CD study of HEW lysozyme and the effects of low pH on its structure.

MANUEL ALMAGRO RIVAS

Biophysical Chemistry. Degree in Chemistry. University of Granada.

M. Almagro: malmriv@correo.ugr.es

Compiled February 12, 2021

In this study we analyse the near-UV spectra of hen egg-white (HEW) lysozyme in order to gather data about its structure. We develop a new model for secondary-structure analysis which outperforms existent models (like K2D3), and use the near-UV spectra of the protein at different pH to discuss how acidity might affect the protein's tertiary structure.

Keywords: **HEW lysozyme, HEWL, circular dichroism, secondary structure determination.**

CONTENTS

1. INTRODUCTION.

Circular dichroism is a widely used technique that allows researchers to gather data referring to both the secondary and tertiary structures of a protein. When it comes to secondary structure, a variety of methods exist that allow for the estimation of the composition of a protein in terms of common structures: α helices, β turns, β sheets, random coils, and so on. The tertiary structure cannot be determined so directly, but comparing several spectra of the same protein in different conditions (pH, ionic strength, temperature...) makes it possible to evaluate qualitatively the changes that might have taken place.

2. THEORETICAL BACKGROUND.

It is known that chiral molecules (those with carbon atoms with four different groups attached) have an effect on polarised light. For circular light, which can be seen as a combination of two components with the same period and a phase difference of $\pi/2$ radians, a delay is introduced so that the light becomes elliptically polarised. There exists a variety of ways to use this effect to study the structure of chiral molecules, but a common way of doing so is using *ellipticity* (θ) which relates the intensity of the two components of elliptically polarised light like shown in Eq. 1

$$\theta(\text{rad}) = \frac{\sqrt{I^-} - \sqrt{I^+}}{\sqrt{I^-} + \sqrt{I^+}} \quad \textbf{(1)}$$

Although it is known that concentration does affect the shape of the spectrum (Brahms & Brahms, 1980), the ellipticity thus measured can be normalised by the concentration of the solution and length of the cuvette used. Since small numbers are usually obtained, it is common to work with millidegrees instead of radians. Such magnitude can be then converted to Molar Ellipticity ($[\theta]$) like shown in Equation 2:

$$[\theta] = \frac{\theta(mdeg) \cdot P_M(g/mol)}{c(g/L) \cdot 10 \cdot l(cm)} \quad \textbf{(2)}$$

This magnitude can be studied in the whole range of UV radiation, but different domains convey very different types of information. The far-UV range of the spectrum corresponds to the aggregated effects of secondary structure elements of the

protein, while the near-UV range shows harder to interpret information about the tertiary structure.

A. Far-UV spectrum.

Some authors, like Brahms & Brahms (1980) have shown that most far-UV spectra can be reconstructed as linear combinations of those spectra reported by proteins with a single motif in their secondary structure. As a result, several tools are already available, like K2D3 (see Andrade-Navarro et al, 2012), which use computational approximations of single-motif peptide spectra to fit an empirically measured far-UV spectrum. Our initial attempt was to run our data through this service, but the results were very poor when compared with the empirical values from the literature (see Tabla 2 for more detail). Therefore, we decided to develop our own model using a similar approximation as Andrade-Navarro et al. (2012), but employing a different dataset of spectra and introducing several measures to account for individual differences between the spectra obtained by different authors. For more details about our model, see Section 3.

B. Near-UV spectrum.

The near-UV spectrum of a protein depends on minor interactions, especially those in which tryptophan, phenylalanine and tyrosine are involved. Disulfide bonds also play a role, as well as many other factors, and therefore it is difficult to obtain information from just one spectrum. Nevertheless, several spectra can be compared against each other to evaluate the possible changes in a protein's tertiary structure. In our case, we will work with the same protein under different pH values in the acid range.

3. SIMON: A NEW METHOD FOR SECONDARY STRUCTURE DETERMINATION.

Due to the bad quality of the estimation obtained with pre-made models, we decided to elaborate our own (***S**tructure **I**dentification by **M**eans of **O**ptimization of a **N**onlinear equation*, SIMON for short) in order to have as much control as possible over the whole process. We extracted data from the paper by Brahms & Brahms (1980). In their work, they obtained the CD far-UV spectra of several proteins with one main component in their secondary structure. Sperm whale myoglobin was used as the source for the α-helix spectrum; a poly(Lys$^+$-Leu-Lys$^+$-Leu) peptide was employed for the β-sheet spectrum; a poly(Pro-Lys$^+$-Leu-Lys$^+$-Leu) peptide was used for the random-coil spectrum; and a poly(Ala$_2$-Gly$_2$) peptide was used, amongst others, for the β-turn spectrum. These spectra can be seen in Figure 1. Each spectrum was interpolated using a smooth fit; this means that we will work from now on with four functions representing these spectra in the [190-250] nm range of the far-UV spectrum. Let these functions be named $f_{\alpha h}$, $f_{\beta s}$, $f_{\beta t}$ and f_r for α-helix, β-sheet, β-turn and random coil, respectively. One can then define a new function, $S(\lambda)$:

$$S(\lambda) = C_{\alpha h} \cdot f_{\alpha h} + C_{\beta s} \cdot f_{\beta s} + C_{\beta t} \cdot f_{\beta t} + C_r \cdot f_r \qquad (3)$$

Where the C_i, $i = \alpha h, \beta s, \beta t, r$ are free scalars. No further constraints are set for them, which is key in our model. Imposing that all of them must add up to one, for example, implies that researchers whose spectra do not match the exact range of $[\theta]$ observed by Brahms & Brahms (1980) will get a poorer prediction. Instead, we perform a least-squares fit of the $S(\lambda)$ function to obtain the best-fit values for the C_i coefficients. Then, each

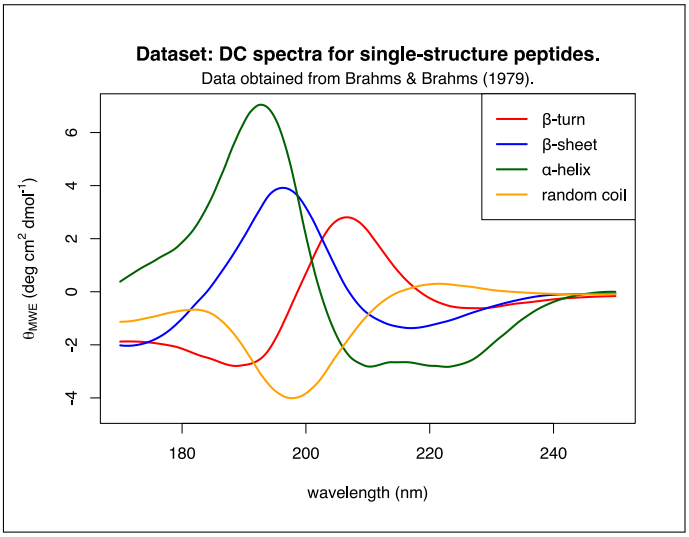

Fig. 1. Spectra of the aforementioned proteins used as representatives for each secondary-structure pattern. Data extracted and redrawn from Brahms & Brahms (1980).

coefficient is normalised like follows:

$$C_i(\%) = \frac{C_i}{\sum_k C_k} \times 100 \qquad (4)$$

Therefore obtaining the resulting percentage for each of the four considered structures. Our model was checked against empirical data, showing good agreement of the estimations and the real secondary structure composition of proteins.

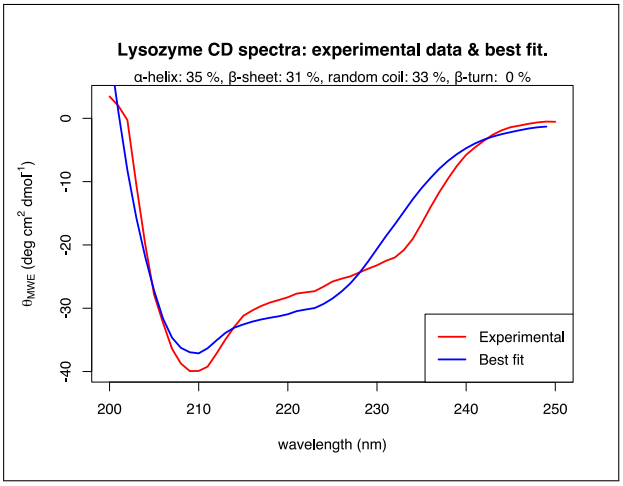

Fig. 2. SIMON's fit for the hen egg-white lysozyme spectrum at pH 3.0.

4. MATERIALS AND METHODS.

The CD spectra were supplied by Prof. José Cristóbal Martínez Herrerías. The near-UV spectra (250-350 nm) were obtained with a 0.5 cm thick cuvette and a HEW lysozyme concentration of 0.5 mg/ml. The far-UV spectra (190-260 nm) were obtained with a 0.2 cm thick cuvette and a concentration of 0.2 mg/ml of the same protein. Both spectra were measured at 25 ^{o}C. Both the solvent and the protein solution spectra were measured and

then scaled according to Eq. 2. Measurements were made at pH 3.0, 3.5, 4.0 and 4.5. We obtained the resulting spectra by substracting the solvent spectra to the solution spectra (see Fig. 4 and 3 for two examples of this).

5. RESULTS.

A. Far-UV spectra.

The far-UV spectra corresponding to four different values of pH (3.0, 3.5, 4.0, 4.5) were computed as mentioned. Since lysozyme is known to preserve its secondary structure elements in pH values much lower than those we have studied here (Babu & Bhakuni, 1997), we assume that the composition of secondary structure will remain constant in our pH range. Thus, it is valid to average the results obtained for the pH values handled here. These four spectra (the graph for pH 3.0 can be seen in Fig. 3 as a representative example) were analysed by our model. Fig. 2 shows the fit performed for the spectra at pH 3.0 as an example, although all of them were computed, but are not shown here due to lack of space. The results reported by our model are shown in Table 1. Using other tools, like K2D3 (see Andrade-Navarro

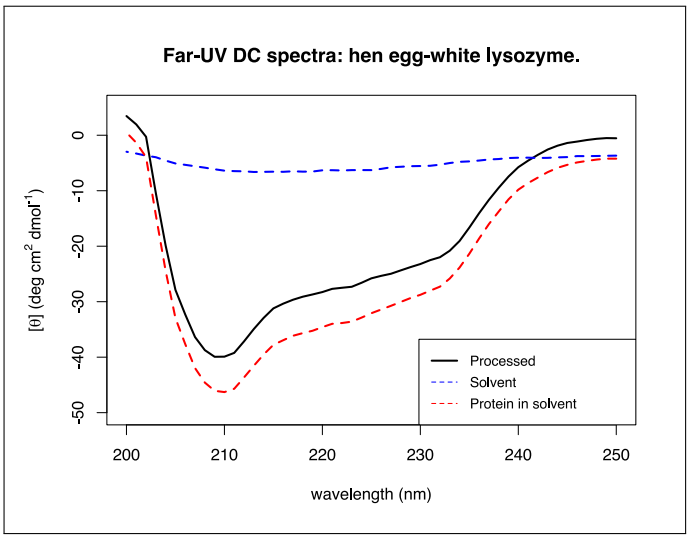

Fig. 3. CD far-UV spectra for HEW lysozyme at pH 3.0.

et al., 2012), the estimation was poorer and less detailed: it was estimated that the protein was comprised of an 86.45% of α-helix and 0.16% β-sheet. This is far from the true values observed empirically. A comparison of these results can be seen in Table 2

	α-Helix	β-Sheet	β-Turn	Random
pH 3.0	35	31	0	33
pH 3.5	40	15	4	40
pH 4.0	50	9	1	40
pH 4.5	35	27	3	34
Global	40±3	21±5	2±1	37±2

Table 1. Our model's results for different pH values. The uncertainty (Δ) was computed as $\Delta = \sigma/\sqrt{N}$, where σ is the standard deviation for each series and $N = 4$ is the number of datapoints.

	α-Helix	β-Sheet	β-Turn	Random
X-ray diffraction	45	19	14	37
SIMON	40±3	21±5	2±1	37±2
K2D3	86.45	0.16	-	-

Table 2. Comparison: results obtained with our model, K2D3 and X-ray diffraction data by Brahms & Brahms (1980). K2D3 did not yield any information about β-turns or random coil composition.

B. Near-UV spectra.

The near-UV CD spectra can be obtained as previously mentioned (see Fig. 4 for an example at pH 3.0). While no immediate conclusions can be drawn from a single spectrum, those for different pH values can be compared against each other to gain some insight. Figure 5 shows the near-UV spectra for several

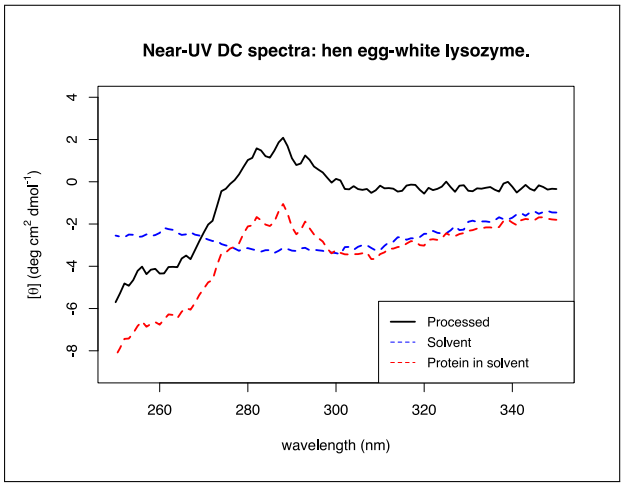

Fig. 4. CD near-UV spectra for HEW lysozyme at pH 3.0.

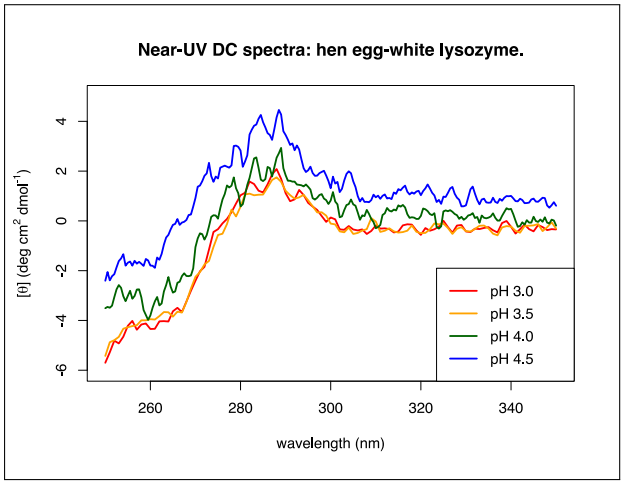

Fig. 5. CD near-UV spectra for HEW lysozyme at different acidic pH values.

pH levels. It is apparent that these spectra are more complex than those in the far-UV range. The spectra for the two lower pH values, 3.0 and 3.5 pH, practically overlap each other. The

two spectra for less acidic pH values maintain the same profile, but they take higher values over the whole range. This could imply higher optical activity dues to the relevant chromophores (tryptophan, phenylalanine and tyrosine) being more exposed to the solvent. That is, the protein changes its tertiary structure at some point between pH 3.5 and pH 4.5, but it is difficult to determine how exactly this happens. This seems to be in agreement with Babu & Bhakuni (1997), who report that HEW lysozyme has lost a significant amount of tertiary structure in very acidic pH values.

6. CONCLUSIONS.

Our method seems to work well for secondary structure determination. Nevertheless, it is obvious that the β-turn estimation (2 ± 1 % theoretical, vs. 14% experimental) was not very satisfactory. This is due, as Brahms & Brahms (1980) recognise in their own paper, due to the variety of potential spectra a β-turn based peptide can show. At least five different variations are shown in their paper, of which we only took one. It is possible thar our model could be improved by adding structures that have not been considered so far, but a more rigourous analysis should be performed in order to avoid overfitting). When it comes to tertiary structure, pH seems to have a marked effect on the protein: the more acidic the solution, the more tertiary structure is lost.

7. APPENDIX: SIMON'S VALIDATION.

Here are some other proteins which helped us learn about our method's ability to recognise secondary structure composition (Table 3). The three spectra have been taken from Lees et al. (2006), and the experimental composition has been taken from Brahms & Brahms (1980). Keep in mind that this is not the result of a thorough analysis or an estimation over various pH levels, but a single run for each protein. (See Figure 6 for an example).

	Model				Experimental			
	α-h	β-s	β-t	r.c.	α-h	β-s	β-t	r.c.
Rubredoxin	25	2	22	51	26	0	23	51
Cytochrome C	45	1	24	30	46	0	23	31
Prealbumin	10	57	12	20	0	60	10	30

Table 3. Some other proteins with which we have tested our model. The columns correspond, in order, to the percentual content in α-helices, β-sheets, β-turns and random coils.

8. BIBLIOGRAPHY.

1. Brahms, S., Brahms, J. (1980). **Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism**. Journal of Molecular Biology, 138(2), 149–178. doi:10.1016/0022-2836(80)90282-x

2. C., Andrade-Navarro, M. A., Perez-Iratxeta, C. (2012). **Prediction of protein secondary structure from circular dichroism using theoretically derived spectra**. Proteins: Structure, Function, and Bioinformatics, 80(12), 2818–2818. doi:10.1002/prot.24168

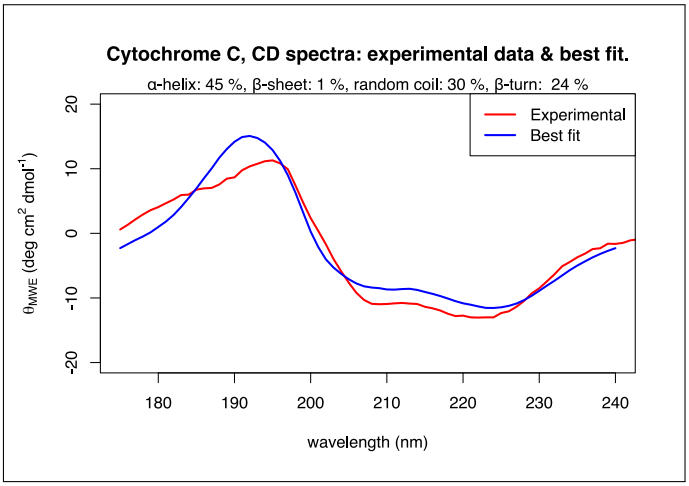

Fig. 6. Example: experimental and fitted CD spectra for Cytochrome C.

3. Babu, K. R., Bhakuni, V. (1997). **Ionic-Strength-Dependent Transition of Hen Egg-White Lysozyme at Low PH to a Compact State and its Aggregation on Thermal Denaturation**. European Journal of Biochemistry, 245(3), 781–789. doi:10.1111/j.1432-1033.1997.00781.x

4. Lees, J. G., Miles, A. J., Wien, F., Wallace, B. A. (2006). **A reference database for circular dichroism spectroscopy covering fold and secondary structure space.** Bioinformatics, 22(16), 1955–1962. doi:10.1093/bioinformatics/btl327