

Chemometric tools for multitechnique and multiexperiment protein process analysis

A. de Juan¹, P. Cutler², S. Navea¹ and R. Tauler³

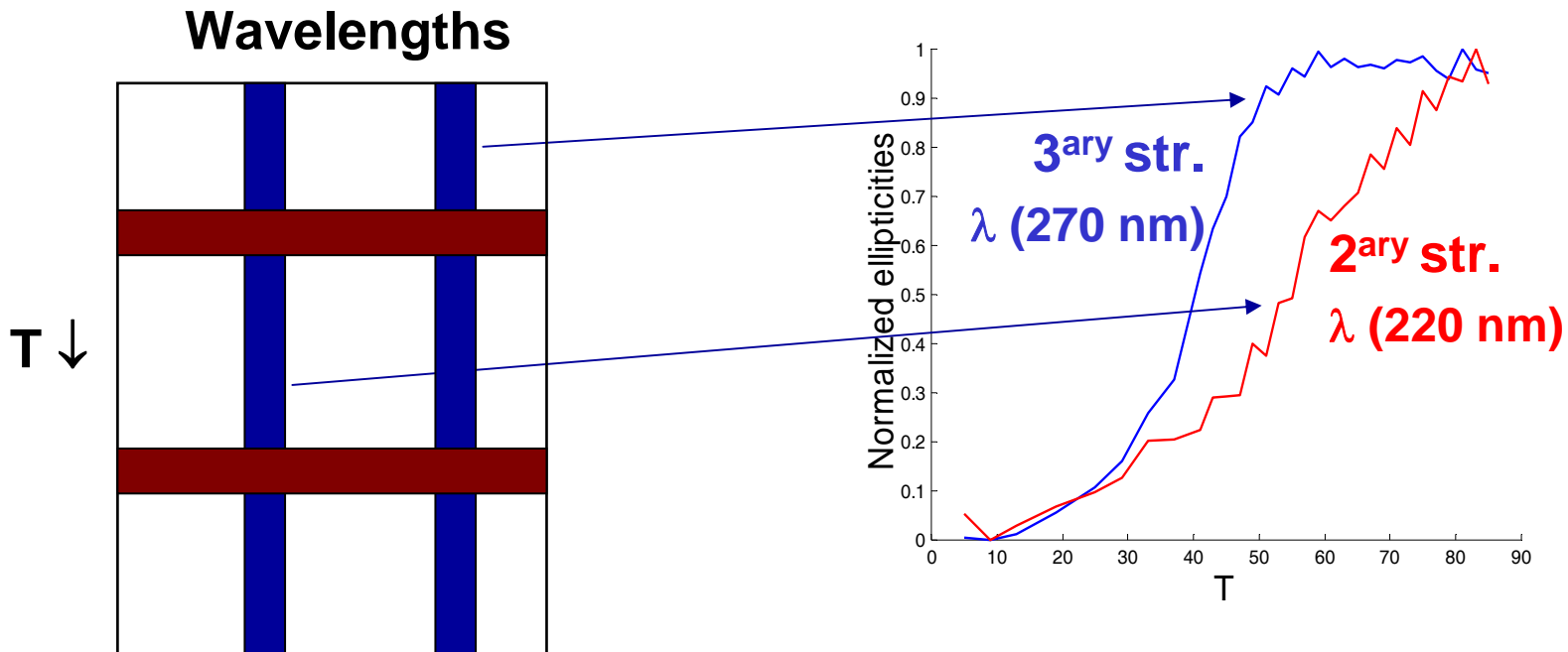
1. Chemometrics group. Dept. of Analytical Chemistry. Universitat de Barcelona. Diagonal, 647. 08028 Barcelona. anna.dejuan@ub.edu
2. Chemistry Department. East Carolina University, Greenville, (U.S.)
3. Department of Environmental Chemistry. Institute of Chemistry and Environmental Research. CSIC Jordi Girona, 16, 08034 Barcelona.

Protein processes. Characteristics.

- The mechanism is unknown, complex and protein-dependent (**no process model**).
- Intermediate conformations in processes cannot be isolated experimentally (**overlapped compound signal**).
- The complete understanding of a process may require the analysis of complex data arrangements (multispectroscopic monitoring, series of experiments performed in different conditions) and the introduction of all available information (**huge amounts of diverse information must be handled**).

Protein process. Classical monitoring (I)

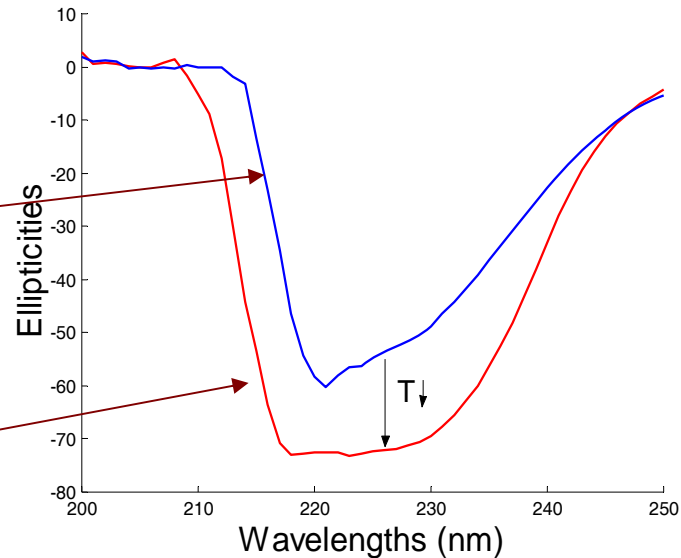
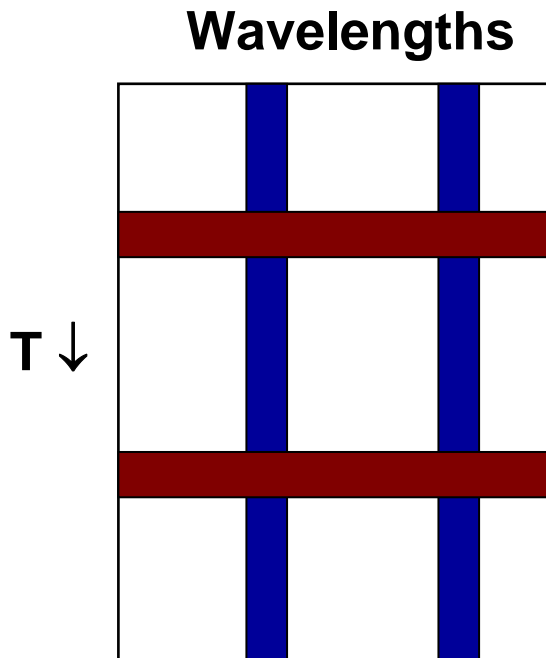
One- λ process monitoring (far- and near-UV CD)



- ✓ Qualitative information on the presence of intermediates (wavelength-dependent).
- ✓ No information on the evolution of protein conformations in the process.
- ✓ No structural information.

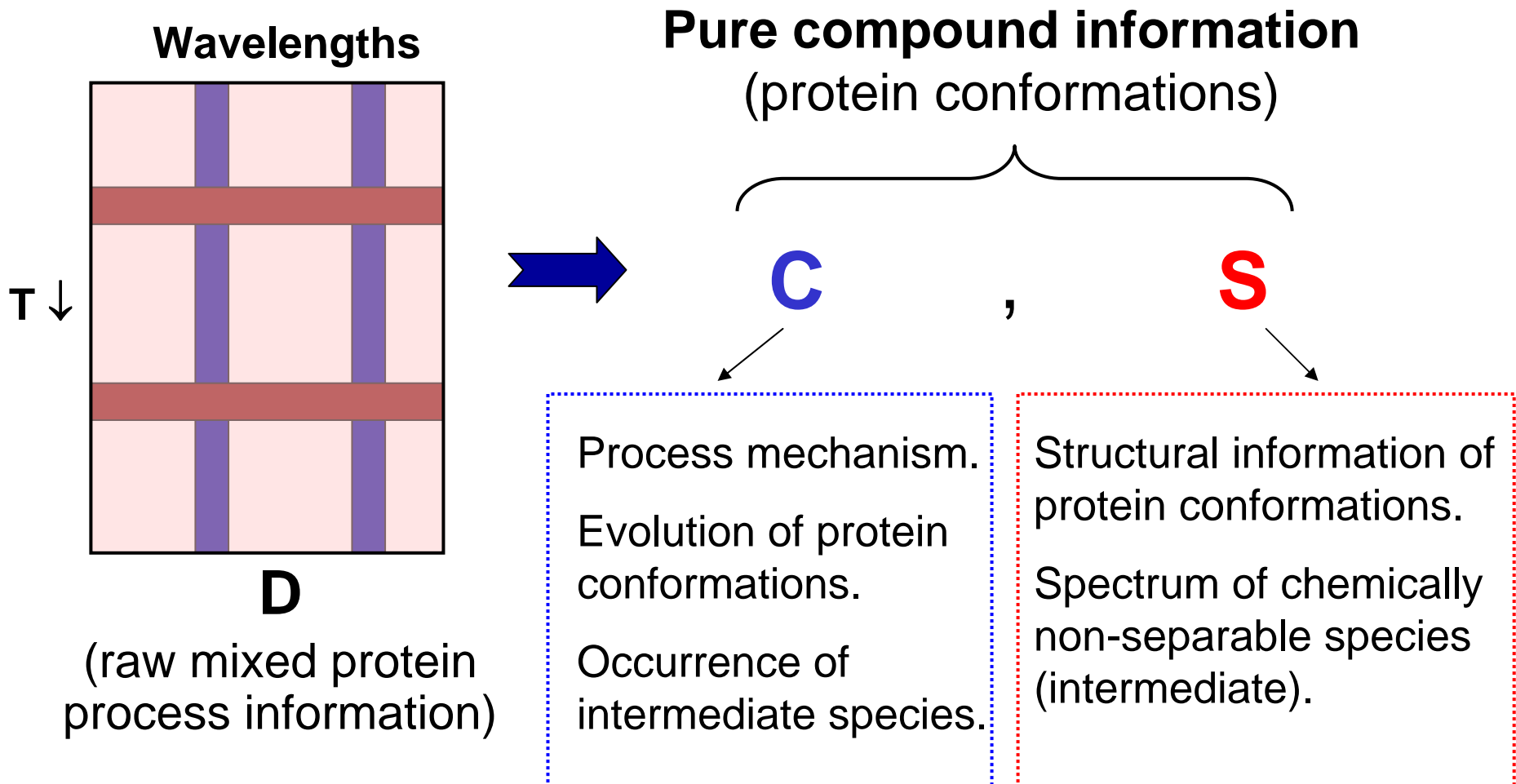
Protein process. Classical monitoring (II)

Deconvolution of spectra (far-UV CD or IR spectra)

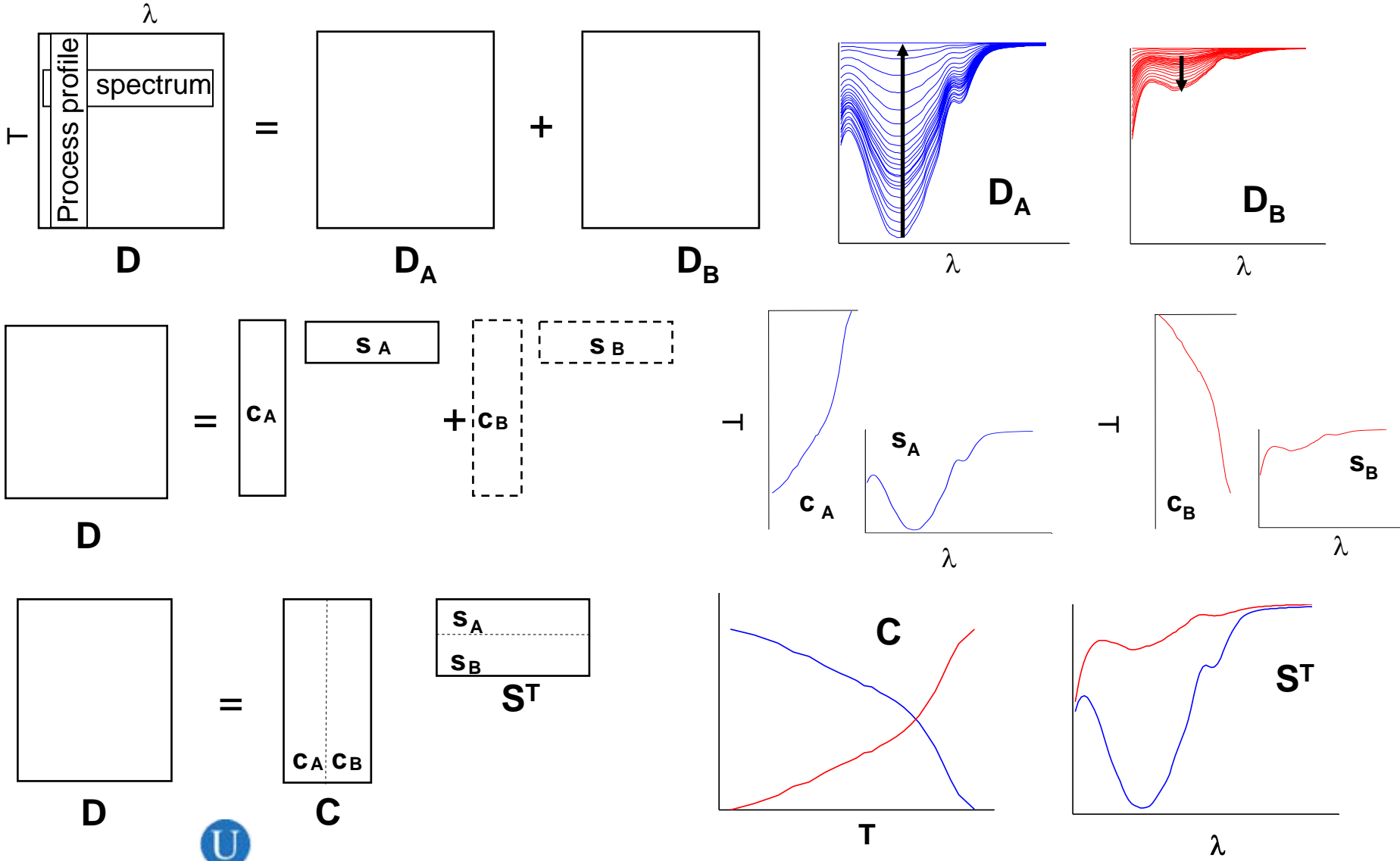


- ✓ Conformational information about abundance of 2^{ary} structural motifs.
- ✓ **Absence of process information (individual spectra).**
- ✓ **Process pathway cannot be known (partial information on process events).**

Protein process. Multivariate monitoring

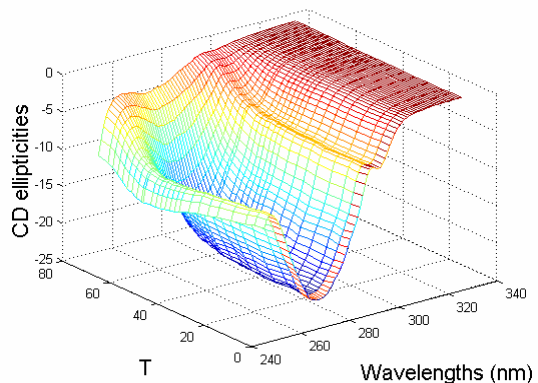
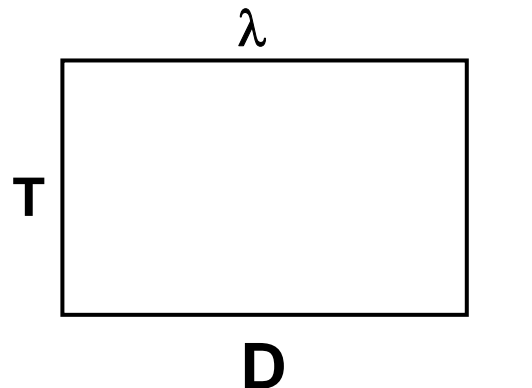


Protein process. The underlying model.

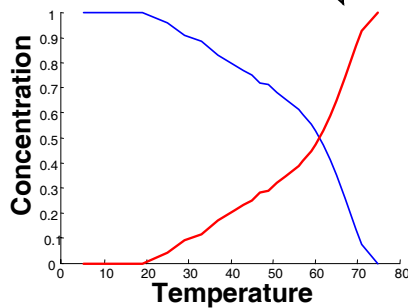
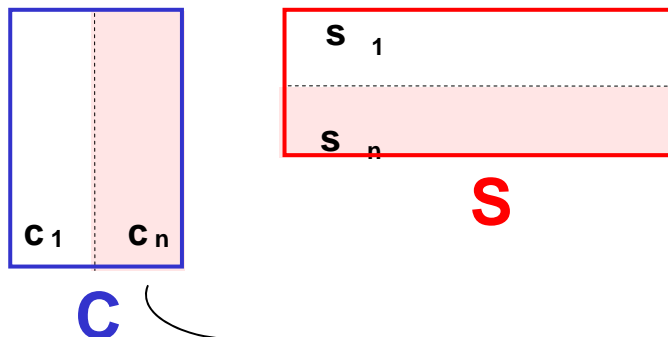


Protein process. The underlying model

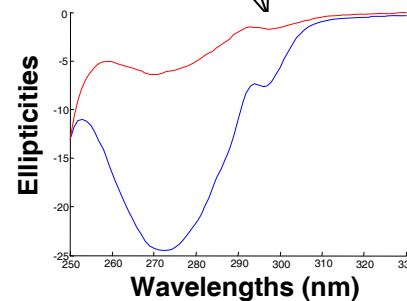
Process monitored (raw mixed information) \longrightarrow Protein conformations (pure contributions)



=



Pure concentration profiles

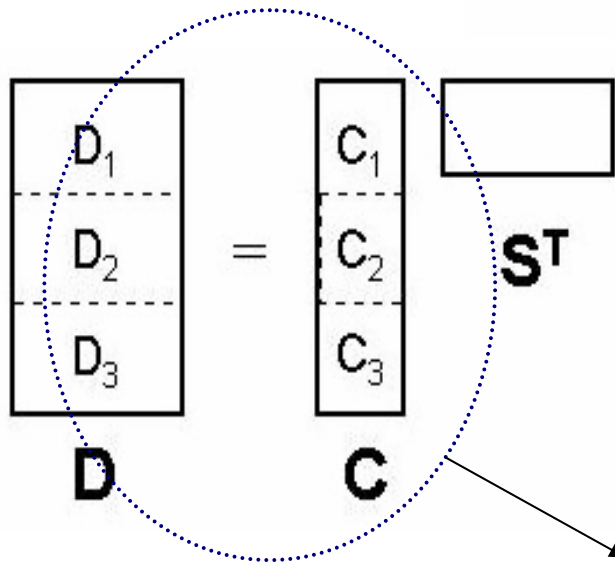
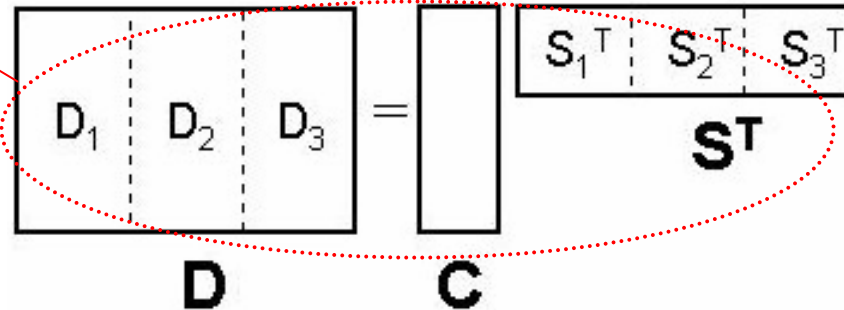


Pure signals

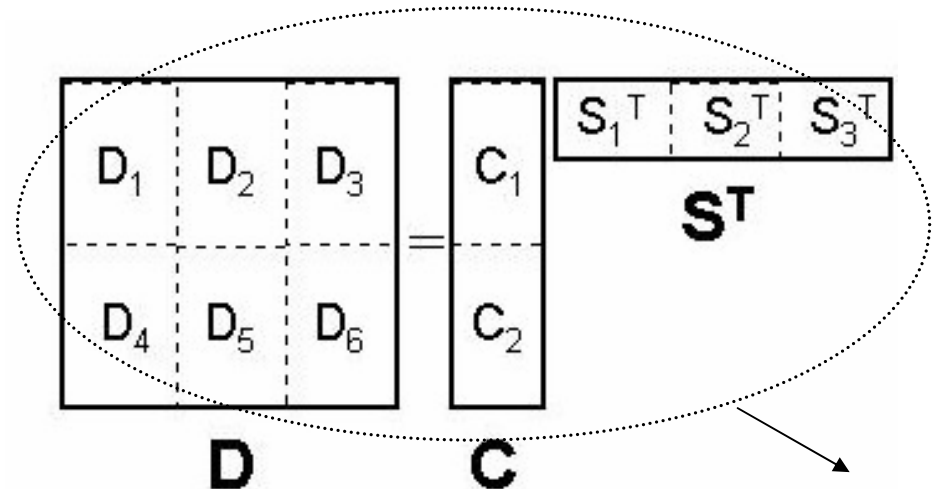
$$D = CS^T + E$$

Protein process. The underlying model

Multitechnique analysis



Multiexperiment analysis



Multiexperiment and multitechnique analysis

Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS)

$$D = CS^T$$

- Determination of the number of components (SVD).
 - Building of initial estimates (**C** or **S^T**) (*EFA, SIMPLISMA, prior knowledge...*)
 - Iterative least squares calculation of **C** and **S^T** subject to **constraints**.
 - Check for satisfactory **CS^T** data reproduction.
- Data exploration**
- Input of external information**

Optimal description of data with **chemically meaningful** profiles.

Constraints

Definition

Any systematic feature fulfilled by the profiles of the compounds in our data set.

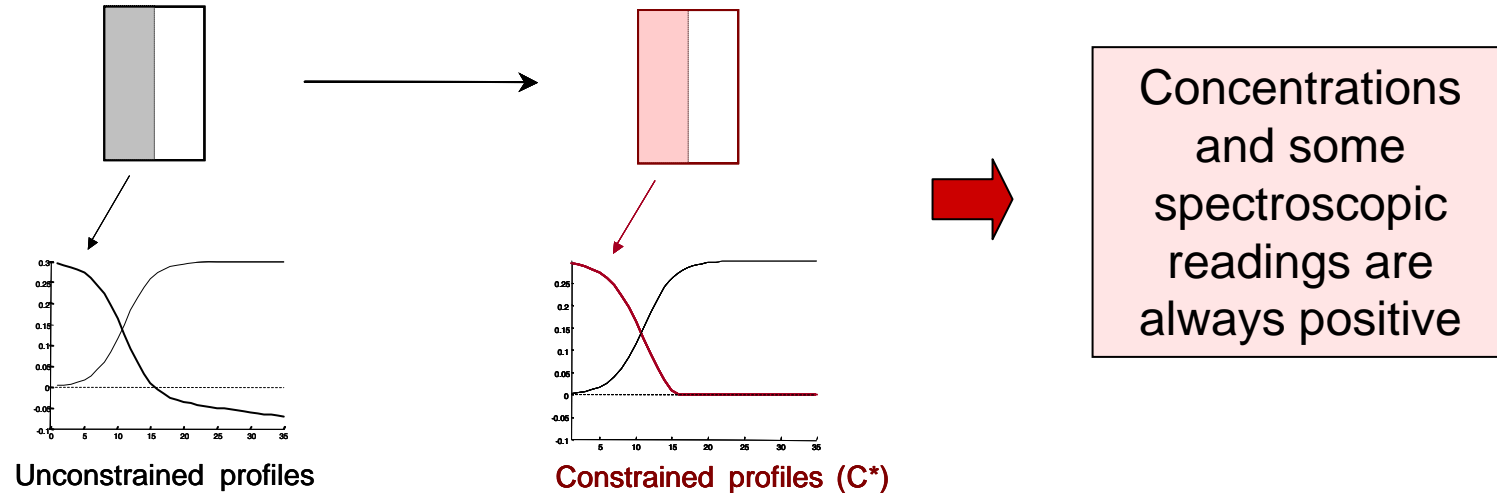
- ✓ Chemical origin
- ✓ Mathematical properties.

Application

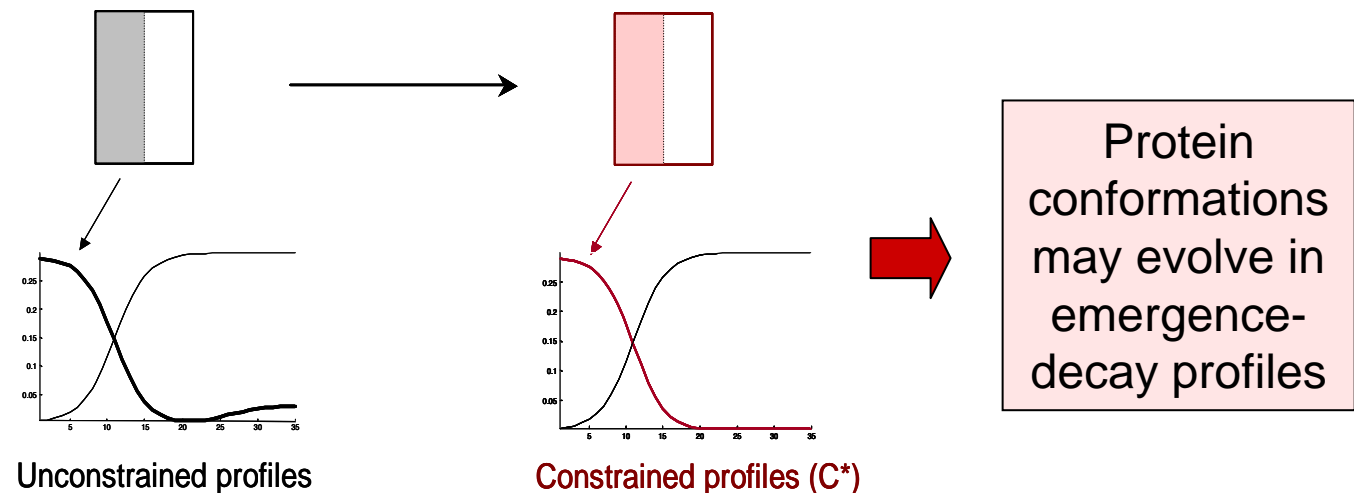
- **C** and **S** can be constrained differently.
- The profiles within **C** and **S^T** can be constrained differently.
- The submatrices in **C** and **S^T** can be constrained differently.

Chemical protein process constraints (I)

Non-negativity (C, S)

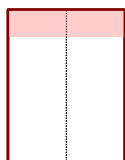
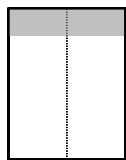


Unimodality (C)



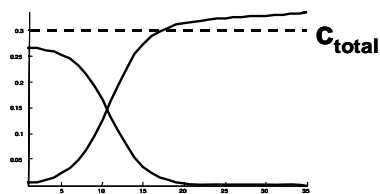
Chemical protein process constraints (II)

Closure (C)

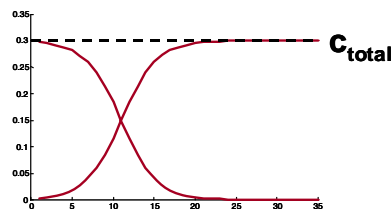


$$\Sigma = C_{\text{total}}$$

All conformations
in a process
respect the mass
balance

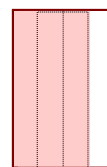
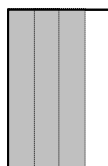


Unconstrained profiles

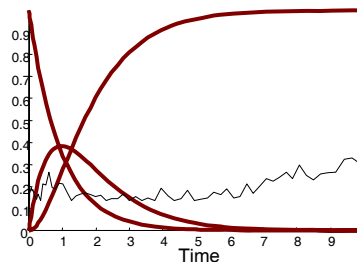
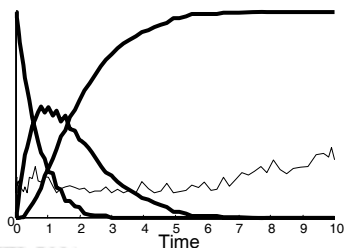


Constrained profiles (C*)

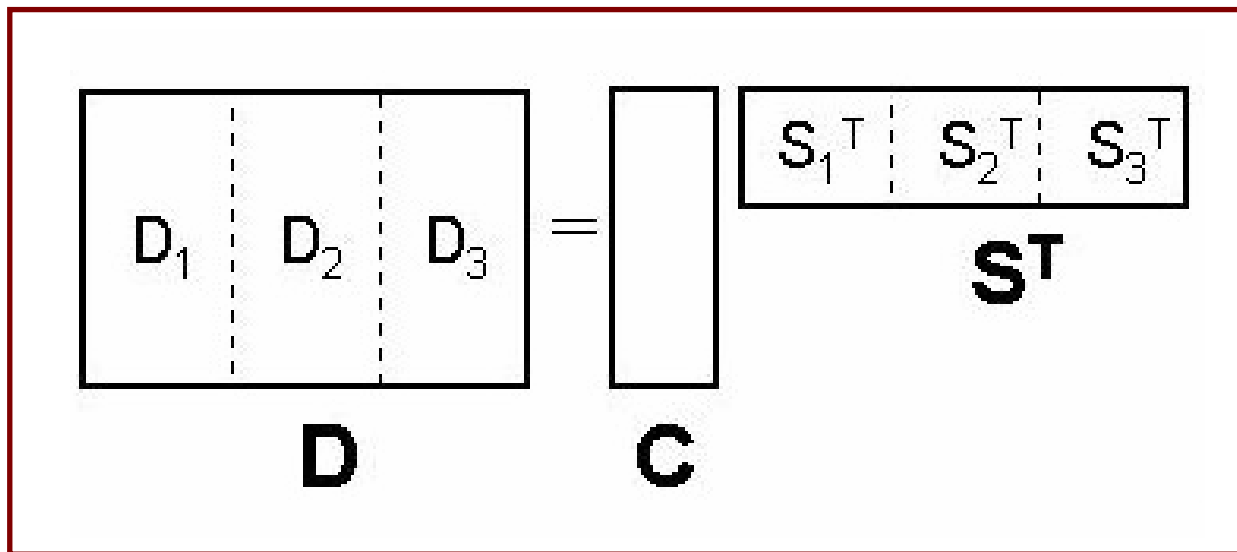
Hard-modeling (C)



Process profiles
may be shaped
according to a
model, if known



Multitechnique protein process analysis



Multitechnique data analysis

- The response linked to each process has a very complete information (**'superspectrum'**).
- The single matrix of process profiles describes the variation of all merged techniques and is much less affected by each technique-specific noise pattern (**robust description of the process**).
- The merged signal overcomes limitations due to large overlap among species in a particular technique (**increase of spectroscopic resolution**).



Multitechnique data analysis. Examples

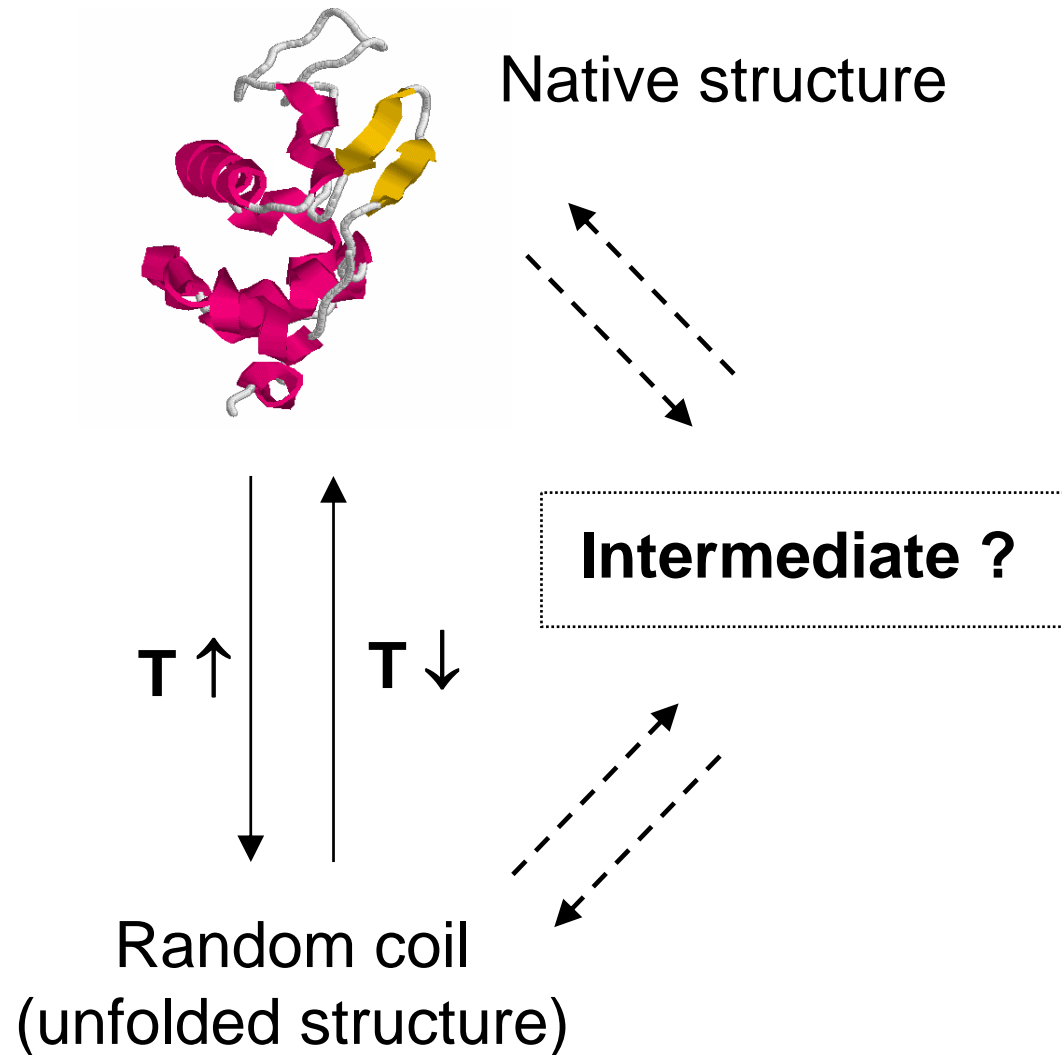
- Detection and modelling of protein folding intermediates.
 - ✓ Thermal-induced α -apolactalbumin unfolding process
 - ✓ Thermal-induced transitions of α -chymotrypsin (CD). Borges et al. *Anal. Chim. Acta* 544 (2005) 159.
- Modeling of protein conformational and oligomeric changes
 - ✓ pH-dependent changes of β -lactoglobulin (CD/MS). Navea et al. *Anal. Chem.* 78 (2006) 4768.
 - ✓ T-dependent changes of β -lactoglobulin (MIR/NIR). Navea et al. *Anal. Chem.* 75(2003) 5592.
- Modelling of folding and heme binding in heme proteins.
 - ✓ pH-dependent transitions of hemoglobin. Muñoz and de Juan. *Anal. Chim. Acta* 595 (2007) 198.



Detection and modelling of protein folding intermediates in thermal-induced unfolding of α -apolactalbumin

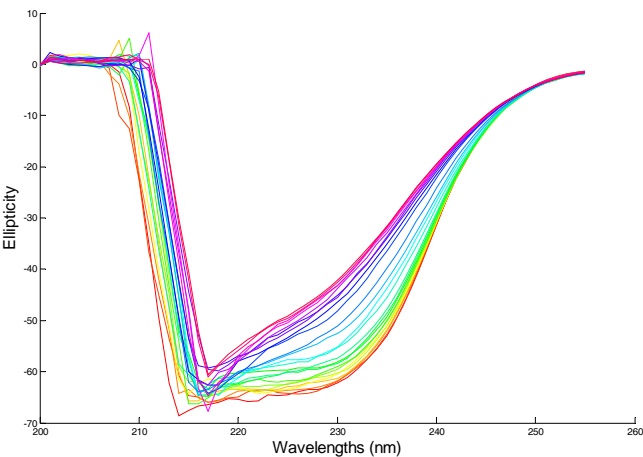


The chemical problem.

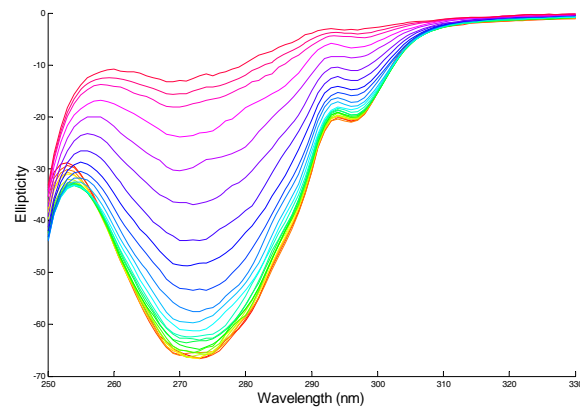


Raw data and data arrangements

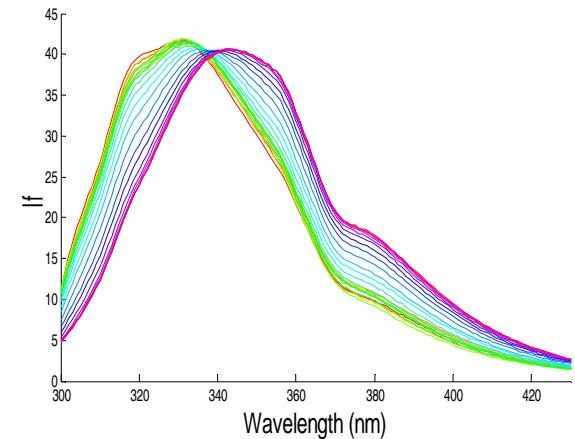
Far-UV CD



Near-UV CD



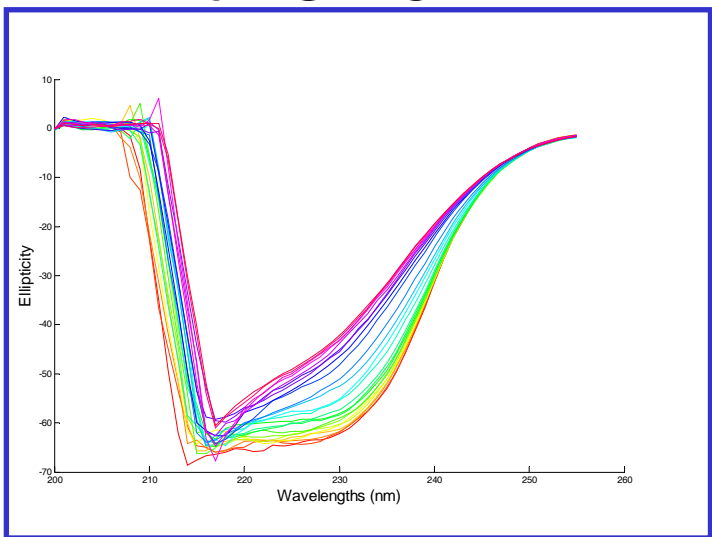
Fluorescence



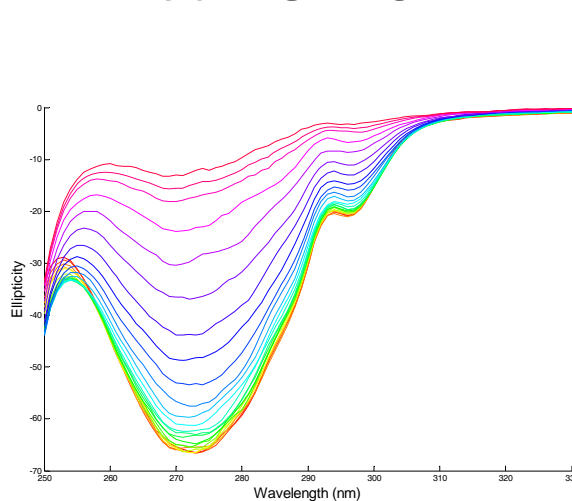
The data arrangement depends on the nature of the information sought from the process

Raw data and data arrangements

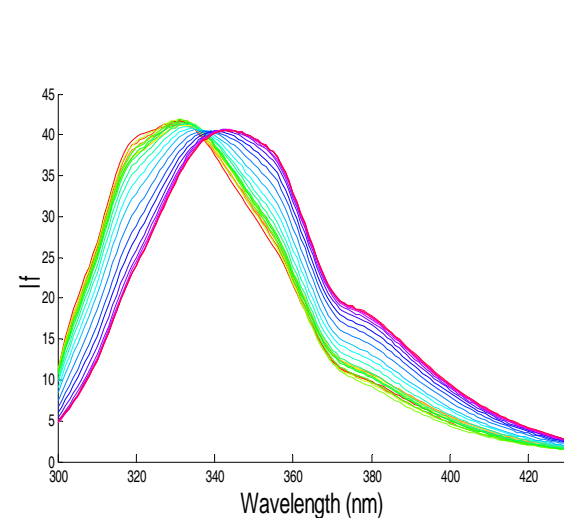
Far-UV CD



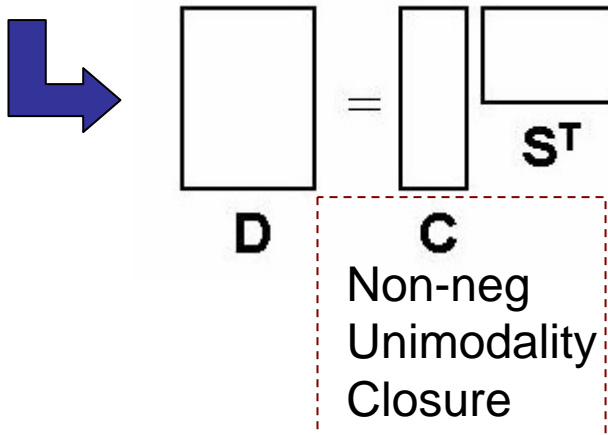
Near-UV CD



Fluorescence

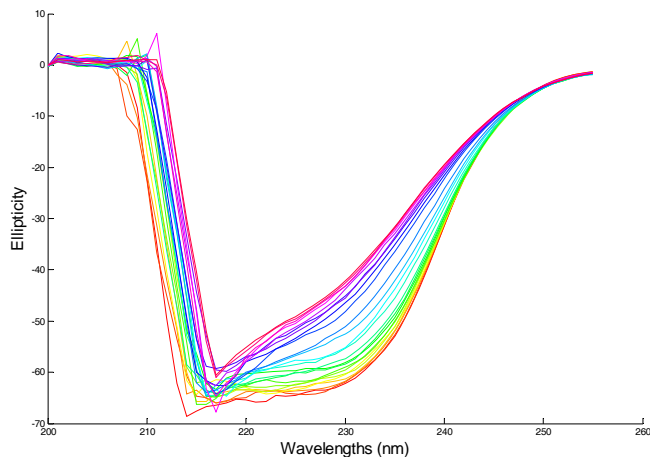


Changes in secondary structure

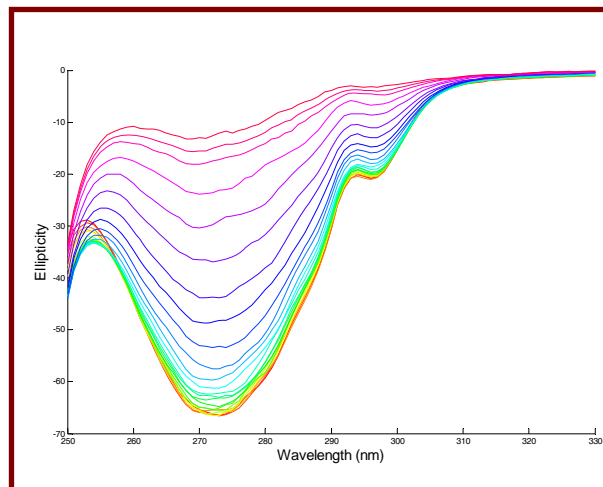


Raw data and data arrangements

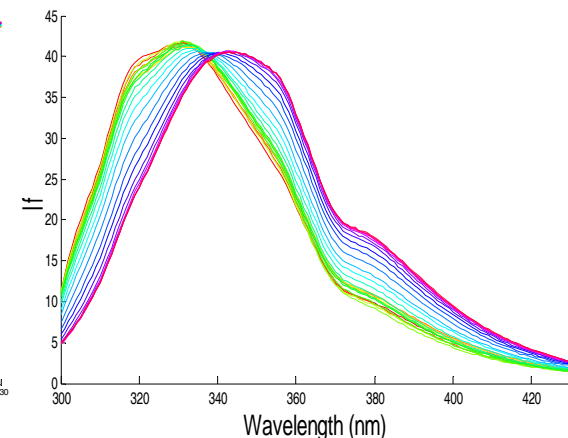
Far-UV CD



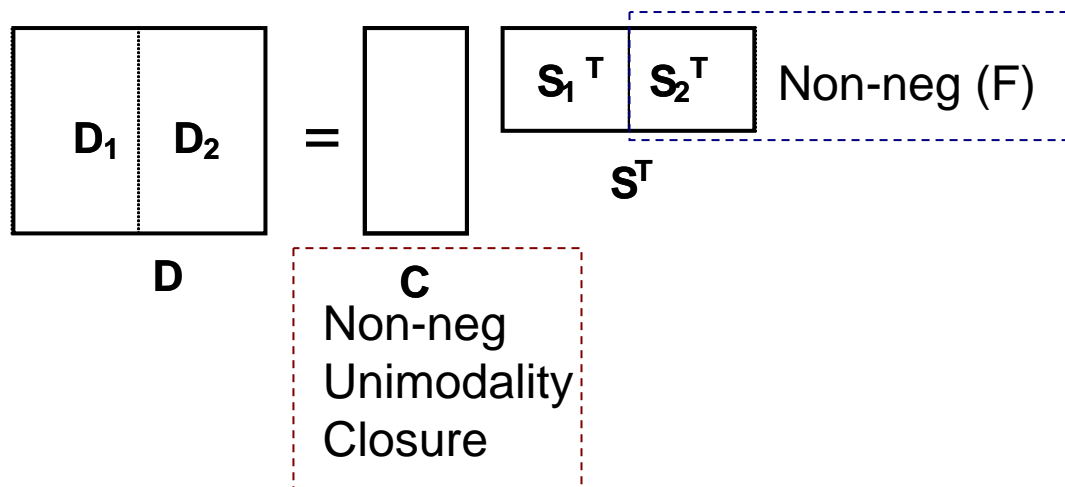
Near-UV CD



Fluorescence

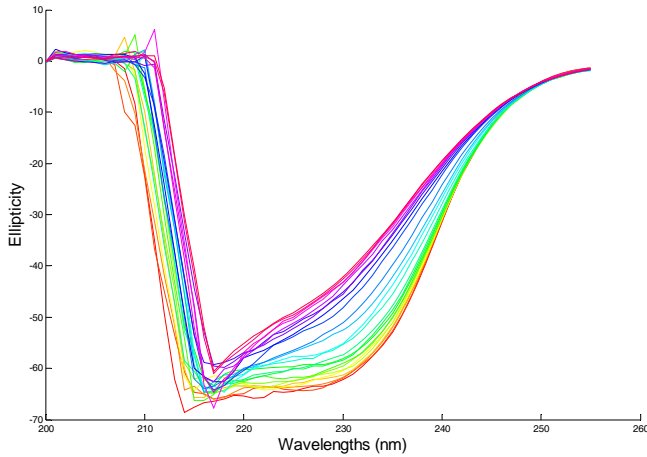


Changes in tertiary structure

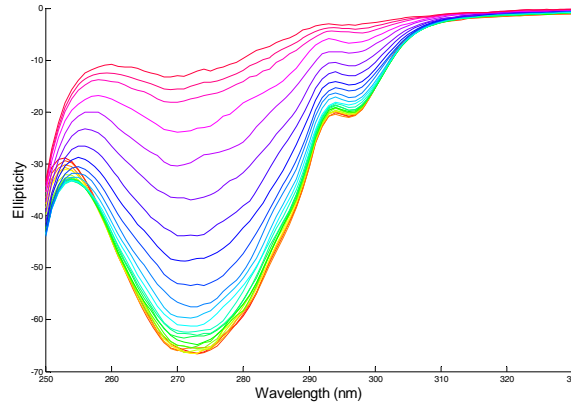


Raw data and data arrangements

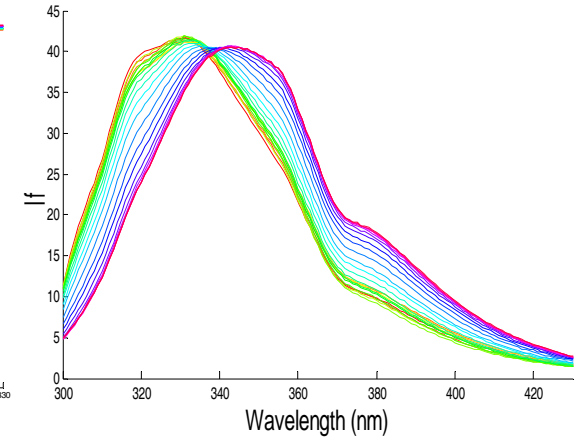
Far-UV CD



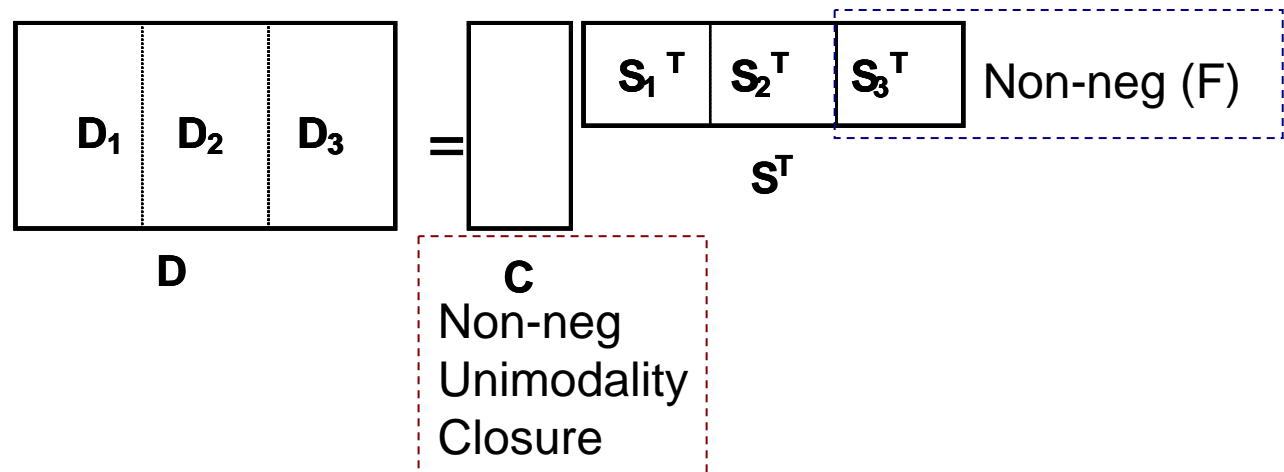
Near-UV CD



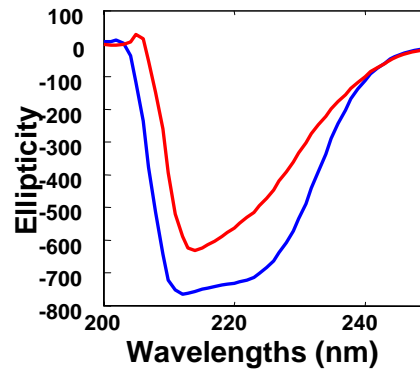
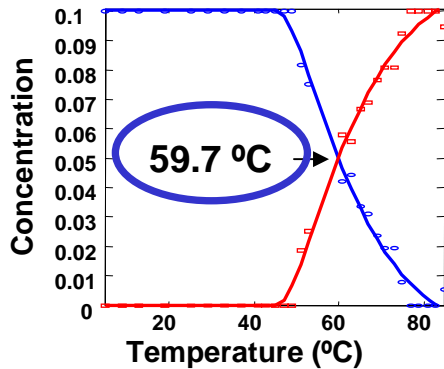
Fluorescence



Protein folding



Changes in the secondary structure

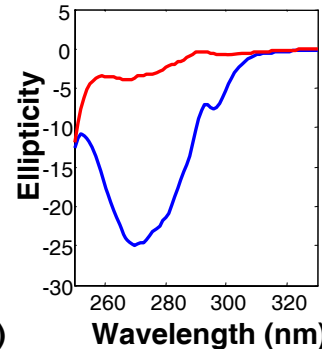
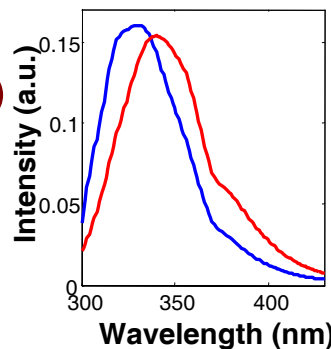
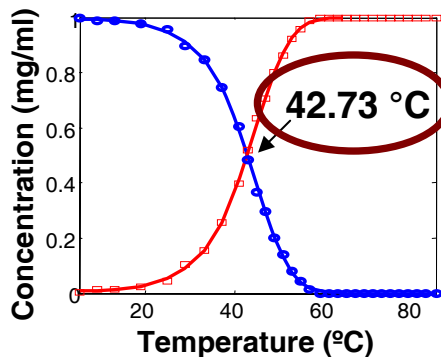


✓ **Loss of 2^{ary} structure:** disappearance of 225 nm band (helical features).

Native form

Unordered form

Changes in the tertiary structure



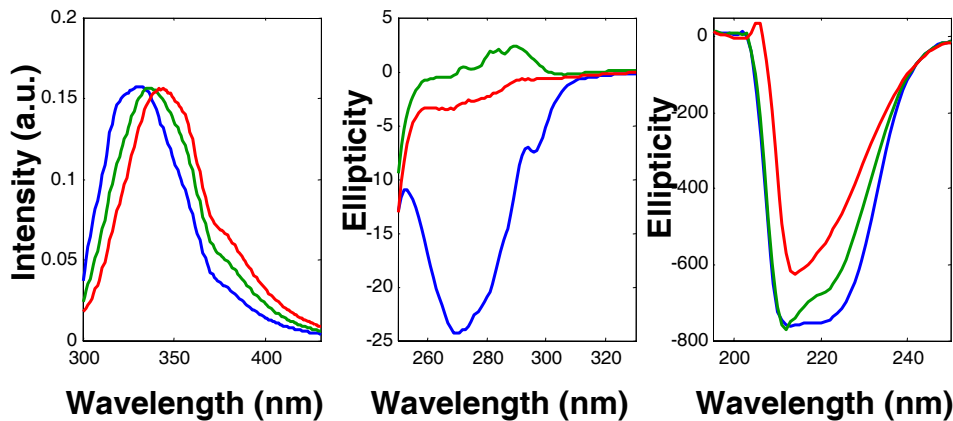
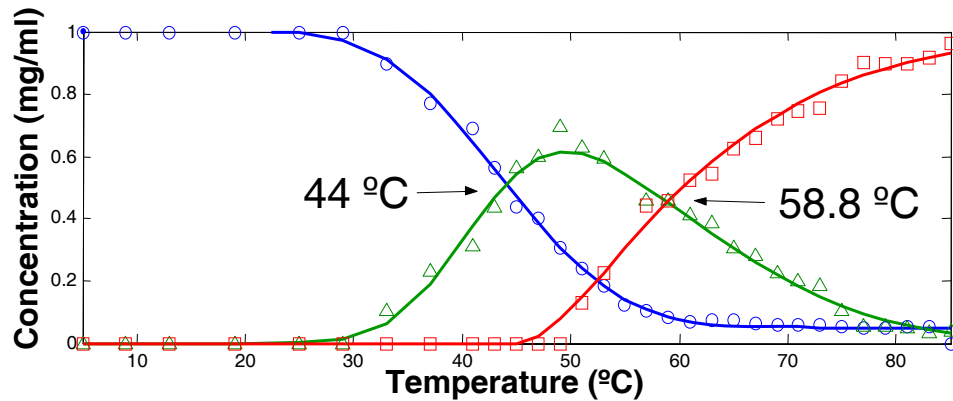
✓ **Loss of tertiary structure:** loss of signal in near-UV CD, red shift in fluorescence band.

Loss of 2^{ary} and 3^{ary} structures is sequential

$$T_{\text{tert}} < T_{\text{sec}} .$$

Intermediate?

Protein unfolding



Native conformation

Intermediate conformation

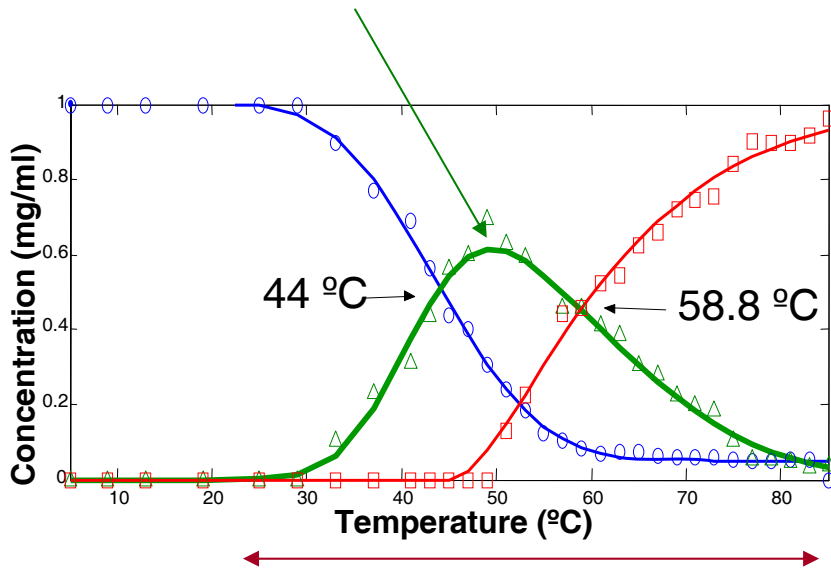
Unfolded conformation

- The protein unfolds in a two-step mechanism.
- The T of half-formation of the 2^{ary} (T_{sec}) and 3^{ary} (T_{tert}) structures agree with previous analyses.
- The spectral shapes of the intermediate are chemically meaningful (2^{ary} structure ordered and 3^{ary} structure unordered).

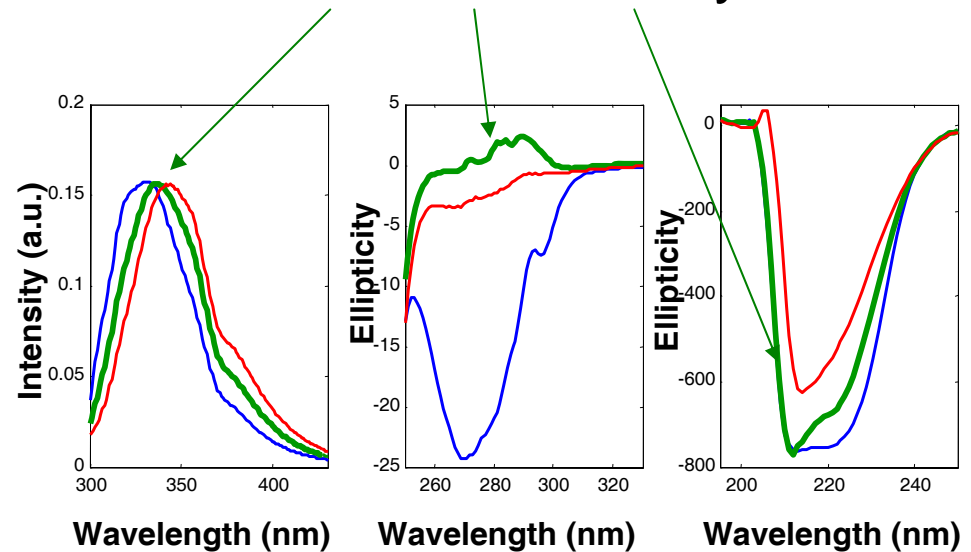


Information on intermediates

Evolution in concentration



Structural identity

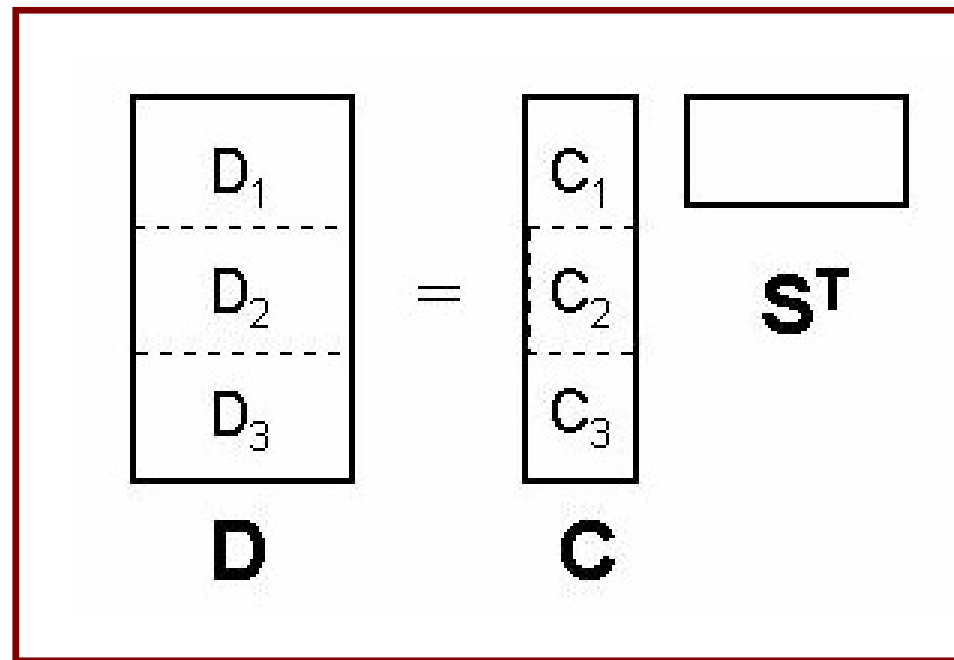


Experimental range of occurrence

Intermediates cannot be isolated by chemical means.



Multiexperiment protein process analysis



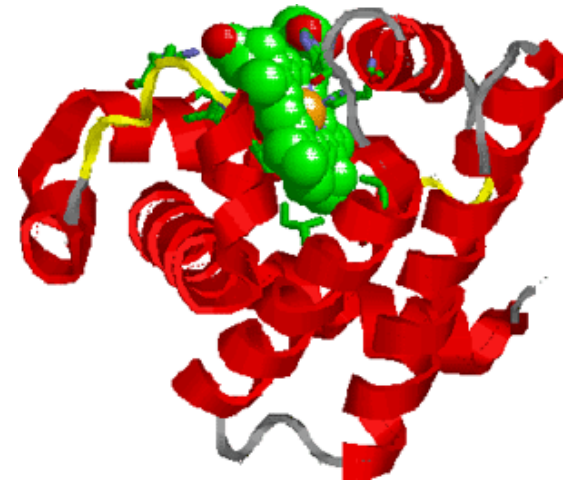
Multiexperiment data analysis

- Possibility of studying the same process induced by different agents. (**comprehensive description of process mechanism**)
- Experiments performed in different conditions, e.g., different protein:ligand ratios can be treated together (**experimental design, better definition of all species in the process**).
- The single matrix \mathbf{S}^T allows for a better definition of the resolved pure spectra. (**better identification of species**).

Multiexperiment data analysis. Examples

- **Steady-state and time-dependent protein folding mechanism.**
 - ✓ **pH- and time-dependent myoglobin unfolding.**
- **Protein-ligand interaction studies.**
 - ✓ **β -lactoglobulin and porphyrin interaction. Navea (poster)**
- **Modelling of conformational transitions in the presence of strong evolving solvent backgrounds.**
 - ✓ **T-dependent β -lactoglobulin conformational changes (MIR/NIR). Navea et al. *Anal. Chem.* 75(2003) 5592.**

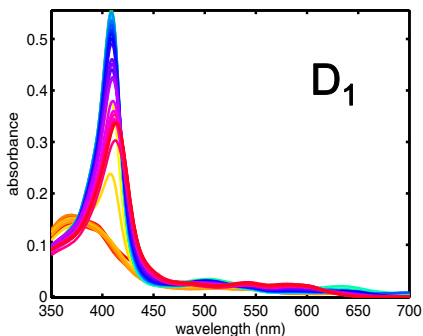
Steady-state evolution and time-dependent myoglobin conformational changes



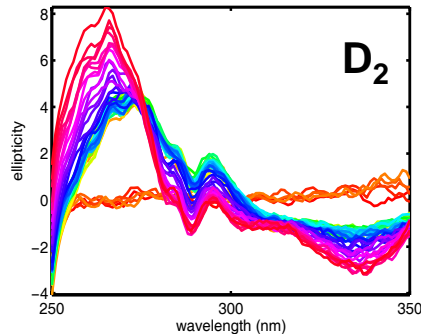
Steady-state studies

(pH-dependent multitechnique analysis)

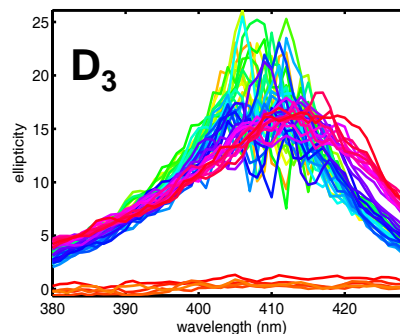
UV



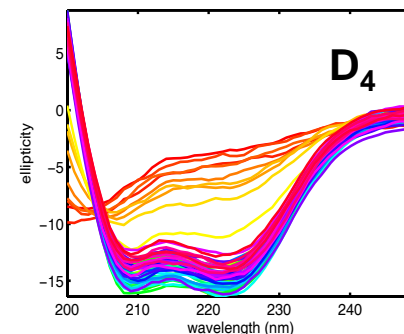
Near-UV CD



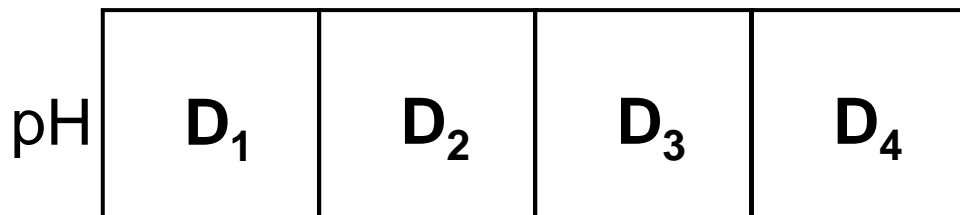
Soret CD



Far-UV CD



pH range: 1.5 - 10

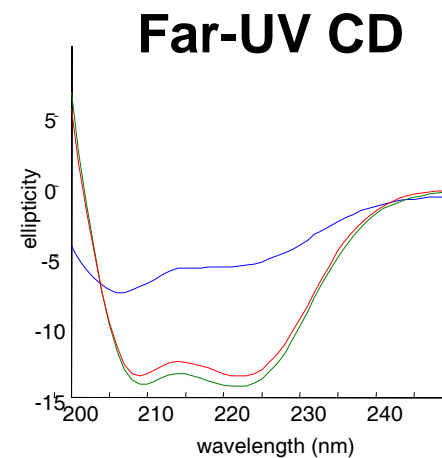
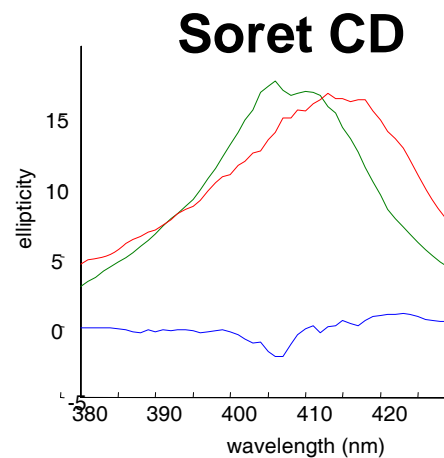
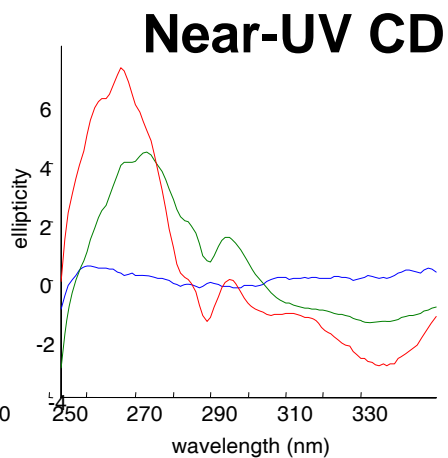
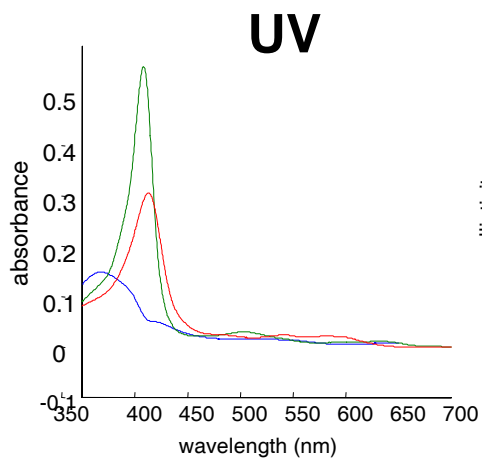
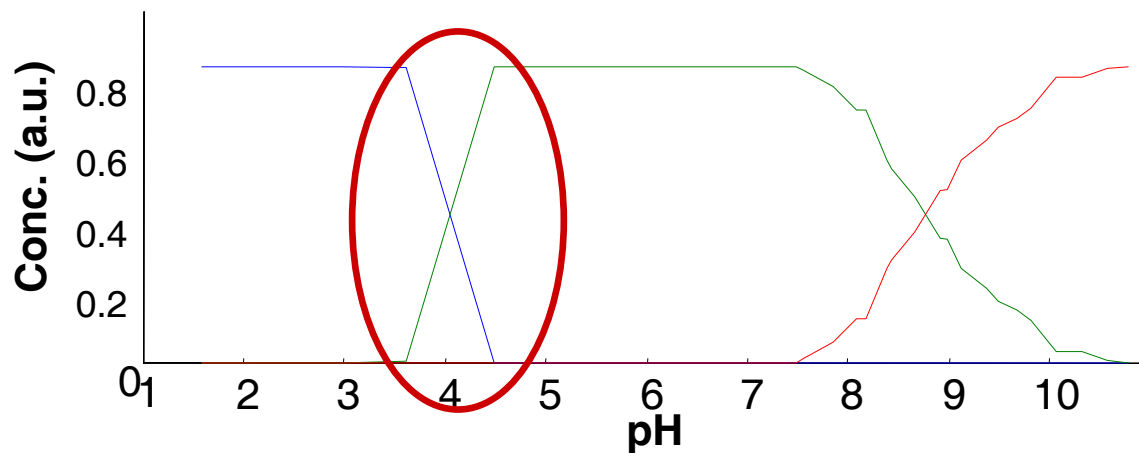


C: n-n, unimodality, closure
S: n-n (UV spectra)



pH-dependent conformations (steady-state)

Unfolded Mb (U) → Native deoxyMb (N) → OxyMb (O)



Mb folding takes place in one-step!!

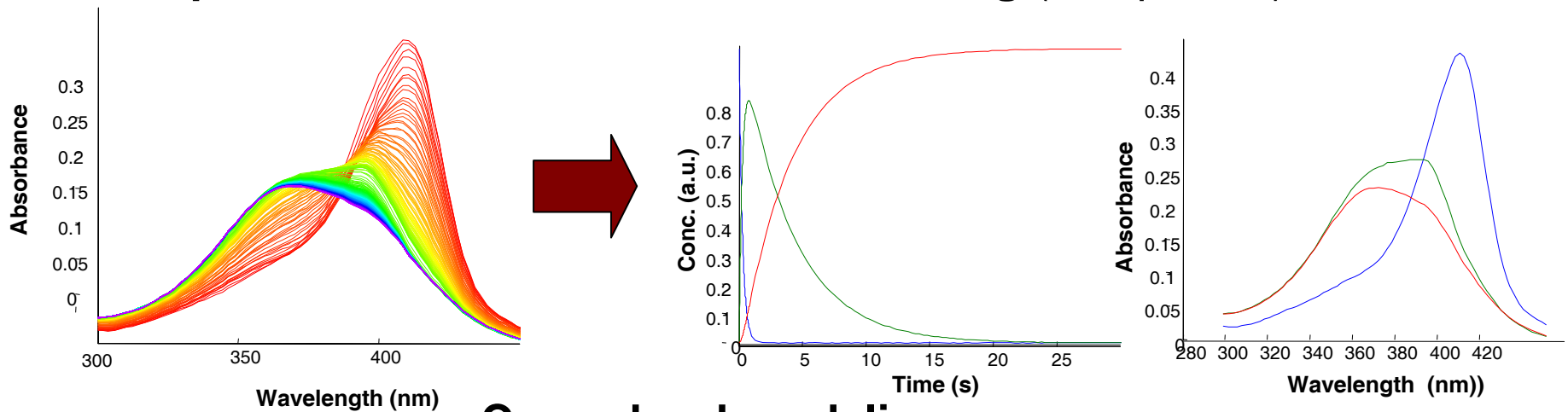
Time-dependent folding studies

pH-jump stopped-flow experiments

✓ Syringe 1: neutral protein solution

✓ Syringe 2: acid.

● Spectra collected over time after mixing (end pH < 3)



C: n-n, hard-modeling

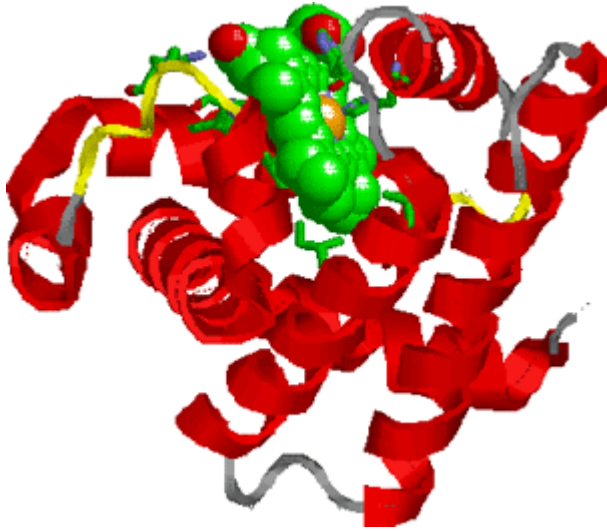
S: n-n

Native Mb (**N**) → Kinetic transient (**I**) → Unfolded (**U**)

Mb unfolding takes place in two-steps!!

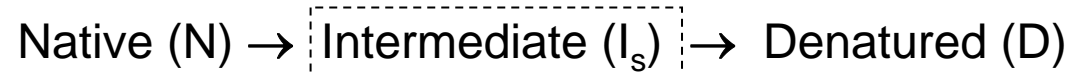


Myoglobin unfolding (multiexperiment analysis)



Mechanism

Steady-state process



Kinetic process

Steady-state process

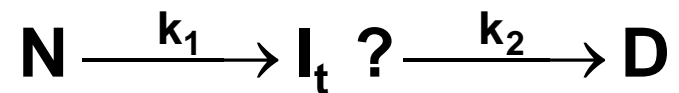
UV spectra, pH range 7.0-2.0



Unknown model

Kinetic process

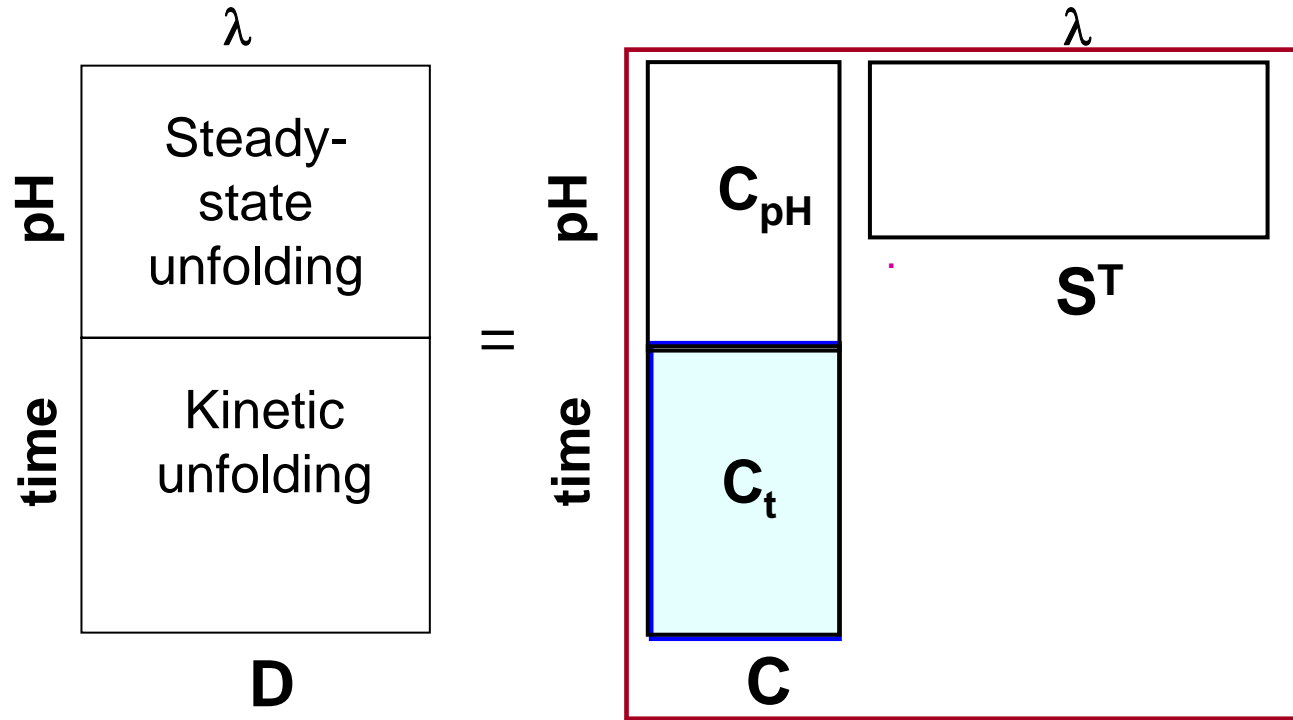
UV spectra, pH-jump stopped-flow



First-order consecutive reactions



Myoglobin denaturation

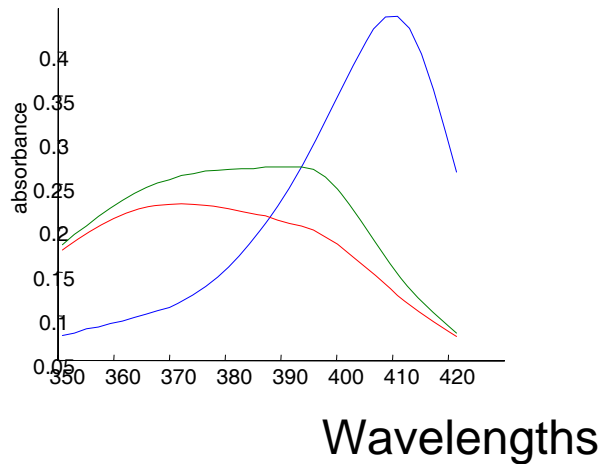
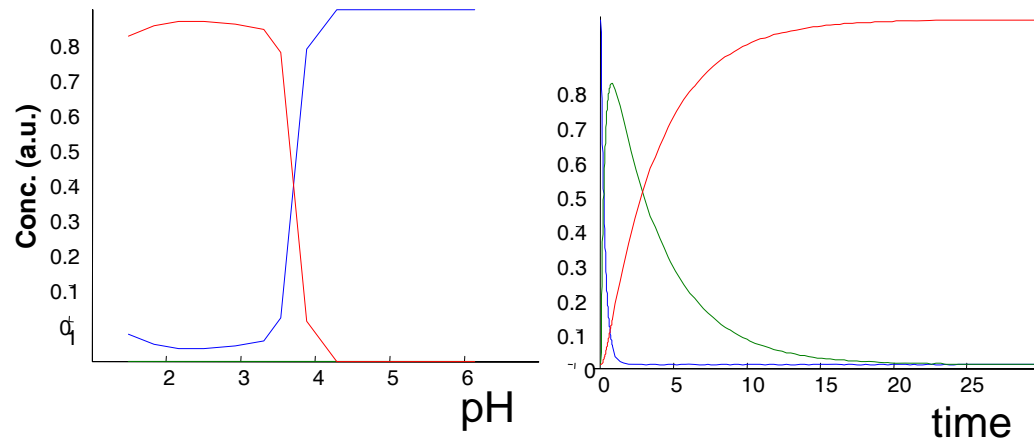


Hard-modelling (kinetic unfolding, 1st order reactions)

Soft-modelling constraints (n-n)

Model-free and model-based
experiments can be analyzed together.

Myoglobin denaturation



Steady-state conditions

Native (**N**) → Denatured (**D**)

Kinetic transient (**I_t**)

Kinetic process

- Formation of a kinetic transient was detected and hard-modelled.
 $k_1 = 4.05 \text{ s}^{-1}$ $k_2 = 0.62 \text{ s}^{-1}$
- No stable intermediates were found in steady-state conditions.



Conclusions

- **Multivariate spectroscopic monitoring and chemometric analysis** are a powerful combination for protein process analysis.
- **Multitechnique analysis** enhances the **robustness of process description** and allows for an **easier identification/discrimination** among protein species.
- **Multiexperiment analysis** provides a **comprehensive description of process mechanisms** based on all possible inducing agents and experimental conditions.



Acknowledgements

- Susana Navea, Romà Tauler, Patrick Cutler, Anna Borges, Glòria Muñoz, Raimundo Gargallo, Joaquim Jaumot.
- Financial support of Spanish government: project CTQ2006-15052-C02-01.

Web page

- **Software:** www.ub.es/gesq/mcr/mcr.htm