Optimal multivalent targeting of membranes with many distinct receptors

Tine Curk[1,2], **Jure Dobnikar**[*1,2,3], **and Daan Frenkel**[*2]

[1]Beijing National Laboratory for Condensed Matter Physics & CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China ; [2]Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, UK; [3]School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

This manuscript was compiled on May 25, 2017

Cells can often be recognised by the concentrations of receptors expressed on their surface. For better (targeted drug treatment) or worse (targeted infection by pathogens), it is clearly important to be able to target cells selectively. A good targeting strategy would result in strong binding to cells with the desired receptor profile, and barely to other cells. Using a simple model, we formulate optimal design rules for multivalent particles that allows them to distinguish target cells based on their receptor profile. We find that: 1) it is not a good idea to aim for very strong binding between the individual ligands on the guest (delivery vehicle) and the receptors on the host (cell). Rather, one should exploit multivalency: high sensitivity to the receptor density on the host can be achieved by coating the guest with many ligands that bind only weakly to the receptors on the cell surface, 2) the concentration profile of the ligands on the guest should closely match the composition of the cognate membrane-receptors on the target surface and 3) irrespective of all details, the effective strength of the ligand-receptor interaction should be of the order of the thermal energy k_BT, where T is the absolute temperature and k_B is Boltzmann's constant. We present simulations that support the theoretical predictions. We speculate that, using the above design rules, it should be possible to achieve targeted drug delivery with a greatly reduced incidence of side effects.

cell targeting | multicomponent | nano-carrier design

Introduction

The fact that most cells can be recognised from the outside is advantageous for the normal functioning of an organism, but it can be a disadvantage when specific cells are being targeted by pathogens. Cells betray their identity (and state of health) by the composition profile of molecules that are exposed on their outer surface. In what follows, we will call these molecules 'receptors', irrespective of whether they are receptors in the biological sense (they are receptors for the ligands that will be used to recognise them). It would clearly be advantageous if diseased cells could be selectively targeted by a drug-delivery vehicle on the basis of their receptor profile. Here, the crucial word is 'selective': we wish to target only those cells that have the correct receptor profile, as binding of drug-delivery vehicles to other cells may lead to undesired side-effects.

Targeted drug delivery is based on identifying a specific marker (peptide, sugar) that is unique to the targeted group of cells. Binding to a single marker type can be effective if this molecule is presented in sufficient quantities on the outer surface of the targeted cell. However, in many cases of practical importance (e.g. many types of cancer), the markers that are known are not unique to cancer cells, but just overexpressed. Over the past 20 years many nanoparticle-based targeting methods have been developed. However, thus far, effective tumour drug delivery is hampered by the lack of reliable, unique markers (1, 2).

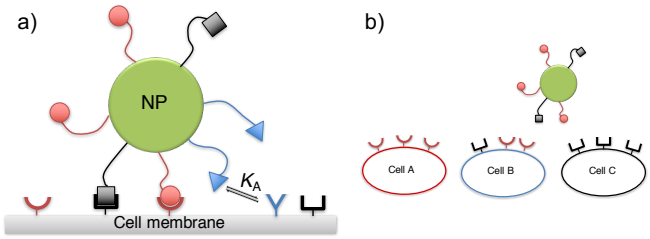


Fig. 1. Schematic pictures representing the multicomponent nano-particle targeting a receptor-decorated membrane. **b)** The challenge: How to target cell B in the presence of cells A and C ?

In order to recognise the simultaneous presence of a mixture of different receptors on the host surface, we need to use a 'guest' particle (e.g. drug-delivery vehicle) that is coated with a mixture of cognate ligands schematically shown on Figure 1. In its simplest form (the binding of dimeric bi-specific antibodies as compared to monomeric antibodies), this problem has been studied theoretically (3) and experimentally (4). The in-vitro experiments showed that the use of bi-specific antibodies lead to a higher specificity than can be achieved with their standard, monomeric counterparts. However, antibodies are not very good at distinguishing between surfaces that have different receptor concentrations. Such selectivity can be achieved by exploiting multi-valency. A series of recent experimental and theoretical papers have shown that multivalent carriers (nanoparticles or polymers) can distinguish target surfaces (cells) on the basis of their receptor concentration, rather than just

Significance Statement

A key challenge in biomedical research is the ability to specifically target cells and tissues. Targeting typically relies on identifying a suitable marker, e.g. a highly expressed receptor, and choosing a ligand that strongly and specifically binds to the marker. However, this procedure fails when a suitable marker unique to the targeted cells cannot be identified, notably in many forms of cancer. We show that properly designed multivalent targeting of multiple cognate receptor types results in a specificity towards a chosen receptor density profile. Thus demonstrating a general route towards targeting cells without particularly dominant markers.

TC, JD and DF designed research; TC performed research; TC, JD and DF wrote the paper.

The authors declare no conflict of interest.

[*]To whom correspondence should be addressed. E-mail: df246cam.ac.uk, jd489cam.ac.uk

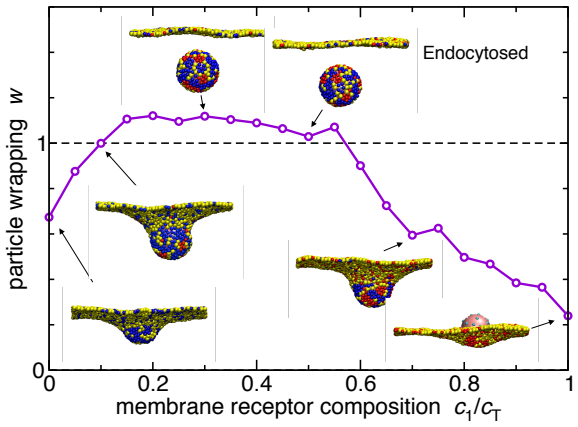


Fig. 2. Simulation results of nanoparticle endocytosis. We consider a system with two ligand types on the particle and two cognate receptor types in the membrane with concentrations c_1 and c_2, respectively. The red and blue beads in the membrane denote receptors of type 1 and 2, respectively. The yellow beads are inert (no binding to the ligands). The total concentration of receptors is kept fixed at $c_1 + c_2 = c_T = 0.4$, but the composition c_1/c_T is varied. The curve shows the coverage of the particle by the membrane beads (particle wrapping). When the wrapping exceeds 1 the particle is fully covered and has therefore undergone endocytosis. Snapshots show corresponding system configurations. The nanoparticle is covered with 40 randomly distributed immobile ligands with a 'ligand' profile $p_1 = 1 - p_2 = 0.3$. The interaction strength of a ligand patch i s determined as $\epsilon_i = \epsilon^* - \ln(p_i)$, with $\epsilon^* = 5k_BT$ for the above results.

on the basis of the presence of a suitable receptor (5–9).

The use of multivalent particles coated with a single type of ligand is very effective, provided that a cognate receptor has been identified that is sufficiently over-expressed in targeted cells. But often the situation is not that clear cut (see Figure 1). In general, it is essential to exploit all the information that we have about the concentration of various receptors on the cell surface and then design guest particles that target this specific receptor profile. In this paper, we show that the design rules for such multi-component targeting are surprisingly simple, and therefore hopefully useful. Specifically, we show that the individual ligand-receptor binding strength needs to be weak, such that when the guest particle is within interaction range of the surface, each ligand is unbound 30% of the time. To target a specific receptor profile selectively, many weak ligands work better than a few strong ones. We derive our results using a simple analytical theory and validate our approach using coarse-grained simulations. As an example, we present simulation results (see Figure 2) showing that correctly functionalised guest particles are more successful at entering cells with matching receptor compositions than those with sub-optimal compositions.

Of course, there is a wide variety of possible host-guest interactions that we could have considered: the receptors could be mobile or immobile, clustered or mixed, etc. In what follows we will assume that the host cell is large compared with the guest particle and that the receptors are mobile on its surface. In this case, the chemical potential of the receptors is effectively fixed and, as shown in the Supplementary Information, the theoretical expressions for the binding free energy become surprisingly simple. The other situations that we mention can also be modelled, and the overall conclusions remain the same, except in the limit where so many guest particles bind to the host surface that they deplete the receptor 'reservoir'. Hence, for the sake of simplicity, we consider just the easiest case. We do not specify the precise nature of the guest particle: it could be a functionalised nano-particle, polymer or a self-assembled DNA-origami structure. Again, for the current analysis the distinction is immaterial. What is important is that the ligands are flexible and can reach multiple receptors on the surface equally well.

Model

The membrane surface is covered by a population of receptors of different types j with concentrations $\mathbf{c} = \{c_j\}$. Similarly, a multivalent particle is characterised by the ligand composition $\{m_i\} \equiv m\{p_i\}$, where m is the total number of ligands on the guest particle and $\mathbf{p}$ specifies their relative profile, i.e. the relative coverage of the particle with ligands of type i is $p_i = m_i/m$. We are interested in the case where guest nanoparticles interact with a fluctuating number of receptors on the surface. Hence, following the Bell adhesion model (10), the number of receptors in contact with the nano-particle is not fixed, but their chemical potential is. In contrast, the number of ligands on the nano-particle that can interact with the surface is fixed. The nano-particle is only attracted to the surface through ligand-receptor interactions. Apart from that, the particle behaves like a hard sphere (see Figure 1). The ligand-receptor binding is valence limited, i.e. only a single ligand can bind to a receptor and vice versa. The model is an extension of the model of ref. (5, 7), generalised to include different ligand/receptor types.

Multicomponent theory. In order to calculate the binding free energy we need to consider all possible bonding combinations between receptors and ligands. To simplify the description, we neglect the interactions between different receptors and we assume that different ligands bind independently (except that no two ligands can bind to the same receptor). The probability that a single ligand i and a single receptor j form a bond depends on the equilibrium constant K_{ij}^A for their association in solution, and on the free-energy cost ΔG_i^{cnf}, which is due to the loss of configurational entropy of the ligand upon binding. The configurational entropy term ΔG_i^{cnf} obviously depends on the distance between the receptor and the grafting point of the ligand, which we capture in the simulations. However, it turns out that the distance-dependence of ΔG_i^{cnf} is not important in a simple theoretical description: in the Supplementary Information we show that we can treat the configurational term as if it were a constant ($\Delta G_i^{cnf} = \Delta \tilde{G}_i^{cnf}$) for all receptors within a distance h_0 of the ligands, and infinite elsewhere. The parameter h_0 represents the ligand-receptor interaction range and is determined by the length of the polymeric linker.

For a ligand within the interaction range of the receptor–decorated surface, the ratio between the probabilities of being in the bound and free (unbound) states is

$$\frac{P_{ij}^{\rm bound}}{P_i^{\rm free}} = c_j K_{ij}^A \frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{h_0} \equiv c_j K_{ij} \ . \qquad [1]$$

We have defined the effective association constant matrix $\mathbf{K}$, which includes the configurational contribution. Note that the first index i in K_{ij} always refers to a ligand, and the second j always to a receptor. Thus, K_{ii} describes the equilibrium constant for binding between ligand i and its *cognate* receptor i.

Emphatically, it does not mean that ligand i and receptor i are the same species. Similarly, K_{ij} describes the 'cross' binding of ligand i with the receptor cognate to ligand j. K_{ij} is, in general, not the same as K_{ji}, which describes the 'cross' binding of ligand j with receptor cognate to ligand i. Using the fact that probabilities must add to unity: $P_i^{\rm free} + \sum_j P_{ij}^{\rm bound} = 1$, we can directly determine the probability that a given ligand is unbound: $P_i^{\rm free} = \left(1 + \sum_j c_j K_{ij}\right)^{-1}$.

As shown in the Supplementary Information, the fact that the receptors do not interact with each other, and are in contact with a reservoir (the remainder of the cell surface) simplifies the expression of the binding free energy between a guest particle and a host membrane. Moreover, we assume that different ligands are uncorrelated, hence the binding free energy of the guest particle is simply a sum of individual ligand contributions, $\Delta F_b = m f_b$, with f_b the binding free energy per ligand:

$$f_b = \sum_i p_i \ln P_i^{\rm free} = -\sum_i p_i \ln\left(1 + \sum_j c_j K_{ij}\right) . \quad [2]$$

We note that this expression could be interpreted as (minus) the cross-entropy between the two distributions p_i and $P_i^{\rm free}$. The total binding free energy for a guest particle near the cell surface is $\Delta F = \Delta F_b + \Delta F_0$ where ΔF_0 is the zero-bond free energy cost of bringing a guest particle into a position to start forming bonds with the host membrane. For what follows, F_0 is unimportant, because we assume that it is the same irrespective of the receptor composition or the ligand profile, and thus it drops out of the expressions for free energy difference that determine the selectivity. We note that the host-guest binding free energy is related to the widely used "avidity" constant $K_{av}^A = e^{-\Delta F/k_B T}/\rho_0$ measuring the association between multivalent entities in units of standard molar concentration $\rho_0 = 1$M.

Furthermore, the same free energy expression (Eq. (2)) also governs the free-energy change upon a passive particle endocytosis (Figure 2) where ligands are not flexible, but the membrane itself is. In this case ΔF_0 simply refers to the particle endocytosis free-energy change when there are no receptors present, and ΔF_b again captures bonding with mobile receptors.

Selectivity optimisation

The expression (2) describes how the binding free energy depends on the receptor composition $\mathbf{c}$, particle profile $\mathbf{p}$ and the interaction matrix $\mathbf{K}$.

Our aim is to design a guest particle that binds more strongly to cells with the specified receptor profile c^*, than to any other. Among all possible receptor compositions $\mathbf{c}$, the targeted composition $\mathbf{c}^*$ should thus be the one with the minimum binding free energy. This yields the condition:

$$\frac{\partial f_b(\mathbf{c},\mathbf{p},\mathbf{K})}{\partial \mathbf{c}}|_{\mathbf{c}=\mathbf{c}^*} = 0 , \quad [3]$$

Note that this equation does not imply that we optimise the receptor composition of the target cell (after all, this composition is given). Rather, we vary the parameters that characterize the guest particle (*viz.* $\mathbf{p}$ and $\mathbf{K}$) to make the guest particle bind more strongly to the target receptor composition than to any other. Since there are several combinations of $\mathbf{p}$ and $\mathbf{K}$ that can satisfy this condition, we need to further select the one that is the most selective, i.e. the one that results in the free energy minimum with the largest curvature.

Our optimization condition is therefore to maximize the selectivity $\mathcal{S}$, defined as

$$\mathcal{S} = \det\left(\frac{\mathbf{H}(f_b)}{|f_b|}\right)_{\mathbf{c}=\mathbf{c}^*} , \quad [4]$$

subject to constraint of Eqn. 3. $\mathbf{H}(f_b)$ in Eqn. 4 is the determinant of the (Hessian) matrix $\mathbf{H}(f_b)$ of second derivatives of the free energy with respect to composition $\mathbf{c}$. As can be seen from Eqn. 4, the selectivity $\mathcal{S}$ is defined as the relative curvature of the free energy functional at its minimum.

It is important to define the selectivity as the relative, rather than the absolute, curvature. The absolute value of the free energy can be trivially controlled by changing the number of bonds m, Eq. (2). Therefore, by optimising for the relative curvature, we obtain the largest possible curvature at a given absolute value of the free energy ΔF_b. The binding can always be made stronger by increasing the total receptor concentration (see Refs. (5–7, 11)), therefore, in order to make a meaningful comparison, we compare systems with the same total receptor concentration $c_T \equiv \sum_i c_i$ on the membranes.

In general, finding the maximal selectivity by solving for the above conditions is non-trivial and must be performed numerically. However, we can greatly simplify the problem and find a closed form solution by also requiring that the binding free energy is optimised with respect to the ligand profile $\mathbf{p}$:

$$\frac{\partial f_b(\mathbf{c},\mathbf{p},\mathbf{K})}{\partial \mathbf{p}}|_{\mathbf{c}=\mathbf{c}^*} = 0 . \quad [5]$$

This is in principle not a necessary condition for selective targeting, however, we expect it to be a practically useful condition; we wish to design guests that are robust to small variations in the ligand profile. Robustness is important for practical applications when the guest particle manufacturing-process-tolerances must also be considered. The additional constraint results in a simple closed-form solution whilst the optimal selectivity only decreases marginally (see Supplementary Information).

We can determine the optimal ligand profile $\mathbf{p}$ and interaction matrix $\mathbf{K}$ analytically. The procedure that we use is discussed in the Supporting Information; here we only outline the main results. Our first result is that all ligands should have the same probability to be unbound

$$P_i^{\rm free} = e^{-\lambda_p} = const. \quad [6]$$

and that each ligand contributes an equal amount, $-\lambda_p k_B T$, to the total binding free energy. Hence, any ligand profile $\mathbf{p}$ will yield the same free energy $f_b = -\lambda_p k_B T$. In a sense, this result is trivial: it simply states that if all ligands are equally likely to bind a small change in the ligand profile will not change the overall host-guest binding free energy. This result should not be viewed as a design rule to 'target' guest particles by cells (in fact, the rule states that, in the optimal case, the cells cannot distinguish between different particles). Rather, we are interested in the opposite problem, namely the targeting of cells by guest particles. That problem does have a unique, non-trivial solution.

Minimizing the free energy functional Eq. (2) with respect to the particle profile $\mathbf{p}$ and targeted composition $\mathbf{c}^*$ determines the optimal ligand profile $\mathbf{p}^*$. For a symmetric interaction matrix $\mathbf{K}$ (or when off-diagonal terms are small) the ligand profile should match the cognate receptor composition: $\mathbf{p}^* = \mathbf{c}^*/c^T$. Finally, the selectivity $\mathcal{S}$ can be decomposed into a part which depends only on λ_p and a reduced Hessian $\hat{\mathbf{H}}$ which does not:

$$\mathcal{S} = \left[\frac{\left(1-e^{-\lambda_p}\right)^2}{\lambda_p}\right]^{d-1} \det(\hat{\mathbf{H}}) \,, \qquad [7]$$

where d is the "dimensionality" (the number of distinct receptor types). Optimising the selectivity: $\frac{\partial \mathcal{S}}{\partial(\lambda_p)} = 0$ we find that the non-trivial solution satisfies $e^{\lambda_p} - 2\lambda_p + 1 = 0$, and hence

$$\lambda_p \approx 1.256\cdots \qquad [8]$$

Eq. (8) is our most important result. It states that the binding free energy of each ligand to the targeted surface should be $f_b = -\lambda_p k_B T \approx -1.3 k_B T$ irrespective of the details of the system. Equivalently, the above equation states that when the guest particle is adsorbed on the targeted surface each ligand should be unbound 30% of the time: $P_i^{\rm free} = e^{-\lambda_p} \approx 0.3$.

Figure 3 shows the variation of the binding free energy with changing receptor composition obtained from the analytical model. In practice we might wish to distinguish a host surface with 20%-80% receptor composition form a surface with inverted 80%-20% composition. In Figure 3 we see that the difference in the corresponding bond strength is about $\Delta f_b \sim 0.5 k_B T$ per ligand. This may not seem to be much, but multivalent guests will hold 10-20 ligands (or more) if we require a total guest binding free energy of the order of $\Delta F_b \sim -10 - 25 k_B T$. The difference thus becomes substantial $\Delta\Delta F_b \sim 5 - 10 k_B T$, corresponding to orders of magnitude difference in adsorption and, therefore, very strong targeting efficacy as shown in Figure 4.

Furthermore, Figure 3**b)** demonstrates that increasing the number of ligand types increases the selectivity because the optimal binding region becomes a smaller fraction of the total parameter space.

Design rules. Our analytical calculations suggest the following simple design rules to make multivalent guest particles that target a particular receptor profile.

- $p_i = c_i/c_T$: ideally, the profile of the nanoparticle should match the density composition of the targeted cell. As shown in the Supplementary Information, this is not a condition on the *average* ligand profile. It really means that, ideally, every nano-particle should have precisely the optimal number of ligands. In fact, if only the averages are fixed and the number of ligands is Poisson distributed, most of the selectivity is lost.

- $K_{ii} = \frac{e^{\lambda_p}-1}{c_i}$, the value of K_{ii} should be inversely proportional to the density c_i. It is useful to avoid cross-binding (i.e. interaction matrix $\mathbf{K}$ should be diagonal). The optimal binding free energy per ligand $\lambda_p \approx 1.3$, which states that ligand binding should be weak, with each ligand independently having the probability of being bound at most 70%.

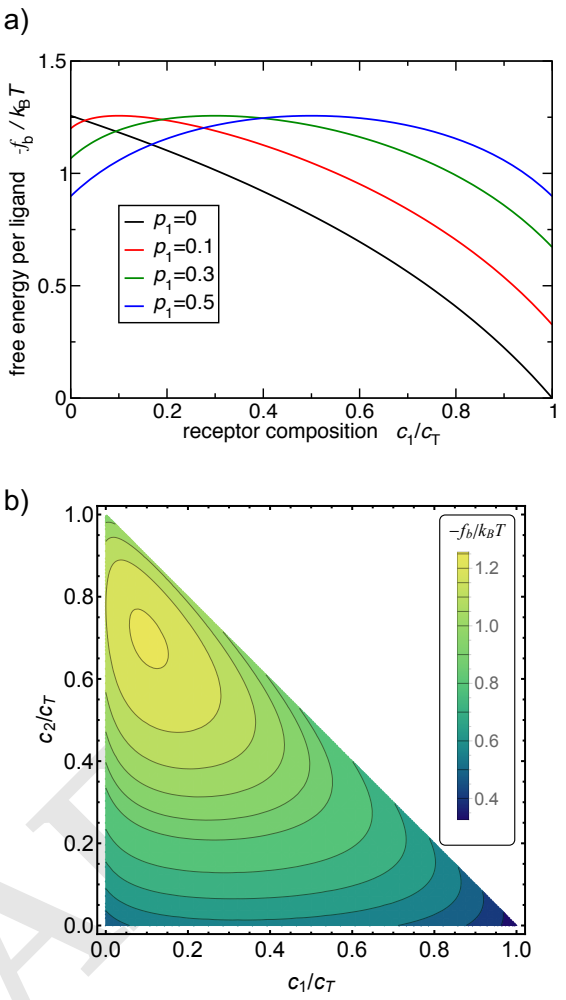


Fig. 3. Optimal targeting using the analytical theory. **a)** The binding free energy f_b per ligand as a function of the cell receptor composition for 2 ligand/receptor types. Different curves correspond to different guest profiles $p_1 = 1 - p_2$. **b)** Targeting with 3 ligand types. Contour plot showing the binding free energy as a function of the receptor composition c_1 and c_2. The ligand profile is chosen as $\mathbf{p} = [0.1, 0.2, 0.7]$. We have used Eq. (2) to calculate the free energy assuming a diagonal interaction matrix $\mathbf{K}$ and optimal $\lambda_p = 1.256$.

- the greater the number of distinct ligand-receptor types, the higher the potential selectivity.

- The overall binding free energy ΔF of the particle is proportional to the number of ligands per guest particle (valency m). Valency should be chosen such to give a desired absolute value of guest adsorption strength, for Langmuir adsorption the optimal will be close to the chemical potential of the guests in solution $\Delta F \sim \mu$.

The major assumption underlying the derivation of the design rules is that all pairs of ligands and receptors can in principle form a bond. This assumption is fulfilled when both the ligands and the receptors are mobile, which is often the case in biological systems. However, our results also apply in other situations with some additional restrictions: **(i)** For mobile ligands on the guest but immobile receptors on the membrane, the theory is relevant when the receptor density is large enough. If $A_{cont} \approx R^2$ is the surface area of the membrane in contact

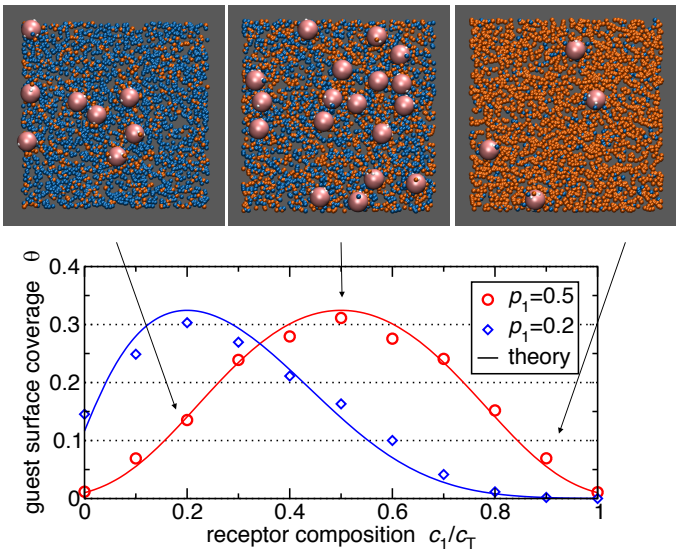


Fig. 4. Grand canonical Monte Carlo simulation results of guest adsorption (symbols) and comparison with analytical theory (solid lines). The simulated surface has two different types of receptors embedded (coloured orange and cyan respectively). The total concentration of receptors is kept fixed at $c_T = 7/R^2$, with R the guest particle radius. Each guest particle has ten ligands (each represented as a soft blob with radius r_b) with a profile of $p_1 = 0.5$ (red circles) and $p_1 = 0.2$ (blue diamonds), cognate bond energy $\epsilon = -4.5k_BT$. The guest chemical potential corresponds to a bulk solution concentration of $\rho = 10^{-6}(2R)^{-3}$. Theoretical adsorption curve (black line) are obtained by inserting the free energy expression Eq. (2) at $\lambda_p = 1.33$ into a standard Langmuir adsorption model assuming each guest occupies an area of $A = (2R + 2r_b)^2$. The surface coverage θ is the number of adsorbed guest per guest area A. The three snapshots correspond to the plotted data (red circles) at composition $c_1/c_T = 0.2, 0.5, 0.9$ and guest profile $p_1 = 0.5$.

with the guest particle, the receptor concentration should be large enough, $c_i > m_i/A_{cont}$, such that all ligands m_i can find their binding partners. **(ii)** In the case of mobile receptors and immobile ligands, the ligand profile p presented to the surface receptors must be independent of the guest particle orientation. This can be achieved either when using long, flexible ligand $h_0 > R$, or when targeting a deformable membrane that can wrap around a particle (see Figure 2), or by carefully uniformly coating the guest particle with ligands such that every "face" presents the same ligand profile. (iii) When both ligands and receptors are immobile, both constraints (i) and (ii) apply and in addition each ligand must be able to find a receptor within its interaction range h_0: $c_i > 1/h_0^2$.

Monte Carlo simulation results

The above analytical model is highly idealised. However, both the coarse-grained simulation results of particle endocytosis (Figure 2) and adsorption (Figure 4) support the predictions of our analytical model. Our simulations clearly show that our design rules, even though derived from a simple model, are nevertheless directly applicable to more complex and realistic systems where ligand interactions, correlations and membrane elasticity cannot be neglected.

The simulation snapshots of multivalent nanoparticle targeting in Figure 4 give a pictorial illustration of the effect of optimising the ligand concentration profile to target a mixed receptor surface. The adsorption isotherm is well captured by the simple analytical model. Furthermore, Figure 5 shows the selectivity, obtained from the analytical theory (Eq. (4)) and simulations, plotted as a function of the ligand binding strength λ_p. Clearly, the simulation results support the analytical value for the optimal ligand strength $\lambda_p \approx 1.3k_BT$.

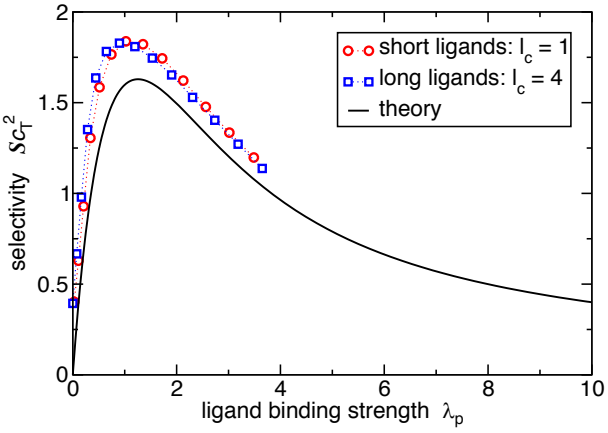


Fig. 5. Selectivity dependence on the mean bond strength λ_p. The theoretical curve is given by Eq. (7). The simulation results were obtained by calculating the free energy profile $f_b^{sim}(\mathbf{c}, \mathbf{p}^*, \mathbf{K})$ of a simulated guest particle and obtaining the relative curvature, and hence the selectivity $\mathcal{S}$, from a parabolic fit to the free energy profile. The mean bond strength from simulations is defined as the free energy per bond at the targeted composition $\lambda_p = -f_b^{sim}(\mathbf{c}^*, \mathbf{p}^*, \mathbf{K})/k_BT$. The simulated guest parameters are chosen according to our analytical design rules, $\mathbf{p}^* = [0.5, 0.5]$.

Conclusions

In this paper, we outline simple rules to design ligand-coated particles that can cell surfaces on the basis of a receptor profile, rather than on the recognition of a single receptor. The receptor profile of a cell surface can be viewed as a 'bar-code' that is selectively recognised by the ligand profile of the guest particle. Here we have shown that properly designed multivalent targeting of multiple cognate receptor types results in a specificity towards a chosen receptor composition. Thus demonstrating a general route towards targeting cells without particularly dominant markers.

We have assumed a generic case where background (untargeted) cell populations contain all possible receptor compositions. However, the selectivity can be increased further if only a few distinct cell populations are present and their receptor compositions are known in advance. In this case the optimal targeting strategy is obtained by maximising the free energy difference between discreet populations rather than the free energy curvature. We also note that, although in this paper we focus on the targeting of cells, our model can also be used to understand how imprinted polymers can be used to sort cells (12, 13).

Materials and Methods

Endocytosis simulations. We perform Monte Carlo simulations with a coarse-grained membrane model (14) and a patchy hard-sphere model (15) for the nanoparticle. The nanoparticle has two different types of circular patches modelling coverage with two different ligand types. The membrane is composed of individual beads which can be either inert (representing normal lipids) or "receptor" beads that bind to the cognate patches on the particle, but are otherwise identical to the inert beads. The receptor beads can interact with

the patches via a square well attraction with width σ equal to the dimeter of individual beads σ, ϵ denotes the well depth. The particle wrapping is calculated as the number of membrane beads within a distance σ of the particle surface N_w, normalised by the fully wrapped triangular lattice: $w = N_w \frac{\sqrt{3}}{2\pi(R+\sigma)^2}$, with $R = 4\sigma$ the particle radius.

The simulations are performed using standard Monte Carlo translational moves in NpT ensemble and no applied external pressure (to be precise, our membrane system is only metastable at zero applied external pressure, the thermodynamically stable configuration is an infinitely large box. However, on a simulation timescale, a flat membrane is stable). The box size is $40\sigma \cdot 40\sigma$ with periodic boundary conditions in lateral directions. The box size in the vertical z-direction is sufficiently large to ensure that none of the particles ever interacts with the hard ceiling or the floor. The simulations started with the particle center-of-mass at height $R+\sigma$ above the membrane and were run for $6 \cdot 10^6$ cycles where in each cycle on average one translational/rotational move per every bead and particle is attempted.

Multivalent particle adsorption. We performed Grand-canonical Monte Carlo simulations where the chemical potential of the guest particles in solution is fixed to a value that results in a desired guest density in solution. Guest particles are represented as hard spheres with radius $R = 3r_b$ and attached polymeric ligand arms that are modeled as a series of soft blobs of size r_b (16). Receptors are represented as points on the hard surface and can bind to ligands with a valence limited harmonic bond and the interaction matrix $\boldsymbol{\epsilon}$ determines the individual binding/unbinding probabilities. Standard Monte Carlo moves are employed to displace and add/delete guests into the system, Rosenbluth sampling is used to change polymeric ligand conformations. The model and technique are an extension of Ref. (5) to multiple ligand/receptor types.

Free energies are calculated using the Wang-Landau sampling technique (17). First, we bias the sampling in the number of formed ligand-receptor bonds to obtain the absolute value of the bound guest free energy relative to a common reference point; a single unbound guest particle within interaction range h_0 of the surface ($h_0 = 3r_b$ for single blob ligands and can be well approximated as the average height of a guest with a single formed bond). We then bias the simulation in the receptor composition to obtain the curvature of the free energy. The selectivity $\mathcal{S}$ is calculated by fitting a quadratic function to the free energy profile and normalising by the absolute value as per Eq. (4). Langmuir isotherm (Figure 4) zero-bond free energy is approximated as the translational entropy of a guest within a lattice site of size A and height h_0: $\Delta F_0 = -k_B T \log(A h_0 \rho_0 N_A)$.

ACKNOWLEDGMENTS. We thank Bortolo Mognetti and Stefano Angioletti-Uberti for in-depth discussions on multivalency, and Anđela Šarič for help with endocytosis simulations. The work was supported by the European Union through ETN NANOTRANS grant 674979 MSC and by the Herchel Smith fund.

1. Mura S, Nicolas J, Couvreur P (2013) Stimuli-responsive nanocarriers for drug delivery. *Nat Mater* 12(11):991–1003.
2. Wilhelm S et al. (2016) Analysis of nanoparticle delivery to tumours. *Nature Reviews Materials* 1:14–26.
3. Caplan MR, Rosca EV (2005) Targeting drugs to combinations of receptors: A modeling analysis of potential specificity. *Annals of Biomedical Engineering* 33(8):1113–1124.
4. Robinson MK et al. (2008) Targeting erbb2 and erbb3 with a bispecific single-chain fv enhances targeting selectivity and induces a therapeutic effect in vitro. *Br J Cancer* 99(9):1415–1425.
5. Martinez-Veracoechea FJ, Frenkel D (2011) Designing super selectivity in multivalent nanoparticle binding. *Proceedings of the National Academy of Sciences of the United States of America* 108(27):10963–10968.
6. Dubacheva GV et al. (2014) Superselective targeting using multivalent polymers. *Journal of the American Chemical Society* 136(5):1722–1725.
7. Dubacheva GV, Curk T, Auzély-Velty R, Frenkel D, Richter RP (2015) Designing multivalent probes for tunable superselective targeting. *Proceedings of the National Academy of Sciences* 112(18):5579–5584.
8. Carlson CB, Mowery P, Owen RM, Dykhuizen EC, Kiessling LL (2007) Selective tumor cell targeting using low-affinity, multivalent interactions. *ACS Chemical Biology* 2(2):119–127.
9. Kiessling LL, Grim JC (2013) Glycopolymer probes of signal transduction. *Chem. Soc. Rev.* 42:4476–4491.
10. Bell G, Dembo M, Bongrand P (1984) Cell adhesion. competition between nonspecific repulsion and specific bonding. *Biophysical Journal* 45(6):1051–1064.
11. Xu H, Shaw DE (2016) A simple model of multivalent adhesion and its application to influenza infection. *Biophysical Journal* 110(1):218–233.
12. Ren K, Zare RN (2012) Chemical Recognition in Cell- Imprinted Polymers. *ACS Nano* 6(5):4314–4318.
13. Ren K, Banaei N, Zare RN (2013) Sorting inactivated cells using cell-imprinted polymer thin films. *ACS Nano* 7(7):6031–6036.
14. Yuan H, Huang C, Li J, Lykotrafitis G, Zhang S (2010) One-particle-thick, solvent-free, coarse-grained model for biological and biomimetic fluid membranes. *Phys. Rev. E* 82:011905.
15. Kern N, Frenkel D (2003) Fluid-fluid coexistence in colloidal systems with short-ranged strongly directional attraction. *The Journal of Chemical Physics* 118(21):9882–9889.
16. Pierleoni C, Capone B, Hansen JP (2007) A soft effective segment representation of semidilute polymer solutions. *The Journal of Chemical Physics* 127(17):171102.
17. Wang F, Landau DP (2001) Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* 86:2050–2053.
18. Schmidt NW et al. (2015) Liquid-crystalline ordering of antimicrobial peptide-DNA complexes controls TLR9 activation. *Nat Mater* 14(7):696–700.
19. Huskens J et al. (2004) A model for describing the thermodynamics of multivalent host-guest interactions at interfaces. *Journal of the American Chemical Society* 126(21):6784–6797.
20. Hunter CA, Anderson HL (2009) What is cooperativity? *Angewandte Chemie International Edition* 48(41):7488–7499.
21. Ercolani G, Schiaffino L (2011) Allosteric, chelate, and interannular cooperativity: A mise au point. *Angewandte Chemie International Edition* 50(8):1762–1768.
22. Krishnamurthy VM, Semetey V, Bracher PJ, Shen N, Whitesides GM (2007) Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *Journal of the American Chemical Society* 129(5):1312–1320.
23. Douglas SM, Bachelet I, Church GM (2012) A logic-gated nanorobot for targeted transport of molecular payloads. *Science* 335(6070):831–834.
24. Speck T (2011) Effective free energy for pinned membranes. *Phys. Rev. E* 83:050901.
25. Martinez-Veracoechea FJ, Bozorgui B, Frenkel D (2010) Anomalous phase behavior of liquid-vapor phase transition in binary mixtures of dna-coated particles. *Soft Matter* 6:6136–6145.
26. Frenkel D, Smit B (2001) *Understanding Molecular Simulation: From Algorithms to Applications*, Computational science series. (Elsevier Science).
27. Varilly P, Angioletti-Uberti S, Mognetti BM, Frenkel D (2012) A general theory of dna-mediated and other valence-limited colloidal interactions. *The Journal of Chemical Physics* 137(9):094108.
28. Angioletti-Uberti S, Varilly P, Mognetti BM, Tkachenko AV, Frenkel D (2013) Communication: A simple analytical formula for the free energy of ligand-receptor-mediated interactions. *The Journal of Chemical Physics* 138(2):021102.

Optimal multivalent targeting of membranes with many distinct receptors

Curk, Dobnikar, Frenkel 10.1073/pnas.XXXXXXXXXX

Supporting Information (SI)

In this SI we provide detailed calculations, derivation of the analytical model, more supporting results and illuminate equivalence between mobile and immobile receptor.

1. Derivation of the analytical results

We wish to minimise the binding free energy per ligand

$$f_b(\mathbf{c},\mathbf{p},\mathbf{K}) = -\sum_i p_i \ln\left(1+\sum_j c_j K_{ij}\right) \quad [1]$$

subject to two constraints

$$\sum_j c_j = c_T, \qquad \sum_i p_i = 1\ . \quad [2]$$

The total number of receptors c_T is fixed and the ligand profile $\mathbf{p}$ is a normalised vector. Obviously, all concentrations and equilibrium constants must be non-negative: $c_j \geq 0$, $p_i \geq 0$, $K_{ij} \geq 0$. The total binding free energy $\Delta F_b = m f_b$ is trivially proportional to the total number of ligands m, hence the overall binding strength can be controlled by varying m. We treat the profile vector $\mathbf{p}$ as continuous which is only approximate since $\mathbf{p}$ is in principle a discrete vector. With m ligands per guest the expected discretisation error in the targeted composition is $\Delta c_i \sim \frac{c_T}{m}$. Using the second order expansion ($H_{ii} \sim d/c_T^2$), we estimate the discretisation error in the total binding free energy is $\Delta F^{err} \sim m\sum_i H_{ii}(\Delta c_i)^2 \sim \frac{d^2}{m} k_B T$, with d the number of distinct targeted receptor types.

We first consider which host profile binds most strongly to a guest particle, that is, we need to determine the receptor composition $\mathbf{c}^*$ that minimizes the binding free energy. This problem is trivially solved using Lagrange multipliers.

$$\nabla_{\mathbf{c}}(f_b + \lambda_c c_T) = 0 \quad [3]$$

which becomes

$$\left(\frac{\partial f_b}{\partial c_l}\right)_{\mathbf{c}=\mathbf{c}*} = -\lambda_c \quad [4]$$

with λ_c the Lagrange multiplier. Inserting Eq. (1) into the above we find

$$\sum_i \frac{p_i K_{il}}{1+\sum_j c_j^* K_{ij}} = \lambda_c \quad [5]$$

which must hold for every receptor type $_l$. Given arbitrary targeted composition $\mathbf{c}^*$ we can chose any $\mathbf{p}$ and $\mathbf{K}$ that satisfy the above equation, and by definition the $\mathbf{c}^*$ will be a minimum.

The condition that the binding free energy is also a minimum with respect to the ligand composition profile on a guest particle $\mathbf{p}$ is, again imposed using a Lagrange multiplier λ_p:

$$\left(\frac{\partial f_b}{\partial p_i}\right)_{\mathbf{c}=\mathbf{c}^*} = -\ln\left(1+\sum_j c_j^* K_{ij}\right) = -\lambda_p\ , \quad [6]$$

which must hold for every ligand type i. We recall that $P_i^{\text{free}} = \left(1+\sum_j c_j^* K_{ij}\right)^{-1}$ is the probability that a ligand of type i is not bound, a simple result follows:

$$P_i^{\text{free}} = e^{-\lambda_p}\ . \quad [7]$$

We shall define relative cross-binding terms κ_{ij} which are determined by the specificity of the ligands and receptors:

$$\kappa_{ij} = \frac{K_{ij}}{K_{ii}} = e^{-(\Delta G_{ij} - \Delta G_{ii})/k_B T}\ , \quad [8]$$

where ΔG_{ij} is the Gibbs free energy of monomeric ligand-receptor dimerisation in solution. The configurational term ΔG_i^{cnf} in the definition of K_{ij} (Eq. (1) in the main text) is the same for all receptors j and cancels out in the above expression.

Therefore κ_{ij} are constants determined by the association matrix ΔG_{ij}, i.e. constants determined by the choice of ligands. We assume that the overall strength of the interaction can be tuned by changing ΔG_i^{cnf} via, for example, the polymer linker length. On the other hand, treating all κ_{ij} as variables (i.e. assuming an infinite continuous space of possible ligand choices) we simply find that maximal selectivity is obtained when individual ligands are maximally specific ($\kappa_{ij} = 0, i \neq j$).

Inserting Eq. (6) into Eq. (5) we find that the solution must satisfy

$$\sum_i p_i K_{ij} = \lambda_c e^{\lambda_p} \ , \qquad [9]$$

additionally Eq. (6) can be rearranged to

$$\sum_j c_j^* K_{ij} = e^{\lambda_p} - 1 \ . \qquad [10]$$

There are $2d$ equations and $2d+1$ unknowns; taking into account that relative off diagonal elements κ_{ij} are constants Eq. (8) and $\sum_i p_i = 1$. The "dimensionality" $d = \text{rank}(\mathbf{K})$ is determined by the rank of matrix $\mathbf{K}$ or, equivalently, the number of distinct receptor types. Therefore, the above equations determine the optimal ligand profile $\mathbf{p}^*$ and all cognate interaction strengths K_{ii}^* up to a constant factor e^{λ_p}.

If $\mathbf{K}$ is a diagonal matrix (only cognate interaction) or, more generally, a symmetric matrix ($K_{ij} = K_{ji}$, i.e. the off diagonal equilibrium constant for the cross-binding of ligand i and receptor j is the same as that between ligand j and receptor i) the above equations imply that the optimal guest profile $\mathbf{p}^*$ must match the targeted composition of the receptors $\mathbf{c}^*$

$$p_i^* = \frac{c_i^*}{c_T} \qquad [11]$$

with $c_T = \sum_j c_j$ the total receptor density and the Lagrange multipliers are related $\lambda_c = \frac{1-e^{-\lambda_p}}{c_T}$. The optimal cognate interaction strength should be inversely proportional to the composition

$$K_{ii}^* = \frac{e^{\lambda_p} - 1}{c_i^*} \qquad [12]$$

using Eq. (10) in the case of no cross-binding. For general non-symmetric matrices Eqs. (11,12) will provide a good approximation whenever cross-binding is weak and cognate interaction dominates: $c_i^* \gg \sum_{j\neq i} \kappa_{ij} c_j^*$.

The only remaining step is to determine the optimal strength of the interaction captured by the constant λ_p. We optimise the selectivity (Eq. (4) of the main text) given by the determinant of the Hessian matrix $\mathbf{H}$. For direct comparison to plotted results (Figures 3 and 4 of the main text) we include the constraint of fixed total receptor concentration: with d receptor types there are only $d-1$ independent concentrations and $c_d = c_T - \sum_j^{d-1} c_j$. Removing this constraint in the definition of $\mathbf{H}$ would only rescale the value of selectivity and would not affect our main results.

An element of the Hessian matrix is obtained by twice differentiating the free energy (Eq. (1))

$$H_{mn} = \left(\frac{\partial^2 f_b}{\partial c_m \partial c_n} \right)_{\mathbf{c}=\mathbf{c}^*} = \sum_i \frac{p_i (K_{im} - K_{id})(K_{in} - K_{id})}{(1 + \sum_j c_j^* K_{ij})^2} \ , \qquad [13]$$

where the indices fall between $1 \leq m, n \leq d-1$. Using Eq. (8) we rewrite Eq. (10) in the form $K_{ii} = \frac{e^{\lambda_p}-1}{\sum_j c_j^* \kappa_{ij}}$, inserting these into the above equation we find that every element of the Hessian matrix decouples into a term that depends only on λ_p:

$$H_{mn} = (1 - e^{-\lambda_p})^2 \, \hat{H}_{mn} \qquad [14]$$

and the remainder $\hat{H}_{mn} = \sum_i \frac{p_i (\kappa_{im} - \kappa_{id})(\kappa_{in} - \kappa_{id})}{(\sum_j c_j^* \kappa_{ij})^2}$ which does not depend on λ_p. In fact the values of all $\hat{H}_{mn}$ are at this point already determined by the solution to Eqs. (9,10). Since λ_p is a constant for the whole matrix $\mathbf{H}$, the determinants are related by a factor: $\det(\mathbf{H}) = (1 - e^{-\lambda_p})^{2d-2} \det(\hat{\mathbf{H}})$.

We use the above relation and $f_b(\mathbf{c}^*, \mathbf{p}, \mathbf{K}^*) = -\lambda_p$ to express the selectivity in terms of the Lagrange multiplier λ_p

$$\mathcal{S} = \det \left(\frac{\mathbf{H}(f_b)}{|f_b|} \right)_{\mathbf{c}=\mathbf{c}^*} = \left[\frac{\left(1 - e^{-\lambda_p}\right)^2}{\lambda_p} \right]^{d-1} \det(\hat{\mathbf{H}}) \ , \qquad [15]$$

with d the number of targeted receptor types. Evidently, the optimal λ_p is given by $\frac{\partial \mathcal{S}}{\partial \lambda_p} = 0$. Since $\hat{\mathbf{H}}$ does not depend on λ_p the derivative is simple to work out. The non-trivial solution satisfies

$$e^{\lambda_p} - 2\lambda_p + 1 = 0, \qquad [16]$$

and is given by the -1^{st} branch of the Lambert W function

$$\lambda_p = -W_{-1}\left(\frac{-1}{2\sqrt{e}} \right) - \frac{1}{2} \approx 1.25643\cdots \qquad [17]$$

approximated to the first 6 digits.

Cross entropy analogy. The probability that a given ligand is unbound can be written as $P_i^{\text{free}} = \left(1 + \sum_j c_j K_{ij}\right)$. The expression for the binding free energy Eq. (1) thus becomes very similar to the cross entropy between distributions p_i and P_i^{free}

$$f_b = \sum_i p_i \ln P_i^{\text{free}} \, . \tag{18}$$

However, $\mathbf{P}^{\text{free}}$ is not a true probability distribution as it is not properly normalised. By defining a normalised distribution $\hat{P}_i^{\text{free}} = P_i^{\text{free}}/a$, with $a = \sum_i P_i^{\text{free}}$ the normalisation constant, the binding free energy becomes

$$f_b = \sum_i p_i \ln \hat{P}_i^{\text{free}} + \ln \sum_i P_i^{\text{free}} = -H(\mathbf{p}, \hat{\mathbf{P}}^{\text{free}}) + E(\mathbf{P}^{\text{free}}), \tag{19}$$

where $H(\mathbf{p}, \hat{\mathbf{P}}^{\text{free}})$ is the cross entropy and $E = \ln \sum_i P_i^{\text{free}}$ is a "cost function" analogous to energy, it measures the overall strength of a bond. If bonds are weak then $P_i^{\text{free}} \sim 1$ and $E > 0$. Conversely, with strong bonds $P_i^{\text{free}} \sim 0$ and E will be negative.

The cross entropy can further on be written as the sum of Shannon entropy and Kullback-Leibler divergence

$$H(\mathbf{p}, \hat{\mathbf{P}}^{\text{free}}) = -\sum_i p_i \ln p_i + \sum_i p_i \ln \frac{p_i}{\hat{P}_i^{\text{free}}} = H(\mathbf{p}) + D_{KL}(\mathbf{p}||\hat{\mathbf{P}}^{\text{free}}). \tag{20}$$

Such that the binding free energy per ligand becomes

$$f_b = E(\mathbf{P}^{\text{free}}) - H(\mathbf{p}) - D_{KL}(\mathbf{p}||\hat{\mathbf{P}}^{\text{free}}). \tag{21}$$

The first term $E(\mathbf{P}^{\text{free}})$ captures the overall bond strength ("energy"), the second $H(\mathbf{p})$ is the Shannon entropy of the ligands, it measures the diversity of the ligands on the guest particle. The Kullback-Leibler divergence $D_{KL}(\mathbf{p}||\hat{\mathbf{P}}^{\text{free}})$ is a measure of the difference between the two distributions $\mathbf{p}$ and $\hat{\mathbf{P}}^{\text{free}}$. Hence, to minimize the free energy: Energy favours strong individual bonds $P_i^{\text{free}} \ll 1$, Entropy favours uniformity $p_i \sim 1/d$, with d the number of ligand types, finally, the Kullback-Leibler divergence favours the two distributions to be as different as possible. The interplay and competition between the three different terms results in simple design principles for optimal targeting.

2. Derivation of the simple analytical model

For mobile receptors the expression for the bound partition function of a multivalent guest particle binding to the receptor decorated membrane is given approximately by

$$Q_b = \prod_i \left(1 + \sum_j c_j K_{ij}\right)^{m_i} , \tag{22}$$

from which we obtain the free energy change due to bond formation

$$\Delta F_b = -k_B T \ln(Q_b) = -m \sum_i p_i \ln \left(1 + \sum_j c_j K_{ij}\right) . \tag{23}$$

We have used this expression to derive our design rules for composition targeting, where c_j is the receptor type j concentration on the host, $m_i = m p_i$ is the number of ligands of type i on the guest particle, with p_i the profile, and K_{ij} are elements of the interaction matrix.

We will start from basic statistical mechanics and show what approximations are necessary to arrive at our simplified expression Eq. (22). Initially we assume that receptors are non-interacting and mobile on the flat host surface and can, therefore, be effectively described as solutes in a 2D ideal solution. Below we also provide a, more tedious derivation, showing that the same result is expected for immobile receptors with fixed but random positions on the host surface. We treat the host membrane as a flat hard surface and the particle coated with flexible polymeric arms, each carrying a ligand. A model of particle endocytosis where the ligands are rigidly attached to the guest particle, but the host membrane is deformable results in the the same free energy expression Eq. (23). Multiple possible combinations of forming bonds are necessary to derive Eq. (22). However, the bonding flexibility and combinatorics can be achieved in many different ways: flexible ligands, flexible receptors or a deformable membrane with mobile receptors.

In our theoretical treatment the guest particle is modeled as a hard sphere with attached polymeric ligand arms, shown schematically on Figure 1 of the main text. A particle is grafted with a total of m polymer linkers, each linker caries a ligand at the tip. The host surface is a flat hard surface with mobile receptors. To arrive at our simple theory, we make a number of approximations: (i) ligand binding is uncorrelated; bound/unbound state of a ligand will not affect the probability that another ligands binds, (ii) Ligands themselves are non-interacting and their positions are independent, for this to hold we assume a mean field approximation where individual ligand grafting points are uncorrelated and are mobile on a particle surface.

The particle is fixed at a height h above the surface. The configurational free energy $\Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'})$ captures the, mainly entropic, effects of displacing the ligand $\mathbf{r}_{i'}$ with respect to the anchor (grafting) point on the particle $\mathbf{r}^a_{i'}$ [*]. If the polymer linker is a flexible polymer the $\Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'})$ will be approximately a quadratic function of the distance $|\mathbf{r}_{i'} - \mathbf{r}^a_{i'}|$. For brevity, we neglect any angular contribution to $\Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'})$, we also neglect the effects of the chemical coupling of ligands to the polymer linker and the interactions between the ligand and the particle. All of these contributions will uniformly change ΔG^{cnf} by a constant value and can be fitted from experiments as discussed below.

The bound partition function of a single ligand i' with its grafting point at position $\mathbf{r}^a_{i'}$ reads

$$q_b(\mathbf{r}^a_{i'}) = \sum_j e^{-\beta \Delta G_{i'j}} c_j \int_S e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'}; z_{i'}=0)} \delta(z_{i'}) d\mathbf{r}_{i'} \ , \qquad [24]$$

where we remember $e^{-\beta \Delta G_{i'j}}$ as the Gibbs free energy interaction matrix from solution, c_j is the surface concentration of receptor type j, the integration is performed over the whole surface S with $\delta(z_{i'})$ is the Dirac delta function constraining the $z_{i'}$ coordinate of the ligand constrained to lie on the surface $z_{i'} = 0$. The particle can freely rotate, therefore, we proceed by integrating the ligand anchor point over the particle surface, the bound partition function of ligand i' with a particle at height h is

$$q^i_b(h) = \int_G q_b(\mathbf{r}^a_{i'}) d\mathbf{r}^a_{i'} = \sum_j e^{-\beta \Delta G_{ij}} c_j \int_G d\mathbf{r}^a_{i'} \int_S e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'}; z_{i'}=0)} \delta(z_{i'}) d\mathbf{r}_{i'} \ , \qquad [25]$$

with $\int_G$ an integral over the guest particle surface. Note that we have used a normal index i denoting a ligand type. For every particular ligand i' that is of the same type i, this integral returns the same value. We implicitly assume an existence of an indicator function that maps every ligand i' to its type i. Unbound partition function of the same ligand when the particle is free in solution is

$$q^i_u = 4\pi r^2_{gp} q_u(\mathbf{r}^a_{i'}) = 4\pi r^2_{gp} \rho_0 \int_V e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'})} d\mathbf{r}_{i'} \ . \qquad [26]$$

For noninteracting unbound ligands the volume integral $\int_V$ of the ligand position $\mathbf{r}_{i'}$ does not depend on $\mathbf{r}^a_{i'}$, therefore, $4\pi r^2_{gp}$ comes from integrating over the guest surface, with r_{gp} the guest particle radius. We neglect that ligands cannot penetrate the surface, therefore, q^i_u does not depend particle height h. $\rho_0 = 1\text{M}$ is the standard concentration with respect to the interaction matrix ΔG_{ij}.

The ratio of the partition functions determines the ratio or probabilities of finding a ligand i in the bound state at height h to unbound state

$$\frac{p^i_{\text{bound}}(h)}{p^i_{\text{unbound}}} = \frac{q^i_b(h)}{q^i_u} = \tilde{K}_i(h) \sum_j e^{-\beta \Delta G_{ij}} c_j \ . \qquad [27]$$

The ligand can be attached to any receptor type j, hence the sum. We have introduced

$$\tilde{K}_i(h) = \frac{\int_C d\mathbf{r}^a_{i'} \int_S e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'}; z_{i'}=0)} \delta(z_{i'}) d\mathbf{r}_{i'}}{4\pi r^2_{np} \rho_0 \int_V e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'})} d\mathbf{r}_{i'}} \ , \qquad [28]$$

which measures the configurational (mostly entropic) cost of localising a ligand i at a surface with the particle at height h above the surface. We expect this cost to increase with increasing height h due to the polymeric linker stretching penalty captured by $\Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'})$. Therefore, the ligand will have an appreciable probability of being bound only when the particle is within an interaction distance h_0 of the surface and binding probability vanishes for large h as $q^i_b(\infty) = 0$. Figure S1 shows the height distribution profiles $p(h)$ for guest particle obtained from simulations. $\tilde{K}_i(h) \propto p(h)$ is proportional to the probability distribution.

The partition function of the whole particle at height h is, for noninteracting ligands, simply a product over all ligand types

$$Q(h) = \prod_i \left(q^i_u + q^i_b(h) \right)^{m_i} \ , \qquad [29]$$

because each ligand independently can be either in a bound or unbound state and we remember m_i as the number of ligands of type i. In order to obtain meaningful predictions for the binding free energy of particle attachment, we must take a ratio of bound to unbound partition functions

$$e^{-\beta \Delta F_b(h)} = Q_b(h) = \frac{Q(h)}{Q(\infty)} = \prod_i \left(1 + \frac{q^i_b(h)}{q^i_u} \right)^m_i = \prod_i \left(1 + \tilde{K}_i(h) \sum_j e^{-\beta \Delta G_{ij}} c_j \right)^{m_i} \ , \qquad [30]$$

which determines the free energy difference $\Delta F_b(h)$ or the normalised partition function $Q_b(h)$ for a particle at height h with respect to a particle free in solution. Eq. (27) shows that this free energy difference is related only to the probabilities of individual ligands being bound or unbound.

[*] we use primes i' to denote a *specific* single ligand, while a standard index i denotes the ligand type

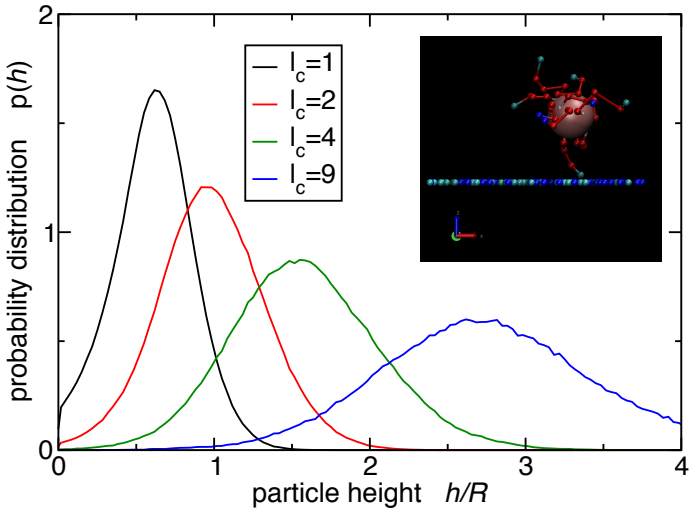


Fig. S1. Simulation results of height probability distribution for a guest particle with a single formed bond. Different curves correspond to varying ligand chain length l_c defined as the number of blobs per ligand chain. The inset shows a representative simulation snapshot of a particle with $m = 10$ ligands and chain length $l_c = 4$ is the number of soft blobs used to represent the ligand polymer linker. The origin ($h = 0$) is defined when the particle is touching the surface; when the particle centre-of-mass is located at radius R above the surface. The particle radius is $R = 3r_b$ with r_b the blob radius. The average height is found at $\langle h \rangle / R = [0.61, 0.99, 1.58, 2.77]$ for the four different chain lengths $l_c = [1, 2, 4, 9]$. System size is $10R \times 10R$ in the lateral directions with receptor concentration $c_T = 1/R^2$, $p_1 = c_1/c_T = 0.5$.

In practice, the relevant measure is the free energy of a bound particle at a surface, hence the above equations needs to be integrated over particle height h to obtain an integrated partition function $\tilde{Q}_b$. We need some sort of a cutoff specifying how we determine a bound particle. One possibility is to consider only particles with at least a single bound ligand

$$\tilde{Q}_b = \int_0^\infty Q_b(h) - 1 \, dh = \int_0^\infty \left[\prod_i \left(1 + \tilde{K}_i(h) \sum_j e^{-\beta \Delta G_{ij}} c_j \right)^{m_i} - 1 \right] dh \, , \qquad [31]$$

another option is to consider all particles within a cutoff height h_{cutoff}

$$\tilde{Q}_b = \int_0^{h_{\text{cutoff}}} Q_b(h) \, dh = \int_0^{h_{\text{cutoff}}} \prod_i \left(1 + \tilde{K}_i(h) \sum_j e^{-\beta \Delta G_{ij}} c_j \right)^{m_i} dh \, . \qquad [32]$$

For sufficiently large h_{cutoff} such that $Q_b(h > h_{\text{cutoff}}) \approx 1$ and dilute solutions (negligible probability to find a particle with no formed bonds within h_{cutoff}) the two forms will return the same value for $\tilde{Q}_b$. We chose the second expression Eq. (32) as our definition of chose as it conveniently enables write the free energy as a sum over individual ligand contributions. In the previous work (5,6,7,18) an alternative definition Eq. (31) was used.

The total free energy change of binding a guest from solution to a surface depends on the surface area size A

$$\Delta F = -k_b T \log \left(A \tilde{Q}_b \rho_0 N_A \right) + \Delta F_{ns} \, , \qquad [33]$$

where ΔF_{ns} includes any non specific surface-guest interactions, i.e. interactions that are not mediated by ligand/receptor binding, ρ_0 and N_A are the standart concentration and Avogadro's constant, respectively.

To arrive at the expression so far we have only assumed that individual ligands bind independently and are non-interacting. In order to obtain our simple analytical model Eq. (22) we must also approximate $\tilde{K}_i$ with a block function of height h_0:

$$\tilde{K}_i(h) = \begin{cases} \frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{\rho_0 h_0} & ; 0 \leq h < h_0 \\ 0 & ; \text{otherwise} \end{cases} \qquad [34]$$

where $e^{-\beta \Delta \tilde{G}_i^{cnf}}$ is the integrated configurational cost of binding a ligand to a surface. Zero height ($h = 0$) is defined when the guest particle touching the surface. We define $e^{-\beta \Delta \tilde{G}_i^{cnf}}$ and h_0 using the zeroth and first moment of $\tilde{K}_i(h)$

$$\frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{\rho_0} \equiv \int_0^\infty \tilde{K}_i(h) \, dh \, . \qquad [35]$$

and

$$h_0 \equiv 2\,\frac{\int_0^\infty h\tilde{K}_i(h)dh}{\int_0^\infty \tilde{K}_i(h)dh}\ . \qquad [36]$$

Gaussian-like distributions (Figure S1) are approximated with a single block of size $h_0 = 2\langle h\rangle$ which defines the guest - surface interaction range. The factor 2 ensures that the calculation is self-consistent: if we insert the block approximation Eq. (34) into the above definition we recover the same value for h_0.

Using the above approximation Eq. (34) in Eq. (32) the integration becomes trivial and we obtain

$$\tilde{Q}_b = h_0 \prod_i \left(1 + \frac{e^{-\beta\Delta\tilde{G}_i^{cnf}}}{\rho_0 h_0}\sum_j e^{-\beta\Delta G_{ij}} c_j\right)^{m_i} + h_{\text{cutoff}} - h_0\ . \qquad [37]$$

In what follows, we chose the cutoff height equal to the interaction range $h_{\text{cutoff}} = h_0$. In a practical experiment the cutoff will most probably be determined by a particular technique used to measure the surface density of adsorbed particles. h_{cutoff} should not be smaller than the interaction range h_0. On the other hand, for very large h_{cutoff}, the term $h_{\text{cutoff}} - h_0$ cannot be neglected. However, this will only introduce a constant offset in the measured adsorbed density.

Finally, we define the equilibrium constant

$$K_{ij} \equiv \frac{e^{-\beta\Delta\tilde{G}_i^{cnf}}}{\rho_0 h_0} e^{-\beta\Delta G_{ij}}\ , \qquad [38]$$

which includes both the configurational Eq. (35) and the association ΔG_{ij} terms. The guest - surface interaction free energy due to bond formation can now be written in a simple form

$$e^{-\beta\Delta F_b} = Q_b = \frac{\tilde{Q}_b}{h_0} = \prod_i \left(1 + \sum_j c_j K_{ij}\right)^{m_i}\ . \qquad [39]$$

We also decompose the total free energy change Eq. (33) into binding part ΔF_b and ΔF_0 which captures the free energy of an unbound particle within lattice size of area A and height h_0:

$$\Delta F = \Delta F_b + \Delta F_0 = \Delta F_b + \Delta F_{ns} - k_B T \log(A h_0 \rho_0 N_A) \qquad [40]$$

where in the last step F_0 can be decomposed into non-specific interactions part ΔF_{ns} and $\log(A h_0 \rho_0 N_A)$ is the translational entropy of an unbound particle within h_0.

Alternative Mean-field definition. Rather than using a series of approximations in the above derivation, we can simply postulate Eq. (39) as the mean-field approximation for the free energy of the real system. Optimal values for the interaction range h_0 and $\Delta\tilde{G}_i^{cnf}$ are determined by fitting Eqs. (39, 38) to experimental (or simulation) data. h_0 determines the height cutoff within which an unbound guest particle must diffuse such that ligands can be treated as independently forming bonds with surface receptors.

The optimal cutoff interaction range h_0 can be obtained from simulations by calculating the free-energy profiles as a function of the number of formed bonds λ. Figure S2 shows such profiles for various choices of h_0 and two polymer lengths: **a)** $l_c = 1$ and **b)** $l_c = 4$. Evidently, changing h_0 affects the free energy of forming the first bond, however, for subsequent bonds the slope of free energy profiles is independent of h_0. To illustrate this Figures S2**c)**,**d)** show the corresponding second order difference $F''(\lambda) = F(\lambda+1) + F(\lambda-1) - 2F(\lambda)$. Second order difference is a very useful measure because it does not depend on the individual ligand binding strength; that is as long as the binding strength is a constant for all ligands. In the case studied the ligand profile and interactions strengths follow the design rules from the main text: $\sum_j c_j K_{ij} + 1 = e^{\lambda_p}$ for all ligands and analytical theory Eq. (39) results in

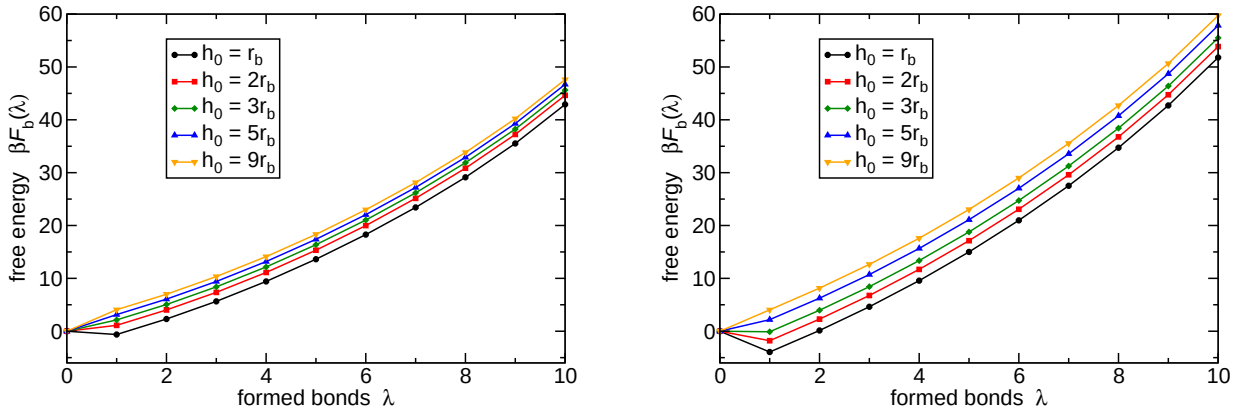
$$F_b(\lambda) = -\log\left[\binom{m}{\lambda}\left(e^{\lambda_p} - 1\right)^{\lambda}\right]\ . \qquad [41]$$

We note the distinction between the number of formed bonds λ and the Lagrange multiplier λ_p which defines the overall bonding strength; these particular characters are used for legacy reasons. Taking the second order difference we find

$$F_b''(\lambda) = \log\left[\frac{(\lambda+1)(m-\lambda+1)}{\lambda(m-\lambda)}\right]\ , \qquad [42]$$

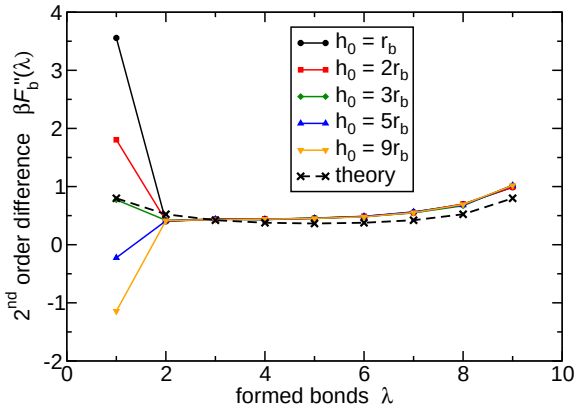
where we remember m as the number of ligands on a particle. Comparing simulation results to the theory (Figures S2**c)**,**d)**) the optimal value for h_0 can be obtained when the theoretical result matches simulations for the second order difference around $\lambda = 1$. The optimal h_0 is $h_0 = 3r_b$ for single blob ligands and $h_0 = 7r_b$ for longer 4-blob ligands, with r_b the blob radius.

Our previous estimate at the effective interaction range h_0 using the first moment Eq. (36) was larger: $h_0 = 3.66r_b$ and $h_0 = 9.48r_b$ for the shorter $l_c = 1$ and longer $l_c = 4$ ligands respectively. The first moment definition Eq. (36) provides a practical route to estimate h_0 without the need for free energy calculations. Alternatively, h_0 could simply be fitted from experiments as outlined below.

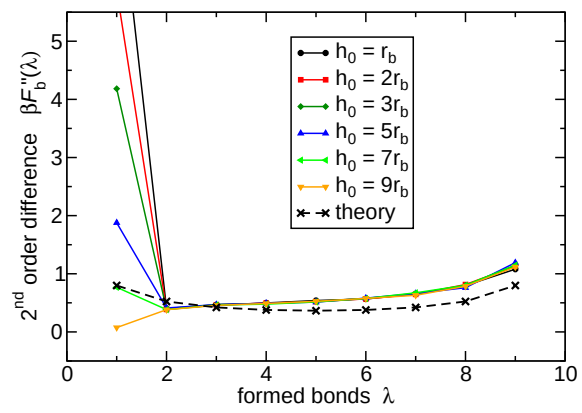


(a) ligand polymer length: $l_c = 1$

(b) ligand polymer length: $l_c = 4$



(c) $l_c = 1$



(d) $l_c = 4$

Fig. S2. Free energies from simulations. **a)** and **b)** show the configurational free energy of a guest particle as a function of the number of formed bonds λ (calculated with Wang-Landau technique at the bond energy $\epsilon = 0$) for various choices of effective interaction range h_0. **c)** and **d)** show the corresponding second order difference together with the theoretical result Eq. (42). The optimal $h_0 = 3r_b$ for short ($l_c = 1r_b$) polymers, and $h_0 = 7r_b$ for longer ($l_c = 4$) polymers, r_b is the blob radius. The number of ligands is always $m = 10$, particle radius $R = 3r_b$; all simulation parameters match Figure S1. Standard errors are smaller than symbols.

Analytical model assumes uncorrelated (independent) ligand binding. h_0 can always be chosen such such that the first and second bond are (or appear to be) uncorrelated. However, tuning h_0 only affects the first bond, subsequent bonds will generally show deviation from the mean-field uncorrelated model. Binding is cooperative (with respect to the uncorrelated ligand binding) if the free energy is below the theoretical value and anti-cooperative above it.

Figure S3**a)** illustrates that, for the guest particle model, the binding becomes cooperative at medium bond numbers ($\lambda \approx 5$) anti-cooperative at larger number of formed bonds ($\lambda \approx 9$). We speculate that such anti-cooperativity is caused by ligand-ligand repulsion. Furthermore, as can be seen on Figure S3**b)**, cooperivity in simulated guest particle binding leads to a small increase in the free energy curvature as compared to the uncorrelated theory. The same data is also used to obtain selectivities $\mathcal{S}$ (i.e. the relative curvature of the free energy) as shown on Figure 5 of the main text. The curvature is obtained by calculating a second order difference around the minimum as $F_b''(c^*) = \frac{1}{(\Delta c)^2}\left(F_b(c^* + \Delta c) + F_b(c^* - \Delta c) - 2F_b(c^*)\right)$, with $\Delta c = 0.1c_T$ and $F_b(c^*)$ is the minimum.

Langmuir adsorption. The free-energy expressions obtained above can be used with Langmuir adsorption model to calculate the experimentally relevant observable; the surface coverage with guest particles. Langmuir model assumes independent adsorption sites and at most a single particle per site, therefore the area A in Eq. (40) must be the excluded area of an adsorbed guest particle. The guest - host avidity association constant is

$$K_A^{av} = e^{-\beta\Delta F}/\rho_0 = Ah_0N_Ae^{-\beta\Delta F_b} \ . \qquad [43]$$

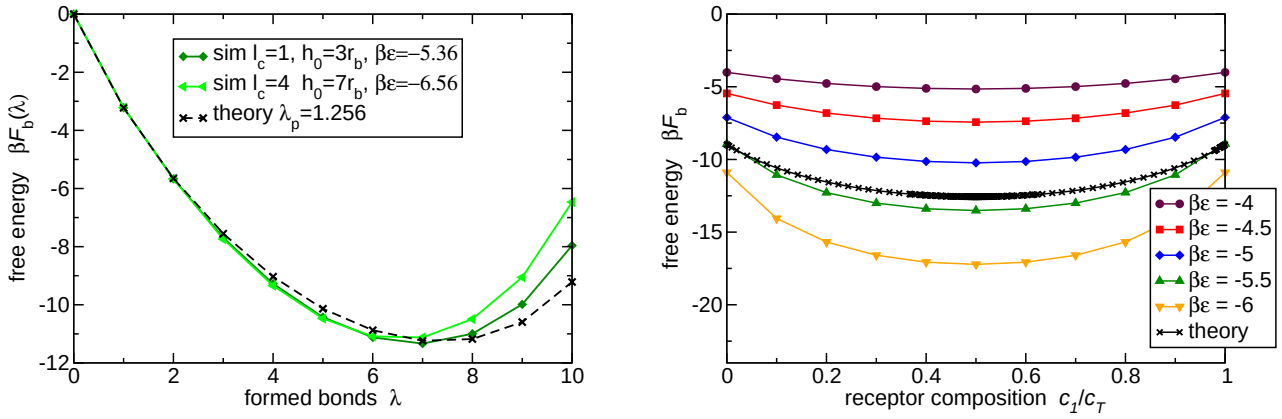


Fig. S3. Theory-simulation comparison. **a)** The free-energy dependance on the number of formed bonds. Optimal h_0 is used resulting in uncorrelated attachment of the first two bonds. ϵ is then determined by fitting to the theoretical result from the design rules, which ensures that the free energy for the first two bonds exactly matches the theory Eq. (41). Because all bonds are of equal strength ϵ, the free energy can be obtained by a linear combination: $F_b(\lambda|\epsilon) = F_b(\lambda|\epsilon = 0) + \lambda\epsilon$, with $F_b(\lambda|\epsilon = 0)$ from Figure S2. **b)** Free-energy as a function of the receptor composition $F_b(c_1)$ was obtained using Wang-Landau calculations (full symbols) and comparison with the theoretical result (black crosses) at $\lambda_p = 1.2563$. Wang-Landau calculations were performed on a single particle without any height constraints but having at least a single formed bond. The absolute value of the free energy is set by $\tilde{F}_b(c_1 = 0.5c_T|\epsilon) = -k_BT \log\left(\sum_{\lambda=1}^{m} e^{-\beta F_b(\lambda|\epsilon)}\right)$, which is then rescaled to include the non-bonded states $F_b = \log(1 + \exp(-\beta\tilde{F}_b))$. The number of ligands is always $m = 10$, particle radius $R = 3r_b$, system size is $10R \times 10R$ in the lateral directions with receptor concentration $c_T = 1/R^2$, $p_1 = c_1/c_T = 0.5$. Simulation parameters match Figure S1. Standard errors are smaller than symbols.

The Langmuir adsorption isotherm

$$\theta = \frac{\rho K_A^{av}}{1 + \rho K_A^{av}} , \qquad [44]$$

gives the number of adsorbed particle per lattice site area A, ρ is the concentration of guest particles in solution. If the solution cannot be treated as ideal, the activity of guest particles should be used instead of concentration. This Langmuir isotherm is shown on Figure 4 of the main text, with the excluded area $A = (2R + 2r_b)^2$ assuming the guest has an effective radius of particle R and the polymeric ligand blob size r_b. The interaction range $h_0 = R$ for single blob ligands from Figure S2.

3. Guide to fitting experiments

In a given multivalent system (e.g. a multivalent particle, linear polymer, star polymer, etc.), we assume that the association matrix between individual ligand-receptor types in solution, ΔG_{ij} or the affinity matrix $K_A^{ij} = e^{-\beta\Delta G_{ij}}$, is known. In principle we could then calculate the binding free energy, or equivalently, the avidity association constant of an adsorbing multivalent entity

$$K_A^{av}\rho_0 = e^{-\beta\Delta F} = e^{-\beta\Delta F_0} \times \prod_i \left(1 + \sum_j c_j K_{ij}\right)^{m_i} , \qquad [45]$$

with $\rho_0 = 1$M the standard concentration. In practice, however, both the zero-bond free energy $\Delta F_0 = \Delta F_{ns} - k_BT \log(a^2 h_0 \rho_0 N_A)$ and the configurational contribution $\frac{e^{-\beta\Delta\tilde{G}_i^{cnf}}}{\rho_0 h_0}$ might be difficult to calculate. But, they could simply be fitted from experiments.

A natural starting point is to neglect non-specific interactions ($\Delta F_{ns} = 0$) and assume that the multivalent guest can be described using the "cloud of ideal ligands" approximation. This approximation only takes into account translational entropy contribution and assumes that unbound ligands can freely explore the entire volume a^3 that is occupied by the multivalent guest; a^3 is the volume of the ligand 'cloud'. For example, in the case of flexible multivalent polymers this volume is equal to the effective volume of the polymer $a^3 = \frac{4\pi}{3}R_g^3$, with R_g the polymer radius of gyration. In the case of a particle based multivalent guests the volume a^3 should match the excluded volume of the particle. Within the "ideal ligand cloud" approximation we obtain $h_0 = a$ and $\Delta\tilde{G}_i^{cnf} = 0$.

Using this approximation and Eq. (38) we rewrite the expression for the binding avidity of the multivalent guest and introduce two fitting constants

$$K_A^{av} = A_{zero} a^3 N_A \prod_i \left(1 + \frac{A_{cnf}}{a} \sum_j c_j K_A^{ij}\right)^{m_i} . \qquad [46]$$

Where we remind ourselves that a is the lateral size of the multivalent guest, c_j the surface molar concentration of receptor of type j on the host, K_A^{ij} the interaction matrix specifying affinity equilibrium constants between a ligand type i and receptor type j from solution and m_i is the ligand valency; the number of ligands of type i on the multivalent guest. The above expression, therefore, predicts the binding avidity depending on the physico-chemical properties of the multivalent guest and the surface concentration of receptors.

The two dimensionless fitting constants A_{zero} and A_{cnf} capture the deviation of the real system from our "cloud of ideal ligands" estimate. Both fitting constants should be viewed as simple correction factors, however, the values of the correction factors need not be close to unity. Previous experiments on hyaluronic acid based multivalent polymers from Refs. (6, 7) determined the equivalent correction factor $U_{poly} = 4.6 k_B T$, which is related to $A_{cnf} = e^{-U_{poly}/k_B T}$.

Moreover, the ratio $\frac{A_{cnf}}{a^3}$ is related to the widely used "effective molarity" approach in rationalising multivalent interactions (19–22). The number of receptors within interaction area a^2 that a multivalent guest "sees" is $n_R^j = a^2 N_A c_j$, therefore, the second term in Eq. (46) can be rewritten by defining the effective molarity $EM = \frac{A_{cnf}}{a^3 N_A}$ which measures the configurational contribution of binding between a ligand and a *particular* receptor within interaction distance a.

Equivalently, some authors (11) fit the effective volume $V_{eff} = 1/EM$ and the 0-bond dissociation constant $K_{D0} = \rho_0 e^{\beta F_0} = \frac{1}{A_{zero} a^3 N_A}$.

4. Poisson fluctuations undermine specificity

We expect that, in practice, any nanoparticle fabrication technique will introduce some heterogeneity or poly-dispersity of the multivalent guest properties. For example, if the ligands are grafted to the particle or polymer by a purely random (Poisson) process, the ligand positions will be distributed on the particle uniformly at random. Moreover, the number of ligands of a specific type per particle will also vary and we expect it to be Poisson distributed. Multivalent guests will, therefore, exhibit heterogeneous binding. In this case it is instructive to calculate the expected value for the bound partition function. We average our expression for the partition function Eq. (39) over the Poisson distribution of the ligands on the particles

$$Q_b(\tilde{\mathbf{m}}, \mathbf{c}, \mathbf{K}) = \sum_{\mathbf{m}=0}^{\infty} \left[\prod_i \left(\frac{e^{-\tilde{m}_i} \tilde{m}_i^{m_i}}{m_i!} \right) Q_b(\mathbf{m}, \mathbf{c}, \mathbf{K}) \right] , \qquad [47]$$

where $Q_b(\tilde{\mathbf{m}}, \mathbf{c}, \mathbf{K})$ is the bound partition function from Eq. (39) and we explicitly wrote as a function of the ligand profile $\mathbf{m}$, the receptor composition $\mathbf{c}$ ahe interaction matrix $\mathbf{K}$. We assume that every ligand type valency m_i is Poisson distributed with mean $\tilde{m}_i$, $\tilde{\mathbf{m}}$ denotes the mean ligand profile vector and $\sum_{\mathbf{m}=0}^{\infty}[\cdot] \equiv \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \dots [\cdot]$ is a nested sum over all m_i.

Inserting Eq. (39) into the expression above and swapping the product and summation order we get

$$Q_b(\tilde{\mathbf{m}}, \mathbf{c}, \mathbf{K}) = \prod_i \left(e^{-\tilde{m}_i} \sum_{m_i=0}^{\infty} \left[\frac{\left[\tilde{m}_i \left(1 + \sum_j c_j K_{ij} \right) \right]^{m_i}}{m_i!} \right] \right) , \qquad [48]$$

where the inner sum can be recognised as the Taylor expansion for the exponential function. Therefore, the final result can be written as a product of independent exponential functions

$$Q_b(\tilde{\mathbf{m}}, \mathbf{c}, \mathbf{K}) = \prod_{i,j} e^{-\tilde{m}_i K_{ij} c_j} . \qquad [49]$$

We call this form the double exponential form because inserting Eq. (38) would yield a double exponential dependance on the bond free energy ΔG_{ij}.

The total binding free energy of this system becomes simply a sum over all possible bond pairs

$$\Delta F_b = -k_B T \log Q(\tilde{\mathbf{m}}, \mathbf{c}, \mathbf{K}) = k_B T \sum_{i,j} \tilde{m}_i c_j K_{ij} = k_B T \, \tilde{\mathbf{m}}^T \mathbf{K} \mathbf{c} . \qquad [50]$$

In the last form on the right is cast using matrix algebra with $\tilde{\mathbf{m}}^T$ being the transpose of vector $\tilde{\mathbf{m}}$.

We stress that for Poisson distributed ligands the binding free energy evidently becomes linear in the receptor composition $\mathbf{c}$. Therefore, the binding free energy can never exhibit a minimum at an arbitrarily chosen composition $\mathbf{c}_0$, regardless of the the ligand profile $\mathbf{m}$ and the interaction matrix $\mathbf{K}$. Hence, for composition specific targeting we need a precise control over multivalent guest fabrication process and synthesis. Multivalent guests must have a well-defined ligand profile with fluctuations in the profile much smaller than the expected Poisson fluctuations. An ensemble of guests with Poisson distributed ligands is not sufficiently selective for specific receptor composition targeting. Therefore, it appears that DNA origami constructs (24), where the geometry of the nano-construct can be almost exactly controlled, would be best suited for receptor composition targeting.

5. Endocytosis multivalent theory

Receptor-mediated particle endocytosis is a slightly more intricate process than mere adsorption. In fact, adsorption to the membrane is usually only the first step, upon which, endocytosis may proceed. Passive endocytosis is a concerted action of membrane bending, receptor recruitment and attachment, neck formation and snapping. We will not try to characterise individual steps in this process. Rather, we only calculate the total free energy which is the main driving force and determinant of passive endocytosis. We show that the free energy change upon receptor-mediated passive endocytosis is governed by the same expression Eq. (23) as multivalent adsorption.

For the derivation we assume that receptors are mobile within the fluid membrane and endocytosis is slow compared to lateral receptor diffusion, hence, membrane receptors remain in a quasi-equilibrium with the ligand-decorated particle during endocytosis. The free energy change of such process ΔF will be exactly the same as if, the particle is first endocytosed without any receptors present, and subsequently the endocytose membrane shell is put in a contact with a reservoir of membrane receptors. Similarly to Eq. (40) the free energy decouples

$$\Delta F = \Delta F_0 + \Delta F_b \, . \qquad [51]$$

ΔF_0 is the no-receptor endocytosis free energy change, it includes the bending penalty and any membrane-particle interactions that are not mediated by receptors. ΔF_b captures all ligand-receptor interactions. We assume that adding receptors does not change the physical properties of the membrane; elastic moduli are constant. Therefore ΔF_0 is a constant independent of receptor compositions. The only remaining step is to show that the receptor interaction part ΔF_b is given by Eq. (23).

Individual ligands attached on the guest particle are in contact with a bath of membrane receptors. Therefore, akin to Eq. (1) in the main text, we can immediately write the ratio of bound-to-unbound grand partition functions for a single ligand

$$\frac{q_b^i}{q_u^i} = \sum_j c_j K_{ij}^A \, \frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{h_0} \, , \qquad [52]$$

where the sum proceeds over all receptor types j. We remember c_j denotes receptor concentration, K_{ij}^A the dimerisation affinity matrix from solution, $\Delta \tilde{G}_i^{cnf}$ is the configuration free energy change upon bond formation and h_0 the length scale. The configurational part $\Delta \tilde{G}_i^{cnf}$ captures possible effects like of pinning the membrane (25) to the particle upon bond formation; effects not present in the solution dimerisation equilibrium constant K_{ij}^A. The length scale h_0 is the thermal roughness (i.e. the root mean square membrane fluctuations in the radial direction) of the wrapped membrane with no bonds present.

Unless strong non-linear pinning effects are present, the correlations between different ligand attachment can be neglected. Therefore, the total partition function of a wrapped guest particle can be written as product over individual ligands. By defining the effective interaction matrix $K_{ij} = K_{ij}^A \, \frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{h_0}$ we obtain a familiar expression

$$Q_b = \prod_i \left(1 + \sum_j c_j K_{ij} \right)^{m_i} \qquad [53]$$

which is identical to Eq. (39).

To derive the above result we presumed that endocytosis is slow compared to lateral receptor diffusion and receptor-ligand binding equilibrates. However, the above expression remains if we relax this assumption. In the opposite limit (slow receptor diffusion) receptors cannot equilibrate and their number is effectively fixed on any given membrane patch. One need only follow the same derivation as in section "Free energy derivation for immobile receptor" above to show that Eq. (53) gives the average partition function for slowly diffusing, but randomly distributed, receptors. Since both limits return the same answer, any intermediate ratio between endocytosis rate and receptor diffusion must result in the same expected value for the binding partition function Eq. (53).

6. Supporting results

Figure S4 shows that optimal value of the binding strength λ_p is independent of the cross-binding terms in $\mathbf{K}$, however, selectivity is diminished. Figure S4**b)** also shows how the optimal ligand profile differs from our simple design rule (guest profile should match the composition: $p_i \propto c_i$) if the cross-binding is strong and the interaction matrix $\mathbf{K}$ is not symmetric. In the particular example on Figure S4**b)**, the type 1 ligand can bind to both receptor types, but the type 2 ligand can only bind to receptor type 2:

$$\mathbf{K} = \begin{pmatrix} K_{11} & K_{11}\kappa_{12} \\ 0 & K_{22} \end{pmatrix} \, , \qquad [54]$$

with the relative cross-binding term $\kappa_{12} = \frac{K_{12}}{K_{11}}$.

Thus far the total concentration of receptors c_T was kept fixed. In Figure S5 we show how varying the total receptor concentration effects guest binding. As expected, increasing the overall receptor concentration (at fixed composition $\mathbf{c}$) will always lead to stronger interaction. Therefore, the total receptor concentration on a targeted cell should not be orders of magnitude smaller compared to un-targeted cells. As a rule of thumb: For any un-targeted cell, there must exist at least

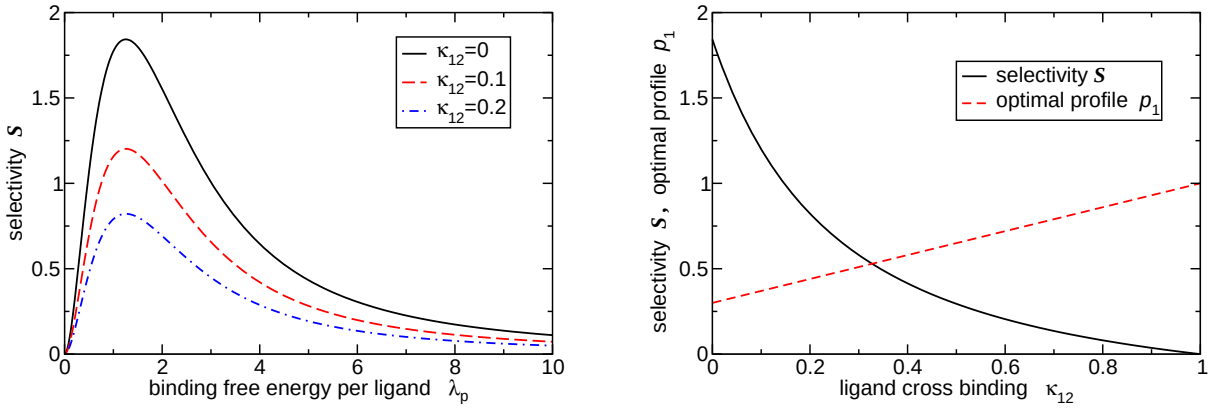


Fig. S4. Selectivity of targeting with 2 ligand types and the effect of cross-binding. **a)** The selectivity $\mathcal{S}$ as a function of bond strength (Lagrange multiplier) λ_p for different magnitudes of cross binding κ_{12}, with fixed $\kappa_{21} = 0$. Cross binding diminishes the selectivity, but the optimal $\lambda_p^{opt} \approx 1.25643$ remains a constant. **b)** The selectivity $\mathcal{S}$ at optimal λ_p^{opt} (the peak value in **a**)) decreases monotonically with cross binding terms κ_{12}. We also plot the optimal ligand profile p_1, calculated by solving Eqs. (9,10) as a function of the cross binding term κ_{12}. Parameters: composition $c_1 = 0.3c_T$, ligand cross binding $\kappa_{21} = 0$, $\kappa_{11} = \kappa_{22} = 1$ by definition.

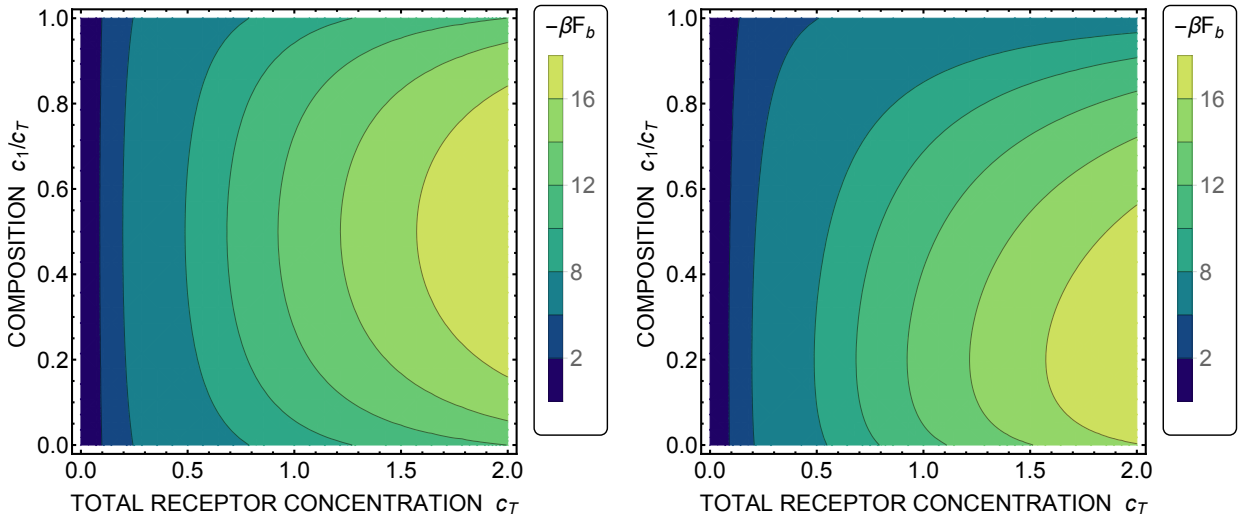
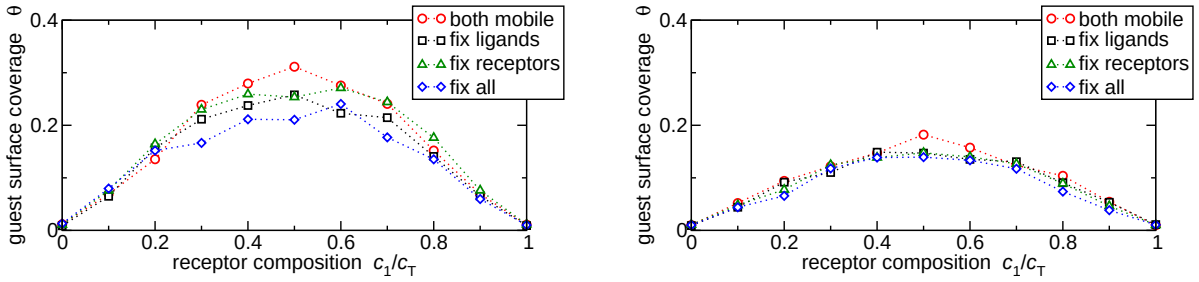


Fig. S5. Effect of changing the total receptor concentration c_T. The two landscape plots show the binding free energy F_b from the analytical theory (Eq. (2) in the main text) for a 2 ligand/receptor type system as a function of the composition c_1 and total receptor concentration c_T. The guest with $m = 10$ ligands was optimised to target a membrane at $c_T = 1$ and composition **a)** $c_1^* = 0.5c_T$ and **b)** $c_1^* = 0.2c_T$. Using the design rules from the main text the guest properties are: $p_i = c_i^*/c_T$ and $K_{ii} = c_T(e^{\lambda_p} - 1)/p_i$, with $\lambda_p = 1.2543$. $K_{ij} = 0$, for $i \neq j$.

one receptor type that is over-expressed on a targeted cell compared to the un-targeted cell. Specific cell-targeting based on receptor recognition is only possible when this condition is satisfied, see Figure 1 of the main text for a pictorial representation.

Simulation results of guest adsorption presented in the main text (Figure 4) were performed with mobile receptors and ligands. Figure S6 shows that mobility is not necessary for composition targeting. Keeping either ligands or receptors (or both) immobile retains the selectivity.

Further numerical optimisation. In our analytical treatment we required that the free energy must be a minimum with respect to the variation of the ligand profile (Eq. (5) in the main text). As already discussed, this might be desirable in practice, but is in principle not a necessary constraint. Removing the constraint somewhat increases the selectivity. Figures S7 and S8 shown the comparison between our reference solution determined by the design rules and the optimal unconstrained solution obtained numerically. Both Figures clearly show that the further optimisation results in a more skewed ligand profile and binding strengths. This solution might be unpractical and hard to realise in a fabrication process while the optimal selectivity



(a) high receptor concentration: $c_T = 7.5/R^2$, low affinity: $\epsilon = -3.5k_BT$ **(b)** low receptor concentration: $c_T = 1/R^2$, high affinity: $\epsilon = -5.5k_BT$

Fig. S6. Effect of receptor and ligand mobility. Immobile receptors are (uniformly) randomly distributed on the surface. Fixed ligands are anchored to an immobile point on the guest particle surface, anchor point locations obey the Boltzmann distribution. **a)** shows the large receptor concentration case; mobile data (red circles) is the same as data on Figure 4 in the main text. **b)** low receptor concentration. Each data point is an average over 5 simulation runs. Parameters correspond to and are described in Figure 4 in the main text.

only increases marginally.

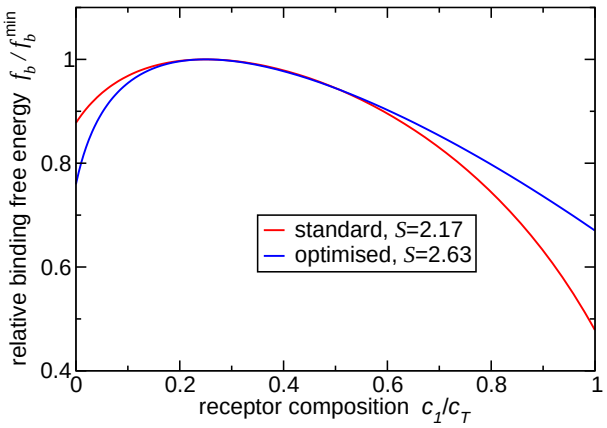


Fig. S7. Optimising the selectivity, 2 ligand/receptor types. The plot shows the binding free energy f_b/f_b^{min} normalised by the minimum. The targeted composition is at $\mathbf{c}^* = [0.25, 0.75]$. The standard result for the design of the particle is $p_1^* = 0.25$, $K_{11}^* = 10$, $K_{22}^* = 3.34$. The optimal numerically obtained solution is more skewed $p_1^{opt} = 0.07$, $K_{11}^{opt} = 26.3$, $K_{22}^{opt} = 0.34$ and the matching free energy per bond $f_b^{min,opt} = 0.34k_BT$ is smaller as compared to the standard result. For simplicity we assumed $c_T = 1$, however, a different value would simply uniformly rescale all resulting K_{ij} values and the plot would remain the same.

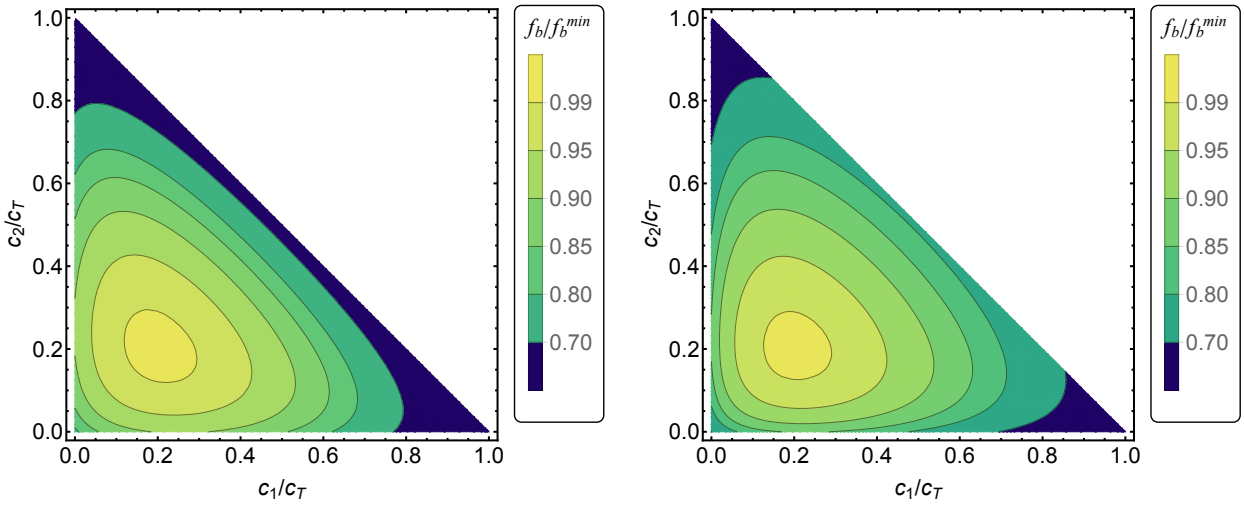
7. Polymer coated particle model

The multivalent particle model used in Monte Carlo simulations is an extension of the model from Ref. (5). Guests are represented as hard spheres with attached polymeric ligands using a coarse-grained polymer model (16). A polymer is represented with a series of Gaussian soft blobs connected with harmonic springs. A great feature of the soft blob model are transferable potentials, we can represent a given polymer by many small blobs, a few larger ones or a single large blob and the interaction potentials do not change. Each polymer chain is represented by N_b soft repulsive blobs with radius of gyration r_b that are connected via harmonic springs

$$U_{ch} = 0.534\, k_BT\, (r/r_b - 0.730)^2\ , \qquad [55]$$

with k_BT the thermal energy and r the centre-to-centre distance. The blob-blob interaction is described as a Gaussian repulsion

$$U_{bb} = 1.75\, k_BT\, e^{-0.80(r/r_b)^2}\ , \qquad [56]$$



(a) standard: $p_1^* = p_2^* = 0.2$, $K_{11} = K_{22} = 12.6$, $K_{33} = 4.2$, $\mathcal{S} = 7.4$

(b) optimised: $p_1^{opt} = p_2^{opt} = 0.104$, $K_{11} = K_{22} = 23$, $K_{33} = 0.8$, $\mathcal{S} = 8.7$

Fig. S8. Optimising the selectivity, 3 ligand/receptor types, analogous to Figure S7. Relative binding free energy depending for: **a)** standardly designed guest particle and **b)** unconstrained solution selectivity by relaxing the. Targeted receptor composition is $\mathbf{c}^* = [0.2, 0.2, 0.6]$ and total concentration $c_T = 1$.

while the blob-surface interaction is modelled as an exponential repulsion:

$$U_{bs} = 3.20\, k_B T\, e^{-4.17(r/r_b - 0.50)}\ . \qquad [57]$$

where r is the perpendicular distance between the blob centre-of-mass and a hard surface. The same potential is also used for blob-particle repulsion. This model accurately describes flexible self-avoiding-walk (SAW) polymers in the scaling regime

Polymers are attached to an anchoring point on a particle surface with a tethering potential (27)

$$U_{tether} = 0.75\, k_B T\, (r/r_b)^2 \qquad [58]$$

which captures the entropic penalty of tethering a flexible polymer to an impenetrable flat surface. The anchoring points can either have fixed positions or the anchors are free to diffuse on the guest surface. Figure S9 presents the ligand coated guest particle model.

Polymer conformations are sampled using standard Monte Carlo single-blob translational moves. The grand-canonical part of the algorithm (insertion and deletion of guest particles) is employed via Rosenbluth-sampling with configurational bias (28). Because multiple polymers can be grafted to a single particle, the Rosenbluth weight of the particle with all tethered polymers must be calculated as $W_{tot} = \prod_i W_i$, with W_i the Rosenbluth weight of individual polymers.

Unless explicitly noted, all simulations were performed with both mobile receptors and ligand anchors. For a single guest - surface interaction, if the dynamical distribution of mobile ligand positions on each guest matches the distribution of fixed ligands on an ensemble of guests and ergodicity is obeyed, the mobile and immobile cases will result in the same average interaction free energy:

$$F^{mobile} = -k_B T \log \left\langle e^{-\beta F^{fix}(\mathbf{r}^a)} \right\rangle_{\mathbf{r}^a}\ , \qquad [59]$$

where $F^{fix}(\mathbf{r}^a)$ is the free energy for given guest with anchor positions $\mathbf{r}^a$. When averaged over an ensemble of guests with different random distributions of $\mathbf{r}^a$ we recover the free energy obtained for a single guest with mobile anchors F^{mobile}. The same argument holds for receptor mobility.

Monte Carlo sampling of valence-limited interactions. Valence limited interactions are employed similarly to polymer adsorption model in Ref. (7). Further examples of modelling valence limited interactions can be found in Refs. (5,29,30).

The last blob in a polymer chain carries a ligand. Hence, guest particle can be viewed as having multiple ligand arms that can "grab" the receptors. The ligand carying blob is identical to other blobs except that it can attach to receptors with binding free energy $\epsilon_{ij} + U_{bond}$, where ϵ_{ij} is a constant interaction matrix specifying interactions between all ligand/receptor types. Tethering potential $U_{bond} = U_{tether}$ (Eq. (58)) captures the bond stretching penalty. Binding is valence limited, one receptor can bind to only a single ligand and vice versa. Receptors are modelled as immobile point-like objects that are randomly distributed on a hard impenetrable surface, there are no interactions between blobs and receptors other then ligand-receptor binding.

A particular state in our simulations is determined by the vector positions of all blobs and arrangement of ligand-receptor bonds. For efficient sampling of states we employ two types of Monte Carlo moves: particle insertion/deletion moves, and single blob/particle translational moves integrated with ligand-receptor binding. Only non-bound guests can be inserted or

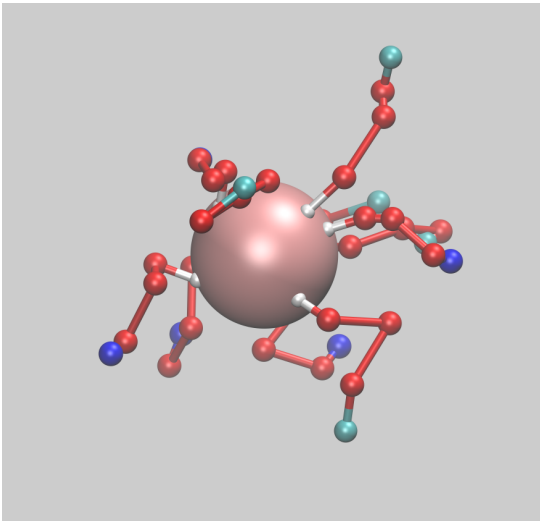


Fig. S9. Simulation snapshot outlining the multivalent guest particle model. The particle is the large sphere, 10 white anchoring points are fixed on the particle's surface. Each anchor has a polymer tethered to it, where polymers are modelled as a series of red blobs connected into a chain. The last blob represents the ligand. 2 distinct ligand types are shown coloured cyan (type 1) or blue (type 2).

deleted. Each adsorption simulation started with an empty box and lasted for about $\approx 10^{10}$ MC cycles, where in each cycle we randomly select to either insert or delete a particle (with probability $p_{ins-del} = 1/(mN_b + 1)$) or move a single blob (with probability $p_{hop} = mN_b/(mN_b + 1)$), with mN_b the total number of blobs per guest. In the adsorption simulations the average number of bound guest particles was determined by averaging over the second half of the simulation run time.

In order to speed up the sampling of the polymer conformations the binding between ligands and receptors is integrated within the blob translational moves. This overcomes the bottleneck which arises when individual blobs are strongly bound to surface receptors. A ligand blob is chosen uniformly at random and all its ligands are made unbound. The binding partition function which counts all possible binding arrangements of ligand i is a sum over all receptors

$$q_i^o = 1 + \sum_j e^{-\beta(\epsilon_{ij} + U_{bond}(r))} \,. \qquad [60]$$

with the bond stretching penalty given by $U_{bond} = U_{tether}$ Eq. (58). The first term on the right (1) captures the unbound ligand state. In our simulation implementation only free receptors within a distance cutoff $r_{bondcut} = 5r_b$ are considered. A new trial position for the given blob is considered and its new internal binding partition function q_b^n is calculated. The move is then accepted with probability

$$p_{o\to n} = \min\left[1, \frac{q_b^n}{q_b^o}\, e^{\beta(U_o - U_n)}\right], \qquad [61]$$

where U_o and U_n are the old and new potential energies of the system determined by Eqs. (55, 56, 57, 58). Regardless of whether the move was accepted or not the blob of interest is still unbound. We now randomly choose a binding state proportional to its Boltzmann weight directly from the partition function q_i^o (rejected move) or q_i^n (accepted move).

8. Free energy derivation for immobile receptors

In the case of mobile receptors the expression for the bound partition function of a multivalent particle to the receptor decorated membrane is given by Eq. (39). Here we show that the same expression is obtained as an expected value even when receptors have fixed positions (for example, due to being attached to the cytoskeleton), provided that receptors are randomly distributed and the surface coverage with guest is small.

We start with the bound partition function of a multicomponent guest binding to a receptor decorated host surface. The positions of receptors and ligands are designated with vectors $\mathbf{r}_R$ and $\mathbf{r}_L$, respectively, n_R and m are the total number of receptors and ligands, respectively. As before, we will denote individual ligands with a prime i' and individual receptors with j' , therefore, $\mathbf{r}_{j'}$ designates a position of receptor j' and $\mathbf{r}_{i'}$ a ligand i' position. We will use a convention that a prime on a script i' denotes a particular binder, while a standard subscript i refers to the type of a binder.

The partition function counting all possible binding configurations for fixed positions of receptors and ligands is a sum over all possible number of bonds λ, and a sum over all possible configurations with λ bonds, $s(\lambda)$. The $\prod_{i'j'(s)} e^{-\beta\Delta G_{ij}} e^{-\beta\Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'})}$ is the Boltzmann factor, with $\Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'})$ the configurational contribution to the bond already discussed in detail above. We have assumed that individual bonds are uncorrelated and the Boltzmann factor is factorised by individual bonds $i'j'(s)$ present in the given binding configuration s. The pair $\{i'j'\}$ defines a bond. We implicitly assume an existence of an indicator function mapping any individual i' or j' to their type i or j; we write ΔG_{ij}, not $\Delta G_{i'j'}$.

The bound partition function of a single ligand linker grafted at position $\mathbf{r}^a_{i'}$ and bound to a specific receptor j' is

$$q_b(\mathbf{r}^a_{i'};\mathbf{r}_{j'}) = e^{-\beta\Delta G_{ij}} e^{-\beta\Delta G^{cnf}(\mathbf{r}_{j'},\mathbf{r}^a_{i'})}. \quad [62]$$

In the bound state the location of the ligand position $\mathbf{r}_{i'} = \mathbf{r}_{j'}$ is the same as the location of the receptor. Similarly to the mobile receptor case above Eq. (25), we proceed by integrating the ligand anchor point over the particle surface, the bound partition function of a ligand of type i bound to receptor j' with the (nano)particle at position $\mathbf{r}_{np}$ is

$$q^i_b(\mathbf{r}_{np};\mathbf{r}_{j'}) = \int_C q_b(\mathbf{r}^a_{i'};\mathbf{r}_{j'})d\mathbf{r}^a_{i'} = e^{-\beta\Delta G_{ij}} \int_C e^{-\beta\Delta G^{cnf}(\mathbf{r}_{j'},\mathbf{r}^a_{i'})} d\mathbf{r}^a_{i'} \ , \quad [63]$$

with $\int_C$ an integral over the particle surface and an index i only means that, when integrated over the particle surface, any ligand of type i will yield identical bound partition function. The unbound partition function q^i_u is not affected by receptor mobility and is given by Eq. (26) derived above for the mobile receptors case. The ratio of partition functions determines the ratio of probabilities of ligand i being bound to a particular receptor j' to being unbound

$$\frac{p^{ij'}_{\text{bound}}}{p^i_{\text{unbound}}} = \frac{q^i_b(\mathbf{r}_{np};\mathbf{r}_{j'})}{N_A q^i_u} = e^{-\beta\left(\Delta G_{ij}+\Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})\right)} \ . \quad [64]$$

Note that this ratio depends on the exact position of the particle $\mathbf{r}_{np}$, not only on the particle height h. We have used the Avogadro's number N_A in the denominator because q^i_u was defined with molar units Eq. (26) via the standard concentration ρ_0. $e^{-\beta\Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})}$ is the integrated configurational cost of forming a bond to a receptor j' with respect to the unbound state, for a particle at position $\mathbf{r}_{np}$.

The partition function of the whole particle at position $\mathbf{r}_{np}$, normalised by the unbound particle in the solution can be written as

$$\frac{Q(\mathbf{r}_{np},\mathbf{r}_R,)}{Q(h=\infty)} = Q_b(\mathbf{r}_{np},\mathbf{r}_R) = \sum_{\lambda=0}^{m}\sum_{s(\lambda)} \exp\left\{ \sum_{\{i'j'\}(s)} -\beta\left(\Delta G_{ij} + \Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})\right)\right\} \ , \quad [65]$$

a sum over all possible number of bonds λ, a sum over all possible bonding configuration $s(\lambda)$ specifying which ligand is bound to which receptor, and finally a sum inside the exponential over all formed bonds $\{i'j'\}$ to obtain the free energy of the bonding configuration, relative to the unbound state.

Integrating over lateral positions of the particle we obtain the average bound partition function for fixed receptors and a particle at height h above the surface

$$\begin{aligned} Q_b(h,\mathbf{r}_R) &= \frac{1}{S}\int_S Q_b(\mathbf{r}_{np},\mathbf{r}_R)\delta(z_{np}-h)d\mathbf{r}_{np} \quad [66]\\ &= \frac{1}{S}\int_S \sum_{\lambda=0}^{m}\sum_{s(\lambda)} \exp\left\{ \sum_{\{i'j'\}(s)} -\beta\left(\Delta G_{ij} + \Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})\right)\right\}\delta(z_{np}-h)d\mathbf{r}_{np} \ , \end{aligned}$$

where the delta function $\delta(z_{np}-h)$ keeps the particle at specified height h. This is a very hard expression to evaluate: firstly, we must explicitly consider all possible bonding arrangements for each and every particle position $\mathbf{r}_{np}$, we cannot assume independent binding as in the mobile case Eq. (30) because a ligand bound to specific receptor j' will prevent another ligand from binding to the same receptor. Secondly, we must integrate over the whole surface.

To form a connection between mobile and immobile receptors we essentially make use of the ergodic hypothesis: the time average of mobile receptors binding to a guest particle (and hence the spatial average over all receptor positions) will be the same as the spatial average over all possible particle positions on the surface with immobile, but randomly distributed receptor positions

$$\frac{1}{S^n}\int_{S^n}(d\mathbf{r}_R)^n Q_b(\mathbf{r}_{np},\mathbf{r}_R) \approx \frac{1}{S}\int_S d\mathbf{r}_{np} Q_b(\mathbf{r}_{np},\mathbf{r}_R) \ , \quad [67]$$

with n the total number of receptors within the surface S. For an infinitely large surface the two integrals will yield an identical result, on the other hand, in a finite sized system with a given $\mathbf{r}_R$ the relation is only approximate. However, the integral over mobile receptors (left hand side of Eq. (67)) will always yield an expected value for $\langle Q_b(h,\mathbf{r}_R)\rangle$ if only the number density of immobile receptors, but not their exact positions $\mathbf{r}_R$, is known.

In the following, we will focus on evaluating the left hand side integral Eq. (67) for an infinitely large surface and prove that it equals to Eq. (30), in this way we also show how a mobile receptor system can be derived by starting from fixed receptor positions $\mathbf{r}_R$. The integral over all receptor positions factorises

$$\begin{aligned} Q_b(h) &= \frac{1}{S^n}\int_{S^n}(d\mathbf{r}_R)^n Q_b(\mathbf{r}_{np},\mathbf{r}_R) \quad [68]\\ &= \sum_{\lambda=0}^{m}\sum_{s(\lambda)}\prod_{\{i'j'\}(s)} e^{-\beta\Delta G_{ij}}\frac{1}{S}\int_S e^{-\beta\Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})}d\mathbf{r}_{j'} \ , \end{aligned}$$

we have used Eq. (65) and converted a sum in the exponential into a product of exponentials. Because the integral over receptor positions decouples, we only need to consider the bound receptors, all unbound receptors contribute a factor of 1 as we use a normalised Eq. (65) partition function. Using Eq. (64) we recognise the integral on the right hand side is the same as Eq. (28) defining $\tilde{K}_i(h)$. Hence, we can write the partition function as

$$Q_b(h) = \sum_{\lambda=0}^{m} \sum_{s(\lambda)} \prod_{\{ij\}(s)} \frac{\tilde{K}_i(h)}{N_A S} e^{-\beta \Delta G_{ij}} \ , \tag{69}$$

which returns a similar expression to Eq. (65), however, explicit dependance on the guest particle and specific receptor positions has been integrated out. The product can now be factorised over types $\{ij\}$. The surface area S should be large enough such that the probability of a particle forming bonds outside this area is negligible, as we shall see below, the expected value of $Q_b(h)$ depends on the receptor density $\frac{n}{S}$, and not on the value of S used in the calculation. Eq. (69) is still very hard to evaluate because we must consider all possible bonding arrangements between n receptors.

The total number of possible bonding arrangements $s(\lambda)$ can be written in terms of the multinomial distribution. We define the number of states Ω for a given number of formed bonds between ij ligand/receptor types λ_{ij}. Note that $\boldsymbol{\lambda}$ is a matrix. The partition function is a sum over all possible matrices $\boldsymbol{\lambda}$

$$Q_b(h) = \sum_{\boldsymbol{\lambda}=0}^{\mathbf{m}} \Omega(\boldsymbol{\lambda}) e^{-\sum_{ij} \beta \epsilon_{ij} \lambda_{ij}} \ . \tag{70}$$

where, for clarity of expressions below, we have defined an effective bond strength as

$$e^{-\beta \epsilon_{ij}} \equiv \frac{\tilde{K}_i(h)}{N_A S} e^{-\beta \Delta G_{ij}} \tag{71}$$

and the sum represents a nested sum over all distinct receptor ligand pairs $\{ij\}$

$$\sum_{\boldsymbol{\lambda}=0}^{\mathbf{m}} [\cdot] = \sum_{\lambda_{11}=0}^{m_1} \sum_{\lambda_{12}=0}^{m_2} \cdots \sum_{\lambda_{21}=0}^{m_1} \sum_{\lambda_{22}=0}^{m_2} \ldots [\cdot] \tag{72}$$

to account for all possible states of distinct bonding arrangements. We note that the maximum term in each sum is set to m_i the number of ligands of type i on the particle, this choice was made for later convenience. As we will see below the density of states Ω will be defined to vanish whenever the number of bonds exceeds the number of receptors: $\Omega = 0$, if there exists a receptor type j such that $\sum_i \lambda_{ij} > n_j$.

Single bond type. Let us first solve the problem in the case of a single ligand and receptor type and calculate the bound partition function

$$Q_b(n, m, \epsilon) = \sum_{\lambda=0}^{m} \Omega(\lambda) e^{-\beta \epsilon \lambda} \tag{73}$$

we have dropped the functional dependance on h for clarity and write it as a function of the number of receptors n and ligands m, and the bond strength ϵ. The dependance on guest height h is implicitly accounted for through ϵ, which is itself a function of h Eq. (71). The density of states is given in terms of binomial coefficients

$$\Omega(\lambda) = \binom{n}{\lambda} \binom{m}{\lambda} \lambda! \tag{74}$$

because we need to choose λ bonds out of n receptors, λ bonds out of m ligands and there are $\lambda!$ ways of binding the chosen receptors/ligands together. We are focusing on the case of a guest particle binding to a host cell, cell being much larger than the particle. The most unbiased assumption we can make for randomly distributed receptors on a host cell, is that the distribution of receptors will be Poisson [†] distributed within every chosen surface S. Therefore, we now Poisson average the partition function Eq. (73) over the number of receptors

$$\begin{aligned} Q_b(\tilde{n}, m, \epsilon) &= \sum_{n=0}^{\infty} \frac{e^{-\tilde{n}} \tilde{n}^n}{n!} Q(n, m, \epsilon) \\ &= e^{-\tilde{n}} \sum_{n=0}^{\infty} \frac{\tilde{n}^n}{n!} \sum_{\lambda=0}^{m} \binom{m}{\lambda} e^{-\beta \lambda \epsilon} \frac{n!}{(n-\lambda)!} \\ &= e^{-\tilde{n}} \sum_{\lambda=0}^{m} \binom{m}{\lambda} e^{-\beta \lambda \epsilon} \sum_{n=0}^{\infty} \frac{\tilde{n}^n}{(n-\lambda)!} \ , \end{aligned} \tag{75}$$

[†] strictly, the distribution will be binomial

where in the second line we have inserted Eq. (74) and in the third line we have swapped the summation order, which is allowed as the variables n and λ are independent. $\tilde{n} = cS$ denotes the mean number of receptors in area S and c is the overall concentration. The last sum can be rewritten by introducing $x = n - \lambda$

$$\sum_{n=0}^{\infty} \frac{\tilde{n}^n}{(n-\lambda)!} = \tilde{n}^\lambda \sum_{x=-\lambda}^{\infty} \frac{\bar{n}^x}{x!} = \tilde{n}^\lambda e^{\tilde{n}} \ , \qquad [76]$$

which is simply a Taylor expansion for the exponential function because terms with negative x are automatically zero by the definition of the factorial (or Γ) function. Inserting into Eq. (75) we get

$$Q_b(\tilde{n}, m, \epsilon) = \sum_{\lambda=0}^{m} \binom{m}{\lambda} \tilde{n}^\lambda e^{-\beta\lambda\epsilon} \ , \qquad [77]$$

a binomial expansion for the function

$$Q_b(\tilde{n}, m, \epsilon) = (1 + \tilde{n}e^{-\beta\epsilon})^m \ . \qquad [78]$$

Inserting our definition of ϵ Eq. (71) we find

$$Q_b(h) = \left(1 + c\tilde{K}(h)\, e^{-\beta\Delta G}\right)^m \ , \qquad [79]$$

which is precisely the expression obtained in the mobile receptor case above, Eq. (30), applied to a single ligand/receptor type.

Multiple components general derivation. We now derive a general expression for the bound state partition function for any number of different ligand/receptor types. The procedure will be very similar to the one presented above for a single component case. We will show that

$$Q_b(\tilde{\mathbf{n}}, \mathbf{m}, \boldsymbol{\epsilon}) = \prod_i \left(1 + \sum_j n_j e^{-\beta\epsilon_{ij}}\right)^{m_i} = \langle Q_b(\mathbf{n}, \mathbf{m}, \boldsymbol{\epsilon})\rangle_{\mathbf{n}} \qquad [80]$$

where $\langle\cdot\rangle_{\mathbf{n}}$ denotes a Poisson average over all elements in $\mathbf{n}$ and $\tilde{\mathbf{n}} = \langle\mathbf{n}\rangle$ is the average of receptor compositions. We continue from Eq. (70)

$$Q_b(\mathbf{n}, \mathbf{m}, \boldsymbol{\epsilon}) = \sum_{\boldsymbol{\lambda}}^{\mathbf{m}} \Omega(\boldsymbol{\lambda}) e^{-\sum_{ij}\beta\epsilon_{ij}\lambda_{ij}}, \qquad [81]$$

where the sum represents nested sums

$$\sum_{\boldsymbol{\lambda}=0}^{\mathbf{m}}[\cdot] = \sum_{\lambda_{11}=0}^{m_1}\sum_{\lambda_{12}=0}^{m_2}\cdots\sum_{\lambda_{21}=0}^{m_1}\sum_{\lambda_{22}=0}^{m_2}\ldots[\cdot] \qquad [82]$$

with the number of states Ω given by a product of multinomial distributions because for each receptor type j we need to choose how many will bind to different ligand types i. Equivalently, we need to choose among m_i ligands how many will get attached to given receptor types i, and repeat the process for each ligand type. Finally, we need to bind ligands and receptors together and there are $\prod_{ij}\lambda_{ij}$ distinct ways of connecting them. The density of states is

$$\begin{aligned}\Omega(\boldsymbol{\lambda}) &= \prod_j \left(\frac{n_j!}{\prod_i(\lambda_{ij}!)\ (n_j - \sum_i \lambda_{ij})!}\right)\prod_i\left(\frac{m_i!}{\prod_j(\lambda_{ij}!)\ (m_i - \sum_j\lambda_{ij})!}\right)\prod_{ij}\lambda_{ij}! \\ &= \prod_j \left(\frac{n_j!}{(n_j - \sum_i \lambda_{ij})!}\right)\prod_i\left(\frac{m_i!}{\prod_j(\lambda_{ij}!)\ (m_i - \sum_j\lambda_{ij})!}\right) \ . \end{aligned} \qquad [83]$$

In the second line we have cancelled out $\prod_{ij}\lambda_{ij}$ which will be convenient later. A similar form for the density of states has been derived by Angioletti-Uberti et. al. (30) when calculating interactions between DNA coated colloids.

The Poisson average is a product of Poisson averages over individual receptor types

$$Q(\tilde{\mathbf{n}}, \mathbf{m}, \boldsymbol{\epsilon}) = \langle Q(\mathbf{n}, \mathbf{m}, \boldsymbol{\epsilon})\rangle_{\mathbf{n}} = \sum_{\mathbf{n}=0}^{\infty}\left[\prod_j\left(\frac{e^{-\tilde{n}_j}\tilde{n}_j^{n_j}}{n_j!}\right)Q(\mathbf{n}, \mathbf{m}, \boldsymbol{\epsilon})\right] \qquad [84]$$

where $\sum_{\mathbf{n}=0}^{\infty}[\cdot] = \prod_j\sum_{n_j=0}^{\infty}[\cdot]$ represents a nested sum over all receptor types j and $\tilde{n}_j$ denotes the mean number of receptors. Inserting Eq. (81) and Eq. (83) into the above equation we obtain a long expression

$$Q = \sum_{\mathbf{n}=0}^{\infty}\left[\prod_j\left(\frac{e^{-\tilde{n}_j}\tilde{n}_j^{n_j}}{n_j!}\right)\sum_{\boldsymbol{\lambda}=0}^{\mathbf{m}}\left[\prod_j\left(\frac{n_j!}{(n_j - \sum_i\lambda_{ij})!}\right)\prod_i\left(\frac{m_i!}{\prod_j(\lambda_{ij}!)\ (m_i - \sum_j\lambda_{ij})!}\right)e^{-\sum_{ij}\beta\epsilon_{ij}\lambda_{ij}}\right]\right] \ , \qquad [85]$$

where we can swap the order of summation over $\mathbf{n}$ and $\boldsymbol{\lambda}$, regroup the terms and cancel out $n_j!$ in the innermost sum to find

$$Q(\tilde{\mathbf{n}},\mathbf{m},\boldsymbol{\epsilon}) = \sum_{\boldsymbol{\lambda}=0}^{\mathbf{m}} \left[e^{-\sum_{ij}\beta\epsilon_{ij}\lambda_{ij}} \prod_i \left(\frac{m_i!}{\prod_j(\lambda_{ij}!)\,(m_i-\sum_j\lambda_{ij})!} \right) \prod_j \left(\sum_{n_j=0}^{\infty} \left[\frac{e^{-\tilde{n}_j}\tilde{n}_j^{n_j}}{(n_j-\sum_i\lambda_{ij})!} \right] \right) \right] . \quad [86]$$

Now we have made progress, the innermost sum can be evaluated as the Taylor expansion for the exponential function $e^{\tilde{n}_j}$ by making use of the substitution $x_j = n_j - \sum_i \lambda_{ij}$, similarly to Eq. (76) the sum simply reduces to

$$\sum_{n_j=0}^{\infty} \left[\frac{e^{\tilde{n}_j}\tilde{n}_j^{n_j}}{(n_j-\sum_i\lambda_{ij})!} \right] = e^{-\tilde{n}_j}\tilde{n}_j^{\sum_i\lambda_{ij}} \sum_{x_j=-\sum_i\lambda_{ij}}^{\infty} \left[\frac{\tilde{n}_j^{x_j}}{x_j!} \right] = \prod_i \tilde{n}_j^{\lambda_{ij}} \quad [87]$$

as the terms in the sum with negative x_j vanish. Inserting this result into Eq. (86) and rearranging the terms we find

$$\begin{aligned} Q(\tilde{\mathbf{n}},\mathbf{m},\boldsymbol{\epsilon}) &= \sum_{\boldsymbol{\lambda}=0}^{\mathbf{m}} \left[\prod_i \left(\frac{m_i!}{\prod_j(\lambda_{ij}!)\,(m_i-\sum_j\lambda_{ij})!} \right) \prod_{i,j} \left(\left(\tilde{n}_j e^{-\beta\epsilon_{ij}} \right)^{\lambda_{ij}} \right) \right] \\ &= \prod_i \left(\sum_{\boldsymbol{\lambda}_i=0}^{m_i} \left[\frac{m_i!}{\prod_j(\lambda_{ij}!)\,(m_i-\sum_j\lambda_{ij})!} \prod_j \left(\left(\tilde{n}_j e^{-\beta\epsilon_{ij}} \right)^{\lambda_{ij}} \right) \right] \right) \\ &= \prod_i \left(1 + \sum_j n_j e^{-\beta\epsilon_{ij}} \right)^{m_i} , \end{aligned} \quad [88]$$

which upon swapping the summation and product can be recognised as the multinomial expansion. Finally, inserting our definition of ϵ Eq. (71), we obtain precisely the expression for the partition function that we have previously found Eq. (30) directly in the grand canonical ensemble for mobile receptors

$$Q(h;\mathbf{c},\mathbf{m},\tilde{\mathbf{K}}) = \prod_i \left(1 + \sum_j c_j \tilde{K}_i(h) e^{-\beta\Delta G_{ij}} \right)^{m_i} . \quad [89]$$

We must now only follow the procedure laid out after Eq. (30) to show that multicomponent binding to immobile (but Poisson distributed) receptors is governed by the same simple expression as in the case of binding to mobile receptors Eq. (39).