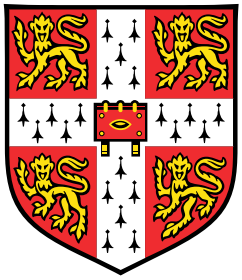
Modelling Multivalent Interactions



Tine Curk

Department of Chemistry
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Darwin College

November 2016

Declaration

I hereby declare that, except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation contains fewer than 60,000 words including appendices, bibliography, footnotes, tables and equations.

Tine Curk
November 2016

Acknowledgements

Foremost, I am very grateful to my mentor Jure Dobnikar, who brought me to the group, and my supervisor Daan Frenkel, who leads by inspiration rather than command and left me plenty of freedom to procrastinate on side projects. This work would not be possible without them.

I would like to thank all past and present members of the Frenkel group, and the wider theory sector, for providing a cosy but stimulating environment for scientific discussions. Particularly Francisco J. Martinez-Veracoechea and Bortolo M. Mognetti for enlightening debates and coaching on statistical mechanics and simulation techniques, and Anđela Šarić who recently introduced me to the all-embracing world of membranes. Furthermore, I thank my experimental collaborators Galina V. Dubacheva, Ralf P. Richter and Gerard C. L. Wong for keeping me connected to the real world. I am grateful to the Herchel Smith fund for generously supporting my PhD studies.

I appreciate my one-and-only Darwin college for providing a relaxed social environment of like-minded graduate students. Special mention goes to my first Darwin bunch: Maria, Romain and Eleni. I am thankful to Darwin sport societies for keeping me physically (and mentally) fit: badminton, squash, and of course the rowing club for also teaching me some discipline and how to work with people.

I am grateful to my family for always being there for me. Finally, I thank my J for checking on me daily and making sure I do not stray from the true path.

Abstract

A Multivalent entity, which could represent a protein, nanoparticle, polymer, virus or a lipid bilayer, has the ability to form multiple bonds to a substrate. Hence, a multivalent interaction can be strong, even if the individual bonds are weak. However, much more interestingly, multivalency enables the design of highly specific interactions using non-specific individual bonds. We attempt to rationalise multivalent effects using simple physical models complemented with numerical simulations. Based on physiochemical characteristics of multivalent binders, we aim to predict the overall strength of interaction and its sensitivity to variation in parameters.

We start with a simple model of homo-multivalency, where all bonds are equivalent. Such systems can exhibit a super-selective response, which denotes the high sensitivity of the strength of multivalent binding to the number of accessible binding sites on the target surface. We present a theoretical analysis of systems of multivalent particles and show that a certain degree of disorder is necessary for super-selective behaviour. Moreover, we formulate a set of simple design rules for multivalent interactions that yield optimal selectivity.

In the second stage, we expand the model to hetero-multivalency, accounting for multiple distinct types of binding partners. We consider targeting of cells based on a density profile of different membrane receptors types and demonstrate, that specificity towards a desired receptor density profile can be obtained. Hence, cells can be reliably targeted in the absence of specific markers. Crucially, we show that for optimal selectivity, individual bonds must be weak.

Finally, we add information about specific geometry and positions of binding sites on the multivalent entity. We focus on molecular imprinting; the process whereby a polymer matrix is cross-linked in the presence of template molecules. The cross-linking process endows the polymer matrix with a chemical 'memory', such that the target molecules can subsequently be recognised by the matrix. We show how the binding multivalency and the polymer material properties affect the efficiency and selectivity of molecular imprinting.

Contents

1

Introduction

> *Pluralitas non est ponenda sine necessitate ("plurality should not be posited without necessity"), Occam's razor*
>
> – William of Ockham

This thesis presents theoretical and computational effort in modelling multivalent interactions in supramolecular chemistry and biochemistry. We will not focus on any particular system in a great detail, instead we try to uncover general behaviour and emergent properties of multivalent interactions. We make extensive use of Occam's razor; trimming away the unnecessary and less important aspects to obtain a simple, clean explanation of the observed phenomena. We also make sure to define the boundaries and applicability of a simple model, and expand on it when deemed necessary.

A prototypical example of a valency limited interaction is a receptor-ligand pair or an antibody-antigen interaction. These interactions are not fundamental - instead they are a consequence of a multitude of weak supramolecular interactions between 2 entities, such as Van der Waals, hydrogen bonding or screened electrostatics. Due to steric repulsion only a single ligand can bind to a single receptor. The interactions have been well studied due to their importance in molecular biology [1–5]. Further examples of valency limited interactions include binding between peptides and nucleic acids [6, 7], or between nucleic acids themselves [8, 9], in which they are known as base pairing interactions.

A supramolecular (e.g. ligand-receptor) valence limited bond is our basic building block. We then explore what properties emerge when multiple such bonds act simultaneously. We assume that interaction strength of a single bond is given; either determined experimentally from binary association of moieties in solution, or calculated from detailed atomistic simulations. In our meso-scopic modelling we neglect all atomistic details

and only consider the relevant information of the ligand-receptor bond: its binding strength and the valence-limited nature of the bond. We then explore the behaviour of multivalent interactions, i.e. how does a simultaneous action of many valence-limited bonds differ from an action of an individual bond. What kind of properties or features emerge when a multitude of interactions act in parallel. Some examples of multivalent systems on a nanoscale include: multivalent nanoparticles [10, 11], polymers [12–16], nested dendrimers [17], or cells and viruses [2, 18–21].

The thesis material is largely self contained and our aim is to understand universal behaviour of multivalent systems, for example, the super-selective response to a variation of system parameters. By comparing our model predictions to experimental and simulation data we show that prototypical simple models can have a surprisingly wide range of applicability: from multivalent particles to polymers and DNA-peptide complexes.

1.1 Background: Ultra-sensitive response

Many processes in biology depend ultra sensitively on variations in one or more of the parameters that control the process. Such ultra-sensitivity manifests itself as an almost switch-like, sigmoidal change in the 'output' when the control parameter crosses a threshold value. Understanding such switch-like behaviour is obviously important to understand many regulatory processes in living systems, but such understanding will also help us design synthetic systems that combine weak supramolecular interactions with high selectivity.

The best known example of ultra sensitivity is attributed to a A. V. Hill, who, in the beginning of the twentieth century, studied the binding of oxygen to haemoglobin. He found the relation between bound oxygen and partial pressure to be sigmoidal [22]. Today this phenomena is explained in terms of allosteric cooperativity whereby the 4 binding sites on haemoglobin do not act independently but are "cooperative", i.e. binding of the first oxygen molecule increases the probability that the second oxygen molecule will bind. Hence, haemoglobin is likely to be either fully loaded with oxygen or empty, which makes haemoglobin an efficient transporter of oxygen between lungs and peripheral tissues. Other examples of ultra sensitivity include the switch-like response of bacterial motors [23], or the switch-like behaviour in gene regulation due to positive feedback loops in nucleosome modification [24]. For more information on this broad topic, the reader is referred to a review by Ferrell [25–27] and references therein.

Ultra sensitive response is usually characterised by a so-called Hill curve:

$$\text{Output} = \frac{\text{Input}^n}{K^n + \text{Input}^n} \ , \tag{1.1}$$

where the Hill coefficient n quantifies the cooperativity of the process: the higher the Hill coefficient, the more sensitive the response.

Due to cooperativity, building blocks that, individually, have limited selectivity can form units that interact selectively. For example, DNA base pairing is highly specific, even though underlying interactions (hydrogen bonding and stacking) are not. Multivalent (or polyvalent) interactions can also lead to an ultra-sensitive response, for example, aggregation of multivalent DNA coated colloids depends sensitively on temperature [11]. Moreover, ligand-receptor or antibody-antigen interactions, are very sensitive to temperature, but also to ion concentration and pH. Internal protein interactions are also multivalent, protein folding and unfolding depends critically on temperature and other external conditions. The functioning of the biochemical machinery in cells relies (mostly) on multivalent supra-molecular interactions. These interactions are very sensitive to external conditions which helps explain reason why living matter (cells, tissues) tends to be very sensitive to temperature, while 'formerly living' matter (say, a piece of wood) is not.

Imagine two multivalent entities at a fixed distance that are connected by a number of bonds (say k). The two entities can dissociate only when all k bonds are broken. We denote the probability that an individual bond is broken by p_1^{unbound} and the probability that all k bonds are broken by p_k^{unbound}. If different bonds do not influence each other, the probability of unbinding is

$$p_k^{\text{unbound}} \sim \left(p_1^{\text{unbound}}\right)^k \ . \tag{1.2}$$

Note that for large 'valencies' k, the relation between p_1^{unbound} and p_k^{unbound} is highly non-linear. In fact, the expression for the ratio between probability $p_k^{\text{unbound}}/p_k^{\text{bound}}$ has a form reminiscent of the Hill equation:

$$\frac{p_k^{\text{unbound}}}{p_k^{\text{bound}}} \sim \frac{\left(p_1^{\text{unbound}}\right)^k}{1 - \left(p_1^{\text{unbound}}\right)^k} \ , \tag{1.3}$$

where the exponent k plays a role similar to that of the Hill coefficient (1.1). The probability of a single bond spontaneously breaking p_1^{unbound} will depend not only on control parameters such as bond strength, temperature, pH of the solution etc., but also

on the number of possible bonding arrangements. Clearly, the unbinding probability (1.2) tends to be very sensitive to any parameter that influences p_1^{unbound}. This example illustrates the physical origins of ultra-sensitive response in multivalent interactions. We shall see below that competition between different bonds modifies the response but retains ultra-sensitivity.

1.2 Cell targeting

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 its receptor profile. Here, the crucial word is 'selective': we wish to target only those cells that have the correct receptor profile - binding of drug-delivery vehicles to other cells leads 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 over-expressed. In 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 [28, 29]. As we will show, the use of multivalent drugs drugs greatly increases the selectivity such that cells can be reliably targeted even when the cognate receptors are only slightly over-expressed on a targeted cell [10, 18]. Multivalent glyco-polymers have been used as selective probes for protein-carbohydrate interactions in a biochemical setting [12–14]. More recently, super-selective targeting was demonstrated in a synthetic system based on host-guest chemistry [15, 16].

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 may not be that clear cut due to variation in expression level of receptors within a population of cells. 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 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. In its simplest form (the binding of dimeric bi-specific antibodies as compared to monomeric antibodies), this problem has been studied theoretically [30] and experimentally [31]. 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.

1.3 Molecular imprinting

The term 'Molecularly Imprinted Polymers' (MIPs) is used to denote polymer matrices that have been "imprinted", *i.e.* cross-linked in the presence of a template molecule, thereby acquiring selective affinity towards its template. MIPs are usually made by free-radical co-polymerisation of ligands and cross-linkers in the presence of template molecules. The molecule-matrix interaction may exploit covalent binding, ionic interactions [32], hydrogen bonding [33], $\pi - \pi$ stacking interactions [34], hydrophobic interactions [35], and metal-ion chelation [36]. In 1930 Polyakov introduced this technique [37] to imprint silica matrices with benzene. However, the technique has only become widely used in recent decades [38–41]. The use of MIPs is related to the fact that they can be designed for highly selective recognition. Moreover, they combine thermal and chemical stability with ease of preparation, and hence low production costs.

MIPs have been used in applications such as solid-phase extraction [42], chiral separation [43], and catalysis [44]. They can act as molecular sensors [39, 41, 45], and mimic antibodies or enzymes [39]. They can selectively bind drugs [46–48], proteins [49], or even whole bacteria [50, 51]. Figure 4.1 shows a schematic representation of the imprinting and subsequent recognition process. The efficiency of the molecular recognition process depends on a number of parameters: 1) the initial ligand concentration c, 2) the template-ligand binding affinity K_D, and 3) the stiffness of the polymer matrix k_h.

Clearly, it is important to maximize the selectivity of MIPs, but in experiments MIPs are often optimized by trial and error. In fact, the theoretical picture is rather fragmented as existing theoretical models for MIPs do not consider the imprinting process as a whole, but rather tend to focus on individual steps in their mode of action [52–58]. Moreover, the atomistic and coarse-grained simulations of molecular imprinting that have been

reported [59–64] focused mostly on specific MIPs and did not explore generic trends that would allow us to arrive at general design principles.

We will apply our multivalent theoretical framework to study molecular imprinting and obtain general design rules for optimal imprinting protocols.

1.4 Note on nomenclature and terminology

We make liberal use of the terms "ligand" and "receptor" with which we shall denote individual binding partners. In our language "receptors" will be found on the substrate surface whilst individual "ligands" are attached to the multivalent entity (say, a nanoparticle or a polymer) that binds to the substrate, shown on Figure 1.1. We use the term "multivalent entity" to denote any moiety that is able to form multiple bonds. The term 'binding site' always denotes an individual monovalent interaction site, equivalent to a single "ligand" or "receptor".

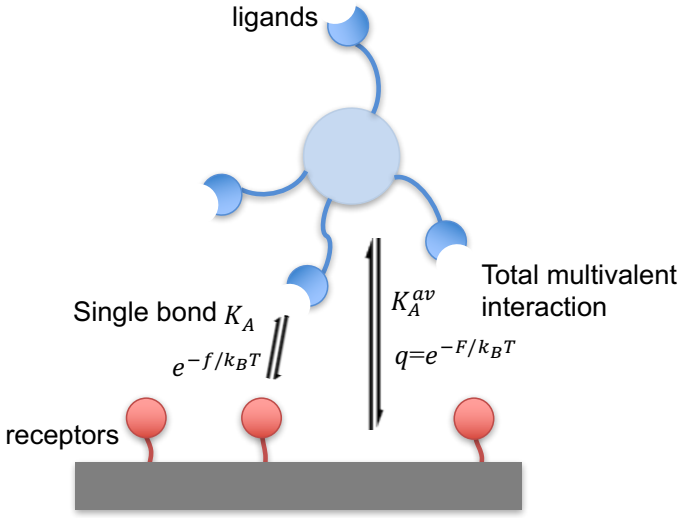


Fig. 1.1 Cartoon representing a multivalent interaction between a nanoparticle with grafted polymeric arms and surface attached receptors. Individual ligand-receptor bond strength is denoted by the free energy f, or equivalently the affinity constant K_A. Most of the work in this thesis focuses on calculating the total multivalent interaction (the binding free energy F, or equivalently the avidity association constant K_A^{av}) via evaluation of the bound partition function q_b.

The binding strength of individual ligand-receptor bonds is characterised by the equilibrium affinity constant K_A or the dissociation constant $K_D = 1/K_A$. The equilibrium constant can be equivalently represented by the dimerisation free energy $\Delta G = k_B T \log(K_D/\rho_0)$ with k_B the Boltzmann constant, T the absolute temperature and

$\rho_0 = 1M$ is the standard concentration. We assume these are known, or can be in principle calculated experimentally, and serve as an input to our modelling. We shall also use the internal equilibrium constant K_{intra} to characterise the strength (or the probability) of forming a bond within a multivalent complex. $K_{intra} = K_A EM$ includes the monomeric affinity K_A but also the so called "effective molarity" EM which was defined as an empirical quantity characterising the efficacy of multivalent interactions. EM can be calculated theoretically as the ratio of the configuration partition functions of the bound/unbound states. Equivalently, we can use the binding free energy $f = -k_B T \log(K_{intra})$, where f includes ΔG but also configurational terms related to a bond formation within a multivalent complex. All of these quantities can (and will) be used to characterise individual bonds. Additionally, we shall use ϵ to denote the binding strength when using coarse-grained models with Monte Carlo simulations.

On the other hand, the avidity equilibrium constant K_A^{av}, or the overall binding free energy $F = -k_B T \log(K_A^{av} \rho_0)$ measures the strength of the overall multivalent interaction, say the interaction between a nanoparticle and a substrate shown in Figure 1.1. Avidity can be seen as the accumulated strength of individual affinities. We shall use n_R and k to denote the number of receptors and ligands, respectively, that can simultaneously participate in the multivalent interaction. The work in this thesis primarily attempts to calculate the avidity K_A^{av} of a multivalent interaction using the knowledge of affinities K_A and the physical properties of the multivalent binders.

We use chemical equilibrium units and statistical mechanics units interchangeably. The modelling and our results can be presented with either notation. However, in the later chapters we shall primarily use statistical mechanics notation as we find its use clearer and resulting expressions more elegant.

The main part of this work is divided into three chapters. On every stage we formulate simple design rules of how to optimise selectivity and specificity of multivalent interactions. In what follows, we shall first focus on the ultra sensitivity of multivalent interaction to the density of "receptors" on a substrate surface. We show how the description of simple chemical equilibria and Langmuir adsorption can be extended to multivalent interactions. In particular, we will derive expressions that show how the binding strength of a multivalent entity (say a ligand-decorated nanoparticle or a multivalent polymer) changes sharply with the concentration of receptors on the substrate surface. We will also present numerical simulations that validate the simple analytical expressions.

Further on we expand our model to multiple ligand/receptor types and consider targeting of cells based on a density profile of different receptors types. We demonstrate that using weak multivalent interactions, the specificity towards a desired receptor density

profile can be obtained. Hence, we show how cells can be targeted in the absence of specific biological markers.

Finally, we focus on rigid multivalency in the context of molecular imprinting, where not only different types of binders, but also their specific location on a multivalent entity, play a role. We show how the binding multivalency and the polymer material properties affect the efficiency and selectivity of molecular imprinting.

Tedious mathematical derivations and evaluation of the integrals has been moved to the appendix to make the main part of the thesis more balanced and readable.

2

Super selectivity

One of the principle objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity.

– J.W. Gibbs

2.1 An emergent property of multivalency

We start our focus on super-selectivity, which denotes the high sensitivity of the strength of multivalent binding to the number of accessible binding sites on the target surface, shown schematically on Figure 2.1. For example, the docking of a multivalent particle to a cell-surface can become very sensitive (super-selective) to the concentration of the receptors to which the multiple ligands can bind.

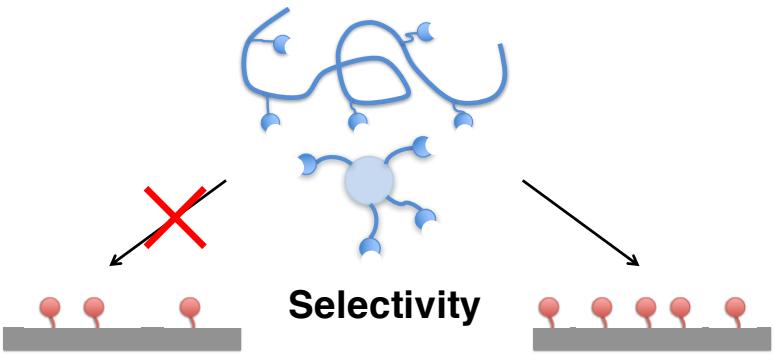


Fig. 2.1 Selectivity denotes the ability of multivalent entities to distinguish between substrates depending on the surface density of binding sites.

We consider a prototypical system of multivalent particles in solution that can adsorb to a receptor-decorated surface. For simplicity, we assume that the surface is flat and much larger than the multivalent particles. Furthermore we assume that these particles are larger than the surface receptors such that each particle can be in range of many receptors sites simultaneously. Adsorption of particles is governed by the well-known Langmuir isotherm which states that the fraction of the surface occupied by particles is

$$\theta = \frac{\rho K_A^{av}}{1 + \rho K_A^{av}} , \tag{2.1}$$

with ρ the molar concentration of particles in solution [1], K_A^{av} is the equilibrium avidity association constant of particles adsorbing to a surface. Note that K_A^{av} is different from the affinity equilibrium constant K_A which specifies chemical equilibria of individual ligand-receptor binding. Avidity (functional affinity) is the accumulated strength of multiple affinities [65].

We aim to understand how the overall avidity constant K_A^{av} depends on the properties of the system, i.e. individual bond affinities K_A, the ligand valency k and number of receptors n_R. The avidity constant includes all possible bound states, and is written as a sum over all possible bonds

$$K_A^{av} = \Omega_1 K_A + \Omega_2 K_A K_{intra} + \Omega_3 K_A {K_{intra}}^2 + ... \tag{2.2}$$

assuming all bonds are equivalent and independent. Later (Chapter 3) we will generalise this expression to different bond types. The first term on the right hand side takes into account all states with a single formed bond with K_A the standard monomeric affinity constant [2]. The second term on the right represents all doubly bound states where K_{intra} is a constant specifying the internal equilibrium between singly and doubly bonded states, the same reasoning applies to triply and other multivalent bonded states. The nature of K_{intra} is further elucidated in Section 2.7 on effective molarity. To obtain the above expression we have assumed that individual bonds are equivalent and form independently, therefore, K_{intra} is a constant, i.e. we ignore (allosteric) cooperative effects where K_{intra} depends on the number of formed bonds. We do this to clearly distinguish multivalent effects (the subject of this chapter) from cooperative effects [66], we note that some authors [67] use the term "chelate cooperativity" to denote multivalent effects.

[1] For non-ideal solutions the density ρ in the Langmuir isotherm (2.1) should be replaced by the fugacity.

[2] K_A is the association equilibrium constant between a monovalent particle (a single ligand attached to a particle) and a single receptor, we assume it can be determined experimentally

Ω_λ is the degeneracy pre-factor, it measures the number of ways in which λ bonds can be formed between two multivalent entities, see Figure 2.2 for representative cartoons. Degeneracy Ω is often labelled as a "statistical pre-factor" which denotes something that should be included for rigour but is otherwise not essential. However, as we will show, it is precisely this degeneracy that gives rise to super-selectivity. The focus of the majority of theoretical papers is on the calculation of the internal equilibrium constant K_{intra} [65, 67–70]. Here, instead, we focus on the degeneracy Ω. We will simply assume that K_{intra} is (or can be) known.

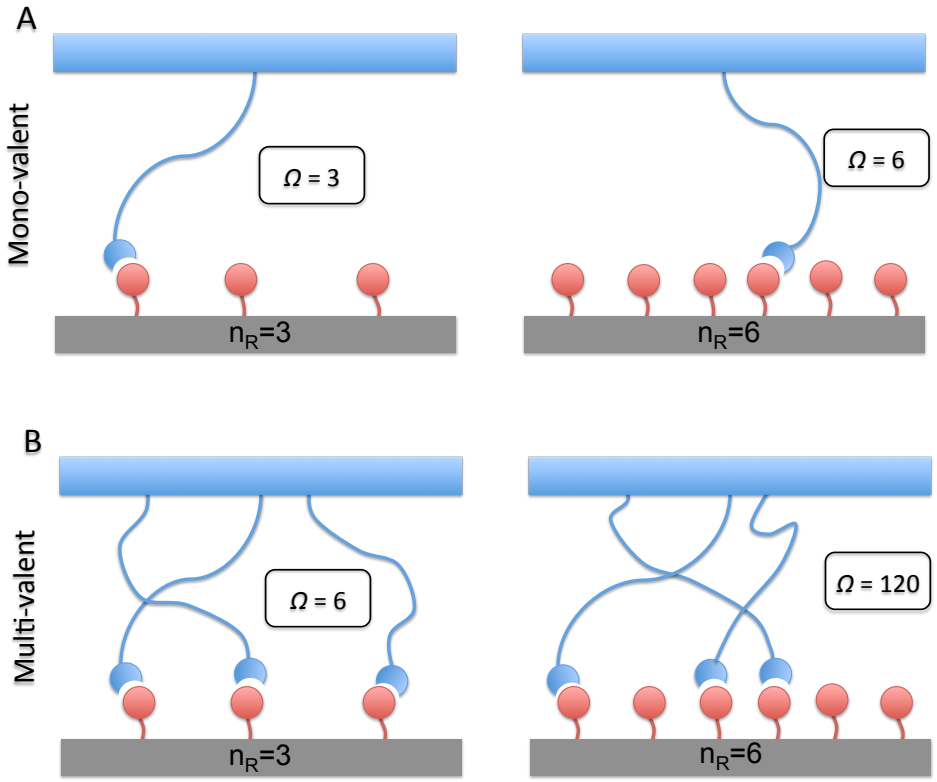


Fig. 2.2 Entropic origin of super selectivity. The pictures show the binding of mono-valent (A) and multi-valent (B) ligands (represented as a blue bar with attached flexible ligands). Receptors are shown as red spheres tethered to the bottom surface. The left panels show a low receptor density ($n_R = 3$) and the panels on the right show a receptor density that is twice as large ($n_R = 6$). In the mono-valent case the number of distinct ways (Ω) to link ligands and receptors grows linearly with the number of receptors n_R, while multivalent ligands show a highly non-linear response: changing n_R from 3 to 6 increases Ω by a factor of 20. In general, the number of binding combinations (degeneracy) Ω is calculated using (2.3).

The degeneracy Ω depends on the spatial arrangement of both ligands and receptors. However, it is instructive to consider first the binding of flexible ligands, where all

k ligands on a particle can bind to n_R receptors (Figure 2.2B). The number of ways (degeneracy) to form λ bonds is

$$\Omega_\lambda = \binom{n_R}{\lambda}\binom{k}{\lambda}\lambda! = \frac{n_R!\ k!}{(n_R-\lambda)!\ (k-\lambda)!\ \lambda!} \tag{2.3}$$

because we need to choose λ ligands out of k and choose λ receptors out of n_R, then there are $\lambda!$ ways of binding the chosen ligands/receptors together. The degeneracy (2.3) becomes a very steep and non-linear function of k and n_R. This form was first considered by Kitov and Bundle [71] and has been applied, among others, to nanoparticle targeting of cells [10] and modelling the adhesion of influenza virus [18].

The avidity constant (2.2), using degeneracy (2.3), can be well approximated by a binomial expansion series [10] when we can approximate $\frac{n_R!}{n_R-\lambda} \approx n_R^\lambda$ for all significant terms in (2.2). This condition is satisfied when the fraction of bound receptors is low $\frac{\lambda}{n_R} \ll 1$, which arrises if the number of receptors is greater than the number of available ligands $n_R \gg k$ or individual bonds are sufficiently weak $K_{intra} \ll 1/k$ [3]. The binomial expansion is summed to yield a simple form:

$$K_A^{av} \approx \frac{K_A}{K_{intra}} \sum_{\lambda=1}^{k} \binom{k}{\lambda} (n_R K_{intra})^\lambda = \frac{K_A}{K_{intra}} \left[(1+n_R K_{intra})^k - 1\right] , \tag{2.4}$$

where, as before, K_A is the monomeric single-bond affinity constant, K_{intra} the internal association constant, and n_R and k are the number of receptors and ligands, respectively. For our purpose it is important to note that for multivalent binding ($k > 1$), K_A^{av} is a steep, non-linear function of n_R (see Figure 2.3).

In a practical system the bond strength K_{intra} will not be a constant because the polymeric ligand arm stretching penalty depends on the exact position of the receptor. Bonds are harder (impossible) to form for receptors far away from the particle. However, to make the problem analytically tractable we will, at this stage, approximate the distance dependant stretching with a step function: within a lattice site ligands can bind the any receptor n_R with equilibrium constant K_{intra}, but cannot bind to receptors outside of the site, see Figure 2.4.

Eq. (2.4) could have also been obtained directly by reasoning that for non-saturated receptors (fraction of bound receptors is low), competition for the same receptor can be ignored. Each ligand can independently bind to any of the n_R receptors (weight

[3] The largest term in (2.2) is obtained by $\Omega(\lambda){K_{intra}}^\lambda \approx \Omega(\lambda+1){K_{intra}}^{\lambda+1}$, which results in $K_{intra} \approx \frac{\lambda}{(k-\lambda)(n_R-\lambda)}$. If the bonds are sufficiently weak: $K_{intra} < 1/k$, the largest term will always arise when the fraction of occupied receptors is low $\frac{\lambda}{n_R} < 0.5$.

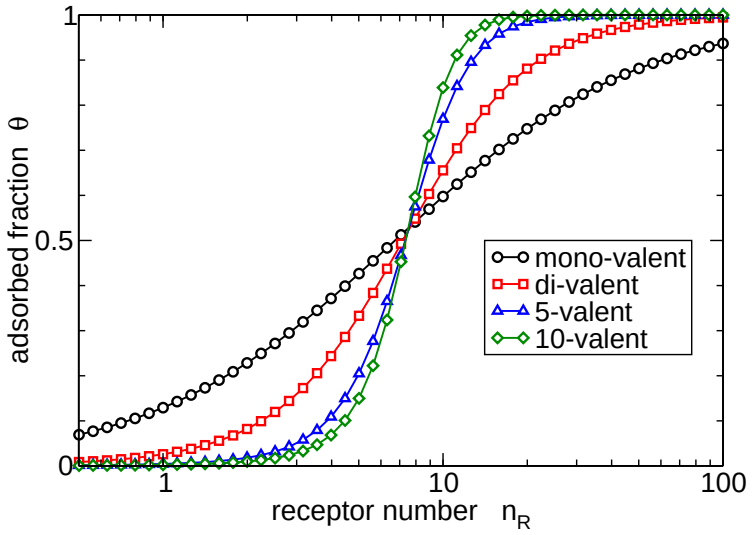


Fig. 2.3 Adsorption profile of multi-valent particles computed using Eqs. (2.1, 2.4). Monovalent adsorption (black circles) $k = 1$ yields the familiar Langmuir isotherm. In contrast, multivalent particles display a steep, sigmoidal response. In the case shown we have chosen the dimensionless activity in solution to be $z \equiv \rho \frac{K_A}{K_{intra}} = 0.001$, the binding affinity of individual bonds decrease as the valency increases from mono-valent to 10-valent: $\log(K_{intra}) = -\beta f = 5$, 1.5, -1, -2, such that the overall avidity K_A^{av} at 50% bound fraction ($\theta = 0.5$) is kept constant for all valencies.

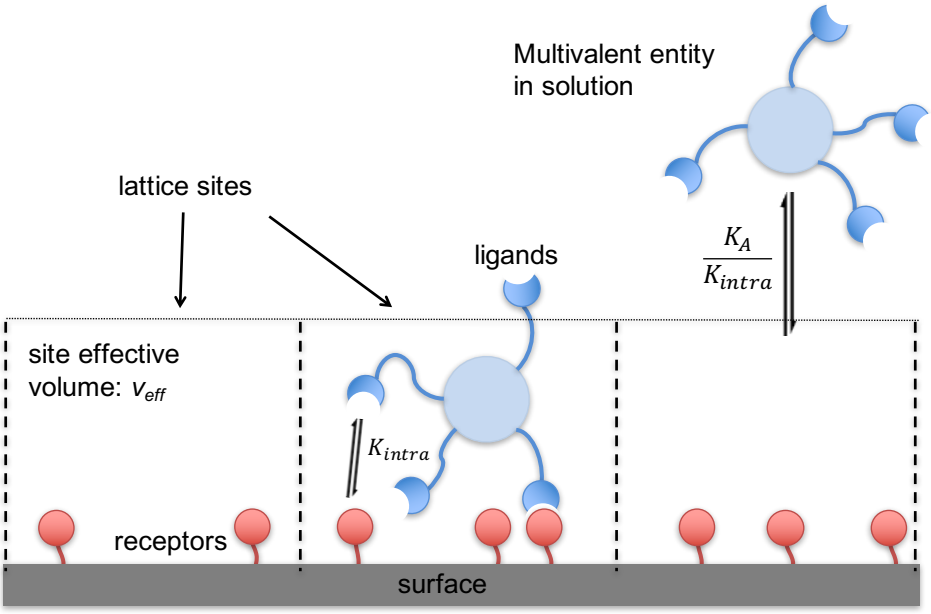


Fig. 2.4 Cartoon showing the multivalent adsorption model. Multivalent entities (particles) adsorb to lattice sites, we assume that within each site the ligands can independently bind to surface attached receptors, but cannot reach any receptor from neighbouring sites. The equilibrium constant for an unbound particle to adsorb from the solution to one of the sites is $\frac{K_A}{K_{intra}} = v_{eff}$, which is related to the effective volume of the site v_{eff}. Once the particle is adsorbed within the site, the ligands can independently form bonds with the surface attached receptors with equilibrium constant K_{intra}.

$n_R K_{intra}$) within the site, alternatively the ligand can be unbound (weight 1). Hence, for systems with a low fraction of bound receptors, the factor $(1 + n_R K_{intra})^k$ accounts (approximately) for all possible states. Furthermore, we subtract 1 because we use the convention that at least a single bond needs to be formed for the multivalent particle to be considered bound. Finally, to obtain the avidity constant K_A^{av}, we multiply our expression the equilibrium constant of adsorbing a non-bound particle to the lattice site: the ratio $\frac{K_A}{K_{intra}}$, which also returns the correct monomeric behaviour (for monomeric binding ($k = 1$) we must have $K_A^{av} = n_R K_A$).

We note that in an idealised system, the ratio $\frac{K_A}{K_{intra}} = v_{eff}$ is related to the effective volume v_{eff} that an unbound particle must enter in order to be able to start forming bonds, see Figure 2.4. The form of equation (2.4) suggest that we can view the multivalent particle adsorption as a 2 step process. First, the particle adsorbs from the solution to the surface and comes into a position to start forming bonds, the equilibrium constant of this process is given by the ratio $\frac{K_A}{K_{intra}}$. Once the particle is in this position, all of the k ligands can independently form bonds with surface receptors.

In the monovalent case ($k = 1$) the avidity constant (2.4) reduces to $K_A^{av} = n_R K_A$ and the standard Langmuir isotherm is obtained. Furthermore, expanding (2.4) in a binomial series and using a maximum term approximation we can insert the maximum term in (2.1) and obtain the phenomenological Hill equation (1.1). In the case of very strong individual bonds ($n_R K_{intra} \gg 1$) virtually all k bonds are formed and the avidity becomes $K_A^{av} \approx n_R^k K_A {K_{intra}}^{k-1}$ [4].

We started by using equilibrium constants as our quantities of choice. However, later we shall switch to statistical mechanics notation using partition function and free energies. We stress that so far both languages are equivalent and interchangeable. Here we provide a translation cheat-sheet between the chemical and statistical mechanical language;

- Monomeric Gibbs hybridisation free energy: $e^{-\beta\Delta G} = K_A \rho_0$
- Binding free energy per single bond once the first bond has been formed: $e^{-\beta f} = K_{intra}$
- Bound state partition function: $q_b = K_A^{av} \frac{K_{intra}}{K_A}$, we note that $\frac{K_{intra}}{K_A} = v_{eff}$ can be represented with a configurational "volume" that a multivalent entity must enter from a solution in order to start forming bonds, see Figure 2.4
- Dimensionless activity of multivalent ligands in solution: $z = \rho v_{eff} = \rho \frac{K_A}{K_{intra}}$

with $\beta \equiv 1/k_B T$ the inverse temperature. Using these identifications, we can rewrite the Langmuir isotherm (2.1)

$$\theta = \frac{z q_b}{1 + z q_b} \,, \tag{2.5}$$

where the bound partition function is given by

$$q_b = \left(1 + n_R e^{-\beta f}\right)^k - 1 \quad . \tag{2.6}$$

This dimensionless notation was used in Refs. [10, 15, 16, 72] and is a more convenient language for theoretical modelling. From now on we shall primarily use this notation.

We have shown how combinatorial entropy (2.3) (also called "avidity entropy" [65]) gives rise to sharp switching behaviour upon a change in receptor concentration n_R (Figure 2.3). Next, we introduce a measure of the sensitivity of the binding of multivalent particles to the surface concentration of receptors:

$$\alpha = \frac{d \log \theta}{d \log n_R} \,. \tag{2.7}$$

[4] this holds for $n_R \gg k$ when (2.4) is applicable even for strong bonds, in general (using (2.2)) the expression would be $K_A^{av} = \frac{n_R!}{(n_R - k)!} K_A {K_{intra}}^{k-1}$

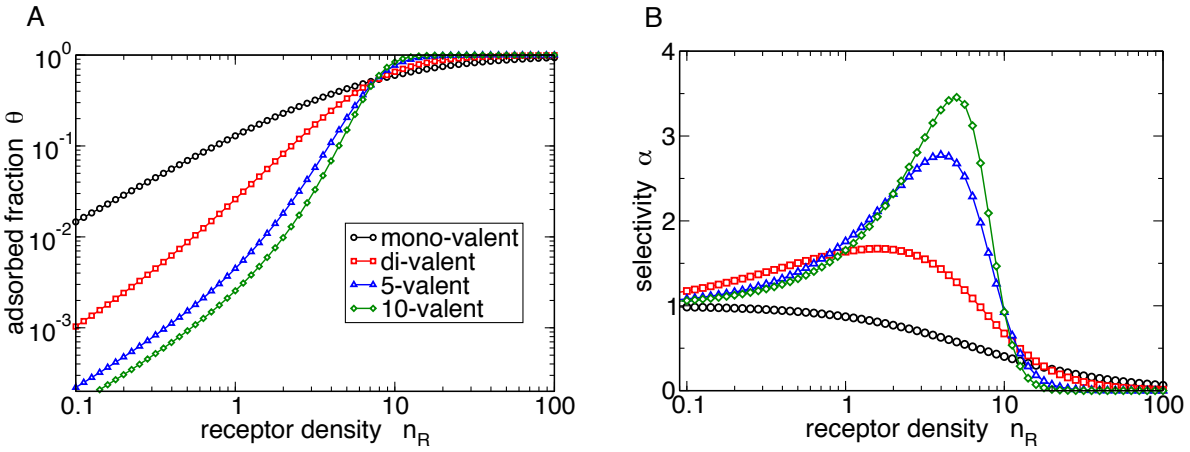


Fig. 2.5 Selectivity. **A**) shows the log-log plot of Figure 2.3, and **B)** shows its slope, i.e. the selectivity $\alpha = \frac{d\log\theta}{d\log n_R}$. We observe that selectivity is typically less than one for mono-valent ligands indicating at most linear response. Multi-valent ligands, on the other hand, exhibit a region with values of α significantly greater than one, thus demonstrating that the number of adsorbed ligands increases faster than linearly with the receptor concentration: in this regime, the system is super-selective.

α is the slope of the adsorption profile in a log-log plot (see Figure 2.5). For mono-valent binding the selectivity α is never larger than one, while in the multi-valent case the selectivity can reach values greater than one, indicating a supra-linear response. Note that for low surface coverage ($\rho K_A^{av} \ll 1$) the selectivity α is equivalent to the effective Hill-coefficient (1.1). However, because we consider all terms (all possible number of bonds) in calculating avidity (2.2), α is not a constant. At very low receptor concentrations the avidity shows a linear dependence on n_R, and $\alpha \approx 1$ [5]. Selectivity then grows with increasing receptor concentration n_R until reaching a peak just before the saturation of the surface ($\rho K_A^{av} \approx 1$). We refer to the region with $\alpha > 1$ as the 'super-selective' region. In this region, a small change in the receptor density n_R causes a faster-than-linear change in adsorption θ.

2.2 Multivalent polymer adsorption

To validate the model for super-elective adsorption described above, we now compare its predictions with experimental data on polymer adsorption. Multivalent glyco-polymers have been used as selective probes for protein-carbohydrate interactions in a biochemical

[5] At sufficiently low receptor concentration, when $n_R k K_{intra} \ll 1$ holds, we expand (2.4) to first order and obtain $K_A^{av} \approx n_R k K_A$.

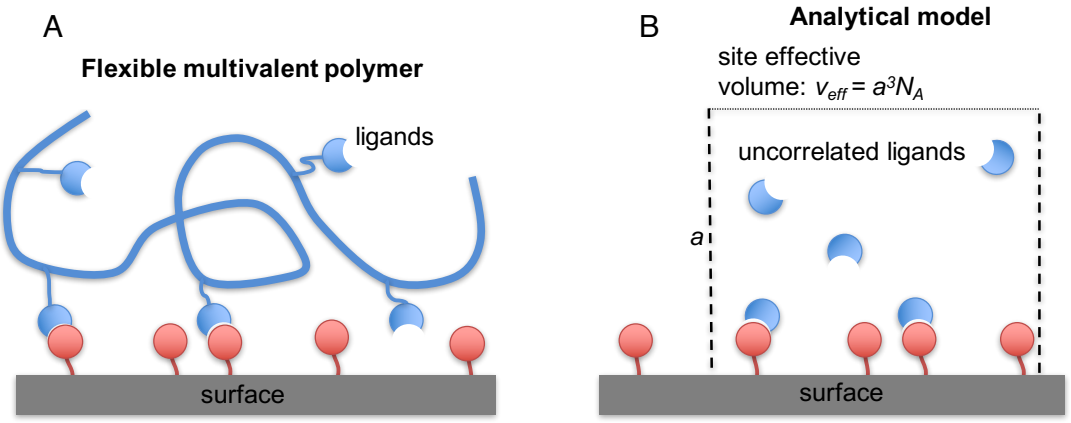


Fig. 2.6 Cartoon showing the multivalent polymer model. **A)** Flexible multivalent polymer close to the receptor decorated surface is modelled as **B)** uncorrelated ligands within a lattice site with volume $v_{eff} = a^3 N_A$ and a the linear lattice size. The ligands can move and bind to receptors independently within the lattice site, but cannot escape the site individually. The model is equivalent to the general multivalent adsorption model (Figure 2.4). However, in the case of flexible polymers we can also analytically estimate the effective volume $v_{eff} = a^3 N_A = \frac{4\pi}{3N_A} R_g^3$, with R_g the polymer radius of gyration in solution.

setting [12–14]. More recently, super-selective targeting was demonstrated in a synthetic system based on host-guest chemistry [15, 16]. We describe multi-valency effects in the case of polymers functionalised with many ligands.

We consider a flexible polymer with a contour length much larger than the persistence length. Ligands are randomly attached along the polymer chain (see Figure 2.6. Similar to the nano-particles case above, a reasonable first assumption is that, due to polymer-chain flexibility, all k ligands on a polymer can bind to any of the n_R receptors within a domain on the surface with lateral dimensions comparable to those of the polymer. For simplicity, we describe the surface as a square lattice. The cells of the lattice have linear dimensions similar to the radius of gyration R_g of the polymer. As in the case of soft multivalent particles, any ligand on the polymer can bind to any receptor in one (and only one) lattice cell. The number of receptors that a polymer can see is then $n_R = \Gamma N_A a^2$, where Γ denotes the molar surface density of receptors.

The calculation of the bound partition function (or the avidity constant) is the same for multivalent polymers or particles (2.4). In the case of flexible polymers we can also estimate the intra association constant as $K_{intra} = K_A / v_{eff}$, with the effective volume $v_{eff} \approx N_A a^3$, the lattice size $a = R_g (4\pi/3)^{1/3}$ and R_g the polymer radius of gyration in solution. This model (and the choice of effective concentration) effectively describes a

multivalent polymer as a "cloud" of ideal gas ligands, i.e. ligands are uncorrelated (can bind independently) but unbound ligands are confined to the lattice site with volume v_{eff}, see Figure 2.6. The model is expected to offer a faithful description of the real system if the mean distance between ligands is larger than the Kuhn segment length such that even consecutive ligands along the polymer chain can be treated as uncorrelated.

Using the above definitions, we find the following expression for the avidity constant of a multivalent polymer:

$$K_A^{av} = a^3 N_A \left[\left(1 + \frac{\Gamma K_A}{a} e^{-\beta U_{poly}} \right)^k - 1 \right] , \tag{2.8}$$

where we have added a correction term U_{poly} which takes into account the deviation of the real system to our "cloud of ideal ligands" approximation, Figure 2.6. This approximation neglects the polymeric degrees of freedom and, consequently, any spatial correlations between ligands. Moreover, we ignore the fact that the binding free energy of ligands to receptors is changed by the coupling of the ligands to the polymer backbone. These approximations will results in an error of order $k_B T$ and we expect U_{poly} to be $\mathcal{O}(k_B T)$.

The analytical model (2.8) captures the scaling relations and design guidelines of multivalent polymer adsorption. However, to obtain the result on Figure 2.7 interpenetration of polymer chains must also considered. To capture the effect of polymer interpenetration the Langmuir expression needs to be extended to include the possible many polymer chains per lattice site. We will switch to the statistical mechanics notation now. The effective bond strength can be obtained using

$$f = -k_B T \ln(K_{intra}) = k_B T \ln \left(\frac{a^3 N_A}{K_A} \right) + U_{poly} . \tag{2.9}$$

We focus on the polymer adsorption dependence on the surface receptor density, therefore we write a grand-canonical partition function $\Xi(n_R)$ for a single surface lattice site as

$$\Xi(n_R) = 1 + \sum_{i=1}^{\infty} z^i q_i \tag{2.10}$$

where i is the number of polymers that are adsorbed at the site, z is the activity of the polymers in solution and q_i is the single site bound state partition function with i polymers occupying the site.

The bound state partition function q_i counts all possible combinations of ligand-receptor bonds. For a single bound polymer the partition function reduces to the familiar

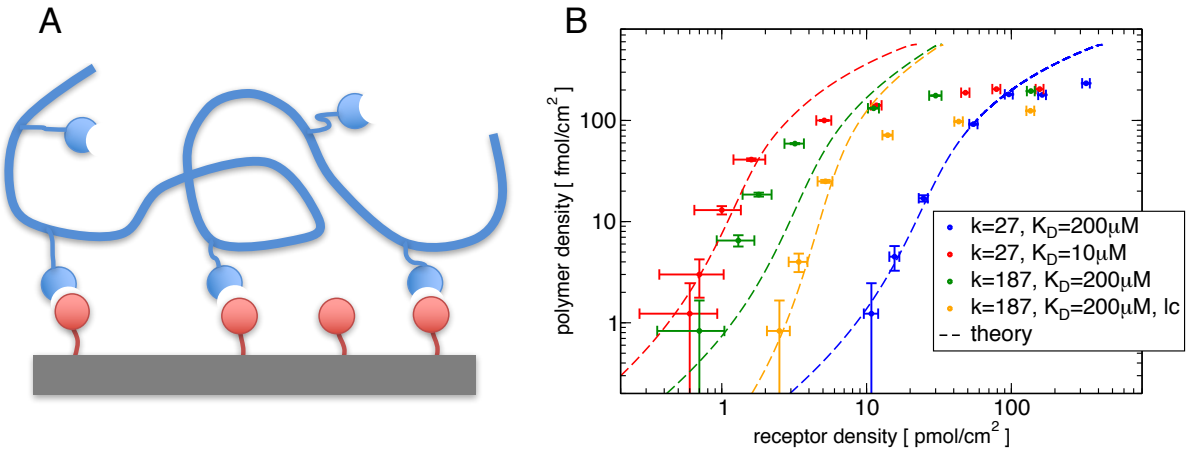


Fig. 2.7 Multivalent polymer adsorption. **A)** Schematic representation of a flexible polymer with attached ligands (blue) that can bind to surface receptors (red blobs). **B)** Experimental adsorption profiles (points) for hyaluronic acid polymers functionalised with β-cyclodextrin hosts (HA-β-CD) binding to surface attached adamantane (affinity $K_D = 1/K_A = 10\mu M$) or ferrocene ($K_D = 200\mu M$) guests, data obtained from Dubacheva et. al. [15, 16]. As can be seen, the theoretical adsorption profiles (dashed lines) match the experimental data well for all valencies (k), affinities (K_D) and polymer concentration studies. In the Figure, "lc" next to orange data points denotes lower concentration of polymers in solution. The value $U_{poly} = 4.6k_BT$ was fitted globally. The theoretical curves overshoot at high surface polymer density because the model of polymer-polymer overlapping repulsion, Eq. (2.12), fails at high polymer densities.

expression (2.3)

$$q_1 = e^{-\beta U_1} \sum_{\lambda=1}^{\min(k,n_R)} e^{-\lambda\beta f} \binom{n_R}{\lambda} \binom{k}{\lambda} \lambda! \tag{2.11}$$

where λ is the number of formed bonds, βf is the free energy gain by forming a single bond and U_1 is the free energy of polymer confinement to the lattice site explained below. The combinatorial factor counts all possible distinct ways of binding together k ligands with n_R receptors using λ bonds. The maximum number of bonds $\lambda_{max} = \min(k, n_R)$ is limited by either the number of ligands or receptors, whichever is lower.

We have also added a free energy cost of polymer interpenetration and steric repulsion from the surface, for i polymers in the lattice site the cost is

$$U_i = A_{dG} i^{9/4} + U_{lat} i. \tag{2.12}$$

where the first term is the des-Cloiseaux law [73] approximation for the free energy of overlapping self-avoiding walk polymers in the semi-dilute regime, A_{dG} is a fitting constant of order unity, on Figure 2.7 the constant was fitted to $A_{dG} = 0.35$. This law fails for large number of polymers, $i \gg 1$, which is the reason why theoretical curves on Figure 2.7 overshoot in the saturated regime at high polymer density. This high polymer density regime is not super-selective, i.e. the surface is saturated, therefore, we did not focus on obtaining a correct theoretical description for the saturated regime.

The second term in (2.12), U_{lat}, captures the the mean field polymer-surface repulsion. The potential of mean force between a self-avoiding-walk polymer and an impenetrable wall is very well approximated up to $10 k_B T$ by an exponential function [74, 75]

$$V_{ps}(r) = A e^{-B(r/R_g - C)}, \tag{2.13}$$

with the constants $A = 3.2 k_B T$, $B = 4.17$ and $C = 0.5$, r is the distance between the surface and polymer centre-of-mass and R_g the polymer radius of gyration. In our theory we divide the surface into cubic sites with size $a = R_g(4\pi/3)^{1/3}$. We now calculate the average potential of mean force acting on a polymer when the polymer is within the lattice site

$$e^{-\beta U_{lat}} = \langle e^{-\beta V_{ps}(r)} \rangle = \frac{1}{a} \int_0^a \exp(-\beta V_{ps}(r)) dr = 0.436 \tag{2.14}$$

and therefore $U_{lat} = -k_B T \log(0.436) = 0.83 k_B T$. In other words, U_{lat} represents a mean-field entropic penalty for the polymer to be located in a lattice site at the surface.

When 2 polymers occupy the same lattice site the first can bind to any of the n_R receptors but the second polymer has less receptors available to it. We write the 2

polymer bound-state partition function as

$$q_2 = \frac{e^{-\beta U_2}}{2!} \sum_{\lambda_1=1}^{\lambda_{max}} e^{-\lambda_1 \beta f} \frac{k! n_R!}{(k-\lambda_1)! \lambda_1! (n_R-\lambda_1)!} \sum_{\lambda_2=1}^{\lambda'_{max}} e^{-\lambda_2 \beta f} \frac{k!(n_R-\lambda_1)!}{(k-\lambda_2)!(\lambda_2)!(n_R-\lambda_1-\lambda_2)!}, \tag{2.15}$$

where λ_1 and λ_2 count the number of bonds formed with the first and second polymer respectively and $\lambda'_{max} = \min(n_R - \lambda_1, k)$ is the maximum number of bonds available to the second polymer. The 2 polymer overlap term βU_2 takes into account the free energy penalty due to polymer excluded volume and the permutation factor 2! accounts for the indistinguishability of the 2 polymer chains. The permutation factor must be included (even if the polymers are not indistinguishable *per se*) because we assume a grand canonical description for the unbound polymers in solution, i.e. a chemical potential, which implicitly includes the permutation factor in the unbound partition function.

We notice that with 2 polymers we have effectively $2k$ ligands in a lattice site so the partition function can be rewritten as

$$q_2 = \frac{e^{-\beta U_2}}{2!} \left[\sum_{\lambda=1}^{\min(2k, n_R)} e^{-\lambda \beta f} \frac{(2k)! n_R!}{(2k-\lambda)! \lambda! (n_R-\lambda)!} - 2 \sum_{\lambda=1}^{\min(k, n_R)} e^{-\lambda \beta f} \frac{k! n_R!}{(k-\lambda)! \lambda! (n_R-\lambda)!} \right], \tag{2.16}$$

where the right hand sum has to be subtracted because each of the polymers needs at least a single bond present to be considered bound. The equality of Eqs. (2.15, 2.16) can be proven by applying the Chu-Vandermonte identity [76]. Inserting (2.11) into (2.16) we find

$$q_2 = \frac{e^{-\beta U_2}}{2!} \left[\sum_{\lambda=1}^{\min(2k, n_R)} e^{-\lambda \beta f} \frac{(2k)! n_R!}{(2k-\lambda)! \lambda! (n_R-\lambda)!} - 2q_1 \right]. \tag{2.17}$$

For the range of parameters studied the number of surface receptors per site n_R will always be much larger then the number of available ligands ik. In limit $n_R \gg ik$ Eq. (2.11) can be simplified (similarly to (2.4)) to

$$q_1 \approx e^{-\beta U_1} \sum_{\lambda=1}^{k} \left(n_R e^{-\beta f} \right)^{\lambda} \frac{(k)!}{(ik-\lambda)! \lambda!} = e^{-\beta U_1} \left[\left(1 + n_R e^{-\beta f} \right)^k - 1 \right]. \tag{2.18}$$

Following the same procedure also (2.17) can be simplified to

$$q_2 \approx \frac{e^{-\beta U_2}}{2!} \left[\sum_{\lambda=1}^{2k} \left(n_R e^{-\beta f} \right)^{\lambda} \frac{(2k)!}{(2k-\lambda)! \lambda!} - 2q_1 \right] = \frac{e^{-\beta U_2}}{2!} \left[\left(1 + n_R e^{-\beta f} \right)^k - 1 \right]^2. \tag{2.19}$$

Extending to any number of polymer we find a simple analytical expression for the bound partition function of i polymers

$$q_i \approx \frac{e^{-\beta U_i}}{i!}\left[\left(1+n_R e^{-\beta f}\right)^k - 1\right]^i \tag{2.20}$$

providing a practical route for the calculation of multivalent polymer interaction. The expression is valid as long as the fraction of bound receptors remains low, i.e. analogous to Eq. (2.4) when $ik \ll n_R$ or $K_{intra} \ll 1/ik$. This condition was satisfied for all results on Figure 2.7. From experimental data, the surface saturated when the number of polymers per site reached about $i^{max} \approx 4$.

We now have the necessary tools to calculate the grand partition function for a single lattice site (2.10) from which the average number of bound polymers $\theta(n_R)$ can be easily calculated as

$$\theta(n_R) = \frac{\partial \ln \Xi(n_R)}{\partial(\beta\mu)} = \frac{\sum_{i=1}^{\infty} i z^i q_i}{1+\sum_{i=1}^{\infty} z^i q_i}\ , \tag{2.21}$$

with μ the chemical potential of polymers in solution. However, we note that a surface is a collection of lattice sites and in general not all sites are equal. It is reasonable to assume that the distribution of receptors on a surface is uniformly random, therefore the number of receptors per lattice site follows a Poisson distribution

$$p(n_R, \Gamma) = \frac{\Gamma^{n_R}}{n_R!} e^{-\Gamma}, \tag{2.22}$$

with Γ the average number of receptors per site. If the surface is composed of N independent lattice sites, the grand partition function for the whole surface $\mathbf{\Xi}(\Gamma)$ reads

$$\mathbf{\Xi}(\Gamma) = \prod_{n_R=1}^{\infty} \Xi(n_R)^{p(n_R,\Gamma)N} \tag{2.23}$$

and the average surface coverage is then given by

$$\langle\theta\rangle = \sum_{n_R=1}^{\infty} p(n_R,\Gamma)\theta(n_R). \tag{2.24}$$

We stress that for the range of parameters studied (Figure 2.7) the number of receptors per site was always large $n_R > 100$ and the Poisson distribution (2.22) is highly peaked around $n_R = \Gamma$. Therefore the relative fluctuations in the surface receptor density are small and we have approximately $\langle\theta\rangle \approx \theta(\Gamma)$. We have included Poisson fluctuations when calculating the theoretical adsorption curves on Figure 2.7.

2.2.1 Simulations

Monte Carlo simulations are performed using a soft-blob model for polymers [77] described in more detail in the Methods chapter. Briefly, the model represents flexible polymers in a good solvent by a series of Gaussian soft blobs connected via harmonic springs. This model accurately describes polymers in the scaling regime, it also assumes that each individual blob represents a polymer in the scaling regime, hence blobs must be large enough to contain at least a few polymer Kuhn segments. On the other hand the model only considers pairwise interactions and does not take into account any 3-body effects. Therefore, the model is appropriate for studying dilute and semi-dilute polymer solutions where the blob density does not exceed 1 blob per blob volume $\rho_{blob} < 3/(4\pi r_b^3)$, with r_b the blob radius of gyration, in which case 3-body effects can be neglected.

A great feature of the soft blob model are the transferable potentials, we can represent a given polymer by series of many small blobs, a few larger ones or a single large blob and the form of the interaction potentials does not change [77, 78]. The universal form of the interaction follows because the interpenetration and surface repulsion of flexible polymers converges to a well defined form when the number of Kuhn segments in the polymer is large [73], the effective potentials are described in the Methods chapter. The radius of gyration of such a polymer is given by

$$R_g = r_b N_b^{\nu}, \tag{2.25}$$

with r_b the blob radius of gyration, N_b the number of blobs per polymer and $\nu = 0.588$ the scaling exponent of a 3D self-avoiding random walk [73].

The ligand-receptor bonding free energy is denoted by ϵ. This includes the pure ligand-receptor interaction, but it also depends on the size of the blob r_b and on the linker properties. We assume that when a ligand is unbound it can explore all of the space within a blob and different ligands in a blob are uncorrelated. We effectively use the same rationale to calculate the single bond strength as in the analytical theory above (2.9), except that now we apply it to individual blobs within a polymer

$$\epsilon = k_B T \ln\left(\frac{4\pi r_b^3 N_A}{3K_A}\right) + U_{poly}(r_b), \tag{2.26}$$

where we find the familiar U_{poly} term which takes into account (i) the correction to our approximation that ligand positions are uncorrelated within each blob and (ii) the effects of the linker between the ligand and the polymer backbone.

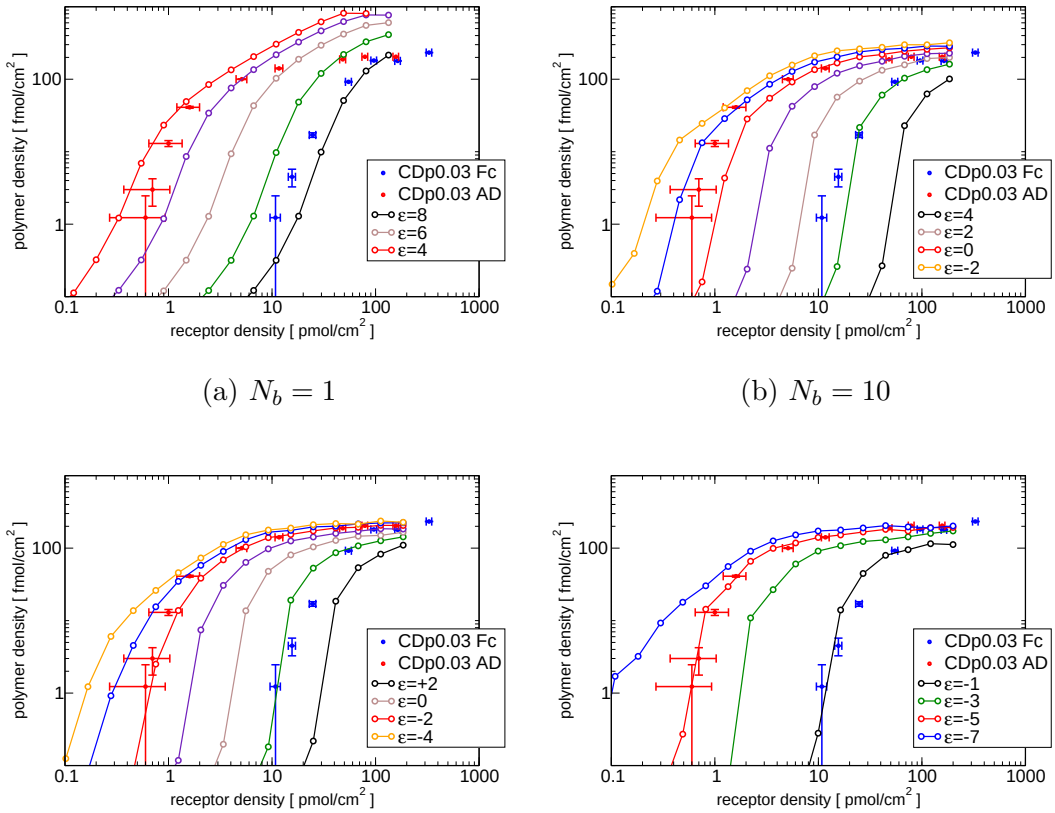
U_{poly} will depend on the chosen size of the blob r_b. The larger the blob the more coarse-grained the polymer representation is and the corrections in (i) become larger, hence we expect U_{poly} to increase with increasing blob size. In the simulation model only receptors within a cut-off distance r_b of a blob centre-of-mass can bind to the ligands of that particular blob. Choosing to represent the polymer with a single blob $N_b = 1$ we obtain $r_b = R_g$ and the expression for the bond free energy (2.26) becomes identical to the one used in analytical theory (2.9) . We expect that theory and simulations will give very similar results. There are still differences between analytical theory and simulations at $N_b = 1$ in that simulations are off-lattice and the simulation model also provides a more accurate account of the blob-surface repulsion.

We show the simulation results and compare them to experimental data on polymer adsorption by Dubacheva et. al. [15, 16]. All parameters in the simulation are determined to correspond to the experimental system: The polymer radius of gyration is assumed to be $R_g = 45$nm, the average number of ligands per polymer is $k_{low} = 27$ for low valency polymer and $k_{high} = 187$ for the high valency polymer, the chemical potential μ of the reservoir is set such as to result in a polymer concentration in bulk solution $\rho = 0.12\mu$M which corresponds to a rescaled density of $\rho^* = 0.025/a^3$.

First we systematically vary the number of blobs per polymer N_b the ligand-receptor binding energy ϵ and the receptor surface density Γ_G. The results for the low valency polymer ($k = 27$) are summarised in the plots on Figure 2.8. Figure 2.9 also shows matching simulation snapshots.

Each simulation was initiated with an empty box and ran for $\sim 10^{11}$ MC cycles. The equilibrium number of polymers was determined by averaging over the number of adsorbed polymers in the second half of the simulation run. We tested for convergence by considering the average number of polymers in intervals of 10^{10} MC cycles. Simulations were not considered converged as long as this number was increasing over subsequent intervals; when the averaged number of polymers started fluctuating around a given value, we considered the simulations sufficiently equilibrated. We also monitored the relaxation of the average number of bonds per polymer, which is related to the brush conformation. Simulations with $N_b = 20$ blobs per polymer reached equilibrium. Using longer polymers ($N_b = 50$) simulations at the largest polymer coverages did not reach equilibrium within $2{\cdot}10^{11}$ MC cycles. However, simulations did converge for lower polymer coverages (polymer density < 100 fmol/cm^2) and the snapshots shown in Figure 2.9 are representative of an equilibrium configuration.

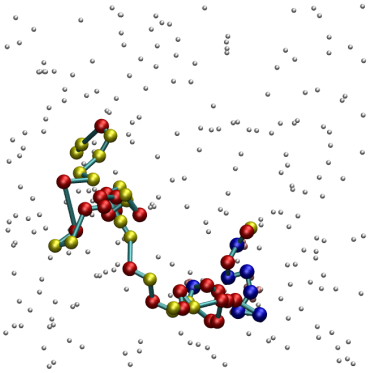
The agreement between simulations and theory is quite remarkable taking into account that the only free parameter in linking simulations and experiments is U_{poly} which simply



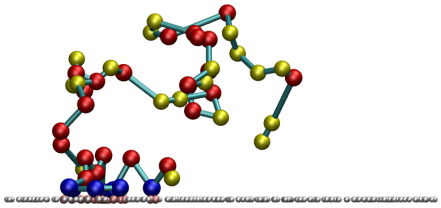
(a) $N_b = 1$ (b) $N_b = 10$

(c) $N_b = 20$ (d) $N_b = 50$

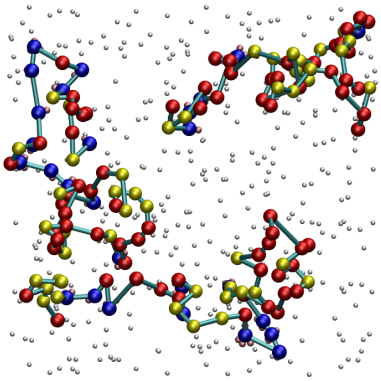
Fig. 2.8 Polymer adsorption simulations. Varying the chain length N_b, while keeping the polymer size R_g fixed. Binding isotherms obtained from simulations at polymer valency $k = 27$ and various chain lengths N_b and ligand-receptor binding free energies ϵ. Separate curves on individual plots correspond to different bond strengths ϵ incremented in steps of $1k_BT$ (only every second curve is labeled). Experimental data [16] on β-Cyclodextrin (CD) functionalised polymers binding to Ferrocene (Fc) or adamantane (AD) functionalised surface is represented with solid points and error bars. Simulation results are shown as solid lines, each data point (circle) corresponds to an individual simulation runs.



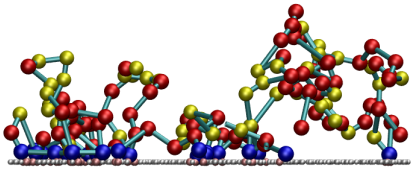
(a) 1 chain, $n_R = 244$, top view



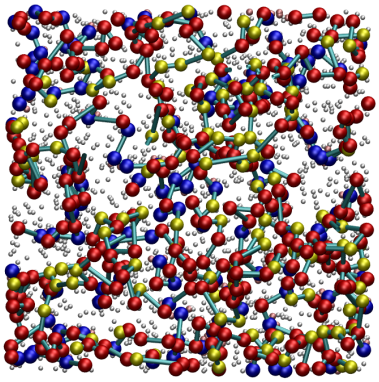
(b) 1 chain, $n_R = 244$, side view



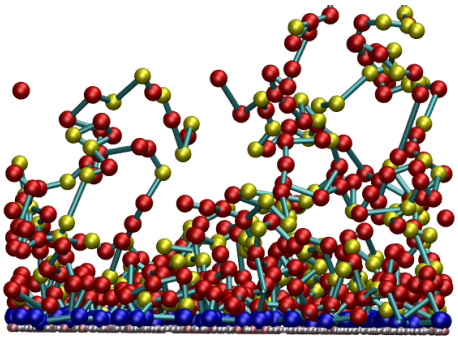
(c) 3 chains, $n_R = 403$, top view



(d) 3 chains, $n_R = 403$, side view



(e) 12 chains, $n_R = 1808$, top view



(f) 12 chains, $n_R = 1808$, side view

Fig. 2.9 Polymer simulation snapshots. The snapshots of system configurations for $N_b = 50$, $\epsilon = -3k_BT$ and $k = 27$ (corresponding to a curve on Figure 2.8**d)**). Different snapshots show different surface densities of receptors in the system, left panel is the top view, and right panel the side view. The colour code is the following: Unbound blobs (red), bound blobs (blue), unbound blobs carrying at least a single ligand (yellow). Receptors are represented as small white points on the surface, bound receptors are pink. The system size is $3R_g$ in lateral directions.

shifts adsorption curves to higher or lower receptor densities. The vertical position of the plots, the slope (selectivity) and the saturation plateau are all accurately predicted by the model. Also the shift in the adsorption isotherm between the high and low affinity polymer is accurately predicted to be $3k_BT$.

We observe that the results for different chain lengths (N_b) are fairly consistent which is to be expected. The simulations capture the plateau in the polymer adsorption which the theory does not. Simulations also overshoot the plateau when we represent the polymer with only a single blob ($N_b = 1$). In this case the blob density is outside the blob-model range $\rho > 3/(4\pi r_b^3)$ for high polymer coverage, the simulation model underestimates the polymer-polymer repulsion. The single blob simulations also fail to take into account the correct polymer conformation at the surface.

On the other hand representing the polymer with large number of blobs also approaches the limits of the model validity because we assume that each blob represents a segment of the polymer which is itself in a scaling regime. With $N_b = 50$ blobs per polymer each blob represents less then 2 Kuhn segments (assuming 64 Kuhn segments per polymer, obtained from experimental data [16] of the contour length $l_c = 900$nm and the Kuhn segment $b = 14$nm) which approaches the the limit of the model validity.

Simulation results consistently show higher selectivity than theoretical data. Figure 2.10 shows the selectivity α (2.7) obtained from the analytical model (Figure 2.7) and simulations (Figure 2.8**c)**). Simulations consistently predict a higher selectivity than the theory. To explain the discrepancy we perform free-energy calculations of the simulated system.

2.2.2 Free energy calculations

Using the soft blob model described above we can perform free energy calculations using the Wang-Landau technique [79] and determine the free energy $F_{wl}(\lambda)$ as a function of the number of formed bonds λ. By comparing the result to the analytical theory we obtain the polymeric degrees of freedom contribution to the free energy per bond: $U_{poly}(\lambda)$ and its dependence on the number of formed bonds λ for various number of blobs per polymer N_b. Using this technique we can only calculate the configurational part $U_{poly}(\lambda)^c$, i.e. the polymer backbone degrees of freedom effect, of the free energy

$$U_{poly}(\lambda) = U^c_{poly}(\lambda) + U_{linker} \ . \tag{2.27}$$

We cannot calculate any effects of the chemical linker between the ligand the polymer backbone, however, the linker contribution U_{linker}, which captures the effects of the linker

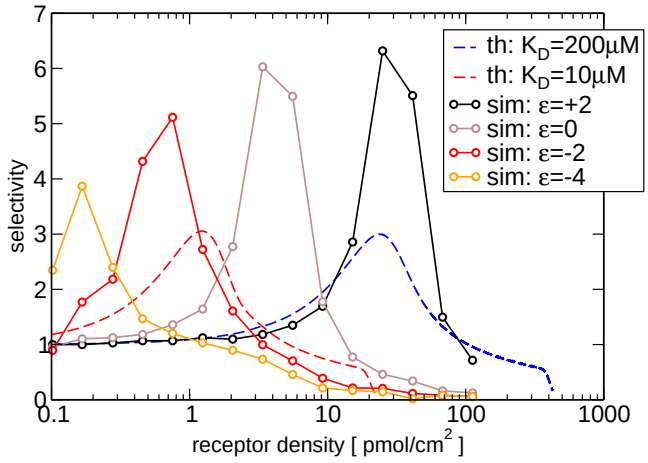


Fig. 2.10 Selectivity comparison between simulations and theory. The curves show the selectivity α (2.7) obtained from theoretical (Figure 2.7) and simulation (Figure 2.8**c)**) data. Simulations (solid lines) consistently predict higher selectivity than the analytical model (dashed lines).

on the single bond strength, is a trivial constant which does not depend on the number of formed bonds.

The configurational contribution measures the effects of additional polymeric degrees of freedom when we change from a single blob picture of the polymer ($U_{poly}(\lambda; R_g)$, captured by the analytical theory), to a more detailed picture $U_{poly}(\lambda; r_b)$, where the polymer is represented by a series of N_b blobs, each of radius r_b:

$$U^c_{poly}(\lambda) = U_{poly}(\lambda; R_g) - U_{poly}(\lambda; r_b) \; . \tag{2.28}$$

Moreover, we cannot calculate any effects of U_{poly} on a length-scale smaller than the blob size r_b, however we expect $U_{poly} \to const.$, a constant independent of the number of formed bonds λ, as the blob size is decreased sufficiently ($N_b > k$). In order to have a consistent comparison with the theory [6], the free energy calculations are performed on a polymer whose centre-of-mass is constraint to lie within $a = R_g(4\pi/3)^{1/3} = 1.61R_g$ of the surface. The particular choice will not affect the free energy profile, it will only affect the free energy of forming the first bond compared to zero bonds $F(1) - F(0)$, as once the polymer is bound the surface, the probability of the polymer centre of mass position spontaneously increasing to beyond $1.61R_g$ of the surface, is negligible.

[6] In the analytical theory (Figure 2.6) we require that an unbound polymer first adsorbs within a lattice site of linear dimension a, in order to start forming bonds.

Once we obtain the free energy profile $F_{wl}(\lambda)$ from simulations we compare it to the full partition function of the analytical theory (2.11). The free-energy calculations were performed at blob-receptor binding energy $\epsilon = 0$, we need to take account of this energy as it depends on the blob size $\epsilon(r_b) = \ln\left(K_D N_A (4\pi r_b/3)^3\right) + U_{poly}(r_b)$. We are only interested in how the configurational cost of binding changes with the number of formed bonds. For a given number of formed bonds λ we equate the free energy obtained from simulations, $F_{wl}(\lambda)$, with the analytical expression (2.11), we obtain

$$e^{-\beta[F_{wl}(\lambda)+\lambda\epsilon(r_b)]} = \binom{n_R}{\lambda}\binom{k}{\lambda}\lambda!\; e^{-\beta\lambda\epsilon(R_g)} \; . \tag{2.29}$$

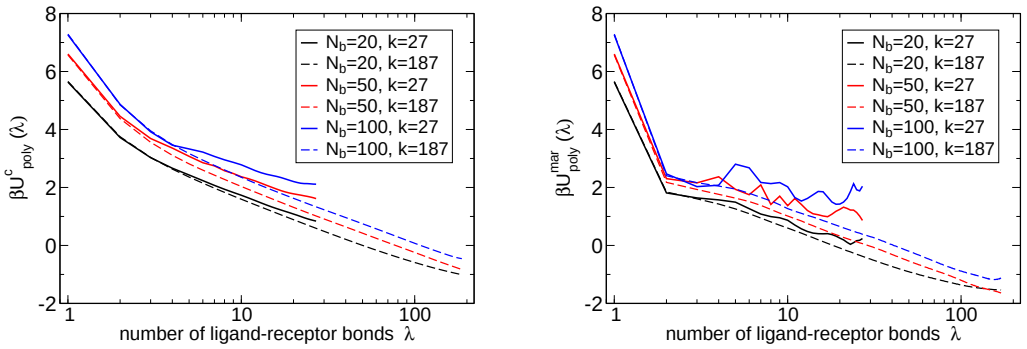
We remember n_R is the number receptors per lattice area a^2, k the number of ligands on the polymer and ϵ the bond free energy (2.26). We rewrite the above equality using (2.26-2.28) and the self avoiding walk relation $R_g = r_b N_b^{\nu}$, with the scaling exponent $\nu = 0.588$, as

$$\beta U^c_{poly}(\lambda) = \frac{1}{\lambda} F_{wl}(\lambda) + \frac{1}{\lambda} \ln\left(\frac{n_R! k!}{(n_R-\lambda)!(k-\lambda)!\lambda!}\right) - 3\nu \ln(N_b). \tag{2.30}$$

U^c_{poly} depends on three terms, the first term on the right is the free energy calculated from simulations, the second term takes into account all possible configurations considered by the theory (2.11) and the third term takes into account the blob size chosen in simulations.

The free energy calculations were performed for various chain lengths N_b, each was averaged over 30 different random realisations of the ligand positions on a chain. The calculated results for U^c_{poly} are shown on Figure 2.11. The first thing we observe is that U^c_{poly} is not a constant as we assumed in analytical theory, rather it is monotonically decreasing with the number of formed bonds λ. This means that subsequent ligand-receptor binding is "cooperative-like" – individual bonds become stronger the more bonds form. The added cooperativity increases the selectivity α of the multivalent polymer binding and explains why the simulations predict higher selectivity then analytical theory (Figure 2.10). This suggests that polymers are inherently more selective then equivalently designed nanoparticles coated with independent ligand arms, for which we expect the configurational cost to be linear[7] in the number of formed bonds. The variation of U_{poly} implies a variation in f and K_{intra} with the number of formed bonds, this means that our initial model (2.2) is only approximate. However, the non constant f does not affect general rules about designing selectivity discussed below, rather it only affects the exact shape of the adsorption curve.

[7]The configurational cost can be written as: $U^c_{poly}(\lambda) = \lambda U^c_{poly}(1)$



(a) (b)

Fig. 2.11 $U^c_{poly}(\lambda)$ calculated from simulations for various chain lengths N_b and two different numbers of ligands per chain: $k = 27$ (full lines), and $k = 187$ (dashed lines). (a) shows U^c_{poly} calculated from (2.30), which corresponds to the average free energy per bond formed. (b) Alternatively we can represent a "marginal" $U^{mar}_{poly}(\lambda)$ defined as $\sum_{i=1}^{\lambda} U^{mar}_{poly}(i) = \lambda U^c_{poly}(\lambda)$, that is the free energy of forming a next bond when there are already λ bonds present. This data is noisier because we essentially take a finite difference derivative of the data in (a). In other words, $U^c_{poly}(\lambda)$ is a cumulative average of $U^{mar}_{poly}(\lambda)$.

Furthermore we observe that U^c_{poly} depends slightly on the blob size $r_b = R_g N_b^{\nu}$ (via keeping R_g fixed and changing N_b). If only a single blob is used $N_b = 1$ there can be no cooperative behaviour. The more "polymer-like" the model (larger the number of blobs in a chain), the large the cooperative effect is. The free energy of the first bond formation depends only on the number of blobs and is logarithmic $U^c_{poly}(1) \propto \log(N_b)$ [80].

The calculated U^c_{poly} decreases slightly with increasing valency, It appears that this could be the explanation for why the analytical theory cannot accurately predict the shift for the high valency polymer, Figure 2.7. The error in analytical theory is $\approx 0.7 k_B T$ which corresponds to the difference in U^c_{poly} between low and high valency on Figure 2.11.

2.3 Which systems are super-selective ?

The discussion thus far focused on selective adsorption of multivalent particles and polymers. We now generalize our treatment and discuss various practical systems. In particular, we will discuss the key role of disorder that is needed to observe super-selective behaviour in multivalent interactions. Specifically, what is needed is that a multivalent entity can bind in many different ways to a receptor-decorated substrate. This kind of

disorder is usually not possible for multivalent interactions on the angstrom or nanometer scale, as the interacting units tend to be effectively rigid on that scale. In contrast, larger supramolecular systems (e.g. the binding of a multivalent polymer to a receptor decorated membrane) can sustain the 'disordered' interactions.

2.3.1 Rigid geometry interactions

A prototypical example of multivalent interactions is the fixed (rigid) geometry multivalency shown in Figure 2.12. Two rigid, multivalent entities bind via multiple bonds: as the geometry is fixed the individual bonds either fit together, or they do not. Examples of this kind of interaction include base pairing interaction in DNA.

Another well-known example of a rigid multivalent interaction is the binding between an enzyme and a substrate. The interaction between a pair of proteins is multivalent, as it involves a number of local interactions of various types (hydrogen bonding, hydrophobic, Van der Waals, electrostatic etc). To a first approximation the enzyme and substrate can be described as rigid objects. This is a simplification as proteins, even in their native state, are not entirely rigid. In any given relative orientation of the ligand to a substrate we find a 2D equivalent of the Figure 2.12. We name this class of multivalent interactions "rigid geometry multivalency".

Due to the lack of flexibility of individual bonds, rigid multivalency will generally not show super-selective behaviour. To understand this, consider a simple one-dimensional example of a sequence of rigidly positioned ligands that bind to a commensurate sequence of receptors. One cannot increase the binding site density on the substrate without breaking the commensurability of the binding. Hence, increasing the receptor density will normally decrease the binding strength. In other words: commensurate lock-and-key interactions are not super-selective. Interestingly, it seems that the ability of rigid multivalent particles to detect commensurate structures is exploited in nature, for instance in the activation of certain Toll-like receptors [72]. We will explore rigid multivalency in the context of molecular imprinting in Chapter 4.

The simplest mean-field model for the commensurable binding case (Figure 2.12) is that every bond pair is equivalent and can be either formed (weight e^{-f/k_BT}) or not (weight 1), and all l bond pairs are independent. The avidity constant K_A^{avfix} of the multivalent interaction is proportional to the bound partition function q_b^{fix} taking into account all possible states

$$K_A^{avfix} \propto q_b^{fix} \approx \left(1 + e^{-f/k_BT}\right)^l \quad . \tag{2.31}$$

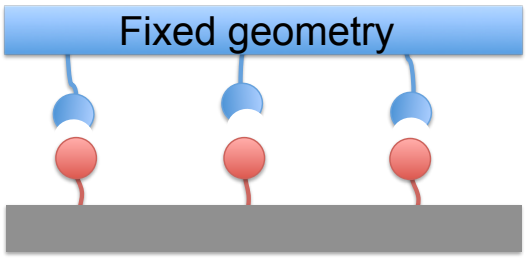


Fig. 2.12 Rigid geometry multivalency. The cartoon presents a prototypical fixed geometry interaction where bonds are commensurate (e.g. DNA base pairing or enzyme-substrate interactions). Such systems generally do not exhibit super-selective behaviour as we cannot increase the (binding site) density on one multivalent entity (substrate) without breaking the commensurability of the bonds. See also Eq. (2.31) and the discussion in the corresponding box.

Evidently the avidity constant is very sensitive to the number of possible bond pairs l, the temperature T and the individual bond strength f. The number of possible pairs l will depend on the geometry of the system. In the simplest model the number of pairs is given by $l = \min[n_R, k]$, it is limited by whichever, substrate or the multivalent ligand has a smaller number of sites. [10]. Hence, rigid geometry multivalent interactions can show super-selective behaviour, but only when the multivalent construct initially had an excess number of binding sites compared to the substrate. Furthermore, when increasing the number of binding sites on the substrate, geometric constraints (commensurability) must be obeyed.

2.3.2 Disordered multivalency

Super-selective behaviour can be exhibited by multivalent systems that can increase the number of possible bonds as the density of receptors increases. As we saw above, fully ordered multivalent systems only bind optimally to commensurate receptor arrangements. To achieve super-selectivity, we typically need some kind of disorder or randomness in the geometry of binding. The ability to increase the number of bonds with increasing receptor density can be due to: (i) long, flexible binders, (ii) mobile receptors, or (iii) random binder positions. Figure 2.13 shows schematic examples of these three cases. Different types of bond disorder will result in different expressions for the bound partition functions (and therefore, for the avidity constants), see Eqs. (2.32 - 2.34). However, they all show similar super-selective behaviour (see Figure 2.14).

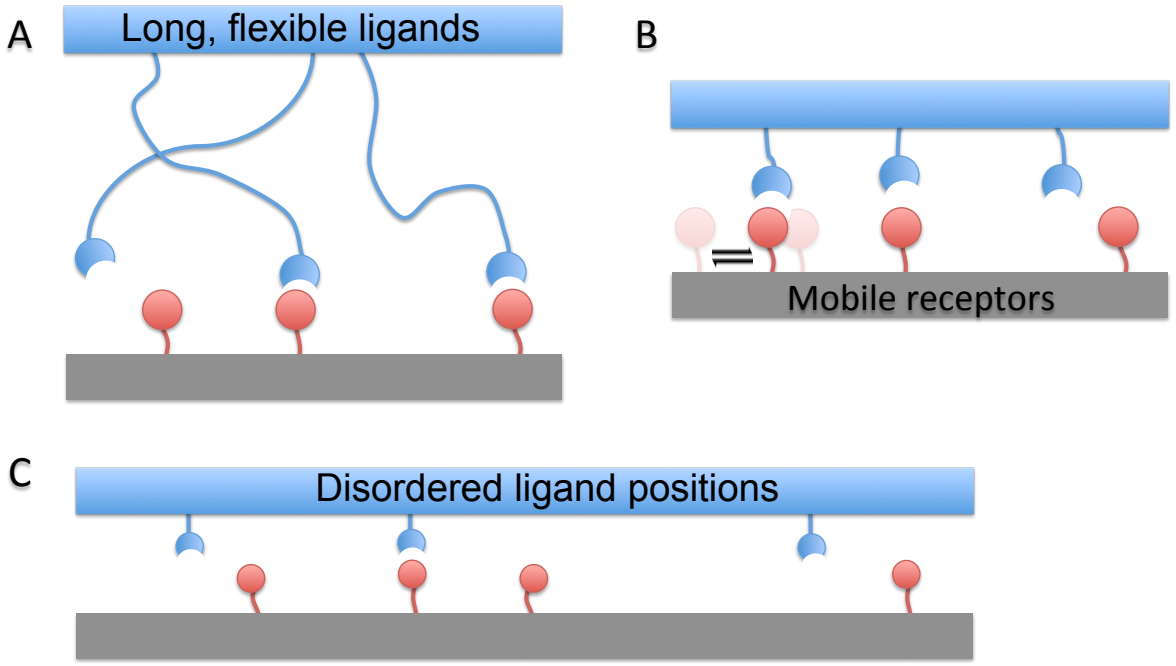


Fig. 2.13 Disordered multivalent systems. Three characteristic types of multivalent interactions are shown: **A)** long, flexible binders, **B)** mobile receptors, **C)** disordered, random positions of individual binders. Different types can behave slightly differently, see the adsorption profiles in Figure 2.14. However, they all exhibit super-selectivity, and consequently, any practical system that is similar to at least one of them, will be super-selective.

Different forms of disorder may cause super-selective behaviour in multivalent systems. The theoretical expressions for the partition function (and hence the avidity constant) of the bound state will depend on the nature of the disorder. Below, we list a few examples:

- Long flexible ligands, Figure 2.13A); the number of ligands and receptors is fixed and all k ligands can reach any of the n_R receptors

$$q_b(n_R,k) = \sum_{\lambda=1}^{\min(n_R,k)} \binom{n_R}{\lambda}\binom{k}{\lambda}\lambda!\; e^{-\beta\lambda f} \;, \tag{2.32}$$

  which is the expression that we have already used above (see Eqs. (2.2,2.3)).

- Mobile receptors (Figure 2.13B)); the number n_R of accessible receptors fluctuates and, neglecting receptor-receptor interactions, will be Poisson distributed with mean $\tilde{n}_R$. Poisson averaging of (2.32) over n_R, we find

$$q_b(\tilde{n}_R,k) = \left(1+\tilde{n}_R e^{-\beta f}\right)^k - 1 \;, \tag{2.33}$$

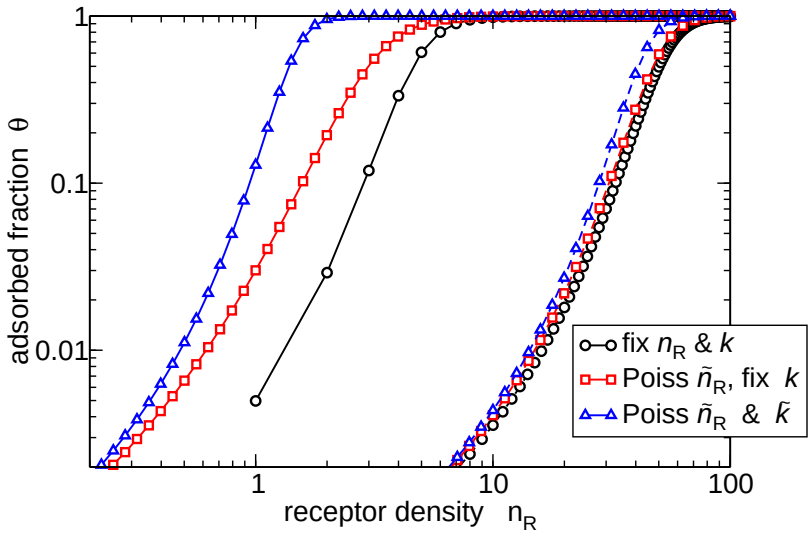


Fig. 2.14 Adsorption profile for different disordered systems, depicted in Figure 2.13. Systems show qualitatively similar behaviour, all are super-selective. For large number of bonds the adsorption profiles converge. We used expressions (2.32,2.33, 2.34) with $k = 5$ and $\beta f = 0$ to compute adsorption isotherm (2.5) in the case of a few strong bonds (solid lines). To represent the case of many weak bonds (dashed lines), we assumed the case of weak binding, we used the same equations but assumed $k = 25$ $\beta f = 5$. The activity was kept fixed at $z = 0.001$.

which is the same equation that we obtained as an approximation before (2.2, 2.6). Mathematical derivation of (2.33) will be provided in the next chapter.

- Colloids, nanoparticles (or cells) with disordered or mobile ligand positions (Figure 2.13C)); a random grafting process of a nanoparticle will result in an approximately Poisson spatial distribution of the grafted ligand positions. Both the number of ligands and the number of receptors are Poisson distributed with mean $\tilde{k}$ and $\tilde{n}_R$ respectively. Poisson averaging (2.33) over k, we find

$$q_b(\tilde{n}_R, \tilde{k}) = e^{\tilde{n}_R \tilde{k} e^{-\beta f}} - 1 \, . \tag{2.34}$$

A comparison between the predicted behaviour of these different systems is shown in Figure 2.14. In the limit of weak bonds $n_R e^{-\beta f} < 1$, the behaviour of all systems converge to the same form. Mathematical derivation of the above expressions is presented in the following chapter.

2.3.3 Receptor mobility

At first sight, it would seem that the case of mobile receptors shown in Figure 2.13B) should be rather different from the immobile case. However, since the receptors are mobile, each ligand can, in principle, bind to any receptor. In this light the two problems become very similar. Another way of looking at the system with mobile receptors is to consider the receptors as a (two-dimensional) 'ideal gas' of particles that can bind to the ligands with an interaction strength f. Up to a concentration-independent term μ_R^0, the chemical potential of these receptors is given by $\mu_R \approx k_B T \log(n_R)$. A small change in the receptor concentration n_R will lead to small change in the chemical potential μ_R, which will alter the probability of each and every individual ligand binding. For multivalent particles a small change per ligand adds up to a large change per particle. If we assume that there are many more receptors than ligands, we can then write the bound-state partition function for k ligands as $q_b \approx \left(1 + C n_R e^{-\beta f}\right)^k - 1$, where the constant C depends only on the concentration-independent part of μ_R; $(C = -k_B T \ln \mu_R^0)$. Clearly, the binding probability depends on n_R, see Refs. [72, 81] for practical examples of super-selectivity with mobile receptors. We note that for dilute receptors the chemical potential is dominated by the translational entropy. Hence, the origin of super-selectivity is entropic, also for mobile receptors.

In the case of many weak binders, $C n_R e^{-\beta f} \ll 1$ and hence we can write $q_b \approx e^{C k n_R e^{-\beta f}} - 1$, obtaining a double exponential dependence on the bond strength f. This highly non-linear dependence can be exploited for immune amplification [72] where the immune response is regulated via TLR9 receptor activation binding to DNA-peptide clusters. Changing the properties of the DNA-peptide cluster (the spacing between DNA helices) causes only a small change in the TLR9 to DNA single bond strength f. However, due to the double exponential dependance, the binding of the whole multivalent cluster, and hence the immune activation, becomes extremely sensitive to small changes in f.

We stress that the relevant number density of mobile receptors n_R is actually the concentration of free (unbound) receptors. Hence, $n_R = n_R^{tot} - n_R^b$ is not a constant, rather, it depends on the concentration of bound receptors n_R^b (i.e. the number of bound receptors per lattice site area), where the total number n_R^{tot} is a constant. The concentration of bound receptors can be obtained by noting that the derivative $\frac{\partial q_b}{\partial \beta \mu_R}$ returns the average number of bonds per adsorbed particle, therefore, the number of bound receptors is

$$n_R^b = \theta \frac{\partial q_b}{\partial(\beta \mu_R)} \,, \tag{2.35}$$

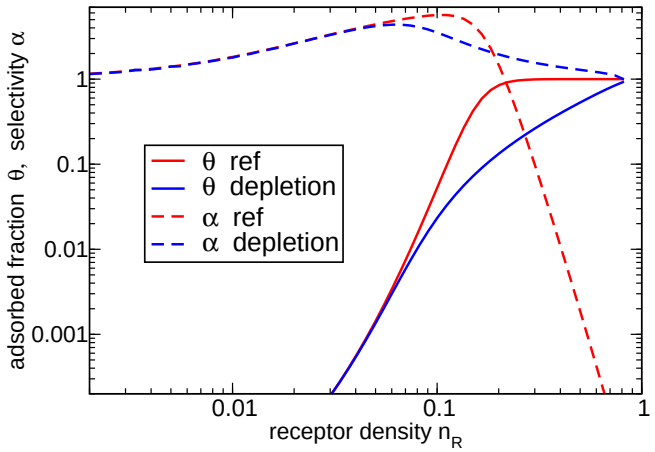


Fig. 2.15 Effect of receptor depletion. Red curves show the reference adsorption profile (solid line) and selectivity (dashed line) where receptors are not depleted $n_R = n_R^{tot}$. Blue curves show that receptor depletion leads to reduced selectivity. In the depletion limited regime each additional multivalent particles binds approximately a fixed amount of receptors. Hence, the relation between the number of adsorbed particles and total number of receptors becomes linear in the depletion limited regime. Parameters: $k = 9$, $f = -3$, $\ln(z) = -15$.

with the chemical potential $\ln(\beta\mu_R) \propto n_R^{tot} - n_R^b$ and q_b the bound partition function (2.33). The above equation must be solved self consistently for the number of formed bonds n_R^b. This leads to anti-cooperative behaviour; adsorbed particles recruit and deplete receptors, inhibiting the adsorption of further particles. Figure 2.15 shows a comparison between adsorption with and without receptor depletion. At higher surface densities of adsorbed multivalent entities the receptors are depleted and selectivity is suppressed. Our previous calculation shown on Figure 2.14 corresponds to a situation where the number of receptors is sufficiently large such that the depletion effects can be neglected $n_R^{tot} \gg n_R^b$. Depletion effect will be strong when the fraction of bound receptors is appreciable.

Finally, for many immobile, but randomly distributed binders, shown in Figure 2.13C) the intuitive reasoning for super-selectivity follows from our initial discussion in the introduction (1.2). Let us consider two ligand/receptor-decorated multivalent nanoparticles, A and B that can attach through ligand-receptor binding. The binding moieties are randomly distributed on both nanoparticles. From a point of view of a particular ligand on particle A, the probability of it binding, denoted by p_{1A}, is to a first approximation linear in the density n_R of complementary receptors on particle B. The number of possible bonds in the contact area is proportional to the number of ligands k in that area. The

net result is that the binding probability depends exponentially on the product kn_R, as would follow from (2.34).

We note that in the cases of fixed short ligands we have only illustrated and discussed the two limiting cases: (i) the perfectly complementary rigid interaction (Figure 2.12) and (ii) Poisson randomly distributed case (Figure 2.13C). Practical systems will fall between these two extremes. As a rule of thumb, small molecules and macromolecules, such as DNA or proteins, or virus capsids have a rather well defined geometry and we expect their interactions to be closer to the rigid geometry case. On the other hand, the spatial distribution of binders (ligands) on entities larger than a few nano-metres is, in general, more disordered; be they man-made such as DNA coated colloids [82–84], or natural such as cells.

We have presented simple analytical models that can be used to rationalise and understand super-selectivity in various multivalent systems. In the case of polymers, the simple model works very well (see Figure 2.7). However, certain systems have been studied in a greater detail. For these cases, more sophisticated (and more complex) models have been developed. For example, cell endocytosis of a virus is mediated by a multivalent interaction between membrane proteins (receptors) and virus capsid proteins (ligands). But to model the process, one should account for membrane elasticity and, in some cases, also for active processes [85]. More detailed models of multivalent polymer adsorption have recently been developed [80, 86]. A theory of valence-limited interactions explicitly taking into account specific positions and different types of tethered binders requires the self-consistent solution of a system of equations [9, 87], the framework was also extended to mobile ligands [88] and multimeric complexes [89]. A complementary approach is based on a saddle-point approximation for the binding free energy [90]. We note that the results presented in these papers support the conclusions about super-selective behaviour that we have obtained here using much simpler models.

2.4 Design principles for super-selective targeting

Clearly, super-selective targeting has important practical applications (as even viruses seem to 'know'). It is therefore important to formulate design principles for achieving optimal super-selectivity. To formulate design rules, we start once again from the simple model described above (2.5): multivalent particle docking to a receptor-decorated surface (e.g. a cell). The density of receptors on the surface is again measured by n_R, the mean number of receptors in the contact area (i.e. the area accessible to a docked particle). In many cases of practical interest, we aim to target only those surfaces (e.g. a cell surface)

that have a receptor concentration above a certain threshold. How should we design the particle to target this surface optimally? Our control parameters are the valency k, the ligand-receptor binding strength f, and the activity of particles in solution z.

In terms of the theoretical expressions (2.5) (2.6), we aim to maximise the selectivity

$$\alpha(n_R) = \frac{\partial \log \theta}{\partial \log n_R} \,. \tag{2.36}$$

at a given desired receptor density n_R. We note that q_b (2.6) and its derivative are increasing functions of n_R, k and $-f$. Hence, we expect the selectivity (slope) to be the highest just before denominator in (2.5) becomes important and the maximal selectivity will be found when $zq_b \approx 1$. Using (2.6) we can solve this equation, which yields a relation between k and f

$$k = \frac{-\log(z)}{\log\left(1 + n_R e^{-\beta f}\right)} \,. \tag{2.37}$$

When $zq \approx 1$ we also have approximately $\theta \approx \frac{1}{2} zq$ and the selectivity becomes

$$\alpha \approx k \frac{n_R e^{-\beta f}}{1 + n_R e^{-\beta f}} = -\log(z) \frac{n_R e^{-\beta f}}{(1 + n_R e^{-\beta f}) \log\left(1 + n_R e^{-\beta f}\right)} \,, \tag{2.38}$$

where, in the last step, we used (2.37).

Expanding the above function to first order for weak/strong binding we find the characteristic behaviour: (i) In the case of strong binding the selectivity is

$$\alpha(n_R e^{-\beta f} > 1) \approx \frac{-\log(z)}{-\beta f + \log(n_R)} \,, \tag{2.39}$$

and in the weak binding limit

$$\alpha(n_R e^{-\beta f} < 1) \approx -\log(z) \,. \tag{2.40}$$

Clearly, the selectivity is maximal in the weak-binding limit and is determined by the logarithm of the activity, see landscape plots in Figure 2.16. In the strong-binding limit, the selectivity decreases with increasing strength of the individual bonds. We remember that $z = \rho \frac{K_A}{K_{intra}}$ and $e^{-\beta f} = K_{intra}$.

The landscape plots of selectivity as a function of the valency k and bond strength f are shown in Figure 2.16. We immediately notice three features: (i) High selectivity appears only in a small region of the parameter space, along the curve predicted by (2.37). (ii) The selectivity reaches a plateau value at large valencies k and weak individual bonds.

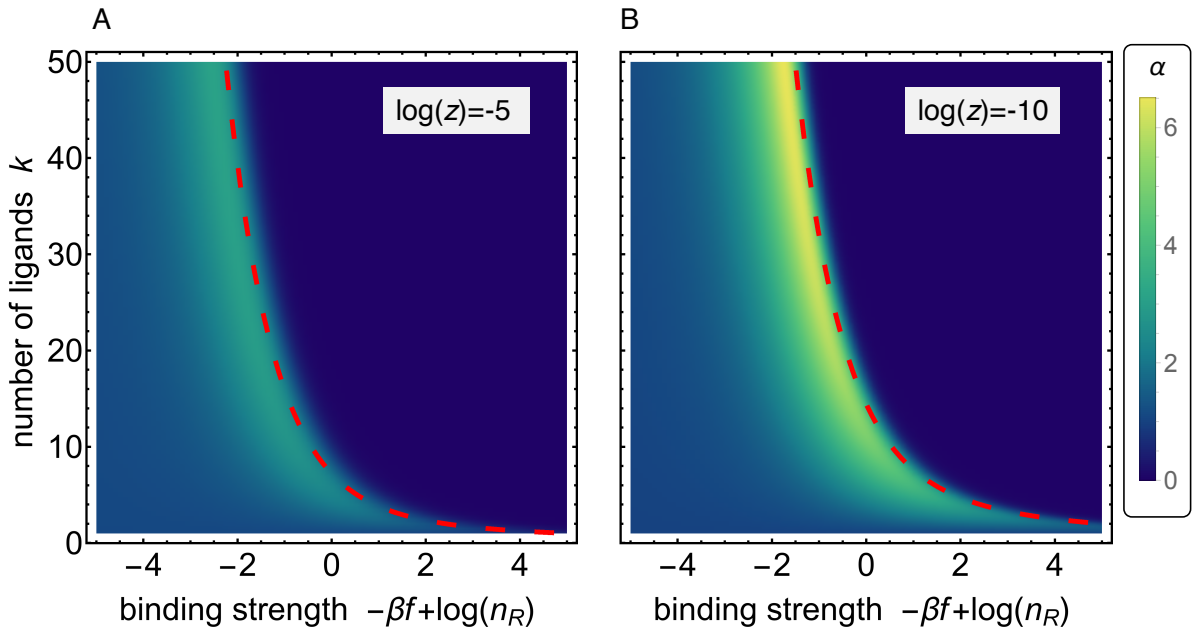


Fig. 2.16 The selectivity landscape as function of the valency k and the rescaled binding strength $-\beta f + \log(n_R)$. The landscape was obtained by calculating the selectivity α using (2.36). The activity of multivalent particles was chosen as: A) $z = \exp(-5)$, and B) $z = \exp(-10)$. Both plots use the same colour scale. The dashed curves represent the approximate optimal selectivity relation (2.37), which rather accurately fits the maximum selectivity region.

(iii) Maximum selectivity is limited by the activity z; lowering the activity (or density) of multivalent particles yields a higher selectivity.

The dimensionless activity $z = \rho \frac{K_A}{K_{intra}}$ depends on the density ρ, but also on the ratio of the equilibrium constants for the formation of the first bond and for the formation of subsequent ligand-receptor bonds in a particle-substrate complex (see Eqn. 2.2). Therefore, even at large densities, selectivity can be substantial if the ratio $\frac{K_A}{K_{intra}}$ is small. This can be achieved by adding a non-specific repulsion between the multivalent entities (for instance, by coating the particle with inert polymer that provides steric repulsion [91]). Such a repulsion would present a barrier to particle association but would not prevent additional bonds from forming once the barrier is overcome: the result would be a reduction in K_A due to repulsion, but as K_{intra} would be less affected, this steric repulsion would decrease the ratio $\frac{K_A}{K_{intra}}$.

Our calculations show that selectivity is suboptimal when using high affinity bonds. However, strong affinity multivalent constructs can still behave super-selectively ($\alpha > 1$) if their activity (concentration) in the solution is low enough (2.39). This suggests that in principle it is possible to design a super-selective system based on very strong affinity interactions, such as the biotin-streptavidin pair. However, extremely low required concentrations and slow kinetics (due to strong affinity and large activation barriers for bond formation and breaking) are likely to be prohibitive for practical applications.

Multivalency leads to super-selectivity, but it also leads to high sensitivity of binding to the variation in other relevant quantities (1.2). Therefore, in practical applications, it is important to control (or, at least know) parameters such as temperature, pH, ionic binding strength when using multivalent particles for selective targeting. The parameter range that yields high selectivity is rather small, see Figure 2.16B). A brute-force 'random' search in design-parameter space is, therefore, unlikely to find the optimal selectivity region. We hope that the theoretical guidelines and design principles set forth in this chapter will enable a more rational design of particles for super-selective targeting.

We condense the results shown in Figure 2.16 and our theoretical considerations (2.38), in a set of simple design rules for multivalent binding that yield maximum selectivity. We use our dimensionless statistical mechanics notation which is straightforwardly converted to chemical equilibrium units using $z = \rho \frac{K_A}{K_{intra}}$ and $e^{-\beta f} = K_{intra}$.

1. The maximal possible selectivity α is limited by the activity of multivalent particles in solution: $\alpha_{max} \sim -\log(z)$ so the activity z of multivalent binders should be small.

2. Many weak bonds are better than few strong ones. The selectivity is determined by the valency k, until a point of saturation given by $k \sim -\log(z)$. Hence, the valency k should be larger [8] than the negative logarithm of the activity: $k > -\log(z)$.

3. the relationship between the ligand number k and binding strength f should be obeyed: $k = \frac{-\log(z)}{\log(1+n_R e^{-\beta f})}$. Together with the above rule, this one states that to achieve maximal selectivity individual bonds should be very weak $K_{intra} = e^{-\beta f} < 1/n_R$. In other words, the fraction of bound receptors/ligands should always remain small.

We restate the main assumptions used to arrive at these design rules: (i) ligands are identical and bind independently, (ii) all ligands of a (surface bound) multivalent construct can reach all surface attached receptors, within a lattice site, but cannot bind to any receptor outside of the site (see Figure 2.4). (iii) Receptors, ligands or particles have no interactions except for the steric (hard-core) repulsion and ligand-receptor affinity.

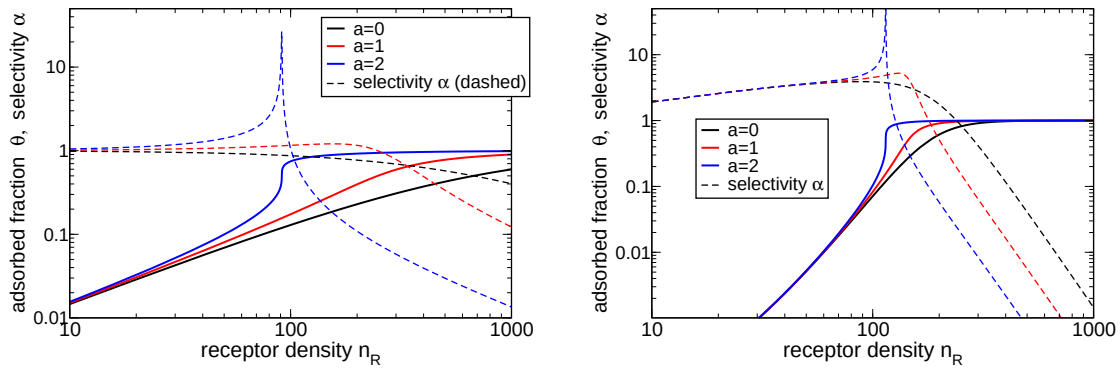
2.5 Langmuir raft

The selectivity can also be increased by a different mechanism to multivalency. In the case of attractive interactions between adsorbing particles, the particles will aggregate and form a raft. We assume that the attractive interactions are too weak to affect or cause aggregation in the bulk solution (bulk density is below the critical micelle concentration). However, upon particle adsorption, the 2D surface density is sufficiently increased for the weak attractive interactions to become important. The simplest way to account for interactions between the nano-particles is to assume a 2D van der Waals-like contribution to the partition function

$$q_{vdW} = e^{-\beta a_{vdW} N^2/A} = e^{-\beta a^*_{vdW} \theta^2} \tag{2.41}$$

with N the number of adsorbed particles per area A and a_{vdW} is the van der Waals attraction parameter. This equation can be rewritten using the surface coverage $\theta = a^2 N/A$ and the rescaled parameter $a^*_{vdW} = a_{vdW} \frac{A}{a^2}$. We have already used a to denote the lattice size, for historical reasons we use the same letter a_{vdW} to denote the van der Waals parameter. The advantage of the van der Waals equation of state is that it can describe a liquid-vapour transition, which, in this context, could be very interesting, as

[8] the first two design rules state that the maximal selectivity is limited by either the valency k or the $-\log(z)$, whichever is smaller.



(a) monovalent, $k = 1$ (b) multivalent, $k = 5$

Fig. 2.17 Attractive interactions between adsorbing particles increase the selectivity. The solid lines correspond to the adsorption profile and the dashed lines its derivative – the selectivity α. **a)** shows the monovalent ($k = 1$) and **b)** multivalent ($k = 5$) particle adsorption. $a^*_{vdW} = 0$ corresponds to the reference adsorption profile (Figure 2.3). The activity of the particles in solution was set at $z = 10^{-5}$. We observe that added attraction increases the selectivity but only in the regime where the surface density is appreciable. The selectivity diverges when we approach the critical point of a lattice gas-solid transition.

the 'response' of the system (adsorbed nano-particle concentration θ) in relation to a 'signal' (f, ρ or n_R) would diverge at the critical point.

We incorporate the van der Waals attraction simply by multiplying the bound partition function (2.6) with the van der Waals term (2.41). The equation must be solved self-consistently because the surface coverage θ appears in (2.41). The adsorption profiles on Figure 2.17 show an increase in selectivity as compared to the reference case ($a^*_{vdW} = 0$). The selectivity diverges when approaching the critical point.

The behaviour of our system is somewhat similar to the 2D Ising model-like transition that can be used to model the effects of receptor cooperativity [92]. The combination of multivalency and inter particle attraction seems to offer the highest selectivity.

2.6 Summary: it is interesting, but is it useful ?

We have shown that weak, multivalent interactions can result in a super-selective behaviour where the overall interaction strength becomes very sensitive to the concentration of individual binders (receptors). We presented a simple yet powerful analytical model that possess good predictive power for designing multivalent interactions. We expect

that, even in cases where the simple model fails quantitatively, the above design rules will still provide a good starting point for designing super-selectivity in practical multivalent systems. Figure 2.18 summarises advantages of weak multivalent interactions in selective targeting.

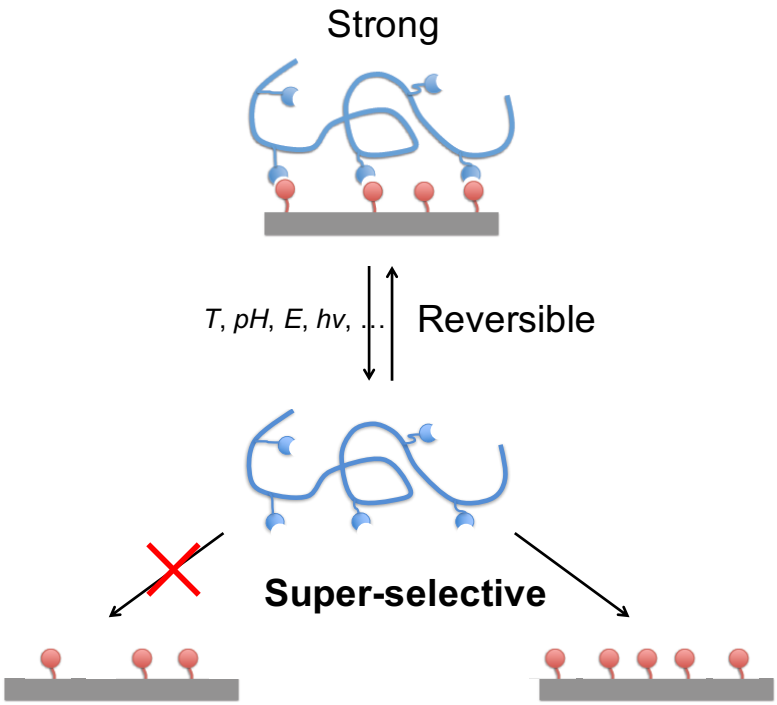


Fig. 2.18 Advantages of many weak bonds. Contrary to strong monovalent antibody-antigen interactions and covalent bonds, multiple weak complexes can be disassembled (one by one) using different environmental stimuli (Temperature, Potential, pH, light), which provides flexibility and reversibility, for example dendrimers [93], stimuli responsive coatings [94], renewable sensors for biomolecules [95], reversible gels [96] or gel-particle glue [97]. Due to multivalency effect the binding of the whole multivalent entity remains strong even when individual bonds are weak. The environmental stimuli can also be used to tune the super-selectivity region to the desired surface density of receptors. For example, we could exploit the acidic extracellular environment of the tumour tissue for drug targeting using multivalent particles.

We can imagine effective purification devices where nano objects of different valencies are passed through super-selective sieves. In the field of material self-assembly, multivalent supramolecular entities could be designed to hierarchically assemble depending on the valency, thus enhancing the precision of self-assembled constructs [83].

The ability to target diseased cells pathogens based on the surface concentration of certain (over)expressed receptors would be of huge practical importance [28]. At present,

the delivery of pharmaceutical compounds to specific cells is usually targeted based on the existence of a specific marker (e.g. a sugar or a peptide fragment) that is unique to the targeted cell type. The current wisdom seems to be to functionalise drugs or drug carriers such that they bind strongly to the specific marker. This strategy is fine if the target cells (e.g. bacteria) are very different from the cells of the ligand, and carry very different markers.

However, the strong-binding strategy becomes problematic if one wishes to target say, cancerous cells, which are usually very similar to our healthy cells and, typically over-express markers that are also present, be it in smaller quantities, on healthy cell surfaces. Examples are the CD44 ('don't eat me' receptor) or the folic receptor. In such cases, a compound that binds strongly to the over-expressed marker will also bind to (and kill) healthy cells. The insensitivity of strong binders to the surface concentration of markers is one of the main reasons why antibiotics can be efficient with few side effects (in most patients), while chemotherapy is directly harmful to our body.

As this chapter outlined, carefully designed multivalent drugs could be targeted super-selectively only to cells with cognate receptor concentration above a certain threshold value [10, 98]. Furthermore, in a living cell, receptor interactions and signalling also play a role which can further enhance the non-linear response of the system [21, 92, 99–102] and shearing, for example blood flow, enriches the system further [103, 104].

In a nutshell, multivalency extends the sensitivity of interactions into the receptor density domain. Moreover, it enables the design of specific, highly selective interactions based on the concentration of ligands or binders, as well as on their chemical nature, thus opening up the possibility for selective targeting with minimal side effects.

2.7 Intermezzo: What is effective molarity ?

Effective molarity (EM) is an empirical concept that is commonly used to relate the kinetics and equilibria of intramolecular and intermolecular reactions [65–67]. It is defined as

$$EM = \frac{{K_A}^{intra}}{{K_A}^{inter}} \ , \tag{2.42}$$

where ${K_A}^{intra}$ and ${K_A}^{inter}$ are the equilibrium association constants. EM has units of molar concentration and is a useful measure of multivalent interactions efficacy, see Figure 2.19. For example, when the concentration ρ of multivalent ligands in solution is high $\rho \gg EM$ multivalent effects are suppressed and ligands will bind monovalently. On the other hand when $\rho \ll EM$ multivalent interactions dominate over monovalent binding. Additionally, EM allows us to de-convolute the intra equilibrium constant into a simple part (K_A) due to bond formation, and a complicated part (EM) related to the change of conformational entropy and free energy upon binding, see Refs. [65–69] for more discussion.

However, it is important not to over-interpret the meaning of "effective" concentrations. The name suggests that we can calculate the internal chemical equilibria of multivalent interactions simply by using some effective concentrations of ligands. That, however, is not quite the case, as the expressions for association equilibrium between two compounds do not carry over to the situation when the numbers involved are small.

Let us consider a prototypical system: Only two particles (ligands) in a box with volume V. The particles can associate with an equilibrium constant K_A that was predetermined for us, see Figure 2.20. We wish to calculate the association probability of these two particles. To obtain the correct result we can calculate the partition functions of the bound/unbound state.

The unbound (configurational) partition function of two molecules in the box is

$$q_u = V^2 \ , \tag{2.43}$$

since we assume both particles are non-interacting and can independently explore the entire box volume V. The bound partition function is

$$q_b = V v_0 e^{-\beta \Delta G} \ , \tag{2.44}$$

with $e^{-\beta \Delta G} = K_A \rho_0$ the dimerisation free energy and $v_0 = \frac{1}{\rho_0 N_A}$ the microscopic volume of the bond and $\rho_0 = 1\text{M}$ the standard concentration. The ratio of the partition functions

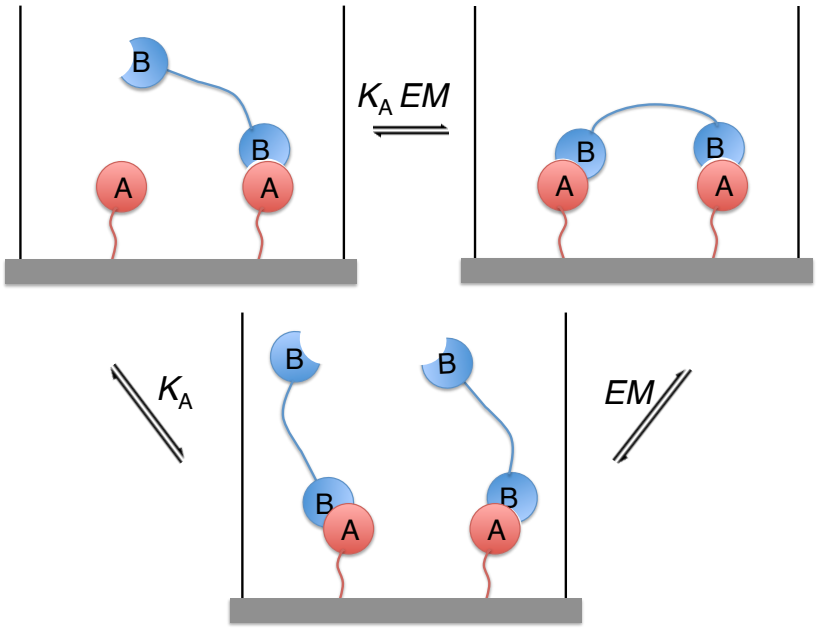


Fig. 2.19 The concept of effective molarity. The above cycle shows the 3 different states that divalent ligands (BB) can bind to two receptors (AA) (unbound state is omitted). We have 3 distinct states and, therefore, need 2 equilibrium constants to characterise the equilibrium properties of the system: K_A and EM. A product of the two is often called an intra association constant $K_A^{intra} = K_A EM$. A useful reference point is that for a divalent ligand/receptor system and saturated receptors, EM determines the concentration of divalent ligands $[BB]$ in solution at which we expect equal number of singly and doubly bonded ligands.

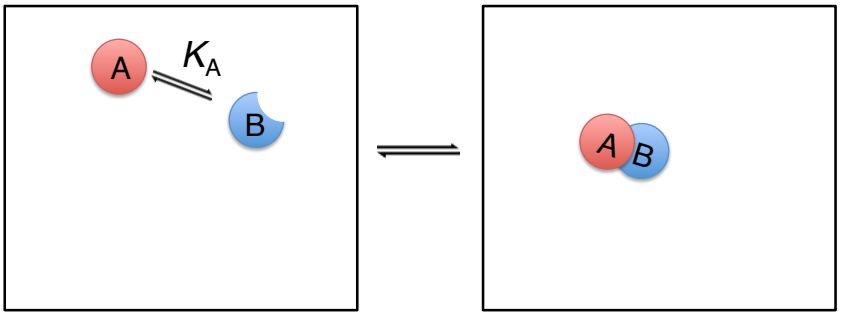


Fig. 2.20 Dimerisation reaction in a small box. We have two particles (a single A-type and a single B-type) in a box with volume V. We assume that, although the particles can bind, they do otherwise behave as an ideal gas. We wish to calculate the relation between probability of dimerisation and equilibrium association constant K_A. Simply calculating effective concentrations of $[A]$, $[B]$ and $[AB]$, and using standard chemical equilibrium equation $\frac{[AB]}{[A][B]} = K_A$ gives a wrong answer.

determines the probability that a dimer is formed.

$$\frac{1-p_u}{p_u} = \frac{q_b}{q_u} = \frac{K_A}{VN_A} \,, \tag{2.45}$$

with p_u denoting the unbound probability and the probability that two particles are bound is simply $p_b = 1 - p_u$.

On the other hand, if we naively make use of the expression for chemical equilibrium in a bulk mixture binary chemical equilibrium, we do not reproduce the correct result. We could simply rationalise that the effective (time averaged) concentration of unbound chemicals is

$$[A] = [B] = \frac{p_u}{VN_A} \,, \tag{2.46}$$

where p_u is the probability that A and B are unbound, V is the box volume and we have added the Avogadro's number N_A to make $[A]$ and $[B]$ a molar concentration. Similarly for the dimerised state $[AB] = \frac{1-p_u}{VN_A}$. Hence, in line with standard chemical dimerisation reaction, we could reason that

$$K_A = \frac{[AB]}{[A][B]} = \frac{1-p_u}{p_u^2}\, VN_A \,, \tag{2.47}$$

which is clearly different from the correct expression (2.45).

Treating the system as a bulk binary reaction is not valid for only two dimerising particles. The approach is valid in the thermodynamic limit where the chemical potential of a molecular species can be related to the logarithm of its concentration. What it boils down to is that Stirling's approximation is valid only for large number of particles $\log N! \approx N \log N - N$, it is clearly wrong when N equals 1 or 2. The same problem occurs when trying to calculate equilibrium constant from molecular dynamics simulations using small system sizes [105]

The above example might seem rather abstract. However, it exposes a potential pitfall of misusing "effective" concentrations. The same pitfall is encountered when calculating binding probabilities of multivalent ligands, because the reactions shown in Figures 2.20 and 2.19 are very similar. For example, one could naively argue that both the unbound ligand (A) and receptor (B) in Figure 2.19 are flexible and can explore some effective volume V and have some effective concentration within this volume. One then applies a "Local chemical equilibrium" (LCE) assumption [82, 106] which, in our simple system is given by Eqs. (2.46, 2.47). But this procedure does not generally give a correct result. It

becomes a good approximation only in the limit of weak binding[9] or a very large valency where the Stirling's approximation becomes applicable.

It should be clear that effective molarity is not really a concentration [10]. Rather, it is a quantity with the dimensions of concentration, defined by Eqn. 2.42. We can view the effective molarity as a measure for the probability that an unbound ligand and receptor would overlap in space (and hence come into position to bind). In an idealised system, neglecting the effects of the linker and orientational correlations in the unbound state, this probability is related to an effective concentration of, say, a ligand (B) as experienced by its complementary receptor (A) [65, 68, 70]. This is exactly the "cloud of ideal ligands" approximation we have used as a starting point for our theory of multivalent polymer adsorption (2.8).

In the case of our simplified system of 2 dimerizing particles (Figure 2.20) the effective concentration c_{eff} of type-A, as experienced by type-B, (or vice versa) is

$$c_{eff} = 1/(VN_A) \ , \tag{2.48}$$

where we recall that V is the box volume. We can think of particle A adsorbing to particle B and the ratio of probabilities of being bound to unbound becomes

$$p_b = K_A c_{eff} p_u \ , \tag{2.49}$$

which is consistent with our correct result (2.45). We could view $c_{eff}p_u$ as the concentration of unbound A.

Applying this concept to dimer adsorption (Figure 2.19) we would find that the empirically calculated effective molarity (2.42) is similar to the theoretical effective concentration $EM \sim c_{eff}$ (in our idealised system they are equal). Therefore, effective concentration, when applied properly, is a useful concept when attempting to theoretically predict equilibria of multivalent binding.

[9] for weak binding $p_u \approx 1$ and Eq. (2.47) becomes a very good approximation to (2.45)

[10] The effective molarity can be calculated via relative concentrations of singly and doubly bound states in solution, see Figure 2.19, but EM as such is a property only of the multivalent construct itself, or a pair of interacting constructs

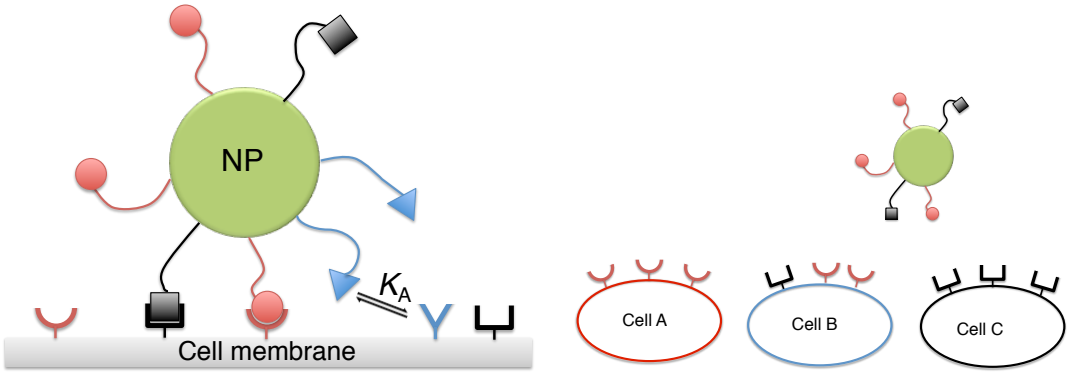
3

Multicomponent targeting

> *This fundamental law* $[\langle A\rangle = \sum_i A_i e^{-\beta E_i}]$ *is the summit of statistical mechanics, and the entire subject is either the slide-down from this summit, as the principle is applied to various cases, or the climb-up to where the fundamental law is derived.*
>
> – R.P. Feynman

In the preceding chapter we have shown that multivalent carriers (nano-particles or polymers) can distinguish target surfaces (cells) on the basis of their receptor concentration, rather than just on the basis of the presence of a suitable receptor. Here we extend our modelling to multiple ligand and receptor types. Our picture remains the same, see Figure 3.1. We essentially extend the number of receptors n_R to a receptor composition vector $\mathbf{n}$, equivalently the number of ligands k is extended into a vector specifying a ligand profile on a particle $\mathbf{k}$. Moreover, the binding strength f or the equilibrium constant K_{intra} is extended to an interaction matrix $\mathbf{K}$ specifying the binding strengths between all ligand/receptor types.

In this chapter we mainly focus on a complementary problem to super-selectivity: the total concentration of all receptors is kept fixed and the composition (the relative fractions of different receptor types is varied). We show that the design rules for such multicomponent targeting are surprisingly simple, and therefore hopefully useful. Specifically, we show that 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 composition 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.



(a) (b)

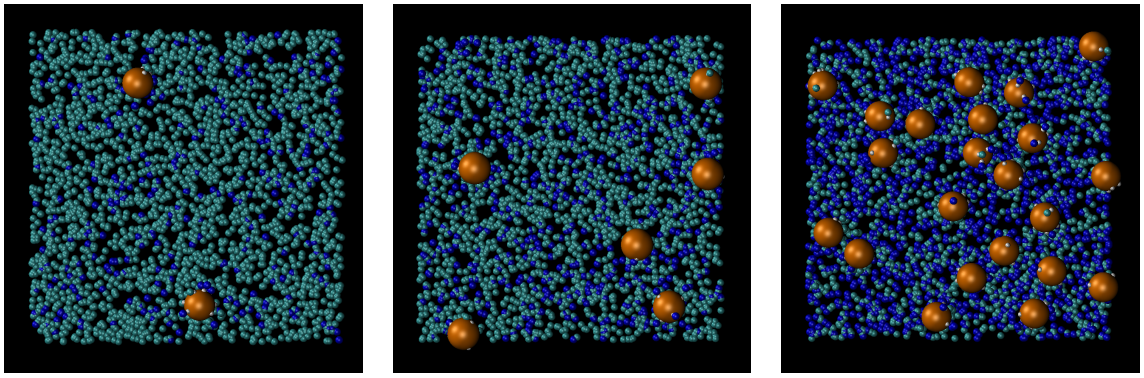
Fig. 3.1 Cartoon pictures representing the multicomponent nano-particle targeting a receptor decorated membrane. **b)** The challenge: How to target cell B in the presence of cells A and C ?

The simulation snapshots of multivalent nanoparticle targeting in Figure 3.2 give a pictorial illustration of the effect of optimising the ligand concentration profile to target a mixed receptor surface. We use analytical theory and simulations to rationalise under what conditions a selectivity towards a specific receptor composition can be obtained. We focus our attention to multicomponent targeting of surfaces, however, the model can also be used to illuminate the mechanism behind the sorting ability of cell imprinted polymers [51, 107].

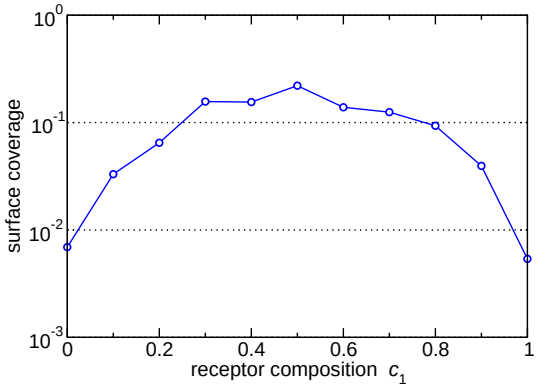
3.1 Model

Multivalent particle carries ligands of different types i and the particle is characterised by the ligand profile (vector k_i). Similarly, the membrane surface is characterised by its membrane receptor composition, i.e. number concentrations of different receptor types on the membrane c_j. Upon particle docking to a cell membrane, ligands can bind to the receptors (see Figure 3.1), the interaction matrix K_{ij} specifies the interaction strength between different ligand - receptor types. The binding is valence limited, only a single ligand can bind to a receptor and vice versa. The particle itself has no interactions with the membrane or the receptors bar the hard sphere repulsion. The model is essentially an extension of the work presented in the previous chapter, generalised to include different ligand/receptor types.

We aim to calculate the binding free energy of a multivalent guest nano-particle to a host membrane and use this knowledge to design a guest particle that can target a specific receptor concentration profile. The binding free energy can be calculated analytically



(a) $c_1 = 0.1$ (b) $c_1 = 0.2$ (c) $c_1 = 0.5$



(d)

Fig. 3.2 Simulated composition targeting. We perform Grand canonical Monte Carlo simulation results of ligand functionalised nanoparticles adsorbing to a receptor decorated flat surface. The surface has 2 different types of receptors embedded (coloured blue and cyan respectively). The total concentration of receptors is kept fixed, but the composition is varied across the three snapshots from **a)** 10% of type 1 receptor (and hence, 90% of type 2) to **b)** 20% type 1 and **c)** 50%, plot on **d**) shows the corresponding adsorption profile. The nanoparticle (NP) has 10 ligands with a profile of $p_1 = p_2 = 0.5$, and bond strength $\epsilon = -3.5 k_B T$. The NPs chemical potential is set to give a volume bulk NP volume fraction of 10^{-5}. Even though total concentration of receptors is constant, the particle preferentially targets the surface with a matching composition. The simulation model used is an extension of Ref. [10]; hard-core nanoparticles with attached soft-blob ligands.

if we assume that ligand binding is uncorrelated (different ligands do not interact) and receptors are approximated as point-like particles diffusing on a 2D membrane. We will first describe the analytical theory and later validate its results in comparison to simulations where ligand arms are modelled explicitly as self avoiding polymers.

3.1.1 Multicomponent theory

To obtain the binding free energy we need to calculate the partition function counting 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 (expect that no two ligands can bind to the same receptor).

The probability that a singe ligand i and a single receptor j form a bond depends on the equilibrium constant K_A^{ij} for the association reaction 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 value of ΔG_i^{cnf} depends on the distance between the receptor and the grafting point of the ligand. However, in the description that we use this is unimportant: as shown in the Supplementary Information, we can treat $\Delta \tilde{G}_i^{cnf}$ as if it is constant for all receptors that are within the range of the ligands, and infinite elsewhere.

For a given ligand grafted within interaction distance of the receptor decorated surface, the ratio between the probabilities of being in the bound (to receptor type j) and unbound states is

$$\frac{P_{\text{bound}}}{P_{\text{unbound}}} = c_j K_A^{ij} \, \frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{h_0} \equiv c_j K_{ij} \tag{3.1}$$

where c_j is the surface concentration of receptors, h_0 is the interaction range determined by the length of the polymeric linker. We have defined the effective association constant matrix K_{ij}, which includes the configurational contribution $\Delta \tilde{G}_i^{cnf}$. Note that the first index in K^{ij} always refers to a ligand i, and the second j always to a receptor. For instance, 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.

Let us first focus on a single ligand of type i. In the case of mobile receptors, the chemical potential of the various receptors is fixed and we can write the ligand partition function (i.e. semi-grand canonical partition function normalised by the unbound state)

as

$$q_{i,1\,\text{ligand}} = 1 + \sum_j e^{-\beta(\Delta G_{ij} + \Delta\tilde{G}_i^{cnf} - \mu_{R,j})} = 1 + \sum_j c_j K_{ij}. \tag{3.2}$$

as the ligand can be either unbound (weight 1) or bound to any receptor type j with bond energy ΔG_{ij}, $\Delta\tilde{G}_i^{cnf}$ is the configurational contribution discussed above and $\mu_{R,j} = k_B T \log(c_j/h_0\rho_0)$ is the chemical potential of receptor type j. For convenience we chose to operate with concentrations c_j and equilibrium constants K_{ij}; matrix K_{ij} has useful properties as we shall see below.

We assume that different ligands are uncorrelated, hence the partition function of the ligands of the guest particle at the membrane is simply a product over all ligands. As shown in the section on free-energy derivation, the fact that the receptors do not interact and are in contact with a reservoir (the remainder of the cell surface) simplifies the expression of the partition function for k_i ligands of type i: $q_i = \left(1 + \sum_j c_j K_{ij}\right)^{k_i}$.

If there are several different ligands, the total partition function (again at constant chemical potential of the receptors) is a product of the expression for the individual types

$$Q_b = \prod_i q_i = \prod_i \left(1 + \sum_j c_j K_{ij}\right)^{k_i}. \tag{3.3}$$

The binding free energy for a guest particle near the cell surface is denoted by $\Delta F_b \equiv -k_B T \ln(Q_b)$. We can also express the average binding free-energy per ligand as

$$f_b = \frac{\beta \Delta F_b}{k} = -\sum_i p_i \ln\left(1 + \sum_j c_j K_{ij}\right), \tag{3.4}$$

where $k = \sum_i k_i$ is the total number of ligands. f_b is simply a sum of the contributions of the individual ligand-receptor types due to chemical recognition. $p_i = \frac{k_i}{k}$ is the particle profile, i.e. specifying the fraction of ligand types on the guest particle.

When computing the binding free energy of a particle to a cell surface, we must also include the free-energy cost ΔF_0 of bringing the host particle sufficiently close to the surface to allow ligand-receptor binds to form. Then the total binding free energy is of the form:

$$\Delta F = \Delta F_b + \Delta F_0, \tag{3.5}$$

where the first term on the right hand side accounts for the free energy due to ligand-receptor bond formation. ΔF_0 will include any non-specific interactions ΔF_{ns} and will

depend on the details of the cell membrane and the guest particle

$$\Delta F_0 = \Delta F_{ns} - k_B T \log(a^2 h_0 \rho_0 N_A) \, . \tag{3.6}$$

$a^2 h_0$ can be recognised as an 'effective volume' that a guest particle must enter from solution in order to be able to start forming bonds, a is the particle size and h_0 the interaction distance discussed above.

In order to directly relate to experimental observations we write the 'avidity' equilibrium constant of a multivalent nanoparticle adsorbing to a receptor decorated surface

$$K_A^{av} = \frac{e^{-\beta \Delta F}}{\rho_0} = Q_b \frac{e^{-\beta \Delta F_0}}{\rho_0} \, . \tag{3.7}$$

To successfully apply our theory to a general multivalent system, two constants would, most probably, need to be determined experimentally. The first one is the configurational contribution to a single bond $\frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{h_0}$, the second constant is F_0 which essentially determines the free energy of a host-guest interaction without forming any bonds. Both constants can be straightforwardly determined by measuring the avidity K_A^{av} (or the apparent host/guest equilibrium constant), details are provided in the section on fitting experiments.

For what follows F_0 (3.6) is unimportant, because we assume that it is the same irrespective of the ligand composition of the guest particle. In that case F_0 drops out of the expressions that determine the selectivity. We expect the constant F_0 assumption to hold as long as the physical properties of different ligands are not too dissimilar, for example, the assumption would break if the disparity polymeric linker length (connecting ligands to the particle) is large.

The probability that a ligand of type i is un-bound is given by

$$p_i^u = \left(1 + \sum_j c_j K_{ij} \right)^{-1} \, , \tag{3.8}$$

With this definition, Eqn. (3.4) can be written as

$$f_b = \sum_i p_i \ln p_i^u \, . \tag{3.9}$$

We note that this expression could be interpreted as (minus) the cross-entropy between the two distributions p_i and p_i^u.

Next, we consider what is the optimal functionalisation for a guest particle to bind to a cell surface with receptor concentrations c_j.

3.1.2 Selectivity optimisation

A necessary condition for a particle to binds preferentially to a cell with a specific receptor density profile c_i, is that the free energy (3.9) is lower for the target host cell than for a cell with any other receptor composition. To make a meaningful comparison, we compare hosts with the same total receptor concentration $c_T \equiv \sum_i c_i$, because the binding can always be made stronger by increasing the total receptor concentration as we have shown in the context of super-selectivity in the preceding chapter.

We then have to minimise the binding free energy

$$f_b = -\sum_i p_i \ln\left(1 + \sum_j c_j K_{ij}\right) \tag{3.10}$$

subject to two constraints

$$\sum_j c_j = c_T, \qquad \sum_i p_i = 1\ . \tag{3.11}$$

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 kf_b is trivially proportional to the total number of ligands k (3.4), hence the overall binding strength can be controlled by varying k.

We first consider which host profile binds most strongly to a guest particle, that is, we need to determine the receptor profile $\mathbf{c}_0$ that minimizes the binding free energy. This problem is trivially solved using Lagrange multipliers.

$$\nabla(f_b + \lambda_c c_T) = 0 \tag{3.12}$$

which becomes

$$\frac{\partial f_b}{\partial c_l} = -\lambda_c \tag{3.13}$$

with λ_c the Lagrange multiplier. Inserting (3.10) into the above and differentiating we find

$$\sum_i \frac{p_i K_{il}}{1 + \sum_j c_j K_{ij}} = \lambda_c \tag{3.14}$$

which must hold for every index $_l$. Given arbitrary $\mathbf{c}_0$ we can chose any $\mathbf{p}$ and $\mathbf{K}$ that satisfy the above equation, and by definition the $\mathbf{c}_0$ will be a minimum.

At this point the problem is under constrained, there are infinitely many possible choices of p_i and K_{ij} that satisfy the above set of equations. We will additionally require that the minimum is as sharp as possible in order to bind specifically to a chosen receptor profile. Hence we will define the selectivity $\mathcal{S}$ as the second derivative (i.e. the Gaussian curvature or, generally, the determinant of a Hessian matrix) at $\mathbf{c}_0$

$$\mathcal{S}(\mathbf{c}_0) = \det\left(\frac{\mathbf{H}(f_b(\mathbf{c}))_{\mathbf{c}_0}}{|f_b(\mathbf{c}_0)|}\right), \tag{3.15}$$

here $\mathbf{H}(f_b(\mathbf{c}))_{\mathbf{c}_0}$ is a Hessian matrix (matrix of second derivatives) evaluated at $\mathbf{c}_0$. The free energy is also a convex function of $\mathbf{c}$, therefore $\mathbf{H}$ is positive and we used an absolute value of the free energy $|f_b(\mathbf{c}_0)|$ in the denominator to define selectivity as the relative curvature of the free energy. It is important to define the selectivity as the relative, rather than the absolute, curvature. The absolute value of the free energy can be controlled by the number of ligands k (3.4), therefore, by optimising for the relative curvature, we obtain the largest possible curvature at an arbitrarily chosen value of the free energy ΔF_b. As one would intuitively expect, a diagonal interaction matrix $K_{ij} = 0, i \neq j$ gives the maximum selectivity (see below). However, the optimal values of diagonal elements K_{ii} or ligand composition p_i are not a trivial function of the composition $\mathbf{c}_0$.

In general, finding the maximal selectivity by solving $\nabla\mathcal{S}(\mathbf{c}_0) = 0$ is non-trivial, see section on numerical optimisation below. However, we can greatly simplify the solution by sacrificing a small amount of selectivity. We achieve this by also requiring that the binding free energy is optimised with respect to the ligand profile $\mathbf{p}$. This is not a necessary condition as we wish to target host cells with guest particles, not target the guest particles with host cells. However, by imposing this additional condition, we greatly simplify the problem, whilst the optimal selectivity only decreases marginally as will be shown below. The condition that the binding free energy is a minimum with respect to the ligand composition profile on a guest particle $\mathbf{p}$ is, again imposed using a Lagrange multiplier λ_p:

$$\frac{\partial f_b}{\partial p_i} = -\ln\left(1 + \sum_j c_j K_{ij}\right) = -\lambda_p, \tag{3.16}$$

which must hold for every i. Using (3.8) we find a simple result

$$p_i^u = e^{-\lambda_p}. \tag{3.17}$$

We recall that p_i^u is the probability that a ligand of type i is not bound. The fact that λ_p in Eq. (3.17) is a constant implies that all ligands should have the same probability

to be bound $p_i^b = 1 - p_i^u$. Hence, any ligand profile $\mathbf{p}$ will yield the same $f_b = -\lambda_p$. In a sense, this result is trivial: it simply states that if all ligands are equally likely to bind, then a small change in the ligand profile will not change the overall host-guest binding free energy. This results is emphatically not 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). But luckily we are interested in the opposite problem, namely the targeting of cells by guest particles. That problem does have a unique, non-trivial solution.

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

$$\kappa_{ij} = \frac{K_{ij}}{K_{ii}} = e^{-(\Delta G_{ij} - \Delta G_{ii})/k_B T} \,. \tag{3.18}$$

We remember ΔG_{ij} as the Gibbs free energy of monomeric ligand-receptor dimerisation in solution. The configurational term ΔG_i^{cnf} in the definition of K_{ij} (Eq. (3.1)) 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.

Inserting (3.16) into (3.14) we find that the solution must satisfy

$$\sum_i p_i K_{ij} = \lambda_c e^{\lambda_p}, \ \forall j, \tag{3.19}$$

additionally (3.16) can be rearranged to

$$\sum_j c_j K_{ij} = e^{\lambda_p} - 1, \ \forall i. \tag{3.20}$$

There are $2d$ equations and $2d+1$ unknowns (taking into account that relative off diagonal elements are constants (3.18) and $\sum_i p_i = 1$) and the "dimensionality" $d = \mathrm{rank}(\mathbf{K})$ is determined the rank of matrix $\mathbf{K}$ or, equivalently, the number of distinct ligand types. Therefore, the above equations determine the ligand profile $\mathbf{p}$ and all interaction strengths K_{ii} up to a constant 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 profile of the particle $\mathbf{p}$ must match the composition of the

receptors **c**

$$p_i = \frac{c_i}{c_T} \tag{3.21}$$

and the Lagrange multipliers are related

$$\lambda_c = \frac{1 - e^{-\lambda_p}}{c_T} \tag{3.22}$$

with $c_T = \sum_j c_j$ the total receptor density. In general, for non-symmetric matrices $\mathbf{K}$, the optimal ligand profile $\mathbf{p}$ can be calculated by simultaneously solving the above set of equations (3.19,3.20). However, Eq. (3.21) will provide a very good approximation whenever cross-binding is weak and cognate interaction dominates: $K_{ii} \gg \sum_{j\neq i} K_{ij}$.

The only remaining step is to determine the optimal strength of the interaction captured by the constant λ_p. We optimise the selectivity (3.15) given by the determinant of the Hessian matrix. An element of the Hessian matrix is obtained by twice differentiating the free energy (3.10)

$$H_{kl} = \frac{\partial^2 f_b}{\partial c_k \partial c_l} = \sum_i \frac{p_i K_{ik} K_{il}}{(1 + \sum_j c_j K_{ij})^2} \ . \tag{3.23}$$

Using Eq. (3.18) we rewrite Eq. (3.20) in the form $K_{ii} = \frac{e^{\lambda_p} - 1}{\sum_j c_j \kappa_{ij}}$. Inserting this expression and Eqs. (3.20,3.18) into the above equation we find that every element of the Hessian matrix decouples into a term that depends only on λ_p and the remainder $\hat{H}_{kl}$

$$H_{kl} = \left(1 - e^{-\lambda_p}\right)^2 \sum_i \frac{p_i \kappa_{ik} \kappa_{il}}{(\sum_j c_j \kappa_{ij})^2} = (1 - e^{-\lambda_p})^2 \hat{H}_{kl} \ . \tag{3.24}$$

We have defined the remainder $\hat{H}_{kl} = \sum_i \frac{p_i \kappa_{ik} \kappa_{il}}{(\sum_j c_j \kappa_{ij})^2}$ which does not depend on λ_p, in fact the values of all $\hat{H}_{kl}$ are at this point already determined by the solution to Eqs. (3.19,3.20) above. Since λ_p is a constant for the whole matrix $\mathbf{H}$, the determinants are related by a factor: $\det(\mathbf{H}) = (1 - e^{-\lambda_p})^{2d} \det(\hat{\mathbf{H}})$.

We use the above relation and $f_b(\mathbf{c}_0) = -\lambda_p$ from Eq. (3.16) to express the selectivity (3.15) in terms of the Lagrange multiplier λ_p

$$\mathcal{S} = \frac{\det(\mathbf{H}(f_b))}{|f_b(\mathbf{c}_0)|} = \left[\frac{\left(1 - e^{-\lambda_p}\right)^2}{\lambda_p}\right]^d \det(\hat{\mathbf{H}}) \ , \tag{3.25}$$

we remember that d denotes the number of distinct ligands. 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

$$e^{\lambda_p} - 2\lambda_p + 1 = 0, \tag{3.26}$$

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

$$\lambda_p = -W_{-1}\left(\frac{-1}{2\sqrt{e}}\right) - \frac{1}{2} \approx 1.25643\cdots \tag{3.27}$$

approximated to the first 6 digits. This is our most important result. It states that the binding free energy of each ligand to the surface should be $f_b \approx -1.3k_BT$ irrespective of the details of the system.

Figure 3.3 shows the variation of the binding free energy with changing receptor composition obtained from the analytical model. To derive this result we have assumed that the interaction matrix is diagonal. Furthermore, Figure 3.4 shows that optimal λ_p is independent of the value of the cross-binding terms in $\mathbf{K}$. The selectivity, however, diminishes with added cross-binding. Figure 3.4**b)** also shows how the optimal ligand profile differs from our simple design rule (3.21) if the cross-binding is strong and the interaction matrix $\mathbf{K}$ is not symmetric. In the particular example on Figure 3.4**b)**, the type 1 ligand can bind to both receptor types, but the type 2 ligand can only bind to receptor type 2.

In practice we might wish to distinguish a surface with 20%-80% composition ($c_1 = 0.2$) form a surface with inverted 80%-20% composition. From Figure 3.3 we see that the difference in the free energy is about $\Delta f_b \sim 0.5k_BT$ per ligand. This does not appear much, however, particles can easily have 10-20 ligands (or more) if we require a total binding strength in the region of $10 - 25k_BT$. The difference in the total thus becomes substantial $\Delta\Delta F \sim 5 - 10k_BT$. Furthermore, Figure 3.5 demonstrates that increasing the number of targeted types increases the selectivity because the optimal region becomes a smaller fraction of the total space with increasing dimensionality.

In the treatment presented we have optimised the curvature of f_b which results in generally applicable design rules. However, knowing in advance exactly the composition $\mathbf{c}$ of possible surfaces, the selectivity can be optimised further. Let us reuse the example above, if we consider only two types of surfaces 20%-80% and 80%-20%, then optimal targeting results from optimising the free energy difference between these two surfaces. This will in principle yield a different result from optimising the curvature at a targeted surface. We will illustrate this in greater detail in the following chapter on molecular imprinting.

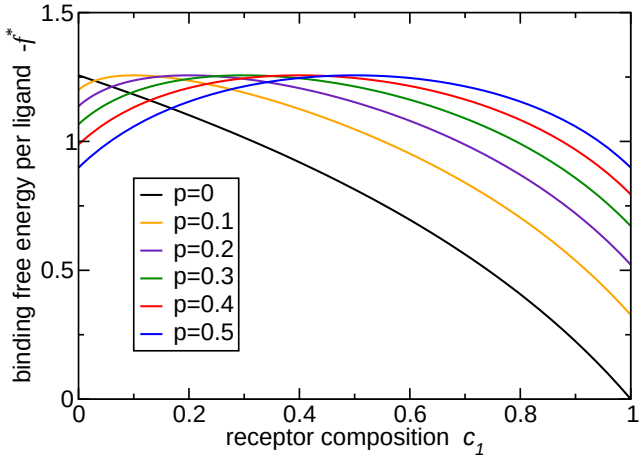
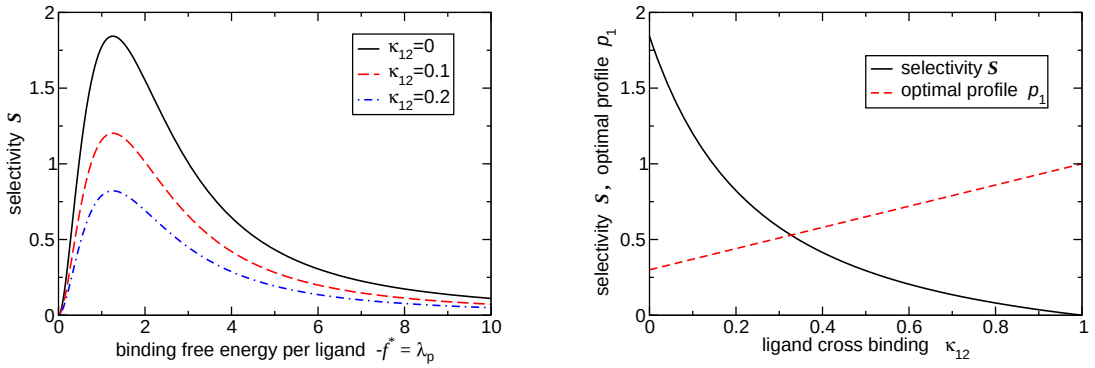
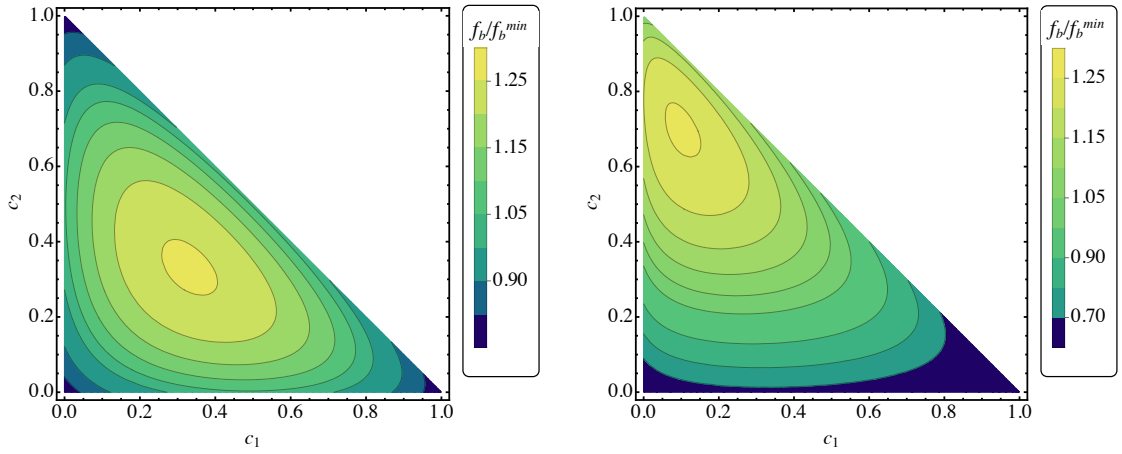


Fig. 3.3 Binding free energy per ligand as a function of the cell receptor composition. Different curves correspond to different particle profiles $p \equiv p_1 = 1 - p_2$ refers to the the fraction of type 1 ligands. We have used (3.10) to calculate the free energy with our design rules (3.19 - 3.22, 3.27). We have 2 ligand and 2 receptor types and only consider cognate interaction $K_{ij} = 0$, for $i \neq j$.



(a) (b)

Fig. 3.4 Selectivity of targeting with 2 ligand types and the effect of cross-binding. **a)** The selectivity $\mathcal{S}$ (3.15) 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 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 (3.19,3.20), as function of the cross binding term κ_{12}. Parameters: $c_1 = 0.3c_T$, $\kappa_{21} = 0$, $\kappa_{ii} = 1$ by definition (3.18).



(a) $p_1 = 0.33, p_2 = 0.33$ (b) $p_1 = 0.1, p_2 = 0.7$

Fig. 3.5 Binding free energy as a function of the composition of ligands on a cell - 3 different ligand-receptor species. Contour plots as a function of the composition of membrane receptors c_1 and c_2, with $c_3 = 1 - c_1 - c_2$ given implicitly, we have assumed $c_T = 1$. The plots were generated using (3.10) and the optimal selectivity rules (3.19 - 3.22, 3.27).

3.1.3 Cross entropy analogy

The expression for the binding free energy (3.9) is very similar to the cross entropy between distributions p_i and p_i^u, however, p_i^u is not a true probability distribution as it is not properly normalised. By defining a normalised distribution $\hat{p}_i^u = p_i^u/a$, with $a = \sum_i p_i^u$ the normalisation constant, the binding free energy becomes

$$\Delta f_b = \sum_i p_i \ln \hat{p}_i^u + \log \sum_i p_i^u = -H(p, \hat{p}^u) + E(p^u), \tag{3.28}$$

where $H(p_i, \hat{p}_i^u)$ is the cross entropy and $E = \log \sum_i p_i^u$ is a "cost function" analogous to energy, it measures the overall strength of a bond. If bonds are weak then $p_i^u \sim 1$ and $E > 0$. Conversely, with strong bonds $p_i^u \sim 0$ and E will be negative. The cross entropy can further on be written as the sum of Shannon entropy and Kullback-Leibler divergence

$$H(p_i, \hat{p}_i^u) = -\sum_i p_i \ln p_i + \sum_i p_i \ln \frac{p_i}{\hat{p}_i^u} = H(p) + D_{KL}(p||\hat{p}^u). \tag{3.29}$$

Such that the binding free energy per ligand becomes

$$\Delta f_b = E(p^u) - H(p) - D_{KL}(p||\hat{p}^u). \tag{3.30}$$

The first term $E(p^u)$ captures the overall bond strength ("energy"), the second $H(p)$ is the Shannon entropy of the ligands, it measures the diversity of the ligands on the particle. The Kullback-Leibler divergence $D_{KL}(p||\hat{p}^f)$ is measure of the difference between the two distributions p and p^u. Hence, to minimize the free energy: Energy favours strong individual bonds $p^u \ll 1$, 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.

3.2 Design rules

Our analytical calculations suggest simple design rules to make multi-valent guest particles that target a particular receptor composition.

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

- It is useful to avoid cross-binding (i.e. K_{ii} should be diagonal) and the value of $K_{ii} = \frac{e^{\lambda_p - 1}}{c_i}$ should be inversely proportional to the density c_i: The constant $\lambda_p = 1.256$, which states that ligand binding should be weak, with each ligand independently having the probability of being bound at most 70%.

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

- The overall binding free energy ΔF of the particle is proportional to the number of ligands per particle (valency k). Valency should be chosen such to give a desired absolute of value host-guest interaction, for Langmuir adsorption the optimal overall free energy will be close to the chemical potential of the guest particles in solution $\Delta F \sim \mu$. This would ensure that a targeted surface would be appreciably covered, but not yet saturated, with multivalent particles.

For clarity, we restate the main model assumptions used to arrive at these design rules: (i) ligands bind independently and are physically equivalent, i.e. all unbound ligands behave exactly the same, the only difference between ligand types is their affinity

to receptors. (ii) all ligands of a (surface bound) particle can reach surface attached receptors, (iii) the surface bound receptors are mobile and their number density is sufficiently large such that the fraction of bound receptors always remains low. Therefore, we treat the membrane as a reservoir for receptors with a constant chemical potential (or a constant density) of unbound receptors. (iv) receptors, ligands or particles have no interactions except for the steric (hard-core) repulsion and ligand-receptor affinity.

3.3 Further numerical optimisations

In our analytical treatment we have over-constrained the binding free energy (3.16); specifying that it must be a minimum with respect to the variation of the ligand profile on the particle. This might be desirable in practice, but is in principle not a necessary constraint. We introduced the constraint to simplify the analytical derivation, removing it increases the selectivity somewhat. However, the result cannot be cast in terms of simple design rules. Figures 3.6 and 3.7 shown the comparison between our reference solution determined by our design rules (using the constraint (3.16)) and the optimal unconstrained solution obtained numerically. Furthermore, Figure 3.6 shows that the numerically optimised solution results in a more skewed optimal ligand profile and binding strengths. This solution might be unpractical and harder to realise in a fabrication process.

3.4 Derivation of the simple analytical model

In the case of mobile receptors the expression for the bound partition function of a multivalent particle to the receptor decorated membrane is given approximately by (3.3):

$$Q_b = \prod_i \left(1 + \sum_j c_j K_{ij}\right)^{k_i} . \tag{3.31}$$

which we used to derive our design rules for composition targeting. We remember that c_j is the receptor type j concentration on the host, $k_i = kp_i$ is the number of ligands of type i on the guest particle and K_{ij} is the interaction matrix.

We will start from basic statistical mechanics and show what approximations are necessary to arrive at our simplified expression (3.31). 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

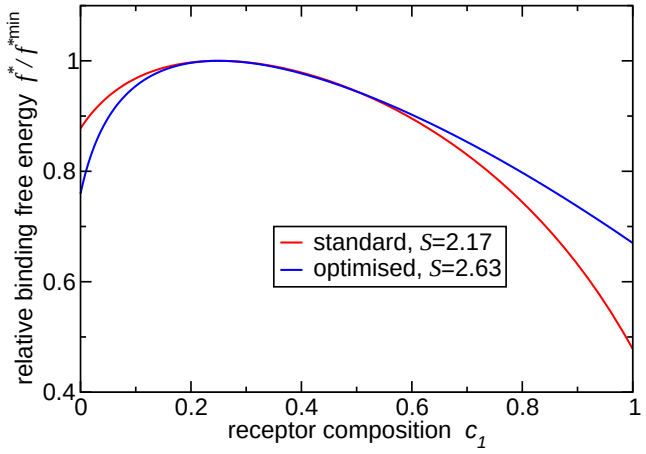
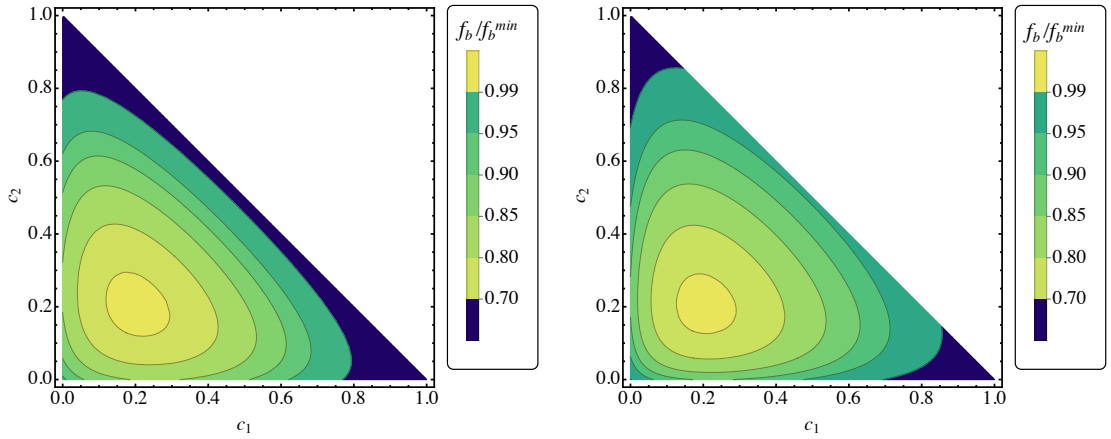


Fig. 3.6 Optimising selectivity, 2 ligand/receptor types. The plot shows the binding free energy f_b/f_b^{min} normalised by the minimum, the targeted composition is at $\mathbf{c}_0 = [0.25, 0.75]$. The standard result for the design of the particle is $p_1^{std} = 0.25$ $K_{11}^{std} = 10, K_{22}^{std} = 3.34$. The optimal numerically obtained solution is more skewed $c_1^{opt} = 0.07$ $K_{11}^{opt} = 26.3, K_{22}^{opt} = 0.34$.



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

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

Fig. 3.7 Optimising selectivity, 3 ligand/receptor types. These landscape plots show the relative binding free energy f_b/f_b^{min}. Selectivity $\mathcal{S}$ can be further increased (beyond our standard design rules) by optimising the compositions p_i of particles and binding strengths. Targeted composition is $\mathbf{c}_0 = [0.2, 0.2, 0.6]$.

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 where the ligands are rigidly attached to a particle, but the host membrane is flexible (either because the membrane is deformable or receptors themselves are flexible polymeric entities), will show qualitatively very similar results. It is important to somehow introduce flexibility in ligand-receptor bonding; multiple possible combinations of forming bonds are necessary for selectivity. However, the bonding flexibility can be achieved in many different ways: flexible ligands, flexible receptors or a deformable membrane with mobile receptors.

In our theoretical analysis the guest particle is treated as a hard sphere with attached polymeric ligand arms, shown schematically on Figure 3.1. The particle is grafted with a total of k polymer linkers, each linker caries a ligand at the tip. The host surface is a flat hard surface with mobile receptors. This is exactly our model used in simulations of multivalent particle adsorption. 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, bar the restriction that at most a single ligand can bind to a receptor. (ii) Ligands themselves are non-interacting and their positions are independent, for this to hold we must assume that ligand grafting points are mobile on a particle surface.

We fix the position of the particle 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 $\mathbf{r}^a_{i'}$. We use primes i' to denote a *specific* single ligand, while a standard index i denotes the ligand type, we also implicitly assume an existence of an indicator function that maps every ligand i' and receptor j' to its type i or j, respectively. 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 specific ligand i' with its grafting point at position $\mathbf{r}^a_{i'}$ reads

$$q_b(\mathbf{r}^a_{i'}) = \sum_j e^{-\beta \Delta G_{ij}} c_j \int_S e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}^a_{i'}; z_{i'}=0)} \delta(z_{i'}) d\mathbf{r}_{i'} \ , \tag{3.32}$$

where we remember $e^{-\beta \Delta G_{ij}}$ 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

surface section S, which must be large enough such that ΔG^{cnf} is negligible outside of this area. $\delta(z_{i'})$ is the Dirac delta function constraining the $z_{i'}$ coordinate of the ligand to lie on the surface $z_{i'} = 0$. We proceed by integrating the ligand anchor point over the colloid surface, the bound partition function of ligand i' with a particle at height h is

$$q_b^i(h) = \int_C q_b(\mathbf{r}_{i'}^a) d\mathbf{r}_{i'}^a = \sum_j e^{-\beta \Delta G_{ij}} c_j \int_C d\mathbf{r}_{i'}^a \int_S e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}_{i'}^a; z_{i'}=0)} \delta(z_{i'}) d\mathbf{r}_{i'} \ , \tag{3.33}$$

with $\int_C$ an integral over the 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.

The unbound partition function of the same ligand is

$$q_u^i = 4\pi r_{np}^2 q_u(\mathbf{r}_{i'}^a) = 4\pi r_{np}^2 \rho_0 \int_V e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}_{i'}^a)} d\mathbf{r}_{i'} \ , \tag{3.34}$$

where the $4\pi r_{np}^2$ comes from integrating over the particle surface, with r_{np} the (nano)particle radius, because for noninteracting unbound ligands the volume integral $\int_V$ of the ligand position $\mathbf{r}_{i'}$ does not depend on $\mathbf{r}_{i'}^a$. Also q_u^i 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 to unbound state

$$\frac{p_{\text{bound}}^i}{p_{\text{unbound}}^i} = \frac{q_b^i(h)}{q_u^i} = \tilde{K}_i(h) \sum_j e^{-\beta \Delta G_{ij}} c_j \ . \tag{3.35}$$

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

$$\tilde{K}_i(h) = \frac{\int_C d\mathbf{r}_{i'}^a \int_S e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}_{i'}^a; z_{i'}=0)} \delta(z_{i'}) d\mathbf{r}_{i'}}{4\pi r_{np}^2 \rho_0 \int_V e^{-\beta \Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}_{i'}^a)} d\mathbf{r}_{i'}} \ , \tag{3.36}$$

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 $\Delta G^{cnf}(\mathbf{r}_{i'}, \mathbf{r}_{i'}^a)$. 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 $q_b^i(\infty) = 0$.

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

$$Q(h) = \prod_i \left(q_u^i + q_b^i(h)\right)^{k_i} \ , \tag{3.37}$$

because each ligand can independently be in either a bound or unbound state and we remember k_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 function

$$e^{-\beta\Delta F_b(h)} = Q_b(h) = \frac{Q(h)}{Q(\infty)} = \prod_i \left(1 + \frac{q_b^i(h)}{q_u^i}\right)_i^k = \prod_i \left(1 + \tilde{K}_i(h) \sum_j e^{-\beta\Delta G_{ij}} c_j\right)^{k_i} , \tag{3.38}$$

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 a solution. Eq. (3.35) shows that this free energy difference is related only to the probabilities of individual ligands being bound or unbound.

In practice, the relevant measure is the free energy of a bound particle at a surface, hence, the above equation needs to be integrated over h. 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

$$e^{-\beta\Delta F_b} = \frac{1}{h_0}\int_0^\infty Q_b(h) - 1 \, dh = \int_0^\infty \left[\prod_i \left(1 + \tilde{K}_i(h) \sum_j e^{-\beta\Delta G_{ij}} c_j\right)^{k_i} - 1\right] dh \ , \tag{3.39}$$

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

$$e^{-\beta\Delta F_b} = \frac{1}{h_0}\int_0^{h_{\text{cutoff}}} Q_b(h) \, dh = \int_0^{h_{\text{cutoff}}} \prod_i \left(1 + \tilde{K}_i(h) \sum_j e^{-\beta\Delta G_{ij}} c_j\right)^{k_i} dh \ . \tag{3.40}$$

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 ΔF_b. We chose the second expression (3.40) as our definition of choice because it conveniently enables writing the free energy as a sum over individual ligand contributions. In the preceding chapter on super-selectivity an alternative definition (3.39) was used. The constant h_0 reflects the interaction range and must be included for dimensionality consistency and our choice of definition of

F_b. The same constant is also present in the zero-bond free energy reference of an adsorbed particle $\Delta F_0 = \Delta F_{ns} - k_B T \log(a^2 h_0 \rho_0 N_A)$ with F_{ns} including any non-specific interactions between a particle and the host surface and a^2 is the excluded area; i.e. the surface area occupied by the particle. The total free energy of a guest particle adsorbed to a host surface

$$\Delta F = \Delta F_b + \Delta F_0 = \Delta F_b + \Delta F_{ns} - k_B T \log(a^2 h_0 \rho_0 N_A) \tag{3.41}$$

is, therefore, independent of the somewhat arbitrary h_0.

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 (3.31) we must also approximate $\tilde{K}_i(h)$ with a step function

$$\tilde{K}_i(h) = \begin{cases} \frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{\rho_0 h_0} & ; h \leq h_0 \\ 0 & ; h > h_0 \end{cases} \tag{3.42}$$

where $e^{-\beta \Delta \tilde{G}_i^{cnf}}$ is the mean configurational cost of binding a ligand to a surface, formally we define it as

$$\frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{\rho_0} = \int_0^\infty \tilde{K}_i(h)\, dh \; . \tag{3.43}$$

The value of h_0 should reflect the interaction range such that $\tilde{K}_i(h)$ is well approximated by a constant value within h_0, but is negligible otherwise. A choice of

$$h_0 \approx \sqrt{\langle h^2 \rangle} \tag{3.44}$$

would yield a good estimate. The mean square height $\sqrt{\langle h^2 \rangle}$ is calculated from the distribution $\tilde{K}_i(h)$ using a weighted average for different ligand types i. If the same polymeric linker is used for all ligands, $\tilde{K}_i(h) = \tilde{K}(h)$ will be the same for all ligand types i.

Inserting the above approximation (3.42) into Eq. (3.40) the integration becomes trivial and we obtain

$$e^{-\beta \Delta F_b} = Q_b = \prod_i \left(1 + \frac{e^{-\beta \Delta \tilde{G}_i^{cnf}}}{\rho_0 h_0} \sum_j e^{-\beta \Delta G_{ij}} c_j \right)^{k_i} + \frac{h_{\text{cutoff}} - h_0}{h_0} \; . \tag{3.45}$$

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 $\frac{h_{\text{cutoff}}-h_0}{h_0}$ cannot be neglected. However, this term will be important only at high concentrations of particles in solution in which case it will introduce a small offset in the measured adsorbed density.

Finally, we define the equilibrium constant

$$K_{ij} = \tilde{K}_i e^{-\beta\Delta G_{ij}} = \frac{e^{-\beta\Delta\tilde{G}_i^{cnf}}}{\rho_0 h_0} e^{-\beta\Delta G_{ij}} , \tag{3.46}$$

which includes both the configurational (3.43) and the association ΔG_{ij} terms. The bound partition function can now be written in the simple form (3.31)

$$e^{-\beta\Delta F_b} = Q_b = \prod_i \left(1 + \sum_j c_j K_{ij}\right)^{k_i} . \tag{3.47}$$

3.5 Guide to fitting experiments

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

$$K_A^{av}\rho_0 = e^{-\beta\Delta F} = e^{-\beta\Delta F_0} \times \prod_i \left(1 + \sum_j c_j K_{ij}\right)^{k_i} . \tag{3.48}$$

In practice, however, both the zero-bond free energy ΔF_0 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 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. (3.46) we rewrite the expression for the binding avidity of a multivalent guest

$$K_A^{av} = A_{zero} a^3 N_A \prod_i \left(1 + \frac{A_{cnf}}{a} \sum_j c_j K_A^{ij}\right)^{k_i} . \tag{3.49}$$

We remind ourselves that a is the lateral size of the multivalent guest, c_j the surface molar concentration of receptor 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 k_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. Furthermore, the values of the correction factors need not be close to unity. Experiments on hyaluronic acid based multivalent polymers presented in the previous chapter (Refs. [15, 16]) 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 $\frac{A_{cnf}}{a}$ is related to the widely used "effective molarity" approach in rationalising multivalent interactions (see section on effective molarity in the previous chapter or Refs. [66–68, 70]). The number of receptors within interaction area a^2 that a multivalent guest "sees" is $n_j = a^2 N_A c_j$, therefore, the second term in (3.49) can be rewritten by defining the effective molarity $EM = \frac{A_{cnf}}{a^3 N_A}$ which measures the configurational contribution to binding between a ligand and a *particular* receptor within interaction distance a. Equivalently, some authors fit the effective volume $V_{eff} = 1/EM$ and the 0-valency dissociation constant $K_{D0} = \frac{1}{A_{zero} a^3 N_A}$ [18].

3.6 Poisson fluctuations undermine specificity

We expect that, in practice, any nanoparticle fabrication technique will introduce some heterogeneity or polydispersity 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 (3.47) over the Poisson distribution of the ligands on the particles

$$Q_b(\tilde{\mathbf{k}}, \mathbf{c}, \mathbf{K}) = \sum_{\mathbf{k}=0}^{\infty} \left[\prod_i \left(\frac{e^{-\tilde{k}_i} \tilde{k}_i^{k_i}}{k_i!} \right) Q_b(\mathbf{k}, \mathbf{c}, \mathbf{K}) \right] , \tag{3.50}$$

where $Q_b(\mathbf{k}, \mathbf{c}, \mathbf{K})$ is the bound partition function from (3.47) and we explicitly write it as a function of the ligand profile $\mathbf{k}$, the receptor composition $\mathbf{c}$ and the interaction matrix $\mathbf{K}$. We assume that every ligand type valency k_i is Poisson distributed with mean $\tilde{k}_i$, $\tilde{\mathbf{k}}$ denotes the mean ligand profile vector and $\sum_{\mathbf{k}=0}^{\infty}[\cdot] \equiv \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \ldots [\cdot]$ is a nested sum over all k_i.

Inserting Eq. (3.47) into the expression above and swapping the product and summation order we get

$$Q_b(\tilde{\mathbf{k}}, \mathbf{c}, \mathbf{K}) = \prod_i \left(e^{-\tilde{k}_i} \sum_{k_i=0}^{\infty} \left[\frac{\left[\tilde{k}_i \left(1 + \sum_j c_j K_{ij} \right) \right]^{k_i}}{k_i!} \right] \right) , \tag{3.51}$$

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

$$Q_b(\tilde{\mathbf{k}}, \mathbf{c}, \mathbf{K}) = \prod_{i,j} e^{-\tilde{k}_i K_{ij} c_j} . \tag{3.52}$$

We call this form the double exponential form because inserting Eq. (3.46) 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

$$\Delta F_b = -k_B T \log Q_b(\tilde{\mathbf{k}}, \mathbf{c}, \mathbf{K}) = k_B T \sum_{i,j} \tilde{k}_i c_j K_{ij} = k_B T \, \tilde{\mathbf{k}}^T \mathbf{K} \mathbf{c} , \tag{3.53}$$

in the last form on the right we have cast the expression in terms of matrix algebra with $\tilde{\mathbf{k}}^T$ being the transpose of vector $\tilde{\mathbf{k}}$.

We stress that in this case the binding free energy is evidently 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 $\mathbf{k}$ 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. Therefore, it appears that DNA origami constructs [108], where the geometry of the nano-construct can be almost exactly controlled, would be best suited for receptor composition targeting.

3.7 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 (3.47). 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. The canonical partition function (fixed number of both ligands and receptors) has previosly been derived by Angioletti-Uberti et. al. [87]. We re-derive the canonical result and use Poisson averaging to obtain our simple expression (3.47).

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 k 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

$$q_b(\mathbf{r}^a_{i'};\mathbf{r}_{j'}) = e^{-\beta\Delta G_{ij}} e^{-\beta\Delta G^{cnf}(\mathbf{r}_{j'},\mathbf{r}^a_{i'})}. \tag{3.54}$$

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 (3.33), 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

$$q^i_b(\mathbf{r}_{np};\mathbf{r}_{j'}) = \int_C q_b(\mathbf{r}^a_{i'};\mathbf{r}_{j'})d\mathbf{r}^a_{i'} = e^{-\beta\Delta G_{ij}} \int_C e^{-\beta\Delta G^{cnf}(\mathbf{r}_{j'},\mathbf{r}^a_{i'})}d\mathbf{r}^a_{i'} \ , \tag{3.55}$$

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. (3.34) 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

$$\frac{p^{ij'}_{\text{bound}}}{p^i_{\text{unbound}}} = \frac{q^i_b(\mathbf{r}_{np};\mathbf{r}_{j'})}{N_A q^i_u} = e^{-\beta\left(\Delta G_{ij}+\Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})\right)} \ . \tag{3.56}$$

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 (3.34) 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

$$\frac{Q(\mathbf{r}_{np},\mathbf{r}_R)}{Q(h=\infty)} = Q_b(\mathbf{r}_{np},\mathbf{r}_R) = \sum_{\lambda=0}^{k}\sum_{s(\lambda)} \exp\left\{\sum_{\{i'j'\}(s)} -\beta\left(\Delta G_{ij} + \Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})\right)\right\} \ , \tag{3.57}$$

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

$$\begin{aligned} Q_b(h,\mathbf{r}_R) &= \frac{1}{S}\int_S Q_b(\mathbf{r}_{np},\mathbf{r}_R)\delta(z_{np}-h)d\mathbf{r}_{np} \\ &= \frac{1}{S}\int_S \sum_{\lambda=0}^{k}\sum_{s(\lambda)} \exp\left\{\sum_{\{i'j'\}(s)} -\beta\left(\Delta G_{ij}+\Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})\right)\right\}\delta(z_{np}-h)d\mathbf{r}_{np}\ , \end{aligned} \tag{3.58}$$

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 (3.38) 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

$$\frac{1}{S^n}\int_{S^n}(d\mathbf{r}_R)^n Q_b(\mathbf{r}_{np},\mathbf{r}_R) \approx \frac{1}{S}\int_S d\mathbf{r}_{np} Q_b(\mathbf{r}_{np},\mathbf{r}_R)\ , \tag{3.59}$$

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 (3.59)) 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 (3.59) for an infinitely large surface and prove that it equals to Eq. (3.38), 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

$$\begin{aligned} Q_b(h) &= \frac{1}{S^n}\int_{S^n}(d\mathbf{r}_R)^n Q_b(\mathbf{r}_{np},\mathbf{r}_R) \\ &= \sum_{\lambda=0}^{m}\sum_{s(\lambda)}\prod_{\{i'j'\}(s)} e^{-\beta\Delta G_{ij}}\frac{1}{S}\int_S e^{-\beta\Delta\tilde{G}^{cnf}(\mathbf{r}_{j'},\mathbf{r}_{np})}d\mathbf{r}_{j'}\ , \end{aligned} \tag{3.60}$$

we have used Eq. (3.57) 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 (3.57) partition function. Using Eq. (3.56) we recognise the integral on the right hand side is the same as Eq. (3.36) defining $\tilde{K}_i(h)$. Hence, we can write the partition function as

$$Q_b(h) = \sum_{\lambda=0}^{k} \sum_{s(\lambda)} \prod_{\{ij\}(s)} \frac{\tilde{K}_i(h)}{N_A S} e^{-\beta \Delta G_{ij}} \ , \tag{3.61}$$

which returns a similar expression to (3.57), 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. (3.61) 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 $\boldsymbol{\lambda}$

$$Q_b(h) = \sum_{\boldsymbol{\lambda}=0}^{\mathbf{k}} \Omega(\boldsymbol{\lambda}) e^{-\sum_{ij} \beta \epsilon_{ij} \lambda_{ij}} \ . \tag{3.62}$$

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

$$e^{-\beta \epsilon_{ij}} \equiv \frac{\tilde{K}_i(h)}{N_A S} e^{-\beta \Delta G_{ij}} \tag{3.63}$$

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

$$\sum_{\boldsymbol{\lambda}=0}^{\mathbf{k}} [\cdot] = \sum_{\lambda_{11}=0}^{k_1} \sum_{\lambda_{12}=0}^{k_2} \cdots \sum_{\lambda_{21}=0}^{k_1} \sum_{\lambda_{22}=0}^{k_2} \ldots [\cdot] \tag{3.64}$$

to account for all possible states of distinct bonding arrangements. We note that the maximum term in each sum is set to k_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 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

$$Q_b(n,k,\epsilon) = \sum_{\lambda=0}^{k} \Omega(\lambda) e^{-\beta\epsilon\lambda} \tag{3.65}$$

we have dropped the functional dependance on h for clarity and write it as a function of the number of receptors n and ligands k, and the bond strength ϵ. The dependance on guest height h is implicitly accounted for through ϵ, which is itself a function of h (3.63). The density of states is given in terms of binomial coefficients

$$\Omega(\lambda) = \binom{n}{\lambda}\binom{k}{\lambda}\lambda! \tag{3.66}$$

because we need to choose λ bonds out of n receptors, λ bonds out of k 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 [1] distributed within every chosen surface area S. Therefore, we now Poisson average the partition function (3.65) over the number of receptors

$$\begin{aligned} Q_b(\tilde{n},k,\epsilon) &= \sum_{n=0}^{\infty} \frac{e^{-\tilde{n}}\tilde{n}^n}{n!} Q(n,k,\epsilon) \\ &= e^{-\tilde{n}} \sum_{n=0}^{\infty} \frac{\tilde{n}^n}{n!} \sum_{\lambda=0}^{k} \binom{k}{\lambda} e^{-\beta\lambda\epsilon} \frac{n!}{(n-\lambda)!} \\ &= e^{-\tilde{n}} \sum_{\lambda=0}^{k} \binom{k}{\lambda} e^{-\beta\lambda\epsilon} \sum_{n=0}^{\infty} \frac{\tilde{n}^n}{(n-\lambda)!} , \end{aligned} \tag{3.67}$$

where in the second line we have inserted (3.66) 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 $x = n - \lambda$

$$\sum_{n=0}^{\infty} \frac{\tilde{n}^n}{(n-\lambda)!} = \tilde{n}^\lambda \sum_{x=-\lambda}^{\infty} \frac{\bar{n}^x}{x!} = \tilde{n}^\lambda e^{\tilde{n}} \tag{3.68}$$

[1] strictly, the distribution will be binomial

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 (3.67) we get

$$Q_b(\tilde{n}, k, \epsilon) = \sum_{\lambda=0}^{k} \binom{k}{\lambda} \tilde{n}^{\lambda} e^{-\beta\lambda\epsilon} \ , \tag{3.69}$$

a binomial expansion for the function

$$Q_b(\tilde{n}, k, \epsilon) = (1 + \tilde{n}e^{-\beta\epsilon})^k \ . \tag{3.70}$$

Using our definition of ϵ (3.63) we find

$$Q_b(h) = \left(1 + c\tilde{K}(h)\, e^{-\beta\Delta G}\right)^k \ , \tag{3.71}$$

which is precisely the expression we have obtained in the mobile receptor case above, Eq (3.38), 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

$$Q_b(\tilde{\mathbf{n}}, \mathbf{k}, \boldsymbol{\epsilon}) = \prod_i \left(1 + \sum_j \tilde{n}_j e^{-\beta\epsilon_{ij}}\right)^{k_i} = \langle Q_b(\mathbf{n}, \mathbf{k}, \boldsymbol{\epsilon})\rangle_{\mathbf{n}} \tag{3.72}$$

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 (3.62)

$$Q_b(\mathbf{n}, \mathbf{k}, \boldsymbol{\epsilon}) = \sum_{\boldsymbol{\lambda}}^{\mathbf{k}} \Omega(\boldsymbol{\lambda}) e^{-\sum_{ij}\beta\epsilon_{ij}\lambda_{ij}}, \tag{3.73}$$

where the sum represents nested sums

$$\sum_{\boldsymbol{\lambda}=0}^{\mathbf{k}}[\cdot] = \sum_{\lambda_{11}=0}^{k_1}\sum_{\lambda_{12}=0}^{k_2}\cdots\sum_{\lambda_{21}=0}^{k_1}\sum_{\lambda_{22}=0}^{k_2}\ldots[\cdot] \tag{3.74}$$

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 k_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

$$\begin{aligned}\Omega(\boldsymbol{\lambda}) &= \prod_j\left(\frac{n_j!}{\prod_i(\lambda_{ij}!)\,(n_j-\sum_i\lambda_{ij})!}\right)\prod_i\left(\frac{k_i!}{\prod_j(\lambda_{ij}!)\,(k_i-\sum_j\lambda_{ij})!}\right)\prod_{ij}\lambda_{ij}! \\ &= \prod_j\left(\frac{n_j!}{(n_j-\sum_i\lambda_{ij})!}\right)\prod_i\left(\frac{k_i!}{\prod_j(\lambda_{ij}!)\,(k_i-\sum_j\lambda_{ij})!}\right)\,. \end{aligned} \tag{3.75}$$

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. [87] in the context of interacting DNA coated colloids.

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

$$Q(\tilde{\mathbf{n}},\mathbf{k},\boldsymbol{\epsilon}) = \langle Q(\mathbf{n},\mathbf{k},\boldsymbol{\epsilon})\rangle_{\mathbf{n}} = \sum_{\mathbf{n}=0}^{\infty}\left[\prod_j\left(\frac{e^{-\tilde{n}_j}\tilde{n}_j^{n_j}}{n_j!}\right)Q(\mathbf{n},\mathbf{k},\boldsymbol{\epsilon})\right] \tag{3.76}$$

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 (3.73) and (3.75) into the above equation we obtain a long expression

$$Q(\tilde{\mathbf{n}},\mathbf{k},\boldsymbol{\epsilon}) = \sum_{\mathbf{n}=0}^{\infty}\left[\prod_j\left(\frac{e^{-\tilde{n}_j}\tilde{n}_j^{n_j}}{n_j!}\right)\sum_{\boldsymbol{\lambda}=0}^{\mathbf{k}}\left[\prod_j\left(\frac{n_j!}{(n_j-\sum_i\lambda_{ij})!}\right)\prod_i\left(\frac{k_i!}{\prod_j(\lambda_{ij}!)\,(k_i-\sum_j\lambda_{ij})!}\right)e^{-\sum_{ij}\beta\epsilon_{ij}\lambda_{ij}}\right]\right], \tag{3.77}$$

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

$$Q(\tilde{\mathbf{n}},\mathbf{k},\boldsymbol{\epsilon}) = \sum_{\boldsymbol{\lambda}=0}^{\mathbf{k}}\left[e^{-\sum_{ij}\beta\epsilon_{ij}\lambda_{ij}}\prod_i\left(\frac{k_i!}{\prod_j(\lambda_{ij}!)\,(k_i-\sum_j\lambda_{ij})!}\right)\prod_j\left(\sum_{n_j=0}^{\infty}\left[\frac{e^{-\tilde{n}_j}\tilde{n}_j^{n_j}}{(n_j-\sum_i\lambda_{ij})!}\right]\right)\right]. \tag{3.78}$$

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 (3.68) the sum simply reduces to

$$\sum_{n_j=0}^{\infty}\left[\frac{e^{\tilde{n}_j}\tilde{n}_j^{n_j}}{(n_j-\sum_i\lambda_{ij})!}\right] = e^{-\tilde{n}_j}\tilde{n}_j^{\sum_i\lambda_{ij}}\sum_{x_j=-\sum_i\lambda_{ij}}^{\infty}\left[\frac{\tilde{n}_j^{x_j}}{x_j!}\right] = \prod_i\tilde{n}_j^{\lambda_{ij}} \tag{3.79}$$

as the terms in the sum with negative x_j vanish. Inserting this result into (3.78) and rearranging the terms we find

$$\begin{aligned} Q(\tilde{\mathbf{n}}, \mathbf{k}, \boldsymbol{\epsilon}) &= \sum_{\boldsymbol{\lambda}=0}^{\mathbf{k}} \left[\prod_i \left(\frac{k_i!}{\prod_j (\lambda_{ij}!)\ (k_i - \sum_j \lambda_{ij})!} \right) \prod_{i,j} \left(\left(\tilde{n}_j e^{-\beta \epsilon_{ij}} \right)^{\lambda_{ij}} \right) \right] \\ &= \prod_i \left(\sum_{\boldsymbol{\lambda}_i=0}^{k_i} \left[\frac{k_i!}{\prod_j (\lambda_{ij}!)\ (k_i - \sum_j \lambda_{ij})!} \prod_j \left(\left(\tilde{n}_j e^{-\beta \epsilon_{ij}} \right)^{\lambda_{ij}} \right) \right] \right) \\ &= \prod_i \left(1 + \sum_j n_j e^{-\beta \epsilon_{ij}} \right)^{k_i} \end{aligned} \tag{3.80}$$

which upon swapping the summation and product can be recognised as the multinomial expansion. Finally, inserting our definition of ϵ (3.63), we obtain precisely the expression for the partition function that we have previously found (3.38) directly in the grand canonical ensemble for mobile receptors

$$Q(h; \mathbf{c}, \mathbf{k}, \tilde{\mathbf{K}}) = \prod_i \left(1 + \sum_j c_j \tilde{K}_i(h) e^{-\beta \Delta G_{ij}} \right)^{k_i} \tag{3.81}$$

We must now only follow the procedure laid out after (3.38) to show that multicomponent binding to immobile (but Poisson distributed) receptors is governed by the same simple expression (3.47).

3.8 Particle endocytosis

The above analytical model is highly idealised. Therefore we also perform explicit simulation of a membrane and nanoparticle endocytosis. We perform Monte Carlo simulations with a coarse-grained membrane model [109] and a patchy hard-sphere model [110] for the nanoparticle. The nanoparticle has 2 different types of circular patches modelling coverage with 2 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 size of individual beads σ, ϵ denotes the well depth.

The simulations are performed using standard Monte Carlo translational moves in a 2D NpT ensemble and no applied external pressure [2]. 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 such that none of the particles ever interacted with the hard ceiling or the floor. The simulations started with the particle right above the membrane and were run for $6 \cdot 10^6$ cycles where in each cycle on average one translational/rotational more per every bead is attempted.

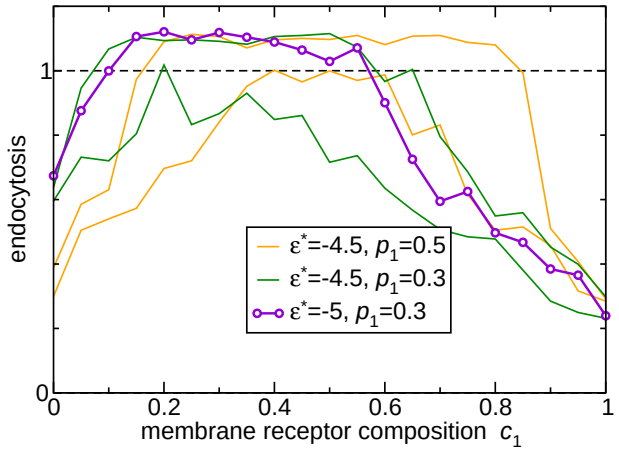
We must stress that we only consider a simple coarse-grained model of passive particle endocytosis. For example, we did not consider any active (irreversible) processes that are also present *in vitro* cell endocytosis [85]. Moreover, we do not take into account any details of receptor activation, receptor clustering and signalling pathways. These additional details should be taken into account in a practical application of drug delivery to specific living cells.

3.9 Summary

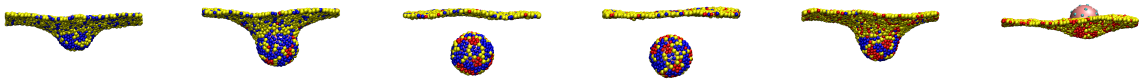
Both the simulation results of particle endocytosis (Figure 3.8) and of the adsorption (Figure 3.2) support the predictions of our analytical model. Simulations clearly show that our design rules, even though derived from a simple model, are nevertheless directly applicable to more complicated and realistic systems where ligand interactions, correlations and membrane elasticity cannot be neglected. We observe that with the chosen parameters the particle is only endocytosed if the receptor composition closely matches the particle ligand profile. Note that the interaction strength in endocytosis simulations (Figure 3.8) was $-\epsilon \approx -5k_BT$, however, we can not directly compare it with the free energy per bond f_b as the latter also includes the cost of recruiting receptor beads and bending the membrane. Both of which will contribute an order of k_BT and we expect the actual f_b in simulations to be close to our $-f_b \sim \lambda_p = 1.3k_BT$ design rule.

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 formulated optimal design rules for multivalent particles that allows them to distinguish target cells based on their receptor profile. We

[2] Strictly, our membrane system is only metastable at no applied external pressure, the thermodynamically stable configuration is an infinitely large box. However, on a simulation timescale, a flat membrane is stable.



(a)



(b) $c_1^* = 0.0$ (c) $c_1^* = 0.1$ (d) $c_1^* = 0.3$ (e) $c_1^* = 0.5$ (f) $c_1^* = 0.7$ (g) $c_1^* = 1.0$

Fig. 3.8 Simulation results of nanoparticle endocytosis. Nanoparticle is covered with 40 randomly distributed patches. There are 2 patch types and 2 cognate receptor types. The total concentration of receptors is kept fixed at $c_T = 0.4$, but the composition is varied as $c_1^* = c_1/c_T$. The curves in **a)** show the coverage of the particle with membrane beads. When the coverage exceeds 1 the particle is fully covered and has, therefore, been endocytosed. Snapshots from **b)** to **g)** correspond to the bold curve in **a)** with circular symbols. The inert beads are coloured yellow, the 1^{st} receptor type is red and the 2^{nd} receptor type is blue. Total coverage of the nanoparticle with circular patches is 0.5, with a 'ligand' profile $p_1 = 1 - p_2 = 0.3$. The particle radius is $R = 4\sigma$ and the interaction strength of patch i s determined as $\epsilon_i = \epsilon^* - \ln(p_i)$, i.e. the more numerous the patches, the weaker the interaction as per our design rule from section 3.2.

found that: 1) it is not a good idea to aim for very strong binding between the guest (delivery vehicle) and 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. Our coarse-grained simulations of guest particle adsorption and endocytosis 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.

Prevalent paradigm in specific interactions in biology is the single ligand-receptor (antigen-antibody) specificity that is used for recognition at the cell level. With multivalent interactions (forming multiple simultaneous ligand-receptor bonds) the specificity is extended to the receptor density, where only cells with cognate receptor concentrations above a certain threshold are targeted. Here we have shown that properly designed multivalent targeting of multiple cognate receptor types results in specificity towards a chosen receptor density profile, thus demonstrating a general route, and limits, of targeting cells without specific markers. In developing the theory we mainly considered the targeting of cells, however, our model can also be used to illuminate the mechanism behind the sorting ability of cell imprinted polymers [51, 107], discussed in detail in the following chapter.

4

Molecularly Imprinted Polymers

Let us descent further...

– a curious mountaineer

The simple multivalent model is now further extended to include the knowledge of specific positions of ligands and receptors. In the preceding chapter we focused on multi-component targeting with flexible geometry, i.e. specific positions of ligands/receptors are unimportant because the nature of the system is such (either flexible ligands or mobile receptors) that the bonding is flexible. We now turn to what was defined in the introductory chapter as the "rigid geometry" multivalency, where specific positions of binding sites on the multivalent entity are important (think interaction between 2 polymers, or between 2 DNA sequences). We apply the theory to study the selectivity of molecularly imprinted polymers.

4.1 Introduction

The term 'Molecularly Imprinted Polymers' (MIPs) is used to denote polymer matrices that have been "imprinted", *i.e.* cross-linked in the presence of a template molecule, thereby acquiring selective affinity towards its template. MIPs are usually made by free-radical co-polymerisation of ligands and cross-linkers in the presence of template molecules. The molecule-matrix interaction may exploit covalent binding, ionic interactions [32], hydrogen bonding [33], $\pi - \pi$ stacking interactions [34], hydrophobic interactions [35], and metal-ion chelation [36]. In 1930 Polyakov introduced this technique [37] to imprint silica matrices with benzene. However, the technique has only become widely used in recent decades [38–41]. The use of MIPs is related to the fact that they can be designed

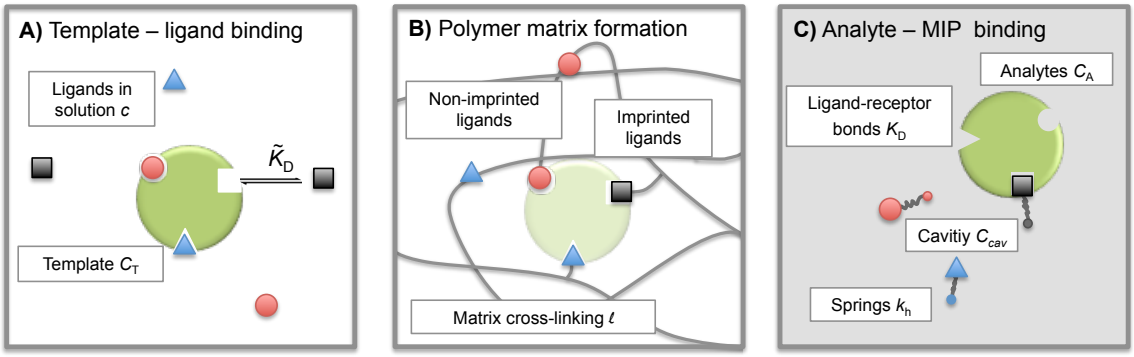


Fig. 4.1 **Schematic representation of the molecular imprinting process: A)** binding of ligands (functional monomers) from the solution to the template with specific receptors (binding sites), **B)** cross-linking and extracting of the template to create an imprinted cavity with ligands attached to the polymer matrix, and **C)** re-binding of an analyte to the cavity.

for highly selective recognition. Moreover, they combine thermal and chemical stability with ease of preparation, and hence low production costs.

MIPs have been used in applications such as solid-phase extraction [42], chiral separation [43], and catalysis [44]. They can act as molecular sensors [39, 41, 45], and mimic antibodies or enzymes [39]. They can selectively bind drugs [46–48], proteins [49], or even whole bacteria [50, 51]. Figure 4.1 shows a schematic representation of the imprinting and subsequent recognition process. The efficiency of the molecular recognition process depends on a number of parameters: 1) the initial ligand concentration c, 2) the template-ligand binding affinity K_D, and 3) the stiffness of the polymer matrix k_h.

Clearly, it is important to maximize the selectivity of MIPs, but in experiments MIPs are often optimized by trial and error. In fact, the theoretical picture is rather fragmented as existing theoretical models for MIPs do not consider the imprinting process as a whole, but rather tend to focus on individual steps in their mode of action [52–58]. Moreover, the atomistic and coarse-grained simulations of molecular imprinting that have been reported [59–64] focused mostly on specific MIPs and did not explore generic trends that would allow us to arrive at general design principles.

Here, we present a generic, coarse-grained statistical mechanics model that captures the key features of molecular recognition. The model provides an integrated description of the MIP formation process and of the subsequent binding of analytes. For the simplest case of divalent particles (i.e. particles with 2 receptors), we derive analytical expressions for the binding free energy and, from that, the adsorption isotherm of analytes on a MIP. For the general case of multivalent particles, we use Monte Carlo simulations to

obtain the adsorption isotherms. Two important measures of the quality of a MIP are: **i)** *how much more efficient is binding of a given analyte to imprinted* (MIP) *than to a non-imprinted matrix* (NIP), and **ii)** *how well can we separate two analytes that are only slightly different* (*e.g.*, of the same size and with the same number - but different spatial pattern - of the receptors). In order to address these questions, we evaluate standard measures of MIP specificity, such as the imprinting factor (*IF*) and the separation factor (*SF*), as a function of matrix and analyte properties and identify the optimal range of the control parameters such as template and ligand concentrations and polymer matrix stiffness. Our work provides insight into the generic features of MIPs operation and leads to a set of simple design principles.

4.2 Model

To construct a simple model for MIPs we describe the template and analyte molecules as impenetrable spheres with a diameter σ. At fixed (but otherwise arbitrary) positions on the surface of these spheres there are binding sites (referred to as 'receptors') that can bind to the functional monomers (referred to as ligands), see Figure 4.1. In step **A** of the MIPs formation, the template receptors bind ligands from the solution, while in step **B** the ligands are tethered via cross-linking, *i.e.* the polymer network is formed and ligands attached to the dangling ends of the network. In a MIP application **C** the receptors on analyte molecules bind to matrix-grafted ligands. The interactions between ligands and receptors are assumed to be valence limited: each receptor can bind to at most one ligand with a 'hybridization free energy' ΔG. Moreover, we assume that individual binding events are uncorrelated, *i.e.* we do not consider allosteric effects between different binding events[1]. We use the term "hybridization free energy" for binding of individual ligands to distinguish it from the free energy change associated with the binding of entire analyte to the cavity. We essentially consider templates and analytes as multivalent entities and our model approach is similar to that of Whitesides et. al. [65]. We treat templates and analytes as shapeless hard spheres, however, the chirality can be encoded by the specific positions of the distinguishable binding sites, see Appendix A for an example of the model application to enantiomeric separation.

In what follows we use the standard notation for the inverse temperature: $\beta \equiv 1/k_B T$ with k_B the Boltzmann constant and T the absolute temperature. Hybridization free

[1]The assumption of no correlations is no longer valid if the two ligands are imprinted very close to each other in a soft matrix, their movement and binding will be correlated. These effects can be added to the model if necessary.

energies for ligand-receptor bonds can be deduced from experimental data on binary association in solution, which yield the dissociation constant $K_D = \rho_0 e^{\beta \Delta G}$, where $\rho_0 = 1$ M is the standard concentration. We note that the ligand-receptor dissociation constant in the pre-polymerization solution $\tilde{K}_D$ can in principle be different from the one in a formed MIP K_D if the conditions such as solvent, pH, salt concentration, or temperature are different. Corrections for the fact that the ligands and receptors have an excluded volume and are tethered to a surface, can be computed (see refs. [111, 9]).

Using standard chemical equilibrium theory we can compute the amount of ligands adsorbing to the templates in step **A**, which depends on the ligand dissociation constant at the conditions of the imprinting phase, $\tilde{K}_D$, on the concentrations of templates C_T and on c, the original concentration of ligands in solution (see Appendix). The fractional occupancy of receptors f_r (the probability that a given receptor is bound to a ligand) is

$$f_r = \frac{\tilde{K}_D + n_r C_T + c - \sqrt{(\tilde{K}_D + n_r C_T + c)^2 - 4 c n_r C_T}}{2 n_r C_T} \,, \tag{4.1}$$

with n_r the number of receptors per template particle. If there are different types of ligand-receptor pairs, and no cross-binding, (4.1) should be applied to each type separately. Subsequent cross-linking (Figure 4.1**B)**) ensures that the adsorbed ligands remain tethered to the matrix after the template has been extracted, resulting in a population of cavities with "imprinted" ligands. Depending on the conditions, the cavities can contain as many ligands as there are receptors on the template surface ($f_r \sim 1$), or fewer in case the ligand-template binding was not saturated (or more when we consider also free non-bound ligands to be part of a cavity).

4.2.1 Polymer matrix elasticity

In our model we must account for the effect of the deformability of the matrix and for the directionality of the ligand-template interaction. To this end, we model MIPs as soft, deformable matrices that contain cavities imprinted by specific ligands (Figure 4.1**C**). The ligands grafted to the matrix can fluctuate around their equilibrium positions due to the thermal fluctuations. We call the equilibrium positions of the imprinted ligands their "anchors". In order to bind to receptors on the analyte, the ligands need to be displaced from their anchors, which increases the elastic free-energy of the matrix by an amount U_h. Following Rubinstein and Colby [112], we assume that the ligand-matrix interaction only depends on the distance between the receptor and the anchor and that it can be

replaced by a harmonic spring:

$$U_h = \frac{k_h}{2} \left| \mathbf{r}^{lig} - \mathbf{r}^{anc} \right|^2 , \tag{4.2}$$

where $\mathbf{r}^{anc}$ is the anchoring position of the ligand and $\mathbf{r}^{lig}$ its actual position. Assuming that the matrix is a linearly elastic medium, we can use the normal mode analysis and relate the effective spring constant k_h to directly measurable macroscopic quantities, *viz.* the bulk modulus B and the shear modulus G.

The deformation energy density of a wave with wave-vector $\mathbf{k}$ and amplitude x_k in an isotropic, homogeneous, elastic medium follows from linear elasticity theory

$$\frac{E}{V} = \frac{1}{2} M \langle x_k^2 \rangle \kappa^2, \tag{4.3}$$

E/V is the energy per volume, M the effective elastic modulus, $\langle x_k^2 \rangle$ the mean square amplitude of the wave and $\kappa = [k_x, k_y, k_z]$ is the wave vector. M is determined by the bulk B and shear G moduli of the polymer matrix: $M_\parallel = B + \frac{4}{3}G$ for longitudinal waves and $M_\perp = G$ for transverse waves. Using equipartition, we can rewrite the above equation as

$$\langle x_\kappa^2 \rangle = \frac{k_B T}{M V \kappa^2}, \tag{4.4}$$

and find the contribution to the mean square displacement for a single normal mode with wave vector κ.

We are interested in the fluctuations in the relative distance between 2 ligands that are anchored at a distance $\mathbf{r}_{ij}$. The mean square fluctuation in the relative displacement of 2 points i and j at distance r_{ij} is given by

$$\langle x_{\kappa,ij}^2 \rangle = \frac{2 k_B T}{M V \kappa^2} \cos(1 - \kappa \cdot \mathbf{r}_{ij}). \tag{4.5}$$

This holds for both longitudinal and transverse modes, in the following we will demonstrate the calculation for longitudinal modes, the calculation for transverse modes is identical except the we replace the modulus $M_\parallel$ with $M_\perp$. To find the mean square displacement contributed by all longitudinal waves we simply sum over all modes as different modes are orthogonal to each other (uncorrelated)

$$\langle \Delta R_{\parallel,ij}^2 \rangle = \sum_\kappa \langle x_\kappa^2 \rangle = \sum_\kappa \frac{2 k_B T}{M_\parallel V \kappa^2} \cos(1 - \kappa \cdot \mathbf{r}_{ij}). \tag{4.6}$$

Converting the sum over wave vectors to an integral, we get

$$\langle \Delta R^2_{\parallel,ij} \rangle = \frac{k_B T}{M_\parallel (2\pi)^3} \int_{k_{min}}^{k_{max}} \frac{2(1-\cos(\kappa \cdot \mathbf{r}_{ij}))}{\kappa^2} d\kappa, \tag{4.7}$$

k_{max} and k_{min} are short and long wavelength cut offs for the elastic modes. The upper limit (the ultraviolet cut off) $k_{max} = \frac{\pi}{\ell}$ is determined by the smallest spacing in the medium ℓ. The meaning of ℓ depends on the nature of the polymeric material: for cross linked gels, ℓ denotes the average distance between adjacent cross-links. For a glassy polymer, ℓ should be of the order of the contact distance between two non-bonded monomers. The lower cutoff $k_{min} = 2\pi/L$ is determined by the system size L. The integral above can be solved evaluated analytically to yield

$$\langle \Delta R^2_{\parallel,ij} \rangle = \frac{k_B T}{M_\parallel (2\pi)^3} 8\pi \left[k_{max} - k_{min} - \frac{Si(k_{max} r_{ij}) - Si(k_{min} r_{ij})}{r_{ij}} \right], \tag{4.8}$$

where Si denotes the sine integral. In a large system we can neglect k_{min} as $k_{min} \ll k_{max}$ and $Si(k_{min} r_{ij}) = 0$. We therefore obtain:

$$\langle \Delta R^2_{\parallel,ij} \rangle = \frac{k_B T}{M_\parallel \pi^2} \left[\frac{\pi}{\ell} - \frac{Si(\pi r_{ij}/\ell)}{r_{ij}} \right]. \tag{4.9}$$

It is instructive to consider a few limiting cases of Eqn. (4.9). First, we consider the case that the distance between the ligands is larger than the mesh size ($r_{ij} \gg \ell$). In that case $Si(\pi r_{ij}/\ell) \approx Si(\infty) = \pi/2$ and we find that $\langle \Delta R^2_{\parallel,ij} \rangle = \frac{k_B T}{M_\parallel \pi} \left[\frac{1}{\ell} - \frac{1}{2 r_{ij}} \right]$. In the opposite limit ($r_{ij} \ll \ell$), we can expand the sine integral to third order $Si(\pi r_{ij}/\ell) \approx \pi r_{ij}/\ell + \frac{1}{18} \left(\frac{\pi r_{ij}}{\ell} \right)^3$ and the first terms cancel out $\langle \Delta R^2_{\parallel,ij} \rangle \approx \frac{\pi k_B T}{18 M_\parallel \ell^3} r^2_{ij}$. The relative fluctuations are proportional to the relative distance.

Adding the longitudinal ($\parallel$) and transverse ($\perp$) contributions, we get

$$\langle \Delta R^2_{ij} \rangle = \langle \Delta R^2_{\parallel ij} \rangle + 2 \langle \Delta R^2_{\perp ij} \rangle, \tag{4.10}$$

The longitudinal and transverse components depend on the elastic properties of the material: the mean square amplitude of parallel fluctuations is $\langle \Delta R^2_{\parallel ij} \rangle \propto M_\parallel^{-1} = (B + 4G/3)^{-1}$, which is smaller than the transverse fluctuations $\langle \Delta R^2_{\perp ij} \rangle \propto M_\perp^{-1} = G^{-1}$. In what follows, we will use an isotropic average of Eq. (4.10) to describe the continuum fluctuations

$$\langle \Delta R^2_{ij} \rangle = 3 \frac{k_B T}{M_e \pi} \left[\frac{1}{\ell} - \frac{Si(\pi r_{ij}/\ell)}{\pi r_{ij}} \right] \tag{4.11}$$

with the effective modulus M_e given by

$$M_e = 3\left(\frac{1}{B + 4G/3} + \frac{2}{G}\right)^{-1}.$$

For flexible gels we can also use ideal-chain statistics to compute the mean-squared amplitude of the fluctuations of polymer strands between anchoring points. In a polymer gel with a Kuhn segment b much smaller than the cross-linking distance $b \ll \ell$, the cross-linking distance is related to the number n of Kuhn segments of the cross-linking chain by $\ell^2 \simeq nb^2$. On very short length scales $r_{ij} \ll \ell$ the gel structure is unimportant and only the single chain statistics is relevant. In this case the fluctuations are found to be of the order

$$\langle \Delta R_{c,ij}^2 \rangle \approx r_{ij}^2. \tag{4.12}$$

Let us compare this result with elastic fluctuations in a gel (Eq. 4.9). The shear modulus of a gel is $G \simeq k_B T/\ell^3$, assuming a phantom network model [112], we remember ℓ is the mean distance between cross-links. For a gel with a Poisson ratio close to zero, we have $B = 2G/3$ and hence the effective modulus becomes $M_e \simeq \frac{6}{5} k_B T/\ell^3$. On short length scales $r_{ij} \ll \ell$ we expand the sine integral to third order and Eq. (4.11) reduces to

$$\langle \Delta R_{ij}^2 \rangle = \frac{5\pi}{36} r_{ij}^2. \tag{4.13}$$

Comparing the equation above and (4.12) we observe that the result obtained are the same except for a pre-factor of order 2. This validates the use of (4.11) for polymer gels.

Now we only need to relate the total fluctuations to the effective spring constant $k_h = \frac{3k_B T}{\langle \Delta R^2 \rangle}$. Individual fluctuations of a ligand are half of the relative ones between 2 ligands $\langle \Delta R^2 \rangle = \frac{1}{2}\langle \Delta R_{ij}^2 \rangle$, because we assumed, in our model, that each ligand is independent from the rest. For brevity we also approximate $r_{ij} \approx \sigma$ the distance between ligands of interest will be about the template (cavity) size, hence the effective spring constant is

$$k_h = \pi M_e \left[\frac{1}{\ell} - \frac{Si(\pi\sigma/\ell)}{\pi\sigma}\right]^{-1}. \tag{4.14}$$

This result can be further simplified, the sine integral function is well approximated by a polynomial $Si(x) \approx x - x^3/18;\ x \ll \pi$ and a constant $Si(x) \approx \pi/2;\ x \gg \pi$. Hence, the

effective spring constant is approximately

$$k_h = \pi M_e \left[\frac{1}{\ell} - \frac{1}{2\sigma}\right]^{-1} ; \ \sigma > \ell \tag{4.15}$$

$$k_h = \frac{18 M_e \ell^3}{\pi \sigma^2} ; \ \sigma < \ell \tag{4.16}$$

We recall that the effective modulus is $M_e = 3\left(\frac{1}{B+4G/3} + \frac{2}{G}\right)^{-1}$, therefore, k_h is fully determined by the properties of the polymer matrix: The bulk B and shear G modulus, the mesh size ℓ and the cavity (template) size σ.

We see that k_h depends weakly on the template (cavity) size σ because the relative fluctuations of 2 ligands will decrease if the two ligands are closer than the mesh size ℓ. For flexible polymer gels $M_e \approx k_B T/\ell^3$ [112] and large particles $\sigma > \ell$ the spring constant is determined simply by the cross-linking distance $k_h \approx \pi k_B T/\ell^2$. Eqn. (4.14) is important because, as we show below, it allows us to predict the effect of the stiffness of the matrix on the selectivity of MIPs. We will focus on the case where particles are rigid, however, the present model can be applied also to soft particles (such as proteins) where receptor positions on the particle itself are fluctuating (fluctuations characterised by spring constant k_h^{part}). In this case we could map (to the first order) a soft particle and a soft matrix to the current model of a hard particle and an "effective" softer matrix $k_h^{eff} = \left(\frac{1}{k_h} + \frac{1}{k_h^{part}}\right)^{-1}$.

In what follows we first focus on a single cavity system and calculate the free energy of binding an analyte. Afterwards we extend the picture to the whole polymer matrix and calculate corresponding binding affinities and adsorption isotherms.

4.3 Binding free energy

The binding free energy F for the analyte-cavity system can be decomposed into a specific interaction part F^{cav} due to the bond formation and a non-specific part F_{ns}, which includes other possible contributions to particle adsorption such as excluded volume, hydrophobic, electrostatic or van der Waals interactions between the particle and the polymer matrix. The non-specific term depends on the particle size, shape and is thus similar for similar analytes. These terms add to the binding free energy, however as we will show below, they cancel out in the ratios defining the imprinting and separation factors – as long as the analyte particles are similar-enough [54]. For differently shaped analytes a study by Simon *et. al.* [113] concluded that the effect of the analyte shape becomes less important when the number of binding sites (receptors) on the analyte is

larger than 2. Therefore, we only evaluate the specific part of the binding free energy due to bond formation. In the divalent case the specific F_2^{cav} can be calculated analytically.

The position $\mathbf{r}_p$ and orientation $\mathbf{\Omega}_p$ of a rigid particle uniquely define the positions of all binding sites on its surface $\mathbf{r}_i^{bs}$. When the particle approaches the cavity, bonds between the binding sites i and the ligands j can be formed - forming a bond results in a lowering of the free energy by an amount equal to the hybridization free energy ΔG_{ij}, but forming a bond also costs free energy because (in general) the ligand must be displaced from its equilibrium position $\mathbf{r}_j^{anc}$ to $\mathbf{r}_j^{lig} = \mathbf{r}_i^{bs}$: the corresponding free energy cost is given by (4.2). To compute the overall binding free energy, we need to compute the ratio of the partition function for the case where one or more ligands are bound to the particle, to the one for the case with no ligands bound.

This expression does not yet include the partition function of the unbound particles: this will later be accounted for through the chemical potential of the free particles. The partition function depends on the distribution of the binding sites on the particle $\mathbf{r}^{bs}$ and on the arrangement of ligand anchors in the cavity $\mathbf{r}^{anc}$. To obtain the partition function we must sum over all particle positions and orientation and over all possible bonding arrangements:

$$\mathcal{R} \equiv \frac{Q_b}{Q_{\text{free}}} = \sum_{\kappa=1}^{\kappa_{max}} \frac{(N_A \rho_0)^{1-\kappa}}{8\pi^2} \left(\frac{\beta k_h}{2\pi}\right)^{\frac{3\kappa}{2}} \sum_{s(\kappa)} \int d\mathbf{r}^p d\mathbf{\Omega}^p e^{-\beta \sum_{ij} (\Delta G_{ij} + \frac{k_h}{2} |\mathbf{r}_i^{anc} - \mathbf{r}_j^{bs}|^2)} \equiv \sum_{\kappa=1}^{\kappa_{max}} q_\kappa \ , \tag{4.17}$$

where N_A denotes Avogadro's constant. The full bound state partition function has been decomposed to a sum over partition functions q_κ for a subset of configurations with κ bonds formed, $s(\kappa)$ denotes all distinct configurations (bonding arrangements) with κ bonds. The maximum number of bonds, κ_{max}, is defined by the total number of binding sites or adjacent ligands, whichever is lower. The partition function $\mathcal{R}$ is directly related to the specific part of the binding free energy:

$$F^{cav} = -k_B T \ln \mathcal{R} \ . \tag{4.18}$$

In (4.17), the terms with $\kappa = 1$ and $\kappa = 2$ can be evaluated analytically:

$$q_1 = \sum_{ij} e^{-\beta \Delta G_{ij}} \ , \tag{4.19}$$

where the sum is over all possible ligand - receptor pairs i, j that can form a single bond.

As a single bond cannot create a static stress in the matrix, q_1 does not depend on the matrix stiffness k_h. A similar result was reported by Tanaka *et. al.* for the case of

imprinted hydrogels [56]. The simplest non-trivial term is the two-bond partition function. For any chosen combination of two binding sites $\mathbf{r}_i^{bs}, \mathbf{r}_{i'}^{bs}$ and two ligands $\mathbf{r}_j^{anc}, \mathbf{r}_{j'}^{anc}$, there are two possible ways of forming two bonds. The total two-bond partition function in a system where there is more than one way to make two bonds is a sum of all possible two-bond pairs $ij, i'j'$,

$$q_2 \equiv \sum_{ij,i'j'} \tilde{q}_2(a_{ii'}, b_{jj'}) e^{-\beta(\Delta G_{ij} + \Delta G_{i'j'})} \ , \tag{4.20}$$

with $b_{ii'} = |\mathbf{r}_i^{bs} - \mathbf{r}_{i'}^{bs}|$ the distance between the binding sites, and $a_{jj'} = |\mathbf{r}_j^{anc} - \mathbf{r}_{j'}^{anc}|$ the distance between the two ligand anchors (see inset of Figure 4.2**a)**). The configurational part of the partition function $\tilde{q}_2(a, b)$ is a solution of a coupled Gaussian integral and can be calculated analytically (see Appendix A):

$$\tilde{q}_2(a, b) = \frac{(\beta k_h)^{\frac{1}{2}}}{4\pi^{\frac{3}{2}} N_A \rho_0} \frac{\sinh\left(\frac{1}{2}\beta k_h ab\right)}{ab} \ e^{-\frac{\beta k_h}{4}(a^2+b^2)} \ . \tag{4.21}$$

Eqn. (4.21) can be rewritten in terms of affinity constants, which is useful in order to connect to the previous wor on multivalent binding [65, 66] and to most of the experimental literature. For a system with two ligands and two binding sites (for simplicity we assume that all bonds are equal: $\Delta G_{ij} \equiv \Delta G$), the analyte - cavity equilibrium association constant is $K_A^{cav} = e^{-\beta F^{cav}}/\rho_0 = 4/K_D + 2K^{intra}/K_D$, where $K_D = \rho_0 e^{\beta \Delta G}$ is the single bond dissociation constant. Note that we have used a symbol K_A^{cav} to denote the binding to the cavity, however, this is equivalent to the avidity constant K_A discussed in the preceding chapters. The internal equilibrium constant (the facilitation of forming the second bond once the first one is present) is $K^{intra} = \tilde{q}_2(a, b)\rho_0/K_D$, with $\tilde{q}_2(a, b)$ precisely the configurational part of the partition function given by (4.21). Our analytical approach thus enables us to calculate the internal association constant K^{intra} for divalent binding. This result goes beyond the scope of molecular imprinting, it is a general solution and applicable to any divalent entity binding within the harmonic approximation (4.2).

In the limit of soft matrices $k_h \ll k_B T/b^2$, where thermal fluctuations are greater then the analyte size, the specificity towards analyte geometry is lost $\tilde{q}_2 = \frac{(\beta k_h)^{3/2}}{8\pi^{3/2}\rho_0 N_A}$. This is the regime where only the number of ligands in the cavity (or the number of receptors on the analyte) remain important, not their spatial positions. In this limit the rotational degrees of freedom can be neglected, such a model of molecular imprinting has been considered in [56, 57]. This regime is equivalent to the all-to-all binding we focused on the previous chapter.

If we assume that all binding cavities are identical and independent, the number of adsorbed analyte particles is determined by the binding free energy F and by the chemical potential μ of the analyte. We recover the simple Langmuir expression for the fraction of occupied cavities:

$$f_{cav} = \frac{e^{\beta(\mu - F)}}{1 + e^{\beta(\mu - F)}} = \frac{e^{\beta(\mu' - F^{cav})}}{1 + e^{\beta(\mu' - F^{cav})}} \tag{4.22}$$

where $\mu' \equiv \mu - F_{ns}$ is the rescaled chemical potential incorporating the non-specific analyte-matrix interactions. In practice, we can either first evaluate the binding free energy F^{cav} (*i.e.* in the case of divalent binding) and from it the adsorption isotherm f_{cav}, or *vice versa* (in all other cases). Since higher order partition function integrals, q_3, q_4 etc., become increasingly complex and are only analytically tractable in rather special (and not very realistic) cases; we use numerical simulations (Grand Canonical Monte Carlo simulations) to compute the binding probability f_{cav}, following the approach of ref. [9], and then invert (4.22) to compute the binding free energy $\beta F^{cav} = \beta\mu' - \log \frac{f_{cav}}{1-f_{cav}}$. The simulation code to compute the binding free energies for arbitrary template and analyte parameters is freely available online at github.com/tc387/mipsp.

4.4 Cavity binding results and discussion

In order to cast our results in the most general form, we introduce the following dimensionless quantities: matrix stiffness $k_h^* = k_h\sigma^2/k_BT$, dissociation constant $K_D^* = K_D N_A \sigma^3$, concentrations $c^* = cN_A\sigma^3$, distances $a^* = a/\sigma$ and analyte-cavity binding free energy $\Delta F^* = F/k_BT + \ln(\sigma^3 N_A \rho_0)$. The single bond hybridization free energy then follows $\Delta G^* = \ln(K_D^*) = \Delta G/k_BT + \ln(\sigma^3 N_A \rho_0)$. For example, if the template size is $\sigma = (\rho_0 N_A)^{-1/3} = 1.18$ nm the rescaled and standard values of equilibrium constant, concentration and free energy remain the same. For dilute solutions of analytes we can assume an ideal solution and the chemical potential becomes $\mu = k_BT \ln(\frac{C_A}{\rho_0})$, with C_A the molar concentration of analytes, rescaling $\mu^* = \ln(C_A^*)$

MIPs that have been cross-linked in the presence of specific template particles will adsorb analytes that have a structure similar to the template: the greater the similarity between template and target particle, the larger the average occupancy f_{cav} of the cavities imprinted by the template particles. In the case of particles with two binding sites, we can compute f_{cav} analytically (Eqs. 4.17-4.22). In Figure 4.2**a)** we compare these analytical calculations with Monte Carlo simulations. The good agreement between analytical result and simulations allows us to validate the Monte Carlo approach. For

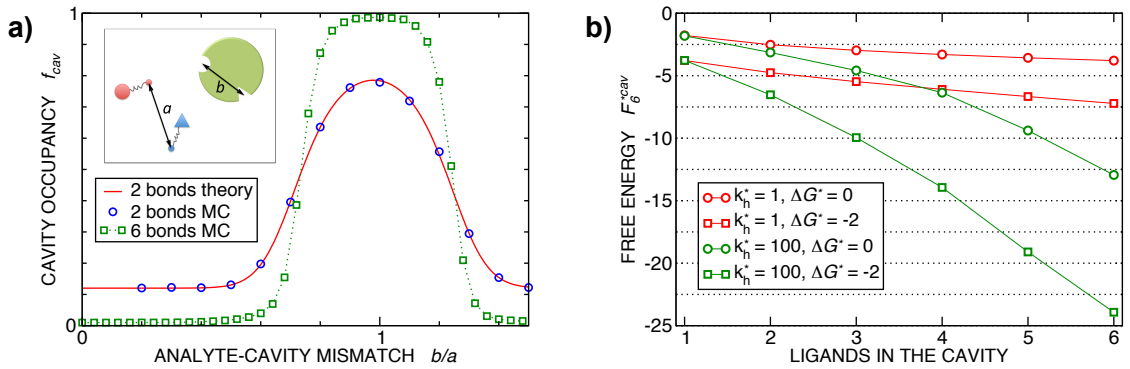


Fig. 4.2 **a) Multivalency.** The cavity occupancy f_{cav} as a function of the imprinting mismatch b/a. The red solid line represents analytical calculations (4.17-4.22). The symbols depict the results of Monte Carlo simulations: blue circles for divalent, green squares for hexavalent case. Parameters: matrix stiffness $k_h^* = 100$, analyte concentration in solution $C_A^* = 3 \cdot 10^{-4}$, and the bond energies in the divalent/hexavalent case $\Delta G_{2b}^* = -5.4$, $\Delta G_{6b}^* = 0$. **b) Incomplete cavities.** The specific binding free energy F_6^{*cav} of a hexavalent analyte as a function of the number of ligands imprinted in the cavity. The results are for two values of k_h^* (red: soft gel, green: stiff gel) and for two values of ligand binding strength ΔG^* (circles: weaker bonds, squares: stronger bonds).

a particle with 6 binding sites, the analytical calculations are no longer tractable, but the MC simulations in Figure 4.2 show that imprinting with a higher valency leads to a much stronger discrimination. In both cases, the cavities have a fixed ligand distribution that is imprinted by the template particle. The analytes are assumed to have the same geometry as the template but their sizes are rescaled by a factor b/a (the "imprinting mismatch"), see inset of Figure 4.2**a)**. Not surprisingly, the average occupancy f_{cav} has a maximum at $b/a \simeq 1$. The difference between two and six binding sites shows up when we increase the mismatch: for the case of two binding sites, a mismatch of 30% decrease f_{cav} only by a factor two. In contrast, for particles with six binding sites, a 30 % mismatch leads to a decrease of f_{cav} by more than an order of magnitude.

Finally, Figure 4.2**b)** illustrates how the binding depends on the number of ligands in the cavity. There are many practical examples where the number of ligands in the cavity is less than the number of receptors on the template. Obviously, we expect to observe 'under-coordinated' cavities in case the cavities are only partially formed. But even if cavities are initially well formed, they may become under-coordinated if the MIP is ground after imprinting (during the grinding, which is a common procedure to enhance the accessibility of cavities in stiff matrices [48], a fraction anchored ligands is likely to be detached from the cavities). Not surprisingly, we find that the binding strength increases with the number of ligands. The dependence is steepest for stiffer matrices (larger k_h).

4.5 MIP characterization

Having derived the coarse-grained elastic model of the polymer matrix and calculated the binding free energy of analytes to imprinted cavities, we now focus on characterizing MIPs with measures used in most of the literature: The binding affinity, the imprinting factor and the separation factor. A formed MIP in principle consists of heterogeneous cavities, therefore a MIP is fully characterized by a binding isotherm (Figure 4.3) or an affinity distribution of cavities, e.g. ref. [48, 114]. However, it is useful to have a single number measure of MIPs. Following Tanaka [56] we define the dimensionless binding affinity as the product of the concentration of cavities C_{cav} and the average equilibrium association constant of binding an analyte to a cavity

$$BA^{cav} \equiv C_{cav}\langle K_A^{cav}\rangle \tag{4.23}$$

where the equilibrium association constant $K_A^{cav} = e^{-\beta F^{cav}}/\rho_0$ is determined by the binding free energy of an analyte binding to the cavity (4.18). The bracket denote a (number weighted) average over all cavities in a MIP taking into account the heterogeneous distribution of cavities where applicable. A MIP is a collection of imprinted cavities, but in principle also other non-imprinted ligands

$$BA^{mip} = BA^{cav} + \mathcal{O}\,. \tag{4.24}$$

The term $\mathcal{O}$ includes corrections due to binding to non-imprinted ligands outside cavities as well as cross-cavity binding (analyte binding to two or more cavities simultaneously). If we assume that cavities are randomly distributed throughout the MIP, the correction term becomes closely related to the binding affinity for non-imprinted polymers (NIPs) described below $\mathcal{O} \approx BA^{nip}$. In chromatography experiments the binding affinity is expressed in terms of the retention factor of the analytes in the column. The retention factor and equilibrium constant are monotonically related [115, 116].

In case of NIPs, there are no imprinted cavities and the average in (4.23) must be taken over a random distribution of ligands in the cross-linked matrix. The result derived in the Appendix A is

$$BA_{n_r}^{nip} = \left(1 + \frac{c}{K_D}\right)^{n_r} - 1 \tag{4.25}$$

for a single ligand-receptor type. We remember that c and K_D are the ligand concentration and ligand-receptor dissociation constant respectively. Heuristically, each receptor can be either free (weight 1), or bound (weight $\frac{c}{K_D}$) and there are n_r independent receptors on the particle. This result and its derivation is conceptually similar to the theory

describing multivalent particles binding to a surface [10]. We observe that, for strong enough binding $\frac{c}{K_D} > 1$ the binding affinity will scale as $BA_{n_r}^{nip} \sim \left(\frac{c}{K_D}\right)^{n_r}$, while for weak binding $\frac{c}{K_D} < 1$ it is approximately linear $BA_{n_r}^{nip} \sim n_r \frac{c}{K_D}$. Such scaling has been observed experimentally for collapsed and swollen hydrogels respectively [56, 57].

In the divalent case we can evaluate the free energy - and thus the binding affinities - analytically (Eqs. 4.17-4.21, 4.25). Assuming that imprinted cavities are randomly distributed throughout the MIP the cross-cavity term becomes equal to BA_2^{nip}, however, single bond terms need to be subtracted to prevent double counting. If the template extraction process is efficient the cavity concentration is equal to the template concentration in imprinting phase $C_{cav} = C_T$. The MIP binding affinity becomes

$$BA_2^{mip} = \frac{2\tilde{q}_2 C_T f_r^2 \rho_0}{K_D^2} + BA_2^{nip} \ , \tag{4.26}$$

with $\tilde{q}_2$ the 2 bond partition function (4.21) and f_r the receptor occupancy fraction (4.1). Heuristically, the first term on the right takes into account divalent binding to cavities, the concentration of doubly functionalized cavities being $f_r^2 C_T$. The second term BA_2^{nip} takes into account all single bond states and divalent binding to all pairs of ligands that do not belong to the same cavity.

In the more general case $\kappa > 2$, BA^{mip} can be computed by numerical simulations. From Eqs. (4.22,4.23) we observe that for low concentrations the binding affinity essentially determines the ratio of concentrations of bound analytes C_B to analytes free in solution C_A

$$BA = \frac{C_B}{C_A} \ . \tag{4.27}$$

On Figure 4.3 we show adsorption isotherms of divalent analytes binding to a MIP and NIP along with representative simulation snapshots. At low concentrations, we observe a good agreement between analytical isotherms (4.25-4.27) and isotherms obtained from numerical simulations. The small discrepancy at low concentration arises because in calculating (4.26) we have assumed that cavities are randomly distributed in the MIP, however, that is only approximate as cavities cannot overlap in our simulations. At higher concentrations the polymer matrix saturates i.e. nearly all ligands are bound. Interestingly, in this regime the non-imprinted matrix can support a slightly greater number of adsorbed analytes than the imprinted matrix. We explain this by observing that in the non-imprinted polymer ligands are mostly far apart and, therefore, act as single binding sites. On the other hand, imprinted ligands tend to be found in pairs (sharing a cavity) and analyte-analyte repulsion makes binding of 2 analytes to a single cavity unfavourable.

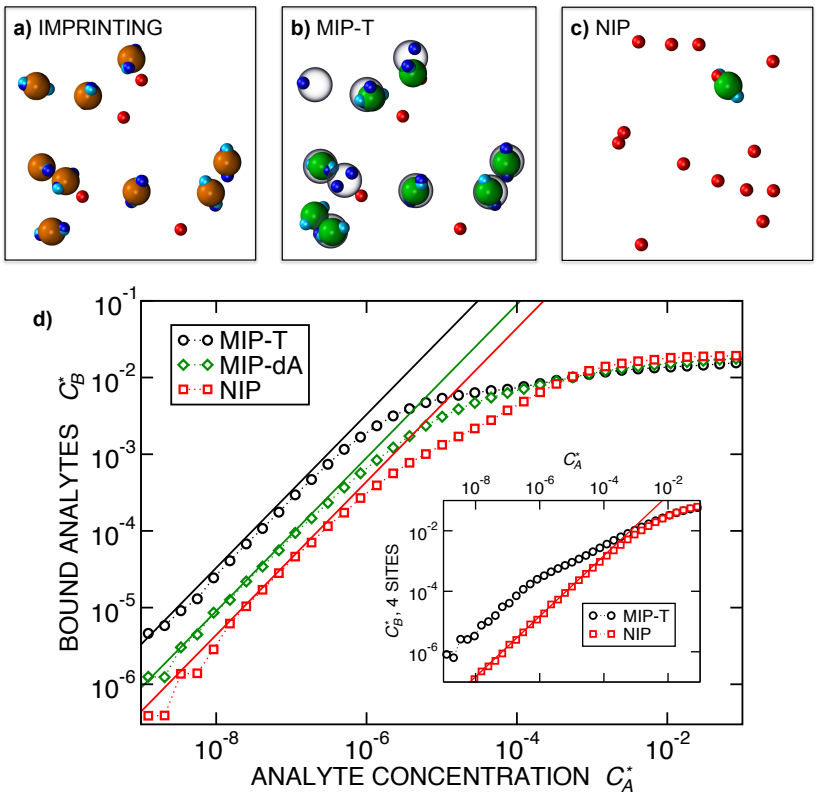


Fig. 4.3 **Simulation snapshots. a)** Imprinting phase: ligands (small balls, red when free and blue when bound) are binding to the divalent templates (orange balls with cyan receptors). The ligand positions are subsequently frozen and templates extracted to form the cavities (MIP) shown on **b)** and **c)** as transparent blobs. **b)** Re-binding of templates to MIP. **c)** Binding of templates to non-imprinted matrix (NIP). Snapshots show only a small part of the simulated system. **d) Adsorption isotherms.** Average number of bound divalent (tetravalent in the inset) analytes C_B^* depending on the concentration of analytes in solution C_A^*. Solid lines depict the analytical prediction (4.25-4.27), and the symbols are for simulation results (black circles: re-binding of templates ($b = a$) to MIP as on snapshot **b)**, green diamonds: binding of different analytes ($b = a/\sqrt{2}$) to MIP, and red squares: binding of templates to a NIP – as on snapshot **c)**). The parameters: $k_h^* = 100$, $c^* = 0.02$, $C_T = 0.01$, $K_D^* = 0.001$. The snapshots represent a configuration of bound analytes at $C_A^* = 10^{-5}$ with the system size $V = (12.6\sigma)^3$. Inset in **d)** shows isotherms of tetravalent particles binding at $c^* = K_D^* = 0.1$.

The binding affinities BA^{nip} and BA^{mip} can be used to evaluate two key quantities used to assess the performance of MIPs in the literature: the imprinting factor $IF(a)$, which is the ratio of the binding affinity of a template a to a MIP and to a NIP, and the separation factor $SF(b_1, b_2; a)$ measuring the ability of a MIP (imprinted by a template a) to distinguish between two different analytes b_1 and b_2:

$$IF(a) \equiv \frac{BA^{mip}(a;a)}{BA^{nip}} , \qquad SF(b_1, b_2; a) \equiv \frac{BA^{mip}(b_1;a)}{BA^{mip}(b_2;a)} . \quad (4.28)$$

$BA^{mip}(b;a)$ here denotes the affinity of the analyte b to the matrix imprinted by the template a. In calculating the binding affinities above we have assumed that all cavities and non-imprinted ligands in the matrix are accessible. For very dense matrices only surface cavities and ligands are accessible, in this case the effective concentrations, and therefore, binding affinities, will be lower. However, this effect is expected to largely cancel out in the ratios defining the imprinting and separation factors.

4.6 Design Principles

We can now return to the original question: *how to design the imprinting process to achieve optimal sensitivity for the desired application?* To arrive at a set of rules, we have summarized the results of our model calculations into a 'phase diagram' that shows the regime where MIPs should function most efficiently (see Figure 4.4**a)**). The control parameters in the phase diagram are the stiffness of the polymer matrix and the concentration of the ligands that are incorporated in the matrix. For generality, we will again use dimensionless quantities defined above. This phase diagram suggests three general design rules for efficient molecular imprinting:

MIP formation. In the MIP formation process (step **A** on Figure 4.1) the ligands should bind to the template in sufficient numbers. The binding of ligands can be calculated assuming chemical equilibrium between ligands and receptors (4.1), a simple rule of thumb would be that the concentration of ligands should be similar or greater than the bond dissociation constant

$$c^* > c^*_{form} = \tilde{K}^*_D . \quad (4.29)$$

Moreover, ligands and template receptors should be in approximately stoichiometric ratio, $c^* \approx n_r C^*_T$, with n_r the number of receptors per template. This rule, which

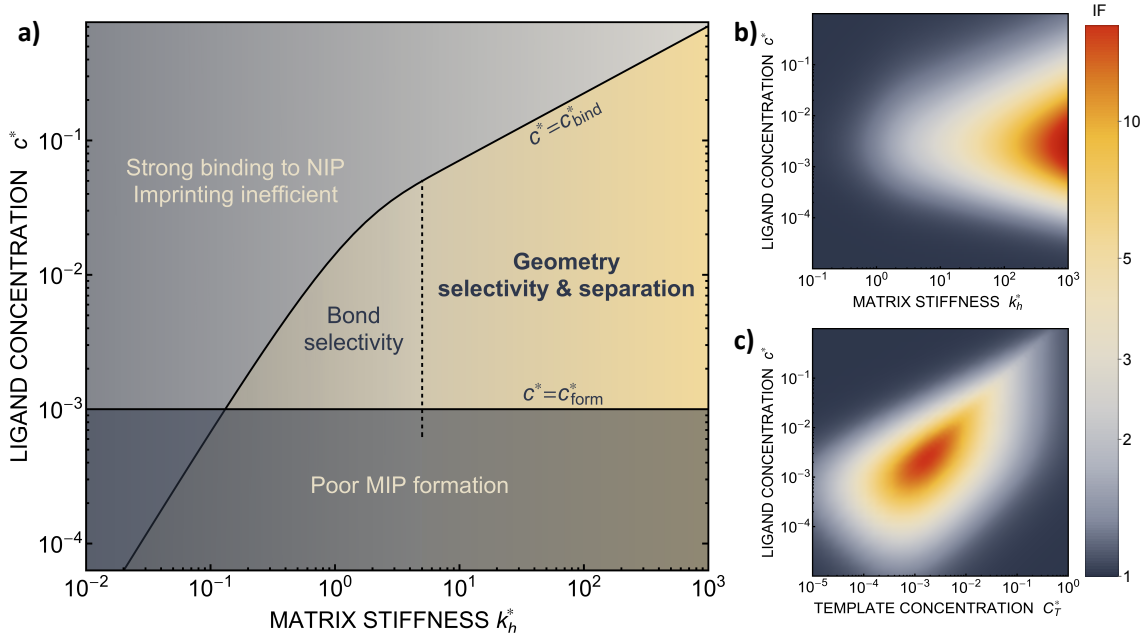


Fig. 4.4 **MIPs design principles. a)** Schematic phase diagram of MIPs efficiency, summarizing the design principles. The ligand concentration for which MIPs are efficient lies between the two limits: $c^* < c^*_{bind}$ (4.30) and $c^* > c^*_{form} = \tilde{K}^*_D$ (4.29). The dashed line drawn somewhat arbitrarily at $k^*_h = 5$ is separating the regions of bond selectivity and geometrical recognition. In **b)** and **c)** the imprinting factor calculated for divalent templates is shown as a function of k^*_h and c^* at a stoichiometric ratio of ligands-receptors $c^* = 2C^*_T$ **(b)**, and as a function of C^*_T and c^* at fixed matrix stiffness $k^*_h = 100$ **(c)**. We have assumed an equal binding strength in the imprinting and binding stage $\tilde{K}^*_D = K^*_D = 0.001$.

is supported by the data in Figure 4.4 **c)**, provides an optimal tradeoff between the formation of multi-ligand cavities on the one hand, and the minimization of the number of non-imprinted ligands on the other hand. A similar empirical observation was reported in systematic experiments by Kim and Spivak [58], and also in lattice model simulations by Shimizu et. al. [64]. If there are many different types of ligands in the solution (binding to different receptors), the rule above should be applied to each of them. The "Poor MIP formation" region $c^* < c^*_{form}$ is shaded on the phase diagram in Figure 4.4**a)**.

MIP binding. Imprinting should make a difference, *i.e.* templates should predominantly bind to the imprinted cavities rather then across them or to randomly distributed ligands, $BA^{cav} > BA^{nip}$. For a divalent template this results in a condition for the concentration of ligands:

$$c^* < c^*_{bind} = \sqrt{k^*_h}(1 - e^{-k^*_h})/8\pi^{3/2} \, . \tag{4.30}$$

The curve $c^*(k^*_h) = c^*_{bind}$ sets the upper bound for the yellow region of efficient MIPs in Figure 4.4**a)**. Additionally, the receptor-ligand binding in step **C** should be strong-enough such that predominantly multiple bonds are formed ($q_2 > q_1$), which results in a similar condition: $K^*_D < c^*_{bind}$.

Cross-linking strength. The matrix stiffness plays a crucial role in the performance of MIPs, the higher the stiffness k^*_h the greater the MIP selectivity, this has also been observed experimentally [117]. In order to separate analytes based on the geometry (receptor patterns) – the polymer matrix has to be stiff enough: $k^*_h > \Delta b^{-2}$, where Δb^* is a measure for the difference in geometry. For divalent particles it is the difference between the receptor distances $\Delta b = b_2 - b_1$ on the 2 analytes we wish to separate. For example, if the relative difference is $\Delta b = 0.1$ then the stiffness should be greater then $k^*_h > 100$. This is marked as **geometry selectivity** in Figure 4.4**a)** and supported by calculations in Figure 4.4**c)**. For example, enantiomeric discrimination of analytes is only possible in the regime of geometric selectivity.

In some applications it is desirable to separate analytes according to receptor composition but not according to the geometry. For instance, if targeting a whole class of analytes with similar receptor composition but varying (or unknown) patterns (*e.g.*, viruses with high mutation rates), geometrical selectivity is unwanted. The optimal regime of matrix stiffness for such applications is the regime of **bond selectivity**: $k^*_h \approx 1$. In this regime imprinting makes a difference, however, the substrate is too flexible to allow for efficient recognition of the geometrical patterns. Hence, MIPs can only discriminate analytes

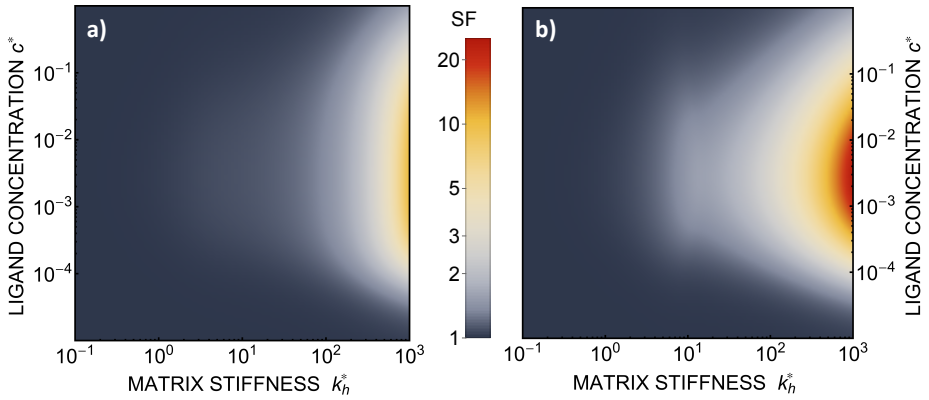


Fig. 4.5 **a) Separation Factor.** The separation factor $SF(b_1, b_2; a)$ for the imprinted template $b_1 = a$ and an analyte with slightly larger inter-receptor distance $b_2 = 1.1a$ as a function of the parameters k_h^* and c^*. **b) Enhancing the selectivity.** The best separation is achieved when the imprinting is slightly mismatched relative to the analyte. The separation factor $SF(b_1, b_2; a^{opt})$ at the optimal value of the imprinted distance a^{opt} as a function of k_h^* and c^*. Parameters are the same as on Figure 4.4**b)**.

with different ligand composition.

In order to test the predictions that follow from Figure 4.4**a)**, we compared them to the analytical results for the divalent binding model (4.25,4.26). To this end, we evaluated the imprinting factor $IF(a)$ for a broad range of ligand concentrations c^*, binding affinities K_D^*, and matrix elasticities k_h^*. Figure 4.4**b)** shows the dependence of $IF(a; a)$ on the matrix stiffness and on the ligand concentration at fixed stoichiometric ratio of ligands and templates in the imprinting phase $c^* = 2C_T^*$. Furthermore Figure 4.4**c)** shows the dependence of IF on the stoichiometry at fixed matrix stiffness, showing that the stoichiometric ratio is close to optimal, the optimal concentrations being $c^{*opt} = 2.4K_D^*$ and $C_T^{*opt} = 1.7K_D^*$. We have assumed that the dissociation constants for individual bonds remain the same in the imprinting and binding phase $\tilde{K}_D^* = K_D^* = 10^{-3}$.

In accordance with our tentative design rules (Figure 4.4), we observe a sharp increase in the imprinting quality in the parameter region where MIPs should function optimally, i.e. for intermediate values of ligand concentrations ($c^* \approx 2K_D^*$) and stiff matrices ($k_h^* \gtrsim 1$).

Within the same parameter space we also computed the separation factor $SF(a, b; a)$ for two slightly different analytes ($b_1 = a$ and $b_2 = 1.1a$) on MIPs imprinted by a template with a separation a between the two receptors (Figure 4.5**a)**). The analyte separation is effective for matrix stiffness $k_h^* \gtrsim 100$. This range can be extended if we compare

analytes that are less similar. Again, the region where analyte separation is most efficient, roughly coincides with the onset of geometrical recognition in Figure 4.4 **a)**.

Somewhat surprisingly, we find that the optimal separation of two analytes, say $b_1 = 1$ and $b_2 = 1.1$, is not best achieved by imprinting with a template identical to the first analyte ($a = b_1$). Rather, the separation factor $SF(b_1, b_2; a)$ can be maximized by designing the cavity with an optimal imprinted distance $a^{opt} < b_1$. This result can be intuitively understood by noting that the binding free energy is approximately a quadratic function of the mismatch close to the minimum. If the imprinting is slightly mismatched, the binding affinity of the chosen analyte is slightly smaller, but at the same time it increases relative to the binding affinity of the other particle resulting in better separation capacity of the MIP. Figure 4.5**b)** displays the separation factor $SF(b_1, b_2; a^{opt})$ for the same analytes as in Figure 4.5**a)** but with the template size a optimized at each point in the parameter space (Appendix A). We can clearly see an increase in the separation capability, however, the qualitative features of the phase diagram in Figure 4.4**a)** remain valid.

4.7 Summary

We have developed a theoretical model of molecular imprinting, which allows us to calculate the performance of imprinted polymers depending on parameters, such as polymer material properties, choice of a template and ligand (functional monomer) concentration. We have explored various factors that determine the quality of molecular imprinting and derived a set of general design principles that can be applied to rationalize specific applications. Our predictions could be studied on a well-defined and tuneable supramolecular system, such as a solution of tetravalent DNA constructs (e.g. Holliday junctions or DNA tetrahedra [118]), which can bind to complementary 'receptor' strands that can be cross-linked into a gel.

The first key observations is that the quality of imprinting depends on the concentration of ligands and templates in the imprinting phase and on the binding strength between them: the optimal imprinting is achieved with a nearly stoichiometric ratio of ligands vs. template receptors, e.g. for tri-valent templates the stoichiometric ratio of templates:ligands should be 1:3. Additionally, initial ligand concentration should be of similar value as the corresponding bond dissociation constant. This provides the optimal tradeoff between, on the one hand efficient cavity formation, and on the other hand selective re-binding of analytes to imprinted cavities.

The second key observation is that stiffer matrices are more selective – suggesting that it should be beneficial to make the gel as rigid as possible. Consequently, polymer gels are not a suitable matrix for efficient molecular imprinting of small molecules: when geometrical recognition is required, stiffer systems such as glassy polymers should be considered. However, the non-specific terms in the free energy – which are not explicitly considered here – will have an opposite effect upon increasing the stiffness of the matrix: slowing down of the kinetics and impeding the particles' access to the cavities. A commonly implemented solution for this problem is grinding the stiff gels in order to expose the imprinted ligands. In this case, much stiffer matrices can be used, however, the procedure inevitably reduces the imprinting quality by grinding-off a fraction of ligands in the cavities. In specific MIP systems, it is likely that the opposing thermodynamic and kinetic trends result in an application-specific optimum gel stiffness.

Moreover, when imprinting soft macromolecules such as proteins or biopolymers onto a hard matrix, the intrinsic softness of the template has a similar effect as a reduced matrix stiffness (within our coarse-grained model a soft template on a hard substrate resembles a hard template on a softer substrate). Extremely large values of the matrix stiffness k_h^* in our model are therefore unlikely to be relevant for experimental realizations. A reasonable choice of the matrix stiffness to design applications seems to be between $10 \lesssim k_h^* \lesssim 100$, i.e. fluctuations of the ligand position relative to the size of the template between 10% and 30%. In such case, for imprinting to work, the strength of the individual ligand-receptor bonds needs to be strong enough. Divalent templates cannot be imprinted effectively unless the bond dissociation constant is $K_D^* \lesssim 10^{-3}$ M. This changes considerably if the templates are multivalent: in the tetravalent case imprinting can be achieved with relatively weak bonds ($K_D^* \sim 0.1$M), while in the hexavalent case the bonds can be extremely weak ($K_D^* \sim 1$M), which is in the realm of hydrogen bonds in aqueous solutions. This suggests that it should be possible to efficiently imprint molecules, such as proteins or drugs, onto polymers in aqueous solutions.

5

Methods

> *A paradox is simply an error out of control; i.e. one that has trapped so many unwary minds that it has gone public, become institutionalized in our literature, and taught as truth*
>
> – E.T. Jaynes

Monte Carlo (MC) methods for numerical sampling of equilibrium distributions are used throughout this work. Methods are described in this chapter for completeness and self-containment of the thesis material. Most MC sampling techniques used are covered in a well-known textbook [119] and will be described only briefly. However, two of the methods employed were customised and optimised for a particular problem: (i) application of the Wang-Landau sampling for the free energy calculations of polymer adsorption, (ii) efficient sampling of valence-limited interactions integrated with translational moves. Therefore, these two methods are described in a greater detail.

All of the simulation algorithms were written by the author in Fortran90 programming language, except for the simulation of particle endocytosis presented in Chapter 3, where the simulation code originally written by Anđjela Šarić in C programming language was used. Python programming language was also used for data analysis. Visual Molecular Dynamics [120] was used for visualisation and production of simulation snapshots. Plots were mostly made with xmGrace and Mathematica.

5.1 Metropolis Monte Carlo

Metropolis Monte Carlo is a Markov chain Monte Carlo scheme for sampling of equilibrium distributions; the appropriate sampling is achieved by accepting or rejecting randomly

proposed moves. Monte Carlo is a very useful method for the study of any equilibrium observables, e.g. thermodynamic quantities like pressure, density etc., of classical many body systems.

The Metropolis algorithm is based on detailed balance, which is a property of equilibrium distributions

$$p_i \Pi_{i\to j} = p_j \Pi_{j\to i}, \tag{5.1}$$

where p_i is the probability to find a system in state i and $\Pi_{i\to j}$ is the probability of transition from state i to state j. Hence, the expected number of transition from state $i \to j$ is the same as the reverse transition $j \to i$.

For a an equilibrium system with a constant number of particles (N), volume (V) and at fixed temperature (T), the probability of a given micro-state i with energy E_i is proportional to the Boltzmann factor $p_i \propto e^{-\beta E_i}$, with $\beta = 1/k_B T$ the inverse temperature and k_B the Boltzmann constant. Using the detailed balance condition (5.1) we find the ratio of probabilities must satisfy

$$\frac{p_i}{p_j} = \frac{\Pi_{j\to i}}{\Pi_{i\to j}} = e^{-\beta(E_i - E_j)} \ . \tag{5.2}$$

The upper equations is satisfied if we use a Metropolis acceptance criterion

$$\Pi_{i\to j} = \min(1, e^{-\beta(E_j - E_i)}) \ . \tag{5.3}$$

The Metropolis choice, however, is not unique; it is the choice of an acceptance rule that results (generally) in the fastest diffusion of the system in the phase space. For particular applications different acceptance schemes become more appropriate. For example, Bennet acceptance ratio can be used for free energy calculations, or the Wang-Landau method described below are both cases of a modified Metropolis scheme that is optimised for the sampling of a particular energy landscape.

In a grand canonical ensemble the entropic (ideal gas) contribution due to having a different number of particles in the system must also be included. The probability to add a particle

$$\Pi_{N\to N+1} \min\left(1, \ \frac{V}{\Lambda^3 (N+1)} e^{\beta\mu} e^{-\beta(E_{N+1} - E_N)}\right) \tag{5.4}$$

and to delete a particle

$$\Pi_{N\to N-1} = \min\left(1, \ \frac{\Lambda^3 N}{V} e^{-\beta\mu} e^{-\beta(E_{N-1} - E_N)}\right), \tag{5.5}$$

where μ is the chemical potential and Λ is the thermal De Broglie wavelength.

5.2 Polymer simulations

For the study of multivalent polymer adsorption we use a coarse-grained polymer model [77]. A polymer is represented with a series of Gaussian soft blobs connected with harmonic springs. Each polymer chain is represented by N_b soft repulsive blobs with radius of gyration r_b that are connected via harmonic springs

$$U_{ch} = 0.534\, k_B T\, (r/r_b - 0.730)^2, \tag{5.6}$$

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

$$U_{bb} = 1.75\, k_B T\, e^{-0.80(r/r_b)^2}, \tag{5.7}$$

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

$$U_{bs} = 3.20\, k_B T\, e^{-4.17(r/r_b - 0.50)}. \tag{5.8}$$

This model accurately describes flexible self-avoiding-walk (SAW) polymers in the scaling regime, we also assume that each individual blob represents a polymer in the scaling regime, hence blobs must be large enough to contain at least a few polymer Kuhn segments. Additionally, polymer must be solvated in a sufficiently good solvent such that the polymer chain can be accurately modelled as a self avoiding walk. On the other hand the model only considers pairwise interactions and does not take into account any 3-body effects. Therefore, the model is appropriate for studying dilute and semi-dilute polymer solutions where the blob density does not exceed 1 blob per blob volume $\rho < 3/(4\pi r_b^3)$ (in this case 3-body effects will be negligible).

A great feature of the soft blob model are the 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. The radius of gyration of such a polymer (isolated in space) is given by

$$R_g = r_b N_b^{0.588}, \tag{5.9}$$

with r_b the blob radius of gyration and N_b number of blobs per polymer.

Polymer conformations are sampled using standard single-blob translational moves. The grand-canonical part of the algorithm (insertion and deletion of polymers in the

system) is employed via Rosenbluth-sampling with configurational bias [119, 121]. The bonding with surface-attached receptor is discussed bellow. Only polymers with no formed bonds can be inserted/deleted.

5.2.1 Valence-limited interactions

It is rather straightforward to include valence limited interaction in a Monte-Carlo scheme, see Refs. [9, 10, 80, 87]. It is much less straightforward to employ with molecular dynamics simulation, which is the reason why we use Monte Carlo as our method of choice.

Each ligand (and each receptor) is represented as a non-interacting point and can be in either of the two states, bound or unbound. First we discuss the application to multivalent polymer adsorption [16].

We simply add ligands to the polymer model discussed above, see Figure 5.1. Assuming we randomly add N_{ligand} ligands per polymer, randomly chosen blobs will carry ligands and others will not. The ligands are assumed to be randomly distributed along the polymer chain and the number of ligands per polymer is Poisson distributed. Blobs can carry more than 1 ligand, where ligand carrier blobs can bind to surface receptors that are within a distance of blob radius r_b. The binding is valence limited, one receptor can bind to only a single ligand and vice versa. Different ligands within a blob are distinguishable. Receptors are modelled as immobile point-like objects that are randomly distributed on the hard surface, there are no interactions between blobs and receptors other then ligand-receptor binding.

5.2.2 Monte Carlo sampling of multivalent polymers

We perform Grand canonical Monte Carlo simulations of polymers using the soft blob model and determine the equilibrium number of polymers bound to the surface. We simulate a box with periodic boundary conditions in x and y directions with the hard surface containing immobile receptors at $z = 0$. The box height was large enough $L_z > 3R_g$ so that bound polymers were never constrained by the box ceiling (results were independent of the box height for $L_z > 3R_g$), typical lateral dimensions were $L_x = L_y \approx 5R_g$. We imagine the simulation box in contact with solution of polymers with a fixed chemical potential μ.

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: polymer insertion/deletion moves, and single blob

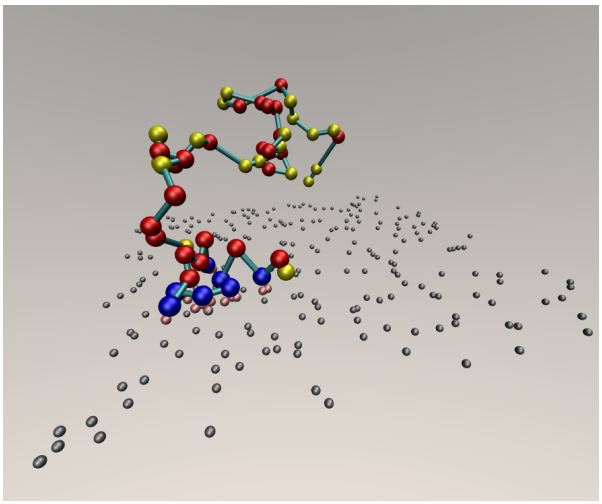


Fig. 5.1 Snapshot of Monte Carlo simulations of a soft blob polymer adsorbing to a surface. Randomly distributed immobile surface receptors (in the xy plane) are represented with grey spheres, the receptors that are currently bound to ligands on a polymer have a pink shade. There is a single polymer in this system, it is represented with blobs of three different colours: (i) normal (inert) blobs are red, (ii) ligand carrying blobs are yellow and (iii) blobs that cary a ligand and are currently bound to a receptor on a surface are blue. The blobs are connected into a chain with turquoise rods.

translational moves integrated with ligand-receptor binding. The polymers are generated or deleted using Rosenbluth sampling with configurational bias [119], only non-bound polymers (without any bonds) can be inserted or deleted. Each simulation started with an empty box and lasted for about $\approx 10^{11}$ MC cycles, where in each cycle we randomly select to either insert or delete a polymer (with probability $p_{ins-del} = 1/(N_b+1)$) or move a single blob (with probability $p_{hop} = N_b/(N_b+1)$). The average acceptance rate for blob displacement was $p_{hop}^{acc} \approx 0.08$ and for polymer insertion/deletion $p_{ins-del}^{acc} \approx 0.01$. The average number of bound polymers (polymers attached to the surface via at least one ligand-receptor bond) was determined by averaging over the second half of the simulation run time.

5.2.3 Optimisation of valence limited polymer sampling

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 blob is chosen uniformly at random and all its ligands are made unbound. The binding partition function which counts all possible ways to make λ bonds is then

calculated

$$q_\lambda = e^{-\lambda\beta\epsilon}\frac{k!n_f!}{(k-\lambda)!\lambda!(n_f-\lambda)!}, \tag{5.10}$$

we arrive at the above expression by noting that we have n_f free receptors within reach (i.e. within blob radius r_b) of the k ligands that the blob carries, we then consider all possible ways of binding then together with λ bonds. The total binding partition function considers all possible numbers of bonds

$$q_b = \sum_{\lambda=0}^{\min(k,n_f)} q_\lambda. \tag{5.11}$$

The above expression is essentially the same as the exact binding partition function used in analytical theory. 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

$$p_{o\to n} = \min\left[1, \frac{q_b^n}{q_b^o}\, e^{\beta(U_o - U_n)}\right], \tag{5.12}$$

where U_o and U_n are the old and new potential energies of the system determined by Eqs. (5.6, 5.7, 5.8).

Regardless of whether the move was accepted or not the blob of interest still has all ligands unbound. We now randomly choose how many bonds to form, the probability to form λ bonds is

$$p_\lambda = \frac{q_\lambda}{q_b} \tag{5.13}$$

where we use the partition functions we calculated for the new (old) blob position (Eq. 5.10,5.11) if the translation move was accepted (not-accepted). Say we chose to form λ^* bonds, we now randomly chose λ^* ligands and λ^* receptor and randomly bind them together. The probability to choose a particular state (bonding arrangement) with λ^* bonds is then

$$p_s = \frac{e^{-\beta\lambda^*\epsilon}}{q_b} \tag{5.14}$$

This algorithm takes advantage of the fact that we can explicitly calculate the partition functions of a blob and it provides a substantial speed up compared to just using the standard translational and single ligand-receptor bind/unbind MC moves. We now show that the algorithm obeys detailed balance. A particular state is determined by the vector positions of all blobs in the system and a particular ligand-receptor bonding arrangement. The equilibrium probability for the system to be in state (1), with potential energy U_1

and number of ligand-receptor bonds λ_1, is proportional to the total (free) energy of that state

$$p_1 \propto e^{-\beta(U_1+\lambda_1\epsilon)} \tag{5.15}$$

The transition probability going from state (1) to state (2) (with potential energy U_2 and λ_2 bonds) is

$$\Pi_{1\to 2} = \min\left[1, \frac{q_{b2}}{q_{b1}}\, e^{\beta(U_1-U_2)}\right] \times \frac{e^{-\beta\lambda_2\epsilon}}{q_{b2}} \tag{5.16}$$

where the first factor corresponds to the probability of moving the blob to a new position (Eq. 5.12) which is multiplied by the probability that we choose a particular bond arrangement at the new position (Eq. 5.14). The reverse transition probability is

$$\Pi_{2\to 1} = \min\left[1, \frac{q_{b1}}{q_{b2}}\, e^{\beta(U_2-U_1)}\right] \times \frac{e^{-\beta\lambda_1\epsilon}}{q_{b1}} \tag{5.17}$$

and the ratio between the two is obtained by noting that the value of $\frac{q_{b2}}{q_{b1}}\, e^{\beta(U_1-U_2)}$ can be either greater or smaller then 1, both cases give the same result

$$\frac{\Pi_{1\to 2}}{\Pi_{2\to 1}} = \frac{e^{-\beta(U_2+\lambda_2\epsilon)}}{e^{-\beta(U_1+\lambda_1\epsilon)}} = \frac{p_2}{p_1} \tag{5.18}$$

which proves that the algorithm obeys detailed balance. The method was also tested to provide the same equilibrium result when compared to standard MC moves.

5.3 Wang-Landau technique

The Wang-Landau algorithm is a very convenient method to calculate the free energy of a simulated system. The method was initially developed for the sampling of the density of states as a function of the energy [79]. However, the method is much more general and can be extended to the sampling of any order parameter.

The Wang-Landau method works by modifying the Metropolis acceptance scheme; introducing a biasing function which is dynamically built and updated during the simulation run in such a way to achieve a uniform sampling of states. Once the uniform sampling in the chosen order parameter is achieved, the biasing function will asymptotically approximate the free energy as a function of that order parameter. This method is very useful for efficient calculations of the free energy, as long as the order parameter is reasonable. By reasonable we mean that the system can diffuse and sample that order parameter.

Consider an ensemble of states with Γ denoting a particular micro-state (configuration) and X is any function of Γ. X could be for example a number density, box volume, position of a designated particle, bond order parameter, or any thermodynamic variable or order parameter. In our application to polymer adsorption X will stand for the number of formed bonds between ligands on the polymer and surface attached receptors. The probability to find our system in a particular configuration Γ given a particular value of X is $p(\Gamma|X)$. The equilibrium probability distribution of X can be obtained by summing the micro-states that yield a particular X

$$p(X) = \sum_{\Gamma} p(\Gamma|X). \tag{5.19}$$

We now introduce a biased probability distribution

$$p^B(\Gamma|X) = \frac{p(\Gamma|X)}{p(X)} \ , \tag{5.20}$$

which results in a uniform probability distribution $p^B(X) = \sum_\Gamma p^B(\Gamma, X) = 1$. We can perform Monte Carlo calculations in the biased ensemble. Detailed balance condition for Monte Carlo (5.1) implies that each pair of states Γ_i and Γ_j must satisfy

$$p^B(\Gamma_i|X_i)\ \Pi^B_{i\to j} = p^B(\Gamma_j|X_j)\ \Pi^B_{j\to i} \ , \tag{5.21}$$

where Π^B is the acceptance probability that gives a biased (uniform) probability distribution of the observable X. We now apply the Metropolis acceptance criterion in the biased ensemble

$$\Pi^B_{i\to j} = \min\left[1, \frac{p^B(\Gamma_j|X_j)}{p^B(\Gamma_i|X_i)}\right] = \min\left[1, \frac{p(\Gamma_j|X_j)}{p(\Gamma_i|X_i)}\frac{p(X_i)}{p(X_j)}\right]. \tag{5.22}$$

Defining the free energy $F(X)$ via the partition function (summing the probabilities of micro-states), the probability of observing X is

$$p(X) \propto \sum_{\Gamma|X} e^{-\beta E_\Gamma} = e^{-\beta F(X)} \ , \tag{5.23}$$

we can write $\frac{p(X_i)}{p(X_j)} = e^{-\beta(F(X_i)-F(X_j))}$ and the acceptance criterion in the biased (uniform) ensemble is

$$\Pi^B_{i\to j} = \min\left[1, \frac{p(\Gamma_j|X_j)}{p(\Gamma_i|X_i)} e^{\beta(F(X_j)-F(X_i))}\right]. \tag{5.24}$$

The normal (unbiased) MC acceptance probability (using Eq. 5.3) would be $\Pi_{i\to j} = \min\left[1, \frac{p(\Gamma_j|X_j)}{p(\Gamma_i|X_i)}\right]$. Hence in order to get a uniform distribution in X we need to multiply the normal (unbiased) MC acceptance probability with the exponential of the free energy difference between the old and new state. This is a general result and therefore it holds for any equilibrium ensemble (e.g. canonical, grand canonical, isothermal isobaric).

We do not know the free energy $F(X)$ of the system in advance, this is precisely the quantity that we are trying to calculate. What the above equations tell us is that if we find a biasing function that will give us a uniform sampling, that biasing function must be $-F(X)$. Wang-Landau algorithm iteratively searches for the biasing function resulting in a uniform sampling.

5.3.1 The iterative algorithm

The practical implementation of the algorithm is obtained by introducing a biasing function in the form $e^{\Psi(X)}$, where X is a discrete variable and X_i denotes a value of variable X with the system in state i. We then use the standard Metropolis acceptance scheme, the acceptance probability from state i to j is

$$\Pi^B_{i\to j} = \min\left[1, \frac{p(\Gamma_j|X_j)}{p(\Gamma_i|X_i)} e^{\Psi(X_j)-\Psi(X_i)}\right] . \tag{5.25}$$

The unbiased ($\Psi(X) = 0$) MC acceptance ratio was $\Pi_{i\to j} = \min\left[1, \frac{p(\Gamma_j|X_j)}{p(\Gamma_i|X_i)}\right]$. In general, our best initial guess at the initial biasing function is $\Psi(X) = 0$, however, any prior knowledge about the free-energy of a particular system can be included. The system is forced into uniform sampling by updating the biasing function on every step in such a way as to make the current state less probable

$$\Psi^n(X_i) = \Psi^o(X_i) - f, \tag{5.26}$$

where f is a parameter specifying how much the biasing function is adjusted on a single step.

We can imagine the workings of the method with a simple 1D picture. We walk on a rough landscape in direction X. On every step we lay a brick with height f and we prefer to walk down rather than up. Repeating this process we would first sample the valleys and lowlands which would be built up by layers of bricks. Once we cover the highest peak of our landscape, we have constructed a brick wall that is roughly flat on the top. Now, simply counting the number of bricks we have used at each point X will tell us the original height of the landscape. The only difference to a high dimensional

picture, is that we need a function that projects every point to some value of X. We walk on a high dimensional landscape, but lay bricks only in X.

The algorithm breaks detailed balance on every step, but at this stage we only wish to build up a biasing function, not simulate the equilibrium distribution of states in the system. The value of f chosen (the brick height) determines the accuracy of the biasing function in k_BT as $\epsilon_i/k_BT \approx \mathcal{O}(fn_{i\rightarrow})$, where k_B is the Boltzmann constant and $n_{i\rightarrow}$ is average acceptance ratio from state i to any other state. The smaller the value of f the better the accuracy, but on the other hand a large f quickly builds up a biasing function.

We usually initialise $f = 1$, then we continuously check the histogram of all visited states $H(X)$, once this histogram is sufficiently flat (say $\frac{\max[H(X)]}{\min[H(X)]} \leq 1.5$), we assign a smaller value to f (e.g. the new $f_n = f_o/2$ is the half of the old f_o) and start the counting of the histogram again, in this way we gradually improve accuracy on $\Psi(X)$. We continue with this scheme until the desired accuracy is reached, in our case this was $f_0 = 10^{-6}$.

Once we obtain a flat histogram $H(X)$ with the desired accuracy the algorithm terminates. At this point we perform a long simulation with a static final value of the biasing function Ψ_{fin}, this scheme obeys detailed balance. We record the final histogram $H_{fin}(X)$ of this simulation which should yield a nearly flat distribution in X. If the final distribution is not flat: either the Ψ_{fin} has not converged, in which case we need to repeat the Wang-Landau scheme, or the final simulation was not long enough and the final sampling was not sufficient. The expectation value for the final histogram is

$$H_{fin}(X) \propto p_{fin}(X) \propto e^{-\beta F(X)} e^{\Psi_{fin}(X)}, \tag{5.27}$$

Hence, we obtain the free energy of the system up to a constant

$$\beta F(X) = \Psi^f(X) - \ln\left(H^f(X)\right) + C\ . \tag{5.28}$$

We applied this algorithm to calculate the free energy of a multivalent polymer as function of the number of formed bonds λ presented in the 2^{nd} Chapter.

5.4 Molecular imprinting simulations

We performed equilibrium Monte Carlo simulations using the “spring" model described in the Chapter 4. We simulate both the imprinting process, and rebinding of analytes into formed mips. The simulated system was a cubic box (volume V) with periodic boundary conditions. In the imprinting phase we perform canonical simulations with fixed number of point-like ligands and hard-sphere template particles. Both template

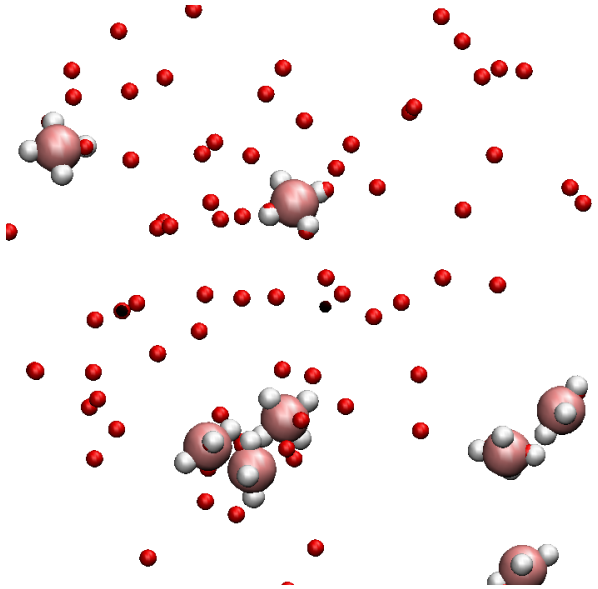


Fig. 5.2 Simulation snapshot. Ligand anchors are represented with red blobs, particles are the big spheres and binding sites on particles are shown as white blobs. In this particular system the ligand anchors are randomly distributed in space and all of the ligands are complementary in binding to all of the binding sites.

particles and ligands are allowed to diffuse around and ligands can bind to the receptors on the template. Ligands and binding sites can be of different types and the interaction matrix $\Delta\tilde{\mathbf{G}}$ defines the hybridization free energies between different types of ligands and binding sites. After this system has equilibrated, we fix the positions of ligands, which we call ligand anchors, and remove the particles, thus an imprinted configuration was formed.

In the analyte binding phase we do not simulate the positions of ligands explicitly, in a given binding event we treat the loss of configurational entropy implicitly using Eq. (A15). We perform Grand-canonical MC simulations where hard-sphere analytes, with specific binding sites on their surface, are inserted into the system from the reservoir with a fixed chemical potential μ (see Figure 5.2 and 4.3 for a depictive simulation snapshot). Analytes are moved around using standard displacement MC moves. We then determine the average number of bound analytes (analytes attached to the ligands with at least one bond) in the system. When simulating the binding of analytes into non-imprinted polymers (NIPs) the ligands are randomly distributed.

In any given Monte Carlo (MC) step a random choice is made between: (i) particle translation and rotation, (ii) particle insertion/deletion move or (iii) binding/unbinding event between ligands and binding sites. The probabilities to chose particular moves

are: 0.4 for translation/rotation, 0.1 for insertion/deletion and 0.5 for binding/unbinding. These numbers were found to give a reasonably good overall performance of Monte Carlo sampling, their exact values are unimportant as we perform equilibrium simulations. The particles are moved around using standard translational moves [119]. Rotations of individual particles are implemented using quaternion algebra, Marsaglia's method [122] is used to generate random quaternions for generating candidate MC moves. The translation and rotation moves are always performed jointly, a translation/rotation is considered a single MC move. Bonds between ligands and binding sites on particles can be made or broken using configurational bias Monte Carlo; a random binding site is chosen, then all possible ligands within a cut-off distance are considered and their Boltzmann weights computed. A binding/unbinding event is chosen with a probability proportional to the weight.

When calculating the free energy of binding we simulated a single cavity site with a restriction that at most a single particle can be present in the system. We calculate the average occupancy of the cavity f_b (the fraction of time the particle is bound in the cavity with at least one bond) and obtain the binding free energy by inverting the Langmuir adsorption isotherm

$$e^{-\beta F} = K_A \rho_0 = (e^{-\beta\mu} + V\rho_0 N_A)\frac{f_b}{1-f_b}. \tag{5.29}$$

In our simulations, the chemical potential μ was chosen such to obtain good statistics. $V = (10\sigma)^3$ is the volume of the simulation box (σ is the particle diameter), the term $V\rho_0 N_A$ in the above equation takes into account the probability of finding a non-bound particle in the system.

5.4.1 Configurational bias

Here we expand on the configurational bias technique. To simulate binding and unbinding events we could use simple Metropolis acceptance criterium, however, for strong bonds, the acceptance ratio would be quite low.

The configurational bias method [119, 121] overcomes this. The idea of the configurational bias method is that for a randomly chosen subset of the system, we consider all possible MC moves and compute all corresponding Boltzmann factors associated with these moves. We then choose a MC move with the probability proportional to this Boltzmann factor. By construction the method is rejection free. Say there are x possible states the subsystem can evolve to, we then we have x possible MC moves, the

probability of a particular move i is

$$p_i = \frac{e^{-\beta \Delta E_i}}{\sum_i e^{-\beta \Delta E_i}}, \tag{5.30}$$

where ΔE_i is the energy difference associated with an i-th move. The drawback of this method is that we have to compute all Boltzmann factors of all possible MC moves. Therefore, the method works well when the set of possible moves is not too large. In our system this was the case, the number of possible binding partners was usually at most 4. For the same reason there was no need to implement more sophisticated MC schemes such as "waste recycling" [123].

For our valence limited interaction sampling, a chosen binding site (receptor) can be bound to any of the nearby ligands or it can be free. We can only consider ligands that are not already bound to another binding site. So we consider all free ligands within some cutoff distance of the binding site. The cutoff distance is chosen such that the total binding energy to any ligand outside of it would be grater than $20k_BT$, and therefore the probability for any such bond to form is negligible. We then compute the probabilities using Eq. (5.30) for all possible binding re-arrangement of a chosen binding site. Irrespective of the initial arrangement, the new state of the binding site is chosen proportionally to the computed probability.

In order to correctly describe the equilibrium properties, the detailed balance needs to be preserved within each MC move. If at some point a chosen binding site was bound to a ligand from outside of the distance cutoff, we need to leave them bound and continue with MC sampling, otherwise the detailed balance would be violated. The reason for the violation being that there is no reverse move since for binding moves we only consider binding sites within a distance cutoff.

References

[1] Richard J. Goldberg. A theory of antibody—antigen reactions. i. theory for reactions of multivalent antigen with bivalent and univalent antibody. *Journal of the American Chemical Society*, 74(22):5715–5725, 1952.

[2] G. I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978.

[3] Charles DeLisi. The biophysics of ligand–receptor interactions. *Quarterly Reviews of Biophysics*, 13:201–230, 1980.

[4] Charles DeLisi. Theory of clustering of cell surface receptors by ligands of arbitrary valence: dependence of dose response patterns on a coarse cluster characteristic. *Mathematical Biosciences*, 52(3):159 – 184, 1980.

[5] Pierre Bongrand. Ligand-receptor interactions. *Reports on Progress in Physics*, 62(6):921, 1999.

[6] P. H. von Hippel and O. G. Berg. On the specificity of dna-protein interactions. *Proceedings of the National Academy of Sciences*, 83(6):1608–1612, 1986.

[7] Leonid Mirny, Michael Slutsky, Zeba Wunderlich, Anahita Tafvizi, Jason Leith, and Andrej Kosmrlj. How a protein searches for its site on dna: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*, 42(43):434013, 2009.

[8] John SantaLucia. A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences*, 95(4):1460–1465, 1998.

[9] Patrick Varilly, Stefano Angioletti-Uberti, Bortolo M. Mognetti, and Daan Frenkel. A general theory of dna-mediated and other valence-limited colloidal interactions. *The Journal of Chemical Physics*, 137(9), 2012.

[10] Francisco J Martinez-Veracoechea and Daan Frenkel. Designing super selectivity in multivalent nano-particle binding. *Proceedings of the National Academy of Sciences of the United States of America*, 108(27):10963–10968, 2011.

[11] Rémi Dreyfus, Mirjam E. Leunissen, Roujie Sha, Alexei V. Tkachenko, Nadrian C. Seeman, David J. Pine, and Paul M. Chaikin. Simple quantitative model for the reversible association of dna coated colloids. *Phys. Rev. Lett.*, 102:048301, 2009.

[12] Kathleen H. Mortell, Ross V. Weatherman, and Laura L. Kiessling. Recognition specificity of neoglycopolymers prepared by ring-opening metathesis polymerization. *Journal of the American Chemical Society*, 118(9):2297–2298, 1996.

[13] Coby B. Carlson, Patricia Mowery, Robert M. Owen, Emily C. Dykhuizen, and Laura L. Kiessling. Selective tumor cell targeting using low-affinity, multivalent interactions. *ACS Chemical Biology*, 2(2):119–127, 2007. PMID: 17291050.

[14] Laura L. Kiessling and Joseph C. Grim. Glycopolymer probes of signal transduction. *Chem. Soc. Rev.*, 42:4476–4491, 2013.

[15] Galina V. Dubacheva, Tine Curk, Bortolo M. Mognetti, Rachel Auzély-Velty, Daan Frenkel, and Ralf P. Richter. Superselective targeting using multivalent polymers. *Journal of the American Chemical Society*, 136(5):1722–1725, 2014. PMID: 24400591.

[16] Galina V. Dubacheva, Tine Curk, Rachel Auzély-Velty, Daan Frenkel, and Ralf P. Richter. Designing multivalent probes for tunable superselective targeting. *Proceedings of the National Academy of Sciences*, 112(18):5579–5584, 2015.

[17] Renato Ribeiro-Viana, Macarena Sánchez-Navarro, Joanna Luczkowiak, Julia R. Koeppe, Rafael Delgado, Javier Rojo, and Benjamin G. Davis. Virus-like glycodendrinanoparticles displaying quasi-equivalent nested polyvalency upon glycoprotein platforms potently block viral infection. *Nat Commun*, 3:1303, 2012.

[18] Huafeng Xu and David E. Shaw. A simple model of multivalent adhesion and its application to influenza infection. *Biophysical Journal*, 110(1):218 – 233, 2016.

[19] Christian Sieben, Christian Kappel, Rong Zhu, Anna Wozniak, Christian Rankl, Peter Hinterdorfer, Helmut Grubmüller, and Andreas Herrmann. Influenza virus binds its host cell using multiple dynamic interactions. *Proceedings of the National Academy of Sciences*, 109(34):13626–13631, 2012.

[20] Patricia G Spear. Herpes simplex virus: receptors and ligands for cell entry. *Cellular microbiology*, 6(5):401–410, 2004.

[21] Joe Grove and Mark Marsh. The cell biology of receptor-mediated virus entry. *The Journal of Cell Biology*, 195(7):1071–1082, 2011.

[22] Archibald Vivian Hill. Proceedings of the physiological society: January 22, 1910. *The Journal of Physiology*, 40(suppl):i–vii, 1910.

[23] Yuhai Tu. The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by maxwell's demons. *Proceedings of the National Academy of Sciences*, 105(33):11737–11741, 2008.

[24] Kim Sneppen, Mille A Micheelsen, and Ian B Dodd. Ultrasensitive gene regulation by positive feedback loops in nucleosome modification. *Molecular Systems Biology*, 4(1), 2008.

[25] James E. Ferrell and Sang Hoon Ha. Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends in Biochemical Sciences*, 39(10):496–503, 2014.

[26] James E. Ferrell and Sang Hoon Ha. Ultrasensitivity part ii: multisite phosphorylation, stoichiometric inhibitors, and positive feedback. *Trends in Biochemical Sciences*, 39(11):556 – 569, 2014.

[27] James E. Ferrell and Sang Hoon Ha. Ultrasensitivity part III: cascades, bistable switches, and oscillators. *Trends in Biochemical Sciences*, 39(12):612–618, 2014.

[28] Simona Mura, Julien Nicolas, and Patrick Couvreur. Stimuli-responsive nanocarriers for drug delivery. *Nat Mater*, 12(11):991–1003, 2013.

[29] Stefan Wilhelm, Anthony J. Tavares, Qin Dai, Seiichi Ohta, Julie Audet, Harold F. Dvorak, and Warren C. W. Chan. Analysis of nanoparticle delivery to tumours. *Nature Reviews Materials*, 1:16014 EP, 2016.

[30] Michael R. Caplan and Elena V. Rosca. Targeting drugs to combinations of receptors: A modeling analysis of potential specificity. *Annals of Biomedical Engineering*, 33(8):1113–1124, 2005.

[31] M. K. Robinson, K. M. Hodge, E. Horak, A. L. Sundberg, M. Russeva, C. C. Shaller, M. von Mehren, I. Shchaveleva, H. H. Simmons, J. D. Marks, and G. P. Adams. 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, 2008.

[32] Borje Sellergren, Bjorn Ekberg, and Klaus Mosbach. Molecular imprinting of amino acid derivatives in macro-porous polymers: demonstration of substrate-and enantio-selectivity by chromatographic resolution of racemic mixtures of amino acid derivatives. *Journal of Chromatography*, 347:1–10, 1985.

[33] L. I. Andersson and K. Mosbach. Enantiomeric resolution on molecularly imprinted polymers prepared with only non-covalent and non-ionic interactions. *Journal of Chromatography*, 516(2):313–322, 1990.

[34] I.R. Dunkin, J. Lenfeld, and D.C. Sherrington. Molecular imprinting of flat polycondensed aromatic molecules in macroporous polymers. *Polymer*, 34(1):77–84, 1993.

[35] Ian A. Nicholls, Olof Ramstrm, and Klaus Mosbach. Insights into the role of the hydrogen bond and hydrophobic effect on recognition in molecularly imprinted polymer synthetic peptide receptor mimics. *Journal of Chromatography A*, 691:349–353, 1995.

[36] Pradeep K. Dhal and Frances H. Arnold. Template-Mediated Synthesis of Metal-Complexing Polymen for Molecular Recognition Cyclization Reactions of Chromium Dienylcarbene. *Journal of the American Chemical Society*, (6):7417–7418, 1991.

[37] M.V. Polyakov. Adsorption properties and structure of silica gel. *Zhurnal fizicheskoi khimii*, (2):799–804, 1931.

[38] H. S. Andersson, A. C. Koch-Schmidt, S. Ohlson, and K. Mosbach. Study of the nature of recognition in molecularly imprinted polymers. *Journal of molecular recognition : JMR*, 9(5-6):675–82, 1996.

[39] Daniela Todorova-Balvay, Ivanka Stoilova, Stoyanka Gargova, and Mookambeswaran a Vijayalakshmi. An efficient two step purification and molecular characterization of beta-galactosidases from Aspergillus oryzae. *Journal of molecular recognition*, 19(4):299–304, 2006.

[40] Lingxin Chen, Shoufang Xu, and Jinhua Li. Recent advances in molecular imprinting technology: current status, challenges and highlighted applications. *Chem. Soc. Rev.*, 40:2922–2942, 2011.

[41] Cameron Alexander, Håkan S. Andersson, Lars I. Andersson, Richard J. Ansell, Nicole Kirsch, Ian A. Nicholls, John O'Mahony, and Michael J. Whitcombe. Molecular imprinting science and technology: A survey of the literature for the years up to and including 2003. *Journal of Molecular Recognition*, 19(2):106–180, 2006.

[42] Chiyang He, Yuanyuan Long, Junlan Pan, Kean Li, and Feng Liu. Application of molecularly imprinted polymers to solid-phase extraction of analytes from real samples. *Journal of Biochemical and Biophysical Methods*, 70(2):133 – 150, 2007.

[43] M. Kempe and K. Mosbach. Molecular imprinting used for chiral separations. *Journal of Chromatography A*, 694(1):3–13, 1995.

[44] Günter Wulff. Enzyme-like catalysis by molecularly imprinted polymers. *Chemical Reviews*, 102(1):1–27, 2002.

[45] Nicholas W. Turner, Christopher W. Jeans, Keith R. Brain, Christopher J. Allender, Vladimir Hlady, and David W. Britt. From 3D to 2D: A review of the molecular imprinting of proteins. *Biotechnology Progress*, 22(6):1474–1489, 2006.

[46] Carmen Alvarez-Lorenzo and Angel Concheiro. Molecularly imprinted polymers for drug delivery. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 804(1):231–245, 2004.

[47] Håkan S. Andersson, Jesper G. Karlsson, Sergey A. Piletsky, Ann Christin Koch-Schmidt, Klaus Mosbach, and Ian A. Nicholls. Study of the nature of recognition in molecularly imprinted polymers, II [1]: Influence of monomer-template ratio and sample load on retention and selectivity. *Journal of Chromatography A*, 848(1-2):39–49, 1999.

[48] George Vlatakis, Lars I. Andersson, Ralf Müller, and Klaus Mosbach. Drug assay using antibody mimics made by molecular imprinting. *Nature*, 361:645–647, 1993.

[49] H. Shi, W. B. Tsai, M. D. Garrison, S. Ferrari, and B. D. Ratner. Template-imprinted nanostructured surfaces for protein recognition. *Nature*, 398(6728):593–597, 1999.

[50] Romana Schirhagl, Eric W. Hall, Ingo Fuereder, and Richard N. Zare. Separation of bacteria with imprinted polymeric films. *The Analyst*, 137(6):1495, 2012.

[51] Kangning Ren and Richard N. Zare. Chemical Recognition in Cell- Imprinted Polymers. *ACS Nano*, 6(5):4314–4318, 2012.

[52] Simcha Srebnik and Ovadia Lev. Toward establishing criteria for polymer imprinting using mean-field theory. *The Journal of Chemical Physics*, 116(24):10967–10972, 2002.

[53] Simcha Srebnik. Theoretical Investigation of the Imprinting Efficiency of Molecularly Imprinted Polymers. *Chemistry of Materials*, 16(5):883–888, 2004.

[54] Lev Sarkisov and Paul R. Van Tassel. Replica Ornstein-Zernike theory of adsorption in a templated porous material: Interaction site systems. *Journal of Chemical Physics*, 123(16):164706, 2005.

[55] Lev Sarkisov and Paul R. Van Tassel. Integral equation theory of adsorption in templated materials: Iinfluence of molecular attraction. *Journal of Physical Chemistry C*, 111(43):15726–15735, 2007.

[56] Kenji Ito, Jeffrey Chuang, Carmen Alvarez-Lorenzo, Tsuyoshi Watanabe, Nozomi Ando, and Alexander Yu. Grosberg. Multiple point adsorption in a heteropolymer gel and the tanaka approach to imprinting: experiment and theory. *Progress in Polymer Science*, 28(10):1489 – 1515, 2003.

[57] Takashi Enoki, Kazunori Tanaka, Tsuyoshi Watanabe, Taro Oya, Takaharu Sakiyama, Yukikazu Takeoka, Kenji Ito, Guoqiang Wang, Masahiko Annaka, Kazuhiro Hara, Rose Du, Jeffrey Chuang, Kevin Wasserman, Alexander Yu. Grosberg, Satoru Masamune, and Toyoichi Tanaka. Frustrations in polymer conformation in gels and their minimization through molecular imprinting. *Phys. Rev. Lett.*, 85:5000–5003, 2000.

[58] Hyunjung Kim and David a. Spivak. New insight into modeling non-covalently imprinted polymers. *Journal of the American Chemical Society*, 125(37):11269–11275, 2003.

[59] Carmelo Herdes and Lev Sarkisov. Computer simulation of volatile organic compound adsorption in atomistic models of molecularly imprinted polymers. *Langmuir*, 25(9):5352–5359, 2009.

[60] E. M. A. Dourado and Lev Sarkisov. Emergence of molecular recognition phenomena in a simple model of imprinted porous materials. *Journal of Chemical Physics*, 130(21):214701, 2009.

[61] Eduardo M. A. Dourado, Carmelo Herdes, Paul R. van Tassel, and Lev Sarkisov. Molecular recognition effects in atomistic models of imprinted polymers. *International Journal of Molecular Sciences*, 12(8):4781–4804, 2011.

[62] Liora Levi, Vladimir Raim, and Simcha Srebnik. A brief review of coarse-grained and other computational studies of molecularly imprinted polymers. *Journal of Molecular Recognition*, 24(6):883–891, 2011.

[63] Ian A. Nicholls, Håkan S. Andersson, Christy Charlton, Henning Henschel, Björn C. G. Karlsson, Jesper G. Karlsson, John O'Mahony, Annika M. Rosengren, K. Johan Rosengren, and Susanne Wikman. Theoretical and computational strategies for rational molecularly imprinted polymer design. *Biosensors and Bioelectronics*, 25(3):543–552, 2009.

[64] Xiangyang Wu, William R. Carroll, and Ken D. Shimizu. Stochastic lattice model simulations of molecularly imprinted polymers. *Chemistry of Materials*, 20(13):4335–4346, 2008.

[65] Estroff Lara A. Krishnamurthy Vijay M. and Whitesides George M. *Multivalency in Ligand Design*, pages 11–53. Wiley-VCH Verlag GmbH and Co. KGaA, 2006.

[66] Gianfranco Ercolani and Luca Schiaffino. Allosteric, chelate, and interannular cooperativity: A mise au point. *Angewandte Chemie International Edition*, 50(8):1762–1768, 2011.

[67] Christopher A. Hunter and Harry L. Anderson. What is cooperativity? *Angewandte Chemie International Edition*, 48(41):7488–7499, 2009.

[68] Jurriaan Huskens, Alart Mulder, Tommaso Auletta, Christian A. Nijhuis, Manon J. W. Ludden, and David N. Reinhoudt. A model for describing the thermodynamics of multivalent host-guest interactions at interfaces. *Journal of the American Chemical Society*, 126(21):6784–6797, 2004. PMID: 15161307.

[69] Alart Mulder, Jurriaan Huskens, and David N. Reinhoudt. Multivalency in supramolecular chemistry and nanofabrication. *Org. Biomol. Chem.*, 2:3409–3424, 2004.

[70] Vijay M. Krishnamurthy, Vincent Semetey, Paul J. Bracher, Nan Shen, and George M. Whitesides. Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *Journal of the American Chemical Society*, 129(5):1312–1320, 2007. PMID: 17263415.

[71] Pavel I. Kitov and David R. Bundle. On the nature of the multivalency effect: A thermodynamic model. *Journal of the American Chemical Society*, 125(52):16271–16284, 2003. PMID: 14692768.

[72] Nathan W Schmidt, Fan Jin, Roberto Lande, Tine Curk, Wujing Xian, Calvin Lee, Loredana Frasca, Daan Frenkel, Jure Dobnikar, Michel Gilliet, and Gerard C L Wong. Liquid-crystalline ordering of antimicrobial peptide-DNA complexes controls TLR9 activation. *Nat Mater*, 14(7):696–700, 2015.

[73] Pierre-Gilles de Gennes. *Scaling concepts in polymer physics*. Cornell University Press, 1979.

[74] A. A. Louis, P. G. Bolhuis, J. P. Hansen, and E. J. Meijer. Can polymer coils be modeled as "soft colloids"? *Phys. Rev. Lett.*, 85:2522–2525, 2000.

[75] Francisco J. Martinez-Veracoechea, Behnaz Bozorgui, and Daan Frenkel. Anomalous phase behavior of liquid-vapor phase transition in binary mixtures of dna-coated particles. *Soft Matter*, 6:6136–6145, 2010.

[76] Wilfrid Norman Bailey. *Generalized hypergeometric series*. Cambridge tracts in mathematics and mathematical physics. Cambridge University Press, 1935.

[77] Carlo Pierleoni, Barbara Capone, and Jean-Pierre Hansen. A soft effective segment representation of semidilute polymer solutions. *The Journal of Chemical Physics*, 127(17), 2007.

[78] Ivan Coluzza, Barbara Capone, and Jean-Pierre Hansen. Rescaling of structural length scales for "soft effective segment" representations of polymers in good solvent. *Soft Matter*, 7:5255–5259, 2011.

[79] Fugao Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86:2050–2053, 2001.

[80] Robin De Gernier, Tine Curk, Galina V. Dubacheva, Ralf P. Richter, and Bortolo M. Mognetti. A new configurational bias scheme for sampling supramolecular structures A new configurational bias scheme for sampling supramolecular structures. *Journal of Chemical Physics*, 244909(2014), 2015.

[81] Lorenzo Albertazzi, Francisco J. Martinez-Veracoechea, Christianus M. A. Leenders, Ilja K. Voets, Daan Frenkel, and E. W. Meijer. Spatiotemporal control and superselectivity in supramolecular polymers using multivalency. *Proceedings of the National Academy of Sciences*, 110(30):12203–12208, 2013.

[82] W. Benjamin Rogers and John C. Crocker. Direct measurements of dna-mediated colloidal interactions and their quantitative modeling. *Proceedings of the National Academy of Sciences*, 108(38):15687–15692, 2011.

[83] Stef A. J. van der Meulen and Mirjam E. Leunissen. Solid colloids with surface-mobile dna linkers. *Journal of the American Chemical Society*, 135(40):15129–15134, 2013. PMID: 24040916.

[84] Yu Wang, Yufeng Wang, Xiaolong Zheng, Étienne Ducrot, Jeremy S. Yodh, Marcus Weck, and David J. Pine. Crystallization of dna-coated colloids. *Nature Communications*, 6:7253 EP –, 06 2015.

[85] Huajian Gao, Wendong Shi, and Lambert B. Freund. Mechanics of receptor-mediated endocytosis. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9469–9474, 2005.

[86] Nicholas B. Tito and Daan Frenkel. Optimizing the selectivity of surface-adsorbing multivalent polymers. *Macromolecules*, 47(21):7496–7509, 2014. PMID: 25400296.

[87] Stefano Angioletti-Uberti, Patrick Varilly, Bortolo M. Mognetti, Alexei V. Tkachenko, and Daan Frenkel. Communication: A simple analytical formula for the free energy of ligand–receptor-mediated interactions. *The Journal of Chemical Physics*, 138(2), 2013.

[88] Stefano Angioletti-Uberti, Patrick Varilly, Bortolo M. Mognetti, and Daan Frenkel. Mobile linkers on dna-coated colloids: Valency without patches. *Phys. Rev. Lett.*, 113:128303, 2014.

[89] Lorenzo Di Michele, Stephan J. Bachmann, Lucia Parolini, and Bortolo M. Mognetti. Communication: Free energy of ligand-receptor systems forming multimeric complexes. *The Journal of Chemical Physics*, 144(16), 2016.

[90] Nicholas B. Tito, Stefano Angioletti-Uberti, and Daan Frenkel. Communication: Simple approach for calculating the binding free energy of a multivalent particle. *The Journal of Chemical Physics*, 144(16), 2016.

[91] Shihu Wang and Elena E. Dormidontova. Selectivity of ligand-receptor interactions between nanoparticle and cell surfaces. *Phys. Rev. Lett.*, 109:238102, 2012.

[92] T. A. J. Duke and D. Bray. Heightened sensitivity of a lattice of membrane receptors. *Proceedings of the National Academy of Sciences*, 96(18):10104–10108, 1999.

[93] Xing Yi Ling, David N. Reinhoudt, and Jurriaan Huskens. Reversible attachment of nanostructures at molecular printboards through supramolecular glue. *Chemistry of Materials*, 20(11):3574–3578, 2008.

[94] Galina V. Dubacheva, Angéline Van Der Heyden, Pascal Dumy, Oznur Kaftan, Rachel Auzély-Velty, Liliane Coche-Guerente, and Pierre Labbé. Electrochemically controlled adsorption of fc-functionalized polymers on beta-cd-modified self-assembled monolayers. *Langmuir*, 26(17):13976–13986, 2010. PMID: 20684518.

[95] Galina V. Dubacheva, Mathieu Galibert, Liliane Coche-Guerente, Pascal Dumy, Didier Boturyn, and Pierre Labbe. Redox strategy for reversible attachment of biomolecules using bifunctional linkers. *Chem. Commun.*, 47:3565–3567, 2011.

[96] Hiroyasu Yamaguchi, Yuichiro Kobayashi, Ryosuke Kobayashi, Yoshinori Takashima, Akihito Hashidzume, and Akira Harada. Photoswitchable gel assembly based on molecular recognition. *Nat Commun*, 3:603, 2012.

[97] Severine Rose, Alexandre Prevoteau, Paul Elziere, Dominique Hourdet, Alba Marcellan, and Ludwik Leibler. Nanoparticle solutions as adhesives for gels and biological tissues. *Nature*, 505(7483):382–385, 2014.

[98] Gregg A. Duncan and Michael A. Bevan. Computational design of nanoparticle drug delivery systems for selective targeting. *Nanoscale*, 7:15332–15340, 2015.

[99] Michael I. Monine, Richard G. Posner, Paul B. Savage, James R. Faeder, and William S. Hlavacek. Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. *Biophysical Journal*, 98(1):48 – 56, 2010.

[100] Tushar Satav, Jurriaan Huskens, and Pascal Jonkheijm. Effects of variations in ligand density on cell signaling. *Small*, 11(39):5184–5199, 2015.

[101] Hao Wu. Higher-order assemblies in a new paradigm of signal transduction. *Cell*, 153(2):287–292, 2013.

[102] Jin Liu, Neeraj J. Agrawal, Andres Calderon, Portonovo S. Ayyaswamy, David M. Eckmann, and Ravi Radhakrishnan. Multivalent binding of nanocarrier to endothelial cells under shear flow. *Biophysical Journal*, 101(2):319 – 326, 2011.

[103] Brian E Collins and James C Paulson. Cell surface biology mediated by low affinity multivalent protein-glycan interactions. *Current Opinion in Chemical Biology*, 8(6):617 – 625, 2004.

[104] Shuqi Chen and Timothy A. Springer. Selectin receptor-ligand bonds: Formation limited by shear rate and dissociation governed by the bell model. *Proceedings of the National Academy of Sciences*, 98(3):950–955, 2001.

[105] Djurre H. De Jong, Lars V. Schäfer, Alex H. De Vries, Siewert J. Marrink, Herman J. C. Berendsen, and Helmut Grubmüller. Determining equilibrium constants for dimerization reactions from molecular dynamics simulations. *Journal of Computational Chemistry*, 32(9):1919–1928, 2011.

[106] Stefano Angioletti-Uberti, Bortolo M. Mognetti, and Daan Frenkel. Theory and simulation of dna-coated colloids: a guide for rational design. *Phys. Chem. Chem. Phys.*, 18:6373–6393, 2016.

[107] Kangning Ren, Niaz Banaei, and Richard N. Zare. Sorting inactivated cells using cell-imprinted polymer thin films. *ACS Nano*, 7(7):6031–6036, 2013.

[108] Shawn M. Douglas, Ido Bachelet, and George M. Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335(6070):831–834, 2012.

[109] Hongyan Yuan, Changjin Huang, Ju Li, George Lykotrafitis, and Sulin Zhang. One-particle-thick, solvent-free, coarse-grained model for biological and biomimetic fluid membranes. *Phys. Rev. E*, 82:011905, 2010.

[110] Norbert Kern and Daan Frenkel. Fluid-fluid coexistence in colloidal systems with short-ranged strongly directional attraction. *The Journal of Chemical Physics*, 118(21):9882–9889, 2003.

[111] Lorenzo Di Michele, Bortolo M. Mognetti, Taiki Yanagishima, Patrick Varilly, Zachary Ruff, Daan Frenkel, and Erika Eiser. Effect of inert tails on the thermodynamics of dna hybridization. *Journal of the American Chemical Society*, 136(18):6538–6541, 2014. PMID: 24750023.

[112] M. Rubinstein and R.H. Colby. *Polymer Physics*. OUP Oxford, 2003.

[113] Ryan Simon, Melissa E. Collins, and David A. Spivak. Shape selectivity versus functional group pre-organization in molecularly imprinted polymers. *Analytica Chimica Acta*, 591(1 SPEC. ISS.):7–16, 2007.

[114] Robert J. Umpleby II, Miguel Bode, and Ken D. Shimizu. Measurement of the continuous distribution of binding sites in molecularly imprinted polymers. *The Analyst*, 125(7):1261–1265, 2000.

[115] Yongqin Lv, Zhixing Lin, Tianwei Tan, Wei Feng, Peiyong Qin, and Cong Li. Application of molecular dynamics modeling for the prediction of selective adsorption properties of dimethoate imprinting polymer. *Sensors and Actuators, B: Chemical*, 133(1):15–23, 2008.

[116] Liqing Wu and Yuanzong Li. Study on the recognition of templates and their analogues on molecularly imprinted polymer using computational and conformational analysis approaches. *Journal of Molecular Recognition*, 17(6):567–574, 2004.

[117] Günter Wulff. Molecular Imprinting in Cross-Linked Materials with the Aid of Molecular Templates— A Way towards Artificial Antibodies. *Angewandte Chemie International Edition In English*, 34(17):1812–1832, 1995.

[118] Russel P. Goodman, Richard M. Berry, and Andrew J. Turberfield. The single-step synthesis of a dna tetrahedron. *Chemical Communications*, 12:1372–1373, 2004.

[119] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Computational science series. Elsevier Science, 2001.

[120] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[121] Jorn Ilja Siepmann and Daan Frenkel. Configurational bias monte carlo: a new sampling scheme for flexible chains. *Molecular Physics*, 75(1):59–70, 1992.

[122] George Marsaglia. Choosing a Point from the Surface of a Sphere. *The Annals of Mathematical Statistics*, 43(2):645–646, 1972.

[123] Daan Frenkel. Speed-up of monte carlo simulations by sampling of rejected states. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17571–17575, 2004.

Appendix A

Molecular imprinting derivations

Any model for molecular imprinting should offer a description of the step in which free ligands undergo equilibrium binding to the template, and of the step in which analytes binds to these ligands, once they have been permanently tethered to the polymer matrix by cross linking. In this Appendix, we provide the statistical mechanical background for both theoretical descriptions.

A.1 Cavity formation theory

We consider a volume V containing a solution of various types of ligands. The number of ligands of type α is denoted by M_α. Before cross-linking, the ligands are free to move. To this solution we add N template particles. Each template particle has various binding sites (receptors) on its surface. In what follows, we will designate ligand types with greek letters $j = \{\alpha, \beta, \cdots\}$, and the complementary binding sites with $i = \{\alpha', \beta', \cdots\}$. The number of binding sites of type α' is denoted by $n_{\alpha'}$. The spatial extent of binding sites and ligands is assumed to be smaller than the template particle, such that different ligands bound to the same particle do not interact with each other. We also assume that ligand binding will not induce conformational changes in the template particles, which would introduce a correlation between different binding events. In equilibrium, an average number m_α ligands of type α will bind to the complementary binding sites on a template particle, as depicted in figure A1.

All binding sites of the same type are equivalent. In thermodynamic equilibrium there will be a chemical equilibrium between the number of free ligands in the solution and the number of ligands bound to the particles. We can easily calculate the occupancy fractions of individual sites. We first write a single-site grand-canonical partition function

for binding site of type i:

$$\zeta_i = 1 + \sum_j e^{\beta(\mu_j - \Delta\tilde{G}_{ij})}, \tag{A1}$$

where the sum goes over all types of ligands present in the system. $\Delta\tilde{G}_{ij}$ is an interaction matrix that specifies the hybridization free energies between a ligand of type j and a binding site of type i in the pre-polymerization solution, μ_j is the chemical potential of ligand type j and $\beta \equiv 1/k_B T$, where k_B is the Boltzmann constant and T the absolute temperature. At a given density of ligands in the solution, the average occupancy of a binding sites of type i with a ligand of type j will be

$$f_i^j = \frac{e^{\beta(\mu_j - \Delta\tilde{G}_{ij})}}{\zeta_i}. \tag{A2}$$

In what follows, we will assume that the total density of ligands in a solution is sufficiently low that we can assume the solution to be ideal. Extension to non-ideal solution is straightforward but the notation is more cumbersome. For dilute solutions the chemical potential is proportional to the logarithm of the density, therefore

$$e^{\beta\mu_j} \approx c_j^f/\rho_0 \tag{A3}$$

with c_j^f the molar concentration of free ligands in solution and $\rho_0 = 1M$ the reference concentration. By binding to the particles the number density of ligands in the bulk is depleted

$$c_j^f = c_j - C_T \sum_i n_i f_i^j. \tag{A4}$$

This equation states that the average number density of free ligands of type j in the solution is equal to the total concentration $c_j = M_j/(VN_A)$ of ligands minus the average number of ligands that are bound to the templates, $C_T = N/(VN_A)$ is the template concentration and N_A the Avogadro's constant. We have assumed that the system is large (macroscopic) such that the fluctuations in the number of free ligands can be neglected $c_j^f = \langle c_j^f \rangle$. Inserting (A1), (A2), (A3) into (A4) we arrive at a system of equations

$$c_j^f = c_j - C_T \sum_i n_i \frac{c_j^f e^{-\beta\Delta\tilde{G}_{ij}}}{\rho_0 + \sum_{j'} c_{j'}^f e^{-\beta\Delta\tilde{G}_{ij'}}}, \tag{A5}$$

which can be solved self-consistently for the densities of free ligands c_j^f.

We now assume that each ligand can bind only to a complementary binding site. For example a ligand of type α can only bind to a binding sites of type α'. This assumption

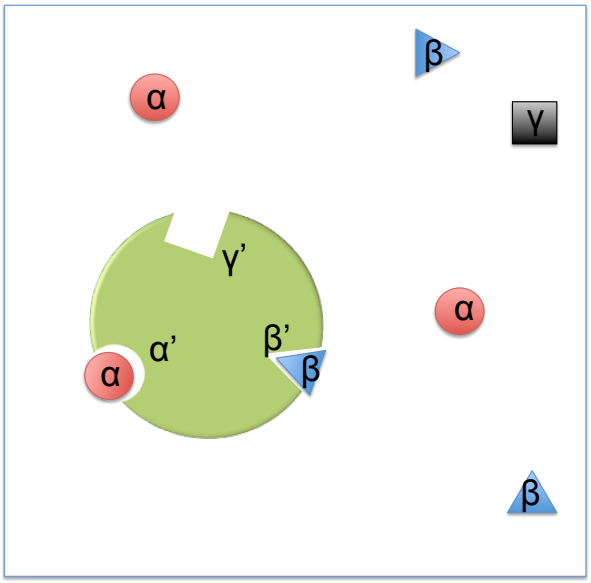


Fig. A1 Free ligand binding. Template particles (big green) are mixed with ligands in a solution, ligands of different types (α, β, γ) can reversibly bind to the specific binding sites on a particle $(\alpha', \beta', \gamma')$. Here we show only a single particle in the box.

implies that $\boldsymbol{\Delta}\tilde{G}$ is a diagonal matrix. The set of equations (A5) then decouple:

$$c_\alpha^f = c_\alpha - C_T n_{\alpha'} \frac{c_\alpha^f e^{-\beta\Delta\tilde{G}_{\alpha'\alpha}}}{\rho_0 + c_\alpha^f e^{-\beta\Delta\tilde{G}_{\alpha'\alpha}}} \tag{A6}$$

This is a quadratic equation that can be solved to yield the ligand concentration c_α^f. The solution of this equation gives us the bulk densities from which it is straightforward to obtain average occupancy fractions $f_{\alpha'}^\alpha$ with Eq. (A2) and (A3), because $f_{\alpha'}^j = 0$ for $j \neq \alpha$,

$$f_{\alpha'}^\alpha = \frac{\tilde{K}_D^{\alpha\alpha'} + n_{\alpha'}C_T + c_\alpha - \sqrt{(\tilde{K}_D^{\alpha\alpha'} + n_{\alpha'}C_T + c_\alpha)^2 - 4c_\alpha n_{\alpha'}C_T}}{2n_{\alpha'}C_T}, \tag{A7}$$

with $\tilde{K}_D^{\alpha\alpha'} = \rho_0 e^{\beta\Delta\tilde{G}_{\alpha\alpha'}}$ the equilibrium dissociation constant of $\alpha - \alpha'$ binding.

Eq. (A6) can be written in a somewhat more transparent form using the standard chemical equilibrium notation

$$[\alpha] = [\alpha_0] - [T] n_{\alpha'} f_{\alpha'}^\alpha \ , \tag{A8}$$

where $[\alpha_0] = c_\alpha$ and $[\alpha] = c_\alpha^f$ are, respectively, the initial and final concentrations of ligand of type α, $[T] = C_T$ is the concentration of the template particles, $n_{\alpha'}$ denotes the number of α' binding sites per template particle. The occupancy fraction of α' binding

sites

$$f^{\alpha}_{\alpha'} = \frac{[\alpha]}{[\alpha] + \tilde{K}_D^{\alpha\alpha'}}, \tag{A9}$$

Inserting the above equation into (A8) we find a quadratic equation in $[\alpha]$ that can be solved to yield

$$[\alpha] = \frac{-\tilde{K}_D^{\alpha\alpha'} - n_{\alpha'}[T] + [\alpha_0] + \sqrt{(\tilde{K}_D^{\alpha\alpha'} + n_{\alpha'}[T] - [\alpha_0])^2 + 4[\alpha_0]\tilde{K}_D^{\alpha\alpha'}}}{2} \tag{A10}$$

Finally, inserting this expression for $[\alpha]$ in (A9), the occupancy fractions $f^{\alpha}_{\alpha'}$ follow.

The imprinted polymer is created by cross-linking the polymer matrix and removing the bound template particles. In our model, the cross-linking process fixes the average positions of the ligands in the (flexible) matrix. If the occupancy fractions in the pre-polymerization solution are close to unity, the resulting imprinted polymer will only contain cavities that contain a number of ligands equal to the number of binding sites on the template. In contrast, for low occupancy fractions, the matrix will contain heterogeneous cavities. In that case we would need to evaluate the expected distribution of different cavities formed and calculate properties of each family separately. For example if $f = 0.5$ and we have 2 binding sites on a particle there will be about 25% of cavities with 2 ligands, 50% with just one and 25% with none. Also the ligands that did not bind to the particle in the pre-polymerization complex will be randomly distributed throughout the polymer matrix and can be treated equivalently to ligands distributed in a NIP (non-imprinted polymer). Therefore to characterize such a MIP, a free energy of binding should be computed for the different cavities expected to form. This process is tractable if the number of receptors on the template is low (say 2 or 3). Alternatively, we can perform Monte Carlo simulations of the pre-polymerization matrix and then calculate the adsorption isotherm of analyte binding, as shown on Figure 4.3.

Heterogeneity of cavities will always diminish the selectivity of MIPs. Therefore, the cavity distribution should be made as homogeneous as possible, i.e $f \approx 1$. A simple rule of thumb would be that the initial concentration of ligands should be at least the sum of the dissociation constant and the concentration of receptors

$$[\alpha_0] > \tilde{K}_D^{\alpha\alpha'} + n_{\alpha'}[T] \ . \tag{A11}$$

This will ensure that the average occupancy of binding sites is greater than $f_{\alpha'} > \frac{1}{2}$.

A.2 Ligand binding

In the polymer matrix the ligands fluctuate around their "anchor" positions and we define $\mathbf{r}_i^{anc} = \langle \mathbf{r}_i^{lig} \rangle$. the potential of mean force is, within the harmonic approximation

$$U_h = \frac{k_h}{2} \left(\mathbf{r}^{lig} - \mathbf{r}^{anc} \right)^2 . \tag{A12}$$

with k_h the effective "spring" constant derived in Chapter 4.

We introduce a hard-sphere like particle with specific binding sites on its surface. Ligands and binding sites can bind together, when they do the system is constrained by $\mathbf{r}_i^{lig} = \mathbf{r}_j^{bs}$, the position of the ligand i is equal to the position of the binding site j on a particle. There is no preferred bond directionality as discussed above; the probability that a ligand binds depends only on the relative distance between the binding site and the ligand anchor. Upon a bond formation the system gains the hybridization free energy ΔG_{ij} it has to pay the harmonic penalty U_h and looses some conformational entropy because of the constraint $\mathbf{r}_i^{lig} = \mathbf{r}_j^{bs}$.

For a given position $\mathbf{r}^p$ and orientation $\mathbf{\Omega}^p$ of the particle we can calculate the probability that a bond is formed. The non-binding partition function (for a single free ligand) is an integral over all possible ligand positions

$$Q_i^u(\mathbf{r}_p, \mathbf{\Omega}_p) = \int d\mathbf{r}_i^{lig} e^{-\frac{\beta k_d}{2} |\mathbf{r}_i^{lig} - \mathbf{r}_i^{anc}|^2} \approx \left(\frac{2\pi}{\beta k_h} \right)^{3/2} , \tag{A13}$$

is a constant. The solution of this Gaussian integral is only approximate due to the fact the volume of the phase space accessible to a fluctuating ligand is reduced if a hard-sphere particle is present in the vicinity (the ligand cannot penetrate into the particle). In performing this approximation we neglect excluded volume effects between particles and free ligands. However, corrections are rather small, for example if a particle surface is located exactly at the free ligand anchoring point, we overestimate Q_0 by about a factor of 2 (because almost half of the available volume is excluded). This in turn translates into the free energy correction of about $\ln 2 \approx 0.7 k_B T$. We stress that this approximation only applies to cases when the particle is partially bound to the cavity (i.e. not all ligands are bound, but the particle is present in the cavity).

The ligand-to-binding-site bound state partition function is

$$Q_{ij}^b(\mathbf{r}_p, \mathbf{\Omega}_p) = v_0 e^{-\frac{\beta k_h}{2} |\mathbf{r}_j^{bs} - \mathbf{r}_i^{anc}|^2} e^{-\beta \Delta G_{ij}} , \tag{A14}$$

where $v_0 \equiv 1/(\rho_0 N_A)$ defined as the standard volume, $\rho_0 = 1\text{M}$ the standard concentration and N_A the Avogadro's number. v_0 can be recognised as the available volume per solute particle in a solution with concentration $\rho_0 = 1\text{M}$. We remember that the position of the binding site is determined by the position and orientation of the particle $\mathbf{r}^{bs} = \mathbf{r}^{bs}(\mathbf{r}^p, \mathbf{\Omega}^p)$.

Using (A14) and (A13) the ratio of probabilities of being bound to non-bound is

$$\frac{p_{ij}(\mathbf{r}_p, \mathbf{\Omega}_p)}{p_0} = \frac{Q^b_{ij}(\mathbf{r}_p, \mathbf{\Omega}_p)}{Q^u_i} = v_0 \left(\frac{\beta k_h}{2\pi}\right)^{3/2} e^{-\frac{\beta k_h}{2}|\mathbf{r}^{bs}_j - \mathbf{r}^{anc}_i|^2} e^{-\beta \Delta G_{ij}} \tag{A15}$$

which depends only on the hybridization free energy ΔG_{ij} and the position of the binding site relative to the ligand anchor.

Following the same procedure as outlined above for one ligand, we can find the probability that two ligands are bound to two binding sites

$$p_{ij,i'j'}(\mathbf{r}_p, \mathbf{\Omega}_p) = p_{ij}(\mathbf{r}_p, \mathbf{\Omega}_p) p_{i'j'}(\mathbf{r}_p, \mathbf{\Omega}_p), \tag{A16}$$

which is simply a product of individual binding probabilities as we assume that different binding events are uncorrelated. This assumption should hold as long as: (i) the distance between ligand anchors is greater than the standard deviation due to fluctuations $a > (\frac{2\pi}{\beta k_h})^{1/2}$ and (ii) the binding sites on a particle are independent, i.e. binding of one site does not introduce conformational (or other) changes in the particle that would affect the binding of a second site.

Further generalization to any number of ligands and binding sites is straightforward and follows the same procedure as described above.

A.3 Analytical free energy calculations

The bound partition function of a particle in a cavity counts all possible states of a particle in a cavity

$$Q^b = \sum_{\kappa=1}^{\kappa_{max}} \left[(q^u_l)^{n_l - \kappa} v_0^\kappa \sum_{s(\kappa)} \int d\mathbf{r}_p d\mathbf{\Omega}_p e^{-\beta \sum_{ij(s)} \left[\Delta G_{ij} + \frac{k_h}{2} |\mathbf{r}^{anc}_i - \mathbf{r}^{bs}_j|^2 \right]} \right], \tag{A17}$$

where outer the sum goes over all possible number of bonds κ and all bonding arrangements $s(\kappa)$ of κ bonds. $q^u_l = (\frac{2\pi}{\beta k_h})^{3/2}$ is the phase volume of a non-bound ligand (A13), n_l is the total number of ligands in the cavity. The inner $\sum_{ij(s)}(...)$ sums the potential and bonding energy of all bound pairs $ij(s)$ of the particular bonding arrangement s.

For an empty cavity, when the particle is free in the solution, all ligands are unbound and the partition function is trivial

$$Q^u = 8\pi^2 v_0 (q_l^u)^{n_l} \; ; , \tag{A18}$$

where $8\pi^2$ comes from integrating over the solid angle and rotations and $v_0 = 1/(\rho_0 N_A)$ is the microscopic volume related to the choice of a standard concentration ρ_0. The binding free energy of a particle (from solution to a cavity) is determined by the ratio of the above partition functions

$$Q^{cav} \equiv \frac{Q^b}{Q^u} = \sum_{\kappa=1}^{\kappa_{max}} q_\kappa = q_1 + q_2 + ... \; , \tag{A19}$$

so that the binding free energy of an analyte-cavity system is simply

$$F^{cav} = -k_B T \ln(Q^{cav}) \; , \tag{A20}$$

or the equivalent equilibrium association constant is simply $K_A = \rho_0 Q$. Each term in the above sum

$$q_\kappa = \frac{v_0^{\kappa-1}}{8\pi^2} \left(\frac{2\pi}{\beta k_h} \right)^{-3\kappa/2} \sum_{s(\kappa)} \int d\mathbf{r}_p d\mathbf{\Omega}_p e^{-\beta \sum_{ij(s)} \left[\Delta G_{ij} + \frac{k_h}{2} |\mathbf{r}_i^{anc} - \mathbf{r}_j^{bs}|^2 \right]} \; , \tag{A21}$$

is a partition function of a particle in a cavity with κ formed bonds.

We now proceed to determine each q_κ separately. A single bond partition function is

$$q_1 = \frac{1}{8\pi^2} \left(\frac{2\pi}{\beta k_h} \right)^{-3/2} \sum_{ij} \int d\mathbf{r}_p d\mathbf{\Omega}_p e^{-\beta \Delta G_{ij} + \frac{\beta k_h}{2} |\mathbf{r}_i^{anc} - \mathbf{r}_j^{bs}|^2} . \tag{A22}$$

We only need to evaluate a Gaussian integral, while integration over orientations gives $\int d\mathbf{\Omega}_p = 8\pi^2$, hence, all pre-factors cancel out and we simply get

$$q_1 = \sum_{ij} e^{-\beta \Delta G_{ij}} \tag{A23}$$

a sum goes over all possible (distinct) pairs: j stands for binding sites and i for ligands. For example if there are 5 ligands in the cavity and 3 binding sites on a particle we have in general 15 possible combination of how to form a single bond.

A.3.1 Evaluation of q_2

In the evaluation of q_2, the imprinting properties and selectivity start emerging. Similar to the one-bond scenario described above, we write the partition function as an integral over all positions $\mathbf{r}_p$ and rotations $\mathbf{\Omega}_p$ of a particle, and additionally summed over all possible combinations of 2 bound pairs $ij, i'j'$, where i, i' stand for ligands and j, j' for binding sites.

$$q_2 = A \sum_{ij,i'j'} \left(e^{-\beta(\Delta G_{ij}+\Delta G_{i'j'})} \int d\mathbf{r}_p d\mathbf{\Omega}_p \; e^{\frac{\beta k_h}{2}(|\mathbf{r}_i-\mathbf{r}_j|^2+|\mathbf{r}_{i'}-\mathbf{r}_{j'}|^2)} \right). \tag{A24}$$

with pre-factor $A = \frac{v_0}{8\pi^2}\left(\frac{2\pi}{\beta k_h}\right)^{-3}$. For later convenience we will generalize q_2 to include all possible positions of the ligand anchors.

$$q_2 = A \sum_{ij,i'j'} \left(e^{-\beta(\Delta G_{ij}+\Delta G_{i'j'})} \int d\mathbf{r}_p d\mathbf{\Omega}_p d\mathbf{r}_i d\mathbf{r}_{i'} \; e^{\frac{\beta k_h}{2}(|\mathbf{r}_i^{anc}-\mathbf{r}_j^{bs}|^2+|\mathbf{r}_{i'}^{anc}-\mathbf{r}_{j'}^{bs}|^2)} p(\mathbf{r}_i, \mathbf{r}_{i'}) \right). \tag{A25}$$

where $p(\mathbf{r}_i, \mathbf{r}_{i'})$ stands for the joint probability to find a ligand anchor i at position $\mathbf{r}_i$ and a ligand anchor i' at $\mathbf{r}_{i'}$. Since the only important variables are the relative distances between binding sites and ligand anchors, we only need to account for all possible relative translations and rotations. There is translational and orientational symmetry in the position and orientation of the particle, in other words, for every arbitrary position and orientation of the particle, the integral over ligand anchor position is the same. Hence we can integrate out the particle's degrees of freedom

$$q_2 = 8\pi^2 V A \sum_{ij,i'j'} \left(e^{-\beta(\Delta G_{ij}+\Delta G_{i'j'})} \int d\mathbf{r}_i d\mathbf{r}_{i'} \; e^{\frac{\beta k_h}{2}(|\mathbf{r}_i-\mathbf{r}_j|^2+|\mathbf{r}_{i'}-\mathbf{r}_{j'}|^2)} p(\mathbf{r}_i, \mathbf{r}_{i'}) \right), \tag{A26}$$

because $\int d\mathbf{r} d\mathbf{\Omega} = 8\pi^2 V$, V being the volume of the cavity, we will later elucidate on V.

We focus on 2 limiting cases. Either there are no correlations between different ligand anchor positions, or the positions are sharply defined. If a polymer matrix is not imprinted that in turn means that there are no correlations between different ligand positions $p(\mathbf{r}_i, \mathbf{r}_{i'}) = 1/V^2$, i.e. the positions of the ligand anchors are random, this is the case of a non-imprinted polymer (NIP). The integral in Eq. (A26) can be separated into two independent 3D Gaussian integrals which are trivial to solve

$$q_2^{nip} = \frac{v_0}{V} \sum_{ij,i'j'} e^{-\beta(\Delta G_{ij}+\Delta G_{i'j'})}. \tag{A27}$$

The partition function is inversely proportional to the volume. This makes sense intuitively because the larger the volume the less likely it is that the distance between two random points will match the distance between the two binding sites on a particle. We stress that Eq. (A27) is a mean-field prediction which gives us the average q_2 in a NIP. In deriving (A27) we have assumed that k_h is a constant, but in fact the relative fluctuations of 2 ligands will, in principle, also depend on their relative distance (4.11). However, the regime we are interested in is the regime of geometrical selectivity $\beta k_h \sigma^2 > 1$ (fluctuations are smaller than the template/cavity size σ). In this case the largest contribution to the integral in (A26) comes when the inter ligand anchor distance is similar to the inter binding site distance on the particle. Hence, we approximate $k_h(|\mathbf{r}_i - \mathbf{r}_{i'}|) \approx k_h(|\mathbf{r}_j - \mathbf{r}_{j'}|) \simeq k_h(\sigma)$, where in the last step we approximated that the inter binding site distance is given by the particle size σ.

On the other hand, an imprinted cavity has well defined distances between ligand anchoring points. Due to translational and orientational symmetry, the positions of the ligand anchors $\mathbf{r}_i$ and $\mathbf{r}_{i'}$ are arbitrary as long as we keep a fixed distance between them $|\mathbf{r}_i - \mathbf{r}_{i'}| = a_{ii'}$, the arrangement of the ligand anchors matches a template molecule with which the cavity was imprinted. The probability to find ligand anchors at given positions is

$$p(\mathbf{r}_i, \mathbf{r}_{i'}) = \frac{1}{4\pi a_{ii'}^2 V} \delta(|\mathbf{r}_i - \mathbf{r}_{i'}| - a_{ii'}), \tag{A28}$$

with $\delta(x)$ a Dirac delta function. The $1/(4\pi a_{ii'}^2 V)$ pre-factor comes in due to normalization $\int p(\mathbf{r}_i, \mathbf{r}_{i'}) d\mathbf{r}_i d\mathbf{r}_{i'} = 1$. Inserting Eq. (A28) into Eq. (A26) we write the partition function as

$$q_2 = \frac{2\pi A}{a_{ii'}^2} \sum_{ii', jj'} e^{-\beta(\Delta G_{ij} + \Delta G_{i'j'})} J(a_{ii'}, b_{jj'}), \tag{A29}$$

where $J(a_{ii'}, b_{jj'})$ is the configurational integral we need to solve

$$J(a_{ii'}, b_{jj'}) = \int d\mathbf{r}_i d\mathbf{r}_{i'} \; e^{\frac{\beta k_h}{2}(|\mathbf{r}_i - \mathbf{r}_j|^2 + |\mathbf{r}_{i'} - \mathbf{r}_{j'}|^2)} \delta(|\mathbf{r}_i - \mathbf{r}_{i'}| - a_{ii'}), \tag{A30}$$

which, as we will see later, depends only on the relative distances between binding sites ($b_{jj'} = |\mathbf{r}_j - \mathbf{r}_{j'}|$) and between ligand anchors ($a_{ii'}$). This is a Gaussian integral with non-linear coupling, however, in this case it can be solved analytically, we write

$$J(a_{ii'}, b_{jj'}) = \int_V d\mathbf{r}_i e^{\frac{\beta k_h}{2}|\mathbf{r}_i - \mathbf{r}_j|^2} \oint_S d\mathbf{r}_{i'} \; e^{\frac{\beta k_h}{2}|\mathbf{r}_{i'} - \mathbf{r}_{j'}|^2} \delta(|\mathbf{r}_i - \mathbf{r}_{i'}| - a_{ii'}), \tag{A31}$$

where the innermost integral is a Gaussian (which is centered around $\mathbf{r}_{j'}$) integrated over a surface of a sphere, the sphere centre being located at $\mathbf{r}_i$. It can be solved by switching

to spherical coordinates

$$\oint_S d\mathbf{r}_{i'}\; e^{\frac{\beta k_h}{2}|\mathbf{r}_{i'}-\mathbf{r}_{j'}|^2}\delta(|\mathbf{r}_i-\mathbf{r}_{i'}|-a_{ii'}) = \frac{4\pi a_{ii'}}{\beta k_h}\frac{e^{-\frac{\beta k_h}{2}(|a_{ii'}|^2+|\mathbf{r}_i-\mathbf{r}_{j'}|^2)}}{|\mathbf{r}_i-\mathbf{r}_{j'}|}\sinh\left(\beta k_h a_{ii'}|\mathbf{r}_i-\mathbf{r}_{j'}|\right), \tag{A32}$$

where $\sinh(x)$ denotes the hyperbolic sine function. Inserting into (A31) we find

$$J(a_{ii'},b_{jj'}) = \frac{4\pi a_{ii'}}{\beta k_h}e^{-\frac{\beta k_h}{2}|a_{ii'}|^2}\int_V d\mathbf{r}_i\frac{e^{-\frac{\beta k_h}{2}(|\mathbf{r}_i-\mathbf{r}_j|^2+|\mathbf{r}_i-\mathbf{r}_{j'}|^2)}}{|\mathbf{r}_i-\mathbf{r}_{j'}|}\sinh\left(\beta k_h a_{ii'}|\mathbf{r}_i-\mathbf{r}_{j'}|\right). \tag{A33}$$

We extend the limits of integration to infinity, transform to spherical coordinates, complete a perfect square in the exponent and it turns out that the integral can be written as a sum of definite Gaussian integrals (error function integrals) which partly cancel out and the final result is

$$J(a_{ii'},b_{jj'}) = 8\left(\frac{\pi}{\beta k_h}\right)^{5/2}\frac{a_{ii'}e^{-\frac{\beta k_h}{4}(d_{ii'}^2+b_{jj'}^2)}}{b_{jj'}}\sinh\left(\beta k_h a_{ii'}b_{jj'}/2\right). \tag{A34}$$

Finally, by inserting Eq. (A34) into (A29), we find the configurational part of the two bond partition function

$$\tilde{q}_2(a_{ii'},b_{jj'}) = \frac{v_0(\beta k_h)^{1/2}}{4\pi^{3/2}}\;\frac{e^{-\frac{\beta k_h}{4}(a_{ii'}^2+b_{jj'}^2)}}{a_{ii'}b_{jj'}}\sinh\left(\beta k_h a_{ii'}b_{jj'}/2\right)\,, \tag{A35}$$

which is a function only of the distances between the two ligands anchors ($a_{ii'}$) and two binding sites on a particle ($b_{jj'}$). q_2 is then obtained by summing over all possible 2 bond configurations

$$q_2 = \sum_{ij,i'j'} e^{-\beta(\Delta G_{ij}+\Delta G_{i'j'})}\tilde{q}_2(a_{ii'},b_{jj'}) \tag{A36}$$

To analyze this result we consider a special case with only 2 ligands and 2 binding sites with $a = a_{ii'} = b_{jj'}$, i.e. the imprinted site perfectly matches the particle, the partition function (A35) is simplified to

$$\tilde{q}_2(a=b) = \frac{v_0(\beta k_h)^{1/2}}{8\pi^{3/2}}\;\frac{1-e^{-\beta k_h a^2}}{a^2}. \tag{A37}$$

Now we distinguish two limits:

1. $\beta k_h a^2 \ll 1$; the thermal fluctuations in position of a free ligand are much grater than the particle size, we expand the exponential to first order, a cancel out and

we find

$$\tilde{q}_2(a=b;\beta k_h a^2 \ll 1) = \frac{v_0(\beta k_h)^{3/2}}{8\pi^{3/2}}, \tag{A38}$$

a constant, which is to be expected, since in this limit the particle effectively feels two springs pulling in the same direction. $\frac{v_0(\beta k_h)^{3/2}}{8\pi^{3/2}}$ is the reduction of phase space when joining together ends of two springs.

2. $\beta k_h a^2 \gg 1$; in the opposite limit the two bonds effectively (translationally and orientationally) confine the particle. The exponential in (A37) is neglected and we find

$$\tilde{q}_2(a=b;\beta k_h a^2 \gg 1) = \frac{v_0(\beta k_h)^{1/2}}{8\pi^{3/2}a^2}. \tag{A39}$$

The partition function falls off with the square of the distance between ligands a. Such scaling is also to be expected because the greater the distance the smaller the orientational fluctuations of the particle. Say S is a surface of a particle that a ligand can explore, the solid angle of possible particle orientations is then $\Omega \propto S/a^2$ for a spherical particle.

A.3.2 Evaluation of q_3 and beyond

Similar to 2 bond case above, we write the 3 bond partition function as an integral over all possible positions of the 3 ligand anchors and a sum over all possible 3 bond $(ij, i'j', i''j'')$ combinations,

$$q_3 = 8\pi^2 V B \sum_{ij,i'j',i''j''} \left(e^{-\beta(\Delta G_{ij}+\Delta G_{i'j'}+\Delta G_{i''j''})} \int d\mathbf{r}_i d\mathbf{r}_{i'} d\mathbf{r}_{i''}\ e^{\frac{\beta k_h}{2}(|\mathbf{r}_i-\mathbf{r}_j|^2+|\mathbf{r}_{i'}-\mathbf{r}_{j'}|^2+|\mathbf{r}_{i''}-\mathbf{r}_{j''}|^2)} p(\mathbf{r}_i,\mathbf{r}_{i'},\mathbf{r}_{i''}) \right), \tag{A40}$$

with a constant $B = \frac{v_0^2}{8\pi^2}\left(\frac{2\pi}{\beta k_h}\right)^{-9/2}$. In the case when the ligand anchors are randomly distributed $p(\mathbf{r}_i,\mathbf{r}_{i'},\mathbf{r}_{i''}) = 1/V^3$ the calculation is simple as we only need to calculate 3 independent Gaussian integrals, similar to procedure for q_2^{nip} we find

$$q_3^{nip} = \frac{v_0^2}{V^2} \sum_{ij,i'j',i''j''} e^{-\beta(\Delta G_{ij}+\Delta G_{i'j'}+\Delta G_{i''j''})}. \tag{A41}$$

In the imprinted case the relative positions of anchors are well defined

$$p(\mathbf{r}_i,\mathbf{r}_{i'},\mathbf{r}_{i''}) = A\delta(|\mathbf{r}_i-\mathbf{r}_{i'}|-a_{ii'})\delta(|\mathbf{r}_i-\mathbf{r}_{i''}|-a_{ii''})\delta(|\mathbf{r}_{i'}-\mathbf{r}_{i''}|-a_{i'i''}) = A\delta^3_{ii'i''}, \tag{A42}$$

here A is the normalization factor

$$A = \left(8\pi^2 a_{ii'}^2 V \sqrt{a_{ii''}^2 - \left(\frac{a_{ii'}^2 + a_{ii''}^2 - a_{i'i''}^2}{2a_{ii'}} \right)^2} \right)^{-1} . \tag{A43}$$

We have not found a general analytical solution of the integral (A40), nor for the cases with more than 3 ligand-receptor pairs.

A.3.3 Non-imprinted polymers

For a non-imprinted polymer (NIP) we have calculated the 2 and 3 bond partition function above, a general expression for κ bonds is

$$q_\kappa^{nip} = (V\rho_0 N_A)^{1-\kappa} \sum_{s(\kappa)} e^{-\beta \Delta G_{s(\kappa)}^{tot}}. \tag{A44}$$

We have used the relation that the microscopic volume is determined by the reference density $v_0 = 1/(\rho_0 N_A)$, V is the "volume" of the cavity and the sum goes over all possible combinations $s(\kappa)$ of κ bonds. As the non-imprinted polymer has no cavities *per se*, V can be chosen arbitrarily. We assume that the ligands (regions in the polymer matrix) accessible to the particles are the same for NIPs and MIPs. For weakly cross-linked matrices nearly all of the ligands should be accessible, while for very dense matrices (mesh size much smaller than the particle size) only the cavities and ligands near the surface will be accessible, as the particles cannot easily diffuse inside the matrix. Therefore, accessibility depends largely on the cross-linking distance and it will be similar for similarly prepared MIPs and NIPs.

If there is only a single type of ligands and receptors in the system (type α) the expected number of ligands in volume V is $N_\alpha = V c_\alpha N_A$ with c_α the ligand concentration in the NIP. As all bonds are equal $\sum_{s(\kappa)} e^{-\beta \Delta G_{s(\kappa)}^{tot}} = s(\kappa) e^{-\beta \kappa \Delta G}$. For a particle with n_α receptors, the number of possible combinations of κ bonds is given by combinatorial. We need to choose κ ligands out of N_α, κ receptors out of n_α and there are κ permutations of binding them together

$$s(\kappa) = \binom{N_\alpha}{\kappa} \binom{n_\alpha}{\kappa} \kappa! = \frac{N_\alpha! n_\alpha!}{(N_\alpha - \kappa)!(n_\alpha - \kappa)! \kappa!} \approx (N_\alpha)^\kappa \binom{n_\alpha}{\kappa} . \tag{A45}$$

Choosing a large volume V we have $N_\alpha \gg n_\alpha \geq \kappa$ and the above can be well approximated with $\binom{N_\alpha}{\kappa} \approx (N_\alpha)^\kappa$. The total NIP partition function is a sum over all possible number

of bonds κ

$$Q_{tot}^{nip} = \sum_{\kappa=1}^{n_\alpha} q_\kappa^{nip} = V\rho_0 N_A \sum_{\kappa=1}^{n_\alpha} \binom{n_\alpha}{\kappa} \left[\frac{N_\alpha e^{-\beta\Delta G}}{V\rho_0 N_A}\right]^\kappa = V\rho_0 N_A \left[\left(1 + \frac{c_\alpha e^{-\beta\Delta G}}{\rho_0}\right)^{n_\alpha} - 1\right] , \tag{A46}$$

which is obtained by summing over all binomial coefficients. This is the partition function of 1 particle in a non-imprinted polymer matrix with volume V. We note that the above result and its derivation is similar to the theory of ligand-decorated nanoparticles binding to receptor-decorated surfaces [10], the two problems are conceptually similar.

It is convenient to express the partition function per individual ligand for easier comparison to MIPs, as there are $N_\alpha = V c_\alpha N_A$ ligands in volume V, the mean partition function per ligand is

$$Q^{nip} = \frac{\rho_0}{c_\alpha}\left[\left(1 + \frac{c_\alpha e^{-\beta\Delta G}}{\rho_0}\right)^{n_\alpha} - 1\right] . \tag{A47}$$

If the particle has different receptors types j and there are different ligand types i in the NIP, the generalized expression becomes

$$Q^{nip} = \frac{\rho_0}{c_{tot}}\left[\prod_j \left(1 + \sum_i \frac{c_i e^{-\beta\Delta G_{ij}}}{\rho_0}\right)^{n_j} - 1\right] , \tag{A48}$$

with $c_{tot} = \sum_i c_i$ the total ligand concentration. We stress that the above is strictly true only for a single particle in the NIP, as it neglects particle-particle interactions and the effect of limited valency. The difficulty is that in a NIP we cannot really talk about independent 'cavities'. However, if the number of particles is small enough such that the fraction of bonded ligands is small, the above expression is correct and it provides a useful analytical way of characterising NIPs (as shown on Figure 4.3).

A.3.4 Summary of analytical results

We write the analyte-cavity bound state partition function as a sum over all possible number of bonds κ as

$$Q^{cav} = \sum_{\kappa}^{\kappa_{max}} q_\kappa . \tag{A49}$$

The binding free energy of analyte to a cavity is

$$F^{cav} = -k_B T \ln(Q^{cav}) \tag{A50}$$

In the case of a single bond we trivially find

$$q_1 = \sum_{ij} e^{-\beta \Delta G_{ij}}. \tag{A51}$$

The imprinted two bond partition function is analytical and is written as a function of the distances between ligand anchors ($a_{ii'}$) and binding sites ($b_{jj'}$)

$$q_2 = \sum_{ij,i'j'} e^{-\beta(\Delta G_{ij} + \Delta G_{i'j'})} \tilde{q}_2(a_{ii'}, b_{jj'}) \tag{A52}$$

with configurational part

$$\tilde{q}_2(a_{ii'}, b_{jj'}) = \frac{(\beta k_h)^{1/2}}{4\pi^{3/2}\rho_0 N_A} \frac{e^{-\frac{\beta k_h}{4}(a_{ii'}^2 + b_{jj'}^2)}}{a_{ii'} b_{jj'}} \sinh\left(\beta k_h a_{ii'} b_{jj'}/2\right) . \tag{A53}$$

For higher number of simultaneous bonds ($\kappa > 2$) we have not found a way to analytically compute the partition functions q_κ.

The general result for a partition function with random distribution of ligand anchors (NIP) is simply a product of independent Gaussian integrals, as there are no cavities in a NIP we normalise the partition function by the concentration of ligands:

$$Q^{nip} = \frac{\rho_0}{c_{tot}} \left[\prod_j \left(1 + \sum_i \frac{c_i e^{-\beta \Delta G_{ij}}}{\rho_0} \right)^{n_j} - 1 \right] , \tag{A54}$$

with c_i the molar concentration if ligand type i, $c_{tot} = \sum_i c_i$ the total concentration of ligands and n_j denotes the number of receptors of type j on a particle.

A.4 Binding affinity calculations

We calculate the binding affinity to NIPs using (A48)

$$BA^{nip} = c_\alpha \langle K_A^{nip} \rangle = \left(1 + \frac{c_\alpha}{K_D} \right)^{n_\alpha} - 1 \tag{A55}$$

for a single ligand-receptor type. In deriving the above we have used $K_A = Q/\rho_0$ and the single bond dissociation constant is $K_D = \rho_0 e^{-\beta \Delta G}$. The above result is trivially extended to multiple bond types

$$BA^{nip} = c_{tot} \langle K_A^{nip} \rangle = \prod_j \left(1 + \sum_i \frac{c_i}{K_D^{ij}} \right)^{n_j} - 1 , \tag{A56}$$

with i and j denoting ligand and bond types respectively and $K_D^{ij} = \rho_0 e^{\beta \Delta G_{ij}}$ is the interaction matrix.

For imprinted polymers the binding affinity is a sum of two contributions: binding to single cavities, and binding to non-imprinted ligands within a MIP or cross-cavity binding (to 2 cavities simultaneously, if they are close enough)

$$BA^{mip} = BA^{cav} + \mathcal{O} \tag{A57}$$

where the cavity binding term is determined by the concentration C_{cav} and free energy F^{cav} of analyte-cavity binding

$$BA^{cav} = C^{cav} \langle K_A^{cav} \rangle = \frac{C_{cav}}{\rho_0} \langle e^{-\beta F^{cav}} \rangle \ . \tag{A58}$$

The 'other' contribution $\mathcal{O}$ includes all possible terms of analyte binding to a MIP that are not included in a single cavity - single analyte picture, namely cross-cavity and binding to non-imprinted ligands within a MIP. This contribution will be substantial if the ligand density is high $c_{tot} N_A \gtrsim (\beta k_h)^{3/2}$. Assuming that imprinted cavities are randomly distributed throughout a MIP, we find

$$\mathcal{O} \approx BA^{nip} \ . \tag{A59}$$

Heuristically, from a ligand's point of view, other ligands belonging to the same cavity have well defined positions, but distances to all ligands in other cavities are random as the cavity positions are random.

For divalent analytes we can calculate the total MIP binding affinity BA_2^{mip} analytically, starting from Eqs. (A26, A27, A28) and considering the ligand probability distribution in the whole MIP, not just in a single cavity. For an imprinted ligand the probability distribution is simply a sum of a delta function (because the other ligand in the same cavity is at a well defined distance) and a constant (all other ligands). The total binding affinity to MIP imprinted by a divalent template is

$$BA_2^{mip} = BA_2^{cav} + BA_2^{nip} - \langle q_1 \rangle \frac{C_{cav}}{\rho_0} \tag{A60}$$

the sum of single cavity binding BA^{cav} and to ligands outside cavities or across 2 cavities BA^{nip}. The average single bond contribution $\langle q_1 \rangle \frac{C_{cav}}{\rho_0}$ needs to be subtracted due to double counting in the first two terms. We expand the single cavity binding affinity as

the sum over single and double bond contributions, using Eqs. (A49, A50,A58),

$$BA_2^{cav} = \frac{C_{cav}}{\rho_0}\left(\langle q_1\rangle + \langle q_2\rangle\right) \tag{A61}$$

If both bonds are equal and independent and template extraction process is efficient, we can write the average 2 bond partition function as

$$\langle q_2\rangle = 2\tilde{q}_2\left(\frac{f\rho_0}{K_D}\right)^2 \tag{A62}$$

with f the occupancy fractions of receptors on a template (given by Eq. A7) that is determined by the chemical equilibrium between ligands and templates in the pre-polymerization solution. In the above $\tilde{q}_2$ is the configurational part of the 2 bond partition function (A52,A53), f is a fraction of doubly functionalized cavities and a factor of 2 comes because there are 2 ways to bind 2-receptor analyte to a 2-ligand cavity, if both bonds are equal. If we have 2 different ligand - receptor pairs $\alpha - \alpha'$ and $\beta - \beta'$, and no cross binding between them, we only need to remove the factor 2 in (A62)

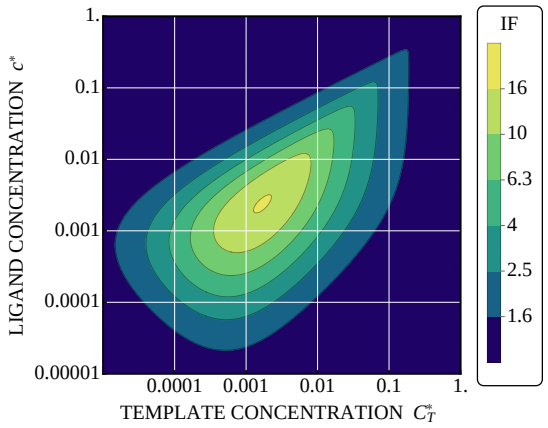
$$\langle q_2^{diff}\rangle = \tilde{q}_2\rho_0^2\frac{f_{\alpha'}^{\alpha} f_{\beta'}^{\beta}}{K_D^{\alpha\alpha'}K_D^{\beta\beta'}} \ . \tag{A63}$$

Inserting (A61) and (A62) into (A60) we find the binding affinity of a divalent analyte into a MIP imprinted with a divalent template and single ligand-receptor bind type (both bonds are equal)

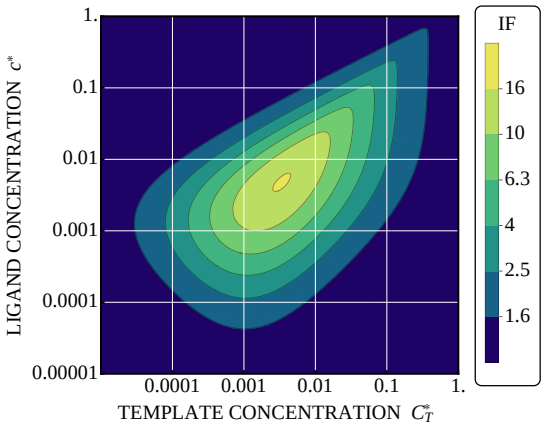
$$BA_2^{mip} = BA_2^{nip} + \frac{2C_T\tilde{q}_2 f^2\rho_0}{K_D^2} \ . \tag{A64}$$

To clarify, BA_2^{nip} is given by (A55), assuming efficient template extraction $C_{cav} \approx C_T$ cavity concentration is given by the template concentration in the MIP formation step, $\tilde{q}_2$ is given by (A53), f by (A7) and $K_D = \rho_0 e^{\beta\Delta G}$ is the single bond dissociation constant. We have used this expression for the binding affinity to calculate the imprinting and separation factor landscapes shown on Figures 4.4 and 4.5 in the main text.

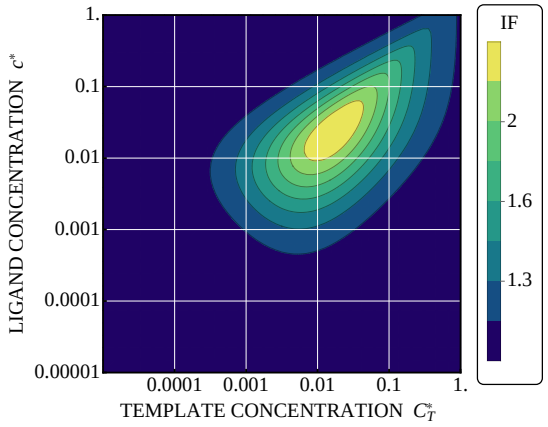
Stoichiometric ratio of ligands to templates in the pre-polymerization solution is close to optimal. On Figure A2 we show Imprinting factor depending on the ligand c and template C_T concentrations in the pre-polymerization phase. We keep the matrix stiffness k_h and the ligand-receptor dissociation constant K_D fixed. The bond strength is the same in the mip formation and analyte binding phase $\tilde{K}_D = K_D$. For divalent templates with 2 equal receptors the optimal imprinting is achieved at ligand concentration $c^* = 2.38K_D^*$ and template concentration $C_T^* = 1.69K_D^*$, regardless of the matrix stiffness k_h. If the



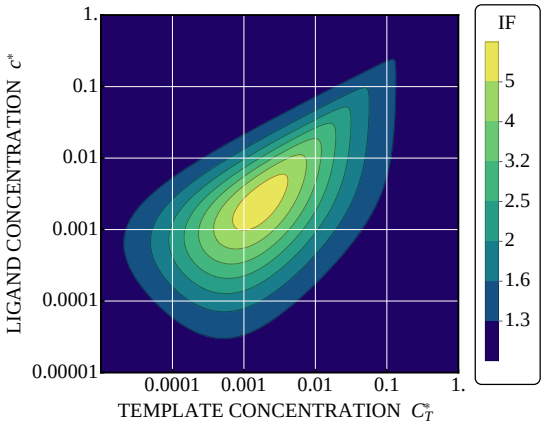
(a) $k_h^* = 100$, $K_D^* = 0.001$



(b) $k_h^* = 100$, $K_D^* = 0.001$, different bond types



(c) $k_h^* = 100$, $K_D^* = 0.01$



(d) $k_h^* = 10$, $K_D^* = 0.001$

Fig. A2 Imprinting factor $IF = \frac{BA_2^{mip}}{BA_2^{nip}}$ for divalent templates as a function of ligand and template concentrations at fixed matrix stiffness k_h^* and bond strength K_D^*. We use rescaled units, described in the main text, such that the phase diagram is valid for any template size. The imprinting factor is calculated using (A55) and (A64). The landscape in **a)** is the same as on Figure 4.4**c)** of the main text. In **b)** we show the imprinting factor for 2 different bonds calculated with (A56, A60,A61,A63), but otherwise identical parameters as **a)**, the ligand concentration axis denotes the total ligand concentration $\frac{c^*}{2} = c_\alpha^* = c_\beta^*$ and both bonds are of equal strength K_D^*. **c)** and **d)** show the effect of changing the bond strength or matrix stiffness relative to **a)**.

two receptors on the template are different (and there is no cross binding) the highest imprinting factor is obtained at the same values, but applied to each type separately: $c^*_\alpha = c^*_\beta = 2.38K^*_D$ and $C^*_T = 3.38K^*_D$.

In summary we make 5 points:

1. We observe that ligands and templates should be in approximately stoichiometric ratio (2 ligands to 1 template for divalent templates).

2. Further on, the ligand concentration should be similar to the ligand-receptor dissociation constant, as discussed in the main text. If the concentration is too low, the MIP formation is inefficient, while for high concentrations analytes (templates) bind strongly already to non-imprinted polymers NIPs.

3. We also observe that using 2 different ligand-receptor types does not bring any benefit over a single type (comparing Figures A2**a)** and **b)**). For example, keeping the overall ligand concentration the same and using 2 distinguishable bonds decreases the binding affinity to NIPs by a factor between 2-4 (A55,A56). However, it also reduces the binding affinity to MIPs (by a factor 2, comparing Eqs. A62 and A63) and it decreases the occupancy fractions in the MIP formation phase (A7). The final result seems to be that using distinguishable ligand-receptor pairs leads to the same imprinting efficiency if the overall ligand and template concentrations are increased by a factor 2.

4. Increasing the bond strength increases the imprinting factor (comparing Figures A2**a)** and **c)**).

5. Changing the matrix stiffness k^*_h does not influence the shape of the phase diagram, it merely rescales the imprinting factor (comparing Figures A2**a)** and **d)**), stiffer matrices offer higher imprinting factors.

A.4.1 Separation factor optimization

Two analytes (say $b_1 = \sigma$ and $b_2 = 0.8\sigma$) can be separated if the matrix is imprinted with one of them (say $a = b_1$). Figure A3**a)** depicts the separation factor as a function of the imprinted distance a for three different values of matrix stiffness (binding free energies are plotted on the inset). Only for very stiff matrices the optimal value is $a^{opt} \approx b_1$, while for softer gels the values are around $a \approx 1.5 > b_1$. This result can be intuitively understood by noting that the binding free energy is approximately a quadratic function of the mismatch close to the minimum. If the imprinting is slightly mismatched, the

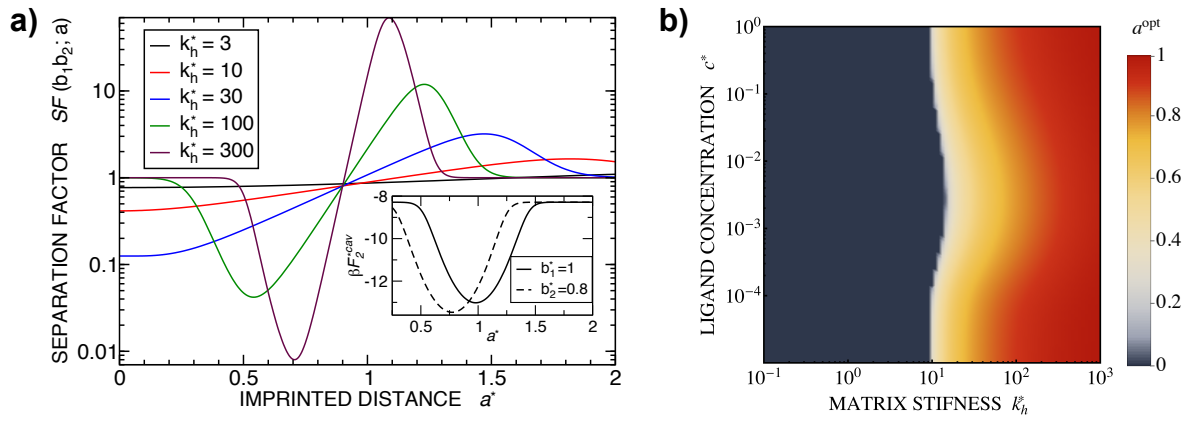


Fig. A3 **Enhancing the specificity. a)** The separation factor for two particles with sizes $b_1^* = 1$ and $b_2^* = 0.8$ as a function of the imprinted distance in the cavity a^* for different values of the gel stiffness k_h^*. The best separation is achieved when the imprinting is slightly mismatched relative to the analyte. The specific part of the binding free energy $\beta F_2^{cav}(a)$ is shown in the inset for each particle. $K_D^* = 0.001$ and we only considered binding to fully functionalized cavities. **b)** Optimal template choice for separation. Landscape of the optimal inter-receptor distance on the template a^{opt}/σ when we wish to separate an analyte with $b_1^* = 1$ from another analyte $b_2^* = 1.1$ (used for *SF* landscape shown on Figure 4.5**b)**).

binding affinity of the chosen analyte is slightly smaller, but at the same time it increases relative to the binding affinity of the other particle resulting in better separation capacity of the MIP.

On Figure 4.5 in the main part we show separation factors between 2 divalent analytes that have a different inter-receptor distance: $b_1 = \sigma, b_2 = 1.1\sigma$. On Figure 5**a)** we have assumed that matrix was imprinted with a template with interceptor distance $a = \sigma$, while on Figure 5**b)** the choice of a template (a^{opt}) was optimized to obtain the largest possible separation factor between the 2 analytes. Here (Figure A3**b)**) we show the optimal a^{opt} yielding that separation factor. For stiff matrices (large k_h^*) the optimal imprinted distance is close the analyte parameter $a^{opt} \simeq b_1$ and this analyte can be effectively also used as a template in imprinting process. In the case of soft matrices, on the other hand, the optimal imprinted distance is lower. Therefore, for efficient separation, a template different then the analyte should be considered.

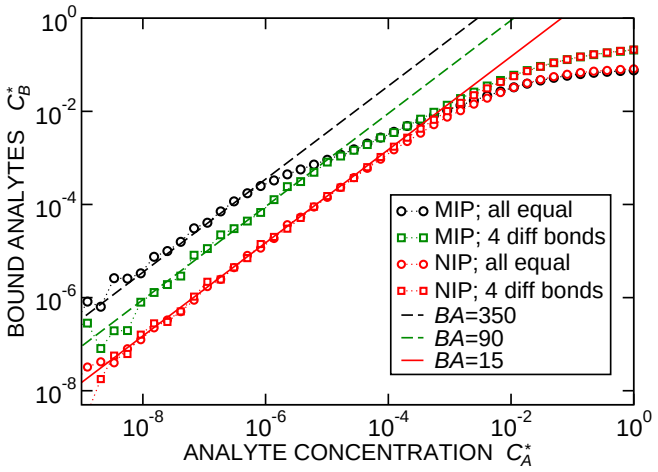


Fig. A4 Binding isotherms obtained from simulations for binding of 4 receptor templates to NIPs (red) and MIPs (black and green). The 4 binding sites are either all equal (circles) or there are 4 distinguishable ligand-receptor pairs (squares), there is no cross binding. Solid red line represents the analytical result (A55), dashed black and green lines are fitted to simulation results in the linear regime. Parameters: $K_D^* = 0.1$, $k_h^* = 100$, $C_T^* = 0.025$, $c^* = 0.1$. In the case of different bonds each ligand type has the same concentration $c_i^* = 0.1$.

A.4.2 Binding affinities from simulations

In Figure 4.3 we showed binding isotherms of divalent analytes into imprinted and non-imprinted polymers. On Figure A4 we also show the binding isotherms for analytes with 4 binding sites (receptors) arranged on a particle on the vertices of a square along a great circle (see Figure A5 below, or Figure 4.3). Even though individual bonds are very weak $K_D^* = 0.1$, the resulting imprinting factors are substantial $IF > 20$. Using 4 distinguishable ligand-receptor pairs does not seem to offer increased selectivity, in fact imprinting factor is decreased in the case shown.

A.5 Enantiomeric separation

Figure A5 illustrates how the binding free energy to a fully functionalized cavity depends on the spatial distribution of the binding sites. The figure shows the results of MC simulations for particles and cavities that have four distinct bonds arranged at the corners of a square. The ligands $\alpha\beta\gamma\delta$ are arranged clockwise. We compare binding of two versions of particles: the matching particle has binding sites $\alpha'\beta'\gamma'\delta'$ arranged clockwise as well, while on the non-matching particle the order is switched to $\alpha'\beta'\delta'\gamma'$. Clearly, the matching particle shows greater affinity to the cavity and by choosing an appropriate regime of the chemical potential μ the two could be effectively separated. This particular example can be viewed as prototypical for enantiomeric separation. Interestingly, non-matching analytes bind strongest to the intermediate matrix stiffness (when the analyte is still able to deform the matrix and satisfy all bonds), while matching analytes (templates) bind more strongly to stiffer matrices.

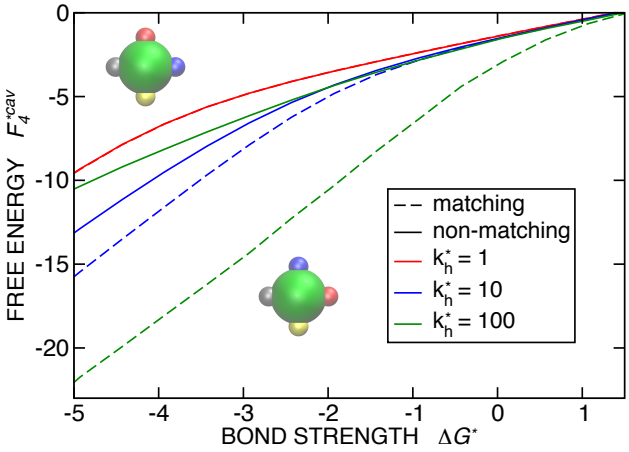


Fig. A5 **Enantiomer separation.** Specific binding free energy βF_4^{*cav} for analyte particles of the same size and four binding sites with either a matching (dashed lines) or non-matching (solid lines) spatial distribution. The matching particle can readily form four bonds, while the non-matching particle with two of the bonds reversed can form all bonds only if the polymer matrix is heavily distorted. Different curves are for different polymer gel stiffness k_h^*. At $k_h^* = 1$ the two curves overlap.

Appendix B

List of Publications

Some of the work included in this thesis has been published in the following papers:
Chapter 1

- T. Curk, J. Dobnikar and D. Frenkel, Design principles for super selectivity using multivalent interactions, arXiv preprint: https://arxiv.org/abs/1608.01222
- N. W. Schmidt=, F. Jin=, R. Lande=, T. Curk=, W. Xian, C. Lee, L. Frasca, D. Frenkel, J. Dobnikar, M. Gilliet, and G. C. L. Wong, *Liquid-crystalline ordering of antimicrobial peptide-DNA complexes controls TLR9 activation*, Nat. Mater. 14, 696 (2015); (= indicates shared first-author contributions)
- G. V. Dubacheva, T. Curk, R. Auzely-Velty, D. Frenkel, and R. P. Richter, *Designing multivalent probes for tunable superselective targeting*, Proc. Natl. Acad. Sci. U.S.A. 112, 5579 (2015)
- G. V. Dubacheva, T. Curk, B. M. Mognetti, R. Auzely-Velty, D. Frenkel, and R. P. Richter, *Superselective targeting using multivalent polymers*, J. Am. Chem. Soc. 136, 1722 (2014)

Chapter 3

- T. Curk, J. Dobnikar, and D. Frenkel, *Rational design of molecularly imprinted polymers*, Soft Matter 12, 35 (2016)

During my PhD studies I have also contributed to the following papers:

- E. Y. Lee, C. K. Lee, N. W. Schmidt, F. Jin, R. Lande, T. Curk, D. Frenkel, J. Dobnikar, M. Gilliet, and G. C. L. Wong *A review of immune amplification via*

ligand clustering by self-assembled liquid-crystalline DNA complexes, Adv. Colloid Interface Sci. (2016)

- J. Wei, J. Dobnikar, T. Curk, and F. Song *The Effect of Attractive Interactions and Macromolecular Crowding on Crystallins Association*, PloS one 11, e0151159 (2016)
- T. Curk, F. J. Martinez-Veracoechea, D. Frenkel, and J. Dobnikar, *Nanoparticle organization in sandwiched polymer brushes*, Nano lett. 14, 2617 (2014)
- R. De Gernier, T. Curk, G. V. Dubacheva, R. P. Richter, and B. M. Mognetti, *A new configurational bias scheme for sampling supramolecular structures*, J. Chem. Phys. 141, 244909 (2014)