More efficient process chromatography
Bosma, Jacob Catrinus

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2001

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 28-02-2019
5

The Effect of Protein Size on Adsorption Equilibria
Abstract

This paper presents a new isotherm for adsorption of proteins: the available area isotherm. This isotherm considers the effect of size and geometric exclusion on adsorption. The single component version is similar to the steric mass action isotherm, the most satisfactory isotherm until now, but it is fundamentally more correct and has one less fit parameter. The multicomponent version predicts a strong exclusion of large molecules by smaller ones.

Introduction

An overview of adsorption isotherms of proteins in liquid chromatography is given by Bellot and Condoret (1993). Generally the equilibrium is described by an adsorption reaction. The Langmuir isotherm is the most well known example and it is probably the most widely used isotherm for protein adsorption. In the derivation of this isotherm it is assumed that a molecule adsorbs on an adsorption site on the surface or, e.g. in the case of ion exchange, that it exchanges with an adsorbed molecule. In the mass action or stoichiometric displacement model (Kopaciewicz, 1983) this approach is extended to the exchange of a protein with a certain number of small ions, this number is called the binding or apparent charge of the protein.

There are some problems with the mass action model. When the mass action model is extended to competitive equilibria of proteins it can easily lose its consistence. This is because the small ion capacity that is fitted for different proteins will not be the same. The steric mass action formalism (Brooks and Cramer, 1992) counters this by simply assuming that the small ion capacity is equal for all proteins: each protein exchanges with a number of small ions and it shields a number of small ions. The number of shielded small ions is usually at least an order of magnitude higher than the number of exchanging small ions. Another problem is the fact that there are various ways of defining the mole fraction of the adsorbed proteins. Should all the adsorbed small molecules be used in the definition, or only the ones that are not shielded, or only the maximum amount that can be replaced by proteins at the same time? A third problem is the extra fitting variable in the mass action isotherm compared to the Langmuir isotherm: the binding charge. The binding charge has an effect on the shape of the isotherm but it is also the variable that accounts for the effect of the ionic strength on the equilibrium. It is usually measured by performing pulse response experiments at low binding strength. However there is no reason to assume that this binding charge is also valid at high binding strength.
Multicomponent equilibria of proteins on ion exchangers have been measured by Skidmore and Chase (1990a,b) and by Garke et al. (1999). In both cases the single component isotherm can be described by a Langmuir isotherm, but the multicomponent Langmuir isotherm fails to describe the competitive equilibrium. In both cases the competitive isotherm underestimates the adsorption of the small protein. Garke et al. (1999) did use an isotherm that could describe the competitive behavior. However the formalism of the isotherm was not consistent with the assumptions (in their equation they assume that the adsorbed concentration of the small protein at the surface available for both proteins equals that at the surface available for the small protein alone).

None of the isotherms mentioned above considers the effect of the protein size and geometrical exclusion. In this paper we propose a new type of isotherm, the available area isotherm, which takes this effect into account. This article first presents the derivation of the isotherm. Then the single component version is discussed by comparing it with the mass action isotherm and the multicomponent isotherm by comparing it with measurements. Finally the conclusions are summarized.

Derivation of the available area isotherm

Concepts and assumptions

The model considers the proteins as disks or cylinders. Each protein has its own radius; other geometrical parameters are not required. The surface of adsorption is considered to be a flat surface with homogeneous binding strength. The disks are distributed randomly at this surface but they are not allowed to overlap. The assumption of random distribution implies that the adsorbed proteins show negligible repulsive or attractive interactions. This is shown in figure 1 for a small part of a surface with two different proteins. Each protein covers a certain fraction of the surface; we give this coverage the symbol $\chi$. ($\chi$ is the Greek letter ‘c’: for coverage). New disks approaching the surface cannot overlap those on the surface. So they only adsorb if their center is at least one radius away from all neighbors. This leads to ‘excluded areas’: these are shown for both protein ‘1’ and the larger protein ‘2’. The fraction of the area that is excluded has the symbol $\varepsilon$ (for excluded). Figure 1 shows that the excluded area is much larger for the larger protein ‘2’ than for ‘1’. The excluded areas show a considerable overlap, even for the sparse coverage shown. The white space left is ‘available area’: this is much larger for the smaller protein than for the larger one. Here the symbol is $\alpha$ (available).
Regular hexagonal packing of the disks gives the highest coverage of the surface. In figure 2 each disk occupies a hexagonal cell: the area of the disk is 0.907 of that of the cell, this is also the maximum coverage, $\chi_{\text{max}}$.

The mathematical derivation starts with the same assumption that is done in the derivation of the Langmuir equation: at equilibrium the adsorption rate and the desorption rate are equal. Like in the Langmuir equation the desorption rate is proportional to the surface coverage. The adsorption rate is proportional to the concentration in the liquid and, unlike in the Langmuir equation, the available area. We obtain for protein '1':

$$k_{\text{ads,1}}c_1\alpha_1 = k_{\text{des,1}}\chi_1 \Rightarrow c_1 = K_1 \frac{\chi_1}{\alpha_1}$$

(1)

Here the $k$’s are rate constants, $c$ is the unadsorbed concentration and $K$ is an equilibrium constant.

In the next paragraphs we derive expressions for the available area. There are two different sets: those for low surface coverage and those for high surface coverage. In the first case, for low surface coverage, a small excluded area is distributed randomly over the available adsorption area. In the second case, for high surface coverage, a small available adsorption area is distributed randomly over the otherwise excluded area. They will be combined using an interpolation formula.
The available area at low surface coverage

At low surface coverage we can use a model that is analogous to the model that we used for describing the concentration dependence of the steric exclusion of proteins from fibrous structures (Bosma and Wesselingh, 2000). It is easy to calculate the area that one protein will exclude for the others. However for even slightly higher concentrations this becomes a problem. Simply adding the excluded areas gives too high an estimate of the excluded area because excluded areas can overlap. Adding the excluded areas and subtracting the random overlap will give too low an estimate because excluded areas do not overlap entirely randomly. This is because there cannot be overlap between adsorbed proteins. In appendix A an equation for the available area is derived with a cell model. The result for protein ‘1’ is:

$$\alpha_{1,\text{low}} = \frac{4\chi_1 + \left(1 + \frac{r_1 + r_2}{r_2}\right)^2 \chi_2}{\chi_1 + \chi_2}$$

(2)

The available area at high surface coverage

At high surface coverage the cell model that we used above breaks down. Here we try a different kind of model. It is based on the ‘Free Volume Theory’ (Cohen and Turnbull, 1959), a theory to estimate the size distribution of holes between spheres positioned at random points in space. We begin with a closely packed array of disks at a surface, and expand this. This is shown in figure 3. Due to the expansion the holes between the disks become larger. We take their size as that of the circle that can be inscribed in the hole. (This model is not suitable for low coverage as shown in the right part of the figure. There the inscribed circles begin to overlap and the idea of holes breaks down.)
Our modification of the free volume theory (which might be called the ‘Free Area Theory’) is worked out in Appendix B. It predicts that the available surface for protein 1 is given by:

\[
\alpha_{1,\text{high}} = n_h S_{\text{free}} \exp \left( -\frac{S_i}{S_{\text{free}}} \right) + n_h \sqrt{\pi} S_{\text{free}} \cdot \text{erf} \left( \frac{S_i}{\sqrt{S_{\text{free}}}} \right) - 1 \]  

(3)

Here \( S_i \) is the area covered by one protein ‘1’ (= \( 4\pi r_i^2 \)), \( n_h \) is the number of holes per surface area and \( S_{\text{free}} \) is the average free area per hole. Definitions of these parameters are given in appendix B.

**Figure 3**  Holes formed in an expanding array of disks

**Obtaining the general isotherm**

We obtain the general isotherm by interpolating between equations 2 and 3. Since the equations can differ over an order of magnitude we take a logarithmic interpolation of the available area. For protein ‘1’ this gives:

\[
\alpha_i = \exp \left( \left( 1 - \frac{\chi_1 + \chi_2}{\sigma \chi_{\text{max}}} \right) \ln (\alpha_{i,\text{low}}) + \frac{\chi_1 + \chi_2}{\sigma \chi_{\text{max}}} \ln (\alpha_{i,\text{high}}) \right) \]

(4)

Here \( \sigma \) is a number (\( \geq 1 \)) that accounts for the increase in the maximum possible surface coverage when the proteins have different sizes. An equation for \( \sigma \) is given in appendix C. Figure 4 gives a graphical illustration of the interpolation. Figure 5 shows the effect of the protein size on the available area: larger proteins have much less available area.
Figure 4  Available area for adsorption versus surface coverage for a single protein

Figure 5  Available area for ‘1’ on a surface covered with ‘2’
Discussion

The single protein isotherm

In this paragraph the available area model is extended to a model for ion exchange. Subsequently the similarities and differences between the available area mass action isotherm, the mass action isotherm and the Langmuir isotherm are discussed.

The Langmuir isotherm is given by:

\[ c_i = K_i \frac{\theta_i}{1 - \theta_i} \]  

(5)

Here \( \theta \) (Greek \( q \)) is the surface coverage defined by: \( \theta = q/q_{\text{max}} = \chi/\chi_{\text{max}} \) and \( q \) and \( q_{\text{max}} \) are the adsorbed concentration and the adsorption capacity. The mass action model gives:

\[ c_i = K_i \theta_i \left( \frac{c_{\text{ion}}}{1 - \theta_i} \right)^z \]  

(6)

Here \( z \) is the binding charge, \( c_{\text{ion}} \) is the small ion concentration in solution and \( 1 - \theta_i \) is proportional to the adsorbed small ion concentration.

The available area isotherm considers the effect of the protein concentration on the equilibrium. The effect of other parameters, such as the ionic strength and the pH, can be accounted for with the equilibrium coefficient. Any model for the equilibrium at infinite dilution can be used to determine the equilibrium coefficient. Some theories for the effect of pH and ionic strength are given by Bosma and Wesselingh (1998) and Jönsson and Ståhlberg (1999).

When we extend the available area isotherm with the mass action model for the effect of ionic strength equation 1 will change to:

\[ k_{\text{ads},i} c_i \alpha_i q_{\text{ion}}^z = k_{\text{des},i} \chi_i c_{\text{ion}}^z \Rightarrow c_i = K_i \left( \frac{c_{\text{ion}}}{q_{\text{ion}}} \right)^z \frac{\chi_i}{\alpha_i} \]  

(7)

Here \( q_{\text{ion}} \) is the local small ion adsorption capacity of the surface. In the mass action model it is assumed that the relevant adsorbed small ion concentration is lower when there is more protein adsorbed. In the available area model we assume that it is given by the local small ion concentration, and therefore is independent of adsorbed protein concentration. This seems fundamentally sounder since, when a protein is allowed close to the surface, the small ion concentration that it ‘sees’ is not affected
by already adsorbed proteins! Effectively equation 7 is the same as equation 1 but
with an equilibrium constant that depends on the ionic strength. When the ionic
strength changes both the mass action and the available area mass action isotherms
will not change their shape, only the scaling of the bulk concentration axis will
change.

![Graph](image)

**Figure 6** Comparison of the mass action isotherm (full line) with the
available area mass action isotherm (dotted line) and the Langmuir
isotherm (broken lines). The mass action isotherms were calculated with
$q_{\text{max}} = 100$, $K = 39$, $z = 4$ and $c_{\text{ion}} = 0.5, 0.1$ and 0.02 respectively from
weak to strong binding. The available area isotherms were fitted to the
mass action isotherms with the parameters $q_{\text{max}} = 117$, $K(q_{\text{ion}})^z = 45$ and
$z = 4$. The Langmuir isotherms were fitted with $(q_{\text{max}} = 50, K = 2)$, $(q_{\text{max}} =
86, K = 0.11)$ and $(q_{\text{max}} = 97, K = 0.01)$ respectively from weak to strong
binding.

It has been observed that the mass action model can describe the effect of ionic
strength on the adsorption equilibrium well with parameters that do not depend on
the ionic strength (Bosma and Wesselingh, 1998 and Karst Lewus and Carta, 1999).
In figure 6 we tried to fit the Langmuir and the available area mass action model to
the mass action model. It can be seen that when the Langmuir model is used, both
the equilibrium constant and the adsorption capacity will depend on the ionic
strength. Figure 6 also shows that the available area mass action isotherm behaves
very similarly to the mass action isotherm. Note that the adsorption capacity
according to the available area model is somewhat larger than the corresponding adsorption capacity according to the mass action model.

The available area model can describe equilibria just as well as the mass action model, and it has two advantages over the mass action model:

- The binding charge in the available area model only describes the effect of ionic strength; in the mass action model it also affects the isotherm shape. At constant ionic strength the available area model has one less fitting parameter.
- According to the mass action model the adsorption rate depends on the total number of small ions available for exchange, while in the available area model we argue that adsorption rate depends on the free surface and on the local small ion concentration. The second assumption seems fundamentally more correct.

**The competitive isotherm**

We found two papers in which experimental results on competitive protein adsorption are published. Skidmore and Chase (1990) measured the adsorption of lysozyme and bovine serum albumin (BSA) on S Sepharose FF in a 0.1 mol/L sodium acetate-acetic acid buffer at pH 5. Garke et al. (1999) measured the adsorption of lysozyme and γ-globulin on Streamline SP in the same buffer. Both found that the competitive Langmuir model underestimates the adsorption of the small protein, lysozyme. Here we only discuss the first set of measurements. The second set gave the same results.

The single component isotherms could be fitted with the available area isotherm with the constants in table I. We assumed that the adsorption surface consists of two parts, an area that is accessible for both proteins and an area that is only accessible for the small protein. The total area that is available for each protein is proportional to the adsorption capacity (in g/L) divided by the radius of the protein. Figure 7 shows that the available area isotherm can accurately describe the single component isotherms. This figure also shows that the total adsorption capacity (expressed in covered area, \( \Sigma q/r_{protein} \)) increases slightly when protein mixtures are adsorbed. This may be explained by the tighter packing that can occur when proteins of different sizes adsorb. Figure 8 shows the competitive adsorption isotherms measured by Skidmore and Chase (1990), together with predictions by the available area isotherm. When equal protein sizes are assumed the predictions are similar to those found by others: the adsorption of the large protein is fairly well predicted but the adsorption of the small protein is clearly underpredicted. This indicates that there is an effect of the protein size on the adsorption equilibrium. When the real sizes of the proteins are used the adsorption of the small protein is predicted accurately but the adsorption of the large protein is under predicted, especially at higher surface
coverage. The introduction of the correction term for the increase in the maximum possible surface coverage when the proteins have different sizes, $\sigma$, has hardly any effect.

**Table I**  Parameters in the two component equilibrium calculations.

<table>
<thead>
<tr>
<th></th>
<th>Skidmore <em>et al.</em>, 1990a,b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BSA</td>
</tr>
<tr>
<td>$M_m$ [kg/mol]</td>
<td>66.3</td>
</tr>
<tr>
<td>$r$ [nm]</td>
<td>2.72</td>
</tr>
<tr>
<td>$q_{max}$ [g/L gel]</td>
<td>195</td>
</tr>
<tr>
<td>$K$ [-]</td>
<td>0.07</td>
</tr>
<tr>
<td>$q_{max, single}/q_{max}$</td>
<td>0.32</td>
</tr>
</tbody>
</table>

The protein radii were estimated with $r = \sqrt[3]{\frac{3}{4\pi pN_{av}} \frac{M_m}{\rho}}$ with $\rho = 1300$ kg/m$^3$.

**Figure 7** Measured (Skidmore *et al.*, 1990) and fitted single component adsorption isotherms of lysozyme (▲) and BSA (■). Also shown are the total adsorbed concentrations of the competitive equilibrium ($\Sigma q/r$ vs. $\Sigma Kc/200$, p) (Skidmore and Chase, 1990). The radii used in the calculations are found in table 1.
Many explanations are possible for the underestimation of the adsorption of the large protein:

• The model at high surface coverage is not suitable if the size differences are large.
• The surface is not homogeneous and has preferential adsorption sites for the large and the small protein.
• Lateral interactions (attraction or repulsion) between adsorbed proteins can be important.

More experiments (also experiments in which the large protein adsorbs strongly) are needed before any competitive isotherm can be discarded. Also the available area model may be improved.

Figure 8  Measured (Skidmore and Chase, 1990) and calculated competitive adsorption equilibrium of lysozyme and BSA. In each experiment the total masses of the proteins in both phases are the same and the liquid holdup during the experiments was 0.974. The broken lines were calculated with the full model, the dotted lines were calculated with the assumption of $\sigma = 1$ and the full lines were calculated while assuming equal protein sizes.
Conclusions

The available area isotherm describes the single component adsorption equilibrium just as well as the mass action isotherm, the ‘best’ isotherm until now. Moreover it has the advantage that the binding charge describes purely the effect of ionic strength and that it seems fundamentally more correct.

Prediction of competitive adsorption isotherms with the available area isotherm is almost as bad as with the mass action model. This article gives more insight into this problem, and we hope that it will bring a solution closer.

References


Appendix A  The available area at low surface coverage

In this appendix an equation for the available area at low surface coverage is derived. A cell model is used. This approach is similar to the one we used for the exclusion of proteins from fibrous structures (Bosma and Wesselingh, 2000).

Suppose we have surface with a certain area, $\chi$, covered by protein ‘2’ and we want to calculate the available area for protein ‘1’. Protein ‘2’ is distributed randomly over the surface with the restriction that the proteins do not overlap. Now suppose that we allow proteins to overlap and imagine an entirely random distribution of protein ‘2’ with the same covered area $\chi$. The total covered area, counting double covered area double etc., will be larger then $\chi$, we call it $\chi_{\text{random}}$. We can then imagine the process of increasing the radius of protein ‘2’ and decreasing the radius of protein ‘1’, while keeping their sum constant. During this process the available area for (the center of) protein ‘1’ will not change. The oversized protein ‘2’ will now occupy a total area $\varepsilon_{\text{random}}$, which includes overlapping excluded areas.

The relation between $\chi_{\text{random}}$ and $\varepsilon_{\text{random}}$ is simple:

$$\frac{\varepsilon_{\text{random}}}{\chi_{\text{random}}} = \left( \frac{r_1 + r_2}{r_2} \right)^2$$

(A1)

Here $r$ is the radius of a protein. The obtained $\varepsilon_{\text{random}}$ is the area excluded for protein ‘1’. It is related to the excluded area, $\varepsilon$, in the same way as $\chi_{\text{random}}$ and $\chi$ are related. This relation can be derived by analyzing the surface with a cell model.

Suppose the surface can be represented by $P$ equally sized cells. We begin with $P$ empty cells, and begin to add protein elements to these at random, until the fraction $\chi$ is filled with protein. During this filling we count two numbers:

1. In the first count we allow each cell to be counted any number of times. A cell filled twice represents an overlap of two proteins. This count is related to $\chi_{\text{random}}$, the surface coverage including overlap.
2. In the second count, a filled cell cannot be counted a second time. This count is related to $\chi$, the surface coverage without overlap.

The first cell encountered is always empty, therefore the number of filled cells according to the second count, begins with:

$$N_1 = 1$$

(A2)

The subscript is the number of filled cells according to the first count. Every next time a cell is filled we have for the second count:
\[ N_{i+1} = N_i + \left( 1 - \frac{N_i}{P} \right) \]  
(A3)

The second term on the right hand is the probability that the cell is still empty. Equation A3 can easily be extended. For example using it 3 times gives:

\[ N_{i+1} = 1 + \left( 1 - \frac{1}{P} \right) + \left( 1 - \frac{1}{P} \right)^2 + \left( 1 - \frac{1}{P} \right)^3 + \left( 1 - \frac{1}{P} \right)^4 N_{i-3} \]  
(A4)

This can be extended and then solved to give:

\[ N_F = \sum_{i=0}^{F-1} \left( 1 - \frac{1}{P} \right)^i = P \left[ 1 - \left( 1 - \frac{1}{P} \right)^F \right] \]  
(A5)

For a large number of cells this yields:

\[ \chi = \lim_{P \to \infty} \frac{N_F}{P} = \lim_{P \to \infty} \left[ 1 - \left( 1 - \frac{1}{P} \right)^{\chi_{random}P} \right] = e^{\chi_{random}} - 1 \]  
(A6)

Here we made use of \( F = \chi_{random}P \) and \( N_F = \chi P \). This can be rearranged as:

\[ \chi_{random} = \ln \left( \frac{1}{1 - \chi} \right) \]  
(A7)

From the known area fraction covered by protein ‘2’, \( \chi \), we calculate the available area fraction for protein ‘1’, \( \alpha_1 \), by first calculating \( \chi_{random} \) from \( \chi \), then \( \epsilon_{1,random} \) from \( \chi_{random} \), then \( \epsilon_1 \) from \( \epsilon_{1,random} \) and finally \( \alpha_1 \) from \( \epsilon_1 \). In this way we obtain:

\[ \alpha_1 = \left( 1 - \chi_2 \right) \left( 1 + r_2 \right)^2 \]  
(A8)

When there is also protein ‘1’ adsorbed we estimate the average of \( \epsilon_{random} \) with a weighted average of the single component expressions. This interpolation is not exact but it works well for the concentration dependence of the partitioning of proteins in gels (Bosma and Wesselingh, 2000). The general relation for the excluded surface for protein ‘1’ at low surface coverage becomes:

Available Area Isotherm
\[ \alpha_{1,\text{low}} = \left(1 - \chi_1 - \chi_2 \right) \frac{4\chi_1 + \left(\frac{n + r_2}{r_2} \right)^2 \chi_2}{\chi_1 + \chi_2} \]  

(A9)

**Appendix B  The available area at high surface coverage**

First we consider the case of one protein ‘1’ adsorbing on a surface almost fully loaded with protein ‘2’.

The highest coverage will be obtained with hexagonal packing of protein ‘2’, as shown in figure 2. In this packing there will be two holes per protein, the covered area is \( \chi_{\text{max}} = 0.907 \).

If the packing is not at the tightest but close to it the holes will have a certain distribution of free area sizes. Let the total range of values of the hole area be divided into small subranges \( i \) having average value \( S_i \). Let \( N_i \) be the number of holes having a surface in the \( i \)th region. We have:

\[
\sum_i N_i S_i = N S_{\text{free}}
\]

(B1)

Here \( S_{\text{free}} \) is the average free area per hole; at the tightest packing the free area is zero and at looser packing it is given by:

\[
S_{\text{free}} = \frac{\chi_{\text{max}} - \chi_2}{n_h}
\]

(B2)

Here \( \chi_{\text{max}} \chi_2 \) is the free area fraction and \( n_h \) is the number of holes per surface area:

\[
n_h = \frac{2\chi_2}{4\pi r_2^2}
\]

(B3)

\( N \) is the total number of holes and it is given by:

\[
\sum_i N_i = N
\]

(B4)

The number of ways of redistributing the surface over the holes without changing the \( N_i \) is:

\[
\Omega = \prod_i \frac{N!}{N_i!}
\]

(B5)
If we require that $\Omega$ be a maximum for given $N$ and $S_{\text{free}}$, we obtain with the method of Lagrange multipliers:

$$N_i = \exp(-\lambda - \beta S_i) \tag{B6}$$

By obtaining the Lagrangian multipliers $\lambda$ and $\beta$ from equations B1 and B4 and passing to the continuum limit for the $S_i$ we obtain for the distribution of holes of size $S$, $p(S)$:

$$p(S) = \frac{1}{S_{\text{free}}} \exp\left(-\frac{S}{S_{\text{free}}}\right) \tag{B7}$$

The distribution tells us that there are many small holes, but fewer and fewer larger ones. Holes with an area more than a few times that of the disk are almost non-existent. The average size of the holes does increase rapidly with an increasing free area. We now consider how much area is available for the other protein ‘1’ on a surface covered with protein ‘2’. We assume that the holes are circular. Protein ‘1’ will only fit in holes with a radius larger than $r_1$. However, even there only part of the hole is available area, because the center of the disk has to be at least one radius away from the side of the hole. This is shown in figure 9. The available area for protein ‘1’ at high surface coverage can then be obtained by integration:

$$\alpha_{1,\text{high}} = n_h \int_{S_i}^{\infty} P(S) \left(\sqrt{S} - \sqrt{S_1}\right)^2 dS \tag{B8}$$

Here $S_i$ is the area of one protein ‘1’ ($= 4\pi r_1^2$), $\left(\sqrt{S} - \sqrt{S_1}\right)^2$ is the area available in one hole with area $S$ for the center of an adsorbing protein ‘1’.

The integral can be solved to give:

$$\alpha_{1,\text{high}} = n_h S_{\text{free}} \exp\left(-\frac{S_1}{S_{\text{free}}}\right) + n_h \sqrt{\pi S_1 S_{\text{free}}} \left[\text{erf}\left(\frac{\sqrt{S_1}}{S_{\text{free}}}\right) - 1\right] \tag{B9}$$

If there are two adsorbing components we propose the following interpolations instead of equations B3 and B9:

$$S_{\text{free}} = \frac{\sigma \chi_{\text{max}} - \chi_1 - \chi_2}{n_h} \tag{B10}$$
\[ n_h = \frac{2}{4\pi} \left( \frac{h_1}{r_1^2} + \frac{h_2}{r_2^2} \right) \]  \hspace{1cm} (B11)

Here \( \sigma \) is a number (\( \geq 1 \)) that accounts for the increase in the maximum possible surface coverage when the proteins have different sizes. An equation for \( \sigma \) is given in appendix C.

**Figure 9** The part of the hole available for the center of disks ‘1’

### Appendix C  The maximum surface packing density of binary protein mixtures

When proteins adsorb on a surface there is a highest possible surface coverage. When only one protein adsorbs this is the number \( \chi_{\text{max}} \). When proteins of different sizes adsorb the highest possible surface coverage will increase because of two effects: smaller proteins can fill up the holes between larger ones and overlap of small and large proteins becomes possible. As far as we know there are no solutions for this problem in the literature. However for the three-dimensional case a solution is given by Westman (1936) and Finkers and Hoffmann (1998). Here we translate this solution to the two dimensional case.

We will assume that we have a mixture of spherical proteins with a binary size distribution. We will calculate the variable \( \sigma \), the maximum surface coverage of the mixture divided by the maximum surface coverage of a single component, \( \chi_{\text{max}} \). This variable will depend on the size ratio of the proteins, \( r \), defined by: \( r = r_{\text{large}}/r_{\text{small}} \) and on the adsorbed composition, \( x \), the surface fraction of the large protein, defined by \( x = \chi_{\text{large}}/(\chi_{\text{large}} + \chi_{\text{small}}) \). Below we first derive an equation for \( \sigma \) which satisfies two boundary conditions. This equation contains a fitting variable \( G \), which depends only on \( r \).
First we approach the problem analogously to the approach of the three-dimensional problem by Westman (1936). In figure 10 we can distinguish two boundary cases. When \( r = 1 \), which will correspond to \( G = 1 \), the proteins are equal and there is no difference with the single component case. When \( r = 0 \), which will correspond to \( G = \infty \), the proteins will occupy the surface independently, because they can very easily overlap (when projected on the surface). \( \sigma \) will then have a maximum of 2 at \( x = 0.5 \) and this point will be connected by straight lines with \( \sigma = 1 \) at \( x = 0 \) and \( x = 1 \). This may be described with an empirical equation of the form:

\[
a^2 + 2Gab + b^2 = 1
\]  
(C1)

For \( a \) and \( b \) we can take the following functions of \( \sigma \) and \( x \):

\[
a = \frac{1}{2} (\sigma + 2x - 3)
\]  
(C2)

\[
b = \frac{1}{2} (\sigma - 2x + 1)
\]  
(C3)

The variable \( G \) will depend on \( r \). When \( r = 1 \) it should be 1 and we get \( \sigma = 1 \); when \( r = 0 \) it should be infinite and we get \( a = 0 \) or \( b = 0 \).

A proper form of \( G(r) \) may be found with the structural ratio concept that Finkers and Hoffmann (1998) used for the three dimensional problem. For the two dimensional problem the proper form will be:

\[
G(r) = 1 + \left( \frac{1 - \chi_{max}}{r^2} \right)^{-k} - (1 - \chi_{max})^{-k}
\]  
(C4)

where \( k \) is an empirical constant. The fitted value for this parameter increases from 0.36 for a wide size distribution to 0.63 for a narrow size distribution. We use the latter value.
Figure 10 The fitting equation (eq. C1) for the maximum surface coverage by mixtures of proteins.