Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Using thermodynamic parameters to calibrate a mechanistic dose-response for infection of a host by a virus

Paul Gale

59 Fairway Avenue, Tilehurst, Reading, Berkshire, UK

ARTICLE INFO

Keywords:
Dose-response
Virus
Entropy
Mucin

ABSTRACT

Assessing the risk of infection from emerging viruses or of existing viruses jumping the species barrier into novel hosts is limited by the lack of dose response data. The initial stages of the infection of a host by a virus involve a series of specific contact interactions between molecules in the host and on the virus surface. The strength of the interaction is quantified in the literature by the dissociation constant \( K_d \) which is determined experimentally and is specific for a given virus molecule/host molecule combination. Here, two stages of the initial infection process of host intestinal cells are modelled, namely escape of the virus in the oral challenge dose from the innate host defenses (e.g. mucin proteins in mucus) and the subsequent binding of any surviving virus to receptor molecules on the surface of the host epithelial cells. The strength of virus binding to host cells and to mucins may be quantified by the association constants, \( K_a \) and \( K_{\text{mucin}} \) respectively. Here, a mechanistic dose-response model for the probability of infection of a host by a given virus dose is constructed using \( K_a \) and \( K_{\text{mucin}} \) which may be derived from published \( K_d \) values taking into account the number of specific molecular interactions. It is shown that the effectiveness of the mucus barrier is determined not only by the amount of mucin but also by the magnitude of \( K_{\text{mucin}} \). At very high \( K_{\text{mucin}} \) values, slight excesses of mucin over virus are sufficient to remove all the virus according to the model. At lower \( K_{\text{mucin}} \) values, high numbers of virus may escape even with large excesses of mucin. The output from the mechanistic model is the probability \( p_1 \) of infection by a single virion which is the parameter used in conventional dose-response models to predict the risk of infection of the host from the ingested dose. It is shown here how differences in \( K_a \) (due to molecular differences in an emerging viral strain or new host) affect \( p_1 \), and how these differences in \( K_a \) may be quantified in terms of two thermodynamic parameters, namely enthalpy and entropy. This provides the theoretical link between sequencing data and risk of infection. Lack of data on entropy is a limitation at present and may also affect our interpretation of \( K_a \) in terms of infectivity. It is concluded that thermodynamic approaches have a major contribution to make in developing dose-response models for emerging viruses.

1. Introduction

Microbiological risk assessment (MRA) requires a dose-response relationship to translate the exposure (i.e. number of pathogen particles entering the host through a given route) into the probability of infection. Infection by an oral pathogen is defined as the multiplication of organisms within the host, followed by excretion (Haas et al., 1999) and, for the purpose of the work here does not include progression of disease or the host acquired immune response. Obtaining dose-response data for humans has generally relied on volunteer challenge experiments e.g. *Cryptosporidium parvum* in students (Okhuysen et al., 1998) or using outbreak data to back-calculate the relationship between measured exposures and infection rates (Teunis et al., 2004). There are limitations to both approaches particularly with emerging pathogens for which the exposure routes may not be fully elucidated, and for pathogens with serious clinical outcomes, e.g. *Zaire ebolavirus* (EBOV).

Furthermore zoonotic viruses emerge through jumping the species barrier from an animal source to humans, e.g. Nipah virus (NiV) and...
EBOV, and in this respect the dose response would be for a one-off event that may be inefficient and difficult to reproduce without large numbers of animals. An additional complication is that the pathogen may adapt to the new host, such that its infectivity increases. This is well established for filoviruses in laboratory animals where the infectivity per plaque-forming unit (pfu) may change by several orders of magnitude with passage (Gale et al., 2016), and has recently been demonstrated for EBOV Makona adapting to humans through an amino acid substitution in its glycoprotein during the recent catastrophic outbreak in West Africa (Diehl et al., 2016; Urbanowicz et al., 2016). That outbreak also raised many questions regarding the unknown potential for companion animals (cats and dogs) to serve either as a reservoir or vector for the virus and so be involved in transmission of EBOV to humans and other animals. The absence of dose-response data for EBOV in humans limits development of MRAs for the risk of infection of citizens in the EU for example from EBOV in illegally imported bushmeat. Indeed, it has been proposed that the infectivity to humans of an EBOV pfu may differ not only from bushmeat samples from different wildlife species (e.g. fruits bats and nonhuman primates) but also from different individuals of the same species depending on the degree of host adaptation (Gale et al., 2016). In effect no two pieces of bushmeat from EBOV-infected wildlife may be the same in terms of infectivity to humans, although this remains to be proved. There is clearly a need for novel approaches to calibrate dose-response relationships for the purposes of MRA for emerging pathogens.

The infection process of a host cell can be broken down into the component steps and modelled mathematically (Handel et al., 2014) and the probability of infection can be expressed as a function of the combined probabilities of each step (Gale et al., 2014). These steps include overcoming the initial host defenses, binding of the virion to its host cell receptor, entry to the host cell (i.e. internalisation and uncoating of the virion), and replication, capsid assembly and budding (Gale et al., 2014). Previously it was demonstrated that a dose-response model could, in part, be parameterized using thermodynamic data for some of the key molecular interactions in the infection process (Gale, 2017). The beauty of thermodynamic data is that they can be measured experimentally by biochemists (in some cases just using molecular components e.g. cloned virus protein and host receptor protein (Wang et al., 2016)) and do not involve live animal or human volunteer studies, which is a major advantage for dangerous pathogens. Furthermore the effect of amino acid substitutions in the host receptors on binding affinity can be measured directly (Yuan et al., 2015). The possibility of applying thermodynamics is further developed here for two of the key steps in the infection process of a host by a virus. The first step modelled is the probability of the virus overcoming the innate host defenses posed by mucin protein molecules and the pathogen recognition receptors (PRRs) produced by the host. Mucins have sugar units on their surface which bind to components on the surface of the virus, for example the haemagglutinin (HA) glycoprotein molecules of influenza virus (de Graaf and Fouchier, 2014) or the VP1 of norovirus (NoV) (de Rougemont et al., 2011). Mucus present in the respiratory tract hampers influenza virus infection and in the case of humans predominantly contains α2,3-sialic acid receptors. Indeed influenza viruses with α2,3 specificity were inhibited by human mucins (de Graaf and Fouchier, 2014). The PRRs include the mannose binding protein (MBP) which has carbohydrate-recognition domains (CRD) which bind to regularly repeating sugar units on pathogen surface (Taylor and Drickamer, 2006). The second step modelled here is the binding of the virus to its specific receptors on the host cell surface. The approach here is developed for a generic faecal/oral virus such as NoV and rotavirus which infects epithelial cells lining the intestine (Boshuizen et al., 2005; de Rougemont et al., 2011), but could be applied to influenza A viruses which are inhaled and infect cells of the trachea and lung (de Graaf and Fouchier, 2014).

This paper first gives an overview of a mechanistic dose response model to introduce two probability parameters, namely the fraction, $F_v$, of virus escaping the mucin defense barrier and the fraction, $F_c$, of host cells with bound virus. The Methods section sets out a difference equation method to model $F_v$ and $F_c$ as a function of the mucin: virus ratio and virus dose in the intestine, respectively. Central to determining $F_v$ and $F_c$ is the strength of binding of the virus to the mucin and host cell as defined by the equilibrium constants $K_{\text{mucin}}$ and $K_r$, respectively. In the Theory section, the application of published data on the binding of the virus surface envelop glycoprotein (GP) to host cell receptor (Cr) molecules or to mucin molecules is reviewed in terms of determining $K_{\text{mucin}}$ and $K_r$ in order to parameterize the dose-response. Particular reference is made to using the dissociation constant $K_d$ which is routinely determined experimentally for virus GPs binding to Cr molecules (Gambaryan et al., 2005; Raman et al., 2014; Yuan et al., 2016). It is then shown how the strength of virus/host cell binding (i.e. the magnitude of $K_r$) may be predicted from changes in two thermodynamic parameters, namely enthalpy (H) and entropy (S). The effects of amino acid changes at the contact surfaces of the virus GP and Cr on the enthalpy are considered with a view to the future parameterization of dose-response models based on genetic sequencing data. Entropy changes are also considered both in terms of virus binding and also in the interpretation of $K_d$ data.

2. Methods

2.1. Overview of the development of a mechanistic dose-response model for infection in the intestine

The model parameters and variables are summarised in Table 1. On ingestion of the initial virus challenge dose, $V_{\text{initial}}$, by the host there are a number of immediate host defences in the mouth and gastrointestinal (GI) tract including the mucus barrier, decoy receptors and the innate immune system that selectively bind and hence remove the virus (McGuckin et al., 2011). For example the histoblood group antigens (HBGAs) are genetically determined glycans to which NoV selectively binds and are present on both decoy receptors in the saliva and on mucin, the major protein component of mucus (Shaneker et al., 2011). The total number of viruses surviving the mucin barrier, and getting through to the intestine is given by

$$V_{\text{intestine}} = F_v \times V_{\text{initial}}$$

(1)

where $F_v$ is the fraction of free virus, i.e. that not bound to mucin. As shown in Fig. 1, $F_v$ can be modelled by two parameters, namely the total number, $M_{\text{total}}$, of mucin molecules in the mucus in the saliva and GI tract and an association constant, $K_{\text{mucin}}$ (defined below) that quantifies the strength of binding of the virus to a mucin molecule. Thus by inserting $F_v$ from Fig. 1 for a given mucin concentration into Eq. (1), the total number of free virus particles in the intestine and available to initiate infection of the epithelium may be modelled. On reaching the intestinal epithelium, a free virus particle binds to the surface of a host cell. The probability of infection of the host, $P_{\text{host}}$, equals the probability of successful infection of at least one cell and is related to the number of cells (C.V) with bound virus by:-

$$P_{\text{host}} = 1 - (1 - P_{\text{cell}})^C\times V$$

(2)

where $P_{\text{cell}}$ is the probability of successful infection of a host cell given a virus has bound to its surface. Thus the more cells with bound virus then the greater the chance that infection will be successful in at least one of them. The probability $P_{\text{cell}}$ depends on ability of the bound virus to enter the cell, replicate and bud (Gale et al., 2014; Gale, 2017) and is not discussed further here. Now

$$C.V = E\times C_{\text{total}}$$

(3)

where $E$ is the fraction of cells with bound virus, and $C_{\text{total}}$ is the number of cells in the host intestinal epithelium. As shown in Fig. 2, $E$ is directly proportional to the total number of virus particles in the intestine, $V_{\text{intestine}}$, and is also dependent on the strength of the binding
Table 1
Summary of model parameters.

| Parameter | Description | Comments |
|-----------|-------------|----------|
| $P_{\text{host}}$ | Probability host organism is infected | Overall objective of dose-response |
| $p_1$ | Probability of infection from ingestion of a single virion by the host | Parameter to be obtained from mechanistic dose response approach developed here for direct use in conventional dose response model (Eq. (16)) |
| $P_{\text{cell}}$ | Probability that a host cell becomes infected given virus has bound to its surface | Not discussed further here |
| $V_{\text{initial}}$ | Challenge dose of virus to the host organism | Oral exposure in MRA. $[V_{\text{initial}}]$ is the concentration of total virus in the simulated intestine at $2.19 \times 10^{-15}$ M |
| $V_{\text{intestine}}$ | Total number of viruses not bound to mucin, and getting through to the intestine to initiate infection of host cell. | Within intestine, $V_{\text{intestine}}$ includes both virus bound to host cells and not bound to host cells, i.e $V_{\text{intestine}} = C.V + V_{\text{free}}$. |
| $V_{\text{Muc}}$ | Number of viruses bound to mucin | This is varied relative to the fixed virus challenge dose $(V_{\text{initial}})$ in Fig. 1 |
| $M_{\text{mucin}}$ | Number of mucin molecules with no bound virus | Probability virus breaks through mucin barrier into intestine |
| $M_{\text{mucin total}}$ | Total number of mucin molecules in the host mucus in the host saliva and GI tract | Probability that cell has virus bound to it |
| $F_{v}$ | Fraction of virus that is not bound to mucin, i.e. survives to infect host cells | In the simulations, $C_{\text{total}}$ is constant at $4.15 \times 10^8$. Concentration of total host cells $[C_{\text{total}}]$ is $2.19 \times 10^{-15}$ M. |
| $C_{\text{free}}$ | Number of host cells without bound virus, i.e. free $[C_{\text{free}}]$ is concentration (M) of free cells (i.e. cell with no virus attached) | $[C_{\text{free}}]$ is concentration (M) of cells with no virus attached |
| $C.V$ | Number of host cells with bound virus $[C.V]$ is concentration (M) of cells with virus attached | $[V_{\text{free}}]$ is concentration (M) of free virus (i.e. not bound to cells) |
| $V_{\text{initial}}$ | Challenge dose of virus to the host organism | Units M. The smaller $K_d$ in magnitude, the stronger the binding. |
| $C_{\text{total}}$ | Number of cells in the host intestinal epithelium | Units M$^{-1}$. Related to reciprocal of $K_a$ (Eq. (17)) |
| $V_{\text{intestine}}$ | Total number of viruses not bound to mucin, and getting through to the intestine to initiate infection of host cell. | Units M. The smaller $K_a$ in magnitude, the stronger the binding. |
| $F_{c}$ | Fraction of host cells $(C_{\text{total}})$ with bound virus | Probability that cell has virus bound to it |
| $F_{c}$ | Fraction of host cells $(C_{\text{total}})$ with bound virus | In the simulations, $C_{\text{total}}$ is constant at $4.15 \times 10^8$. Concentration of total host cells $[C_{\text{total}}]$ is $2.19 \times 10^{-15}$ M. |
| $p_{\text{cell}}$ | Probability that a host cell becomes infected given virus has bound to its surface | Overall objective of dose-response |
| $p_1$ | Probability of infection from ingestion of a single virion by the host | Parameter to be obtained from mechanistic dose response approach developed here for direct use in conventional dose response model (Eq. (16)) |
| $K_{\text{mucin}}$ | Association constants for binding of virus to cell and mucin respectively | Not discussed further here |
| $K_{\text{d}}$ | Dissociation constant measured experimentally between individual virus molecules and host molecules. | Probability virus breaks through mucin barrier into intestine |

interaction between the virus and the epithelial cell surface as quantified by the association constant, $K_a$ which is now defined.

Expressing the fraction, $F_{v}$, of viruses not bound to mucus and the fraction, $F_{c}$, of host cells with bound virus in terms of the association constants, $K_{\text{mucin}}$ and $K_a$.

Each binding process is represented by a dynamic equilibrium. Thus, within the given volume of the intestine, free virus $(V_{\text{free}})$ and host cells with no bound virus $(C_{\text{free}})$ are in dynamic equilibrium with cells with bound virus $(C.V)$ as represented by:-

$$C_{\text{free}} + V_{\text{free}} \leftrightarrow C.V$$

The association constant, $K_a$, is expressed in terms of the concentrations (Gale, 2017) as:-

$$K_a = \frac{[C.V]}{[V_{\text{free}}][C_{\text{free}}]}$$

(5)

where the square brackets, $[ ]$, represent the concentration in moles dm$^{-3}$ (M). The term “in dynamic equilibrium” means the process is reversible (Handel et al., 2014) such that free virus binds to free host cells to form C.V with a rate, $k_{on}$, and the C.V complexes then dissociate into $C_{\text{free}}$ and $V_{\text{free}}$ at a slower rate, $k_{off}$, depending on the strength of binding. Thus the association constant, $K_a$ (and similarly $K_{\text{mucin}}$) may also be written in terms of the association/dissociation rates as:-

$$K_a = \frac{k_{on}}{k_{off}}$$

(6)

Visualising $K_a$ (or $K_{\text{mucin}}$) in terms of on/off rates may be easier conceptually and has implications when host or virus factors selectively change $k_{on}$ or $k_{off}$ (see below) thus affecting $K_a$ (or $K_{\text{mucin}}$) according to Eq. (6). As shown by Gale (2017), the fraction, $F_{c}$, of host cells with
bound virus is given by:-

$$F_i = \frac{[C \cdot V]}{[C_{free}] + [C \cdot V]}$$

Substituting $[C \cdot V]$ with Eq. (5) and rearranging gives $F_i$ in terms of the free virus concentration $[V_{free}]$ and $K_{s}$:-

$$F_i = \frac{[V_{free}]}{[V_{free}] + 1/K_{s}}$$  (7)

Similarly, when the virus enters the host there will be a dynamic equilibrium between virus that is not bound to mucin and hence reaches the intestinal epithelium ($V_{intestine}$), free mucin ($Muc_{free}$) and virus bound to mucin ($V \cdot Muc$).

$$V_{intestine} + Muc_{free} \leftrightarrow V \cdot Muc$$  (8)

The strength of the binding is reflected by the association constant, $K_{mucin}$, between virus and mucin and is expressed as:-

$$K_{mucin} = \frac{[V \cdot Muc]}{[V_{intestine}] [Muc_{free}]}$$  (9)

The fraction of free virus ($F_i$) is given by:-

$$F_i = \frac{[V_{intestine}]}{[V_{intestine}] + [V \cdot Muc]}$$  (10)

Replacing $[V \cdot Muc]$ with Eq. (9) and rearranging gives $F_i$ in terms of the free mucin concentration and $K_{mucin}$:-

$$F_i = \frac{1}{1 + K_{mucin} [Muc_{free}]}$$  (11)

### 2.2. Modelling how $F_i$ varies with dose of virus in the intestine

The objective is to construct a plot of $F_i$ as a function of $V_{intestine}$ as shown in Fig. 2. Eq. (7) may be used with $[V_{free}]$ representing $V_{intestine}$ when the total number of virus particles, $V_{intestine}$, greatly exceeds the number of host cells ($C_{total}$) such that $[V_{free}]$ is relatively unaffected by virus binding and therefore $[V_{free}]$ approximates $V_{intestine}$. This is also acceptable for low binding affinity viruses as in the species barrier model of Gale (2017) such that very little virus is bound even at high virus doses. However, in many natural infection processes (e.g. through drinking water), the host may be challenged by very low numbers of pathogen which bind with high affinity to host cells such that $[V_{free}]$ is greatly diminished compared to $V_{intestine}$ and tends to zero as all the virus is bound. The problem is that Eq. (7) cannot be expressed mathematically in terms of $V_{intestine}$. The solution adopted here to model $F_i$ in terms of $V_{intestine}$ comprises three steps. The first step involves setting up a model for a host intestine so that the concentrations of virus and host cells may be defined in order to calculate $K_{s}$ using Eq. (5). In the second step, a difference equation approach is used to produce a range of $[V_{free}]$, $[C_{free}]$ and $[C \cdot V]$ combinations at six different total virus ($V_{intestine}$) doses. From these concentrations, $K_{s}$ values and $F_i$ values are calculated and are plotted in Fig. 3. In the third step, for each of the six $V_{intestine}$ doses, the $F_i$ is read off Fig. 3 for given $K_{s}$ values. $F_i$ is then plotted against $V_{intestine}$ for each $K_{s}$ in Fig. 2.

#### 2.2.2. Virus doses used in the simulation

Simulations were undertaken for six virus doses ($V_{intestine}$) from $1.0 \times 10^{3}$ to $4.15 \times 10^{11}$ virions in the model intestine (Fig. 2), representing virus: host cell ratios from $6:1$ to $1000:1$ and concentrations from $5.3 \times 10^{-21}$ M to $2.2 \times 10^{-12}$ M.

#### 2.2.3. Difference equation approach to model $F_i$ against $K_{s}$ for each virus dose (Fig 3)

For low virus doses (i.e. $V_{intestine} < C_{total}$), $K_{s}$ were calculated using $[C \cdot V]$, $[V_{free}]$ and $[C_{free}]$ in Eq. (5) over the full range of virus binding i.e. from one bound virus to all $V_{intestine}$ viruses being bound. According to Eq. (4), each cell can only bind one virus, and the number of cells with bound virus therefore equals the number of viruses bound to cells which was calculated as:-

$$C \cdot V = V_{intestine} - V_{free}$$  (12)

over the range of $V_{free}$ from 0 (all viruses bound) to $V_{intestine}$ (no viruses bound). For high virus doses (i.e. $V_{intestine} > C_{total}$) the simulation was run for $V_{free}$ from 0 to $C_{total}$ (such that C.V is always positive) because the model in Eq. (4) assumes each cell can only bind one virus. For each C.V, the number of free cells, $C_{free}$, was calculated as:-

$$C_{free} = C_{total} - C \cdot V$$

Values for $[C_{free}]$, $[V_{free}]$ and $[C \cdot V]$ were calculated as $C_{free}$, $V_{free}$ and C.V respectively, divided by the volume of the intestine (in dm$^3$) and L and then used to calculate $K_{s}$ in Eq. (5) over the range of C.Vs. For each virus challenge dose, values of $F_i$ are plotted as a function of $K_{s}$ in
Fig. 3 where $F_v$ is calculated as:

$$F_v = \frac{C \cdot V}{C_{\text{total}}},$$  

(13)

and $C_{\text{total}}$, is constant at $4.15 \times 10^9$ cells (Table 1).

2.3. Modelling how $F_v$ varies with ratio of mucin to virus: surviving the mucin defence

The fraction of free virus, $F_v$, represents the probability that virus is not bound to the mucin and in effect escapes the mucin barrier. The objective is to model $F_v$ as a function of the mucin to virus ratio over a range of $K_{\text{mucin}}$ values as shown in Fig. 1. The simple approach to model $F_v$ is to use Eq. (11) with $[\text{Muc}_{\text{free}}]$ representing the total concentration of mucin in the intestine, $[\text{Muc}_{\text{total}}]$, which can be measured in a host experimentally. As discussed below, this fails at mucin: virus ratios of $<1:1$. Therefore a difference equation approach was used. For this, the challenge dose, $V_{\text{initial}}$, in the host intestine was fixed at $4.15 \times 10^8$ virions and the fraction of virus not bound to mucin, $F_v$, was then calculated for ten values of $\text{Muc}_{\text{total}}$ ranging from $10^3$ mucin molecules to $4.15 \times 10^{12}$ mucin molecules, representing: virus ratios ranging from $2.4 \times 10^{-6}:1$ to $10,000:1$. This was done using the same stages as described above for virus binding to host cells for seven $K_{\text{mucin}}$ values ranging from $10^9$ to $10^{22} \text{ M}^{-1}$. Thus for the number of free mucin molecules, $\text{Muc}_{\text{free}}$, ranging from 0 (i.e. all mucin molecules bound to virus) to $\text{Muc}_{\text{total}}$ (i.e. all mucin molecules free of virus), values of $V_{\text{Muc}}$ and $V_{\text{intestine}}$ were calculated as:

$$V_{\text{Muc}} = \text{Muc}_{\text{total}} - \text{Muc}_{\text{free}},$$  

(14)

$$V_{\text{intestine}} = V_{\text{initial}} - V_{\text{Muc}},$$  

(15)

By dividing by the volume of the intestine and $L$, the values of $V_{\text{Muc}}$, $V_{\text{intestine}}$ and $\text{Muc}_{\text{free}}$ were converted to corresponding concentrations, namely $[V_{\text{Muc}}]$, $[V_{\text{intestine}}]$ and $[\text{Muc}_{\text{free}}]$, from which $K_{\text{mucin}}$ Values were calculated using Eq. (9). For each $K_{\text{mucin}}$ value, $F_v$ was calculated as:

$$F_v = \frac{V_{\text{intestine}}}{V_{\text{initial}}},$$  

(15)

and a plot (not shown) of $F_v$ versus $K_{\text{mucin}}$ constructed in the same was as for virus binding to host cells in Fig. 3. From that plot, values of $F_v$ for each of the ten $\text{Muc}_{\text{total}}$ were read off for a given $\text{Muc}_{\text{total}}$ and plotted against the mucin molecule:virus ratio (i.e. $\text{Muc}_{\text{total}}/V_{\text{initial}}$) in Fig. 1.

2.3.1. Effects of stochasticity

Stochasticity in the challenge dose ($V_{\text{initial}}$) would be addressed in the exposure calculation and is outside the scope of this work. The mechanistic dose-response model developed here using the difference equation approach is not affected by stochasticity because the values of $V_{\text{Muc}}$ and $V_{\text{intestine}}$ calculated in Eqs. (12) and (14) respectively are “given” integers, and thus $F_v$ and $F_v$ calculated by Eqs. (13) and (15) respectively are exact for each integer $V_{\text{Muc}}$ and $V_{\text{intestine}}$ ($C_{\text{total}}$ and $V_{\text{initial}}$ being constant integers in the simulation). The probability of infection of the host according to Eq. (2) is calculated for each integer value of $V_{\text{Muc}}$ from 0, 1, 2, … to $C_{\text{total}}$ or $V_{\text{intestine}}$ (depending on which is lower) and is not affected by stochasticity. The effect of stochasticity is considered here for the use of Eq. (7) to calculate $F_v$ and where $V_{\text{Muc}}$ is a fraction.

2.3.2. Spatial heterogeneity in the model of the intestine

The model assumes that the free pathogens and mucins are homogeneously distributed within the lumen of the intestine, such that the concentrations at equilibrium are constant along the 1 m length of the simulated intestine. The spatial heterogeneity of the pathogen in the lumen is not known and could vary depending on the pathogen distribution in the food, water or faecal/vomit contamination ingested by the host. For example, ingestion of a small amount of faeces laden with virus could give much higher virus concentrations at certain parts of the lumen, depending on the degree of mixing within the lumen. In contrast ingestion of a $0.314 \text{ dm}^3$ volume of water contaminated with pathogen could give a homogeneous (Poisson) distribution along the 1 m length. This is not considered further here. Spatial heterogeneity will exist in the mucin concentration because the mucus forms a layer lining the intestine wall (McGuckin et al., 2011). This needs consideration in the further development of this mechanistic approach.

3. Results

Here an intestine is simulated with an internal volume of $0.314 \text{ dm}^3$ and a total of $4.15 \times 10^8$ susceptible cells in the intestinal epithelium. In Fig. 1, an oral challenge dose of $4.15 \times 10^8$ viruses is administered, and the fraction of these escaping the mucin barrier and reaching the epithelium is modelled. Fig. 3 shows the fraction of epithelium cells with bound virus as a function of $K_{\text{v}}$ for six doses of virus that have got through the mucin barrier ranging from 1000 virions to $4.15 \times 10^{11}$ virions. In Fig. 2, the points from Fig. 3 are replotted in the form of a dose-response which relates the fraction of host cells with bound virus to the dose of virus that has got through the mucin barrier. These dose-response relationships are presented for a range of $K_{\text{v}}$ values from $10^5$ to $10^{20} \text{ M}^{-1}$ in Fig. 2.

3.1. Assessing the magnitude of $K_{\text{mucin}}$ needed for an effective mucus barrier

The fraction, $F_v$, of virus escaping the mucus barrier is plotted as function of the mucin: virus ratio for a range of $K_{\text{mucin}}$ values from $10^9$ to $10^{22} \text{ M}^{-1}$ in Fig. 1. The horizontal dotted line in Fig. 1 represents just one free virus remaining in the $0.314 \text{ dm}^3$ volume of the simulated intestine. At values of $F_v$ below this line there is < 1 free virus in the simulated intestine and in effect all of the $4.15 \times 10^8$ virions in the initial challenge dose ($V_{\text{initial}}$) are bound to mucin. Thus, the effectiveness of the mucin barrier can be assessed simply in terms of $K_{\text{mucin}}$ and the mucin: virus ratio required to bring $F_v$ to below this line in Fig. 1. At the very high $K_{\text{mucin}}$ value of $10^{22} \text{ M}^{-1}$ all of the virus is bound as the mucin: virus ratio exceeds 1:1. However, at progressively lower $K_{\text{mucin}}$ values, less and less of the virus is bound at a given mucin: virus ratio. Thus even at $K_{\text{mucin}}$ values as high as $10^{18}$ and $10^{20} \text{ M}^{-1}$ large numbers of viruses (20,330 and 180 respectively) are still free at mucin: virus ratios of 10:1. According to the model in Fig. 1, in the $0.314 \text{ dm}^3$ volume of simulated intestine, large excesses of mucin over virus are needed to make a significant impact on the fraction of free virus at the lower $K_{\text{mucin}}$ values. For example, at $K_{\text{mucin}}$ values of $10^{13} \text{ M}^{-1}$ and $10^{17.2} \text{ M}^{-1}$ 4% and 50% of the virus is free, respectively, at mucin: virus ratios of 1000:1. Thus mopping up virus is relatively inefficient and, except at very high $K_{\text{mucin}}$ values, large numbers of virus may break through. This reflects the dilution in the large volume of the simulated intestine. The main conclusion from Fig. 1 is that as the mucin: virus ratio increases from < 1:1 through 1:1 to > 1:1 (i.e. going from left to right along the x-axis) for high $K_{\text{mucin}}$ values ($>10^{17} \text{ M}^{-1}$), then the fraction of free virus falls dramatically and non-linearly, with all of the virus being mopped up at $K_{\text{mucin}}$ of $\sim 10^{22} \text{ M}^{-1}$.

3.1.1. The virus binding capacity of the mucin is finite

As expected at mucin: virus ratios < 1:1, the fraction of free virus approaches 100%, even with very high $K_{\text{mucin}}$ values (Fig. 1). This simply reflects the fact that there is not enough mucin capacity to mop up all the virus such that very high virus challenge doses overwhelm the mucin barrier. With very high $K_{\text{mucin}}$ values ($10^{22} \text{ M}^{-1}$), a very slight excess of mucin over virus is sufficient to mop up all the virus. At intermediate $K_{\text{mucin}}$ values ($10^{18}–10^{20} \text{ M}^{-1}$) mopping up the last remaining viruses is less efficient and proportionately higher mucin: virus ratios are required at lower $K_{\text{mucin}}$ values.
3.1.2. Simplifying the approach: using Eq. (11) to model Fc in MRA

The total mucin: virus ratio is also expressed as the concentration of total mucin, [Muctotal], on the x-axis scale of Fig. 1. This is relative to the virus concentration in the intestine which is fixed at 2.2 × 10^{-15} M. Thus 4.15 × 10^8 virus particles in a volume of 0.314 dm^3 on dividing by L represent a concentration of 2.2 × 10^{-15} M. The dashed lines in Fig. 1 show Fc, as calculated by Eq. (11) for each K_mucin value using [Muctotal] as an approximation for [Mucfree]. This is an appropriate approximation at high mucin to virus ratios. However, at mucin: virus ratios of <1:1, [Mucfree] becomes much less than [Muctotal] as all the mucin is bound, particularly at high K_mucin Values, and Eq. (11) is not applicable. As an example, the arrow in Fig. 1 shows that Eq. (11) predicts that 4.4% of virus is free (i.e. 95.6% of virus is bound) when there is 10,000-fold more virus particles than mucin molecules. This is clearly not possible. At lower K_mucin values, Fc decreases linearly with increasing mucin: virus ratio and may be modelled by Eq. (11) in good agreement with Fc calculated by the difference equation approach (symbols in Fig. 1). Thus Eq. (11) holds at low K_mucin values or at high mucin: virus ratios (irrespective of K_mucin).

This is important because high mucin: virus ratios might be expected in a natural infection situation even with the high virus challenge dose used in the simulation here. From the practical point of developing MRA methodology, applying Eq. (11) is easier than the complicated difference equation approach developed for the symbols in Fig. 1. Thus understanding the limitations of Eq. (11) is important. In summary Eq. (11) fails at mucin: virus ratios of <1:1 at the higher K_mucin values, as represented by the arrow, but is generally applicable at lower K_mucin values particularly at higher mucin: virus ratios. The results show that the value of K_mucin is critical not only in determining whether Eq. (11) can be used in MRA but also in assessing how effective the mucin barrier is in mopping up viruses.

3.2. Fc is related to both Ka and the virus dose in the intestine

The fraction of host cells with bound virus increases with the dose of virus in the simulated intestine for all K_a values until at high virus doses the host cells become saturated (Fc → 1) with virus (Fig. 2). At low K_a values (<10^{13} M^{-1}) unrealistically high virus doses (>10^{12}) are required for saturation of the host cells. From Fig. 3, Fc increases linearly with K_a at a given virus dose but, not surprisingly, is limited by the virus dose at V_{intestine} C_{total} ratios of <1:1. Thus, there are 4.15 × 10^8 cells in the simulated intestine, and for the virus doses of 10^3 and 10^6 virions, the maximum values achievable for Fc are 2.4 × 10^{-6} and 2.4 × 10^{-3} respectively when all those virions are bound. At K_a values of >10^{15} M^{-1}, the binding is so strong that all the virions in the dose are bound (Fig. 3) and Fc reaches its plateau. At V_{intestine} C_{total} ratios >1 (i.e. the 4.15 × 10^9 and 4.15 × 10^{11} viruses in Fig. 3 representing V_{intestine} C_{total} ratios of 100:1 and 1000:1, respectively) saturation of the host cells occurs at high K_a values with progressively lower K_a values required for saturation of the host cells as the virus dose increases. Thus increasing K_a above certain values has no effect on Fc, either because all the viruses are bound at low virus doses (i.e. V_{intestine} C_{total} ratios <1) or because all the host cells are saturated at high virus doses (i.e. V_{intestine} C_{total} ratios >1). This is borne out in Fig. 2 which shows that for K_a values >10^{15} M^{-1}, the dose-response curves are superimposed such that Fc is limited by available dose and not by the K_a. The main conclusion of Figs 2 and 3 is that the value of K_a over the range 10^9 to 10^{15} M^{-1} is critical in affecting Fc.

3.2.1. Simplifying the approach: using Eq. (7) to model Fc in MRA

The total virus dose in the intestine (V_{intestine}) may be approximated to V_{free}, converted to a concentration and used as [V_{free}] in Eq. (7) to calculate Fc, as an alternative to the difference equation approach used to construct Fig. 2. However, this simplification fails at K_a values greater than 10^{15} M^{-1} (results not shown) for which Eq. (7) over estimates Fc by orders of magnitude at low virus doses. The results show that Eq. (7) is appropriate to calculate Fc up to K_a values of at least 10^{13} M^{-1} irrespective of dose.

3.3. Demonstration of the potential application of the model

The application of the model is demonstrated in Table 2 for four scenarios representing different combinations of initial challenge dose and affinity of the virus for the cell (as determined by K_a). Thus the K_a is 10^{10} M^{-1} for the low affinity binding virus and 10^{13} M^{-1} for the high affinity binding virus. These broadly reflect those measured experimentally (Nunes-Correia et al., 1999) for low binding and high affinity binding sites for influenza A virus H5N1 binding to canine kidney cells (Table 3). Also as discussed below, this difference of 1000-fold in K_a could reflect the result of a single amino acid change in the GP or Cr molecule affecting a salt bridge at the GP/ Cr interface. For the purpose of the demonstration, p_{cell} is assumed to be 0.1, i.e. a cell with bound virus has a 10% chance of successful infection. For these scenarios, the predicted probability of the host being infected, p_{host}, ranges from 4.4 × 10^{-3} to 0.988. Fc is constant at 0.19 (Table 2) and is calculated using Eq. (11) with a K_mucin of 10^{-13} M^{-1} (at which Eq. (11) is appropriate) and a [Mucfree] of 4.25 × 10^{-9} M. This value of [Mucfree] is calculated using a mucin mass concentration of 1.7 mg/ml for saliva (Kejriwal et al., 2014), a dilution factor of 100-fold in the food or water matrix and a mucin protein molecular weight of 4000,000 Daltons (Resimer and Sheehan 2012). It is solely calculated for demonstration purposes in Table 2 in the absence of published data. Fc in Table 2 varies for each scenario and is calculated using Eq. (7) with [V_{free}] calculated from V_{intestine} (in the 0.314 dm^3 volume of the simulated intestine) and L. The values of K_a used in these scenarios are low enough such that Eq. (7) is applicable (see above) as can be seen for the K_a = 10^{13} M^{-1} line from the difference equation approach in Fig. 2 where Fc ≈ 1.0 × 10^{-7} for a V_{intestine} of 1904 virions in agreement with Table 2.

3.3.1. Effect of stochasticity

The approach using Fc from Eq. (7) predicts fractions of a C.V (Table 2) for both low dose scenarios and also for the high dose, low affinity scenario. In reality CV must be an integer as in the difference equation approach. The issue of stochasticity arises in the model.
which p host = 0.5. It was found that for the high a
for the ID50 is 6.6 and, being much greater than 1, is in the range where
(Ka = 1.010 M
infection of a host is written as:-

\[ \text{stochasticity is less of an issue. The classic dose response model for} \]
application in Table 2 when C.V is a fraction. To investigate the effect of
stochasticity, the mechanistic model developed in Table 2 is used to
determine the infectious dose 50% (ID50), i.e. the value of \( V_{\text{initial}} \) for
which \( p_{\text{host}} = 0.5 \). It was found that for the high affinity virus
(Ka = 1.013 M
The values of \( p_{\text{host}} \) are also calculated in Table 2 by using \( p_1 \) (obtained
from the mechanistic model as described above) directly in Eq.
(16) to estimate the risk from the initial challenge dose \( (V_{\text{initial}}) \). The
results are identical with those predicted by the mechanistic method.

\[ p_{\text{host}} = 1 - (1 - p_1) V_{\text{initial}} \]  

(16)
This not only provides further evidence that stochasticity is not an issue but also demonstrates how the output from the mechanistic model can be applied directly to conventional dose-response models.

4. Theory

4.1. Parameterising the model: determining the virus/host cell association constant, \( K_a \)

Three approaches for parameterising \( K_a \) are set out in Table 3. In a few cases, the \( K_a \) is determined experimentally by directly measuring virus binding to cells (Table 3). Approach 2 developed here is to estimate \( K_a \) from published \( K_d \) values which are routinely measured experimentally. Thus, the magnitude of \( K_a \) for binding of a virus to its host cell is determined by the strength of the interaction(s) between viral surface components, typically the viral glycoprotein (GP) in the case of enveloped viruses such as EBOV, and the receptors (Cr), e.g. T-cell immunoglobulin and mucin domain protein 1 (TIM-1) for EBOV, on the human host cell surface (Yuan et al., 2015). Biochemists quantify the binding affinity between molecules in terms of the dissociation constant (Price and Dwek, 1979):

\[
K_d = \frac{[GP][Cr]}{[GP \cdot Cr]}
\]

for the reversible dissociation process given by:-

\[
GP \cdot Cr \leftrightarrow GP + Cr
\]

The magnitude of \( K_d \) in units of mol/dm\(^3\) (M) may be determined experimentally between purified parts of the virus GP and the host Cr using surface plasmon resonance (SPR) in which one protein “partner” is immobilized to a chip surface and changes in refractive index are used to measure binding of the other “partner” which is free in solution. Thus, Yuan et al. (2015) immobilized EBOV-GP to a chip and used SPR to measure a \( K_d \) of \( 2.67 \times 10^{-10} \) (M) for EBOV, on the human host cell surface (Yuan et al., 2015). Biochemists quantify the change in entropy on whole virus binding, \( \Delta S_{sw} \) (discussed below) \( K_a \) may be expressed as:-

\[
K_a = f \left( \frac{1}{K_d^4} \Delta S_a \right)
\]

(17)

The value of \( G \) is affected by temperature and pressure, and for this reason \( G \) is used by biochemists. Thus providing the temperature and pressure are constant, as assumed here for virus in the host intestine, then any changes in \( G \) relate only to those changes in energy within the system of interest, namely from the molecular interactions between virus and host during infection within the simulated intestine.

4.1.2. Estimating \( K_a \) through the thermodynamic parameters, enthalpy and entropy

Approach 3 in Table 3 is to calculate the association constant \( K_a \) through the changes in two thermodynamic parameters, namely enthalpy (H) and entropy (S) on binding. Thus \( K_a \) is related to \( \Delta G \) for association of the virus with the cell (Gale, 2017; Kastrits and Bonvin, 2013) according to:

\[
\Delta G_a = -RT\ln K_a = \Delta H_a - T\Delta S_a
\]

(18)

which on rearranging gives

\[
K_a = e^{\frac{\Delta G_a}{RT}} = e^{\frac{-\Delta H_a}{RT} + \frac{\Delta S_a}{R}}
\]

(19)

where \( \Delta H_a \) and \( \Delta S_a \) are the changes in enthalpy and entropy, respectively, of the virus/host cell system on association (Carneiro et al., 2002; Gale, 2017). Therefore in principle knowing \( \Delta H_a \) and \( \Delta S_a \) would enable calculation of \( K_a \) and hence \( F_c \) for a given dose of virus from Fig. 2. Bostrom et al. (2011) demonstrated how the binding affinity of an antibody could be broken down into the enthalpy and entropy terms. How this approach could be applied to a virus during infection with the aim of calibrating dose-response models is summarized as Approach 3 in Table 3. To the author’s knowledge there are no data for \( \Delta H_a \) and \( \Delta S_a \) for viruses binding to host cells. However, Carneiro et al. (2002) measured the forces between vesicular stomatitis virus (VSV) and an artificial phospholipid (PL) bilayer (as opposed to a host cell) and obtained values for \( \Delta H_a \) and \( \Delta S_a \) of \(-4958\) kJ/mol and \(-16,062\) kJ/mol/k respectively. Carneiro et al. (2002) explain the huge value for \( \Delta H_a \) as the result of the cumulative \( \Delta H_a \) from multiple interactions due to the dense packing of G protein on the VSV virus surface. Indeed Carneiro et al. (2002) estimate seven G proteins’ binding per VSV and calculate the \( \Delta H_a \) and \( \Delta S_a \) values for binding of a single G protein at \(-694\) kJ/mol and \(-2196\) kJ/mol/k respectively.

4.2. The importance of understanding entropy changes during virus binding to a cell for calculating \( K_a \)

Approaches 2 and 3 in Table 3 are dependent on estimating \( \Delta S_a \). As shown by Carneiro et al. (2002) for VSV binding to PL bilayers, the decrease in entropy is huge (Table 3), demonstrating the importance of understanding \( \Delta S_a \) in estimating \( K_a \). Unlike \( \Delta H_a \) which is based on molecular contacts at the binding faces, \( \Delta S_a \) involves changes in order and mobility and is difficult to visualise, but may be broken down into a number of components which act additively (Bostrom et al., 2011). Thus

\[
\Delta S_a = \Delta S_{solvent} + \Delta S_{fold} + \Delta S_{n} + \Delta S_{mem}
\]
where $\Delta S_{\text{solvent}}$ is the change in entropy of the water solvent molecules, $\Delta S_{\text{conf}}$ is the change in the internal conformational freedom of the proteins, $\Delta S_{R}$ is the change in rotational and translational freedom of the virus, and $\Delta S_{\text{membrane}}$ is an entropic pressure associated with bringing two membranes close together, as for example, when the virus envelope approaches the host cell membrane prior to fusion (Sharma, 2013). The likely contributions from the four entropy component terms are summarized in Table 3.

4.2.1. The change in conformational entropy ($\Delta S_{\text{conf}}$): intrinsically disordered proteins in viruses and their role in ligand binding

Many ligand binding proteins have intrinsically disordered regions (IDRs) which undergo a disorder-to-order transition on or near the interface on binding the ligand (Fong et al., 2009) decreasing the entropy such that $\Delta S_{\text{conf}}$ is negative (Rajasekaran et al., 2016). Viral proteins are rich in intrinsic disorder (Dolan et al., 2015) and intrinsically disordered proteins serve as host cell receptors for viruses. An example is the ephrin receptor ligand binding domain (Fong et al., 2009) which also serves as receptor for Hendra virus (HeV) and NiV (Xu et al., 2012). Thus the unbound ephrin receptor contains partially disordered loops. In the complex (bound to ephrin) these loops are ordered to form the ligand-binding channel (Fong et al., 2009). This raises the question of whether the ephrin receptor undergoes a disorder-to-order transition on HeV/NiV binding. Interestingly Xu et al. (2012) show from the X-ray crystal structure that the binding of HeV G protein involves the movement of a tryptophan “latch” on the ephrin-B2 receptor. There would be a change in the conformational entropy $\Delta S_{\text{conf}}$ associated with this during formation of the HeV G protein/ephrin-B2 receptor complex.

According to Eq. (17), increasing the number of GP/Cr contacts during virus binding increases $K_a$, and hence the infectivity according to Fig. 3. Counterintuitively, in the case of EBOV GP pseudoviruses, increasing the surface density of GP actually decreases the infectivity by about 10-fold, perhaps reflecting steric hindrance from tightly packed GP proteins (Mohan et al., 2015) blocking a conformational change that gives an increase in $\Delta S_{\text{conf}}$ required for binding of GP to TIM-1 (Gale, 2017). This is consistent with the observation that EBOV GPs are well separated in space on the EBOV filament surface (Beniac et al., 2012) thus allowing plenty of space for conformational changes without steric hindrance. A demonstration of how changes in $\Delta S_{\text{conf}}$ could explain the 10-fold increase in infectivity reported by Mohan et al. (2015) on decreasing the GP density may be explored with Eq. (19). Thus increasing $S_{\text{conf}}$ by 20 J/mol/K (irrespective of the values of $\Delta H_{\text{ex}}$ and $T$) increases $K_a$ by 11-fold. A decrease in conformational entropy ($\Delta S_{\text{conf}}$) due to steric hindrance of 20 J/mol/K represents the ordering of just four amino acid residues on the basis of the $\sim 6.1$ J/mol/K per residue proposed by Rajasekaran et al. (2016) for disordered proteins. Therefore blocking the disordering of just four amino acids in the EBOV GP could explain the 10-fold decrease in infectivity of EBOV GP pseudoviruses with high GP densities.

Booth et al. (2013) show that filovirus filaments are very flexible, more flexible than filamentous paramyxoviruses and much more flexible than rhabdoviruses. Thus a flexible filament, making multiple GP/Cr contacts, can adopt many more conformations in 2D-space when bound to a host cell surface than a rigid rod of the same length. A bound filament will therefore have a higher entropy than a stiff rod, and will have a higher $K_a$. A flexible filament can also test out multiple GP/Cr contacts, thus maximising contacts and making $\Delta H_{\text{ex}}$ more negative in magnitude which increases $K_a$ according to Eq. (19). Indeed Booth et al. (2013) suggest that the extreme aspect ratio of filaments may be an adaptation that enhances cellular attachment.

4.2.2. The change in rotational and translational entropy ($\Delta S_{R}$)

The association of two species to form a bound complex, e.g. the binding of a ligand to a protein or the adsorption of a peptide on a lipid membrane, always involves an entropy loss (Ben-Tal et al., 2000). This no doubt applies to the binding of viruses to cells. Consideration of $\Delta S_{R}$ may be of particular importance for binding of filoviruses (Gale, 2017) which comprise repeating modular units linked together into single very long filaments (Benniac et al., 2012; Booth et al., 2013). Thus Gale (2017) argued that the prior immobilisation of multiple virus units through linking together into a polyprotid filament enhanced cell binding compared to that of single virus units in the case of filoviruses through a less unfavourable $\Delta S_{R}$ term. Furthermore for this reason, natural polyprotid filovirus filaments may bind more strongly to cells than suggested by results from experiments to determine filovirus infectivity using EBOV GP-expressing pseudoviruses, which are spherical and single.

4.3. Using enthalpy to relate changes in amino acid sequence at the binding faces of the virus GP and host receptor protein to a change in $K_a$

While lack of data on $\Delta H_{\text{ex}}$ limits the use of Approach 3 in Table 3 to determine $K_a$, Eq. (19) may be applied to determine the impact of amino acid changes in GP and Cr on $K_a$ and hence link genetic sequencing data to the risk of infection. The ultimate aim would be to predict how such sequence changes affect infectivity, host range and jumping the species barrier from the perspective of the dose-response. This could include assessing the relative risk of novel strains of virus with mutations in the GP and also assessing the risk of a “standard” virus jumping the species barrier into a novel host which may have amino acid differences in its Cr relative to the “standard” host. The starting point is having information on the infectivity of the “standard” virus in the “standard” host and also crystal structure data of the “standard” virus GP bound to the “standard” host Cr. Crystal structure data are available for a number of viral GPs docked to their human host cell Cr at atomic detail including EBOV and MERS-CoV (Zhao et al., 2016; Wang et al., 2016; Lu et al., 2013). Knowing the crystal structure of the GP/Cr protein complex assists in understanding how changes in amino acids at the binding interface of GP and Cr affect $\Delta H_{\text{ex}}$ and hence $K_a$ through Eq. (19). Thus although $\Delta H_{\text{ex}}$ itself may not be known, it may be possible to predict the change in $\Delta H_{\text{ex}}$ (i.e. $\Delta H_{\text{ex}}$) for an amino acid substitution from basic biophysics knowledge of the energies of the intermolecular forces that hold proteins together. These are typically salt bridges and hydrogen bonds as shown for MERS-CoV bound to its Cr (Lu et al., 2013). An example of how this could be applied is demonstrated in Table 5 for changing amino acids involved in salt bridges. In a salt bridge a negatively charged amino acid residue on one protein interacts electrostatically with a positively charged residue on the other protein resulting in a strong attraction. Salt bridges, e.g. between histidine 30 (positively charged) and aspartate (Asp) 70 (negatively charged) in bacteriophage T4 lysozyme contribute $\sim 11$ to $\sim 21$ kJ/mol to $\Delta G$ (Anderson et al., 1990; Dong and Zhou, 2002). For the purpose of the demonstration in Table 5 it is assumed that the $\Delta H_{\text{ex}}$ for a salt bridge formation is $\sim 17.8$ kJ/mol as this gives a 1000-fold change in $K_a$.

4.3.1. EBOV adaptation to humans through changes in GP

Amino acid changes in the EBOV GP affect its host specificity. Thus, EBOV Makona in the 2013 West Africa outbreak adapted to humans (and in doing so became less infectious to bats) through substituting alanine (A) with valine (V) at residue 82 of the GP (Diehl et al., 2016; Urbanowicz et al., 2016). This is referred to as the A82V substitution. However, the changes in infectivity to humans (as measured by EBOV GP pseudoviruses infecting human cell lines) appear to be relatively
small and in the range of 2-fold to 4-fold. A 4-fold increase in infectivity due to a 4-fold increase in $K_a$ in Fig. 3 would reflect a 4-fold increase in $K_a$ and a change in $\Delta G_a$ according to Eq. (18) of $-3.6 \text{ kJ/mol}$. This would probably reflect changes in $\Delta H_a$ (and not the $T \Delta S_a$ term) since residue 82 on the EBOV GP is between residues 80 and 83 which make multiple atom contacts directly with six surface residues on the NPC1 residue 82 on the EBOV GP is between residues 80 and 83 which make.

In the human NPC1 protein (Cr), an Asp at residue 502 protrudes from the surface and contacts the EBOV GP in the docked complex (Zha0 et al., 2016). The natural presence of phenylalanine (Phe) at residue 502 (instead of Asp) in the NPC1 of the African Straw-coloured bat (Eidolon helvum) appears to protect this species from EBOV (Ng et al., 2015). Indeed, replacing Phe at residue 502 of E. helvum NPC1 with Asp completely restored its binding to EBOV GP (Ng et al., 2015). From the X-ray structure, the presence of Phe at residue 502 of the NPC1 in E. helvum would cause severe clashes on binding EBOV GP (Zha0 et al., 2016) such that $\Delta H_a$ in the E. helvum NPC1/EBOV GP complex would be significantly less negative than that for binding to human NPC1 and perhaps even positive (reflecting a repulsion). Indeed, determining $\Delta H_a$ would enable the effect of the Asp to Phe substitution at residue 502 on $K_a$ and hence on the infectivity of EBOV to E. helvum to be assessed through Eq. (19) as in Table 5.

Hoffmann et al. (2016) developed VSV pseudoviruses with filovirus GPs expressed on their surfaces to test the efficiency of GP-mediated cell entry into cell lines of different bat species. Hoffmann et al. (2016) demonstrated that the efficiency of entry of EBOV-GP VSV pseudovirus into E. helvum cell lines was markedly reduced compared to that with GPs from other filoviruses and at standard virus levels no infection was detected. However, at high levels of (i.e. undiluted) virus, EBOV-GP VSV pseudovirus did show significantly increased infection compared to the negative controls. Thus EBOV GP is capable of mediating entry into E. helvum cells, albeit with low efficiency. This is entirely consistent with the dose-response model simulated in Fig. 2 which predicts for low $K_{0.5}$ ($10^{5}-10^{7} \text{ M}^{-1}$) that very small fractions (10^{-7} to 10^{-9}) of the host cells have bound virus at very high virus doses, i.e. $10^{11}$ to $10^{12}$. Thus some infection would be expected in E. helvum on the basis of Eqs. (2) and (3) and the probability of infection which, although low at low $K_a$ values, would never be zero because there is no threshold $K_a$ below which binding does not occur (Fig. 3). Supporting this, Hayman et al. (2010) reported EBOV antibodies in one of 256 E. helvum bats tested in Ghana and suggested this resulted from EBOV infection.

### 4.4. Parameterising the model: determining values for $K_{\text{mucin}}$

Values for $K_{\text{mucin}}$ may be derived from published data on the strength of interactions between lectins and glycans. Lectins are proteins which bind selective glycan groups (Taylor and Drickamer, 2006). Viruses such as influenza virus, NoV and rotavirus behave as lectins by binding selectively to glycans. Indeed, $K_{\text{mucin}}$ is related to the number of lectin/glycan interactions and their respective $K_a$ in the same way as $K_a$ according to Eq. (17). Dam and Brewer (2010) report $K_{\text{mucin}}$ of $2.0 \times 10^{-10} \text{ M}$ for the lectin soy bean agglutinin binding to porcine submaxillary mucin which has multiple GalNAc sugars to bind. Similarly Vataira macrocarpa lectin has a $K_{\text{mucin}}$ of $1.0 \times 10^{-10} \text{ M}$ with porcine submaxillary mucin. Acting singly (i.e. using $n = 1$ in Eq. (17)), these $K_{\text{mucin}}$ would translate into $K_{\text{mucin}}$ values of $\sim 10^{-10} \text{ M}^{-1}$ at which very high excesses of mucin would be needed to achieve removal of the virus according to Fig. 1. There are two mechanisms by which the magnitude of $K_{\text{mucin}}$ could be increased:-

1. Through multiple binding interactions; and
2. Through irreversible binding.

The effect of multiple contacts on the strength of binding has been shown for binding of the mannose binding protein (MBP) to carbohydrate structures on the surfaces of pathogens. Each carbohydrate-recognition domain (CRD) of the MBP interacts only with the terminal sugar residue in an oligosaccharide chain (Taylor and Drickamer, 2006). The $K_a$ for interaction with a high mannose oligosaccharide is $\sim 10^{-3} \text{ M}$, i.e. low affinity. High-affinity binding of MBP requires interaction of multiple CRDs with multiple terminal mannoside residues. The trimeric structure of MBP presents a cluster of CRDs for interaction with appropriately spaced terminal mannoside residues on the pathogen surface. Thus the arrays of terminal sugar residues on the surfaces of microorganisms can interact with multiple sites.
simultaneously. As shown in Eq. (17), a three way interaction involving three binding sites with $K_d$ of $1.0 \times 10^{-3}$ M can result in an overall $K_a$ of up to $1.0 \times 10^9$ M$^{-1}$ for a multivalent ligand (i.e. pathogen surface) with appropriately spaced terminal sugar residues (Taylor and Drickamer, 2006). In the case of influenza virus A, the binding site of each HA polypeptide is relatively shallow and interacts primarily with the terminal sialic acid residues linked to galactose (Taylor and Drickamer, 2006). The affinity for monomeric sialosides is weak ($K_d \sim 10^{-3}$ M), but binding to cell surfaces is enhanced by the simultaneous interaction of multiple HA-binding sites with multiple sialic acid residues on the target host cell (Taylor and Drickamer, 2006). The cumulative effect of $n$ binding sites each with the same $K_d$ is given by the $n^{th}$ power (Eq. (17)). Thus seven interactions with $K_d \sim 10^{-3}$ M, for example, would give a $K_a \sim 10^{21}$ M$^{-1}$. It is concluded that mucins and MBP PRRs of the innate immune system could bind virus with sufficient affinity such that slight mucin excesses could remove all the virus (Fig. 1). Furthermore, $K_{mucin}$ could become higher if the virus were irreversibly bound for example by being “folded in” to the inside of mucous droplets such that it can no longer exchange with free virus on the outside. In effect $k_{off}$ would tend to 0 in Eq. (6) and hence $K_{mucin}$ would tend to infinity. It should be noted that the magnitude of $K_d$ (and hence $K_a$ and $K_{mucin}$) is constant for a given molecular interaction at a given temperature. However, the magnitude of $K_{mucin}$ may change during the infection process through both host mechanisms and virus mechanisms as is now discussed.

4.4.1 Changes in $K_{mucin}$ through host effects

Mucin glycosylation changes during infection such that mucins isolated from rotavirus-infected mice at 4 days post infection were more potent at inhibiting rotavirus infection than mucins from control mice (Boshuizen et al., 2005). Interestingly there are also age-dependent differences in mucin quantities, composition, and/or structure which alter the antiviral capabilities of the small intestine mucins (Boshuizen et al., 2005). Furthermore, mucin production by the host increases during infection by rotavirus with mucin-coding mRNA levels peaking at 1 day post infection (Boshuizen et al., 2005). Thus both the mucin: virus ratio and the magnitude of $K_{mucin}$ could change during progression of the infection, hence affecting $F_v$ in Fig. 1.

4.4.2 Changes in $K_{mucin}$ through virus effects

There are viral mechanisms that remove the virus from the mucin so increasing the concentration of free virus. For example in the case of influenza virus, the neuraminidase removes sialic acids from glycans, which enables virus particles to be removed from the cell surface after assembly and from decoy receptors e.g. in mucus (Guo et al., 2017). This has two effects on virus binding in Fig. 1. First it increases the rate of removal of the virus ($k_{off}$ in Eq. (6)) and hence decreases $K_{mucin}$. Second it decreases the concentration of mucin thus decreasing the mucin: virus ratio. From Fig. 1 both these act synergistically to increase the fraction of free virus.

In the case of influenza viruses, the HA protein binds to sialoside receptors on the host cell surface, preferentially binding to sialic acids linked to a penultimate galactose (Guo et al., 2017). The balance between activities of HA and neuraminidase proteins has a critical role in optimal viral fitness, tropism and transmission (Guo et al., 2017). The opposing effects of HA and neuraminidase in influenza A virus infectivity could be modelled through their effect on $K_{mucin}$ (through $k_{on}$ and $k_{off}$ in Eq. (6)) and the mucin: virus ratio in Fig. 1. Handel et al. (2014) use differential equations to model how sticky an influenza virus should be to maximize fitness, with stickiness representing the balance between attachment and detachment. Shanker et al. (2011) proposed that local flexibility in part of the NoV surface protein could allow the virus to disassociate from salivary mucin-linked HBGA in the changing microenvironment (pH for example) during its passage through the GI tract to subsequently associate with HBGAs linked to intestinal epithelial cells. In effect $k_{off}$ rates increase in Eq. (6) such that $K_{mucin}$ varies in different parts of the host.

5. Discussion

This is a concept paper which presents a generic approach to use data from biochemistry and molecular biology to parameterize dose-response models for MRA through thermodynamic equations. In the model developed here a host intestine is simulated with $4.15 \times 10^8$ cells in a volume of $0.314$ dm$^3$. The initial infection process is broken down into two stages, namely escape of the virus from the innate host defense barriers (Fig. 1) and the subsequent binding of any remaining virus to its specific receptor on the host cell (Fig. 2). The two parameters, namely $F_v$ and $F_{cell}$ which quantify these stages represent probabilities which together with $p_{cell}$ can be used with Eqs. (1)–(3) to give a complete dose-response model which calculates the risk of infection of the host from a given challenge dose ($V_{init}$). (Although $V_{intestine}$ from Eq. (1) does not appear in Eqs. (2) or (3), it is used in Fig. 2 to estimate $F_v$ for Eq. (3)). The difference equation approach developed here is tedious to apply to MRA and applying Eq. (7) and Eq. (11) would greatly simplify the calculation of $F_v$ and $F_{cell}$ respectively for MRA purposes. However, using Eq. (7) for $F_v$ fails at low doses of a virus which has a high binding affinity for the host cells ($K_d > 10^{-15}$ M$^{-1}$) and the difference equation approach has to be used as in Fig. 2. Using Eq. (11) for $F_{cell}$ is appropriate for MRA if the mucin concentration exceeds the virus concentration in the host by a factor of > 10 or if $K_{mucin} < 10^{-15}$ M$^{-1}$ (Fig. 1).

The mechanistic approach developed here may be used to determine the ID$_{50}$ (see Table 4) from which it is easy to calculate $p_{cell}$, the risk to the host from ingestion of a dose of one pathogen. The parameter $p_{cell}$ may then be used directly in conventional dose-response models (Eq. (16)) thus linking the application of the mechanistic model developed here to MRA. The work here appears to demonstrate that stochasticity in the mechanistic model is not an issue. If stochasticity issues were to arise for low doses of very high affinity pathogens (i.e. those with very high $K_d$), the approach should be to use the difference equation approach to determine the ID$_{50}$ from which $p_{cell}$ can then be calculated and used directly in Eq. (16). Thus, depending on course of the value of $p_{cell}$ C.V. is much greater than 1 for the ID$_{50}$ (Table 4) and stochasticity issues are minimised.

Although more information on the concentrations of mucin proteins in the saliva and intestine is needed, lack of data on the thermodynamic parameters is the main limiting factor. Approach 3 (Table 3) which uses Eq. (19) to calculate $K_d$ is limited by the current lack of data for $\Delta H_d$ and $\Delta S_d$ for the virus/host cell interaction and Approach 2 using published $K_d$ data should therefore be pursued in the absence of experimental $K_d$ data from Approach 1. While data are available for $K_d$ for GP/CR binding for several viruses, Approach 2 using Eq. (17) to estimate $K_d$ is also limited by lack of information on $\Delta S_d$ for binding of the whole virus to cells. This is important because the magnitude of $\Delta S_d$ for immobilization of a whole virus is likely to be huge as demonstrated for VSV binding to artificial PI bilayers (Table 3). Lack of information on $\Delta S_d$ is therefore a major data gap, and furthermore may also have wider implications in affecting our ability to interpret measured $K_d$ values in terms of the actual infectivity of the virus. Thus it is the $K_d$ as determined by Eq. (17) taking into account $\Delta S_d$ for whole virus binding which defines infectivity (see results from Table 2) and not the $K_d$ alone. This is because $\Delta S_d$ for binding of the whole virus to cells will be much greater than the $\Delta S_d$ for binding of a soluble protein fragment as in $K_d$ determinations by SPR. Indeed the $\Delta S_d$ for binding of VSV is a huge $-16,662$ J/mol/K compared to the $\Delta S_d$ and $\Delta S_{conf}$ terms for binding of an antibody to its antigen (Table 3) which is comparable to GP binding to its Cr in terms of molecular sizes.

The model developed here based on Eq. (4) does not allow for multiple virus particles to bind to a single host cell. However, the cell membrane of the host cell typically contains many copies of Cr and in theory each host cell could therefore bind multiple copies of the virus.
Furthermore, the virus particles themselves could be ingested as a clump. Indeed in the case of EBOV, many genome copies may be linked together into long polyprotein filaments (Beniac et al., 2012) which could affect the thermodynamics of binding (Gale 2017). Eq. (4) could be modified to accommodate multiple viruses binding to a single host cell (as described for ligands binding to a protein by Price and Dwek (1979)) if the number of Crs per host cell is known. It should be stressed that homogeneity in both the mucin and pathogen is assumed here for simplicity. In further developing this mechanistic model, data would be needed not only on the heterogeneity of the mucin concentration but also on the statistical distribution for the number of viruses bound per cell. For example the distribution could be Poisson such that for a ratio of 1 virus bound per host cell (representing a high \( V_{\text{infectious}} \), some host cells have zero, most have one while a few have two, three or even four viruses bound. Alternatively the distribution could be over-dispersed, such that a few host cells have thousands of bound virus while most have none. Clearly, over-dispersion would decrease C.V compared to the Poisson distribution. In the model presented here for demonstrating proof of principle, the modelling of \( F_c \) as function of \( V_{\text{infectious}} \) assumes each host cell binds just one virus through Eq. (4). This is appropriate for two reasons. First, exposure to faecal/oral pathogens through routes such as water may involve very low numbers of pathogen per person (see Gale 2017), such that it would be unlikely that more than one pathogen would bind to the same host cell. Second, it maximises C.V and hence \( p_{\text{host}} \). Thus, the binding of more than one virus to a host cell may be waste of virus resource because C.V in Eq. 2 is not maximised. Binding of two viruses to the same cell will not affect \( p_{\text{cells}} \) unless there is co-operation between the two viruses in some way, such that one facilitates infection by the other. In this sense, the value of \( p_{\text{host}} \) predicted through the mechanistic dose-response here is worst case. It should be noted that Gale (2017) proposed that formation of polyloid filaments in the case of EBOV did indeed enhance cell binding by optimising \( \Delta S_{fi} \) effect representing a co-operation between viral particles in the infection process. This demonstrates the importance of not only considering the statistical distribution aspects of virus clumps but also the thermodynamic implications on their binding.

As shown in Fig. 1, the probability of the virus escaping the mucin barrier (as represented by \( F_c \)) is controlled by the mucin: virus ratio and the mucin/virus binding affinity as represented by \( K_{\text{mucin}} \). The range for which \( K_{\text{mucin}} \) is biologically significant is \(< 10^{22} \text{M}^{-1}; 10^{22} \text{being the } K_{\text{mucin}} \text{ Value at which there is no free virus at mucin: virus ratios } > 1:1 \) according to the simulation in Fig. 1. Values for \( K_{\text{mucin}} \) could be high due to multiple contacts with repeating units and “folding in” of bound virus into the interior of the mucus. Reading the x-axis from right to left in Fig. 1 suggests a threshold effect for virus dose, as the binding capacity of the mucin is exceeded at high virus doses. This is not considered further here, but is of interest for the development of MRA which generally assumes there is no threshold dose. The approach developed here in Fig. 1 for mucins could also be applied to modelling the probability of the virus being inactivated by other components of the host innate immune system such as the PRRs, which include the MBP, which recognize repeating units on the pathogen surface. Once bound to the MBP, the virus would be taken up (phagocytosis) through the complement system by macrophages and destroyed (Taylor and Drickamer, 2006).

The probability of a host cell having bound virus (as represented by \( F_c \)) increases linearly with dose (Fig. 2). The biologically significant range for \( K_a \) is \( 10^{15} \text{M}^{-1} \) within which \( F_c \) increases linearly with increasing \( K_a \) (Fig. 3). \( K_a \) values measured by SPR are in the \( 10^{-6} \text{M} - 10^{-12} \text{M} \) range for GP/Cr interactions for MERS-CoV and influenza virus suggesting very strong binding (Liu et al., 2013; Raman et al., 2014). Even for EBOV GP with \( K_a \) in the \( 10^{-4} \text{M} \) to \( 10^{-9} \text{M} \) range (Yuan et al., 2015; Wang et al., 2016), high \( K_a \) values could be achieved through multiple contacts (Eq. (17)) although this may be offset by the \( \Delta S_{fi} \) term (as discussed above). Increasing the \( K_a \) above \( 10^{15} \text{M}^{-1} \) does not further increase the fraction (\( F_c \)) of host cells with bound virus (Fig. 3) such that \( F_c \) versus virus dose curves are superimposed in Fig. 2 at \( K_a > 10^{15} \text{M}^{-1} \). At such high \( K_a \), receptor binding is not likely to be rate-limiting in the infection process. Thus Xu et al. (2012) found that the stability of the HeV-G/ephrin-B2 association does not strongly correlate with the efficiency of viral entry, suggesting that, for ephrin-B2-expressing cells, viral attachment is not the rate limiting step in the viral entry process. This is consistent with highly efficient binding (i.e. large \( K_a \) such that further increasing \( K_a \) does not increase the fraction of cells with bound virus, \( F_c \), according to Fig. 3. Hence C.V (through Eq. (3)) is constant and not rate-limiting, while a subsequent part of the infection process accommodated in \( p_{\text{cell}} \) is less efficient than \( F_c \), such that \( p_{\text{cell}} \) is very small and controls the overall value of \( p_{\text{host}} \) according to Eq. (2). Indeed, the magnitude of \( K_a \) is much more important when receptor binding is inefficient such as in the jumping of the species barrier into a novel host by an emerging virus as suggested by Gale (2017). This is confirmed by Fig. 3 with \( F_c \) increasing linearly with \( K_a \) at lower \( K_a \) values. Gale (2017) demonstrated that virus binding to its receptor is only important in controlling the probability of infection of a host cell if it is a major barrier, i.e. relatively inefficient. The \( K_a \) for a virus jumping the species barrier into a novel host species would be expected to be very low perhaps due to electrostatic repulsion as for Scenarios A and B in Table 5. The subsequent adaptation of the virus to its new host through mutation would increase \( K_a \) as suggested here in Table 5 and reported for EBOV in humans (Diehl et al., 2016; Urbanowicz et al., 2016).

The magnitude of \( \Delta H_{fi} \) is largely controlled by how good the molecular fit is between the virus GP surface and the Cr surface and depends on hydrogen bonds and salt bridges. As shown in Table 5, changes in \( \Delta H_{fi} (\Delta \Delta H_{fi}) \) may be used to predict changes in \( K_a \) providing the molecular basis of the GP/Cr interaction is available at atomic detail to enable some assessment of \( \Delta H_{fi} \). The infectivity of the new virus/host combination relative to the “standard” virus/host would be directly proportional to the change in \( K_a \) (as demonstrated from results in Table 2). Thus, thermodynamics provides the link between changes in the virus/host identified through sequencing data and the risk of infection. This in theory allows predictions of the infectivity of emerging virus strains or the susceptibility of novel hosts to be assessed on the basis of the effects of changes in sequence data. It is concluded that thermodynamic approaches have a major contribution to make in developing dose-response models for emerging viruses.

Conflict of interest

None declared.

Disclaimer

The views expressed in this paper are those of the author and not necessarily those of any organisations.

References

Anderson, D.E., Bectel, W.J., Dahlquist, F.W., 1990. pH-Induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of 14 lysozymes. Biochemistry 29, 2403-2408.
Beniac, D.R., Melito, P.L., deVarennes, S.L., Hiebert, S., Rabb, M.J., Lamboo, L.L., Jones, S.M., Booth, T.F., 2012. The organisation of Ebola virus reveals a capacity for extensive modular polyplody. PLoS ONE 7, e29608.
Ben-Tal, N., Honig, B., Ragazzari, C.K., Ben-Shaul, A., 2000. Association entropy in adsorption processes. Biophys. J. 79, 1180-1187.
Booth, T.F., Rabb, M.J., Beniac, D.R., 2013. How do filovirus filaments bend without breaking? Trends Microbiol. 21, 583-593.
Bostrom, J., Haber, L., Koening, P., Kelley, R.F., Fuh, G., 2011. High affinity antigen recognition of the dual specific variants of herceptin is entropy-driven in spite of structural plasticity. PLoS ONE 6, e17887.
Boshuizen, J.A., Reimerink, J.H.J., Korteland-van Male, A.M., van Ham, V.J.J., Borma, J., Geerke, G.J., Koopmans, M.P.G., Buller, H.A., Dekker, J., Einerhand, A.W.C., 2005. Homoeostasis and function of goblet cells during rotavirus infection in mice. Virology 337, 210-221.
Carneiro, F.A., Bianconi, M.L., Weissmuller, G., Stauffer, F., Da Poian, A., 2002.
