

## Supporting information

### Hydrogen bonding aggregation in acrylamide: theory and experiment

Elena Patyukova<sup>1\*</sup>, Taylor Rottreau<sup>1</sup>, Robert Evans<sup>1</sup>, Paul D. Topham<sup>1\*</sup>, and  
Martin J. Greenall<sup>2\*</sup>

<sup>1</sup>*Aston Institute of Materials Research, Aston University, Birmingham, B4 7ET, UK*

<sup>2</sup>*School of Mathematics and Physics, University of Lincoln, Brayford Pool, Lincoln, LN6 7TS, UK*

\* Emails: patyukova@gmail.com, p.d.topham@aston.ac.uk,  
mgreenall@lincoln.ac.uk

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 704459.



# Contents

<b>1</b>	<b>Models</b>	<b>S3</b>
1.1	Model 0 . . . . .	S3
1.1.1	Model 0m . . . . .	S4
1.1.2	Model 0g . . . . .	S4
1.2	Model 1 . . . . .	S5
1.2.1	Model 1m . . . . .	S8
1.2.2	Model 1g . . . . .	S8
1.2.3	Model 1s . . . . .	S9
1.3	Model 2 . . . . .	S11
1.3.1	The proof . . . . .	S12
1.3.2	Model 2m . . . . .	S13
1.3.3	Model 2g . . . . .	S13
1.3.4	Model 2s . . . . .	S14
1.4	Model 3 . . . . .	S15
1.4.1	Model 3m . . . . .	S17
1.4.2	Model 3g . . . . .	S17
1.4.3	Model 3s . . . . .	S17
1.5	Model 4 . . . . .	S18
1.5.1	Model 4m . . . . .	S20
1.5.2	Model 4g . . . . .	S20
1.5.3	Model 4s . . . . .	S21
1.6	Model 5 . . . . .	S21
1.6.1	Model 5m . . . . .	S22
1.6.2	Model 5g . . . . .	S24
1.6.3	Model 5s . . . . .	S24
1.7	Model 6 . . . . .	S24
1.7.1	Model 6m . . . . .	S28
1.7.2	Model 6g . . . . .	S30
1.7.3	Model 6s . . . . .	S30
1.8	Model 7 . . . . .	S32
1.8.1	Model 7m . . . . .	S35
1.8.2	Model 7g . . . . .	S36
1.8.3	Model 7s . . . . .	S36
1.9	Model 8 . . . . .	S36
1.9.1	Model 8m . . . . .	S38
1.9.2	Model 8g . . . . .	S40
1.9.3	Model 8s . . . . .	S40
1.10	Model 9 . . . . .	S40
1.10.1	Model 9g . . . . .	S42
1.11	Model 10 . . . . .	S42
1.11.1	Model 10g . . . . .	S42
1.12	Model 11 . . . . .	S43
1.12.1	Model 11g . . . . .	S43
<b>2</b>	<b>Selection of the best model</b>	<b>S43</b>
<b>3</b>	<b><math>\delta\chi_{HB}(f)</math> dependences for a block copolymer with polyacrylamide in random mixing approximation</b>	<b>S44</b>

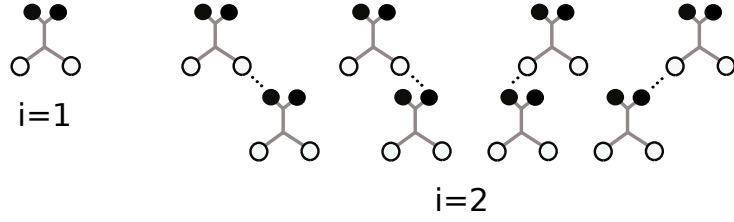


Figure S1: Schematic representation of the amide group and all four ways of forming a hydrogen bonded linear dimer from identical molecules. Hydrogen bonds are represented by dotted lines. In Model 0, we assume that only unimers and linear dimers are present in the solution.

## 1 Models

### 1.1 Model 0

In model 0, we assume that the solution contains only monomers and linear dimers (see Figure S1), and that chemical equilibrium with respect to hydrogen bonding association is described by a single association constant  $K$ . Then, the total concentration of the solution and the concentration of unimers are related to each other by the equation

$$c = c_1 + 2c_2, \quad (1)$$

where the concentration of dimers is  $c_2 = 4Kc_1^2$ . The factor of 4 appears here because there are four ways to form a dimer from two identical molecules (due to the presence of two hydrogens in the  $\text{NH}_2$  group and two lone electron pairs on the oxygen).

The free energy of hydrogen bonding of a system of volume  $V$  with  $N$  molecules and  $M$  dimers can be written as

$$F_{\text{HB}} = M\epsilon - kT \ln(p^M \Xi), \quad (2)$$

where  $\epsilon$  is the energy of a hydrogen bond,  $p = C/V$  (where  $C$  is a constant) is the probability that two molecules will meet and orient with respect to each other to form a bond, and  $\Xi$  is the combinatorial number of ways to form  $M$  dimers out of  $N$  molecules, such that

$$\Xi = \frac{N!2^M}{(N-M)!} \frac{(N-M)!2^M}{(N-2M)!} \frac{1}{M!} = \frac{N!4^M}{(N-2M)!M!}, \quad (3)$$

where the first factor is the number of ways of choosing  $M$  acceptor molecules, the second factor is the number of ways of choosing  $M$  donor molecules, the factor  $2^M$  takes into account that each molecule has two hydrogens (in the case when we assume that two bonds per oxygen are possible then an additional factor of  $2^M$  appears), and the last factor takes into account the indistinguishability of bonds.

Minimizing the free energy with respect to  $M$  yields

$$\frac{M}{4(N-2M)^2} = \frac{K}{V}, \quad (4)$$

where  $K = C \exp(-\epsilon/kT)$ , or, in terms of concentrations,

$$\frac{m}{4(c-2m)^2} = K. \quad (5)$$

As the concentration of unimers in the system is  $c_1 = c - 2m = c - 8Kc_1^2$ , we get  $m = c_2 = 4Kc_1^2$  and  $c = c_1 + 8Kc_1^2$  in agreement with Equation 1.

Let us now try to determine the association constant by fitting the dependence of the concentration on the height of the  $3530\text{cm}^{-1}$  peak.

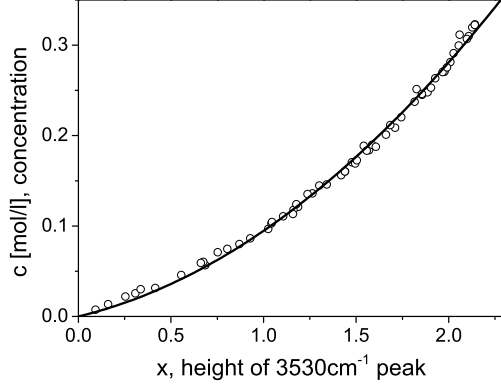


Figure S2: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 0m.

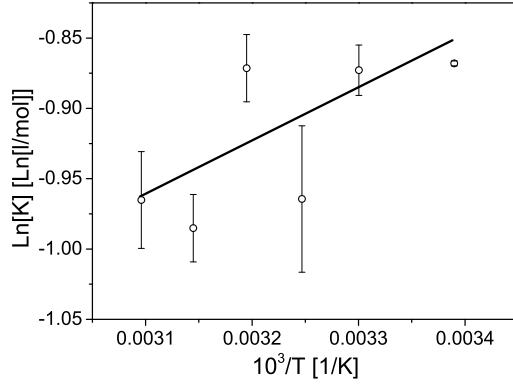


Figure S3: Dependence of  $\ln K$  on  $1/T$  for model 0m.

### 1.1.1 Model 0m

First, we assume that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the  $\text{NH}_2$  group in unimers. In this case,  $c_1 = Ax$ , where  $x$  is the height of the peak and  $A$  is some constant. Substituting this in Equation 1 gives the fitting equation

$$c = Ax + 8KA^2x^2. \quad (6)$$

A fit of the dependence of peak intensity on concentration at  $T = 22^\circ\text{C}$  is shown in Figure S2. The quality of the nonlinear fit can be quantified by the Akaike Information Criterion, on which further details are given at the end of this document. The value of this quantity for the current fit is  $\text{AICc} = -457.6$ .

According to the definition of the association constant in our model,  $K = C \exp(-\epsilon/kT)$ ,  $\ln K$  should depend linearly on inverse temperature  $1/T$ , and this dependence, together with a linear weighted fit, is shown in Figure S3. This yields estimates for the model parameters of  $\ln C = -0.84 \pm 3 \text{ ln[l/mol]}$  and  $\epsilon = -1.03 \pm 1.8 \text{ kcal/mol}$ . The poor quality of the fit in this case is reflected in the low value of the coefficient of determination,  $r^2 = 0.07$ .

### 1.1.2 Model 0g

It is now assumed that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of free  $\text{NH}_2$  groups. In this case,  $Ax = c_1 + 4Kc_1^2$ , and the fitting equation is

$$c = \frac{1 + 16AKx - \sqrt{1 + 16AKx}}{8K}. \quad (7)$$

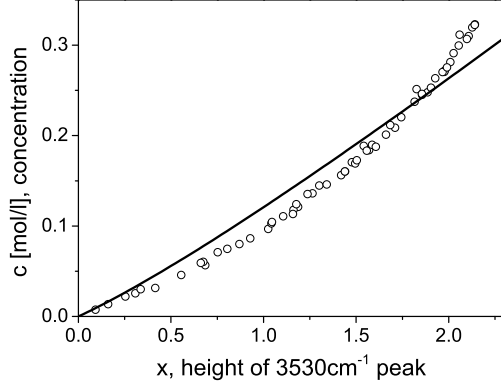


Figure S4: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 0g.

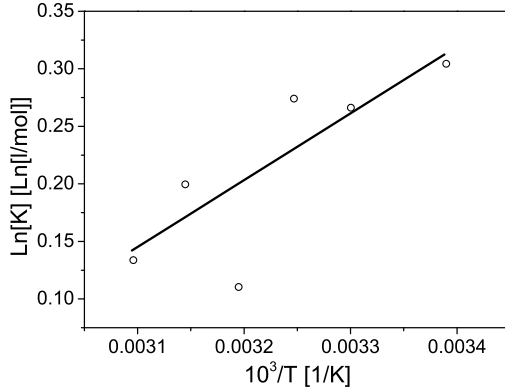


Figure S5: Dependence of  $\ln K$  on  $1/T$  for model 0g. The error bars are too large to be shown.

A fit of peak intensity versus concentration at  $T = 22^\circ\text{C}$  is shown in Figure S4. This fit is visibly less successful than that for model 0m, and AICc takes the higher value of  $-298$ .

The dependence of  $\ln K$  on  $1/T$  is shown in Figure S5. The estimates of the model parameters are  $\ln C = -1.65 \pm 0.5 \ln[\text{l/mol}]$  and  $\epsilon = -1.15 \pm 0.3 \text{ kcal/mol}$ . The quality of the fit is better than in model 0m, and this is shown by the higher value of the coefficient of determination,  $r^2 = 0.75$ . However, we note that this apparent improvement may be offset by the very large error bars on  $\ln K$ , which probably result from the poor quality nonlinear fit in Figure S4.

The free energy density due to hydrogen bonding in terms of the volume fraction of hydrogen bonding molecules  $\phi = cv$  and the dimensionless association constant  $K' = K/v$  has the form

$$f_{\text{HB}} = \frac{(\sqrt{1 + 32K'\phi} - 1)^2}{64K'} + \phi \ln \frac{\sqrt{1 + 32K'\phi} - 1}{16K'\phi}. \quad (8)$$

## 1.2 Model 1

In model 1, we assume that we have one bond per oxygen, two bonds per  $\text{NH}_2$  group, one association constant, and no cycles. This means that the aggregates are tree-shaped (see Figure S6).

The free energy of hydrogen bonding can be written as

$$F = M\epsilon - kT \ln(p^M \Xi), \quad (9)$$

where  $M$  is the number of hydrogen bonds,  $p$  is the probability that a donor and an acceptor form

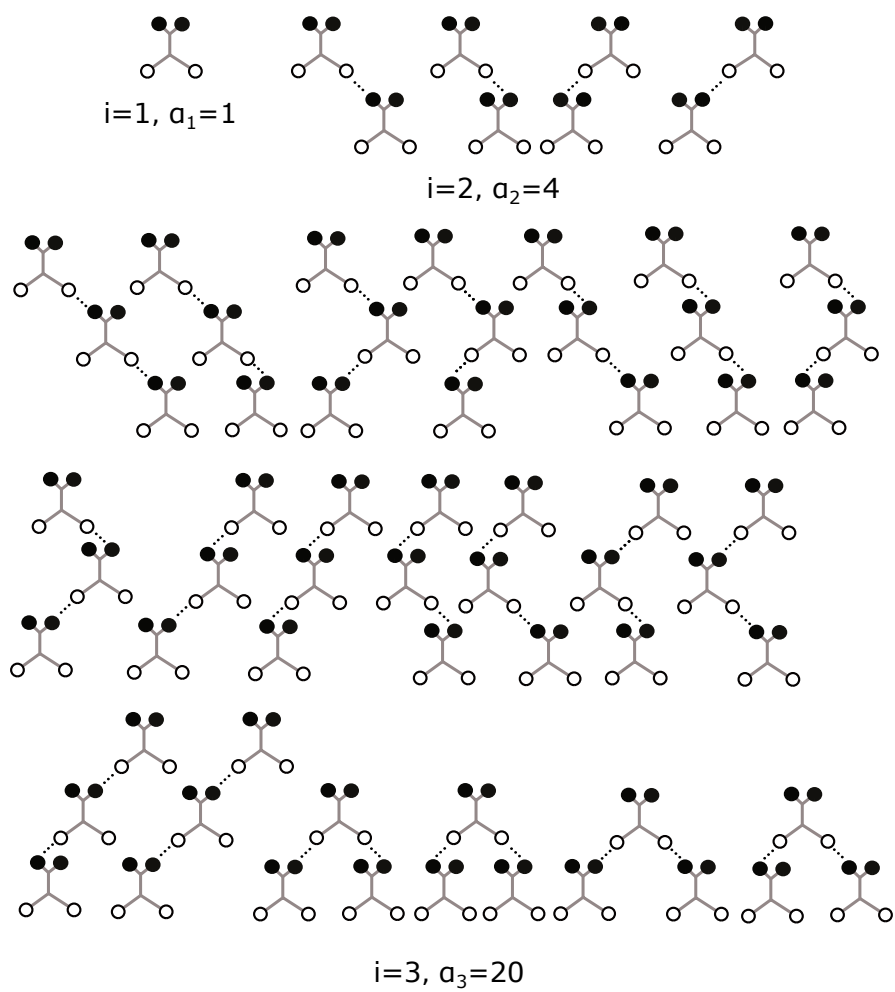


Figure S6: Schematic representation of possible aggregates with size up to  $i = 3$  in model 1.

a bond, and  $\Xi$  is the number of ways to form  $M$  bonds, given in this case by

$$\Xi = \frac{N!2^M}{(N-M)!} \frac{2N!}{(2N-M)!} \frac{1}{M!}, \quad (10)$$

where the first factor is the number of ways to choose an acceptor, the second is the number of ways to choose a donor, and the final factor takes into account the fact that all bonds are identical. Substituting Equation (10) into Equation (9) and using Stirling's formula gives

$$\frac{F}{kT} = M \left( \frac{\epsilon}{kT} - \ln p \right) + N \ln \frac{(N-M)(2N-M)^2}{4N^3} + M \ln \frac{Me}{(N-M)(2N-M)}. \quad (11)$$

After minimization with respect to  $M$ , we find, in terms of concentrations  $m = M/V$  and  $c = N/V$ ,

$$\frac{m}{2(c-m)(2c-m)} = K. \quad (12)$$

Let us suppose that the concentration of aggregates of size  $i$  can be expressed as

$$c_i = \alpha_i K^{i-1} c_1^i, \quad (13)$$

where the  $\alpha_i$  are unknown coefficients. Then, for the concentration of bonds  $m$  and total concentration  $c$  we have

$$m = \sum_{i=1}^{\infty} (i-1) c_i \quad (14)$$

and

$$c = \sum_{i=1}^{\infty} i c_i. \quad (15)$$

In order to find  $\alpha_i$ , we substitute expressions 14 and 15 into Equation 12 and equate coefficients in front of like powers of  $c_1$ . Using this method, we can calculate the values of the coefficients, which in this case are (starting from  $i = 1$ ) 1, 4, 20, 112, 672, 4224, 27456, ... Using the On-Line Encyclopedia of Integer Sequences<sup>1</sup>, we can assume that the general formula for a term of this sequence most probably has the form (in the main paper a more elegant way to get this result is described)

$$\alpha_i = 2^{i-1} \frac{(2i)!}{(i+1)!i!}. \quad (16)$$

The sequence  $\beta_i = (2i)!/(i+1)!i!$  is known as the Catalan numbers. It is known that these numbers represent the number of different rooted binary trees with  $i+1$  leaves. In our case, we have an additional factor of  $2^{i-1}$ , since each molecule apart from the root can be added in two ways to form a bond with one of the free hydrogens because there are two bonding sites on the oxygen. All aggregates allowed in this model with size up to  $i = 3$  are shown in Figure S6. So we can say that the physical meaning of  $\alpha_i$  is the number of ways to form an aggregate of size  $i$  out of  $i$  molecules.

It is also interesting to note that, by looking at Figure S6, it can be seen that the aggregates can be built recursively from each other, so a generating function  $G(z)$  can be written as

$$G(z) = 1 + 4zG(z) + 4z^2G(z)^2, \quad (17)$$

where  $z = Kc_1$  is the multiplicative factor that is introduced when the size of the aggregate is increased by one. Solution of this equation gives

$$G(z) = \frac{1 - 4z - \sqrt{1 - 8z}}{8z^2}, \quad (18)$$

and this generating function can be expanded with respect to  $z$  to give the values of  $\alpha_i$ .

With this expression for  $\alpha_i$  in hand, the total concentration of the solution can be calculated as

$$c = \frac{1 - 4Kc_1 - \sqrt{1 - 8Kc_1}}{8K^2c_1\sqrt{1 - 8Kc_1}}. \quad (19)$$

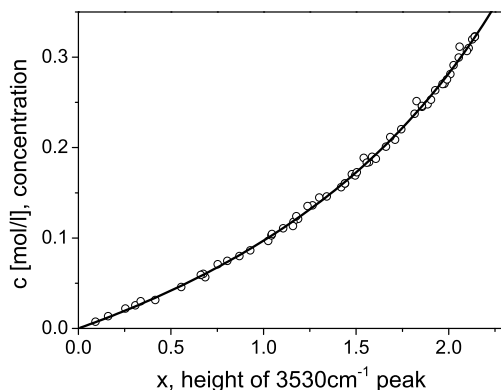


Figure S7: Fit of the dependence of the total concentration on the height of  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 1m.

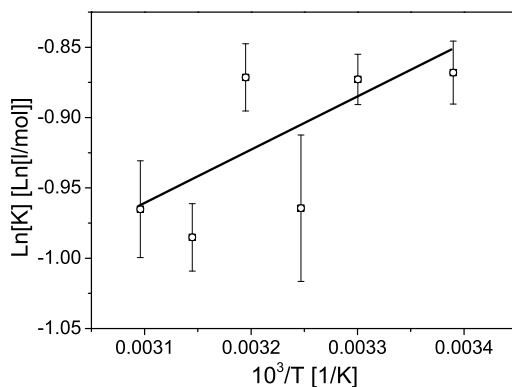


Figure S8: Dependence of  $\ln K$  on  $1/T$  for model 1m.

### 1.2.1 Model 1m

Let us assume first that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the  $\text{NH}_2$  group in unimers. In this case,  $c_1 = Ax$ , where  $x$  is the height of the peak and  $A$  is some constant. Substituting this in Equation 19 gives the fitting equation:

$$c = \frac{1 - 4KAx - \sqrt{1 - 8KAx}}{8K^2Ax\sqrt{1 - 8KAx}}. \quad (20)$$

The results of this fit are shown in Figures S7 and S8. The value of AICc for the fit in Figure S7 is  $-503.8$ . The dependence of  $\ln K$  on  $1/T$  is shown in Figure S8, and this fit gives estimates of the model parameters of  $\ln C = -2.14 \pm 0.6 \ln[\text{l/mol}]$  and  $\epsilon = -0.79 \pm 0.2 \text{ kcal/mol}$ , with a coefficient of determination of  $r^2 = 0.56$ .

### 1.2.2 Model 1g

In model 1g, it is assumed that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of  $\text{NH}_2$  free groups both in free molecules and in aggregates. Therefore, in order to find a fitting equation, we need to calculate the concentration of free groups. However, it turns out that it is impossible to do this in the framework of model 1 because the number of free groups depends on the structure of the aggregate. However, model 5 (to be introduced later) does include the relevant information about the structure of the aggregate, and reduces to model 1 when its two association constants are set equal to each other. Therefore, the calculation of the fitting expression is carried



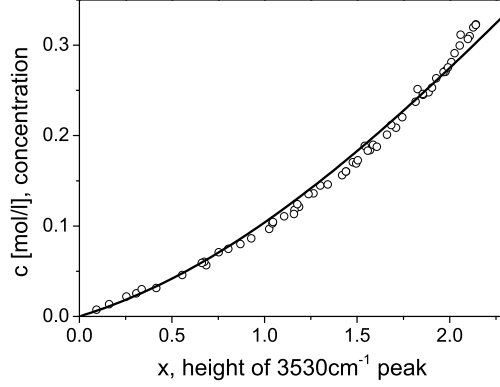


Figure S9: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 1g.

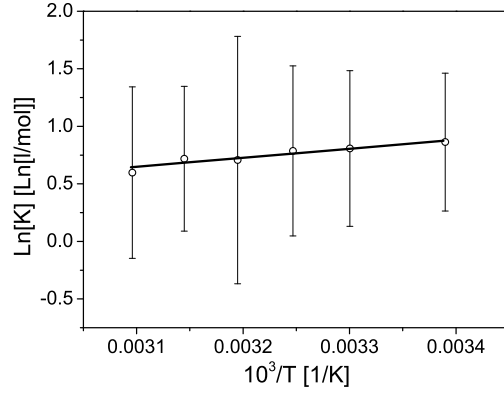


Figure S10: Dependence of  $\ln K$  on  $1/T$  for model 1g.

out in model 5, and  $K_1$  and  $K_2$  are both set equal to  $K$  at the end. This gives

$$c = \frac{1 - 4AKx + 16A^2K^2x^2 - (1 - 4AKx)\sqrt{1 + 16A^2K^2x^2}}{8AK^2x}. \quad (21)$$

The results of the fit are shown in Figures S9 and S10. The quality of fit in Figure S9 can be characterized by the parameter  $\text{AICc} = -388.9$ , and the fit shown in Figure S10 estimates the model parameters to be  $\ln C = -1.77 \pm 0.4 \ln[\text{l/mol}]$  and  $\epsilon = -1.55 \pm 0.3 \text{ kcal/mol}$ , with  $r^2 = 0.9$ .

### 1.2.3 Model 1s

Here we assume that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of  $\text{NH}_2$  free groups in free molecules and dimers. In this case, the relation between  $c_1$  and  $x$  takes the form  $Ax = c_1 + 4Kc_1^2$ .

The fitting results are shown in Figures S11 and S12. The value of  $\text{AICc}$  for the fit in Figure S11 is  $-503.9$ , and the estimates for the model parameters corresponding to Figure S12 are  $\ln C = -1.75 \pm 0.5 \ln[\text{l/mol}]$  and  $\epsilon = -0.78 \pm 0.3 \text{ kcal/mol}$ , with  $r^2 = 0.66$ .

The free energy density of hydrogen bonding in model 1 in terms of the volume fraction of hydrogen bonding molecules  $\phi = cv$  and dimensionless association constant  $K' = K/v$  has the form

$$f_{\text{HB}} = m + \phi \ln \frac{(\phi - m)(2\phi - m)^2}{4\phi^3}, \quad (22)$$

where  $m$  is a solution of the equation  $m/[2(\phi - m)(2\phi - m)] = K'$ .

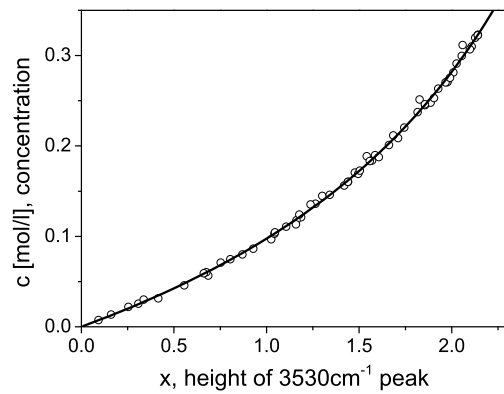


Figure S11: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 1s.

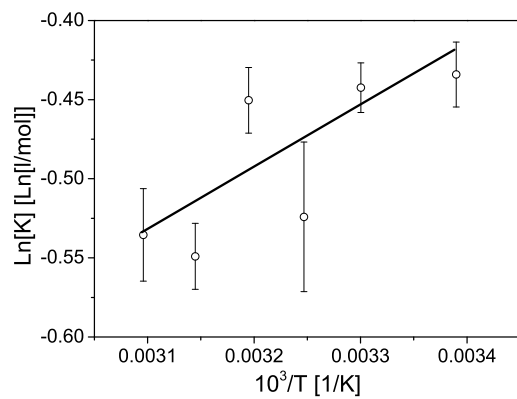


Figure S12: Dependence of  $\ln K$  on  $1/T$  for model 1s.

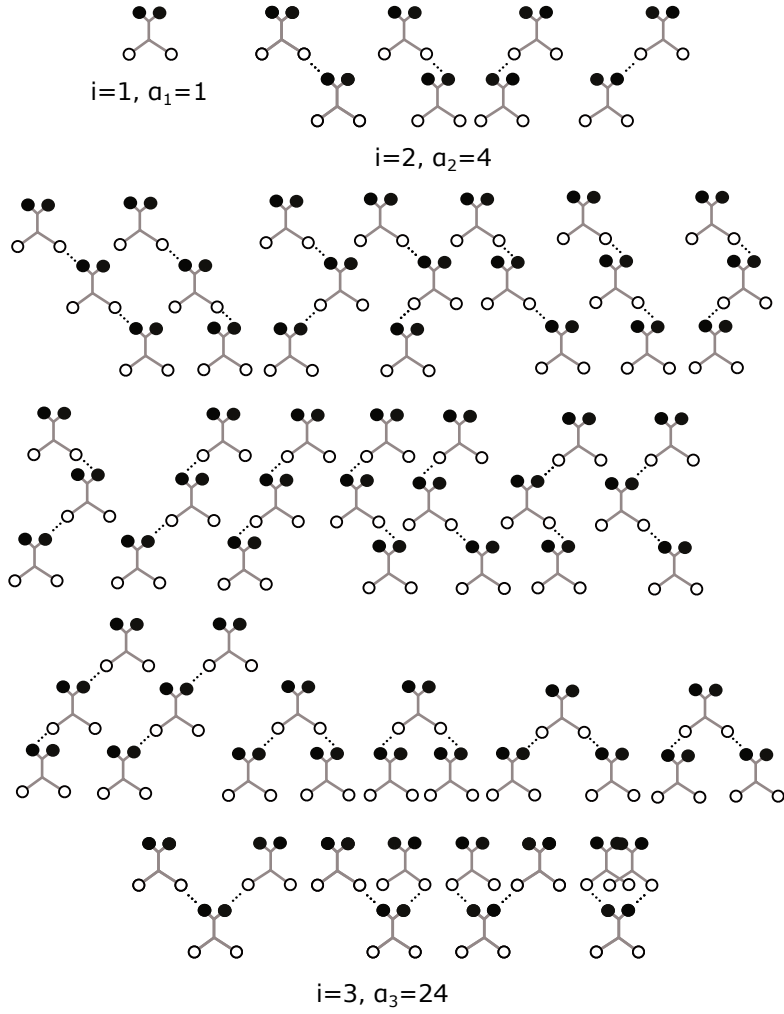


Figure S13: Schematic representation of possible aggregates with size up to  $i = 3$  in model 2.

### 1.3 Model 2

In model 2, the following assumptions are made: oxygen can form two bonds, the  $\text{NH}_2$  group can form two bonds, there are no cycles and there is one association constant. The range of allowed aggregates for this case with sizes up to  $i = 3$  is shown in Figure S13.

The free energy of hydrogen bonding can be written as

$$F = M\epsilon - kT \ln(p^M \Xi), \quad (23)$$

where  $M$  is the number of hydrogen bonds and the number of ways to form these bonds is

$$\Xi = \left( \frac{2N!}{(2N - M)!} \right)^2 \frac{1}{M!}. \quad (24)$$

Minimization of the free energy yields

$$\frac{m}{(2n - m)^2} = K. \quad (25)$$

Next, it is assumed that the concentration of aggregates of size  $i$  can be written as

$$c_i = \alpha_i K^{i-1} c_1^i, \quad (26)$$

so that the total concentration is

$$n = \sum_{i=1}^{\infty} i c_i, \quad (27)$$

and the concentration of bonds is

$$m = \sum_{n=1}^{\infty} (n-1) c_n. \quad (28)$$

Then, we substitute Equations 27 and 28 into Equation 25 written in the form  $m = K(2n - m)^2$  and, equating coefficients in front of like powers of  $c_1$ , we can calculate the first terms of the  $\alpha_i$  sequence. The OEIS tells us that this sequence is probably that known as A000309, whose  $i^{\text{th}}$  term is given by

$$\alpha_i = 2^i \frac{(3i)!}{(2i+1)!(i+1)!}. \quad (29)$$

In the following section, a proof is given that the sequence A000309 is indeed the set of coefficients  $\alpha_i$  that satisfy Equation 25. Using this result for the  $\alpha_i$ , the total concentration can be written down as

$$n = c_1 \cdot {}_3F_2 \left( \left[ \frac{4}{3}, \frac{5}{3}, 2 \right], \left[ \frac{5}{2}, 3 \right], \frac{27Kc_1}{2} \right). \quad (30)$$

### 1.3.1 The proof

*This proof was provided by Mark van Hoeij.*

Let us consider  $\alpha_i$ ,  $m$  and  $n$  defined according to Equations 29, 28, and 27 respectively. Then, let us define a function

$$y = 2n - m = \frac{1}{K} \sum_{i=1}^{\infty} (i+1) \alpha_i x^i, \quad (31)$$

where  $x = Kc_1$ . Next, introduce  $M = Km$  and  $Y = Ky$ . Then, Equation 25 is equivalent to

$$M = Y^2. \quad (32)$$

If a function  $Z$  is defined by

$$Z = 1 + Y = \sum_{i=0}^{\infty} (i+1) \alpha_i x^i \quad (33)$$

(the difference with  $Y$  is that the summation starts from 0), it can be verified that

$$(27x - 2) x Z'' + (54x - 3) Z' + 6Z = 0. \quad (34)$$

Indeed, since

$$3Z' = 3 \sum_{i=0}^{\infty} (i+1) i \alpha_i x^{i-1} = 3 \sum_{i=0}^{\infty} (i+1) (i+2) \alpha_{i+1} x^i \quad (35)$$

and

$$2xZ'' = 2 \sum_{i=0}^{\infty} (i+1) i (i-1) \alpha_i x^{i-1} = 2 \sum_{i=0}^{\infty} i (i+1) (i+2) \alpha_{i+1} x^i, \quad (36)$$

Equation 34 is equivalent to

$$(27i^2 + 27i + 6) \alpha_i = (i+2) (2i+3) \alpha_{i+1}, \quad (37)$$

which is true for  $\alpha_i$  defined by Equation 29. Differentiating Equation 34 gives

$$(27x - 2) x Z''' + (108x - 5) Z'' + 60Z' = 0, \quad (38)$$

or, since  $Y' = Z'$ ,

$$(27x - 2) x Y''' + (108x - 5) Y'' + 60Y' = 0, \quad (39)$$

and it can be found that  $u = Y^2$  satisfies the equation

$$(27x - 2)^2 x^3 u'''''' + 2x^2 (270x - 13) (27x - 2) u'''''' + 35x (2592x^2 - 246x + 5) u'''''' + (204120x^2 - 13440x + 140) u'''' + (146160x - 5040) u'' + 20160u' = 0. \quad (40)$$

To show this, we note that, if we take  $Y^2$  and differentiate repeatedly, then all the resulting expressions can be written in terms of products of  $Y$ ,  $Y'$  and  $Y''$ , because all instances of  $Y'''$  can be eliminated with Equation 39. Next, we verify that  $M = \sum_{i=1}^{\infty} (i-1) \alpha_i x^i$  also satisfies Equation 40. Since both  $M$  and  $Y^2$  satisfy Equation 40, the functions  $M$  and  $Y^2$  are equal if the first six terms in their expansions in powers of  $x$  coincide (because the differential equation is sixth order), which we can check by direct computation.

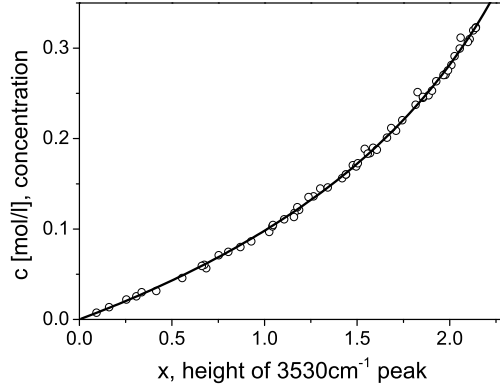


Figure S14: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 2m.

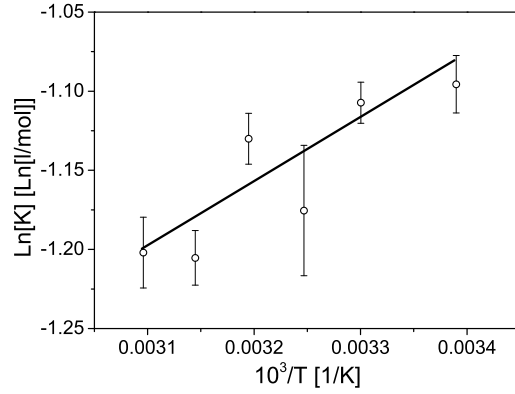


Figure S15: Dependence of  $\ln K$  on  $1/T$  for model 2m.

### 1.3.2 Model 2m

Let us assume first that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the  $\text{NH}_2$  group in unimers. In this case,  $c_1 = Ax$  where  $x$  is the height of the peak and  $A$  is some constant. Substituting this in Equation 30 gives

$$n = Ax \cdot {}_3F_2 \left( \left[ \frac{4}{3}, \frac{5}{3}, 2 \right], \left[ \frac{5}{2}, 3 \right], \frac{27KAx}{2} \right). \quad (41)$$

The results of the fit are shown in Figures S14 and S15. The value of AICc in Figure S14 is  $-502.5$ , and the estimates of the model parameters given by the fit in Figure S15 are  $\ln C = -2.5 \pm 0.3 \ln[\text{l/mol}]$  and  $\epsilon = -0.8 \pm 0.2 \text{ kcal/mol}$  with  $r^2 = 0.75$ .

### 1.3.3 Model 2g

Let us assume now that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of  $\text{NH}_2$  free groups. As in the case of model 1g, we need to include more information about the aggregate in order to calculate the concentration of free groups. If we distinguish between molecules with both hydrogens bonded ( $N_2$  is the number of such molecules) and only one bonded hydrogen in the amide group ( $N_1$  is the number of such molecules) we can write the number of ways to form

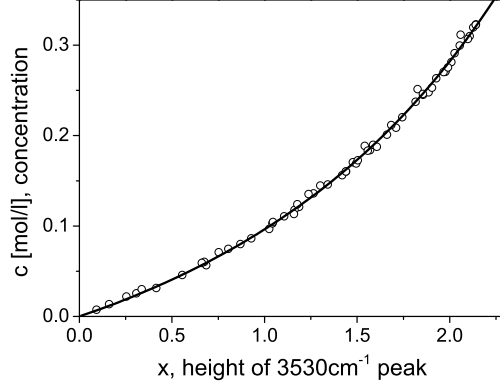


Figure S16: Fit of the dependence of the total concentration on the height of the 3530cm<sup>-1</sup> peak at  $T = 22^\circ\text{C}$  with model 2g.

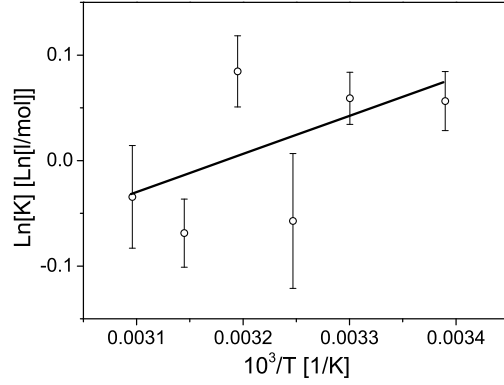


Figure S17: Dependence of  $\ln K$  on  $1/T$  for model 2g.

bonds as (see model 6 for more details)

$$\begin{aligned} \Xi &= \frac{2N!}{(2N-M)!} \frac{N!2^{N_1}}{(N-N_2-N_1)!N_1!N_2!} \\ &= \frac{2N!}{(2N-M)!} \frac{N!2^{2N-M-2N_f}}{N_f!(2N-M-2N_f)!(N_f+M-N)!}, \end{aligned} \quad (42)$$

where  $M = N_1 + 2N_2$  and  $N_f = N - N_1 - N_2$ . Minimizing the free energy and eliminating  $M$  gives

$$c = \frac{n_f}{(1-2Kn_f)^2} = \frac{Ax}{(1-2KAx)^2}, \quad (43)$$

where, according to our peak attribution assumption,  $n_f = Ax$ .

The results of this fit are shown in Figures S16 and S17. The value of AICc for the fit in Figure S16 is  $-499.7$ , and the fit in Figure S17 yields estimates for the model parameters of  $\ln C = -1.2 \pm 0.7 \ln[/math>/mol] and  $\epsilon = -0.7 \pm 0.5 \text{ kcal/mol}$ , with  $r^2 = 0.38$ .$

### 1.3.4 Model 2s

Let us assume here that the 3530cm<sup>-1</sup> peak corresponds to the out-of-phase vibrations of the free NH<sub>2</sub> group in unimers and dimers. In this case,  $Ax = c_1 + 4Kc_1^2$ , where  $x$  is the height of the peak and  $A$  is a constant. Substituting this in Equation 30 yields

$$c = \frac{-1 + \sqrt{1 + 16AKx}}{8K} \cdot {}_3F_2 \left( \left[ \frac{4}{3}, \frac{5}{3}, 2 \right], \left[ \frac{5}{2}, 3 \right], \frac{27}{16} \left( -1 + \sqrt{1 + 16AKx} \right) \right). \quad (44)$$

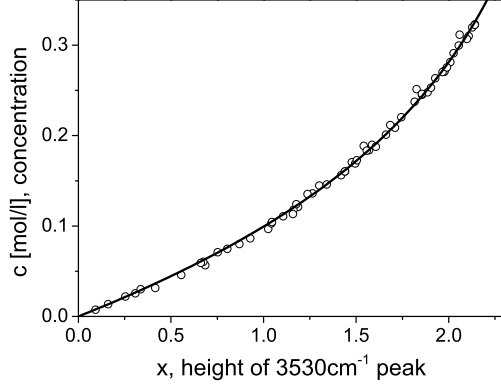


Figure S18: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 2s.

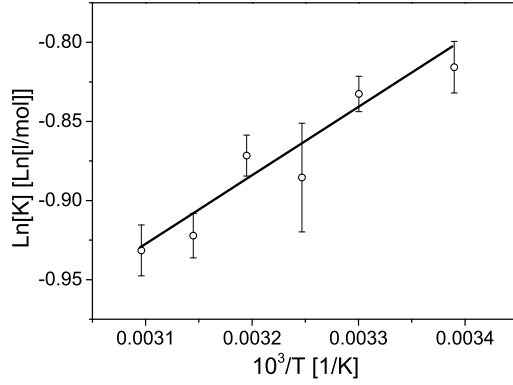


Figure S19: Dependence of  $\ln K$  on  $1/T$  for model 2s.

The results of the fit are shown in Figures S18 and S19. The value of AICc for the fit shown in Figure S18 is  $-494.2$ , and the fit in Figure S19 yields estimates for the model parameters of  $\ln C = 0.50 \pm 0.2 \ln[\text{mol}]$  and  $\epsilon = -0.86 \pm 0.1 \text{ kcal/mol}$ , with  $r^2 = 0.92$ .

The free energy due to hydrogen bonding in model 2 in terms of the volume fraction of hydrogen bonding molecules  $\phi = cv$  and the dimensionless association constant  $K' = K/v$  has the form

$$f_{\text{HB}} = m + 4\phi \ln \frac{(2\phi - m)}{2\phi}, \quad (45)$$

where  $m$  is a solution of the equation  $m/(2\phi - m)^2 = K'$ .

## 1.4 Model 3

In this model, it is assumed that there is only one bond per oxygen and one bond per  $\text{NH}_2$ , there are no cyclic dimers and there is only one association constant. The range of allowed aggregates with aggregation numbers up to  $i = 3$  is shown in Figure S20; aggregates are chain-like in this case.

The number of ways to form bonds in model 3 can be written as

$$\Xi = \left( \frac{N!}{(N-M)!} \right)^2 \frac{4^M}{M!}, \quad (46)$$

and minimization of the free energy gives

$$\frac{M}{4(N-M)^2} = \frac{K}{V}. \quad (47)$$

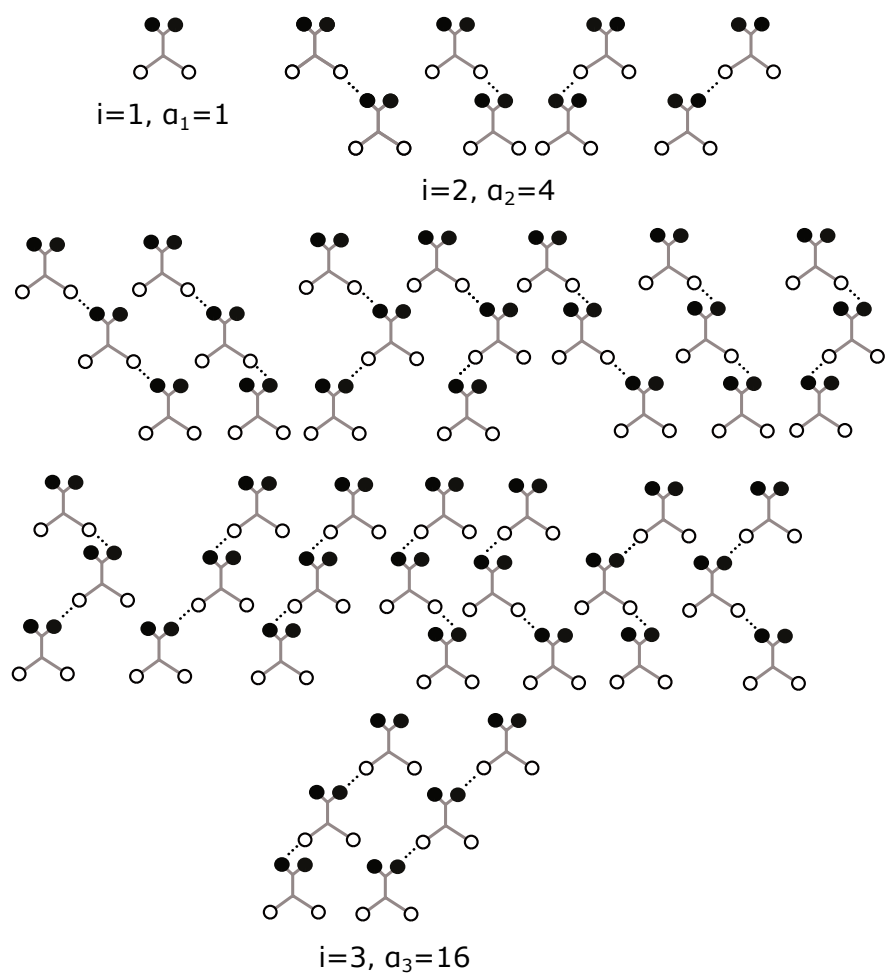


Figure S20: Schematic representation of allowed aggregates with size up to  $i = 3$  in model 3.



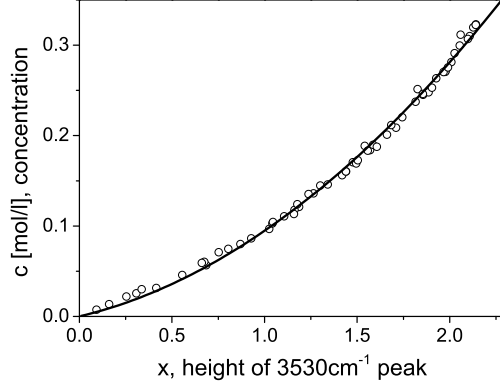


Figure S21: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 3g.

In this case, the concentration of aggregates of size  $i$  is

$$c_i = 4^{i-1} K^{i-1} c_1^i, \quad (48)$$

so for the total concentration we have

$$c = \frac{c_1}{(1 - 4Kc_1)^2}. \quad (49)$$

The concentration of free groups is given by  $n_f = n - m$ , so the relation between the total concentration and the concentration of free groups is

$$n = n_f + 4Kn_f^2. \quad (50)$$

#### 1.4.1 Model 3m

Let us first assume that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the  $\text{NH}_2$  group in unimers. In this case,  $c_1 = Ax$ , where  $x$  is the height of the peak and  $A$  is some constant. The fitting equation is then

$$c = \frac{Ax}{(1 - 4KAx)^2}. \quad (51)$$

This expression is the same as the fitting equation for model 2g, but with an association constant that is two times smaller. This means that the fitting results are the same, with the values of the model parameters and statistical measures being  $\text{AICc} = -499.7$ ,  $\ln C = -1.9 \pm 0.7 \ln[\text{l/mol}]$ ,  $\epsilon = -0.7 \pm 0.5 \text{ kcal/mol}$ , and  $r^2 = 0.38$ .

#### 1.4.2 Model 3g

In this model, it is assumed that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the free  $\text{NH}_2$  groups. Here,

$$c = Ax + 4KA^2x^2. \quad (52)$$

The results of the fitting procedure are shown in Figures S21 and S22. The value of the  $\text{AICc}$  parameter for the fit in Figure S21 is  $-457.6$ , and the estimates of the model parameters given by the fit in S22 are  $\ln C = -0.15 \pm 3 \ln[\text{l/mol}]$  and  $\epsilon = -1.0 \pm 1.9 \text{ kcal/mol}$ , with  $r^2 = 0.07$ .

#### 1.4.3 Model 3s

Let us first assume that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the free  $\text{NH}_2$  groups in unimers and dimers. In this case,  $Ax = c_1 + 4Kc_1^2$ , where  $x$  is the height of the peak and  $A$  is some constant. The fitting equation in this case is

$$c = \frac{-1 + \sqrt{1 + 16KAx}}{2K(-3 + \sqrt{1 + 16KAx})^2}. \quad (53)$$

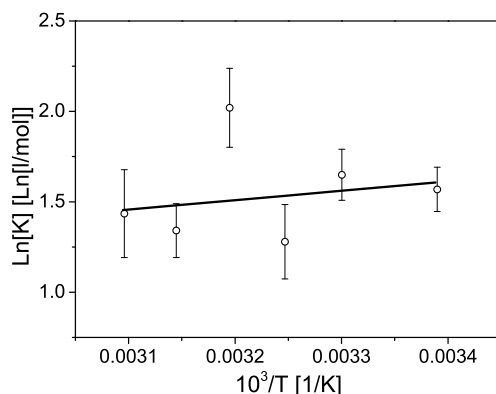


Figure S22: Dependence of  $\ln K$  on  $1/T$  for model 3g.

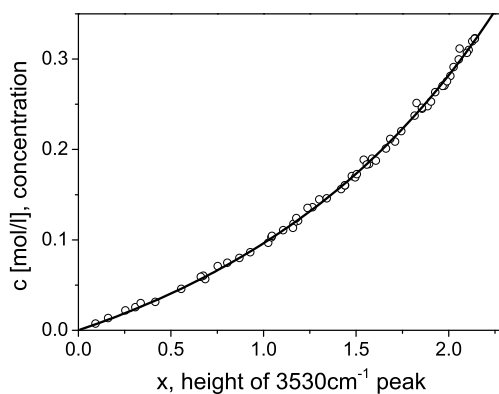


Figure S23: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 3s.

The results of the fitting procedure are shown in Figures S23 and S24, and the values of the model parameters and statistical measures are  $\text{AICc} = -499.3$ ,  $\ln C = 1.58 \pm 0.8 \ln[\text{l/mol}]$ ,  $\epsilon = -0.7 \pm 0.5 \text{ kcal/mol}$ , and  $r^2 = 0.37$ .

The free energy density of hydrogen bonding in model 3, written in terms of the volume fraction of hydrogen bonding molecules  $\phi = cv$  and the dimensionless association constant  $K' = K/v$ , has the form

$$f_{\text{HB}} = m + 2\phi \ln \frac{(\phi - m)}{\phi}, \quad (54)$$

where  $m$  is a solution of the equation  $m/4(\phi - m)^2 = K'$ .

## 1.5 Model 4

Model 4 is analogous to model 1 but with one bond allowed per  $\text{NH}_2$  group and two bonds allowed per O in each acrylamide molecule (see Figure S25). This means that the expression for the relation between  $c$  and  $c_1$  is the same in model 4 as in model 1. However, the number of free  $\text{NH}_2$  groups in model 4 is different from model 1 and is equal to the number of aggregates as there is only one free  $\text{NH}_2$  per aggregate. We can then write that

$$n_f = \sum_{i=1}^{\infty} c_i = \sum_{i=1}^{\infty} \frac{2^{i-1} (2i)!}{i! (i+1)!} K^{i-1} c_1^i = \frac{1 - 4c_1 K - \sqrt{1 - 8c_1 K}}{8c_1 K^2}. \quad (55)$$

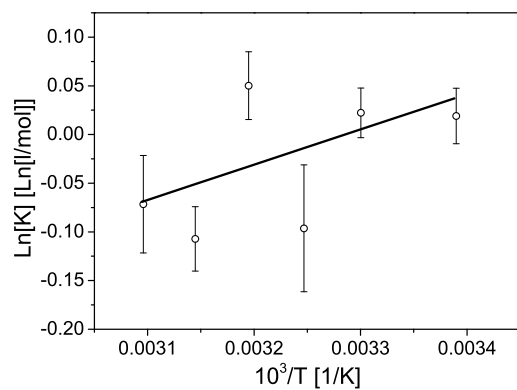


Figure S24: Dependence of  $\ln K$  on  $1/T$  for model 3s.

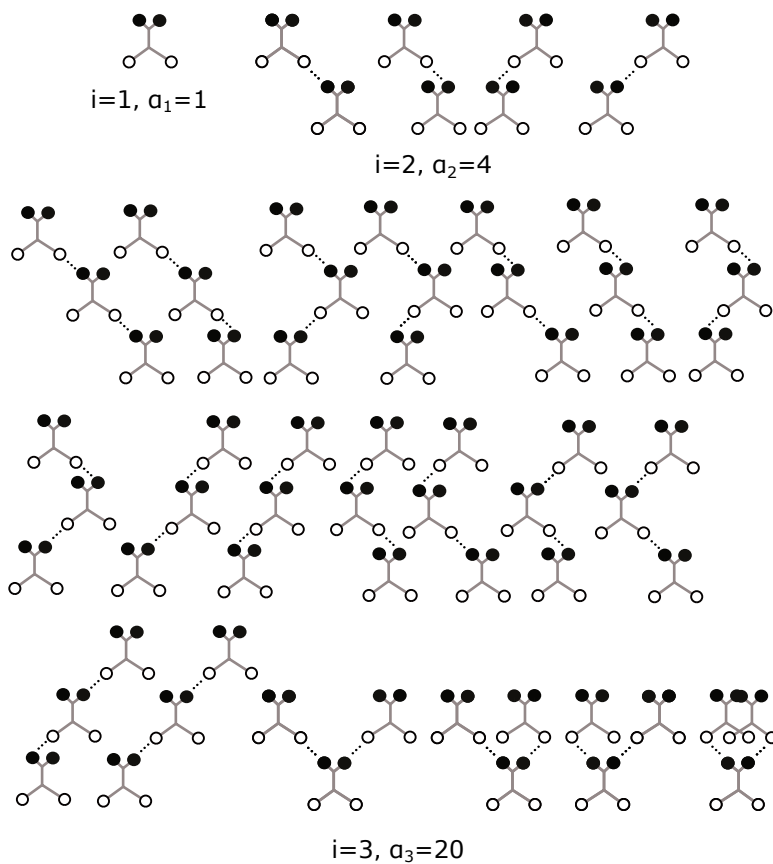


Figure S25: Schematic representation of possible aggregates with size up to  $i = 3$  in model 4.

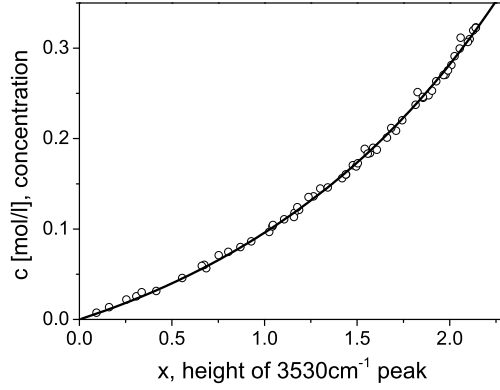


Figure S26: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 4g.

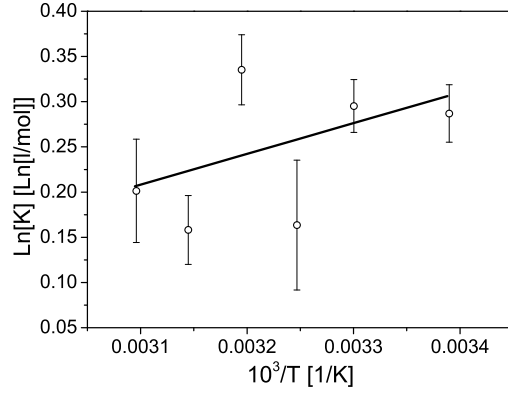


Figure S27: Dependence of  $\ln K$  on  $1/T$  for model 4g.

Substituting this expression into Equation 19 gives

$$c = \frac{n_f + 2Kn_f^2}{1 - 2Kn_f}. \quad (56)$$

### 1.5.1 Model 4m

First, we assume that the  $3530\text{cm}^{-1}$  peak corresponds to out-of-phase vibrations of free acrylamide molecules. The fitting equation is the same as in model 1m, so we do not repeat the fitting procedure here.

### 1.5.2 Model 4g

With the free groups assumption, we can write fitting equation as

$$c = \frac{Ax + 2KA^2x^2}{1 - 2KAx}. \quad (57)$$

The results of the fit are shown in Figures S26 and S27. The quality of the fit in Figure S26 is characterized by the parameter  $\text{AICc} = -495.8$ , and the estimates for the model parameters resulting from the fit in Figure S27 are  $\ln C = -0.8 \pm 0.9 \ln[\text{l/mol}]$  and  $\epsilon = -0.7 \pm 0.5 \text{ kcal/mol}$ , with  $r^2 = 0.28$ .

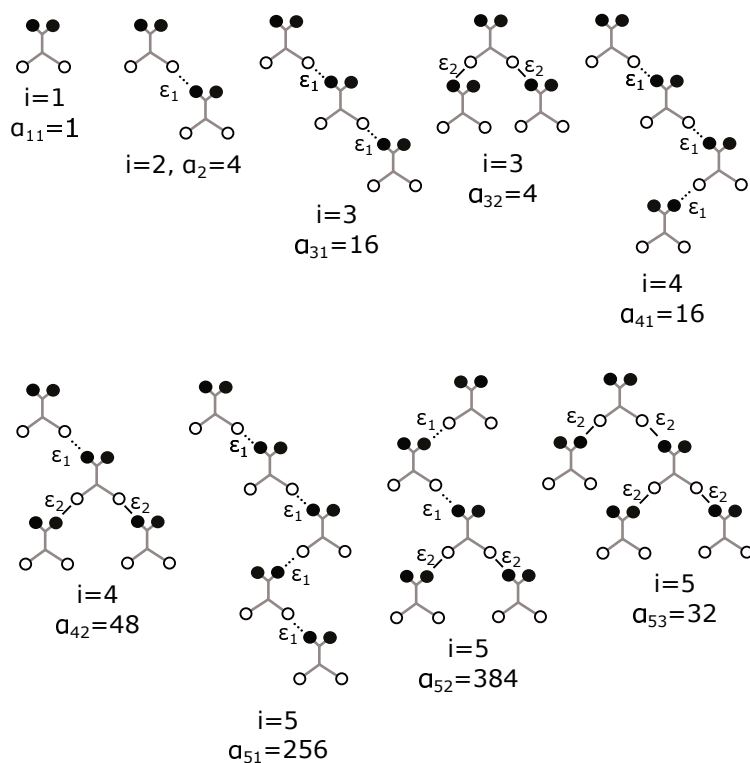


Figure S28: Schematic representation of aggregates with size up to  $i = 5$  in model 5. In contrast to the figures for one-parameter models, only configurations of aggregates with different energies are shown. The quantity  $a_{ij}$  gives the number of different configurations for an aggregate of  $i$  molecules that contains  $2(j-1)$   $\epsilon_2$  bonds.

### 1.5.3 Model 4s

In this case, we assume that the  $3530\text{cm}^{-1}$  peak corresponds to out-of-phase vibrations of free  $\text{NH}_2$  groups in unimers and dimers. The fitting equation and all fitting results are the same as for model 1s.

## 1.6 Model 5

Model 5 is the first model with “cooperativity” we consider. In this model, we allow one bond per oxygen, two bonds per  $\text{NH}_2$  group, no cycles, and two association constants (corresponding to bond energies  $\epsilon_1$  and  $\epsilon_2$ ) that depend on the bonding state of the  $\text{NH}_2$  group in the donor molecule (see Figure S28).

Let us denote the number of molecules with donors involved in  $\epsilon_1$  bonds as  $N_1$  (or, in other words, the number of donor molecules with one bonded hydrogen) and the number of molecules involved in  $\epsilon_2$  bonds as  $N_2$  (in other words, the number of donor molecules with two bonded hydrogens). Correspondingly, the number of  $\epsilon_1$  bonds is  $M_1 = N_1$  and number of  $\epsilon_2$  bonds is  $M_2 = 2N_2$ .

The number of ways to form  $M_1$  bonds with energy  $\epsilon_1$  and  $2N_2$  bonds with energy  $\epsilon_2$  can be written as

$$\Xi = \frac{N!2^{N_1+2N_2}}{(N-N_1-2N_2)!} \frac{N!2^{N_1}}{(N-N_1-N_2)!N_1!N_2!}, \quad (58)$$

where the first factor is the number of ways to choose an acceptor for  $M = N_1 + 2N_2$  bonds, and the second factor is the number of ways to choose  $N_1$  and  $N_2$  donor groups out of  $N$  molecules, taking into account the fact that in molecules with only one bonded hydrogen, this hydrogen can be chosen in two ways.

The free energy of hydrogen bonding in model 5 is

$$F_{\text{HB}} = \epsilon_1 M_1 + \epsilon_2 M_2 - kT \ln \left( p_1^{M_1} p_2^{M_2} \Xi \right), \quad (59)$$

and minimizing this with respect to  $M_1$  and  $M_2$  gives

$$\frac{N_1}{4(N - N_1 - 2N_2)(N - N_1 - N_2)} = \frac{K_1}{V} \quad (60)$$

and

$$\frac{N_2}{4(N - N_1 - N_2)(N - N_1 - 2N_2)^2} = \frac{K_2^2}{V^2}. \quad (61)$$

It is interesting to note that the following equality exists:

$$\frac{(2N)!}{(2N - M)!M!} = \sum_{N_2=0}^{M/2} \frac{N!2^{M-2N_2}}{(N - M + N_2)!(M - 2N_2)!N_2!}, \quad (62)$$

which shows that model 5 reduces to model 1 when  $\epsilon_1 = \epsilon_2$  – a property that we make use of in our calculations on model 1g above.

Now, we look for a relation between the total concentration and the concentration of unimers, and assume that concentration of aggregates of size  $i$  with  $2(j - 1)$   $\epsilon_2$ -bonds has the form

$$c_{ij} = \alpha_{ij} K_2^{2(j-1)} K_1^{i-1-2(j-1)} c_1^i, \quad 1 \leq j \leq (i + 1)/2. \quad (63)$$

Substituting this expression in Equations 60 and 61 gives

$$\alpha_{ij} = \frac{2^{2i-2j}}{(j - 1)!j!} \frac{(i - 1)!}{(i - 2j + 1)!}, \quad (64)$$

and the dependence of the total solution concentration on the concentration of unimers can then be calculated to be

$$c = \sum_{i=1}^{\infty} \sum_{j=1}^{(i+1)/2} i \alpha_{ij} K_2^{2(j-1)} K_1^{i-1-2(j-1)} c_1^i = \frac{1 - \sqrt{1 - \frac{16K_2^2 c_1^2}{(1-4K_1 c_1)^2}}}{8K_2^2 c_1 \sqrt{1 - \frac{16K_2^2 c_1^2}{(1-4K_1 c_1)^2}}}. \quad (65)$$

Alternatively, the coefficients  $\alpha_{ij}$  can be found from the generating function for the family of trees shown in Figure S28. If we denote  $c_1 K_1$  as  $z_1$  and  $c_1 K_2$  as  $z_2$ , we can write down the equation for the generating function as

$$G = 1 + 4Gz_1 + 4z_2^2 G^2, \quad (66)$$

which can be solved to find

$$G = \frac{1 - 4z_1 \pm \sqrt{(1 - 4z_1)^2 - 16z_2^2}}{8z_2^2} \quad (67)$$

where the required expression is that with the negative root. It is straightforward to verify that expansion of this expression in powers of  $z_1$  and  $z_2$  will yield the values of  $\alpha_{ij}$  given by Equation 64. We also can see that if we put  $z_1 = z_2$  we will recover the generating function for model 1, as would be expected from the fact that the two models are equivalent when  $\epsilon_1 = \epsilon_2$ .

The concentration of free groups is given by  $n_f = c - n_1 - n_2$  and eliminating  $n_1$  and  $n_2$  from Equations 60 and 61 gives

$$c = \frac{1 - 4K_1 n_f + 16K_2^2 n_f^2 - (1 - 4K_1 n_f) \sqrt{1 + 16K_2^2 n_f^2}}{8K_2^2 n_f}. \quad (68)$$

### 1.6.1 Model 5m

Let us assume first that the  $3530\text{cm}^{-1}$  peak corresponds to out-of-phase vibrations of the  $\text{NH}_2$  group in free molecules. In this case, the fitting equation is

$$c = \frac{1 - \sqrt{1 - \frac{16A^2 K_2^2 x^2}{(1-4AK_1 x)^2}}}{8AK_2^2 x \sqrt{1 - \frac{16A^2 K_2^2 x^2}{(1-4AK_1 x)^2}}}. \quad (69)$$

The results of the fit are shown in Figures S29 and S30. The value of the AICc parameter for the nonlinear fit in Figure S29 is  $-501.7$ , and the estimates of the model parameters given by the fits in Figure S30 are  $\ln C_1 = -19.6 \pm 5 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -11 \pm 3 \text{ kcal/mol}$ ,  $\ln C_2 = 3.4 \pm 1 \ln[\text{l/mol}]$ , and  $\epsilon_2 = 2.4 \pm 0.7 \text{ kcal/mol}$ . In all our two-parameter models, we have two coefficients of determination, which in this case are given by  $r_1^2 = 0.76$  and  $r_2^2 = 0.73$ .

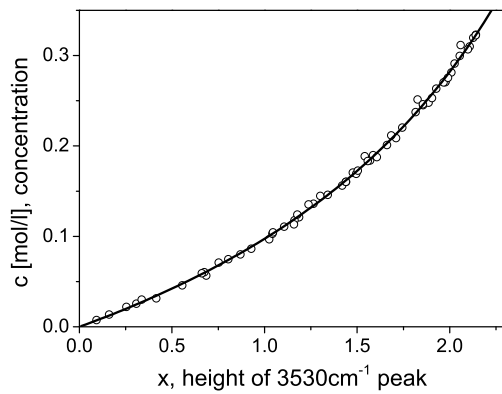


Figure S29: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 5m.

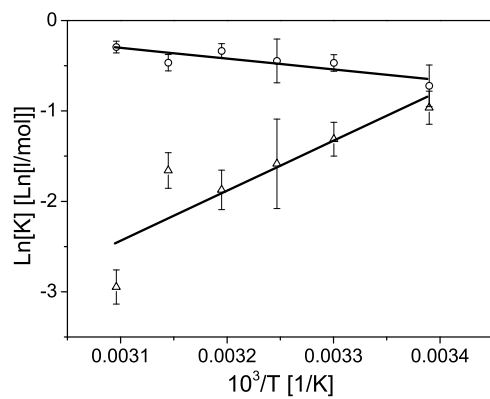


Figure S30: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 5m.

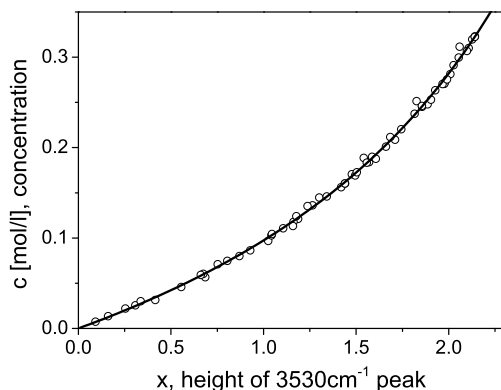


Figure S31: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 5s.

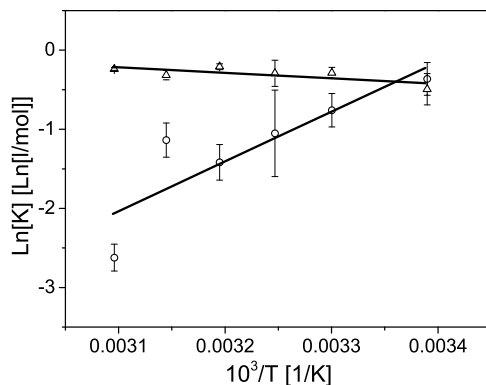


Figure S32: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 5s.

### 1.6.2 Model 5g

Now we assume that the  $3530\text{cm}^{-1}$  peak corresponds to out-of-phase vibrations of free  $\text{NH}_2$  groups. In this case, the fitting equation is

$$c = \frac{1 - 4AK_1x + 16A^2K_2^2x^2 - (1 - 4AK_1x)\sqrt{1 + 16A^2K_2^2x^2}}{8AK_2^2x}. \quad (70)$$

In this case, we were unable to obtain any results, as the fitting procedure did not converge.

### 1.6.3 Model 5s

We also check the possibility that the  $3530\text{cm}^{-1}$  peak does not correspond to a single species (such as all free molecules or all free groups), but instead corresponds to absorption by free groups in some subset of aggregates. Here we check the subset composed of unimers and dimers, so  $Ax = c_1 + 4K_1c_1^2$ .

The results of the fit are shown in Figures S31 and S32. The value of AICc for the fit in Figure S31 is  $-501.7$ , and the estimates of the model parameters given by the fits in S32 are  $\ln C_1 = -21.5 \pm 6 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -12.5 \pm 3.5 \text{ kcal/mol}$ ,  $\ln C_2 = 1.95 \pm 1 \ln[\text{l/mol}]$ , and  $\epsilon_2 = 1.3 \pm 0.6 \text{ kcal/mol}$ , with  $r_1^2 = 0.76$  and  $r_2^2 = 0.56$ .

### 1.7 Model 6

In model 6, we allow one bond per oxygen, two bonds per  $\text{NH}_2$  group, no cycles and two association constants. However, in contrast to model 5, the association constant is now determined by the



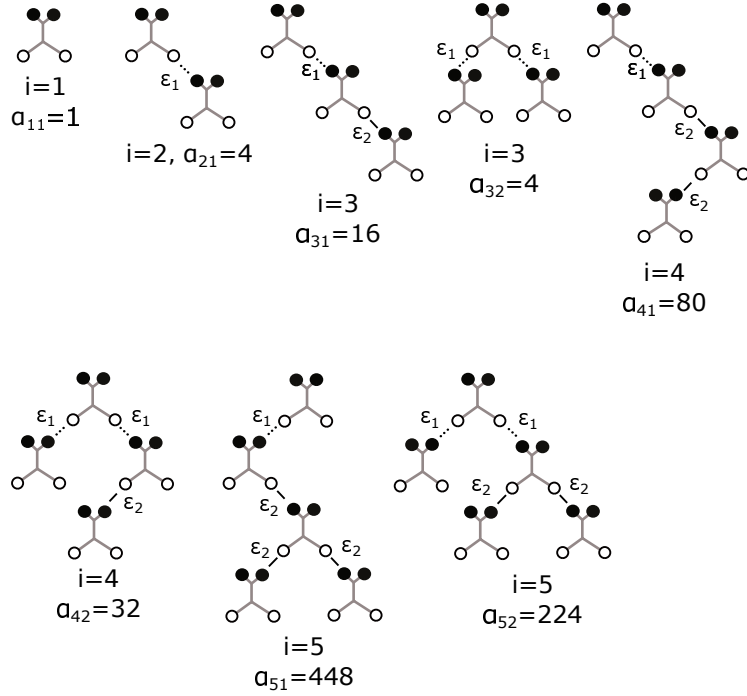


Figure S33: Schematic representation of aggregates with size up to  $i = 5$  in model 6. Only aggregates with different total bond energies are shown.

bonding state of the acceptor in the donor molecule (see Figure S33). Let the bond energy be denoted by  $\epsilon_1$  in the case when the oxygen in the donor amide group is free and by  $\epsilon_2$  otherwise. Then, the number of bonds with energy  $\epsilon_1$  is  $M_1$ , the number of bonds with energy  $\epsilon_2$  is  $M_2$ , the number of molecules with donors involved in  $\epsilon_1$  bonds is  $N_1$ , and the number of molecules involved in  $\epsilon_2$  bonds is  $N_2$ . The number of ways to form bonds can then be written as

$$\Xi = \frac{N!2^{M_1+M_2}}{(N - M_1 - M_2)!} \frac{(2N - 2M_1 - 2M_2)!}{(2N - 3M_1 - 2M_2)!} \frac{(2M_1 + 2M_2)!}{(2M_1 + M_2)!} \frac{1}{M_1!M_2!}, \quad (71)$$

where the first factor in  $\Xi$  is the number of ways to choose an acceptor for  $M_1 + M_2$  bonds. This uses the assumption that there is only one bond per oxygen. The second factor is the number of ways to choose a donor for  $\epsilon_1$  bonds, with  $2N - 2M_1 - 2M_2$  giving the number of hydrogens in molecules with free acceptor groups. The third factor is the number of ways to choose a donor for  $\epsilon_2$  bonds. The hydrogens for these bonds should be chosen from molecules with bonded acceptors. The number of such molecules is  $M_1 + M_2$  and they contain  $2M_1 + 2M_2$  hydrogens. As usual, the last term accounts for the indistinguishability of the bonds.

Minimization of the free energy gives

$$\frac{M_1 (2M_1 + M_2)^2 (N - M_1 - M_2)}{2 (M_1 + M_2)^2 (2N - 3M_1 - 2M_2)^3} = \frac{K_1}{V} \quad (72)$$

$$\frac{M_2 (2M_1 + M_2) (N - M_1 - M_2)}{2 (M_1 + M_2)^2 (2N - 3M_1 - 2M_2)^2} = \frac{K_2}{V}. \quad (73)$$

We can notice that there are two types of aggregates: those with one  $\epsilon_1$  bond and those with two  $\epsilon_1$  bonds. Then, the concentrations of aggregates with size  $i$  and either one or two  $\epsilon_1$  bonds can be written as

$$c_{i1} = \alpha_{i1} K_1 K_2^{i-2} c_1^i, \quad i \geq 2 \quad (74)$$

$$c_{i2} = \alpha_{i2} K_1^2 K_2^{i-3} c_1^i, \quad i \geq 3, \quad (75)$$

and the total concentration of acrylamide and concentrations of each type of bond as

$$c = c_1 + \sum_{i=2}^{\infty} i c_{i1} + \sum_{i=3}^{\infty} i c_{i2} \quad (76)$$

$$m_1 = \sum_{i=2}^{\infty} c_{i1} + \sum_{i=3}^{\infty} 2c_{i2} \quad (77)$$

$$m_2 = \sum_{i=2}^{\infty} (i-2) c_{i1} + \sum_{i=3}^{\infty} (i-3) c_{i2}. \quad (78)$$

Substituting these expressions into Equations 72 and 73 leads to the following expressions for  $\alpha_{i1,2}$ :

$$\alpha_{i1} = 2^i \frac{(2i-2)!}{(i-1)!i!}, i \geq 2 \quad (79)$$

$$\alpha_{i2} = 2^{i+1} \frac{(2i-3)!}{(i-3)!(i+1)!}, i \geq 3. \quad (80)$$

For the dependence of the total concentration on the concentration of unimers we have

$$c = \frac{16c_1^2 K_1 K_2^3 - 16c_1^2 K_2^2 K_1^2 + K_1^2 - 4c_1 K_1^2 K_2}{8c_1 K_2^4 \sqrt{1 - 8c_1 K_2}} + \frac{8c_1^2 K_2^4 - 16c_1^2 K_1 K_2^3 - K_1^2 + 8c_1^2 K_1^2 K_2^2}{8c_1 K_2^4}. \quad (81)$$

Another way to determine the coefficients  $\alpha_{i1}$  and  $\alpha_{i2}$  is to find an expression for the generating function, as was done in for models 1 and 5. In the current case, we have

$$G(z_1, z_2) = 1 + 4z_1 G_1(z_2) + 4z_1^2 G_1^2(z_2), \quad (82)$$

where

$$G_1(z_2) = \frac{1 - 4z_2 - \sqrt{1 - 8z_2}}{8z_2^2}, \quad (83)$$

the generating function for model 1.

Now, let us turn to the calculation of the number of free groups. In order to do this, it is necessary to distinguish molecules with one hydrogen bond per  $\text{NH}_2$  group and two hydrogen bonds per  $\text{NH}_2$  group. We denote the number of molecules with one bonded hydrogen and free oxygen by  $N_1$  (Figure S34) and the number of molecules with two bonded hydrogens and free oxygen by  $N_3$ , so that  $M_1 = N_1 + 2N_3$ . Similarly, we denote the number of molecules with one bonded hydrogen and bonded oxygen by  $N_2$ , and the number of molecules with two bonded hydrogens and bonded oxygen, by  $N_4$ , so that  $M_2 = N_2 + 2N_4$ . Then, the number of ways to form bonds is

$$\Xi = \frac{N! 2^{M_1+M_2}}{(N - M_1 - M_2)!} \frac{(N - M_1 - M_2)! 2^{N_1}}{(N - M_1 - M_2 - N_1 - N_3)! N_1! N_3!} \frac{(M_1 + M_2)! 2^{N_2}}{(M_1 + M_2 - N_2 - N_4)! N_2! N_4!}. \quad (84)$$

After simplification, this becomes

$$\Xi = \frac{N! (N_1 + N_2 + 2N_3 + 2N_4)! 4^{N_1+N_2+N_3+N_4}}{(N - 2N_1 - 3N_3 - N_2 - 2N_4)! (N_1 + 2N_3 + N_4)! N_1! N_2! N_3! N_4!}. \quad (85)$$

Also, for simplicity, we assume first that there are four association constants and correspondingly four different bond energies, so we can write for the free energy of hydrogen bonding

$$F_{\text{HB}} = \epsilon_1 N_1 + \epsilon_2 N_2 + 2\epsilon_3 N_3 + 2\epsilon_4 N_4 - kT \ln \left( p_1^{N_1} p_2^{N_2} p_3^{2N_3} p_4^{2N_4} \Xi \right). \quad (86)$$

Minimization of this expression gives

$$\frac{N_1 (N_1 + 2N_3 + N_4)}{4 (N_1 + N_2 + 2N_3 + 2N_4) (N - 2N_1 - 3N_3 - N_2 - 2N_4)^2} = \frac{K_1}{V} \quad (87)$$

$$\frac{N_2}{4 (N_1 + N_2 + 2N_3 + 2N_4) (N - 2N_1 - 3N_3 - N_2 - 2N_4)} = \frac{K_2}{V} \quad (88)$$

$$\frac{N_3 (N_1 + 2N_3 + N_4)^2}{4 (N_1 + N_2 + 2N_3 + 2N_4)^2 (N - 2N_1 - 3N_3 - N_2 - 2N_4)^3} = \frac{K_3^2}{V^2} \quad (89)$$

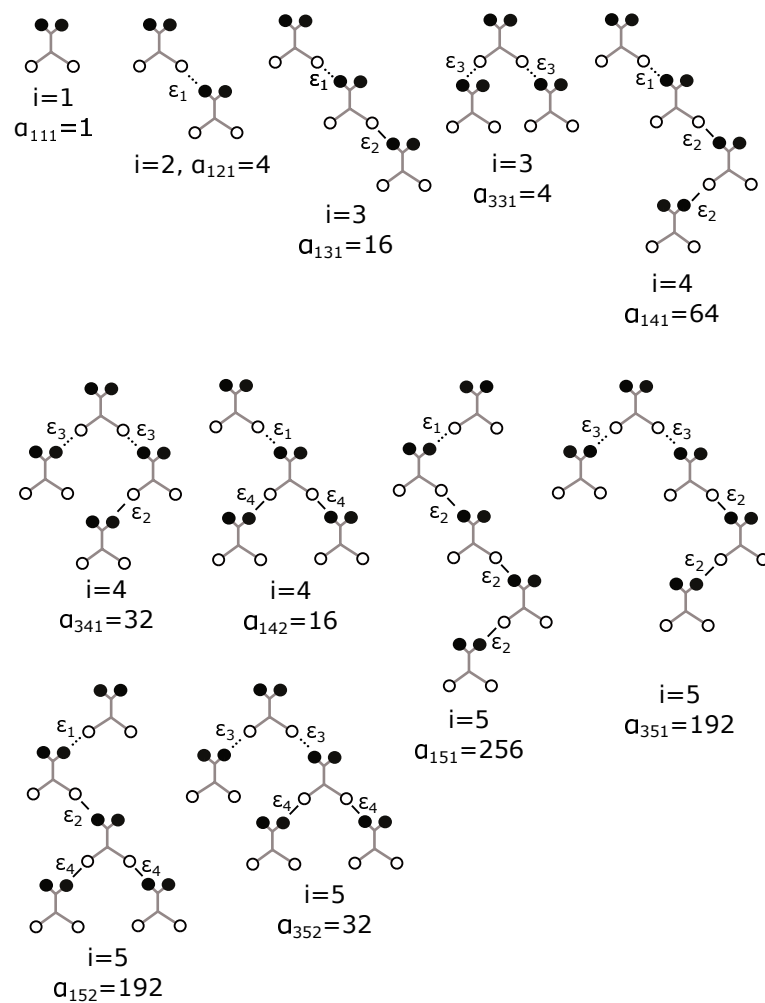


Figure S34: Schematic representation of aggregates with size up to  $i = 5$  in the modification of model 6 with four association constants. Only aggregates with different total bond energies are shown.

$$\frac{N_4 (N_1 + 2N_3 + N_4)}{4(N_1 + N_2 + 2N_3 + 2N_4)^2 (N - 2N_1 - 3N_3 - N_2 - 2N_4)^2} = \frac{K_4^2}{V^2}, \quad (90)$$

and the number of free groups is given by  $N_f = N - N_1 - N_2 - N_3 - N_4$ .

We notice that, since only one bond per oxygen is allowed, all aggregates have either one  $\epsilon_1$  bond or two  $\epsilon_3$  bonds, which lie at the “root” of each aggregate. Furthermore, we note that the distribution of  $\epsilon_2$  and  $\epsilon_4$  bonds is very similar to the original model 6. Then, we can label different aggregates by the set of three numbers  $\{1ij\}$  or  $\{3ij\}$  where the first letter denotes the type of “root” (type 1 bonds or type 3 bonds),  $i$  is the number of molecules in the aggregate and  $j$  the number of type 4 bonds in aggregate. The concentrations of aggregates can then be written as

$$c_{1ij} = \alpha_{1ij} K_1 K_2^{i-2j} K_4^{2j-2} c_1^i \quad (91)$$

$$c_{3ij} = \alpha_{3ij} K_3^2 K_2^{i-1-2j} K_4^{2j-2} c_1^i \quad (92)$$

and the concentrations of molecules in different bonding states can be calculated to be

$$c = c_1 + \sum_{i=2}^{\infty} \sum_{j=1}^{i/2} i c_{1ij} + \sum_{i=3}^{\infty} \sum_{j=1}^{(i-1)/2} i c_{3ij} \quad (93)$$

$$c_1 = \sum_{i=2}^{\infty} \sum_{j=1}^{i/2} c_{1ij} = \sum_{j=1}^{\infty} \sum_{i=2j}^{\infty} c_{1ij} \quad (94)$$

$$c_3 = \sum_{i=3}^{\infty} \sum_{j=1}^{(i-1)/2} c_{3ij} = \sum_{j=1}^{\infty} \sum_{i=2j+1}^{\infty} c_{3ij} \quad (95)$$

$$c_2 = \sum_{j=1}^{\infty} \sum_{i=2j}^{\infty} (i-2j) c_{1ij} + \sum_{j=1}^{\infty} \sum_{i=2j+1}^{\infty} (i-1-2j) c_{3nj} \quad (96)$$

$$c_4 = \sum_{j=1}^{\infty} \sum_{i=2j}^{\infty} (j-1) c_{1ij} + \sum_{j=1}^{\infty} \sum_{i=2j+1}^{\infty} (j-1) c_{3ij}. \quad (97)$$

Additionally, we can write that

$$n_f = c_1 + \sum_{j=1}^{\infty} \sum_{i=2j}^{\infty} j c_{1ij} + \sum_{j=1}^{\infty} \sum_{i=2j+1}^{\infty} (j+1) c_{3ij}. \quad (98)$$

Substituting these expressions into Equations 87, 88, 89, 90 gives

$$\alpha_{1ij} = 2^{2i-2j} \frac{1}{j! (j-1)!} \frac{(i-2)!}{(i-2j)!}, i \geq 2j \quad (99)$$

$$\alpha_{3ij} = 2^{2i-2j-1} \frac{1}{(j+1)! (j-1)!} \frac{(i-2)!}{(i-2j-1)!}, i \geq 2j+1. \quad (100)$$

We can now evaluate the sums and calculate the dependence of  $n_f$  on  $c_1$ .

Since we assumed that equilibrium is described by two association constants in model 6, we put  $K_3 = K_1$  and  $K_4 = K_2$  and find

$$n_f = \frac{c_1 (4c_1 K_1 K_2^2 + K_2^2 \sqrt{1 - 8c_1 K_2} + K_1^2 (1 - 4c_1 K_2 - \sqrt{1 - 8c_1 K_2}))}{K_2^2 \sqrt{1 - 8c_1 K_2}}. \quad (101)$$

As we know the dependence of both  $c$  and  $n_f$  on  $c_1$ , we have parametrically defined a function  $c(n_f)$ , which we can use to fit experimental data in case of the free groups assumption.

### 1.7.1 Model 6m

Here, it is assumed that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the  $\text{NH}_2$  group in free molecules. In this case, we put  $c_1 = Ax$  and substitute it into Equation 81.

The fitting results are shown in Figures S35 and S36. The value of the information criterion for the fit shown in Figure S35 is  $\text{AICc} = -501.7$ , and the estimates of the model parameters given by the fits in Figure S36 are  $\ln C_1 = -9.9 \pm 1 \ln[\text{mol}]$ ,  $\epsilon_1 = -5.2 \pm 0.8 \text{ kcal/mol}$ ,  $\ln C_2 = -0.62 \pm 0.5 \ln[\text{mol}]$ , and  $\epsilon_2 = 0.1 \pm 0.2 \text{ kcal/mol}$ , with  $r_1^2 = 0.91$  and  $r_2^2 = 0.03$ .

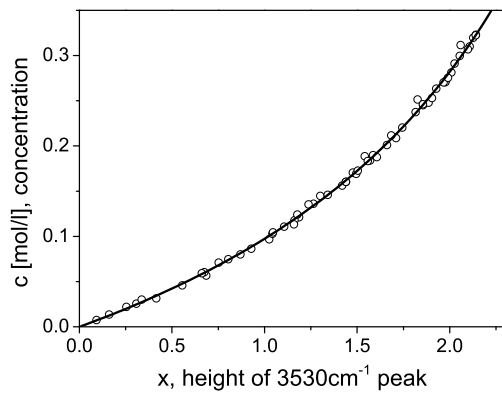


Figure S35: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 6m.

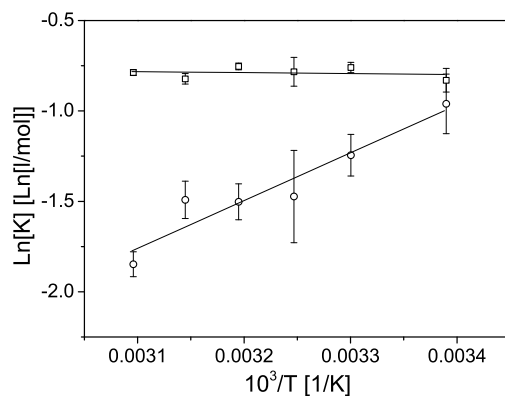


Figure S36: Dependence of  $\ln K_1$  (circles) and  $\ln K_2$  (squares) on  $1/T$  for model 6m.

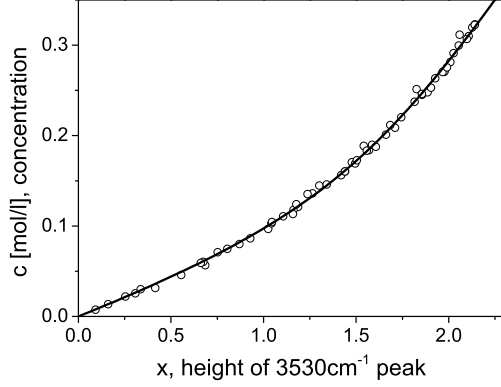


Figure S37: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 6g.

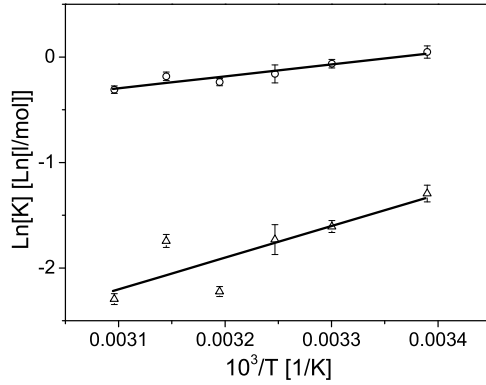


Figure S38: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 6g.

### 1.7.2 Model 6g

In this section, it is assumed that the  $3530\text{cm}^{-1}$  peak corresponds to absorption by free groups, so that  $n_f = Ax$ . The results of the fitting procedure are shown in Figures S37 and S38. For the fit in Figure S37, we have that  $\text{AICc} = -493.8$ . The estimates of the model parameters for the fits in Figure S38 are  $\ln C_1 = -11.5 \pm 5 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -5.9 \pm 3 \text{ kcal/mol}$ ,  $\ln C_2 = 3.5 \pm 0.6 \ln[\text{l/mol}]$ , and  $\epsilon_2 = -2.2 \pm 0.3 \text{ kcal/mol}$ , with  $r_1^2 = 0.71$  and  $r_2^2 = 0.90$ .

### 1.7.3 Model 6s

For the assumption  $Ax = c_1 + 4K_1c_1^2 + 4K_1^2c_1^3$  ( $\text{NH}_2$  groups in unimers, dimers and trimers without  $\epsilon_2$ -bonds), we have the fitting results shown in Figures S39 and S40. The quality of fit in Figure S39 can be characterized by  $\text{AICc} = -501.7$ . The estimates of parameters for the fits in Figure S40 are  $\ln C_1 = -11.89 \pm 2 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -6.7 \pm 1.1 \text{ kcal/mol}$ ,  $\ln C_2 = -3.18 \pm 0.3 \ln[\text{l/mol}]$ , and  $\epsilon_2 = -1.6 \pm 0.2 \text{ kcal/mol}$ , with  $r_1^2 = 0.91$  and  $r_2^2 = 0.95$ .

The free-energy density of hydrogen bonding in model 6 in terms of the volume fraction of hydrogen bonding molecules  $\phi = cv$  and dimensionless association constants  $K'_1 = K_1/v$ ,  $K'_2 = K_2/v$  has the form

$$f_{\text{HB}} = m_1 + m_2 + \phi \ln \frac{(2\phi - 3m_1 - 2m_2)^2}{4\phi(\phi - m_1 - m_2)}, \quad (102)$$

where  $m_1$  and  $m_2$  can be calculated by taking the sums in Equations 77 and 78 to be functions of the association constants and  $c_1$ . Then, we can consider  $c_1$  as a parameter and now have a parametrically defined function  $f_{\text{HB}}(\phi)$ .

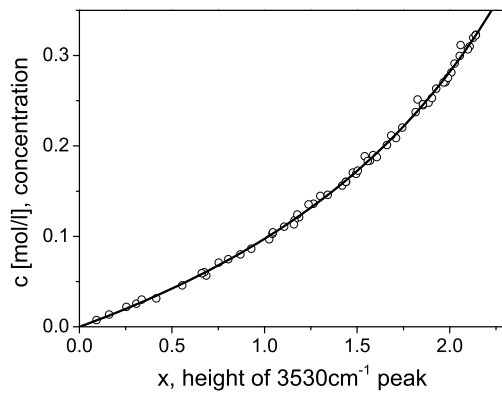


Figure S39: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 6s.

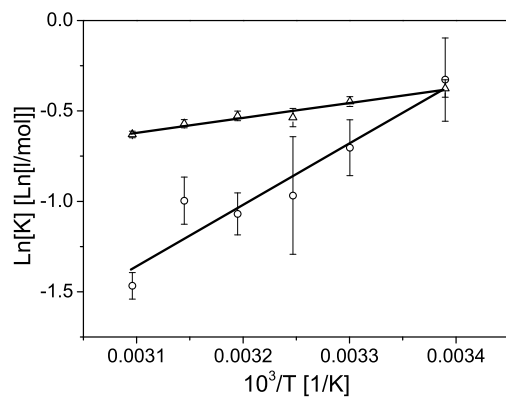


Figure S40: Dependence of  $\ln K_1$  (circles) and  $\ln K_2$  (squares) on  $1/T$  for model 6s.

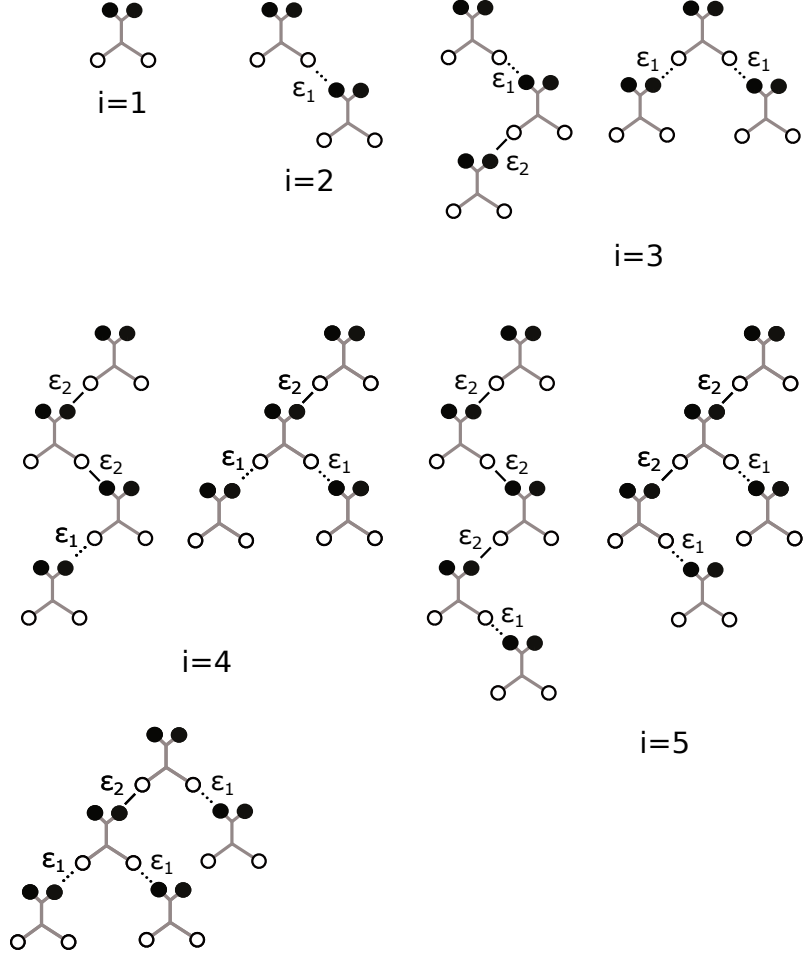


Figure S41: Schematic representation of aggregates with size up to  $i = 5$  in model 7. Only configurations corresponding to different energies are shown.

## 1.8 Model 7

In model 7, we again allow one bond per oxygen, two bonds per  $\text{NH}_2$  group, no cyclic dimers and two association constants. However, in contrast to models 5 and 6, the association constant is now determined by the bonding state of the  $\text{NH}_2$  group in the acceptor molecule (see Figure S41).

We denote the energy of a bond formed by an acceptor with a free donor group by  $\epsilon_1$  and the energy of a bond formed by an acceptor molecule with one or two bonded hydrogens by  $\epsilon_2$ . As in some previous models, we will first consider a more detailed case (with three association constants), in which the situations corresponding to one bonded hydrogen in an acceptor molecule ( $\epsilon_2$ ) and to two bonded hydrogens in an acceptor molecule ( $\epsilon_3$ ) differ from each other (see Figure S42). We write  $M_1$  for the number of bonds with energy  $\epsilon_1$ ,  $M_2$  for the number of bonds with energy  $\epsilon_2$ , and  $M_3$  for the number of bonds with energy  $\epsilon_3$  (see Figure S42). Similarly, we write  $N_1$  for the number of free molecules,  $N_2$  for the number of molecules with a free oxygen and one bonded hydrogen,  $N_3$  for the number of molecules with a free oxygen and two bonded hydrogens,  $N_4$  for the number of molecules with a bonded oxygen and both hydrogens free,  $N_5$  for the number of molecules with a bonded oxygen and one bonded hydrogen, and  $N_6$  for the number of molecules with a bonded oxygen and two bonded hydrogens. Among these values, the following relations exist:  $M_1 = N_4$ ,  $M_2 = N_5$ ,  $M_3 = N_6$ ,  $N = \sum_{i=1}^6 N_i$ ,  $M_1 + M_2 + M_3 = N_2 + 2N_3 + N_5 + 2N_6$ .

In this case the number of ways to choose acceptors is

$$\frac{N! 2^{N_4 + N_5 + N_6}}{(N - N_4 - N_5 - N_6)!} \quad (103)$$



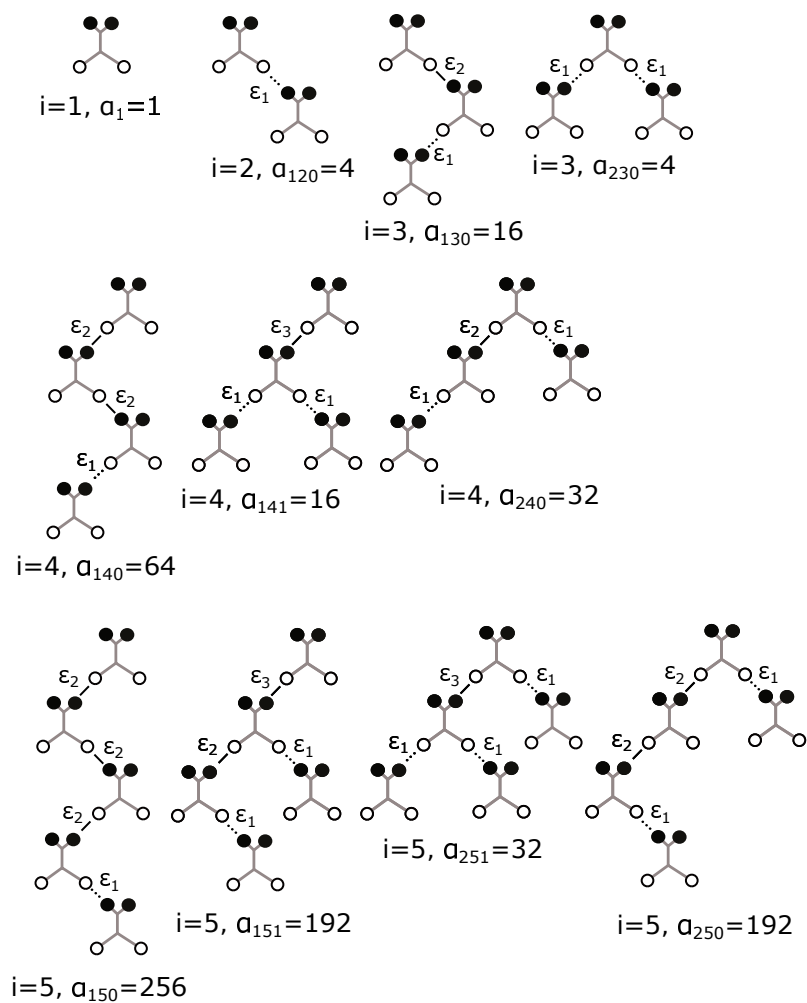


Figure S42: Schematic representation of aggregates with size up to  $i = 5$  in model 7. Only configurations corresponding to different energies are shown.

The number of ways to choose donors is (this expression is effectively the same thing as Equation 62)

$$\frac{(N - N_4 - N_5 - N_6)! 2^{N_2} (N_4 + N_5 + N_6)! 2^{N_5}}{N_1! N_2! N_3! N_4! N_5! N_6!}, \quad (104)$$

and finally we have

$$\Xi = \frac{N! 2^{2N-2N_1-2N_4} (N_4 + N_5 + N_6)!}{N_1! (2N - 2N_1 - 3N_4 - 2N_5 - N_6)! (N_1 + 2N_4 + N_5 - N)! N_4! N_5! N_6!}. \quad (105)$$

Minimization of the free energy yields the following set of equations:

$$\frac{4N_1 (N_1 + 2N_4 + N_5 - N)}{(2N - 2N_1 - 3N_4 - 2N_5 - N_6)^2} = 1 \quad (106)$$

$$\frac{4N_4 (N_1 + 2N_4 + N_5 - N)^2}{(2N - 2N_1 - 3N_4 - 2N_5 - N_6)^3 (N_4 + N_5 + N_6)} = \frac{K_1}{V} \quad (107)$$

$$\frac{N_5 (N_1 + 2N_4 + N_5 - N)}{(2N - 2N_1 - 3N_4 - 2N_5 - N_6)^2 (N_4 + N_5 + N_6)} = \frac{K_2}{V} \quad (108)$$

$$\frac{N_6}{(2N - 2N_1 - 3N_4 - 2N_5 - N_6) (N_4 + N_5 + N_6)} = \frac{K_3}{V}. \quad (109)$$

For the concentrations of aggregates, we may write

$$C_{ijk} = \alpha_{ijk} K_1^{i-1-j-k} K_2^j K_3^k c_1^i, \quad (110)$$

where  $0 \leq j \leq i-2$  and there are two series of  $k$  values that satisfy  $2k+j+2=i$  for  $j$  and  $i$  both odd or even and  $2k+j+3=i$  otherwise. Based on these relations, one index may be removed and two series of concentrations introduced instead:

$$C_{1ik} = \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i, \quad 0 \leq k \leq \infty, 2k+2 \leq i \leq \infty \quad (111)$$

$$C_{2ik} = \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i, \quad 0 \leq k \leq \infty, 2k+3 \leq i \leq \infty. \quad (112)$$

Therefore, we can write

$$c = c_1 + \sum_{k=0}^{\infty} \sum_{i=2k+2}^{\infty} i \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i + \sum_{k=0}^{\infty} \sum_{i=2k+3}^{\infty} i \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i \quad (113)$$

$$n_1 = c_1 \quad (114)$$

$$n_4 = \sum_{k=0}^{\infty} \sum_{i=2k+2}^{\infty} (k+1) \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i + \sum_{k=0}^{\infty} \sum_{i=2k+3}^{\infty} (k+2) \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i \quad (115)$$

$$\begin{aligned} n_5 = & \sum_{k=0}^{\infty} \sum_{i=2k+2}^{\infty} (i-2k-2) \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i \\ & + \sum_{k=0}^{\infty} \sum_{i=2k+3}^{\infty} (i-2k-3) \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i \end{aligned} \quad (116)$$

$$n_6 = \sum_{k=0}^{\infty} \sum_{i=2k+2}^{\infty} k \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i + \sum_{k=0}^{\infty} \sum_{i=2k+3}^{\infty} k \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i \quad (117)$$

$$\alpha_{1ik} = 4^{i-1-k} \frac{(i-2)!}{(i-2-2k)! k! (k+1)!} \quad (118)$$

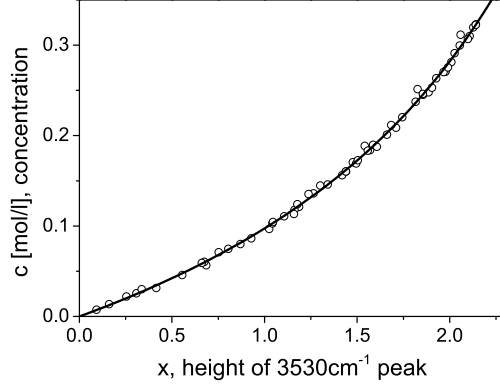


Figure S43: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 7m.

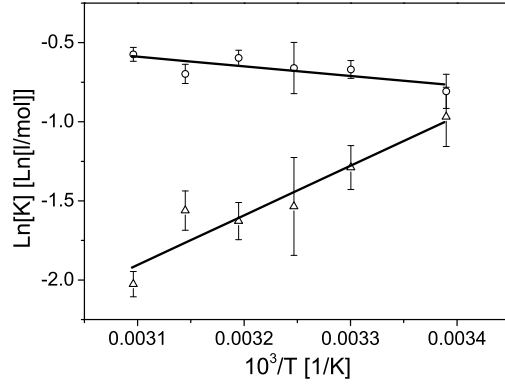


Figure S44: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 7m.

$$\alpha_{2ik} = 4^{i-2-k} \frac{(i-2)!(2k+2)}{(i-3-2k)!(k+1)!(k+2)!} \quad (119)$$

By summing the series, assuming that  $K_2 = K_3$ , it can be shown that

$$c = \frac{1 - 4c_1K_2}{8c_1K_2^2\sqrt{1 - 8c_1K_2 - 16c_1^2(K_1 - K_2)K_2}} + \frac{8c_1^2K_2(K_2 - K_1) - 1}{8c_1K_2^2}. \quad (120)$$

Finally, the number of free groups, which is the sum of  $c_1$  and  $n_4$ , can be calculated as

$$n_f = c_1 - \frac{c_1K_1}{K_2} + \frac{c_1K_1}{K_2\sqrt{1 - 8c_1K_2 - 16c_1^2(K_1 - K_2)K_2}}. \quad (121)$$

### 1.8.1 Model 7m

Here, it is assumed that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the  $\text{NH}_2$  group in free molecules. In this case, we put  $c_1 = Ax$  and substitute it into Equation 120.

The fitting results are shown in Figures S43 and S44. The quality of fit in Figure S43 can be characterized by  $\text{AICc} = -501.7$ , and the estimates of the model parameters for the fits in Figure S44 are  $\ln C_1 = -11.6 \pm 2 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -6.2 \pm 1 \text{ kcal/mol}$ ,  $\ln C_2 = 1.3 \pm 0.8 \ln[\text{l/mol}]$ , and  $\epsilon_2 = 1.2 \pm 0.5 \text{ kcal/mol}$ , with  $r_1^2 = 0.9$  and  $r_2^2 = 0.62$ .

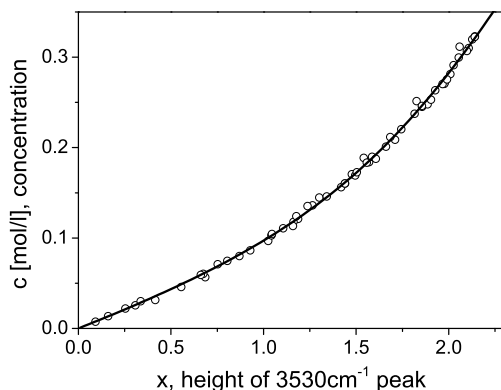


Figure S45: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 7g.

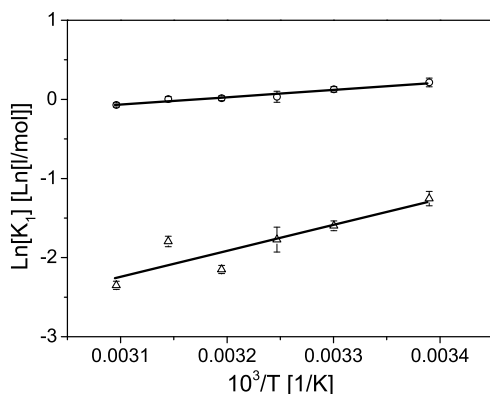


Figure S46: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 7g.

### 1.8.2 Model 7g

Next, it is assumed that the  $3530\text{cm}^{-1}$  peak corresponds to absorption by free groups, so that  $n_f = Ax$ .

The results of the fitting procedure are shown in Figures S45 and S46. The value of AICc for the fit in Figure S45 is  $-497.4$ . The estimates of the model parameters for the fits in Figure S46 are  $\ln C_1 = -12.4 \pm 3 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -6.5 \pm 1.6 \text{ kcal/mol}$ ,  $\ln C_2 = -2.9 \pm 0.3 \ln[\text{l/mol}]$ , and  $\epsilon_2 = -1.8 \pm 0.2 \text{ kcal/mol}$ , with  $r_1^2 = 0.80$  and  $r_2^2 = 0.95$ .

### 1.8.3 Model 7s

In the case of the assumption  $Ax = c_1 + 4K_1c_1^2 + 4K_1^2c_1^3$  ( $\text{NH}_2$  groups in unimers, dimers and trimers without  $\epsilon_2$ -bonds) the fitting results shown in Figures S47 and S48 are found. The value of the information criterion for the fit in Figure S47 is  $\text{AICc} = -501.7$ , and the estimates of the model parameters given by the fits in Figure S48 are  $\ln C_1 = -14.4 \pm 2 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -8.2 \pm 1.4 \text{ kcal/mol}$ ,  $\ln C_2 = -1.5 \pm 0.3 \ln[\text{l/mol}]$ , and  $\epsilon_2 = -0.6 \pm 0.2 \text{ kcal/mol}$ , with  $r_1^2 = 0.9$  and  $r_2^2 = 0.69$ .

## 1.9 Model 8

In model 8, we allow only one bond per oxygen and one bond per  $\text{NH}_2$  group, so all aggregates are supposed to be linear. We also assume that the association equilibrium is described by two association constants and there are no cyclic dimers. We suppose that the bond has energy  $\epsilon_1$  if the acceptor of donor molecule is free and  $\epsilon_2$  otherwise (see Figure S49). It is worth mentioning that this model does not change if we assume that the energy of the bond is defined by the bonding

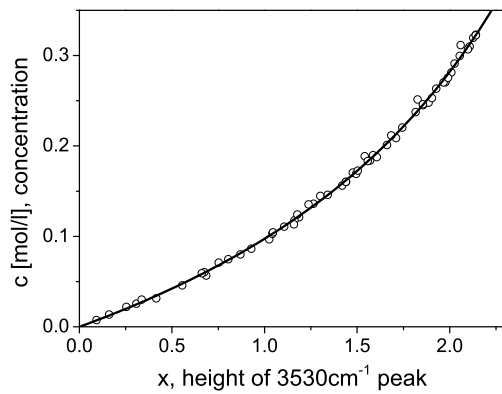


Figure S47: Fit of the dependence of the total concentration on the height of the 3530cm<sup>-1</sup> peak at  $T = 22^\circ\text{C}$  with model 7s.

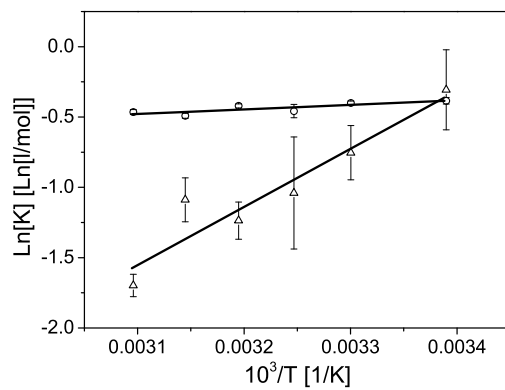


Figure S48: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 7s.

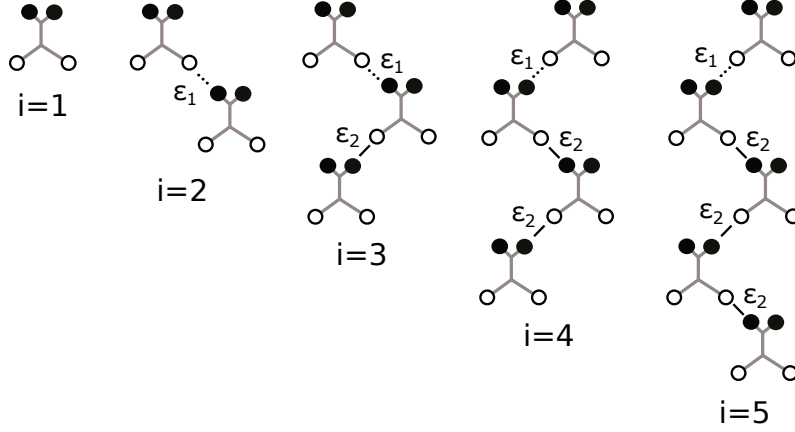


Figure S49: Schematic representation of aggregates with size up to  $i = 5$  in model 8. Only configurations corresponding to different energies are shown.

state of the donor group in the acceptor molecule. The number of free molecules will be denoted by  $N_0$ , the number of molecules with one bonded hydrogen and free oxygen by  $N_1$ , the number of molecules with one bonded hydrogen and one bond per oxygen by  $N_2$ , and the number of molecules with one bond per oxygen and free hydrogens by  $N_3$ . Then, the number of ways to form bonds is

$$\Xi = \frac{N!2^{N_3+N_2}}{(N - N_3 - N_2)!} \frac{(N_0 + N_1)!2^{N_1}}{N_1!N_0!} \frac{(N_3 + N_2)!2^{N_2}}{N_2!N_3!}, \quad (122)$$

where the first factor is the number of ways to choose acceptors, the next factor is the number of ways to choose donors for bonds with energy  $\epsilon_1$ , and the following factor is the number of ways to choose bonds with energy  $\epsilon_2$ .

Taking into account that  $N_3 = N_1$  and  $N_0 = N - 2N_1 - N_2$ , we have

$$\Xi = \frac{N!(N_1 + N_2)!2^{2N_1+2N_2}}{(N - 2N_1 - N_2)!(N_1!)^2 N_2!}. \quad (123)$$

Minimization of the free energy gives

$$\frac{N_1^2}{4(N_1 + N_2)(N - 2N_1 - N_2)^2} = \frac{K_1}{V} \quad (124)$$

$$\frac{N_2}{4(N_1 + N_2)(N - 2N_1 - N_2)} = \frac{K_2}{V}. \quad (125)$$

In this model it is straightforward to guess that

$$c_i = c_1^i 4^{i-1} K_1 K_2^{i-2}. \quad (126)$$

Then, the total concentration can be calculated as

$$c = \sum_{i=1}^{\infty} i c_i = c_1 + \frac{8c_1^2 K_1 (1 - 2c_1 K_2)}{(1 - 4c_1 K_2)^2}. \quad (127)$$

For the concentration of the free groups,  $n_f = c - n_1 - n_2$ , we have

$$n_f = c_1 + \frac{4K_1 c_1^2}{1 - 4K_2 c_1}. \quad (128)$$

### 1.9.1 Model 8m

Let us assume first that the  $3530\text{cm}^{-1}$  peak corresponds to the out-of-phase vibrations of the  $\text{NH}_2$  group in free molecules. In this case, we put  $c_1 = Ax$  and substitute it into Equation 127.

The fitting results are shown in Figures S50 and S51. The quality of fit in Figure S50 can be characterized by  $\text{AICc} = -501.5$ , and the estimates of the model parameters yielded by the fits in Figure S51 are  $\ln C_1 = -12.3 \pm 2 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -6.5 \pm 1 \text{ kcal/mol}$ ,  $\ln C_2 = 0.1 \pm 0.6 \ln[\text{l/mol}]$ , and  $\epsilon_2 = 0.3 \pm 0.5 \text{ kcal/mol}$ , with  $r_1^2 = 0.89$  and  $r_2^2 = 0.18$ .

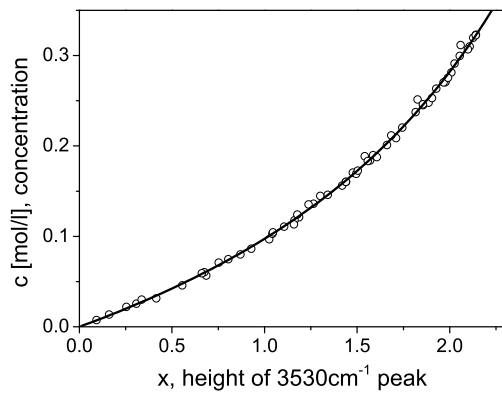


Figure S50: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 8m.

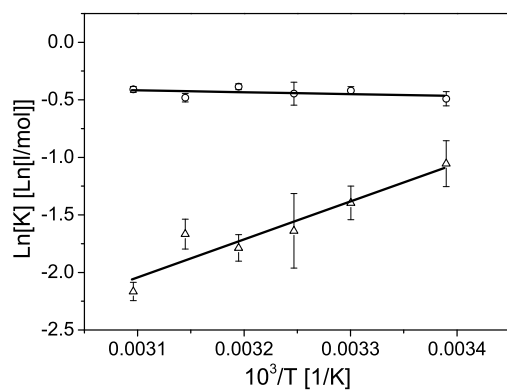


Figure S51: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 8m.

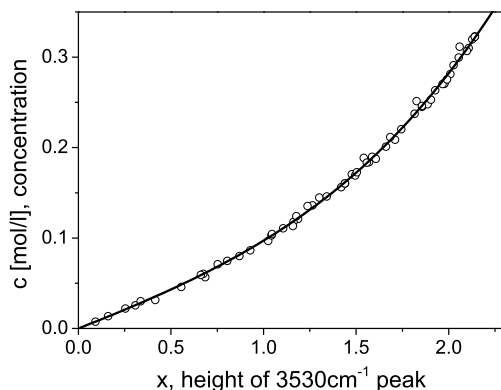


Figure S52: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 8g.

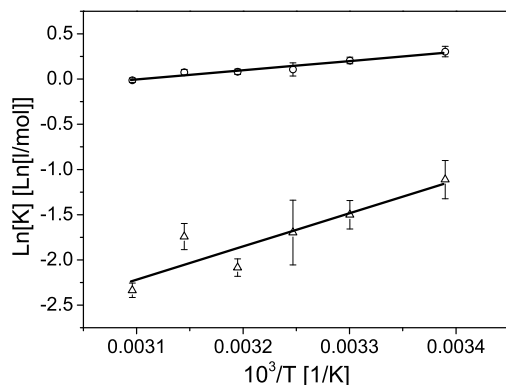


Figure S53: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 8g.

### 1.9.2 Model 8g

Next, it is assumed that the  $3530\text{cm}^{-1}$  peak corresponds to absorption by free groups, so that  $n_f = Ax$ .

The fitting results are shown in Figures S52 and S53. The value of the AICc parameter characterizing the quality of fit in Figure S52 is  $-499.5$ , and the estimates of the model parameters resulting from the fits in Figure S53 are  $\ln C_1 = -13.6 \pm 3 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -7.3 \pm 1.6 \text{ kcal/mol}$ ,  $\ln C_2 = -3.2 \pm 0.4 \ln[\text{l/mol}]$ , and  $\epsilon_2 = -2.0 \pm 0.2 \text{ kcal/mol}$ , with  $r_1^2 = 0.83$  and  $r_2^2 = 0.95$ .

### 1.9.3 Model 8s

In the case of the assumption  $Ax = c_1 + 4K_1c_1^2$  ( $\text{NH}_2$  groups in unimers and dimers without  $\epsilon_2$ -bonds), the fitting results shown in Figures S54 and S55 are found. The quality of fit in Figure S54 can be characterized by  $\text{AICc} = -501.4$ . The estimates of the model parameters given by the fits in Figure S55 are  $\ln C_1 = -14.2 \pm 2 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -7.9 \pm 1.4 \text{ kcal/mol}$ ,  $\ln C_2 = -2.4 \pm 0.3 \ln[\text{l/mol}]$ , and  $\epsilon_2 = -1.3 \pm 0.2 \text{ kcal/mol}$ , with  $r_1^2 = 0.88$  and  $r_2^2 = 0.93$ .

### 1.10 Model 9

In this model, we allow two bonds per oxygen and only one bond per  $\text{NH}_2$  group. We assume that the energy of the bond is determined by the bonding state of the oxygen. If the oxygen forms only one bond, then its energy is  $\epsilon_1$ , and if it forms two then its energy is  $\epsilon_2$ . This means that this model is analogous to model 5 with the roles of the acceptor and donor group exchanged. Therefore, all equations are the same, apart from the value of the number of free  $\text{NH}_2$  groups. It is clear that in



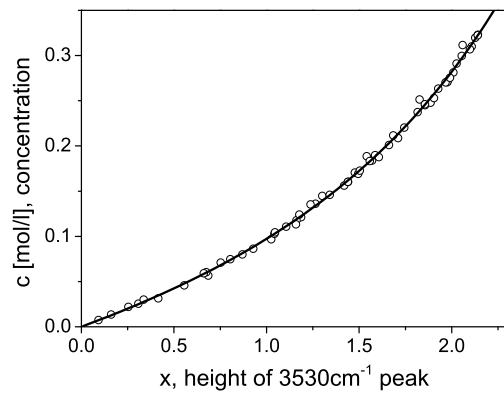


Figure S54: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 8s.

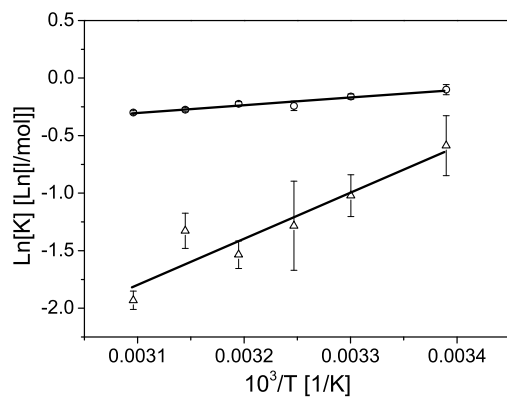


Figure S55: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 8s.

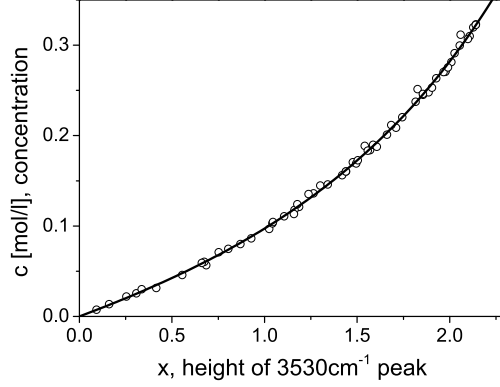


Figure S56: Fit of the dependence of the total concentration on the height of the  $3530\text{cm}^{-1}$  peak at  $T = 22^\circ\text{C}$  with model 10g.

case of model 9 there is only one free group per each aggregate, so we have

$$n_f = \frac{1 - 4c_1K_1 - \sqrt{1 - 8c_1K_1 + 16c_1^2K_1^2 - 16c_1^2K_2^2}}{8c_1K_2^2}. \quad (129)$$

We do not repeat the fitting for models 9m and 9s as these cases are equivalent to models 5m and 5s.

### 1.10.1 Model 9g

We assume that the  $3530\text{cm}^{-1}$  peak corresponds to absorption by free groups, so that  $n_f = Ax$ . The fitting results are not shown, because at higher temperatures they give a negative value of  $K_1$ .

## 1.11 Model 10

In this model, we allow two bonds per oxygen and only one bond per  $\text{NH}_2$  group. We assume that the energy of the bond is determined by the bonding state of the  $\text{NH}_2$  group in the acceptor molecule. If the hydrogen in the acceptor molecule is free, then the bond energy is  $\epsilon_1$ ; otherwise, it is  $\epsilon_2$ .

Therefore, this model is analogous to model 6 with the roles of the acceptor and donor groups exchanged, and all equations are the same, the only difference being the value of the number of free  $\text{NH}_2$  groups. It is clear that, in the case of model 9, there is only one free group per each aggregate, so we have

$$n_f = c_1 + \frac{K_1(1 - 4K_2c_1 + \sqrt{1 - 8c_1K_2})}{2K_2^2} + \frac{K_2^2(1 - 8c_1K_2 + 8c_1^2K_2^2 - \sqrt{1 - 8c_1K_2} + 4c_1K_2\sqrt{1 - 8c_1K_2})}{8c_1K_2^4}. \quad (130)$$

### 1.11.1 Model 10g

Here, we assume that the  $3530\text{cm}^{-1}$  peak corresponds to absorption by free groups, so that  $n_f = Ax$ .

The results of the fitting procedure are shown in Figures S56 and S57. The quality of fit in Figure S56 can be characterized by  $\text{AICc} = -501.1$ . The estimates of the model parameters for the fits in Figure S57 are  $\ln C_1 = -11.3 \pm 2 \ln[\text{l/mol}]$ ,  $\epsilon_1 = -6.3 \pm 1.0 \text{ kcal/mol}$ ,  $\ln C_2 = -4.2 \pm 0.5 \ln[\text{l/mol}]$ , and  $\epsilon_2 = -2.4 \pm 0.3 \text{ kcal/mol}$ , with  $r_1^2 = 0.90$  and  $r_2^2 = 0.95$ .

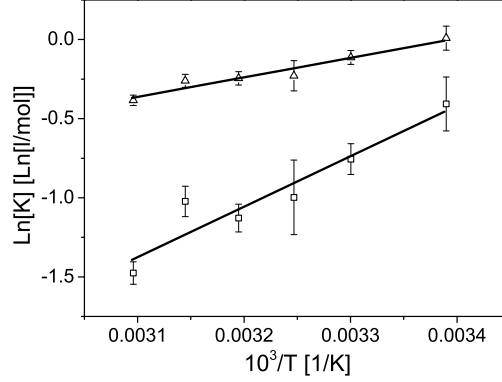


Figure S57: Dependence of  $\ln K_1$  (triangles) and  $\ln K_2$  (circles) on  $1/T$  for model 10g.

## 1.12 Model 11

In this model we allow two bonds per oxygen and only one bond per  $\text{NH}_2$  group. We assume that the energy of the bond is determined by the bonding state of the acceptor in the donor molecules. If the oxygen in the donor molecule is free then the bond energy is  $\epsilon_1$ ; otherwise, it is  $\epsilon_2$ .

Therefore, this model is analogous to model 7 with the roles of the acceptor and donor group exchanged. All equations are the same, except for the value of the number of free  $\text{NH}_2$  groups. Again, as in the two previous models, there is only one free group per aggregate

$$n_f = \frac{1 - 4K_2c_1 - 8K_2c_1^2(K_2 - K_1) - \sqrt{1 - 8K_2c_1 - 16K_2(K_1 - K_2)c_1^2}}{8K_2^2c_1}. \quad (131)$$

### 1.12.1 Model 11g

Here we assume that the  $3530\text{cm}^{-1}$  peak corresponds to absorption by free groups, so that  $n_f = Ax$ . However, in this case, the fits do not converge.

## 2 Selection of the best model

To compare the quality of fit of different models we use Akaike's information criterion<sup>2,3</sup>. It states that the best model is that with the smallest value of

$$AIC = 2k - 2 \ln(L(A, K_1, K_2, \sigma)), \quad (132)$$

where  $L(A, K_1, K_2, \sigma)$  is a likelihood function and  $k$  is the number of model parameters (4 in two-parameter models and 3 in one-parameter models because  $\sigma$  is also included in the context of the likelihood function). The likelihood function for the model with parameters obtained by minimization of the sum of squared deviations is usually written in the form

$$L(A, K_1, K_2, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_i - \text{model}(x_i, A, K_1, K_2))^2}{2\sigma^2} \right\}. \quad (133)$$

In this case, minimization of the sum of squared deviations is equivalent to maximization of the logarithm of the likelihood function. Then, for the AIC value we have

$$\begin{aligned} AIC &= 2k + n \ln(2\pi) + n \ln(\hat{\sigma}^2) + \frac{(y_i - \text{model}(x_i))^2}{2\hat{\sigma}^2} \\ &= 2k + n \ln(2\pi) + n \ln \frac{\text{RSS}}{n} + n, \end{aligned} \quad (134)$$

where the estimate for  $\hat{\sigma}$  is  $\hat{\sigma}^2 = \text{RSS}/n$  and the sum of the squares of the residuals is  $\text{RSS} = \sum_{i=1}^n (y_i - \text{model}(x_i))^2$ .

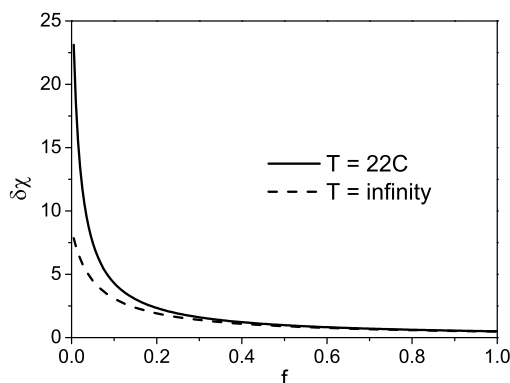


Figure S58: Dependence of  $\delta\chi$  on  $f$  for model 1m at  $T = 22^\circ\text{C}$  and in the limit of infinite temperature.

For a finite sample size there exists the following correction:

$$\text{AICc} = \text{AIC} + \frac{2(k+1)(k+2)}{n-k-2}, \quad (135)$$

which is useful when we compare models with different number of parameters.

The relative probabilities of the two models with values of the information criterion given by  $\text{AIC}_1$  and  $\text{AIC}_2$  respectively can be estimated as

$$\exp\left(\frac{\text{AIC}_1 - \text{AIC}_2}{2}\right) \quad (136)$$

So, if  $\text{AIC}_1 - \text{AIC}_2 = -501.7 + 493.9 = -7.8$  then model with value  $\text{AIC}_1$  is 49.4 times more probable than the model with value  $\text{AIC}_2$ .

### 3 $\delta\chi_{HB}(f)$ dependences for a block copolymer with polyacrylamide in random mixing approximation

Let us do calculations of  $\delta\chi_{HB}$  for a diblock copolymer with a polyacrylamide block based on our models of hydrogen bonding association and the values of association constants deduced from fitting IR data.

$$\delta\chi_{HB}(f) = -\frac{1}{2} \frac{d^2 f_{HB}(f)}{d\phi_B^2} \quad (137)$$

First, the dependence of  $\delta\chi$  on  $f$  for models with one association constant is considered. The graph for model 1m is shown in Figure S58. For other models, the plots appear qualitatively the same, so they are not shown here. The largest absolute value of  $\delta\chi_{HB}$  for one-constant models is attained in the limit as  $f \rightarrow 0$  and  $\delta\chi$  then monotonically decreases as the fraction of hydrogen-bonding block increases. The decrease of  $\delta\chi$  with  $f$  is intuitively expected, since, if more neutral segments are mixed in with the network of hydrogen-bonded segments, then more hydrogen bonds need to be broken, and more energy needs to be spent in doing so. It can also be seen that changing the temperature leads to a decrease of the maximal value of  $\delta\chi$  at  $f \rightarrow 0$ , although this value is still very high even at infinite temperature. It is also interesting to note that at  $f > 0.4$  there is no change in  $\delta\chi$  with temperature and that  $\delta\chi$  depends only on the volume fraction  $f$ .

In the case of models with two association constants, the behavior at small volume fractions is qualitatively different. The dependence of  $\delta\chi$  on the volume fraction of the hydrogen-bonding block calculated for model 6s at different temperatures is shown in Figure S59. As the temperature is increased, a peak at a finite value of  $f$  appears, which moves to the right as the temperature grows further. However, the behavior of  $\delta\chi$  at  $f > 0.4$  is similar to the one-parameter models: there is a gradual decrease of  $\delta\chi$  as the volume fraction of the hydrogen bonding block is increased and little difference between the curves calculated for different temperatures. It is also interesting

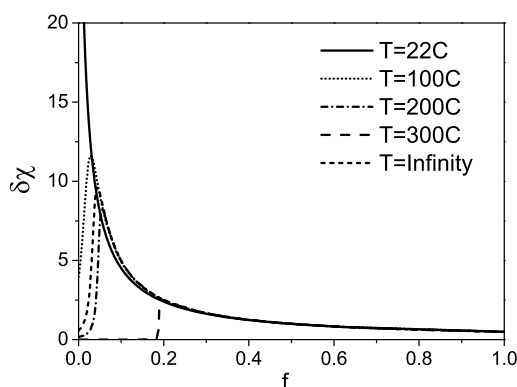


Figure S59: Dependence of  $\delta\chi$  on  $f$  for model 6s at different temperatures.

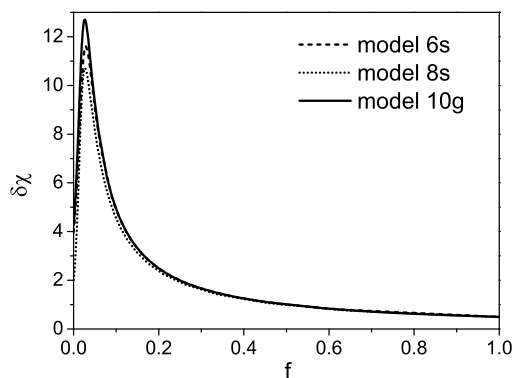


Figure S60: Dependence of  $\delta\chi$  on  $f$  for all good models at  $T = 100^\circ\text{C}$ .

to note that the values of  $\delta\chi$  for all good two-parameter models are close to each other not only qualitatively, but also quantitatively (see Figure S60).

It can be seen that our current predictions for  $\delta\chi$  of polyacrylamide are unrealistically high. This is especially true for small values of  $f$ . However, these small values of  $f$  are in fact never reached in polymer systems due to the non-randomness of mixing that we discussed above. In addition, it is important to re-emphasize this analysis as just an initial step on the way to the application of the association model approach to describe hydrogen bonding interactions in block copolymers.

## References

- [1] The On-Line Encyclopedia of Integer Sequences, published electronically at <https://oeis.org>, May 2018.
- [2] Akaike, H. In *Selected Papers of Hirotugu Akaike*; Parzen, E., Tanabe, K., Kitagawa, G., Eds.; Springer New York: New York, NY, 1998; pp 199–213.
- [3] Sugiura, N. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Stat. Theory Methods* **1978**, *7*, 13–26.