CHEMICAL EQUILIBRIA STUDIES USING MULTIVARIATE ANALYSIS METHODS

Joaquim Jaumot, Ramon Eritja, Agnieszka Stępień, Raimundo Gargallo

Dept. de Química Analítica, Universitat de Barcelona, Barcelona, Spain

joaquim@apolo.qui.ub.es

INTRODUCTION

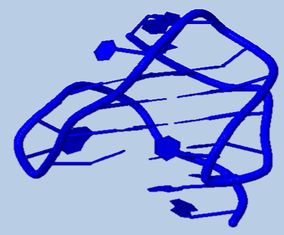
Multivariate Resolution methods are widely used in the analysis of evolving processes or complex mixtures[1].

Traditionally, analysis of the data of a process has been carried out by fitting the measured signal to a proposed physicochemical model (approach known as hard-modeling) because the data is forced to fulfill the postulated model. On the other hand, it appears the soft-modeling approaches in which the data is analyzed using only natural constraints (such as non-negativity or closure) without the need of postulating any chemical model. In fact, one of the goals of this kind of methods is the recovery of a reliable chemical model from the resolved profiles. Finally, hybrid methods that combine the main advantages of soft- and hard-modeling methods have been proposed. These methods allow applying the natural constraints usual in the soft-modeling methods together with a constraint in which a hard equilibrium model is defined. This approach allows maintaining the flexibility in the resolution of the soft-modeling methods and, also, allows minimizing the effects of the rotational and intensity ambiguities.

In this work, a comparison of the usefulness of these different approaches is shown when analyzing experimental data corresponding to spectroscopically monitored equilibria of two DNA sequences..

EXPERIMENTAL

Thirty-two base-long oligonucleotides with sequences 5'-CCT CCC CCT CTT CCC TCT TCC CAC ACC GCC CT-3' (kras-C) and 5'-AGG GCG GTG TGG GAA GAG GGA AGA GGG GGA GG -3' (kras-G) were used in this work. The sequences kras-C and kras-G are located within the *kras* oncogene and may form two complex structures, known as *i*-motif and G-quadruplex, respectively[2]. Interaction studies were carried out using the ligand TMPyP4.

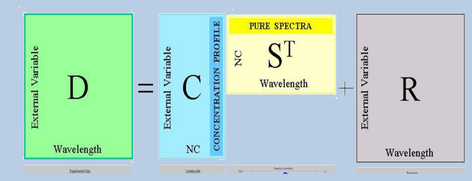


Acid-base and mole-ratio experiments have been carried out to study the acid-base equilibria of the considered oligonucleotides and the interaction of these with the ligand. Monitoring of the acid-base titrations was carried out in-line by adjusting the pH of solutions containing the oligonucleotides. Mole ratio experiments were carried out by progressively adding small volumes of ligand stock solution to a 2 μM oligonucleotide solution. Experimental conditions were as follows: 25 ºC, pH 7.1 or 5.0, and 150 mM ionic strength (adjusted with KCl).

DATA ANALYSIS METHODS

Resolution methods allow the decomposition of the initial data matrix **D** into the product of two data matrices **C** and $\mathbf{S}^T$, each of them including the pure response profiles associated with the row and column direction of the initial data matrix solving the Equation represented in the figure at right.

Three different resolution methods have been tested to resolve the data under analysis:



MCR-ALS (Pure Soft- Modelling)

Soft-modeling methods does not require the postulation of a chemical model. This procedure solves the equation iteratively by an Alternating Least Squares (ALS) algorithm[3]. This optimization requires initial estimations either of **C** or $\mathbf{S}^T$ and allows the application of several constraints in the spectra and or concentration profiles to give more chemical meaning to the optimal mathematical solution.

Estimation of the $\mathbf{C}_0$ or $\mathbf{S}^T_0$ matrices (using EFA, pure variable detection methods) for a *k* number of components.

a) Estimation of $\mathbf{C}_{calc}$ matrices under constraints (non-negativity, unimodality, closure, ...): $\mathbf{C}_{calc}=\mathbf{D}\mathbf{S}_0(\mathbf{S}^T_0\mathbf{S}_0)^{-1}$

b)Estimation of $\mathbf{S}^T_{calc}$ matrix under constraints:

$\mathbf{S}^T_{calc}=(\mathbf{C}^T_{calc}\mathbf{C}_{calc})^{-1}\mathbf{C}^T_{calc}\mathbf{D}$

Determination of the residuals:

$\mathbf{E}=\mathbf{D}-\mathbf{C}_{calc}\mathbf{S}^T_{calc}$

Optimization of **C** and $\mathbf{S}^T$ by minimization of **E** by repeating second and third steps until convergence

MCR-Hybrid (Hard-/Soft- Modelling)

Hybrid methods combine the advantages from Hard- and Soft- modeling methods:

- HM allows to minimize the ambiguities related to Factor Analysis methods.
- SM allows to analyze systems in which there are interfering species which behavior can not be described using a model.

The resolution procedure is similar to those described for the pure SM approach but using a new HM constraint[4]:

Estimation of the $\mathbf{S}^T_0$ matrices (using EFA, pure variable detection methods) for a *k* number of components.

a) Estimation of $\mathbf{C}_{calc}$ matrices under constraints :

$\mathbf{C}_{calc}=\mathbf{D}\mathbf{S}_0(\mathbf{S}^T_0\mathbf{S}_0)^{-1}$

b) Estimation of $\mathbf{S}^T_{calc}$ matrix under constraints:

$\mathbf{S}^T_{calc}=(\mathbf{C}^T_{calc}\mathbf{C}_{calc})^{-1}\mathbf{C}^T_{calc}\mathbf{D}$

Hard-modeling constraint
Forces the profiles in **C** to fulfill a model

Determination of the residuals:

$\mathbf{E}=\mathbf{D}-\mathbf{C}_{calc}\cdot\mathbf{S}^T_{calc}$

Optimization of **C** and $\mathbf{S}^T$ by minimization of **E** by repeating second and third steps until convergence

EQUISPEC (Pure Hard- Modelling)

Hard-modeling methods require the fulfillment of a postulated chemical models (kinetic, acid-base , complexation, ...)[5].

This model is defined by:

- stoichiometries of all species involved in the proposed equilibria
- approximate values of the equilibrium constants

The resolution procedure used to solve the $\mathbf{D}=\mathbf{C}\mathbf{S}^T$ equation can be summarized in the following steps:

Building up of the **C** matrix according to the postulated physicochemical model: $C_{model}=f(x,model)$

Estimation of pure spectra matrix:

$\mathbf{S}^T_{calc}=\mathbf{C}^+_{calc}\cdot\mathbf{D}=(\mathbf{C}^T_{calc}\cdot\mathbf{C}_{calc})^{-1}\cdot\mathbf{D}$

Determination of the residuals:

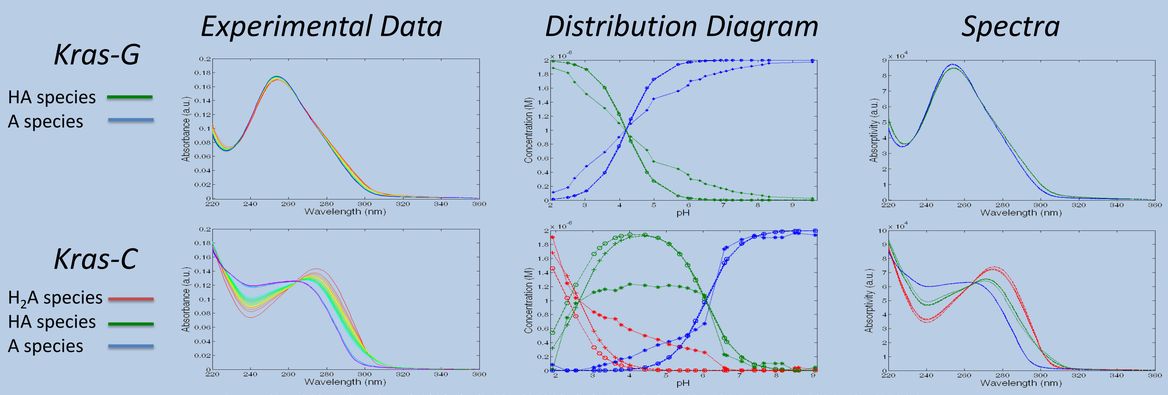
$\mathbf{E}=\mathbf{D}-\mathbf{C}_{calc}\cdot\mathbf{S}^T_{calc}=\mathbf{D}-(\mathbf{C}_{calc}\cdot\mathbf{C}^+_{calc}\cdot\mathbf{D})$

Optimization of **C** and $\mathbf{S}^T$ by minimization of **E** by changing the *k* values.

RESULTS

Acid-Base Equilibria

The three different approaches have been tested in the analysis of kras-C and kras-G acid-base titrations:



Lines: MCR-ALS (dotted), MCR-Hybrid (dashed), EQUISPEC (dash-dot). Symbols: MCR-ALS (*), MCR-Hybrid (+), EQUISPEC (o)

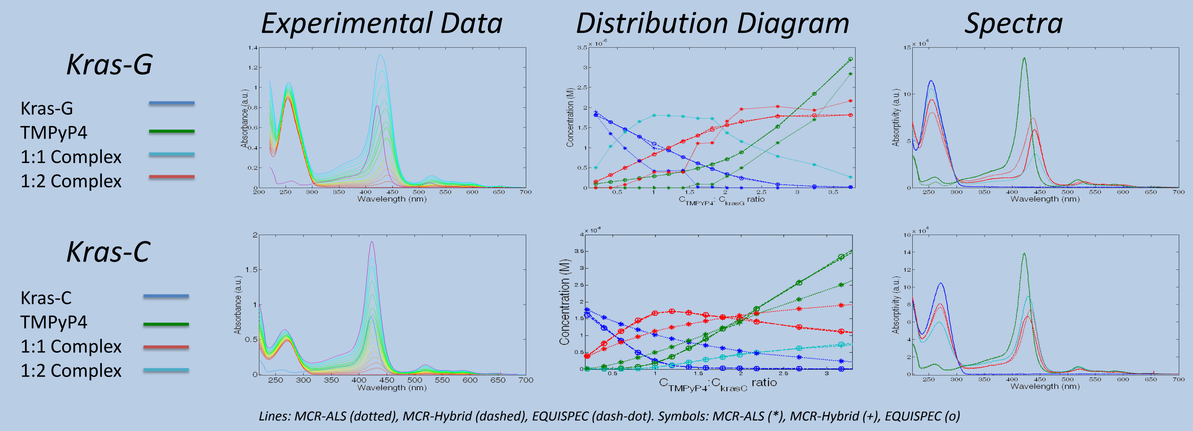
In alls case, pKa values of the acid-base system can be obtained.

MCR-Hybrid and EQUISPEC provides the pKa value as a result of the optimization while the MCR-ALS pKa value is obtained from the crossing point of the concentration profiles.

	Kras-G Titration	Kras-C Titration	
	pKa	pKa1	pKa2
MCR-ALS	4.19	2.65	6.15
MCR-Hybrid	4.19(0.06)	2.63(0.04)	6.08(0.04)
Equispec	4.18(0.09)	2.34(0.15)	6.06(0.05)

Complexation Equilibria

The three different approaches have been tested in the analysis of the titration of kras-C and kras-G with the TMPYP4 ligand:



Lines: MCR-ALS (dotted), MCR-Hybrid (dashed), EQUISPEC (dash-dot). Symbols: MCR-ALS (*), MCR-Hybrid (+), EQUISPEC (o)

MCR-Hybrid and EQUISPEC provides very similar results that fitted models with 1 or 2 ligands. However, classical MCR-ALS provides very different results (for instance, different number of species).

In this case, complexation constants can also be obtained:

	KrasG : TMPyP4 Mole-Ratio		KrasC : TMPyP4 Mole-Ratio	
	Log(β1)	Log(β2)	Log(β1)	Log(β2)
MCR-ALS	-	-	-	-
MCR-Hybrid	-	12.9 (0.8)	7.4(1.4)	12.7.(1.5)
Equispec	-	13.0 (0.6)	7.6 (1.6)	12.9(1.6)

CONCLUSIONS

We have compared three different chemometrical approaches (soft-, hard- and hybrid modelling methods) that can be used to study the equilibria of complex processes, specifically, acid-base and complexation equilibria. However, it has been shown that the application of a hybrid approach allows exploiting the advantages of the two different methods. Thus, this method shows flexibility in the analysis (for instance, modelling an unknown interference or a complex baseline drift) as a soft-modelling method and also allows proposing a chemical model as a pure hard-modelling method for the species involved in the equilibrium minimizing the factor analysis ambiguities.

REFERENCES

1. Jaumot, J; Vives, M; Gargallo, R. Analytical Biochemistry, 327(1), 1-13, 2004.
2. Cogoi, S; Xodo, LE. Nucleic Acids Research, 34(9), 2536-2549, 2006.
3. Tauler, R. Chemometrics and Intelligent Laboratory Systems, 30(1), 133-146, 1995
4. Diewok, J; de Juan, A; Maeder, M, et al. Analytical Chemistry, 75(3), 641-647, 2003
5. Dyson, RM; Kaderli, S; Lawrance, GA, et al. Analytica Chimica Acta, 353(2-3), 381-393, 1997

ACKNOWLEDGEMENTS

This research was supported by the Spanish Ministerio de Ciencia e Innovación (grant number CTQ2009-11572) and the Generalitat de Catalunya (grant number 2009-SGR-45).