Electronic Supplementary Material (ESI) for Soft Matter.

Predicting the size and morphology of nanoparticle clusters assembled by biomolecular recognition

P. Palacios-Alonso,[†] E. Sanz-de Diego,[†] Raúl P. Peláez,[‡] A. L. Cortajarena,[¶,§] F. J. Teran,[†] and Rafael Delgado-Buscalioni[*,⊥]

[†]*iMdea Nanociencia, Campus Universitario de Cantoblanco, 28049 Madrid, Spain*

[‡]*Dpto. Física Teórica de la Materia Condensada, Universidad Autónoma de Madrid, 28049 Madrid, Spain*

[¶]*CIC biomaGUNE-BRTA, 20014, Donostia-San Sebastián, Spain*

[§]*Ikerbasque, Basque Foundation for Science, Bilbao, Spain*

[‖]*Nanobiotecnología (iMdea-Nanociencia), Unidad Asociada al Centro Nacional de Biotecnología (CSIC), 28049 Madrid, Spain*

[⊥]*Dpto. Física Teórica de la Materia Condensada, Universidad Autónoma de Madrid, 28049 Madrid, Spain. Institute of Condended Matter (IFIMAC).*

E-mail: rafael.delgado@uam.es

Supporting information

S.1. Bethe theory of percolation

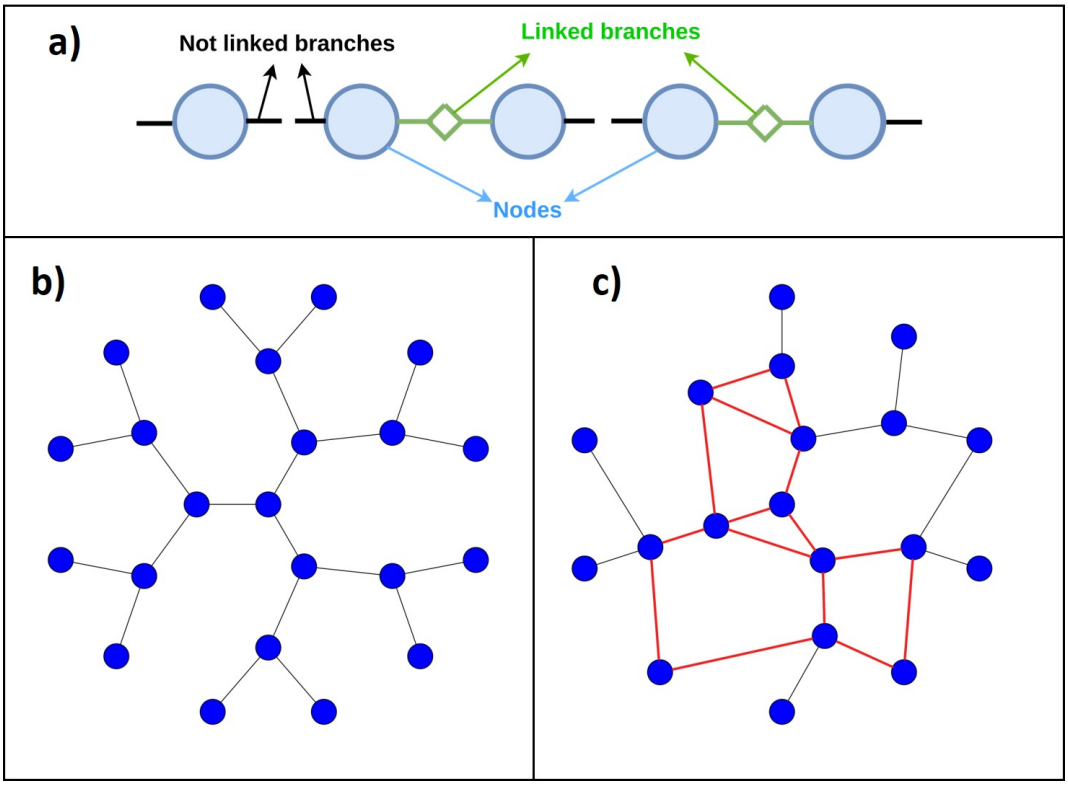

Figure S1: a) Representation of a 1D (f=2) Bethe lattice, each node (blue circles) have 2 branches, that can be formed a link (green) or not (black). b) Bethe lattice with functionality f = 3 c) Real lattice with some cyclic structures remarked in red.

The percolation theory studies the behavior of a network of interconnected nodes that can be linked with each other (see Figure S1 a,b). It describes the phase transition between the state where in the network there are some small clusters not connected between them, when the number of links are small, and the state where most nodes are connected in an infinite cluster (gel phase). In most real lattices (Figure S1 c) it is not possible to make analytical predictions, however in Bethe's lattice[1] it is possible to obtain some analytical results about the distribution of clusters, the gelation probability, the cluster size, etc. For more details on the following derivations the reader is refereed to Rubinstein and Colby textbook.[2]

A Bethe lattice is defined as an infinite connected cycle-free graph where each node has a fixed number of branches f, that can be forming link between to nodes with a probability p. We can use the cycle-free assumption to obtain a prediction for the mean number of neighbors that a node has in a lattice composed by s particles. In a cluster with s particles the number of formed bonds will be $s-1$, consequently on average each particle will be connected to,

$$n_{neigh} = 2\frac{s-1}{s} \tag{1}$$

Where the 2 arises from the fact that each bonds connects two particles.

Percolation in 1-D

The simplest case that we can study is a Bethe lattice with f = 2 (Figure S1 a). In this case the clusters will be one dimensional chains composed by a given number of nodes.

Every chain, independently of its size, will necessarily have two free branches, and all the other branches will be occupied. In fact, a cluster of size s, will have s occupied branches and 2 free-ones. If the probability that a link is formed is p, the distribution of chains of size s, will be,

$$n_s(p) = p^s(1-p)^2 \tag{2}$$

In a 1-D Bethe lattice, the only way to observe an infinite cluster is that all the nodes were linked, consequently the phase transition will occur at a probability $p = p_c = 1$. If we define $\xi = -\frac{1}{\ln(p)}$, we can rewrite equation X as a function of ξ and p_c as,

$$n_s(p) = (p_c - p)^2 \exp(-s/\xi) \tag{3}$$

where ξ is the connectivity length of the clusters. The Taylor expansion of $\ln(p)$ near p_c, is,

$$\ln(p) \approx \ln(1) + \frac{1}{p_c}(p - p_c) = -|p - p_c| \tag{4}$$

Thus, near the phase transition the connectivity length diverges as a power law,

$$\xi = \frac{1}{|p - p_c|} \tag{5}$$

Percolation in a general Bethe lattice

All the results previously obtained can be generalized to a Bethe lattice with an arbitrary number, f of neighbors per node. If two nodes can be linked with a probability p, it is not difficult to see that the mean number of first neighbors to the central node is $C_1 = fp$, the number of connections to the second generation $C_2 = p(f-1)C_1$ and similarly, the number of connections of the N generation is $C_N = (p(f-1))^N C_1$. When N tends to infinity, this number diverges if $p(f-1) > 1$, which provides the critical value of the binding probability $p = p_c$ for the formation of a gel phase (infinite cluster),

$$p_c = \frac{1}{f-1} \tag{6}$$

Note that if f=2, this condition implies a very strong binding affinity ($p_c \to 1$), meaning the formation of an infinitely large linear chain. Using geometric arguments, it can be seen that a clusters of size $s > 0$ (i.e. having s occupied nodes) will have $s_f = 2 + s(f-2)$ free nodes (e.g. for f=2, this corresponds to the two free nodes in the chain ends). The fraction of clusters of size s follows a Binomial distribution,

$$n_s = g_s p^s (1-p)^{2+s(f-2)} \tag{7}$$

Where g_s is the number of combinations of placing s_f nodes in the cluster of size s, i.e,

$$g_s = \frac{f(fs-s)!}{s!(fs-2s+2)!} \tag{8}$$

We can calculate the connectivity length writing $n_s(p)$ as a function of $n_s(p_c)$,

$$n_s(p) = n_s(p_c)\frac{1-p}{1-p_c}^2 \left(\frac{p(1-p)^{f-2}}{p_c(1-p_c)^{f-2}}\right)^s \tag{9}$$

which can be also written as,

$$n_s(p) \propto \exp(-s/\xi) \tag{10}$$

with ξ defined as,

$$\xi(p) = \frac{-1}{\ln\left[p(1-p)^{(f-2)}\right] - \ln\left[p_c(1-p_c)^{(f-2)}\right]} \tag{11}$$

We can expand in Taylor series the logarithm near p_c, to obtain how the connectivity length diverges near p_c. Let,

$$h(p) = \ln\left(\frac{p(1-p)^{f-2}}{p_c(1-p_c)^{f-2}}\right) \tag{12}$$

then:

$$\frac{d}{dp}h(p) = \frac{1}{p} - \frac{f-2}{1-p} \tag{13}$$

$$\frac{d^2}{dp^2}h(p) = -\frac{1}{p^2} - \frac{f-2}{(1-p)^2} \tag{14}$$

$$\tag{15}$$

Evaluating h(p) and their derivatives in $p = p_c$, we obtain $h(p_c) = \frac{dh}{dp}(p_c) = 0$ and $\frac{d^2h}{dp^2}(p_c) =$, and substituting $f - 2 = 1/p_c$ (eq 31) we obtain,

$$h(p) \approx -\frac{1}{2}\frac{(p-p_c)^2}{p_c^2(1-p_c)} \tag{16}$$

Finally, we obtain that ξ diverges near p_c as,

$$\xi \propto |p - p_c|^{-2} \tag{17}$$

From equation X and substituting the expression of g_s we can prove that n_s decays as a power law,

$$n_s \propto s^{-\tau} \tag{18}$$

With $\tau = 5/2$. That value can be obtained by substituting equation X and X +2 in equation X+1, taking logarithms on both sides of the equation and using the Stirling approximation $\ln(n!) \approx n\ln(n) - \sqrt{n}\ln(2\pi n)$.

S.2. Cluster sizes and Flory exponents

We use Eq. 6 to fit the results for the cluster gyration radius against the number of particles in the cluster s. Results are collected in Table S1 and Figure S2.

Table S1: Coefficients used to fit the cluster gyration radius using Eq. 6 for several values of the average number of receptors per IONP $\bar{f}$. The Flory exponent (large cluster limit) is ν_∞.

$\bar{f}$	c_1	c_2	c_3	ν_∞
2	9.75	0.45	24.29	0.500 ± 0.007
3	22.01	0.85	30.57	0.480 ± 0.004
4	22.29	0.85	30.57	0.480 ± 0.003
8	23.45	1.2	29.08	0.492 ± 0.005
12	21.24	1.1	28.50	0.493 ± 0.004

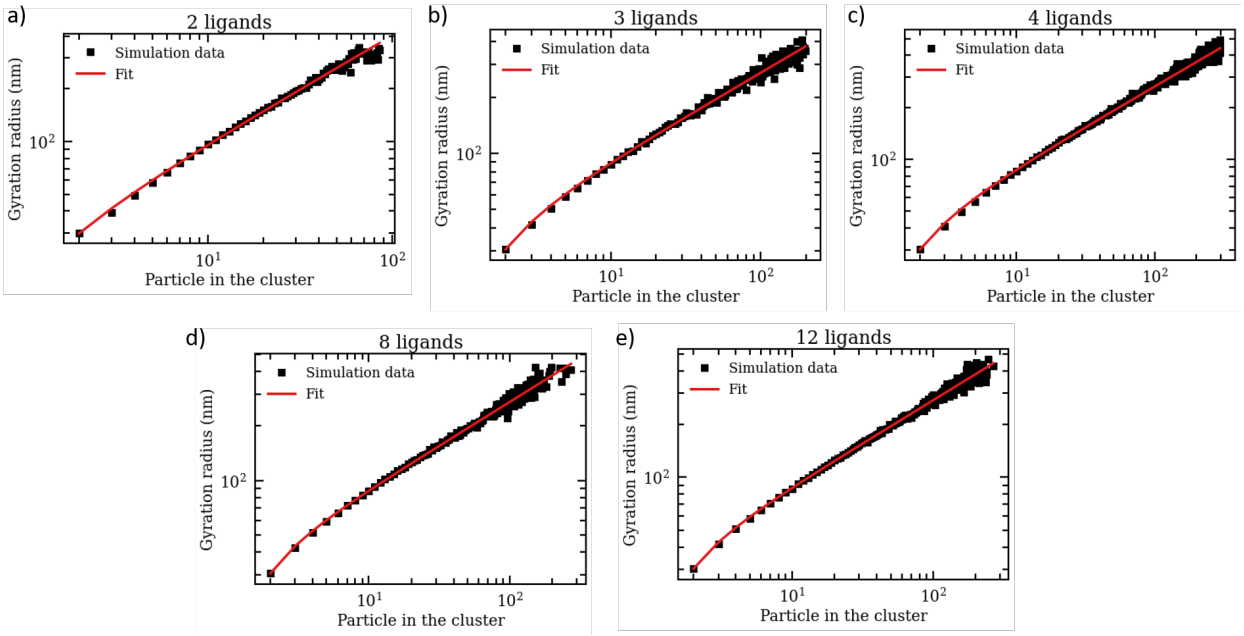

Figure S2: Gyration radius of NMP clusters against cluster size. Fits to Eq. 6 are shown in red.

S.3. Cluster size distributions

We present extra results for $n(s)$, the distribution of clusters of mass s and recall that in Bethe's theory $n(s)$ is normalized according to

$$\sum_{s=1}(s-1)n(s) = \frac{n_{bonds}}{fN} \tag{19}$$

where n_{bonds} is the number of bonds. Under the assumption that bonds are irreversible the number of bonds is equal to the number of analytes if the system is not saturated of analytes ($c_a < \frac{f}{2}c_p$) and to the number of initially-free receptors, if $c_a > \frac{f}{2}c_p$. Note that the factor $(s-1)$ in the left side of the equation 19 corresponds to the number of links in a cluster with s particles (in the original formulation of the Bethe model, s denotes the number of links). The cluster distribution can be calculated from the list of clusters as,

$$n(s) = \frac{N^s_{cluster}}{fN} \tag{20}$$

Where $N^s_{cluster}$ is the number of clusters composed by s particles. Extracting the distribution of clusters requires a large number of aggregates, which involve much more computational time than determining the mean hydrodynamic size of the system. We repeated each run between 10 and 150 times, depending on how fast $n(s)$ tends to 0 (distributions that decay more slowly require longer statistics). Typically the total number of analyzed clusters was about tens of thousands.

Results for $n(s)$ obtained from simulations of IONPs with average functionality $\bar{f} = \{8, 12\}$ are shown in Figure S3. Red lines correspond to Eq. 7 using the effective functionality f_{eff} indicated in the legend. Blue lines are the distributions for a Bethe lattice with fixed $f = \bar{f}$.

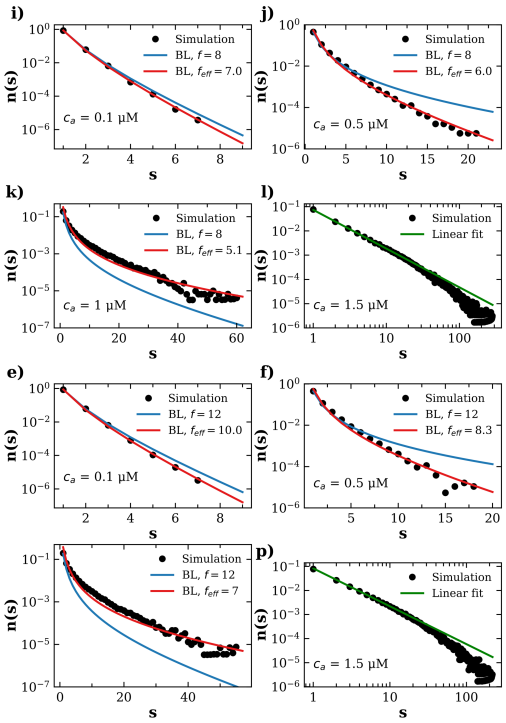

Figure S3: Cluster distributions $n(s)$ (proportional to the number of clusters of size s) for $\bar{f} = 8$ (a-d) and $\bar{f} = 12$ (e-h) at increasing analyte concentration. Nanoparticle concentration fixed at $c_p = 1.27$ μM.

S.4. Receptor distribution

Results for the statistics of the number of receptors per nanoparticle are shown in Fig. S4, comparison is made between a Poisson distribution, a binomial distribution and random sampling.

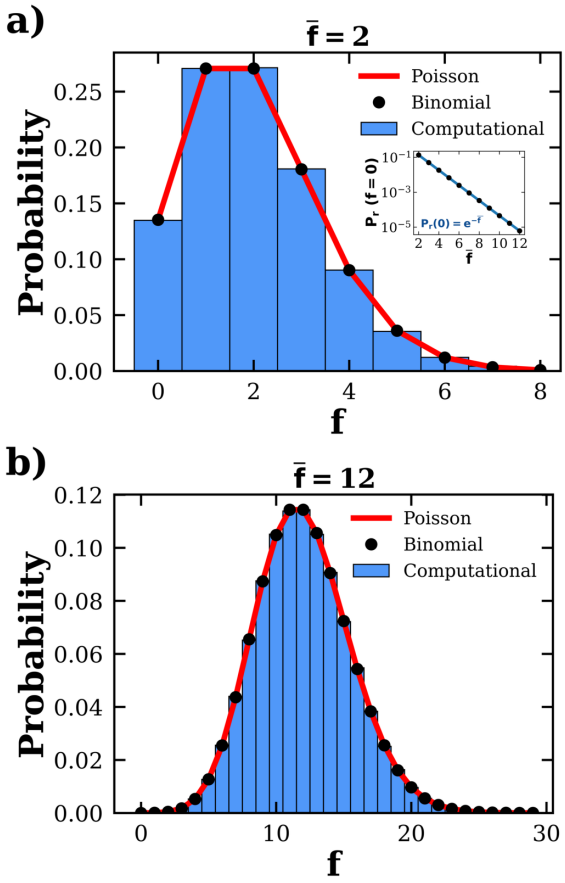

Figure S4: Distribution of receptors on the IONP: comparison is made between a Poisson distribution, a binomial distribution and random sampling (see text). a) $\bar{f} = 2$ and b) $\bar{f} = 12$. The inset in (a) shows the probability of having particles without receptors ($f = 0$).

S.5. Distribution of occupied receptors

In order to obtain an analytical equation to predict the probability of a particle having f_{occ} receptors in a system with a particle concentration of c_p, an analyte concentration of c_a, and a mean number of receptors per particle $\bar{f}$, we use:

$$P(f_{occ}) = \sum_{f=f_{occ}}^{\infty} P_r(f)\ P(f_{occ}|f) \tag{21}$$

Where $P_r(f)$ is given by equation 2 of the main text and $P(f_{occ}|f)$ is the probability that a particle with f receptors has f_{occ} occupied receptors. This probability can be calculated using the notion that the probability that a receptor is occupied is equal to the number of analytes in the system divided by the total number of receptors in the system, $p_{occ} = c_a/(\bar{f}c_p)$. If we assume that both the number of analytes and receptors are large enough such that this probability does not change after the formation of one bond, then the probability $P(f_{occ}|f)$ is given by a binomial distribution:

$$P(f_{occ}|f) = \binom{f}{f_{occ}} p_{occ}^{f_{occ}}\ (1-p_{occ})^{f-f_{occ}} \tag{22}$$

Substituting equations S22 and 2 into equation S21, and then reordering the terms, we obtain:

$$P(f_{occ}) = \frac{p_{occ}^{f_{occ}}\ e^{-\bar{f}}}{f_{occ}!} \sum_{f=f_{occ}}^{\infty} \frac{(1-p_{occ})^{f-f_{occ}}}{(f-f_{occ})!}\ (\bar{f})^f \tag{23}$$

Multiplying and dividing by $(\bar{f})^{f_{occ}}$ results in:

$$P(f_{occ}) = \frac{(\bar{f}\ p_{occ})^{f_{occ}}\ e^{-\bar{f}}}{f_{occ}!} \sum_{f=f_{occ}}^{\infty} \frac{(1-p_{occ})^{f-f_{occ}}}{(f-f_{occ})!}\ (\bar{f})^{f-f_{occ}} \tag{24}$$

Finally, by recognizing that the term inside the summation is equal to the exponential of $\bar{f}(1-p_{occ})$ and using the fact that $p_{occ} = c_a/(\bar{f}c_p)$, we obtain the final expression:

$$P(f_{occ}) = \frac{e^{-c_a/c_p}}{f_{occ}!}\left(\frac{c_a}{c_p}\right)^{f_{occ}} \tag{25}$$

Figure S5 shows the distribution of occupied receptors in two cases $c_a = c_p$ and in a media saturated with analytes $c_a = 5c_p$.

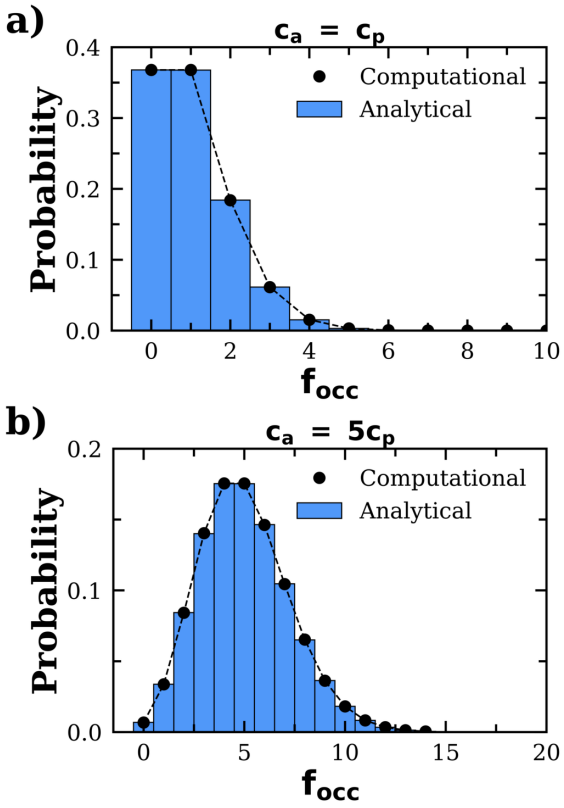

Figure S5: Distribution of occupied receptors (primary links) on the IONPs obtained using Eq. 12 and random sampling when: a) the concentration of particles is equal to the concentration of analytes $c_a = c_p$ and b) under saturation $c_a = 5c_p$.

S.6. Magnetization field (MH) curves of the IONPs

The MH curves of the nanoparticles where measured in quasi-static conditions at 4K and 300K using a SQUID. IONPs present a superparamagnetic behaviour (in absence of field the system is not magnetized) and hence in absence of field the magnetic interactions between IONPs are absolutely negligible.

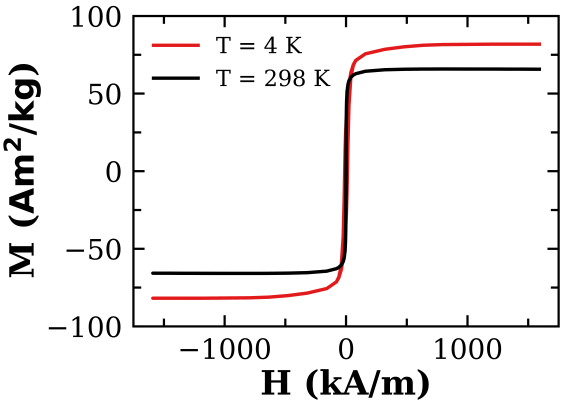

Figure S6: Mass normalized M-H curves of the nanoparticles employed in the experiments measured with a SQUID at 4K and 300K under quasi-static conditions

S.7. Number of formed bonds as a function of time

With the purpose of ensuring that we were running the simulations for a time long enough to reach the final state of the aggregates we measured the number of links formed in the system as a function of time. In Figure S7 we can see that for all the concentrations of analytes we reached a plateau in the number of bonds, that means that all links that can be formed, are formed during the simulation, so we are reaching the final state.

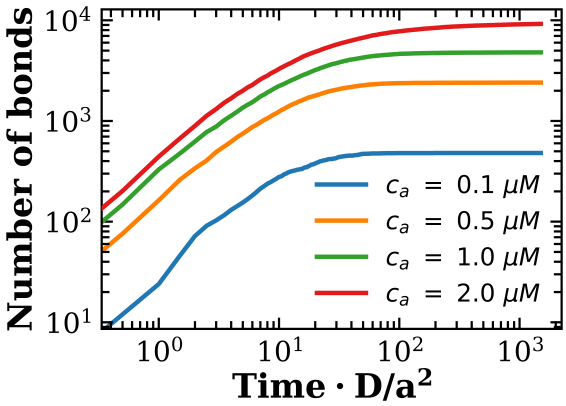

Figure S7: Evolution of the number of bonds created in a simulation over time, for different values of analytes concentration and $\bar{f} = 4$ receptors per particle.

S.8. Schematic of two bonded IONPs

The interaction between two bonded IONPs is modelled using a harmonic interaction that causes a force and a torque on each particle that depends on the parameters indicated in Figure S8.

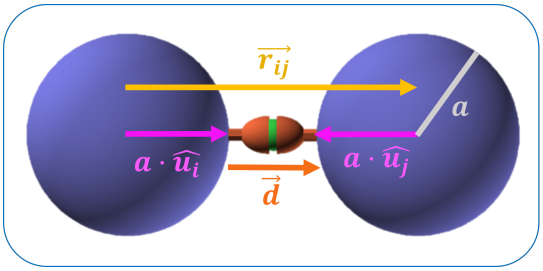

Figure S8: Geometrical parameters involved in the calculation of the bonding forces and torques. The receptor length is $\ell \approx 2$ nm.

References

(1) Bethe, H. A. Statistical theory of superlattices. *Proc. R. Soc. Lond. A 150*, 552–575.

(2) Rubinstein, M.; Colby, R. *"Polymer physics"*; Oxford University Press, 2003.