DEVELOPING AND TESTING MODELS FOR ANTIBODY KINETICS IN SURFACE PLASMON RESONANCE EXPERIMENTS

by

H.A.J Moyse

supervised by

Dr. N. Evans

Thesis

Submitted to the university of Warwick

for the degree of

Master of Science
Abstract

In this thesis three novel models for antibody binding in surface plasmon experiments, that account for heterogenous binding dynamics, have been created. These have been structurally analysed, and the best has been fitted to data for an experiment on commercial anti A IgM, and demonstrated to give a dramatically lower RSS than the existing models.

The structural analysis has demonstrated that meaningful estimates of parameters can be made even when dealing with clinical samples, where the concentration of analyte is unknown.
Contents

1 Introduction ............................................. 1
   1.1 Existing models ...................................... 3

2 methodology ............................................. 4
   2.1 Experimental methods ............................... 5
   2.2 Extend existing models ............................ 6
      2.2.1 New models .................................. 8
   2.3 Structural Identifiability ......................... 9
   2.4 Testing models on monoclonal data .............. 12
   2.5 Modeling data with multiple binding types ...... 13

3 Discussion .............................................. 15

4 Conclusion and Outlook ............................... 17

5 Acknowledgements .................................... 18

A Appendix .............................................. 20
   A.1 Identifiability analysis ............................ 20
   A.2 Programs .......................................... 21
      A.2.1 matlab ..................................... 21
      A.2.2 Facsimile .................................... 23
Nomenclature

ABO  Blood group system

DSAs  Donor specific antibodies

HLA  Human leukocyte antigen

L  Langmuir model

LT  Langmuir model with transport equation

ODE  Ordinary differential equation

PDE  Partial differential equation

QSS  Langmuir model with a the quasi steady state approximation of the transport equation

RSS  Residual sum of squares

SDLN  Standard deviation of natural logarithm

SGI  Structurally globally identifiable

SLI  Structurally locally identifiable

SU  Structurally Unidentifiable

SPR  Surface plasmon resonance
1 Introduction

One strategy to reduce a patient’s risk of transplant rejection would be the use of tailored immunosuppresant drugs. Before such drugs can be used, a better understanding of the levels of each type of antibody a transplant can tolerate is needed. This thesis presents the use of existing mathematical models and new mathematical models to evaluate the dynamics of antibody binding in surface plasmon resonance (SPR) experiments with commercial anti-A monoclonal IgM. These fits and models are one of the necessary steps towards identifying and characterising potential risks in individual patients, and allowing tailored immunosuppresant drugs[17].

Currently a patient receiving a kidney transplant would undergo crossmatch to test for compatibility. Such a test would be failed if donor specific antibodies (DSAs), antibodies that would attack the transplant, were found in the recipient; either associated with their ABO blood type or their human leukocyte antigen tissue type (HLA). In such cases pretransplant antibody removal can be used to allow the operation[8]. However this carries a large immunological risk, with heightened chances of infection dysfunction as well as the existing risk of rejection. As a result significant improvements could be made using immunosuppresant drugs designed for the individual; these would act specifically on the antibodies that would attack the transplant, lowering and maintaining them at safe levels. This would lead to lower rejection rates, and as a result a greater availability of donor organs and shorter waiting times for donor organs.

The data analysed in this thesis was taken using SPR experiments conducted with the ProteOn XPR36 platform (manufactured by Bio Rad,See[10] and [19] fora detailed explanation of this kind of technology). SPR experiments were used rather than the established method for measuring ABO specific antibody levels, haemaglutenation (HA) because of established problems with reproducibility[1]. Historically SPR experiments have been used to estimate the binding coefficients of cultured antibodies in [6] and [13].

A number of mathematical models have been developed to describe the dynamics of a reaction between a flow of one reactant across a sensor surface. The most complex of these involve a system of partial differential equations (PDEs) that in include the effects of fluid mechanics
on the reactants, these models are used to develop appropriate ordinary differential equations (ODEs) that can be fitted to data. Examples of this include [14] and [4] who consider the effects of a boundary layer between the well mixed analyte and the ligand where reactions take place, and use this as a boundary condition for the transport equation; and [5, 20] who both consider the effects of three dimensional receptor layer that analyte must diffuse through before reacting. The simplest models include a single differential equation and, assume that in a short period of time the analyte becomes well mixed, and the rate the reactants bind is solely dependant on the laws of mass action. However, it has been demonstrated that the interactions measured are determined not only by the availability of both reactants but the effects of transport processes[14], as a result the use of these models in parameter fitting can introduce systematic errors [2]. A compromise between the two levels of detail was proposed by [14], by assuming that the analyte on the boundary layer is in a quasi steady state, a model may be constructed consisting of only a single differential equation in terms bound concentration. This model was shown to be a good approximation of the physical system under certain conditions [5].

The models discussed were designed to deal with experiments where only one binding reaction is taking place, but antibodies have multiple binding domains, some of which binding and unbinding at different rates [16]. This problem would be exacerbated in clinical conditions where the analyte may be made up of antibodies of different isotypes. As a result, additional models were developed for this thesis which allow for different binding types.

Whilst there may be a number of processes occurring in an SPR experiment, the output we are given is purely in terms of the concentration of analyte bound to the sensor. It is necessary to consider the identifiability of a new model before fitting parameters, and establish that there are not multiple sets of parameters that would give an identical output. Even in the case where each parameter of a model is physically meaningful there may not be enough information in the output to uniquely determine the parameters for a model [9]. Additionally, because the models we are developing need to be useful for clinical data where the concentration of analyte may be unknown special care has to be taken to ensure that their parameters can be meaningfully estimated from the output of an experiment where the analyte concentration would be unknown.
Table 1: Parameters, variables and sets

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>Concentration of analyte at the surface</td>
<td>ng/mm$^3$</td>
</tr>
<tr>
<td>$B$</td>
<td>Average bound area of the surface</td>
<td>ng/mm$^2$</td>
</tr>
<tr>
<td>$h$</td>
<td>Quotient of the volume in contact with the surface over the surface area</td>
<td>mm$^3$</td>
</tr>
<tr>
<td>$k_a$, $k_d$</td>
<td>Constants of association and disassociation</td>
<td></td>
</tr>
<tr>
<td>$K_M$</td>
<td>Transport coefficient</td>
<td></td>
</tr>
<tr>
<td>$R$</td>
<td>Maximum density of bound analyte</td>
<td></td>
</tr>
<tr>
<td>$I$</td>
<td>Inlet concentration of analyte</td>
<td></td>
</tr>
<tr>
<td>$C_T$</td>
<td>Analyte sample concentration</td>
<td></td>
</tr>
</tbody>
</table>

For non-linear systems, such as the models discussed here, there are a number of techniques for preforming a structural identifiability analyses, those based on smooth transitions between models with identical outputs[15], differentiable algebra [12, 18], uniqueness of the Taylor series expansion of the output [7]. The models created in this thesis are such that an equation relating the output to its derivatives, rather than the state variables of the system, may be obtained, as a result a method similar to that used in [3] may be applied.

1.1 Existing models

Of the available models in the literature discussed three were selected, and are presented in this section: one that assumes analyte on the boundary of the receptor is at the same concentration as that free flowing, the Langmuir model (L); one that uses a transport equation and allows analyte on the boundary to vary, Langmuir with Transport (LT)[14]; one that doesn’t allow the concentration of the analyte on the boundary to vary, but includes the effects of transport, the Langmuir model with a quasi state approximation of the transport equation (QSS)[5, 14]. These models are presented using the parameters and variables outlined in 1.

The Langmuir model (L) is defined as:

$$B(t) = k_a C_T (R - B(t)) - k_d B(t)$$

(1)
it was first proposed in [11], and forms the basis for the more recent models.

The Langmuir with Transport (LT) model is defined as:

\[
\begin{align*}
\dot{C}(t) &= -k_a C(t) (R - B(t)) + k_d B(t) + K_M (C_T - C(t)) \\
\dot{B} &= k_a C_T (R - B) - k_d B
\end{align*}
\]

(2) (3)

it was first proposed in [14]. In this paper the transport equation, a differential equation governing \( C(t) \), the concentration of the analyte at the surface was added to the system, and was demonstrated to improve fits to simulated data.

The quasi steady state approximation of the Langmuir model (QSS) is

\[
\dot{B} = \frac{k_a C_T (R - B) - k_d B}{(k_a/K_M) (R - B) + 1}
\]

(4)

it was also first proposed in [14], but in [5] it was demonstrated to be a good approximation of a fluid dynamics model up to \( O(Da^2) \) where \( Da \) the Damköhler number, the quotient of reaction velocity to diffusion velocity in the boundary layer, is defined as

\[
Da = \frac{k_a R}{k_M h_d}
\]

(5)

and \( h_d \) is a constant that incorporates the effects of the receptor layer. This model may be derived by setting \( \dot{C} = 0 \) and substituting the solution of eq.2 for \( C \) into eq.3.

2 methodology

The method that was used in this project is outlined here:

1. Extend existing models
2. Analyse identifiability of extended models
3. Fit parameters with single binding types

4. Select models

5. Fit parameters with multiple binding types

Each one of the above is given a full discussion in a corresponding subsection. Before the modeling can be discussed it is also necessary to discuss the way the data was gathered, and the experimental procedures involved.

2.1 Experimental methods

The experiments were conducted on the XPR 36 SPR platform (manufactured by Bio-rad), in which reactions take place along a $6 \times 6$ grid of analytes and ligands, that intersect at interaction spots as shown in 1. Analyte is pumped through six channels, and sequentially encounters the stationary ligands. At each reaction spot an individual SPR experiment is conducted, giving us a time series in terms of the average bound area of the surface, with data points every 0.9 seconds. We term the output from the $i^{th}$ analyte and $j^{th}$ ligand $y_{ij}$.

The experiments were conducted using two ligands, trisaccheride amine and trisaccheride linker on two separate lanes of a carboxylated SPR chip. A single analyte was used, commercially available anti-A monoclonal IgM, but it was put in five different dilutions (1:10, 1:20, 1:50, 1:75, 1:100) on the first five channels respectively. As a result we have time series data for 10 separate simultaneous binding reactions ($y_{ij} : i \in (1, ..., 5), j \in \{a, l\}$).
Each one of these time series consists of several phases, firstly a 7.2 second period before the analyte begins interacting with the ligands $T_0$; a 120 second association phase $T_1$, when analyte is pumped through an inlet into the channels; and a 667.5 second dissociation phase $T_2$, after analyte ceases to be pumped.

In Figure 2 we see the outputs of these experiments plotted against time. The bound concentration rises sharply through $T_1$ and reaches a maximum at 127.2 seconds, the beginning of $T_2$. From there on the bound concentration descends, sharply at first, but at a decreasing rate. Throughout these changes, at all time, the bound concentration of the experiments on the channels with lower dilutions remain higher than the bound concentrations of the experiments with higher dilutions for the same ligand. Additionally, at all time points, all the bound concentrations remain higher on the amine rather than the Linker.

### 2.2 Extend existing models

Two deal with the experimental outputs outlined in the above it is necessary to introduce two new concepts to the models outlined in section 1.1. We now use an extended notation, including indices $i$ and $j$ corresponding to the analyte and ligand respectively, which is summarised in Table 2. Notably the parameters $d_i$ and $h_j$ are presented with only one index each, because they are the same either for all ligands or for all analytes.
As well as restating these models I shall also be listing the parameters of these models, corresponding to the number of analytes, ligands and reaction spots and giving the number of parameters required to deal with the data outlined in section 2.1. This enumeration is necessary because the numbers of parameters will become important when we analyse their identifiability and fit these models.

The Langmuir model is

\[
\dot{B}_{ij} = \begin{cases} 
  k_{aj}d_iC_T(R_{ij} - B_{ij}) - k_{dj}B_{ij} & t \in (T_1) \\
  -k_{dj}B_{ij} & t \notin (T_1)
\end{cases}
\]

and has 24 unknown parameters: two unique to each ligand \(k_{aj}, k_{dj}\); and two \(R_{ij}, C_{ij}\) unique to each interaction spot.

The Langmuir with Transport model (LT) is

\[
\dot{B}_{ij} = k_{aj}C_{ij}(R_{ij} - B_{ij}) - k_{dj}B_{ij}
\]

\[
h_j\dot{C}_{ij} = \begin{cases} 
  -k_{aj}C_{ij}(R_{ij} - B_{ij}) + k_{dj}B_{ij} + K_{Mij}(d_iC_T - C_{ij}) & t \in (T_2) \\
  -k_{aj}C_{ij}(R_{ij} - B_{ij}) + k_{dj}B_{ij} - C_{ij}K_{Mij} & t \notin (T_2)
\end{cases}
\]

with 31 unknown parameters: three for each ligand \(k_{aj}, k_{dj}, h_j\); two for each spot \(R_{ij}, K_{Mij}\), and one for each analyte channel \(d_i\).

The Langmuir model with the quasi steady state approximation of the transport equation (QSS) is

\[
\dot{B}_{ij} = \begin{cases} 
  k_{aj}d_iC_T(R_{ij} - B_{ij}) - k_{dj}B_{ij} \left(\frac{k_{aj}}{K_{Mij}}\right) & t \in (T_1) \\
  \frac{-k_{dj}B_{ij}}{(k_{aj}/K_{Mij})(R_{ij} - B_{ij})+1} & t \notin (T_1)
\end{cases}
\]

with 29 unknown parameters: two for each ligand \(k_{aj}, k_{dj}\), two for each reaction spot \(R_{ij}, k_{Mij}\) and one for each analyte \(I_i\).
Table 2: Parameters, variables and sets

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_{ij}$</td>
<td>output</td>
<td>$10^{-3} \text{ ng/mm}^3$</td>
</tr>
<tr>
<td>$C_{ij}$</td>
<td>Concentration of analyte at the surface</td>
<td>ng/mm$^3$</td>
</tr>
<tr>
<td>$B_{ij}$</td>
<td>Average bound area of the surface</td>
<td>ng/mm$^2$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>conversion factor between output and average bound area</td>
<td></td>
</tr>
<tr>
<td>$h_j$</td>
<td>Quotient of the volume in contact with the surface over the surface area</td>
<td>mm$^?$</td>
</tr>
<tr>
<td>$k_{aj}, k_{dj}$</td>
<td>Constants of association and dissociation</td>
<td></td>
</tr>
<tr>
<td>$k_{Mij}$</td>
<td>Transport coefficient</td>
<td></td>
</tr>
<tr>
<td>$R_{ij}$</td>
<td>Maximum density of bound analyte</td>
<td></td>
</tr>
<tr>
<td>$I_i$</td>
<td>Inlet concentration of analyte</td>
<td></td>
</tr>
<tr>
<td>$C_T$</td>
<td>Analyte sample concentration</td>
<td></td>
</tr>
<tr>
<td>$d_i$</td>
<td>Dilution factor</td>
<td></td>
</tr>
<tr>
<td>$t_1, t_2$</td>
<td>Start and finish times of the association phase</td>
<td></td>
</tr>
</tbody>
</table>

2.2.1 New models

To deal with the heterogeneous binding established in [16], a new index, $k$, representing the type of binding undergone in each reaction, is created. As we have done with the previous models in this section we shall also number the parameters needed for these models to deal with the data outlined in section 2.1.

We extend the Langmuir without transport for $n$-binding types (referred to as Ln):

$$
\dot{B}_{ijk} = \begin{cases} 
    k_{ajk}C_{ij}(R_{ij} - \sum_{k=1}^{n} B_{ijk}) - k_{dj}B_{ijk} & t \in (T_1) \\
    -k_{dj}B_{ijk} & t \notin (T_1) 
\end{cases}
$$

(10)
giving us $10(n + 1)$ parameters in total.

We extend the Langmuir with transport for $n$-binding types (referred to as LTn)
\[
\dot{B}_{ijk} = -k_{ajk}C_{ij} \left( R_{ij} - \sum_{k=1}^{n} B_{ijk} \right) + k_{dj}B_{ijk} \tag{11}
\]

\[
h_j \dot{C}_{ij} = \begin{cases} 
- \sum_{k=1}^{n} [k_{ajk}C_{ij} (R_{ij} - \sum_{k=1}^{n} B_{ijk}) + k_{dj}B_{ijk}] - C_{ij}k_{Mij} & t \in T_1 \\
- \sum_{k=1}^{n} [k_{ajk}C_{ij} (R_{ij} - \sum_{k=1}^{n} B_{ijk}) + k_{dj}B_{ijk}] + k_{mij}(d_iC_T - C_{ij}) & t \in T_1 \end{cases} \tag{12}
\]

giving us \(5(2n + 6)\) parameters in total.

We extend the Quasi steady state model for n-binding types (referred to as QSSn)

\[
\dot{B}_{ijk} = \begin{cases} 
\frac{k_{ajk}(K_{Mij}d_iC_T + \sum_{k=1}^{n} k_{dj}B_{ijk})(R_{ij} - \sum_{k=1}^{n} B_{ijk})}{\sum_{k=1}^{n} k_{ajk}(R_{ij} - \sum_{k=1}^{n} B_{ijk}) + k_{Mij}} - k_{dj}B_{ijk} & t \in (T_1) \\
\frac{k_{ajk}(\sum_{k=1}^{n} k_{ajk}B_{ijk})(R_{ij} - \sum_{k=1}^{n} B_{ijk}) + k_{Mij}}{\sum_{k=1}^{n} k_{ajk}(R_{ij} - \sum_{k=1}^{n} B_{ijk}) + k_{Mij}} - k_{dj}B_{ijk} & t \notin (T_1) \end{cases} \tag{13}
\]

for each curve there are \(5(2n + 5)\) parameters in total

notably there are lots of parameters, so to make fitting easier we deal with the cases where \(n = 1\) or \(2\).

### 2.3 Structural Identifiability

In an SPR experiment we do not necessarily know the values taken by our state variables \((B_{ij}, C_{ij})\) at any point in time. We, however, do have an output for each interaction spot on the chip which is related to our state variables by

\[
y_{ij} = \alpha B_{ij} \tag{14}
\]

where \(\alpha\) is the conversion factor from the units of \(B_{ij}\) to the response units of the sensograms (1 RU = \(10^{-3} ng/mm\)).

As a result it is not immediately apparent whether or not there are more than one parameter sets that will give us the same evolution of \(y_{ij}\) during the time of the experiment. Writing this more formally, it is not apparent if given a vector of parameters \(p\) taken from the set of all possible vectors of parameters \(\Omega \subset \mathbb{R}\), if \(y_{ij}(p; t) = y_{ij}(\bar{p}; t) \rightarrow p = \bar{p}\).
Before we can discuss the results that have been obtained for the three models we are dealing with initially we have to establish some standard definitions relating to identifiability.

**Definition 1.** Two parameter vectors \( p \) and \( \bar{p} \) are *indistinguishable* if they give the same outputs \( y_{ij}(t; p) = y_{ij}(t; \bar{p}) \) for \( \forall t \geq 0 \). We write this equivalence relationship with \( p \sim \bar{p} \).

**Definition 2.** A parameter \( p_i \) is *locally identifiable* if there is a neighbourhood, \( N(p) \), of the parameter vector where \( \bar{p} \in N(p), p \sim \bar{p} \Rightarrow p_i = \bar{p}_i \).

**Definition 3.** A parameter \( p_i \) is *globally identifiable* if it is locally identifiable for the neighbourhood \( \Omega \), that is \( \bar{p} \in \Omega, p \sim \bar{p} \rightarrow p_i = \bar{p}_i \).

**Definition 4.** A parameter \( p_i \) is *unidentifiable* if there does not exist a neighbourhood where it is locally identifiable.

These terms can be used to make definitions relating to a model as a whole

**Definition 5.** A systems model is *structurally globally identifiable* (SGI) if all of its parameters are globally identifiable.

**Definition 6.** A systems model is *structurally locally identifiable* (SLI) if all of its parameters are locally identifiable, but not all of its parameters are globally identifiable.

**Definition 7.** A systems model is *structurally unidentifiable* (SU) if one or more of its parameters are unidentifiable.

In this thesis a variation on the approach developed by [3] is used to classify the models LTn, Ln and QSSn, as SGI, SLI or SU. Firstly the set of derivatives for the differential equations defining each model (eq. 10 for Ln, eq.11 and 12 for LTn and eq.13 for QSSn ) is solved to produce a single equation relating the output \( y_{ij} \) to a quotient of multinomials in terms of \( y_{ij} \) and its time derivatives, i.e.

\[
y_{ij}(t; p) = \frac{M_1(\dot{y}_{ij}, \ddot{y}_{ij}, \ldots; p)}{M_2(\dot{y}_{ij}, \ddot{y}_{ij}, \ldots; p)}.
\]

For a parameter vector \( \bar{p} \) to be indistinguishable from \( p \) these quotients would be equal, giving:
known \( C_T \)

<table>
<thead>
<tr>
<th>Model</th>
<th>Globally identifiable parameters</th>
<th>Locally identifiable parameters</th>
<th>Unidentifiable parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln</td>
<td>( R_{ij}, k_{M_{ij}} )</td>
<td>( k_{a_{ijk}}, k_{d_{ijk}} )</td>
<td>-</td>
</tr>
<tr>
<td>LTN</td>
<td>( R_{ij} )</td>
<td>( k_{d_{ijk}} )</td>
<td>( k_{M_{ij}}, k_{a_{ijk}}, h_{ij} )</td>
</tr>
<tr>
<td>QSSn</td>
<td>( R_{ij}, k_{M_{ij}}, k_{M_{ij}} )</td>
<td>( k_{a_{ijk}}, k_{d_{ijk}} )</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Identifiability of parameters for known \( C_T \)

unknown \( C_T \)

<table>
<thead>
<tr>
<th>Model</th>
<th>Globally identifiable parameters</th>
<th>Locally identifiable parameters</th>
<th>Unidentifiable parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln</td>
<td>( R_{ij}, k_{M_{ij}} )</td>
<td>( k_{a_{ijk}}, k_{d_{ijk}}, C_T )</td>
<td></td>
</tr>
<tr>
<td>LTN</td>
<td>( R_{ij} )</td>
<td>( k_{d_{ijk}} )</td>
<td>( k_{M_{ij}}, k_{a_{ijk}}, h_{ij}, C_T )</td>
</tr>
<tr>
<td>QSSn</td>
<td>( R_{ij}, k_{M_{ij}}, k_{M_{ij}} )</td>
<td>( k_{a_{ijk}}, k_{d_{ijk}} )</td>
<td>( C_T )</td>
</tr>
</tbody>
</table>

Table 4: Identifiability of parameters for unknown \( C_T \)

\[
M_1(\tilde{y}_{ij}, \ddot{y}_{ij}, \ldots; \tilde{p})M_2(\tilde{y}_{ij}, \ddot{y}_{ij}, \ldots; p) = M_1(\tilde{y}_{ij}, \ddot{y}_{ij}, \ldots; p)M_2(\tilde{y}_{ij}, \ddot{y}_{ij}, \ldots; \tilde{p}). \tag{16}
\]

For this equality to be true the coefficient of each product of derivatives must be equal to a corresponding quotient on the opposite side. As a result we get a new set of equations relating the elements of \( p \) to the elements of \( \tilde{p} \). If this set of equations only has the trivial solution \( \tilde{p} = p \) then the model is SGI, if it has a set of distinct solutions other than the trivial one it is SLI and it is SU otherwise.

In the appendix (A.1) the maple worksheets that were used to apply this method to LTN, Ln and QSSn for \( n = 2 \). The results obtained are summarised in tables 3 and 4. In these tables, the binding constants are never globally identifiable because the labelling of one binding type as \( k = 1 \) and the other as \( k = 2 \) is arbitrary, so the labels can always be swapped. Additionally, the binding constants are unidentifiable in the LTN model for unknown \( C_T \) making it inappropriate for clinical data.

We make the choice of using only two binding types because of the length of time required for parameter fitting for each model. The methods outlined here could with comparatively
little work be extended for $n = 3$ or a general case.

### 2.4 Testing models on monoclonal data

A precursor to fitting models to clinical data would be fitting them to data from experiments conducted with monoclonal antibodies and demonstrating that they give a dramatically improved fit.

Parameter estimation was performed in two stages, first the models $L$ and $QSS$ were fitted, secondly a the model that fitted the data best was selected and, its extended version (either $Ln$, $LTn$ or $QSSn$) was fitted. Notably the $LT$ model was not fit, this is because it would be structurally unidentifiable for clinical data, and the parameters obtained would be meaningless. Fits were conducted using facsimile for windows (MCPIA software UK).

Because each parameter represents a physical quantity of an unknown order of magnitude for each model a large number of starting parameter vectors were created for each model (441 for $L$ and 9261 for $QSS1$), with each parameter taking every exponent of 10 from $10^{-4}$ to $10^{2}$. These were then read in, and the residual sum of squares for each model with each parameter vector was calculated. For each model the best 100 starting parameter vectors was then run 1000 times, with facsimile completing an average of 45 simulation runs per model. Of these best fits the best fit for each model was then ran 1000 times more.

In figure 3 presents fits to the data of the Langmuir and quasi steady state models. Whilst they appear similar there is a large difference in residual sum of squares; whilst the Langmuir model has an RSS of 1,578, the quasi steady state model has an RSS of 941. The key differences in fit that can be seen in fig 3 are that the Langmuir model predicts much faster growth in bound area in the association phase and peaks earlier and lower than the Quasi steady state model on all ten curves.

In table 5 we see the parameters from the fits of both models as well as their standard deviation of natural logarithm (SDLN). All parameters have SDLNs below 0.1 and as a result are well determined. Interestingly the values of the association constants in the Langmuir model are an order of magnitude smaller than those from the QSS model, this is likely because
Figure 3: Fits of Langmuir and Quasi steady state models (red) to experimental data (blue).

The Langmuir model assumes that the boundary concentration of analyte is the same as the concentration of the analyte outside the boundary, i.e. is more available, and binds slower.

2.5 Modeling data with multiple binding types

The extended version of the QSS model, QSSn, was selected to be fit next. This was done for two primary reasons: QSS gave a dramatically better fit than L and it allows for the effects of a boundary layer.

Start points were created by selecting the parameter values for the best 50 starting points for the QSS model, each one of these parameter vectors was turned into 225 new parameter vectors with different combinations of orders of magnitude for the 4 new parameters, each one of these starting points was read in to Facsimile and its RSS was calculated. The top 100 of these starting points were then ran 1000 times and the top parameter vector of those was ran a further 1000 times.

Because each of the elements in a parameter vector represents a physically meaningful quantity that we would not expect to vary by multiple orders of magnitude between two models it was considered to be acceptable to use the best starting points from the QSS model rather than creating all new starting points for the QSSn model. This step decreased the number
<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>SDLN</th>
<th>value</th>
<th>SDLN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{aa}$</td>
<td>$5.2 \times 10^{-3}$</td>
<td>0.01</td>
<td>$2.74 \times 10^{-2}$</td>
<td>0.07</td>
</tr>
<tr>
<td>$k_{al}$</td>
<td>$5.0 \times 10^{-3}$</td>
<td>0.02</td>
<td>$2.61 \times 10^{-2}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$k_{da}$</td>
<td>$4.6 \times 10^{-2}$</td>
<td>0.01</td>
<td>$6.05 \times 10^{-4}$</td>
<td>0.03</td>
</tr>
<tr>
<td>$k_{dl}$</td>
<td>$5.0 \times 10^{-2}$</td>
<td>0.01</td>
<td>$6.89 \times 10^{-4}$</td>
<td>0.07</td>
</tr>
<tr>
<td>$R_{1a}$</td>
<td>1.26</td>
<td>0.01</td>
<td>1.20</td>
<td>0.03</td>
</tr>
<tr>
<td>$R_{2a}$</td>
<td>1.37</td>
<td>0.01</td>
<td>1.14</td>
<td>0.04</td>
</tr>
<tr>
<td>$R_{3a}$</td>
<td>1.06</td>
<td>0.01</td>
<td>$6.29 \times 10^{-1}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$R_{4a}$</td>
<td>1.21</td>
<td>0.01</td>
<td>$6.07 \times 10^{-1}$</td>
<td>0.02</td>
</tr>
<tr>
<td>$R_{5a}$</td>
<td>1.30</td>
<td>0.01</td>
<td>$6.29 \times 10^{-1}$</td>
<td>0.02</td>
</tr>
<tr>
<td>$R_{1l}$</td>
<td>$9.50 \times 10^{-1}$</td>
<td>0.00</td>
<td>$8.98 \times 10^{-1}$</td>
<td>0.00</td>
</tr>
<tr>
<td>$R_{2l}$</td>
<td>1.03</td>
<td>0.01</td>
<td>$8.51 \times 10^{-1}$</td>
<td>0.00</td>
</tr>
<tr>
<td>$R_{3l}$</td>
<td>$7.69 \times 10^{-1}$</td>
<td>0.01</td>
<td>$4.40 \times 10^{-1}$</td>
<td>0.01</td>
</tr>
<tr>
<td>$R_{4l}$</td>
<td>$8.90 \times 10^{-1}$</td>
<td>0.01</td>
<td>$4.12 \times 10^{-1}$</td>
<td>0.01</td>
</tr>
<tr>
<td>$R_{5l}$</td>
<td>1.00</td>
<td>0.01</td>
<td>$4.16 \times 10^{-1}$</td>
<td>0.01</td>
</tr>
<tr>
<td>$k_{M1a}$</td>
<td>$3.47 \times 10^{-3}$</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M2a}$</td>
<td>$6.52 \times 10^{-3}$</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M3a}$</td>
<td>$7.79 \times 10^{-3}$</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M4a}$</td>
<td>$1.15 \times 10^{-3}$</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M5a}$</td>
<td>$1.49 \times 10^{-3}$</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M1l}$</td>
<td>$2.57 \times 10^{-3}$</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M2l}$</td>
<td>$4.87 \times 10^{-3}$</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M3l}$</td>
<td>$6.36 \times 10^{-3}$</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M4l}$</td>
<td>$1.15 \times 10^{-2}$</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{M5l}$</td>
<td>$1.31 \times 10^{-2}$</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Table of estimated parameters and standard deviations of natural logarithms
of starting points required to cover the possible orders of magnitude the parameters of the QSSn model by an order of magnitude.

In figure 4 we see fits from the QSSn model, visually it looks much closer to the data than either of the preceding models, correspondingly it has a significantly lower residual sum of squares, 261, less than a third of that of the preceding model. Whilst there is only a small distance between the model and the data there seem to be some interesting patterns in the errors. As time increases the model increasingly over predicts the loss of bound analyte, and the data appears to be reaching an asymptote, suggesting that some of the binding may be irreversible.

We present the parameters of this model in table 6.

3 Discussion

It has been demonstrated that the QSSn model offers dramatically better fits than either the L or QSS models. Additionally it remains SLI even if the total sample concentration $C_T$ is unknown. As a result it could be used for clinical data, and in such data could be expected to preform much better than previously existing models.
<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>SDLN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{aa1}$</td>
<td>$3.1 \times 10^{-3}$</td>
<td>0.07</td>
</tr>
<tr>
<td>$k_{al1}$</td>
<td>$2.3 \times 10^{-2}$</td>
<td>0.05</td>
</tr>
<tr>
<td>$k_{da1}$</td>
<td>$3.23 \times 10^{-2}$</td>
<td>0.03</td>
</tr>
<tr>
<td>$k_{dl1}$</td>
<td>$5.01 \times 10^{-4}$</td>
<td>0.07</td>
</tr>
<tr>
<td>$k_{aa2}$</td>
<td>$3.30 \times 10^{-3}$</td>
<td>0.03</td>
</tr>
<tr>
<td>$k_{al2}$</td>
<td>$2.9 \times 10^{-2}$</td>
<td>0.07</td>
</tr>
<tr>
<td>$k_{da2}$</td>
<td>$3.61 \times 10^{-4}$</td>
<td>0.03</td>
</tr>
<tr>
<td>$k_{dl2}$</td>
<td>$3.78 \times 10^{-2}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$R_{1a2}$</td>
<td>1.51</td>
<td>0.04</td>
</tr>
<tr>
<td>$R_{2a}$</td>
<td>1.98</td>
<td>0.04</td>
</tr>
<tr>
<td>$R_{3a}$</td>
<td>2.17</td>
<td>0.01</td>
</tr>
<tr>
<td>$R_{4a}$</td>
<td>2.50</td>
<td>0.02</td>
</tr>
<tr>
<td>$R_{5a}$</td>
<td>2.81</td>
<td>0.00</td>
</tr>
<tr>
<td>$R_{1l}$</td>
<td>1.05</td>
<td>0.00</td>
</tr>
<tr>
<td>$R_{2l}$</td>
<td>1.10</td>
<td>0.00</td>
</tr>
<tr>
<td>$R_{3l}$</td>
<td>1.12</td>
<td>0.00</td>
</tr>
<tr>
<td>$R_{4l}$</td>
<td>1.15</td>
<td>0.01</td>
</tr>
<tr>
<td>$R_{5l}$</td>
<td>1.18</td>
<td>0.01</td>
</tr>
<tr>
<td>$k_{M1a}$</td>
<td>$9.47 \times 10^{-2}$</td>
<td>0.96</td>
</tr>
<tr>
<td>$k_{M2a}$</td>
<td>$3.09 \times 10^{-2}$</td>
<td>0.12</td>
</tr>
<tr>
<td>$k_{M3a}$</td>
<td>$2.05 \times 10^{-2}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$k_{M4a}$</td>
<td>$3.51 \times 10^{-2}$</td>
<td>0.07</td>
</tr>
<tr>
<td>$k_{M5a}$</td>
<td>$3.27 \times 10^{-2}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$k_{M1l}$</td>
<td>$1.20 \times 10^{-2}$</td>
<td>0.76</td>
</tr>
<tr>
<td>$k_{M2l}$</td>
<td>$1.63 \times 10^{-2}$</td>
<td>0.09</td>
</tr>
<tr>
<td>$k_{M3l}$</td>
<td>$1.19 \times 10^{-3}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$k_{M4l}$</td>
<td>$1.15 \times 10^{-2}$</td>
<td>0.05</td>
</tr>
<tr>
<td>$k_{M5l}$</td>
<td>$1.31 \times 10^{-2}$</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 6: Table of estimated parameters and standard deviations of natural logarithms
The key improvement that the QSSn model offers is that the peaks it predicts are close both in the time that they occur and in height to those observed in the experiments at the end of the association phase. This improvement supports the work of [16]. Together that paper and this thesis give strong grounds to conclude that IgM has heterogeneous binding kinetics.

One criticism of the analysis is that models with varying numbers of parameters were compared, and the ones with least RSS were selected despite having a greater number of parameters. This could create two potential problems: firstly the dramatic decrease in RSS between models could be explained as purely a result of introducing new parameters, rather than the parameters themselves being physically meaningful; secondly the increase in the number of parameters makes model fitting slower. If a model such as QSSn was adopted for estimating binding kinetics in a clinical environment each analysis would take dramatically longer than it would with L.

4 Conclusion and Outlook

In this thesis models have been created, structurally analysed and fitted to data. In particular the QSSn model has been demonstrated to provide a better fit to data from a commercial (IgM) antibody sample than any of the established models discussed, and for two binding types ($n = 2$) it has shown to have identifiability properties making it appropriate for use on a clinical samples.

However it also needs to be fitted to other commercial antibody types and clinical data before it could be accepted as the standard model for antibody binding in SPR experiments. A second hurdle will be extending the identifiability analysis for three binding types and a general case, this is because in clinical experiments there may be an unknown number of binding types. Additionally if models with increasing numbers of parameters are used, methods will be needed to estimate the parameters more efficiently, and models will have to be compared using statistical concepts like the Bayesian Information - Criterion or Akaike information criterion, which will require a detailed assessment of the likelihoods involved, and the non-normal errors observed.
5 Acknowledgements

This project was conducted on funding from the EPSRC, and with the aid of

References


A Appendix

A.1 Identifiability analysis
A.2 Programs

A.2.1 matlab

Programs are presented in this order

1. runFACSIM.m
2. startpointsQSS1.m
3. startpointsL1.m
4. compare.m
5. topfew.m
6. startpointsQSS2.m

The program runFACSIM.m allows the user to repeatedly run a single .fac file.

The programs startpointsQSS1.m and startpointsL1.m creates a number of folders containing different .txt files representing parameter values that are read in by a .fac file in the same folder, they differ in that they create parameter values for use by different models. Notably both of these create these starting parameters based on a number of rules, such as that the binding coefficients of a single reaction are of the same order of magnitude, whilst these may be false in reality, facsimile is capable of fitting parameters extremal quickly and varying them across orders of magnitude. These both output an index which is to be used with compare.m or topfew.m

The program compare.m finds the best parameter vectors from a set of files, and runs a number (n_top) of the best files a number of times (n_runs). It also checks for files are missing, have errors or that haven’t been run before, so the user can investigate any issues. This program outputs an ordered index which lists the folders with the best parameters to the worst parameters, excluding those that have problems.

The program topfew.m finds uses much of the code as compare.m to find the best files from an index. Rather than running the best files, it moves them to a new directory. This allows
the user to create a single set of parameter vectors for the QSS model, and copy the best ones to a new directory where they form the basis for the QSSn model.

The program startpointsQSS2.m creates a number of folders containing different .txt files representing parameter values that are read in by a .fac file in the same folder. It differs from the other startpoints programs because it needs to have the address of the files created by topfew.m, it uses the good combinations of parameters from the QSS1 model as the parameters it shares with that model, and in a new set of files combines these good parameters with values for the parameters unique to this model.
A.2.2 Facsimile

Templates for programs that are edited by matlab and copied into directories where they can be ran are presented in this order

1. HarryL1.fac
2. HarryQSS1.fac
3. HarryQSS2.fac

These are presented in an abbreviated form, and exclude the data that would be necessary to run them.