Mathematical model for fragmentation of bacterial inclusion bodies [1]

R. Cubarsi, M.M. Carrió[2] and A. Villaverde[2]
Dept. Matemàtica Aplicada i Telemàtica
Universitat Politècnica de Catalunya, Barcelona, Spain
[2]Institut de Biologia Fonamental
Universitat Autònoma de Barcelona, Bellaterra, Spain

[1]This work was presented at the *3ecm Third European Congress of Mathematics*, held 10-14 July, 2000, Barcelona, Spain, and it has been supported by Grant BIO98-0527 (CICYT).

Mathematical model for fragmentation of bacterial inclusion bodies

R. Cubarsi[1], M. M. Carrió[2] and A. Villaverde[2]
[1] Dept. Matemàtica Aplicada i Telemàtica, Universitat Politècnica de Catalunya, Barcelona, Spain; [2] Institut de Biologia Fonamental and Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Spain

Abstract

Bacterial inclusion bodies are microscopic, ovoid-shaped aggregates of insoluble protein. Under protease exposure a digestion process is produced that reveals a variable fragmentation rate, not compatible with a surface-restricted erosion of body particles, or an uniform sensibility to the fragmentation agent. The modelling and fitting of experimental data is performed in two steps. (a) Due to poor estimation of protein amounts only first derivatives can be numerically evaluated, and a non-linear first-order fragmentation model is adopted. Although it is a very good approximation for intermediate points, the asymtotic behaviour of the solution is inconsistent with the fragmentation process. (b) The solution of previous kinetic modelling is used to compute higher-order derivatives in intermediate points and to adopt a higher-order lineal model for the overall interval with protein fragmentation. The resulting model consists in a superposition of Poisson processes associated with several populations of protein with different fragmentation resistance. Numerical estimation of model constants is also described and discussed. In particular, an iterative method of weighted least squares is used in order to obtain minimum variance parameters.

CORRESPONDING AUTHOR:
Rafael Cubarsi
Dept. Matemàtica Aplicada i Telemàtica
Universitat Politècnica de Catalunya, Campus Nord
Jordi Girona, 1-3; E08034-Barcelona; Spain
Phone: 34-3-401-5995, Fax: 34-3-401-5981
E-mail: rcubarsi@mat.upc.es

1991 MATHEMATICS SUBJECT CLASSIFICATION: 92, 60, 62, 65.
KEY WORDS: mathematical modelling, stochastic processes, finite mixture distributions.

1. Introduction

Inclusion bodies (IBs) are refractile aggregates of misfolded, insoluble protein commonly observed upon targeted gene overexpression in bacterial cells [1]. IBs are not homogeneous in protein composition. Many cell folding-assistant proteins and proteolytic fragments of the main polypeptide component are found embedded as variable fractions of the total protein content [2, 3]. Despite this heterogeneity, transmission electron microscopy has often pictured IBs as amorphous aggregates lacking defined structural organisation [4].

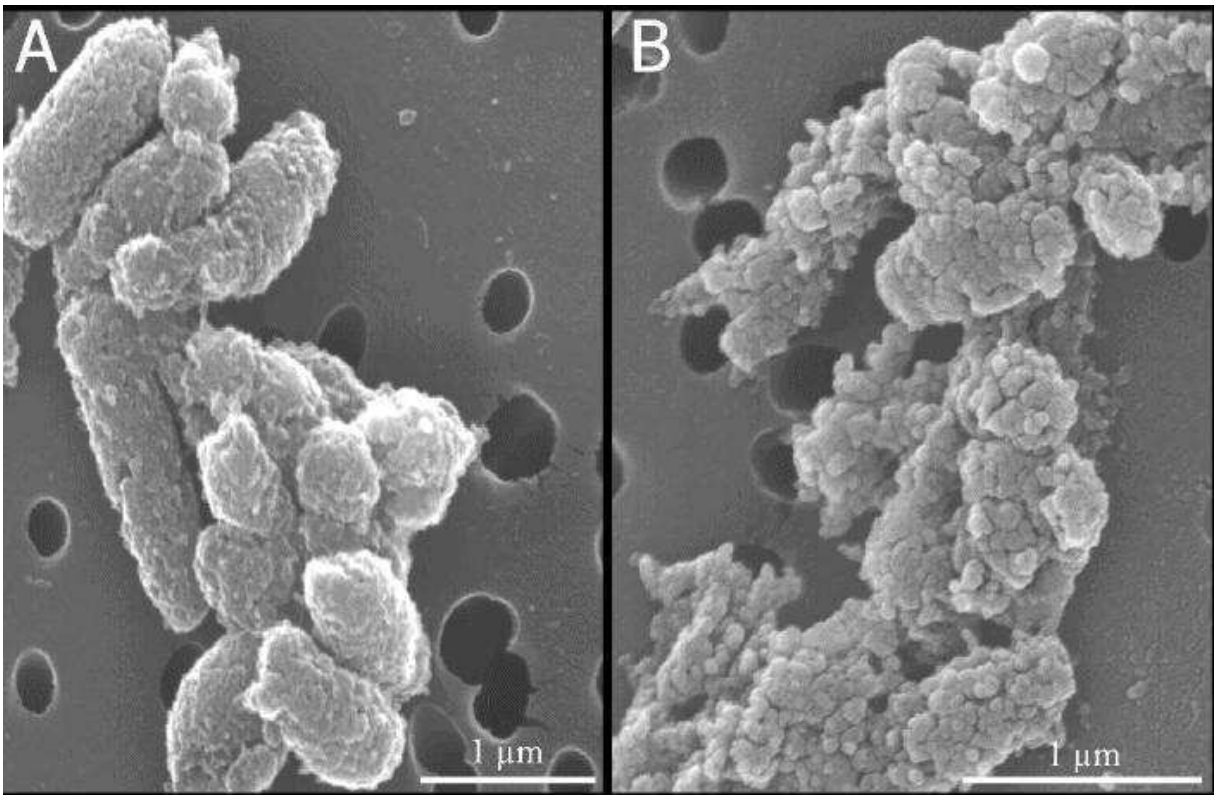

Figure 1: Scanning electron micrographs of purified VP1LAC protein before (A) and after 5 min (B) of being submitted to in vitro proteolytic digestion by incubation with trypsin. Similar images were obtained with other protein inclusion bodies. These pictures are representative of a large number of examinations.

By combining scanning electron microscopy and a kinetic modelling of IBs protein digestion during trypsin treatment unexpected architectural features of inclusion bodies (Fig. 1) and the coexistence of distinct populations of aggregated protein with different conformational states is observed [5 − 10]. Henceforth we summarize the mathematical modelling of bacterial IBs digestion process.

2. Fragmentation rate

In order to investigate in more detail the mechanics of protein digestion and the putative influence of body volume, several digestion experiments were performed [11] and proteolysis of IB intact recombinant proteins was modelled as follows. Being $N(t)$ the non-fragmented protein (in absolute densitometric units) at time t, fragmentation of IBs protein can be generally described from a non-linear differential equation of the following form,

$$\frac{1}{N}\frac{dN}{dt} = -qN(t)^{p-1};\ p > 0,\ q > 0 \tag{1}$$

Since experimental data does not provide enough good estimations for higher-order derivatives a first-order fragmentation law was adopted. The left hand side of foregoing equation is usually referred as fragmentation rate, $R(t)$, and its oposite, $D(t) = -R(T)$, is known as digestion rate. Depending on $p > 1$, $p = 1$, or $p < 1$, $D(t)$ is an increasing, constant, or decreasing positive function of $N(t)$, respectively.

The relevant parameter p is obtained for each individual experiment (Table 1). Before introducing the definitive modelling two possible interpretations are presented:

• In principle, a Poisson stochastic process would be expected to account for the digestion kinetics [12]. In this case the probability density function for the time interval from trypsin addition up to the first molecular proteolytic event would be given by Eq. 1 rendering $p = 1$. However, such behaviour would imply a constant digestion rate, with an expected life for the intact IB of $< t >= 1/q$, which is clearly inconsistent with the experimental data (Fig. 2-a).

• An alternative possibility of a concentric layer erosion (a surface-restricted proteolytic attack of IB protein) also predicts a digestion profile that does not fit with the experimentally obtained kinetics (Fig. 2-b). Also note that, in this case, p would be $2/3$ and that the approximated p values for all the performed experiments are indeed higher than one (Table 1).

Although the approach from Eq. 1, with $p > 1$, provides a satisfactory explanation for low and intermediate times of fragmentation process, the asymtotic behaviour of $D(t)$ is inconsistent with data coming from long time fragmentation experiments, where a nearly constant and positive di-

gestion rate is measured. Note that, if $p > 1$, from Eq. 1,

$$\lim_{t\to\infty} D(t) = 0 \tag{2}$$

as it is shown in Fig.3. Hence the modelling of IBs fragmentation must give account of:

a) A similar behaviour to Eq. 1 for low and intermediate times.
b) A non-null asymtotic digestion rate.
c) An interpretation in terms of Poisson process.

3. Mixture model

A more plausible hypothesis is given by assuming a heterogeneous nature of the IB protein in which more than one different protein species (1, 2,... n, with increasing proteolytic susceptibility) would coexist. Under protease exposure, each of these species follows an individual Poisson process for fragmentation resulting in distinguishable expected lives. According to this model, the composition of IBs is described as a mixture of exponential density functions for the specific time interval up to the fragmentation event,

$$N(t) = \sum_{i=1}^{n} N_i e^{-q_i t};\ 0 < q_1 < q_2 < ... < q_n \tag{3}$$

where $N_i > 0$ $(i = 1, ..., n)$. The characteristic parameter for each protein component is the partial digestion rate q_i, inverse of the expected life (although sometimes the half-life $T_i = \frac{ln2}{q_i}$ is used as component describing parameter), and the fraction $n_i = \frac{Ni}{N(0)}$ corresponds to the mixing proportion. Therefore in Eq. 3 the components are ordered in decreasing expected life.

Obviously the digestion rate $D(t)$ assotiated with Eq. 3 is a positive decreasing function of time. At the beginig of experiment, $t = 0$, the detected digestion rate $D(0)$ is the mean value of the partial ones,

$$q_0 = \frac{1}{N(0)} \sum_{i=1}^{n} N_i q_i \tag{4}$$

while at the end of the experiment next equality is fulfilled,

$$\lim_{t\to\infty} D(t) = q_1 \tag{5}$$

This asymtotic behaviour of the digestion rate is shown in the fitting curves of Figs. 2 and 3 for some of the experiments. Moreover, for any time t, the following relationship is satisfyied

$$q_0 > D(t) > q_1,\ 0 < t < \infty \tag{6}$$

While a two-component mixture explains qualitatively the degradation kinetics, for most experiments the minimal number of protein species required to actually account for the observed data is three (Table 1).

Protein	p	T_1	n_1	T_2	n_2	T_3	n_3
VP1LAC(5hI)	1.91 ± 0.11	1	1.5	3.5	65.5	26.5	33
VP1LAC(5hII)	1.62 ± 0.06	3.5	8	11	68	77	24
LACVP1(3hI)	2.46 ± 0.37	1	27	4	27	50.5	46
LACVP1(3hII)	1.72 ± 0.22	1	24	4.5	23	22.5	53
LACVP1(5hI)	3.74 ± 1.03	-	-	2.7	58	64	42
LACVP1(5hII)	1.88 ± 0.32	1.6	34	6	15	43	51
LACVP1(24hI)	1.81 ± 0.63	4	19	10.5	42	75.5	39
LACVP1(24hII)	1.55 ± 0.31	6.5	18	19.5	56	103.5	26

Table 1: Protein composition and stability of several fragmentation experiments, namely estimated parameter p for first-order approximation (Eq. 1), half-lives T_i (in min), and mixing proportions n_i for the mixture model.

4. Determination of component parameters

It is well known that the function described in Eq. 3 satisfies a n-order linear homogeneous differential equation with constant coefficients ($a_i;\ i = 1, ..., n$), namely

$$F(N(t), N'(t), ..., N^{(n)}(t); a_1, a_2, ..., a_n) = 0 \tag{7}$$

where the coeficients are non-linear functions of the component parameters,

$$a_i = a_i(q_1, q_2, ..., q_n);\ i = 1, ..., n \tag{8}$$

From numerical evaluations of $N(t), N'(t), ..., N^{(n)}(t)$ at intermediate degradation times, a linear overdeterminate system of equations can be built in order to estimate the coeficients a_i. However the numerical estimation of derivatives directly from experimental data leads us to non-consistent results. For this reason, the time derivatives in Eq. 7 are explicitly evaluated from Eq. 1,

$$\frac{d^k N(t)}{dt^k} = (-q)^k \prod_{m=1}^{k-1} (mp - m + 1) N(t)^{(kp-k+1)},\ k > 1 \tag{9}$$

In fact, the fragmentation rate in the mixture model can be approximated by the following function with a similar behaviour,

$$r(t) = -A(N(t)^{p-1} + B);\ p > 1;\ A, B > 0 \tag{10}$$

so that $AB = q_1$, which is an estrictly concave and decreasing continuous function, which does not present the asymtotic inconsistency of Eq.2. Nevertheless, for times far from the asymtotic behaviour, namely $t < t_l$ with $B << N(t_l)^{p-1}$, Eq. 1 becomes a very good approximation of Eq. 10 rending $B = 0$ and, hence, of the mixture fragmentation rate.
Then Eq. 7 is converted into the following form

$$G(N(t), p, q; a_1, a_2, ..., a_n) = 0 \tag{11}$$

which is definitively used in order to estimate the coeficients a_i.

On the other hand, the component parameters q_i are obtained as real roots of the characteristic equation associated with Eq. 7, namely $P(x; a_1, a_2, ..., a_n) \in R[x]$ so that

$$P(q_i; a_1, a_2, ..., a_n) = 0;\ i = 1, ..., n \tag{12}$$

Finally the constants N_i, involved in Eq. 3, are obtained in order to determine the mixing proportions n_i. From experimental protein amounts $N(t)$

at different times, and with the partial digestion rates q_i already known, above values are estimated by solving a linear overdeterminate system of equations associated with Eq. 3.

• Comments about the numerical procedure:

- The fragmentation rate $\frac{1}{N}\frac{dN}{dt}$ in Eq. 1, for the set $(t, N(t)); t = t_1, ..., t_m$; is computed from a three-point estimation of the derivative $g'(t) = (ln(N(t))'$ at intermediate points $t_i = t_2, ..., t_{m-1}$ as follows [13] ,

$$g'(t_i) = \frac{(t_{i-1} - t_i)^2 g(t_{i+1}) - ((t_{i-1} - t_i)^2 - (t_{i+1} - t_i)^2) g(t_i) - (t_{i+1} - t_i)^2 g(t_{i-1})}{(t_{i+1} - t_i)(t_{i-1} - t_i)(t_{i-1} - t_{i+1})}$$

-The following linear and more convenient form of Eq. 1 is used in order to estimate the parameters p and q:

$$ln|(ln(N(t))'| = ln(q) + (p-1)ln(N(t))$$

-All overdeterminate systems of equations have been solved from an iterative wheighted least squares method, according to the algorithm described in [14] (this Conference). In order to obtain minimum variance parameters [15], weights are evaluated from the inverse covariance matrix of experimental errors.

References:

[1] F.A.O. Marston. Biochem. J. 240 (1986) 1-12.
[2] U. Rinas, J.E. Baily. Appl. Microbiol. Biotechnol. 37 (1992) 609-614.
[3] U. Rinas. Biotechnol. Prog. 12 (1996) 196-200.
[4] G.A. Bowden, A.M. Paredes, G. Georgiou. Biotechnology 9 (1991) 725-730.
[5] M.M. Carrió, J.L. Corchero, A. Villaverde. Biophys. Biochem. Acta 1434 (1999) 170-176.
[6] J.L. Corchero, E. Viaplana, A. Benito, A. Villaverde. J. Biotechnol. 48 (1996) 191-200.
[7] J.L. Corchero, A. Villaverde. Biotechnol. Bioeng. 58 (1998) 625-632.
[8] E. Corchero, R. Cubarsi, S.-O. Enfors, A. Villaverde. Biophys. Biochem. Res. Commun. 237 (1997) 325-330.
[9] J.X. Feliu, A. Villaverde. FEBS Lett. 434 (1998) 23-27.
[10] M.M. Carrió, J.L. Corchero, A. Villaverde. FEMS Microbiol. Lett. 169 (1998) 9-15.
[11] M.M. Carrió, R. Cubarsi, A. Villaverde. FEBS Lett. 471 (2000) 7-11.
[12] A. Papoulis. Probability, Random Variables, and Stochastic Processes, MacGraww-Hill, Singapore (1989).
[13] D. Kincaid, W. Cheney. Numerical Analysis. Brooks-Cole Publishing Co. Belmont, California (1991).
[14] R. Cubarsi, J.L. Corchero, P. Vila, A. Villaverde. Modelling and parameter estimation of gene expression and cell growth in batch cultures. This Conference No. 488, section 7.
[15] A. Stuart, J.K. Ord. Kendall's Advanced Theory of Statistics, Vol. 1, Distribution Theory. Ch. Griffin & Co., London (1987).

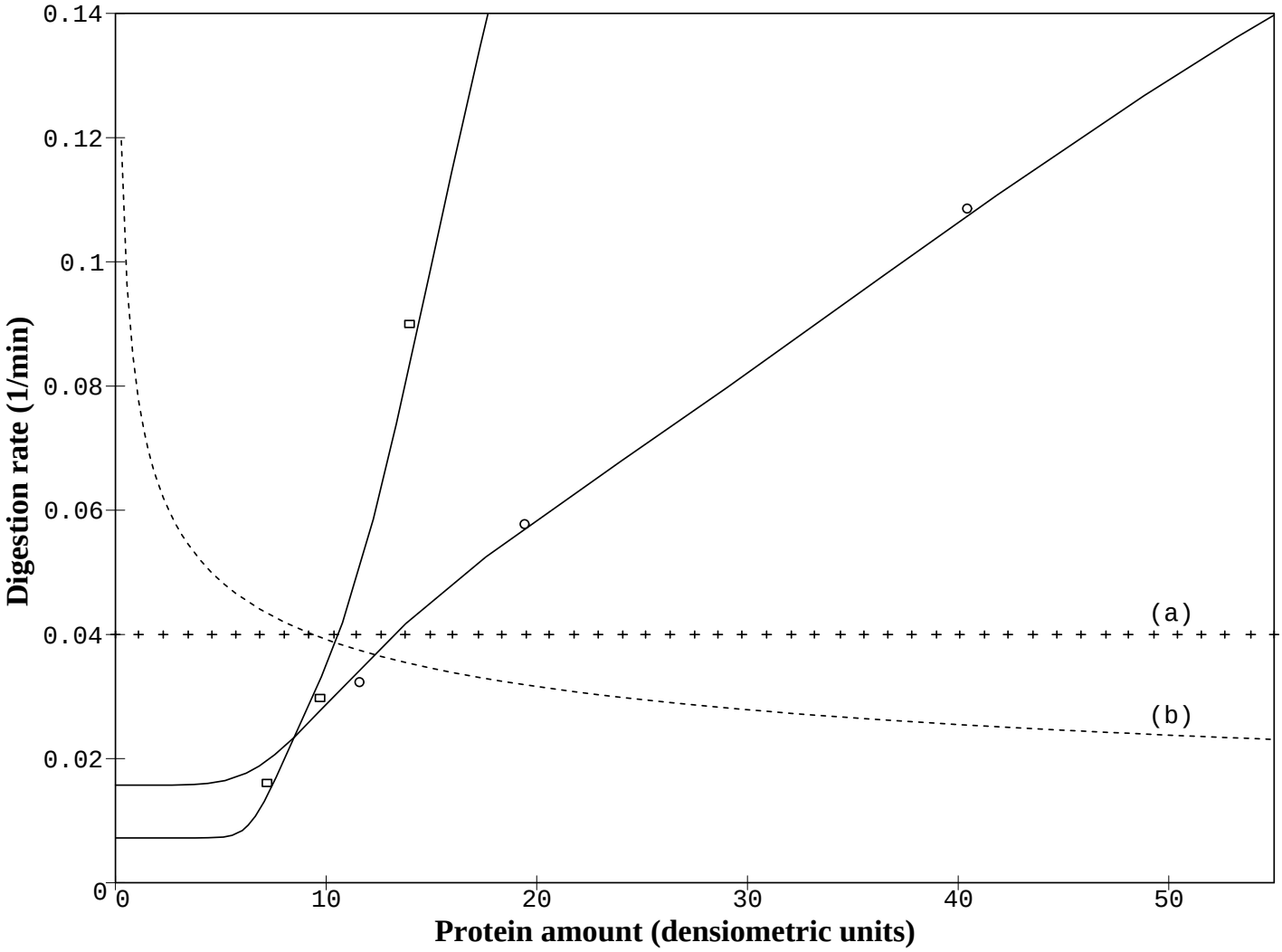

Figure 2: Digestion rates of IB protein versus protein amount for VP1LAC(5hI) (circles) and LACVP1(5hI) (squares). In addition expected rates are plotted according to different models of IB digestion, namely constant rate (a, dashed line), surface-restricted erosion (b, continuous line), and mixture of protein species with constant but distinguishable digestion rates, as modelled from the experimental data for both VP1LAC(5hI) (continuous plot) and LACVP1(5hI) (dashed plot).

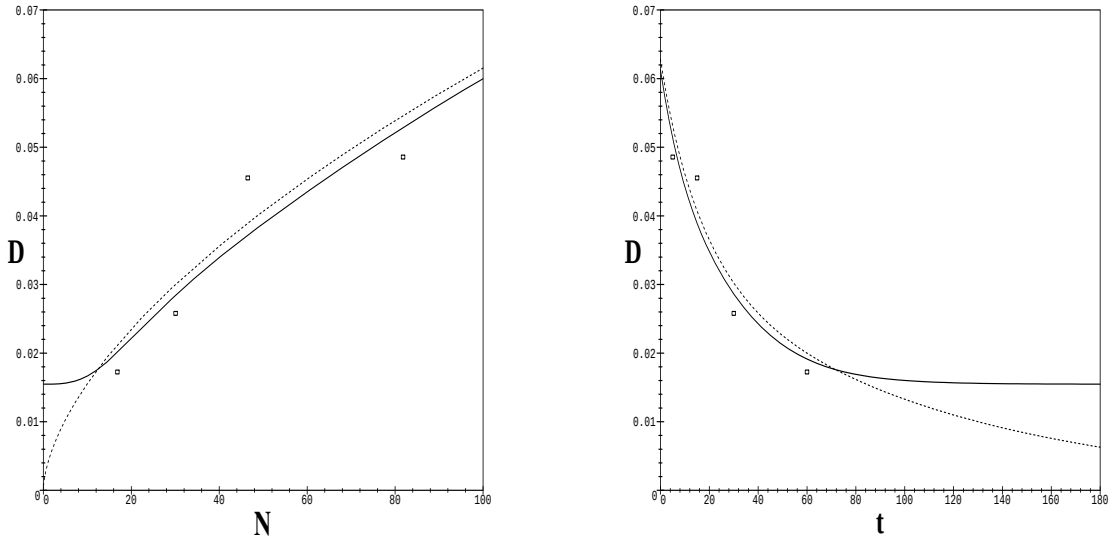

Figure 3: Digestion rates D of IB protein versus protein amount N (left), and versus time t (right), obtained from intermediate point estimation for IBs of VP1LAC(5hII). The approaches from the first-order model, Eq.1, (dashed line), and from a three component mixture model (continuous line) are compared. Both models provide a good fitting for intermediate points, although the asymptotic behaviour of the digestion rate in the mixture model must be remarked. In this case the digestion rate at protein value $N = 0$ and $\lim_{t\to\infty} D(t)$ indicate the digestion rate for most protease-resistant protein form found.