June 2010 ISSN : 0974 - 7478 Volume 6 Issue 1

Trade Science Inc.

Macromolecules

An Indian Journal

Full Paper

MMAIJ, 6(1), 2010 [30-34]

# Calculation of equilibrium intrinsic binding constant and cooperativity coefficient of polynucleotide poly(I)–poly(C) protonation

**A.G.Kudrev**

**St.Petersburg State University, 198504, Universitetskii pr. 26, St. Petersburg, (RUSSIA)**

**E-mail : kudrevandrei@mail.ru**

*Received: 9th March, 2010 ; Accepted: 19th March, 2010*

## ABSTRACT

In the present work the pH-induced formation of double stranded poly(inosinic)–poly(cytidylic) acid has been studied by means of the hard modelling method based on the results of multivariate data analysis soft modelling method. The equilibrium model, describing species distribution changes as functions of solution acidity has been proposed. Proposed approach has been used to determine intrinsic protonation constants of poly(I-)–poly(C) polynucleotide system in aqueous solutions such as cooperativity coefficient ($\omega$ = 6.0) and a proton binding constant ($lgK_g$ = 8.92). 

## KEYWORDS

Binding constants;
Poly(I)-poly(C);
Multivariate curve resolution.

## INTRODUCTION

Extraction of quantitative information for solution equilibria systems with help of chemometric methods based on Factor Analysis (FA), is very attractive because they are free from mass-action law constraints and do not require an initial postulation of a chemical model. One of the first works in this area were devoted to the application of Evolving Factor Analysis (EFA)[1,2]. It has been demonstrated for several systems how to build up abstract concentration profiles of species at equilibrium and how to calculate equilibrium constants directly from the concentration profiles obtained by EFA, after assigning the proper stoichiometric coefficients to the individual species by means of chemical reasoning[1].

Subsequent systematic studies have revealed that the Alternating Least Squares (ALS) Multivariate Curve Resolution (MCR) method[3-7], based on FA[8] can be easily adapted to the analysis of spectrometric data of equilibria systems. In these works the emphasis was mainly devoted to show the scope of the recovery without ambiguities of the concentration and spectra profiles of the spectrally active species in solution.

To determine equilibrium parameters such as cooperativity coefficient and a proton binding constant to a polymer a hard model should be used that require mass-action law constraints and an initial postulation of a chemical model.

In the present work the combination of MCR which is soft modeling method and Mc-Ghee-von Hippel (GH) hard model for study of equilibria of biopolymer-ligand interaction in solutions is considered. The tandem of soft and hard modelling for study of equilibria with a polymers is performed in the same way as in previously described calculations of equilibrium constants and stoichiometric coefficients for ordinary nonpolymeric equilibria systems[9,10]. Proposed MCR-GH method has been applied for evaluation of protonation constants of deprotonated poly(inosinic)–poly(cytidylic) acid, {poly(I-)-poly(C)} in aqueous solution.

## Computational details

The mathematical decomposition of spectrophotometrically measured matrices of absorption $A_i$ into concentration and spectra matrices, $C_i^{calc}$ and $S_i^{T\,calc}$ is carried out using the singular value decomposition (SVD) method[11]. The SVD gives a mathematical abstract decomposition of the original data matrix $A_i$ for the selected number of contributions or components which best reproduce the original data matrix within a certain noise level.

$$\mathbf{A_i = U_i \cdot S \cdot V_i^T + \tau_i = A_i^* + \tau_i} \quad (1)$$

where diagonal matrix S, of the same dimension as A and with nonnegative diagonal elements in decreasing order, $U_i$ and $V_i^T$ are unitary matrices[11] and $\tau_i$ is the residual error matrix containing the rest of variance not explained by $U_i$ and $V_i^T$. $A_i^*$ is the reproduced data matrix using the selected number of components. Some of the noise present in the original data matrices $A_i$ (that which is not correlated with the estimated components in score and loading matrices) is removed in the reproduced data matrices $A_i^*$. Score ($U_i$) and loading ($S \cdot V_i^T$) matrices obtained by eq. (1) are usually very different to those really responsible of the data variance in matrices $A_i$. In fact, there are an infinite number of possible matrix decomposition using eq. (1) and there is not an unique solution. There are rotational and intensity ambiguities in a matrix decomposition such as that described in eq. (1). Principal Component Analysis (PCA) finds that solutions which explain maximum variance and are orthonormal. Under these constraints the solution found by PCA is unique. However, this solution is rarely the true one since these two conditions are not fulfilled by most of the physical systems, such as the chemical equilibrium systems under study in the present work. The PCA solutions have not a direct physical meaning; they are only pure mathematical solutions. As the interest of the data analysis in this work is the obtaining of the set of profiles which really caused the observed data variance of the data matrices $A_i$ the PCA reproduced data matrices $A_i^*$ is further decomposed as:

$$\mathbf{A_i^* = C_i^{calc} \cdot S_i^{Tcalc}} \quad (2)$$

This eq. (2) is solved iteratively by an ALS. In the present study the iterative optimization of the estimations of $C_i^{calc}$ and $S_i^{Tcalc}$ matrices is performed in two steps. An initial estimation of the concentration matrix is available, $C^{inic}$ from EFA, the corresponding spectra matrix is estimated by Least Squares (LS)[12]. After initial estimation of the species spectra matrix we can solved equation for $C_i^{calc}$. Since $A_i^*$ are noise filtered for the particular number of principal components, the iterative calculations are more stable than using the original $A_i$ matrices. Eq. (2) iteratively solved under the constraints of nonnegativity, unimodality and closure as described in[3-7,12]. The set of possible solutions is highly restricted in this way. However, the correct resolution of the system can only be achieved without ambiguities (within noise level) when selectivity or local rank conditions[7,13] are present. The application of the ALS-MCR method to each data matrix $A_i$ gives a new set of matrices $S_i^{Tcalc}$ and $C_i^{calc}$. If ambiguities were totally solved, the set of $S_i^{Tcalc}$ and $C_i^{calc}$ matrices calculated in this way, would be a set of normally distributed estimation matrices with mathematical expectations $S^{T\,theor}$ and $C^{theor}$. Once convergence of the iterative ALS optimization is achieved, data fitting can be tested using the expression for the percentage of data lack of fit:

$$\mathbf{lof} = \left[\frac{\mathbf{Trace}[(\mathbf{A^{ALS} - A}) \cdot (\mathbf{A^{ALS} - A})^T]}{\mathbf{Trace}[\mathbf{A \cdot A^T}]}\right]^{1/2} \cdot 100\% \quad (3)$$

where A and $A^{ALS}$ are respectively the input and reproduced data by the ALS optimization, Trace is the sum of diagonal elements of a matrices. When convergence is achieve data fitting should give lof values close to the noise levels.

For the recovering of the theoretical model govern the matrix of concentration profiles we have to fit $C_i^{calc}$ with matrices of species fractions were formed by polymer and a ligand X. For an equilibria with a polymeric component the appropriate model function can be obtained using statistical or matrix methods, as it has been shown previously[16-22]. In the present study McGhee-von Hippel conditional probabilities model[16] was used to calculate the fractions of infinite polymer species. In accordance with GH model the ligand-ligand interaction is only allowed between nearest neighbours, bound without intervening free lattice residues. This restriction results in three distinguishable types of ligand binding sites and consequently four spectrally different species of monomers. To isolated site (1) a ligand binds with an intrinsic association constant $K_g$. To singly

contiguous site (2) ligand binds with constant $K_g\omega$. And to double contiguous site (3) the binding constant is $K_g\omega^2$. The equilibrium constant of moving ligand from an isolated site to singly contiguous site and from singly contiguous site to double contiguous site is so called cooperativity parameter $\omega$.

As it was shown in[22] the matrix of fractions of spectroscopically distinguishable forms of monomer in polymer may be expressed as

$$\mathbf{C_{mon}=[C_{mon}(0); C_{mon}(1); C_{mon}(2); C_{mon}(3)]} \quad (4)$$

then, under the above conditions, the components of this matrix $C_{mon}$ can be find with help of the method of conditional probabilities.

$$\mathbf{C_{mon}(0) = 1 - m\bar{n}_g = [Mon]} \quad (5)$$

$$\mathbf{C_{mon}(1) = \left\{\frac{(b_m f)(ff)^{m+1}}{(fb_1)}\right\} m\bar{n}_g K_g [X]} \quad (6)$$

$$\mathbf{C_{mon}(2) = \left\{\frac{2(fb_1)}{(ff)}\right\} C_f(1)\omega} \quad (7)$$

$$\mathbf{C_{mon}(3) = \left\{\frac{(fb_1)}{2(ff)}\right\} C_f(2)\omega} \quad (8)$$

Here[X], [Mon]- equilibrium concentration of free ligand and monomer; the conditional probabilities are determined by the following expressions.

$$\mathbf{(b_m f) = \frac{(m-1)\bar{n}_g - 1 + \Phi}{2\bar{n}_g(\omega - 1)}};$$

$$\mathbf{(fb_1) = \frac{(m-1)\bar{n}_g - 1 + \Phi}{2\bar{n}_g(\omega - 1)[Mon]}};$$

$$\mathbf{(ff) = \frac{(2\omega - 1)[Mon] + \bar{n}_g - \Phi}{2(\omega - 1)[Mon]}};$$

$$\mathbf{\Phi = \sqrt{(1-(m+1)\bar{n}_g)^2 + 4\omega\bar{n}_g[Mon]}}.$$

Using the common data of potentiometry and spectrofotometry, it is possible to find in each experimental point j of titration curve the value of apparent stability constant of complex monomer-ligand which is convenient for fitting procedure

$$\mathbf{\lg K = \lg(\bar{n}_g/(1-\bar{n}_g)) - \lg[X]} \quad (9)$$

where degree of saturation of monomers can be expressed by: $\bar{n}_g = \frac{\sum_{p=1}^{3} C_{mon}(p)}{m}$; summation runs over fractions of all monomer species connected with a ligand, m-average numbers of monomers occupied by a ligand.

## The procedure of fitting

In the present study the intrinsic binding parameters are estimated by ordinary LS. Standard Marquardt algorithm[23] is used to find minimum of objective function M. Function M in matrix notation is

$$\mathbf{M = \sqrt{\frac{Trace[(\lg K_{exp} - \lg K_{calc}) \times (\lg K_{exp} - \lg K_{calc})^T]}{Trace[(\lg K_{exp} \times \lg K_{exp})^T]}}} \quad (10)$$

here $\lg K_{exp}$-experimental data, $\lg K_{calc}$-calculated in accordance with eq. (9). The assumption of equal population variances were checked using standard F-test[24].

The reliability of the proposed MCR-GH method for simulated numerical examples was shown previously[25].

## RESULTS AND DISCUSSION

Calculations of protonation constants of poly(I-)-poly(C) from experimental data. Data of spectrofotometric study of acid-base titration of poly(I)-poly(C) system, published in[26] has been used as an experimental example. Concentration profiles $C^{calc}$ was calculated with help of ALS-MCR from measured changes of spectrum of poly(I-)-poly(C) aqueous solution as function of pH. In the present work these profiles has been analysed with the aim to calculate intrinsic binding constant and parameter of cooperativity of formation of poly(I)-poly(C)-complex. Plot of $\lg K_{exp}$ as function of pH calculated in accordance with eq. (9) is shown on figure 1. It can be seen, that apparent constant correlated with the degree of protonation of polymer. As was shown in[26] changes UV and CD spectra can be explained by acid-base equilibrium involving hypoxantine nitrogen base i.e. formation of complex Mon ($I\text{-}H^+$). The parameters of McGhee & von Hippel model calculated from full set of experimental data by means of OLS are given in TABLE. But parameter m for this formal model has

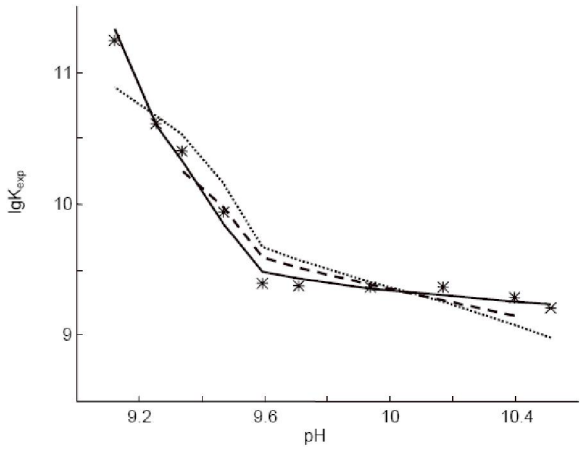


**Figure 1 : Experimental apparent binding constant as function of pH of aqueous solution of poly(I)-poly(C). The stars corresponds to the experimental data (calculated with help of ALS-MCR ); Solid line is obtained from fitting according GH model ( $lgK_{gh}$ = 9.21; ω = 1.44; *m* = 0.65 ) Dotted line is obtained from fitting according GH model ($lgK_{gh}$ = 8.66; ω =13.7 ; $m_{fixed}$ = 1 ); Dashed line obtained as least noise sensitive part of $lgK_{exp}$ = *f* (pH), allocated using F-test and fitting according GH model ( $lgK_{gh}$ = 8.92; ω = 6.0; $m_{fixed}$ =1 )**

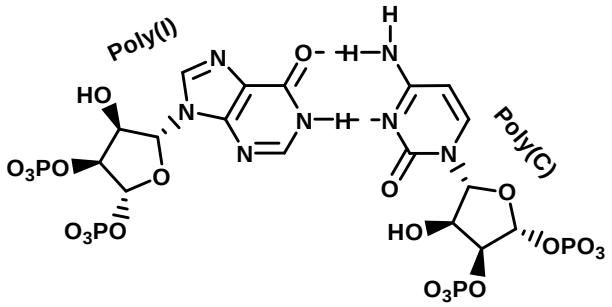


**Figure 2 : Nitrogenated base pairing in poly(I)-poly(C)**

**TABLE 1 : Intrinsic protonation parameters of poly(I-) and poly(I-)-poly(C)**

| Polymer | pH | Model | $lgK_g$* | ω | m | PE (%) | ref. |
|---|---|---|---|---|---|---|---|
| Poly(I-) | 6÷11 | - | 9.05 | 0.63 | 1 | 1.38 | [26,27] |
| | 7.3÷11 | + | 8.99 | 0.79 | 1 | 0.28 | [26,27] |
| Poly(I-)-Poly(C) | 9.1÷10.5 | - | 9.21 | 1.44 | 0.61 | 0.6 | * |
| | 9.1÷10.5 | - | 8.66 | 13.7 | 1(f) | 2.1 | * |
| | 9.2÷10.2 | + | 8.92 | 6.0 | 1(f) | 0.62 | * |

***present work, f - the parameter is fixed, (+)- exclusion of noise sensitive interval of pH**

unsuitable physical meaning. Most probably that ligand ($H^+$) interacts with single monomeric unit. In this case parameter m should be equal to 1. With using of this constraint as it seen from figure 1 theoretical function significantly disagree with experimental curve mostly at low pH. Allocation of pH interval of equal population variances leads to new parameters of protonation in this region. As can be seen from figure 1 difference of experimental and designed functions $lgK_{exp}$=f(pH) less for new parameters. The value of lack of fit between simulated and experimental curves is rather small, therefore we can conclude, that model of protonation of poly(I-)-poly(C) in aqueous solution from formal statistical point of view has been adequately recognised in determined pH interval. More complex, than in chosen model, behaviour poly(I)-poly(C) at high degree of saturation by the ligand $H^+$ can be due to, for example, beginning of the following step of protonation.

The values of intrinsic protonation constants logarithms $lgK_g$ for poly(I-)-poly(C) and poly(I-)[27] are nearly the same as can be seen in TABLE. In the previous study of the acid-base equilibria of the poly(I)-poly(C) acid in wide pH range were shown the existence of four different species, depending on the degree of protonation of the different kinds of nitrogenated bases. The species which appears at pH > 9 is related to the mixture of poly(I-) and poly(C), in the pH range 5-9, the double stranded poly(I)-poly(C) species predominates. As has been proposed previously[28-30] such species has a structure similar to that of natural double stranded nucleic acid stabilised because of the formation of protonated inosine-deprotonated cytosine base pairs as presented on figure 2.

## CONCLUSIONS

It is well established that poly(A)-poly(U) cooperative zipper mechanism of double helix formation requires three base pair nucleus[31]. Based on the results presented here it is possible to propose that the same mechanism govern double helix formation in the course of decreasing pH in the system poly(I-)-poly(C) in aqueous solution. The first step is association of single base pair. A proton traps by poly(I-) molecular structures first leads to single protonated complexes (with neutral inosinic base) were formed simply by statistical occupation of the free bases as a result of random distribution. Observed protonation constant value comparable with corresponding value of pure poly(I-) in aqueous solution. The neutral inosinic base forms pair with cytidylic base of poly(C) as shown on figure 2. Isolated base pair is rather unstable, but addition of

another, neighbouring and stacked base pair follows with increasing stability. This conform the high value of cooperativity coefficient. A third base-pair is stacked on top of the first two and creates a suitable nucleus from which further addition of stacked base pairs leads to stepwise construction of a helix. The formation of triple stacked base pairs has been loaded in the model as double contiguous site formation. The contiguous site formation in the poly(I)-poly(C) provides by high cooperativity. That is why growth of the double helix the same as for poly(A)-poly(U) can be proposed as spontaneous due mainly to geometrical constraints of sugar-phosphate backbone.

## REFERENCES


**[1]** H.Gampp, M.Maeder, C.J.Meyer, A.Zuberbuhler, Talanta, **33**, 943 **(1986)**.

**[2]** H.Gampp, M.Maeder, C.J.Meyer, A.Zuberbuhler; Talanta, **32**, 1133 **(1985)**.

**[3]** J.Jaumot, R.Eritja, R.Tauler, R.Gargallo; Nucleic Acids Res., **34**, 206 **(2006)**.

**[4]** M.Vives, R.Tauler, R.Eritja, R.Gargallo; Anal. Bioanal.Chem., **387**, 311 **(2007)**.

**[5]** N.Khan, A.Avino, R.Tauler, C.González, R.Eritja, R.Gargallo; Biochimie, **89**, 1562 **(2007)**.

**[6]** R.Tauler, A.De Juan; In: Practical Guide to Chemometrics, 2nd Edition, Taylor & Francis Group, LLC, Chapter 11, **(2006)**.

**[7]** http://www.ub.edu/mcr/ntheory.htm.

**[8]** E.R.Malinowski, D.E.Howery; Factor Analysis in Chemistry, 2nd Ed., Wiley, NewYork, **(1991)**.

**[9]** A.Kudrev; Rus.J.Coord.Chem., **25**, 152 **(1999)**.

**[10]** A.Kudrev; Rus.J.Inorg.Chem., **46**, 758 **(2001)**.

**[11] (a)** G.H.Golub, F.Van Loan; Matrix Computation, 2nd Edn, Johns Hopkins Univ.Press, London, **(1989)**; **(b)** B.M.Wise; PLS Toolbox, Version 1.3 for Use with Matlab™, **(1992)**.

**[12]** A.de Juan, Y.Vander Heyden, R.Tauler, D.L.Massart; Analyt.Chim.Acta, **346**, 307 **(1997)**.

**[13]** R.Manne, Chemom; Intel.Lab.Systems, **27**, 89 **(1995)**.

**[14]** R.M.Dyson, S.Kaderli, G.Lawrance, M.Maeder, A.D.Zuberbuhler; Anal.Chim.Acta, **353**, 381 **(1997)**.

**[15]** A.Izquierdo-Ridorsa, J.Saurina, S.Hernandez-Cassou, R.Tauler; Chemom.and Intel.Lab.Systems, **38**, 183 **(1997)**.

**[16]** J.D.McGhee, P.H.Von Hippel; J.Mol.Biol., **86**, 469 **(1974)**.

**[17]** T.L.Hill; Cooperativity Theory in Biochemistry, Steady State and Equilibrium Systems, N.Y.Springer Verlag, 459 **(1985)**.

**[18]** J.L.Breslow, D.M.Crothers; Biochem., **20**, 3547 **(1981)**.

**[19]** G.Schwarz, F.Watanabe; J.Mol.Biol., **163**, 467 **(1983)**.

**[20]** D.Yu.Nechipurenko, G.V.Gursky; Biofizika, **48**, 773 **(2003)**.

**[21]** J.L.Garcés, C.Rey-Castro, C.David, S.Madurga, F.Mas, I.Pastor, J.Puy; J.Phys.Chem.B, **113**, 15145 **(2009)**.

**[22]** A.Kudrev; Rus.J.of General Chem., **72**, 1501 **(2002)**.

**[23]** D.W.Marquardt; J.Soc.Ind.Appl.Math., **11**, 431 **(1963)**.

**[24]** I.N.Bronshtain, K.A.Semendyaev; Spravochnik Po Matematike (Reference Book for Mathematics), Nauka, Moscow, **(1986)**.

**[25]** A.G.Kudrev; Biophyzika, **47**, 439 **(2002)**.

**[26]** A.Izquierdo-Ridorsa, E.Casassas, R.Gargallo, I.Marques, R.Tauler, Reactive Funct.Polymers, **27**, 1 **(1995)**.

**[27]** A.Kudrev; Rus.J.Jeneral Chem., **73**, 23 **(2003)**.

**[28]** R.Gargallo, F.Cuesta-Sanches, A.Izquierdo-Ridorsa, D.L.Massart; Anal.Chem., **68**, 2241 **(1996)**.

**[29]** D.Thiele, W.Guschlauer, A.Favre; Biochim. Biophys.Acta, **272**, 22 **(1972)**.

**[30]** D.Thiele, W.Guschlauer; Biopolymers, **8**, 361 **(1969)**.

**[31]** W.Saenger; Principles of Nucleic Acid Structure, N.Y., Springer-Verlag, 141 **(1988)**.