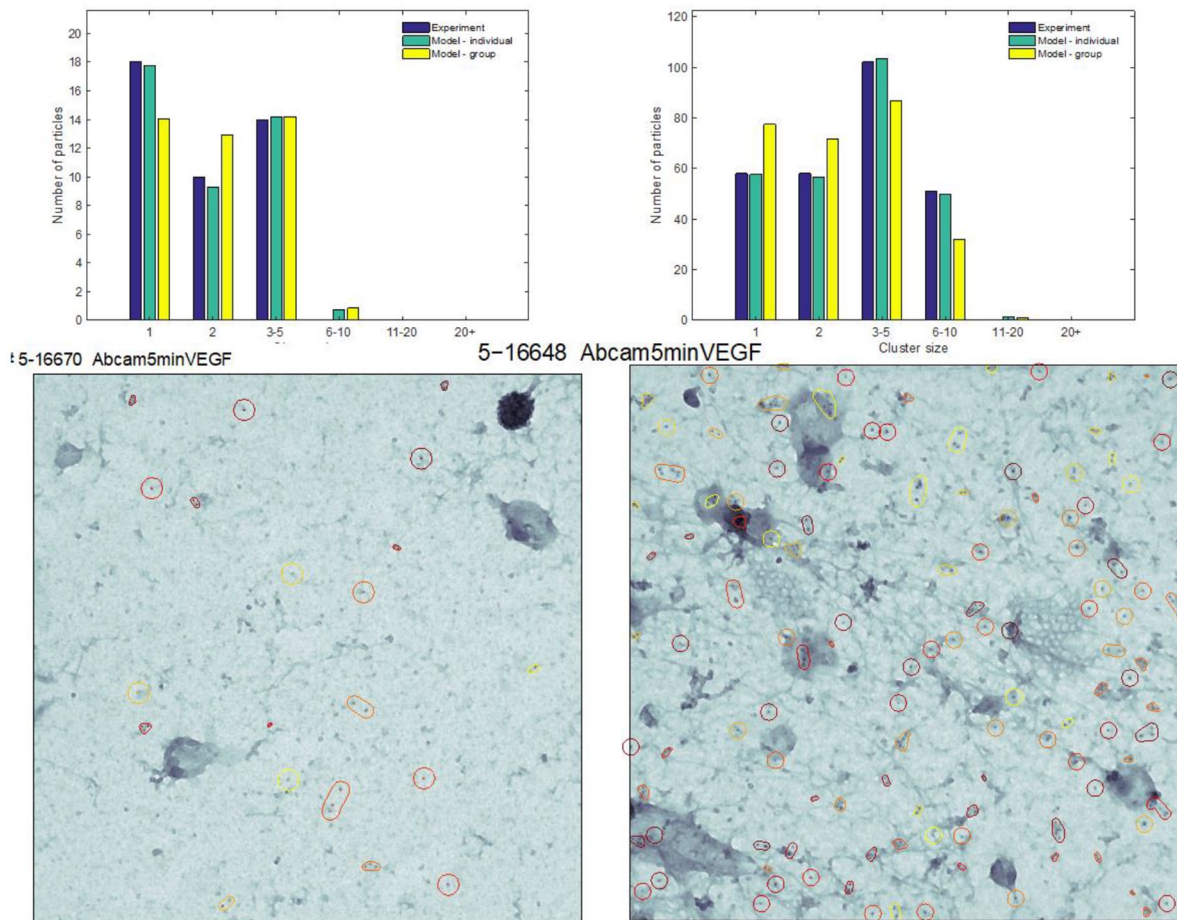


Figure 8. Example of individual and group fits for groups 1 and 2. Micrographs provided by the courtesy of the Wilson Lab. Colored points indicate receptors; contours identify clusters and their footprint. Bar plots show the number of particles by cluster size, comparing the actual values and the model predictions using the best-fit parameters. The individual fits are always closer but the group fits are also reasonable.

approximation, one may identify clusters of size 2 or larger with the presumed domains. A measure of the accumulation of receptors is the number of clustered receptors as a fraction of the total compared with the area occupied by clusters as a fraction of the total area of an image (Figure 4), that is, the relative density of receptors in clusters versus overall. The resulting enrichment factors are typically between 100...1000. Individual domains range from $\frac{1}{2500} \dots \frac{1}{160}$ of the entire image area (equivalent to $\frac{1}{50} \dots \frac{1}{12.5}$ of the diameter).

Importantly, the enrichment factor is consistently high across the data set, including many images with a large average cluster size. This indicates that the enrichment effect is not primarily due to dimers (which would appear as clusters of 2 points, with a small footprint).

In summary, a model for the emergence of clusters should explain why more than 50% of the observed particles tend to collect in a small fraction of the total area, with densities of 1 to 3 orders of magnitude higher than the average, in individual clusters numbering from 2 to 200 particles. The small physical size (typically 0.025% of the image area) of individual clusters precludes an explanation based on “accidental” proximity.

Domain-based models for cluster size distributions

We developed simple models to explain the observed cluster sizes, based on the random placement of particles into the area of interest (corresponding to one image) in the presence of a set of domains; each particle places either in one of the domains or in the rest of the area. For a given drawing (placing all the particles), the particles in one domain correspond to an observed cluster of the corresponding size; particles outside domains are identified with singletons (or clusters of size 1). For a given image with N_p particles, we compare the expected distributions of particles into clusters of different sizes with the observed distribution.

We considered models with 1 or 2 types of domains. One type of domain is characterized by the total number of domains N_D (of this type in the entire image) and the expected fraction of particles that place in this type of domain. We then compute the expected number of particles outside domains (identified as singletons), and the expected number of particles in domains that contain 1, 2, etc, particles.

These expectations are then compared with the actual number of particles in clusters of the corresponding sizes. As the number of larger clusters was typically small, we used the total

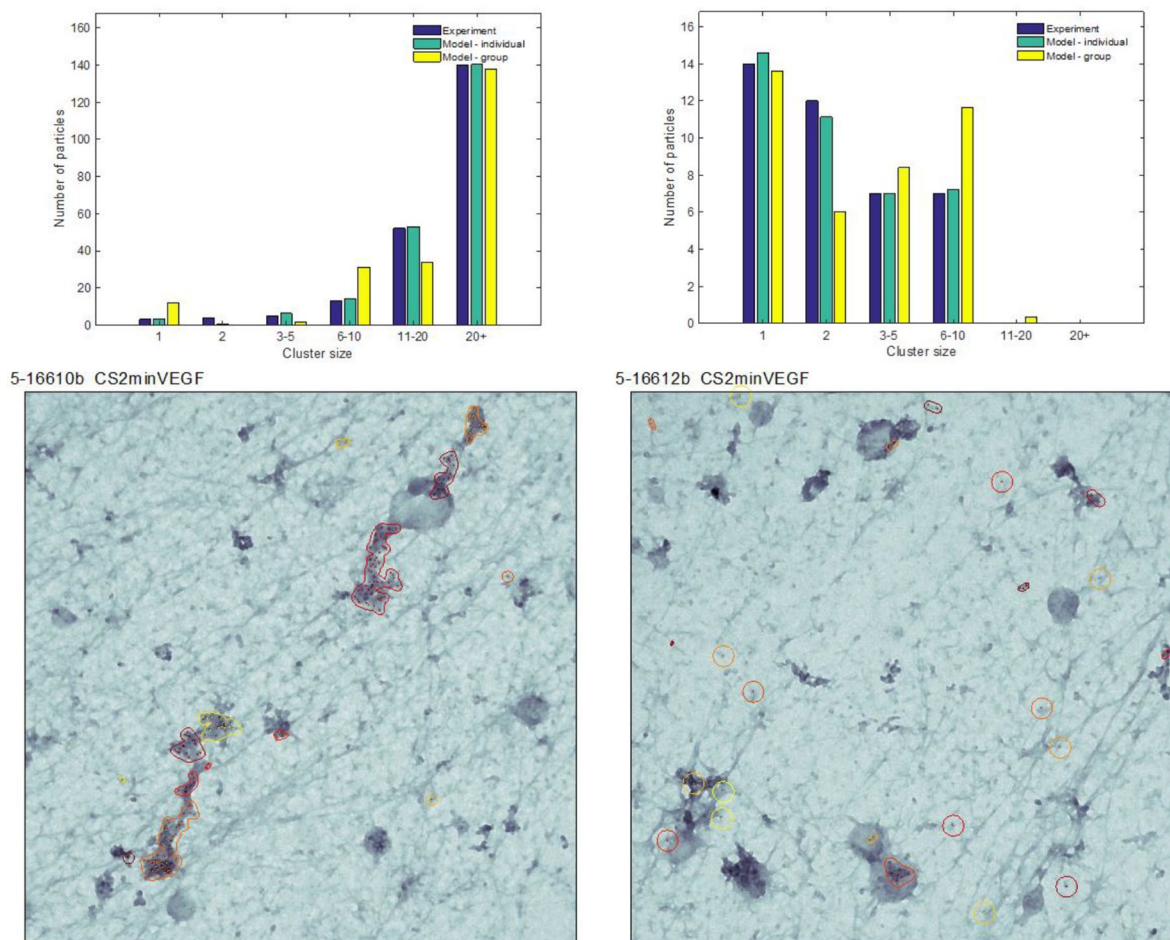


Figure 9. Example of individual and group fits for groups 3 and 4. Micrographs provided by the courtesy of the Wilson Lab. Colored points indicate receptors; contours identify clusters and their footprint. Bar plots show the number of particles by cluster size, comparing the actual values and the model predictions using the best-fit parameters. The individual fits are always closer but the group fits are also reasonable.

number of particles in clusters within ranges of sizes: 1, 2, 3-5, 6-10, 11-20, and higher than 20. We performed fits by minimizing the (square) difference between the actual and expected number of particles per cluster size interval.

Some of the images could be very closely matched by a model with only one type of domain, but a significant subset could not. We were able to fit closely to the remaining cluster size distributions after extending of the model to 2-domain types, each with their number of domains and assigned fraction of particles. The parameters inferred from individual model fits (with 1- or with 2-domain types) are distributed over wide ranges. To identify narrower parameter values, we attempted to identify groups of images with similar features, and then tried to fit all images in a group with a single set of parameters. We provided detailed results where the 81 images were partitioned into 6 groups. We briefly comment on these results below.

Individual model fits of cluster size distributions

We performed individual fits for each image, using models with 1-domain type and models with 2 different types of domains. The fitting procedure aimed to minimize the minimal square

distance (equation (13)) for each image, over the model parameters, using a MCMC algorithm. The resulting minimal square distances for each image with a 1- and 2-domain model are compared in Figure 5. While some images are well fit by the 1-domain model, a large fraction of the images are better (or significantly better) fit with the 2-domain model. The model-predicted distribution of particles by domain/cluster size is compared with the observed distribution for the 3 sample images. The sample image 5-16636 (labeled B) is a good illustration of how a model with one type of domain can fail. The distribution of moderate size clusters (with 3-20 particles) can be reasonably well reproduced with one type of attractive domain; however, the fact that more than 70 particles are in domains >20 while there are none in domains between 11 and 20 is exceedingly unlikely in a single binomial distribution.

Fitting groups of images to a single model. The estimated model parameters vary over a wide range. This reflects the variability of the input data (images) but is partially due to the fact that our distance function does not always provide a sharp optimum for the model parameters, so multiple parameters sets could provide comparably good fits for the same image.

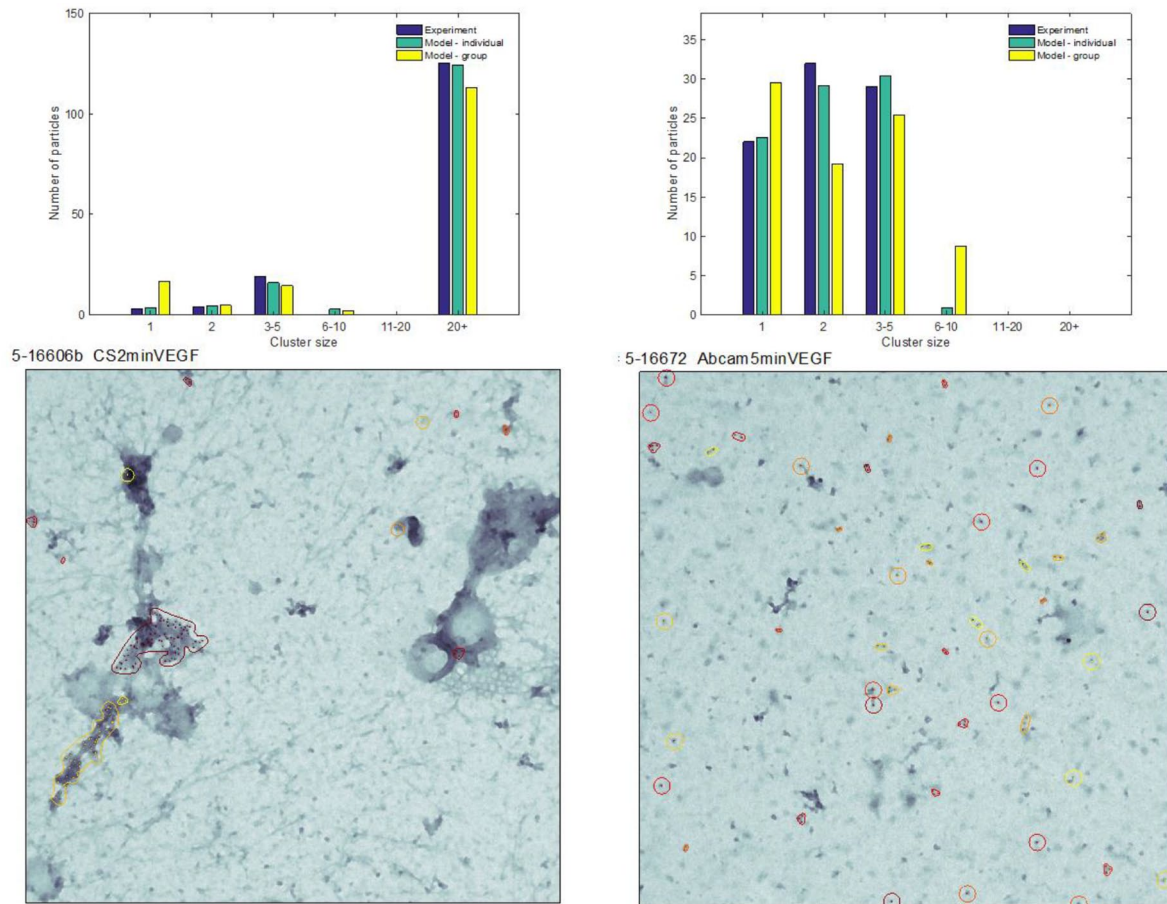


Figure 10. Example of individual and group fits for groups 5 and 6. Micrographs provided by the courtesy of the Wilson Lab. Colored points indicate receptors; contours identify clusters and their footprint. Bar plots show the number of particles by cluster size, comparing the actual values and the model predictions using the best-fit parameters. The individual fits are always closer but the group fits are also reasonable.

We leveraged this property by searching for parameters sets that are optimized simultaneously for a set of images. Briefly, we performed the previous MCMC fitting procedure, but using the same model/parameter set for *all images* in a group (which may be the entire data set or part of it). The procedure is otherwise identical to the single image fit. We seek to minimize the sum of all square distances (equation (14)), or the *group distance*.

Attempts to fit the entire set of images to a single model failed to achieve reasonably low distances. This led us to adopt a group approach where the set of 81 images was partitioned into groups. The process of identifying the groups was similar to k -means clustering, in that we began with a given partition of the image set into k groups and performed repeated parameter optimization passes, followed by re-assignment of images to the closest group (centroid). The initial partition was obtained by actual k -means clustering applied to feature vectors that contain various characteristics of the images.

The results of the group fitting process are illustrated in Figure 7 and the resulting parameter sets (the number of domains of each type and the corresponding individual success probabilities, equation (16)) are given in Table 1. For each of

the 6 groups, a representative image and the corresponding fit bar plots are shown in Figures 8 to 10.

Conclusions

We summarize our results as follows. Nearest-neighbor distance-based measures indicate that the observed clustering of VEGF receptors is significant. The receptors accumulate in the small regions of diameter from 50 to 200 nm containing from 2 to 3 up to a 100 receptors. We hypothesize that the observed clusters are the result of preferential random placement of receptors (enrichment) in preexisting areas of the membrane. We estimate the enrichment factor typically in the range of 100 to 1000.

To validate our hypothesis of domain-induced clustering, we devised a simple model to predict the distribution of cluster sizes. The key element is that particles are first partitioned among 2 or 3 sectors: free (not in a domain) and confined in domains of up to 2 types. Free particles correspond to singletons. Confined particles are distributed randomly among the domains of the respective type, thus one given domain will receive a number of particles given by a binomial distribution. We were able to fit all images with the 2-domain-type model, but only some of them with 1-domain type. The resulting

individually fitted parameters are distributed over a wide range of values. We were able to fit groups of images with one set of parameters (per group).

Acknowledgements

The authors thank Professor Bridget Wilson. The original set of micrographs was obtained in the Wilson Lab and Professor Wilson provided guidance and made contributions in developing the methods used here to identify clusters and their geometric footprint.

Author Contributions

AMH and JSE conceptualized the project; EG, AMH, and MJW conducted the research, performing the formal analysis, software development, validation, and visualization of the results; EG, AMH, and MJW wrote the article. All authors reviewed the article.

ORCID iD

Emine Güven  <https://orcid.org/0000-0001-9324-0879>

REFERENCES

- Subramanian N, Torabi-Parizi P, Gottschalk RA, Germain RN, Dutta B. Network representations of immune system complexity. *Wiley Interdiscip Rev Syst Biol Med.* 2015;7:13-38.
- Hsieh MY, Yang S, Raymond-Stintz M, et al. Stochastic simulations of ErbB homo- and hetero-dimerization: potential impacts of receptor conformational state and spatial segregation. *IET Syst Biol.* 2008;2:256-272. doi:10.1049/iet-syb.
- Hsieh M, Yu Yang S, Raymond-Stintz MA, Edwards JS, Wilson BS. Spatio-temporal modeling of signaling protein recruitment to EGFR. *BMC Syst Biol.* 2010;4:1-19.
- Ritchie K, Kusumi A. Single-particle tracking image microscopy. *Methods Enzymol.* 2003;360:618-634.
- Wilson BS, Pfeiffer JR, Raymond-Stintz MA, et al. Exploring membrane domains using native membrane sheets and transmission electron microscopy. *Methods Mol Biol.* 2007;398:245-261.
- Kusumi A, Nakada C, Ritchie K, et al. Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: high-speed single-molecule tracking of membrane molecules. *Annu Rev Biophys Biomol Struct.* 2005;34:351-378.
- Mac Gabhann F, Popel AS. Dimerization of VEGF receptors and implications for signal transduction: a computational study. *Biophys Chem.* 2007;128:125-139.
- Barua D, Hlavacek WS, Lipniacki T. A computational model for early events in B cell antigen receptor signaling: analysis of the roles of Lyn and Fyn. *J Immunol.* 2012;189:646-658.
- Radhakrishnan K, Halász Á, McCabe MM, Edwards JS, Wilson BS. Mathematical simulation of membrane protein clustering for efficient signal transduction. *Ann Biomed Eng.* 2012;40:2307-2318.
- Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol.* 2001;2:127-137. doi:10.1038/35052073.
- Birk DA, Barbato J, Mureebe L, Chaer RA. Current Insights on the biology and clinical aspects of VEGF regulation. *Vasc Endovascular Surg.* 2010;42:517-530.
- Karamysheva AF. Mechanisms of angiogenesis. *Biochemistry (Mosc).* 2008;73:751-762.
- Güven E, Wester MJ, Wilson BS, Edwards JS, Halász ÁM. Characterization of the experimentally observed clustering of VEGF receptors. In: Češka M, Šafránek D, eds. *Computational Methods in Systems Biology: 16th International Conference, CMSB 2018, Brno, Czech Republic, September 12-14, 2018 Proceedings, Lecture Notes in Bioinformatics 11095.* Springer; 2018:75-92. doi:10.1007/978-3-319-99429-1_5
- Chen Y, Short C, Halász ÁM, Edwards JS. The impact of high density receptor clusters on VEGF signaling. *Electron Proc Theor Comput Sci.* 2013;2013:37-52.
- Espinoza FA, Oliver JM, Wilson BS. Using hierarchical clustering and dendrograms to quantify the clustering of membrane proteins. *Bull Math Biol.* 2011;74:190-211.
- Pryor MM, Steinkamp MP, Halasz AM, et al. Orchestration of ErbB3 signaling through heterointeractions and homointeractions. *Molecular Biology of the Cell.* 2015;26:4109-4123.
- Kerketta R, Halász ÁM, Steinkamp MP, Wilson BS, Edwards JS. Effect of spatial inhomogeneities on the membrane surface on receptor dimerization and signal initiation. *Front Cell Dev Biol.* 2016;4:81. doi:10.3389/fcell.2016.00081
- Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9:671-675.
- Zhang J, Leiderman K, Pfeiffer JR, Wilson BS, Oliver JM, Steinberg SL. Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale electron-dense probes and electron microscopy. *Micron.* 2006;37:14-34.
- Jain AK, Dubes RC. *Algorithms for Clustering Data* (Advanced Reference Series). Upper Saddle River, NJ: Prentice-Hall; 1988:218.
- Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika.* 1970;57:97-109. doi:10.1093/biomet/57.1.97.