Research Article

Journal of PeptideScience

Received: 24 April 2009 Revised: 30 June 2009 Accepted: 1 July 2009 Published online in Wiley Interscience: 18 September 2009

(www.interscience.com) DOI 10.1002/psc.1169

The inherent flexibility of peptides and protein fragments quantitized by CD in conjunction with CCA+

Imre Jákli[a] and András Perczel[a,b]*

ECD spectroscopy is traditionally used for rapid, non-atomic level structure analysis of natural products such as peptides and proteins. Unlike globular proteins, peptides less frequently adopt a single 3D-fold in a time average manner. Moreover, they exhibit an ensemble of conformers composed of a multitude of substantially different structures. In principle, both ECD- and vibrational circular dichroism (VCD)-spectroscopy are sensitive enough to pick up structural information on these dynamic ensembles. However, the interpretation of the raw spectral data of these highly dynamic molecular systems can be cumbersome. The herein presented *C*onvex *C*onstraint *A*nalysis Plus method, or CCA+ for short (http://www.chem.elte.hu/departments/protnmr/cca/), provides a unique opportunity for spectral ensemble analysis of peptides, glycopeptides, peptidomimetics, and other foldamers. The precision and accuracy of the approach is presented here through different peptide model systems. An interesting temperature and pH dependent folding and unfolding of a miniprotein (e.g. Tc5b variant) is also described. Analysis of CD spectra sets strongly affected by solvent and ion type is also introduced to account for severe environmental-induced structure influencing effect(s). The deconvolution makes always possible the quantitative data analysis even when the interpretation of the deconvolution resulted in pure CD curves is complex.

Keywords: peptide; structure determination; circular dichroism; spectral deconvolution; secondary structural content; inherent flexibility; foldamer

Introduction

UV or ECD spectroscopy is traditionally used for rapid, non-atomic level structure analysis of peptides and proteins. For proteins their secondary structure content is commonly estimated and their structural changes are regularly monitored through ECD. Post-experimental data analyzing methods are well established and commonly used for the semi-quantitative interpretation of the implicit structural information. On the contrary, for peptides the same type of ECD-based structure elucidation suffers from inherent limitations. Unlike globular proteins, peptides less frequently adopt a single 3D-fold in a time average manner. Moreover, they exhibit an ensemble of conformers composed of a multitude of substantially different structures. Even longer polypeptides, miniproteins (e.g. Tc5b [1,2], betanova [3], or cyclic oligopeptides present a 3D-fold of increased dynamics detectable on several timescales of motion. In principle, both ECD- and vibrational circular dichroism (VCD)-spectroscopy are sensitive enough to pick up structures forming these dynamic ensembles. However, the interpretation of the raw spectral data of these highly dynamic molecular systems can be cumbersome. Unlike for proteins, where the secondary structural elements provided typical CD-reference curves are well defined, for peptides a similar reference spectra arsenal is not yet uncovered. Thus, for proteins by either using a preset ensemble of reference CD curves (methods based on any linear combination type strategy), or applying an *in situ* deconvolution, the way is opened for a quantitative or semi-quantitative CD-based structure elucidation. In contrast to the above, for oligo- and polypeptides, both the conformational properties and the spectral features of the different secondary structural elements are less uniform. In fact, they are substantially different, presenting 'distortions' on a large scale. Thus, it seems unlikely that any set of reference- or base-CD curves determined *a priori* can accurately and uniformly be presented and used for structural elucidation of peptides. The only general solution is to apply an '*in situ*' deconvolution on the spectral ensemble of the investigated molecular system. In this way, one could generate the right set of pure component CD curves, adequately describing the spectral set of interest for analysis. The herein presented *C*onvex *C*onstraint *A*nalysis Plus method, or CCA+ for short, provides a unique opportunity for spectral ensemble analysis of peptides, glycopeptides, peptidomimetics, foldamers, etc. Besides their structural variability, temperature, solvent type, ionic strength, temperature, etc. significantly influence their spectral outcome. This is because there is no chance to present a uniformly applicable pure CD curve set. However, such deconvolution type analysis can handle complex problems and provide a quantitative

* *Correspondence to: András Perczel, Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary. E-mail: perczel@chem.elte.hu*

a *Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary*

b *Protein Modeling Group HAS-ELTE, Institute of Chemistry, Eötvös Loránd University, H-1538, Budapest, P.O. Box 32, Hungary*

Abbreviations used: *Aad, α-aminoadipicacid; Aca, ε-aminocaproicacid; Ava, δ-aminovalericacid; ECD, Electronic Circular Dichroism; IPP, ion per peptide ratio; SDK, software development kit; TFE, trifluoroethanol.*

Journal of PeptideScience

answer, even if some of the deconvoluted pure component CD curves are hard or impossible to assign to structure(s).

Peptide chemists are continuously searching for foldamers (e.g. type I β-turn, α- or 3_{10}-helix) which could result in characteristic CD curves. Recently, Reed and Reed [4] reported a set of reference curves (for the PEPFIT program) for nine typical secondary structural elements based on carefully selected peptide models. Among others [5–7], we also have made attempts at finding peptide sequences which provide a single and time average structure [8,9]. For example, cyclo[Gly-Pro-Ser(OtBu)-Gly-Ava) turned out to be an excellent and almost ideal model for the type I β-turn as analyzed by X-ray crystallography [10]. However, even such a model which presents an ideal type I β-turn in the solid state, in solution the very same molecule possesses a conformational mixture as demonstrated by NMR-spectroscopy. Thus, the recorded CD curve can hardly be attributed to the type I β-turn reference CD spectrum. Moreover even these, close to ideal molecular structures, are subjected to backbone distorting 'environmental' effects, such as pH, ionic strength, and intermolecular interactions. Because of these and additional problems, the strategy of 'hunting for rigid structural building blocks' should be altered and the approach suggested here could to be taken. Namely, the *in situ* deconvolution and analysis (see Materials and Methods Section) of the spectral data set of relevance. As there is no universal CD spectral data set of foldamers making up peptides in general, they should not be chased.

The first ECD analysis programs were made rather simple, by using a command line interface according to the computer environment available at that time ([11,12] etc.). Nowadays graphical interface for a user-friendly program is a requirement. There are several ECD analyzer programs that run on a desktop computer (DICHROPROT [13], CDNN [14], CDTool [15]) but an increasing number of internet-based services also become available (K2D [16], K2D2 [17], DICHROWEB [18]). DICHROPROT [13], CDTool [15], and DICHROWEB [18], ACDP [19]). Some of them integrate several command line based methods and utilities for data manipulation. Several approaches were developed in the past to analyze CD spectrum: ridge regression [20], singular value decomposition [12,21], variable selection [22], locally linearized model [23], cluster analysis [24], self-consistent method [25–27], principal component factor analysis [28], neural network [14,16,17,29,30], matrix descriptor [31].

The reference set obtained from such a deconvolution is again not universal, but can provide a good description of the particular set of CD curves, enough in most cases. The major drawback of such analysis is that the resulting pure CD curves must be assigned each and every time, based on their shape or by any additional information (e.g. NMR data). Our CCA+ deconvolution algorithm is quick and efficient to fulfill such a task and the user has a single remaining task to complete, namely, the interpretation of the out come. In this article, we are showing the performance of the algorithm, the precision and accuracy of the approach taken as well as its versatile nature. We are focusing here on the analysis of CD spectra sets strongly affected by solvent and ion type. It will be shown that deconvolution makes always possible the quantitative data analysis, even when the interpretation of the pure CD curves is complex or ambiguous.

Materials and Methods

Most spectrum analyzing methods are based on the simple assumption called as the *additivity* rule (Eqn. 1), namely, that an ECD spectrum can be calculated by summing up the contributions of the different spectral properties:

$$ECD(\lambda) = \sum_{k=1}^{N} f_k B_k(\lambda) \tag{1}$$

where $ECD(\lambda)$ is the ECD spectrum as a function of the wavelength, f_k is the fraction of the $B_k(\lambda)$ base spectrum. The constraints are $\sum f_k = 1$ and $f_k \geq 0$. It is also assumed that different secondary structure elements have a specific spectral shape. The CCA deconvolution algorithm by Perczel *et al*. [11] minimizes the following expression (2):

$$\left[\sum_{j=1}^{N} ECD_j^m(\lambda) - \sum_{j=1}^{N} ECD_j^c(\lambda)\right]^2 = \left[\sum_{j=1}^{N} ECD_j^m(\lambda) - \sum_{j=1}^{N}\sum_{i=1}^{P} f_{i,j} \cdot B_i(\lambda)\right]^2 \tag{2}$$

where ECD^m is the set of recorded ECD spectra (of N elements), ECD^c is the calculated spectral ensemble, determined from the linear combination of B_i base curves each with $f_{i,j}$ coefficients. The number of base curves (P) is an *a priori* parameter. The sum of linear combination coefficients is:

$$\sum_{i=1}^{P} f_{ij} = 1, \quad \text{where} \quad j = 1, 2, \ldots, N \tag{3}$$

and

$$f_{ij} \geq 0, \quad \text{where} \quad i = 1, 2, \ldots, P \quad \text{and} \quad j = 1, 2, \ldots, N \tag{4}$$

Because of the 'convex' boundary condition, during the minimization, all $f_{i,j}$ coefficients should be within a minimal volume of a simplex, defined by the base curves as vectors in the Euclidean space. (Note that the simplex is a line segment in 1D-, a triangle in 2D-, and a tetrahedron in the 3D-space.) This method is very rapid and provides a fully quantitative result. It can also depict rather small spectral changes and it is able to monitor structural shifts in a quantitative manner, just from the measured raw spectral data (e.g. set of ion titration, temperature scans). The simplex is used here in two contexts: (i) the simplex is considered as the third restraint for minimization (beside those of Eqns. 3 and 4), (ii) the 'representation' of the vector space defined by the 'pure component-spectra' vectors. All analyzed 'spectra' as a 'point' should be within this simplex after minimization as determined by the deconvolution process. The inherent properties of the simplex are closely related to our convex boundary conditions: If the corners are the pure/base components, then all points within the simplex represent the linear combination of the pure components, like in a barycentric view. The distance from the corners will indeed provide the component percentages. If we use this picture, the deconvolution process fits the simplex with the smallest possible volume around all data 'points', the latter ones defined by the spectra. When optimization completed, all calculated points satisfy the convex boundary conditions.

The CCA+ program was written in Windows™ 32-bit SDK development environment so it runs on desktop computers using the Windows™ operating system. It implements features

necessary for the data processing: input preparation, data import and minimal data manipulation. Both the Linear Combination of spectra (Lincomb) and the CCA+ method can be used for ECD spectral analysis, whereas the test generator, an integrated part of the program, helps the validation and benchmarking of the algorithm.

The main screen has two fields: The upper part is the spectrum window, whereas the lower part contains the list of spectrum or result components. For CCA analysis, first an input set of spectra (data matrix) has to be created by selecting spectra of choice. Second, the appropriate method (CCA+) and the input parameters should be set. The deconvolution procedure, which is an iterative method lasting only for a couple of seconds, results in both the calculated pure spectra (a set of B_i) and also their weights or linear combination coefficients ($f_{i,j}$) (Eqn. 2). To verify and visually check the results, the originally recorded raw (ECD^m) and the back calculated (ECD^c) spectra can be superimposed and presented in the result file. A detailed help and a step by step tutorial helps the user to learn the different features of the program as well as the first steps of the deconvolution method. The current version of the program is available free of charge via download, after registration, from http://www.chem.elte.hu/departments/protnmr/cca/.

The first spectrum library composed CDs of a miniprotein, namely, those of Tc5b-D9Aad. A total of 68 UV-CD spectra were recorded in the temperature range of 5–85 °C (temperature jump by 5 °C) at the following pH values: 2.5, 4.5, 5.5, and 6.5. As the original peptide of Neidigh *et al*. [1], that of Tc5b-D9Aad (H-^{1}NLYIQWLK-Aad-^{10}GGPSSGRPPP^{20}S-OH also exhibits a composition of different secondary structural elements (Aad stands for α-amino adipic-acid replacing Asp^{9} of the original sequence.) Owing to its relatively small size, it can easily be thermodenaturated especially at lower pH. This example stands for longer polypeptides having a compact structure at lower temperature, losing it relatively easily at increased T.

Most peptides could contain shorter fragments of different secondary structural elements, such as short α- or 3_{10}-helices, β-sheets (e.g. β-nova peptides [3]) and hairpins (e.g. SETE [32–34]). Typically, the shorter the sequence the larger its flexibility and plasticity is. One particular secondary structure often adopted by shorter sequences is β-turns of different type. In the second example, we are presenting a set of peptide sequences which form different β-turns (type I, type II, etc.) in form of conformational equilibrium, easily influenced by their amino acid composition [9,35,36] and by the solvent used: water (H_2O), trifluoroethanol (TFE), and acetonitrile (CH_3CN or ACE). The dielectric constants of TFE and CH_3CN are relatively low and somewhat mimic the molecular environment β-turns buried in the core of a larger protein. However, H_2O with $\varepsilon \sim 78$ (at 25 °C) can stand for turns located on or close to the surface of a protein. This spectrum library contains a total of 88 spectra of 30 different linear and cyclic oligopeptides (Table 1). The size of the model compounds varies from three to six amino acid residues, both linear and cyclic model peptides were introduced, with polar and apolar side chains.

β-turn model peptides introduced above can specifically and aspecifically bind metal ions. Especially aspecific ion binding of backbone carbonyls introduced conformational shifts are of general interest, as they may occur at any circumstance. In our third example, we are presenting data on the effect of the ion binding of β-turn sequences. Besides the biologically most common Ca^{2+}, the influence of the Al^{3+} ions on the conformation is also studied. The latter ion, Al^{3+}, is often claimed to be involved in the aggregation of some neuropeptides connected to Alzheimer disease [37]. This peptide library contains both four pseudo cyclo-hexapeptides, namely, cyclo[(δ)Ava-Gly-Pro-Xxx-Gly] where Xxx=Ser, Ser(O^{t}Bu), Thr and Thr(O^{t}Bu), and additional model systems of interest: cyclo [Pro-Ala-(ε)Aca], cyclo [Pro-Ala-(ε)Aca-Pro-Ala-(ε)Aca] [10,35,36]. A total of 73 CD curves were recoded in the 185–260 nm spectral range, using physiologically relevant 1–5 equivalent (IPP) ion concentration in a low dielectric media (TFE).

Results and Discussion

Precision and Accuracy

To show the efficiency of the algorithm for structure elucidation of dynamic peptides, first we are to demonstrate its performance on controlled data sets. Based on the additivity rule (Eqn. 1), we have generated 'controlled' CD data sets using the linear combination of typical CD spectra. These controlled spectral ensembles are realistic in the sense that both the shape of the component CD curves and their weights are similar to those of peptides and proteins. Nevertheless, they are labeled as 'controlled' ones as we know *a priori* the outcome of the analysis, as 'mixing' was completed in a controlled manner. Using the four CD curves of Poschner *et al*. [38], namely, that of α-helix, β-sheet, general β-turn, and unordered, provided in the peptide training set created for CDNN [14], we can generate any 'controlled' set of peptide CD libraries (Figure 1). Following the deconvolution one has to check, both the spectral properties of the resulted in base curves by correlating them with those of Poschner's, as well as the similarity of the deconvoluted spectral weights with those set initially (Figure 2). (Very similar tests were previously carried out successfully for proteins [11,39].)

The benchmark test for peptides was completed for two different types of CD data sets, but for both cases the Poschner's four pure component CD curves were used. The differences between the two runs are in the manner how the spectral coefficients are generated initially. During the first one (hereafter referred as *random*), weights of each spectrum have been selected in a completely independent and random way and subsequently normalized to make the sum of them equal to the unity. However, according to the second protocol, the weight of one of the four components was 'enhanced' (from herein we use the terminology enhanced) in the following way: this weight (f_1) was selected randomly between 0% and 100%, while that of the second one between 0% and (100 $-$ f_1)%, etc. Thus, the maximal value of the forthcoming components (f_n) can only be (100 $-$ $\sum_{k=1}^{n-1} f_k$)%. Generating the conformational weights according to the second 'randomizing protocol', one selected component is clearly enhanced. To make all components equally 'preferred', components were randomly reshuffled. To make the difference visible between the two 'random selection' strategies, two data sets of 20 spectra were made visible on 3D-barycentric plots (note the difference in 'dot' distribution on Figure 3a and c). The very pure components are located at the corners of these tetrahedrons (Figure 3b), whereas a dot inside the simplex corresponds to the actual weights used for linear combination of the component pure CD curves. The mixing coefficient is the distance of a point from the plane opposite of a corner.

The spectral set generated according to the first so-called *random* manner stands for the very 'realistic' situation, namely, where neither the pure component spectrum nor its 'close neighbor' is explicitly among the curves to be deconvoluted.

Journal of PeptideScience

Table 1. Deconvolution resulted in component percentage (all values in %) of the CD curves of both linear and cyclic peptide models, recorded in water (H_2O), acetonitrile (ACE), and trifluoroethanol (TFE)

Peptide[a]	H_2O				TFE				ACE			
	Rnd[b]	βI-III	β-II	βIII-I	Rnd[b]	βI-III	β-II	βIII-I	Rnd[b]	βI-III	β-II	βIII-I
Ac-Pro-Ser-NHMe	42	5	34	19	35	7	36	21	29	9	45	18
Boc-Ile-Thr-NHMe	27	3	47	23	26	2	41	31	25	1	39	36
Boc-Leu-Thr-NHMe	22	1	51	27	25	1	46	28	23	1	42	34
Boc-Pro-Asn-NHMe	33	4	46	18	27	0	36	37	28	3	31	38
Boc-Pro-Asp-NHMe	33	5	44	19	33	5	31	30	28	5	40	27
Boc-Pro-Gln-NHMe	33	3	47	17	29	2	41	28	31	3	37	29
Boc-Pro-Glu-NHMe	32	2	47	19	30	3	40	27	28	5	41	26
Boc-Pro-Ser-NHMe	33	2	46	19	29	3	45	23	26	4	45	25
Boc-Pro-Thr-NHMe	34	1	46	18	33	3	39	25	30	3	41	26
Boc-Val-Asn-NHMe	24	0	61	15	–	–	–	–	21	2	34	43
Boc-Val-Asp-NHMe	23	2	46	28	27	3	36	34	23	3	38	35
Boc-Val-Gln-NHMe	29	2	51	18	25	1	45	29	24	1	49	27
Boc-Val-Glu-NHMe	31	2	46	21	24	2	47	27	31	5	39	25
Boc-Val-Ser-NHMe	29	2	46	23	25	2	48	26	24	2	46	29
Boc-Val-Thr-NHMe	29	3	44	24	26	1	43	29	25	0	40	35
Piv-Pro-Ser-NHMe	41	3	37	19	30	3	46	22	30	1	51	17
Average:	**31**	**2**	**46**	**21**	**28**	**3**	**41**	**28**	**27**	**3**	**41**	**29**
Standard deviation:	*5*	*1*	*6*	*4*	*3*	*2*	*5*	*4*	*3*	*2*	*5*	*7*
Boc-Pro-D-Ser-NHMe	16	1	74	9	18	5	76	2	10	5	84	1
Boc-Val-D-Ser-NHMe	14	2	78	6	14	3	79	4	13	3	84	0
Average:	**15**	**2**	**76**	**8**	**16**	**4**	**77**	**3**	**12**	**4**	**84**	**0**
Standard deviation:	*2*	*1*	*3*	*2*	*3*	*1*	*2*	*1*	*2*	*2*	*0*	*1*
Cyclo[Pro-Ala-Aca]	8	40	24	27	0	33	33	34	4	35	36	24
Cyclo[Pro-Ser-Aca]	12	25	42	21	19	7	57	17	15	19	48	17
Cyclo[Pro-Ser(OBzl)-Aca]	14	36	35	15	10	33	42	15	11	39	45	6
Cyclo[Pro-Thr-Aca]	20	37	0	43	–	–	–	–	21	31	3	45
Cyclo[Pro-Thr(OBzl)-Aca]	10	40	47	3	0	55	45	0	5	35	55	5
Average:	**13**	**36**	**30**	**22**	**7**	**32**	**44**	**16**	**11**	**32**	**37**	**19**
Standard deviation:	*5*	*6*	*19*	*15*	*9*	*20*	*10*	*14*	*7*	*7*	*20*	*16*
Cyclo[Gly-Pro-Ser-Gly-Ava]	25	2	58	15	34	7	41	18	36	8	43	14
Cyclo[Gly-Pro-Ser(O^{t}Bu)-Gly-Ava]	27	9	40	25	41	4	42	14	52	6	22	20
Cyclo[Gly-Pro-Thr(O^{t}Bu)-Gly-Ava]	30	11	38	22	41	4	40	16	57	0	21	23
Average:	**27**	**7**	**45**	**21**	**39**	**5**	**41**	**16**	**48**	**5**	**29**	**19**
Standard deviation:	*2*	*4*	*11*	*5*	*4*	*2*	*1*	*2*	*11*	*4*	*12*	*4*
Cyclo[Pro-Ala-Aca]$_2$	41	12	25	22	30	15	27	29	61	16	0	24
Cyclo[Pro-Ser-Aca]$_2$	36	16	24	24	30	12	39	19	39	15	21	24
Cyclo[Pro-Ser(OBzl)-Aca]$_2$	46	34	17	4	59	33	0	8	43	38	19	0
Cyclo[Pro-Thr(OBzl)-Aca]$_2$	55	19	20	7	57	22	12	9	49	17	16	18
Average:	**44**	**20**	**21**	**14**	**44**	**21**	**19**	**16**	**48**	**22**	**14**	**16**
Standard deviation:	*8*	*10*	*4*	*10*	*16*	*9*	*17*	*10*	*9*	*11*	*10*	*11*

The molecules are sorted into five groups (from up to down): linear polypeptides, D-amino acid containing linear models, 4-, 6-, and 8-member ring containing cyclic peptides. The average and the standard deviation of the component percentage for each molecule group are reported (figures with bold and italic face).

[a] For model selection, see refs 9,35,36.

[b] **Rnd**; atypical or unordered structure, **βI-III;** the sum of type I and III β-turns, **βII;** type II β-turn content, **βIII-I**, the sum of type I and III β-turn component.

Clearly in this case all points are located nearby the middle part of the tetrahedron (see Figure 3a). On the contrary, spectral data set generated by the 'enhanced' approach results in a close to 'ideal' spectral ensemble to be deconvoluted, as among the CD curves one or more of the pure CD components are present. (Note that one or several points are located nearby the corner(s) of the simplex (see Figure 3c). For most highly dynamic peptides, a CD spectral set resembling to the former case is expected to occur, at any conditions.

One clear goal of such a test run is to determine the total number of CD spectra, *N*, required for a successful deconvolution (Figure 2). (Success is reported in terms of R^2.) We have completed

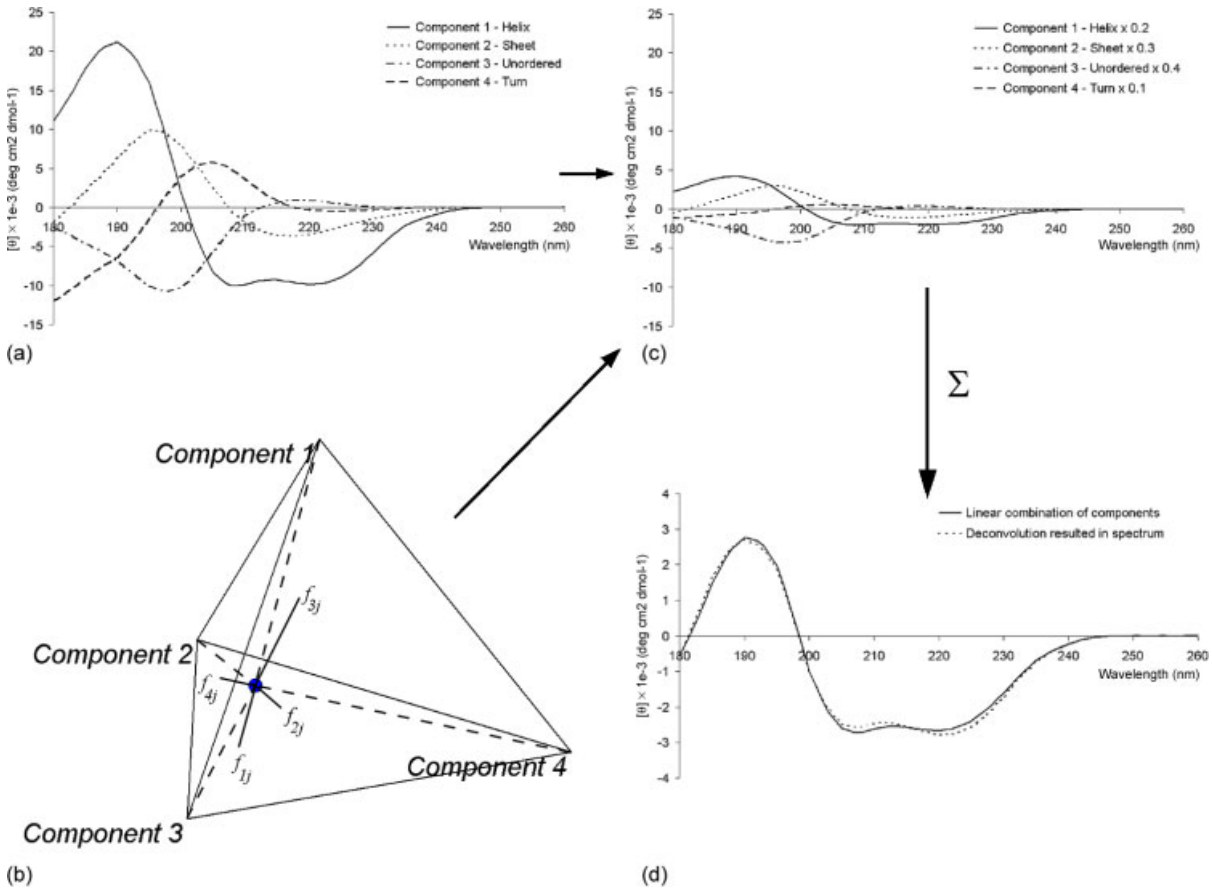

Figure 1. The construction of a typical element of the 'controlled' data set: (a) The four base-CD curves of Poschner *et al*. [38], used here to generate this data set. (b) Representation of the four mixing coefficients (f_{1j}, f_{2j}, f_{3j}, and f_{4j}, j = 1-N) in a tetrahedron/simplex. (c) The four base-CD curves multiplied by these four mixing coefficients. (d) Summing up these base curves provides a CD spectrum (solid line) of the 'controlled' data set. The deconvolution resulted in appropriate CD curve (dashed line). This figure is available in colour online at www.interscience.wiley.com/journal/jpepsci.

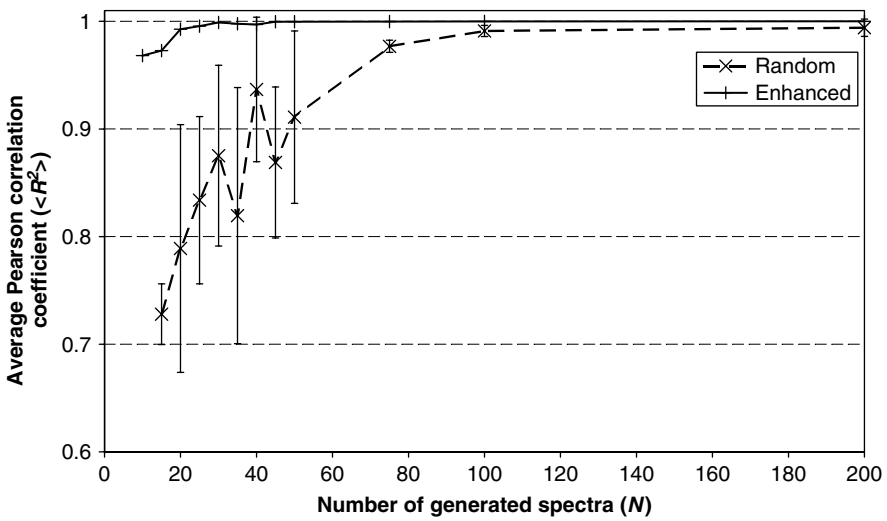

Figure 2. The average Pearson correlation coefficients ($<R^2>$) of the *a priori* generated and the deconvolution resulted in component percentages as function of the number of CD curves (N) in the data set. Both for the 'enhanced' (+) and the 'randomly' (X) selected coefficient methods (see text for more) were tested.

Journal of PeptideScience

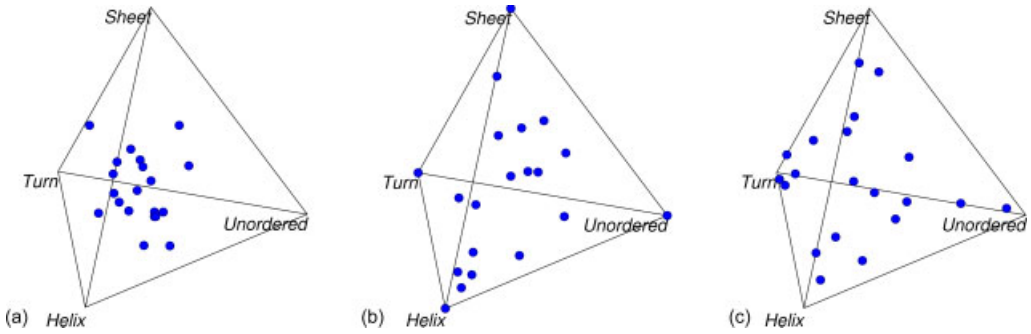

Figure 3. Graphical representation of three sets of component percentages, by using tetrahedral plots (created with CSpace [40]), all for a spectrum ensemble composed of 20 CD curves. First tetrahedron (a) contains a completely random coefficient selection, the second one (b) contains coefficients recommended by Poscher *et al*. [38], the last (c) represents a pure component 'enhanced' set. This figure is available in colour online at www.interscience.wiley.com/journal/jpepsci.

runs from $N = 15, 20, 25$, etc. up to 200. (This means, that by taking Poschner's four pure CD curves and by generating suitable weight coefficients, a total of 15, 20 25, etc. spectra containing data sets were generated and subsequently deconvoluted.) For error and statistic purposes, data generation was repeated five consecutive times, both for the 'random' and for the 'enhanced' cases. Finally, a single data set containing 20 'controlled' CD spectra was also calculated not randomly, but by using the coefficients reported by Poschner *et al*. [38] (Figure 3b). (We hereafter refer this test run as 'CDNN', as this input data are the same as the one used as training set for the CDNN program of Poschner *et al*. [38].) As mentioned earlier, the success of the 'run' is measured by the similarity of the initial and of the deconvoluted weight matrices, reported in term of *Pearson* correlation coefficient: $0 \leq R^2 \leq 1$.

By plotting the *Pearson* correlation coefficients against the number of individual CD curves (***N***) (Figure 2) show that if N is large enough ($N > 100$), then R^2 reaches 1, reporting the perfect identity of the generated and deconvoluted spectrum weights. (As five parallels, for all data sets were deconvoluted, the average R^2 values with their standard deviations are plotted (see error bar on Figure 2). Even though both data generation techniques, random and enhanced, result in a successful deconvolution at 'high' N, the way of their convergence to 1 is different. For datasets where special emphasis ensures that nearby regions of the 'corners' or in other words close to pure component curves are among the spectra to be deconvoluted (Figure 3c), full success is obtained at $N \sim 30$ and above (black line in Figure 2). Thus, in such a fortunate situation, where, for example, the CD curve of a pure secondary structural element is explicitly present in the spectral ensemble, deconvolution can be completed even by using a few CD curves only. However, as there is no foolproof method to determine in advance whether such a 'lucky situation' holds, as described above data were generated for the less advantages cases, called as the fully *random* situation (dashed lines in Figure 2 and Figure 3a). Clearly, the algorithm succeeds even in this case, however, it requires many more CD spectrum to be used for 'extraction'. As we see in the Introduction section, peptides are typically conformationally flexible entities, thus exhibiting multiple conformers in a time average manner. Therefore, their CD curve typically reflects to a conformational mixture and thus, very different conditions (T, pH, co-solvent, ionic strength, etc.) should be probed and a large set of different CD curves should be assembled for a successful analysis. (Here the name 'random' reflects to the uncertainty of knowing whether the pure component curve is among the CD curves to be analyzed.)

Finally, in a 'random' data set there is a chance, that nearby the corners of the simplex no dots are localized. Thus, the calculated simplex cannot unambiguously be restored, especially if N is too small (Figure 2). For example, for the ensemble of 20 'controlled' spectra, the average R^2 varies strongly, depending on which data generation method is used. For the 'enhanced' case $R^2 = 0.963$ and for the 'CDNN' peptide training set, $R^2 = 0.946$, deconvolution is completed well. However, for the 'random' method generated test set R^2 is significantly lower; 0.789. However, even in the latter situation it is not the CCA+ algorithm which is performing less well, but the nature of the data set makes unable to ensure a perfect deconvolution! In this case, there are many different possibilities to fit the simplex on the dots, among which some fit resembles more (higher R^2) and some less to the original one (*cf.* the high error bar at $N = 20$ gray line Figure 2). Nevertheless, by increasing the number of CD curves R^2 increases and around $N = 70$ R^2 goes above 0.9! There is no guarantee that the increase in size of the dataset will help in general, but there is a chance that as for the test case in a real world one can find such a condition, where pure component(s) could become prevalent.

In practice, the CCA+ method can be used for analyzing ECDs and VCDs of peptides with high confidence. However, as for most cases the close to 'ideal' situation will not hold. Thus, it is strongly recommended to use spectral data set as large and as diverse as possible. Fortunately, the algorithm is powerful enough to complete deconvolution fast, within few seconds even for data sets comprising few hundred of individual CD curves. The hereafter examples shows how deconvolution can be completed for real cases and how the analysis of flexible peptides can be used for structure elucidation.

Peptide Spectrum Libraries

It is understood [1] that even shorter polypeptides, which can adopt self-interacting secondary structural elements, and thus can shield hydrophobic resides in their core, could be called as proteins. These miniproteins, especially those not containing disulfide bonds, look ideal molecular systems for testing thermo- and pH-stability of foldamers. As our first case, the ECD spectrum library of miniprotein Tc5b-D9Aad is analyzed. A total of 68 CD spectra were recorded spanning a broad temperature and pH range (from 5 °C to 85 °C) and pH (2.5, 4.5, 5.5, and 6.5), to decipher its fold stability. Following the deconvolution, the

spectral ensemble (Figure 4) (see Materials and Methods) seems to contain only two distinct spectral components (Figure 5). Monitoring their interconversion as function of pH and T can help us to quantitatively monitor the complex procedure of folding and unfolding of Tc5b-D9Aad. Although it is known from our NMR structural data (to be published) that this miniprotein contains perhaps three different secondary structural elements (e.g. α-helix, 3_{10} helix, poliproline-II [PPII] and unordered), due to the ECD spectral similarity of the first and of the second two categories, the analysis by using two pure components holds.

The first pure component curve (Figure 5) can be attributed to the 'unordered' or dynamically averaged ECD curve, in conjunction with PPII. The second component corresponds to the 'ordered' contribution, dominated in this miniprotein by the α-(a total of 7–9 residues) and by the 3_{10}-(3–5 residues) helices. The continuous interconversion between the 2-folded states can be monitored by CCA+. At low temperature ($T = 5\,^{\circ}$C) and at physiological pH (pH = 6.5), the miniprotein is folded the most. As expected, with the increase of the temperature the relative contribution of the 'disordered' part increases monotonically (Figure 6) up to approximately 80%, signaling that at increasing with T the miniprotein continuously unfold. Furthermore, the same miniprotein shows a clear stability decrease with pH. It is not yet fully uncovered, whether only the Aad9. . .Arg16 salt bridge is weakened by dropping the pH, or additional structural factors also play in a role. Nevertheless, at pH approximately 2.5 Aad becomes protonated, the 3_{10}-helix stabilizing salt bridge vanishes, and thus the 'disordered' spectral contribution becomes higher (~50%) even at a lower temperature (5 °C). At low pH, due to the lack of the above described salt bridge at elevated temperature (e.g. 50 °C) the contribution of the 'disordered' structural part becomes dominant (>90%). In conclusion, the Trp-cage miniprotein stability depends both on temperature and on pH. As a consequence of it, the 'disordered' spectral contribution can remain relatively low, <50%, even at 50 °C. Note taking advantages of the CCA+ protocol, only by simply looking at the raw ECD curves (Figure 4), it is harder to find out and quantify such stability range, now clearly depicted by the deconvolution. However, one has to bear in mind that by enlarging the CD data set, with the introduction of more models or additional factors the same molecular picture could be altered, but the semi-quantitative conclusion will remain. Finally, if we tried to monitor the same spectral change by using, for example, the K2D2 program [17], we can detect only a 10–15% change in the α-helix and in the β-sheet content of the miniprotein. (As K2D2 program warns us, that 'the distance is too big', really it cannot be used with high confidence to complete such an analysis.) Interestingly enough, the use of the Lincomb module of our CCA+ program, in conjunction with the peptide reference data set of PEPFIT program [4] seems to be a better choice. However, the analysis of the linear combination component percentages provides a less pronounced answer.

Oligo- and polypeptides often adopt mixtures of turns and loops as secondary structural elements. As these sequences present only residual structural preferences, as they are conformational

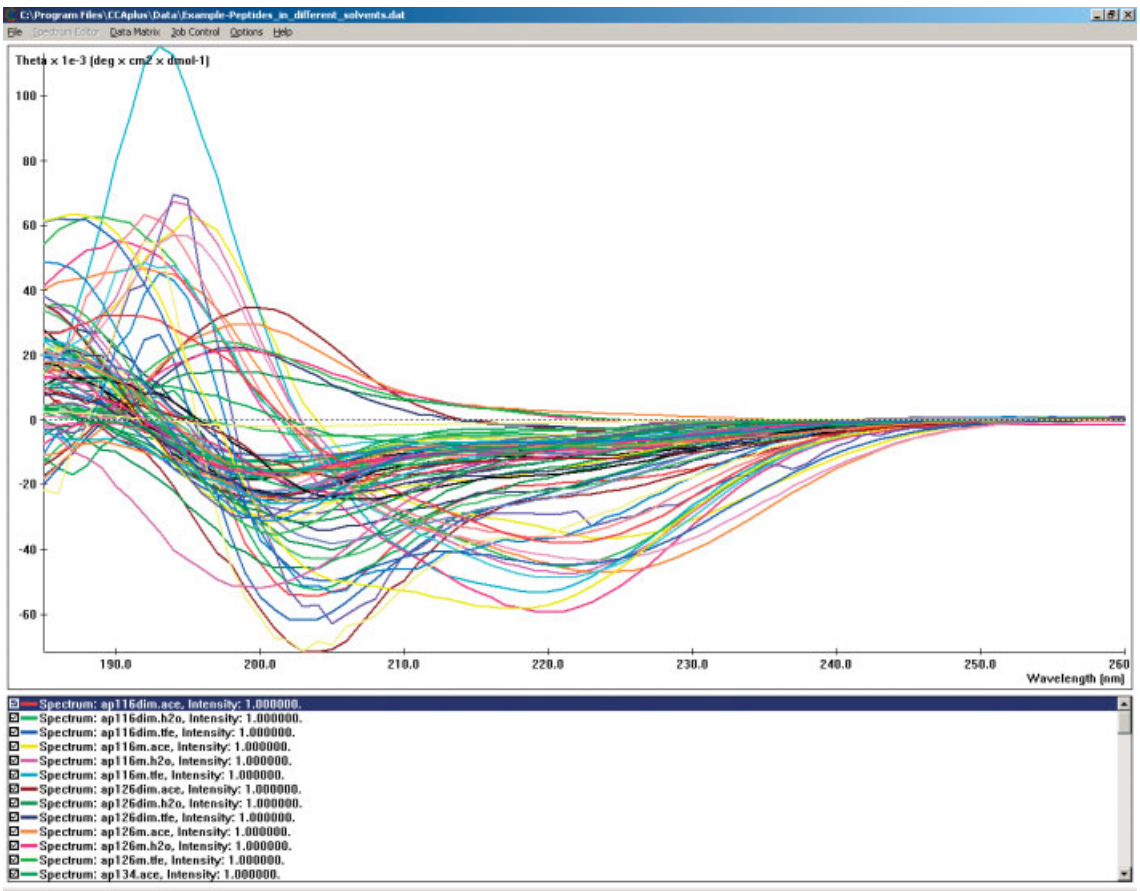

Figure 4. The ECD spectra ensemble measured at pH 2.5, 4.5, 5.5, and 6.5 in the range of 5–85 °C for Tc5b-Aad miniprotein. (A total of 68 somewhat different spectral conditions were established and jointly analyzed to decipher inherent conformational properties of the present miniprotein).

Journal of PeptideScience

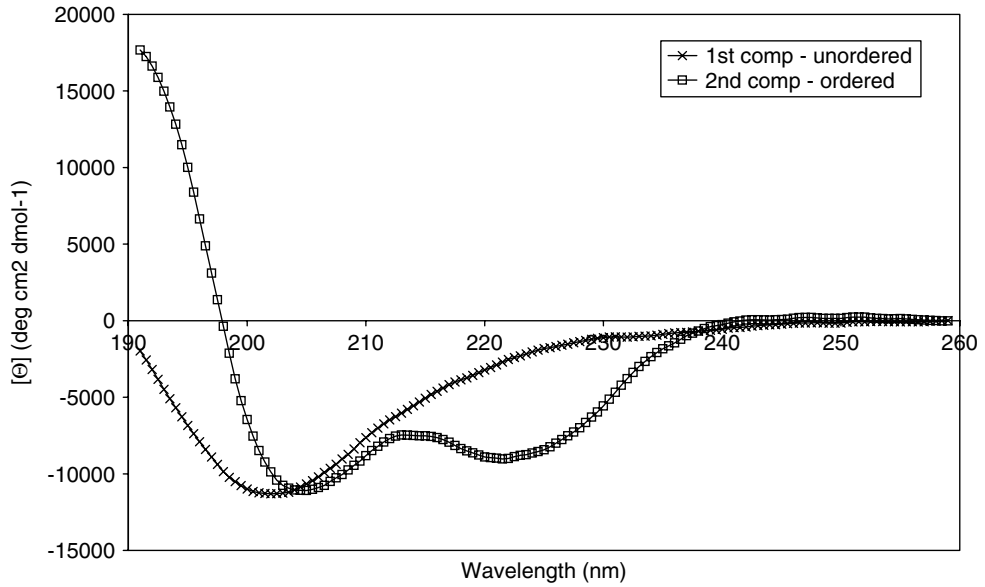

Figure 5. The two base curves obtained from the deconvolution of Tc5b-Aad miniprotein's ECD spectral ensemble. (For the same miniproptein, a total of 17 × 4 = 68 spectral conditions are jointly analyzed.) (Legend: × first component 'unordered' structure related base curve, □-second component assigned to ordered structure).

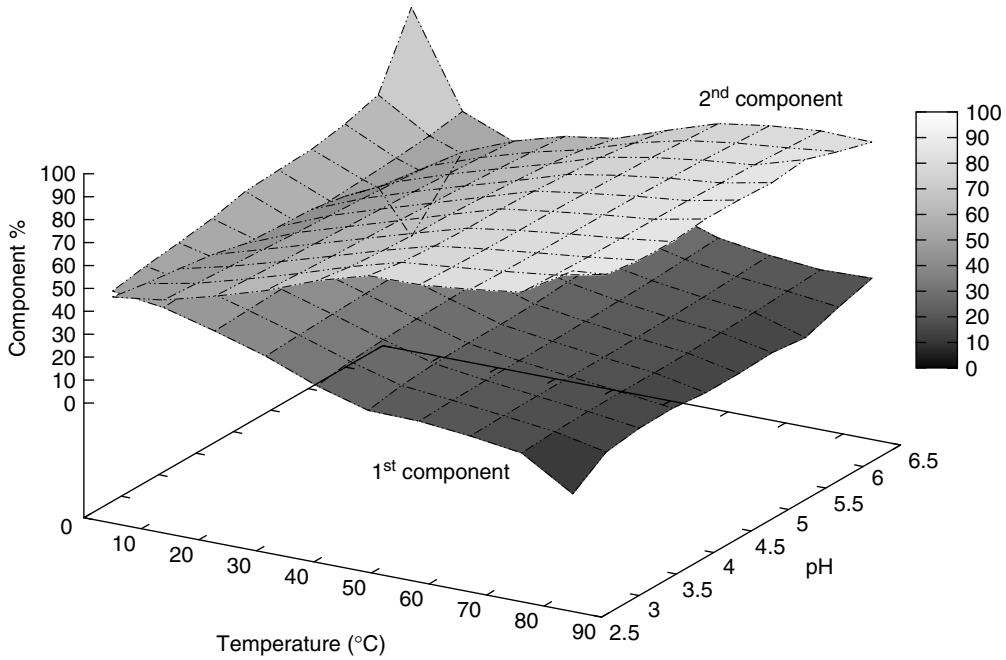

Figure 6. The spectral contribution (%) of the first component and second component base curves for the Tc5b-Aad 'miniprotein'. The component contribution plotted as function of the temperature and pH.

ensembles, there is a low chance to obtain any of the component structures in its pure form. Instead, by altering their amino acid composition, their size and sequence length and by varying their molecular environment (by changing the solvent), the relative ratio of the residual secondary structural elements can be shifted. By recording CD spectra of these models as function of the above variables, a large spectral set can be obtained and deconvoluted. The second data set, which comprises several linear and cyclic oligopeptide models of different ring size, most of them recorded in three different solvents was analyzed. Results interpreted here is the one where four pure- or base component curves (Figure 7) were extracted. (Deconvolution completed with five component curves provides hardly interpretable base curves, while that of three or less base component curves results in a poor description of the spectral ensemble.) The first pure CD spectrum extracted has a 'U-shape' and conventionally assigned to the so-called unordered or unstructured component (Figure 7). (Its shape is slightly distorted at lower frequency, probably because of the presence of D-amino-acid residues within the models.) The second and the fourth pure CD component curves were assigned to the sum of the type I plus III β-turns. As the backbone torsion angles of the latter two types of β-turns are similar, their CD curves,

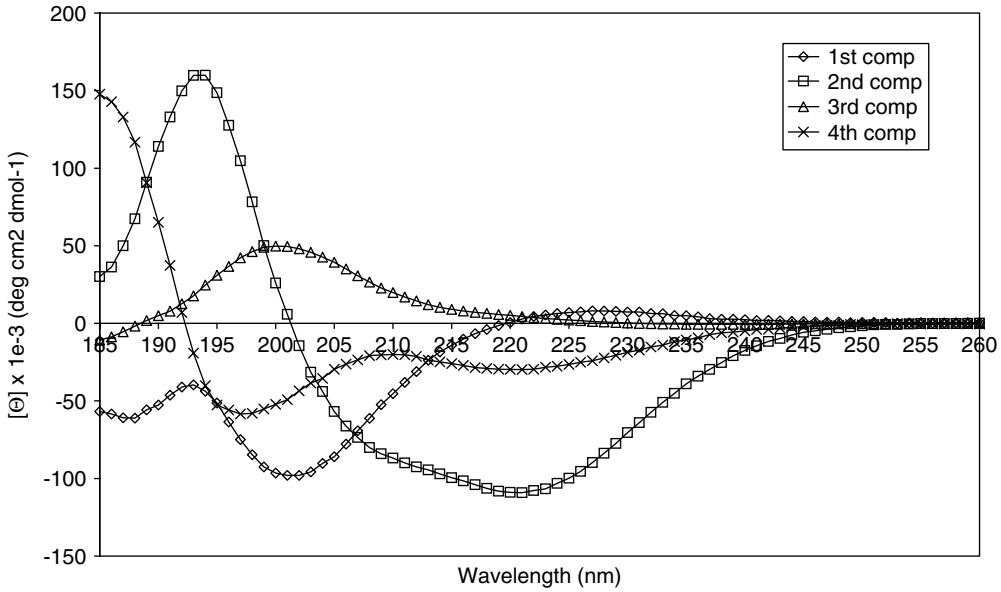

Figure 7. Deconvolution resulted in base curves of the linear and cyclic peptide data set measured in acetonitrile, trifluoroethanol, and water. (Legend: ◊-first component: non-typical spectrum, □-second component: C-type spectrum of type I or III β-turns, Δ-third component: type II β-turn, ×-fourth component: another C-type spectrum for type I or III β-turns).

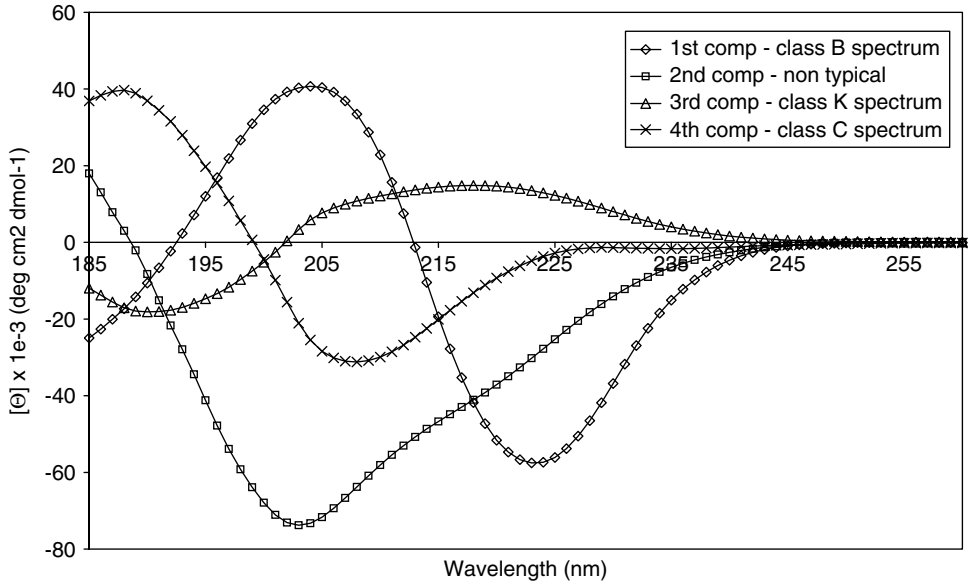

Figure 8. Base curves obtained from the Ca^{2+} and Al^{3+} ion titration of cyclic β-turn models. (Legend: ◊-first component: class B spectrum [41], □-second component: non-typical, Δ-third component: class K spectrum [42], ×-fourth component: class C spectrum [41]).

both of type C according to Woody *et al*. [41] are also of great similarity. (The spectral properties of the second and of the fourth pure component CD curve are alike, but the former one is blue-shifted.) The third pure component resembles to a type B CD curve, typically associated with type II β-turns [41], predominant for D-residue containing models. Based on the structural vicinity of the 30 different peptides jointly analyzed here, the following subdivision was made: linear peptides, linear peptides with a D-amino acid residue plus four-, six-, and eight-residue long cyclopeptides. (For all these five structural categories both the 'average' component contribution (%) and their standard deviations were calculated and reported (Table 1.) Based on our deconvolution completed at three different solvents, the following general conclusions can be drown: (i) As expected, all linear and cyclic models exhibit conformation ensembles of regular β-turns and unstructured or highly flexible backbone folds. (ii) For the L-amino acid containing linear models, the contribution of the type II β-turn structure is higher (41–46%) than the sum of type I and III β-turns (24–32%) and does not change with the solvent. The relative contribution of the atypical or 'unordered' component is in between the two (27–31%). Surprisingly, the weight of the fourth pure component curve, associated with the relative percentage of type I (III) β-turns is modest (25 $\pm$ 4%) even though sequences such as Pro-Ser, Val-Asp are believed to enhance type I (III) β-turn formation. (iii) In the second category, where a D-amino acid residue is introduced into the (i + 2)-position of the model, -Pro-D-Ser- and -Val-D-Ser-, as

Journal of PeptideScience

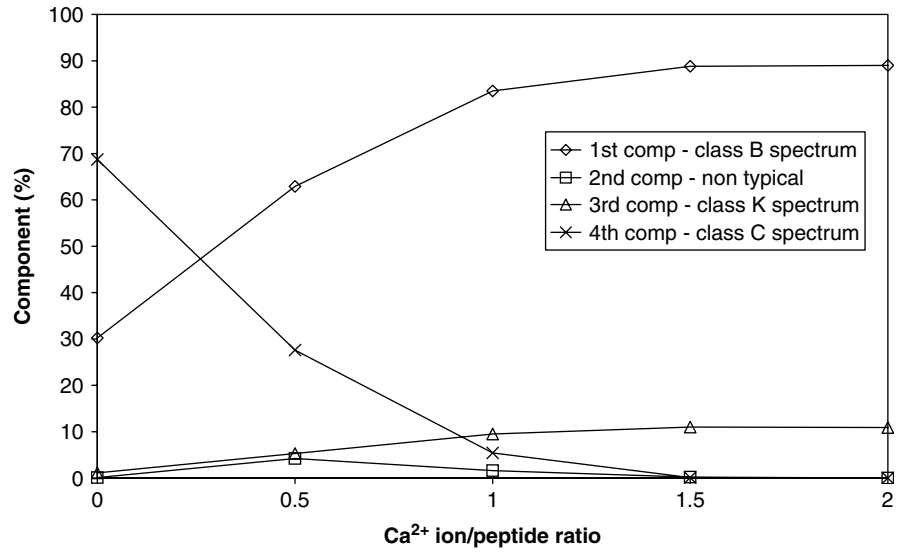

Figure 9. Base curve coefficients obtained from the Ca^{2+} ion titration of cyclo[(ε)Aca-Pro-Ala]. (Legend: ◊-first component: class B spectrum [41], □-second component: non-typical, Δ-third component: class K spectrum [42], ×-fourth component: class C spectrum [41]).

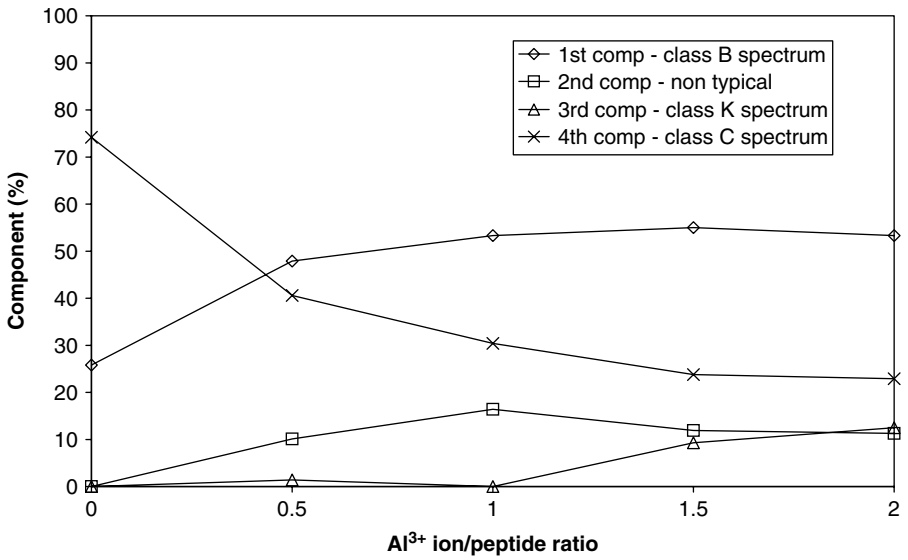

Figure 10. Base curve coefficients obtained from the Al^{3+} ion titration of cyclo[(ε)Aca-Pro-Ala]. (Legend: ◊-first component: class B spectrum [41], □-second component: non-typical, Δ-third component: class K spectrum [40], ×-fourth component: class C spectrum [41]).

expected the relative conformational weight of the type II β-turn became predominant; 76% and further increases (84%) as the dielectric constant of the solvent decreases. The conformational contribution of type II β-turn in these models was estimated by NMR to be around 80% [8]. (iv) As expected, the tightest ring containing cyclo-pseudo-tetrapeptides (cyclo[Xxx-Yyy-Aca]) have the lowest unstructured content: $10 \pm 3\%$ at all three solvents. As detected by CD, in the conformational equilibrium of these tight systems, the contribution of the type I (III) β-turn (the sum of the second and fourth component curves Figure 7) is high: ~48–58%, the highest in H_2O. However, even at such a tight ring size a low but significant amount of type II β-turn is present at any solvent as also described based on NOEs [9]. (v) Further enlargement of the ring size has a marginal effect; (4–6; cyclo[Gly-Pro-Xxx-Gly-Ava]) the conformational ensemble remains with the slight dominance of the type II β-turn. (Note that in the solid state these models can adopt a perfect type I β-turn structure as revealed by X-ray crystallography [10].) They show strong resemblance to the conformational equilibrium described above for the linear model systems in water. By using solvents with lower dielectric constants (e.g. ACE), the type II β-turn contribution decreases, most likely due to the enlargement of conformational motion. This is nicely monitored by the increase of the 'Rnd' component: 27 → 48% (Table 1). (vi) The further enlargement of the ring size further increases the backbone flexibility of pseudo-octapeptides of the following kind; cyclo[Xxx-Yyy-Aca-Xxx-Yyy-Aca]. Once again, ECD data with the help of CCA+ provide a quantitative measure of it. Because of the increased internal mobility, the relative contribution of the 'Rnd' term (first pure ECD component) is the highest (~45%). In conclusion, for the presently analyzed β-turn models the effect of solvent, even as dissimilar as H_2O and ACE, is minor. The backbone conformational equilibrium shifts but only moderately.

On the contrary, the ring size and the amino acid composition of the models have a more significant influence on their β-turn content. Because of the ECD spectral deconvolution, it became evident that the secondary structure content correlates well with the NMR data and always present conformational mixtures [35,36]. This is in fact the very reason, why in spite of all efforts the reference compounds of the above β-turn types could not yet be synthesized. Nevertheless, by using the CD spectra of the above models, the deconvolution of the spectral ensemble by CCA+ provides a relatively accurate and plausible description of these inherent flexibility β-turn models.

In our third example, the ion type and concentration-induced structure modifying effect on peptide fold is quantitatively analyzed by CD. Selected four-, six-, and eight-membered cyclic-pseudopeptides were further analyzed in the presence of Ca^{2+} and Al^{3+} ions. The CCA+ deconvolution of a total of 73 ECD curves recorded for six models at different ion type and ratios (Tables 2 and 3) resulted in the following global picture. Based on the shape of the 'base curves' (Figure 8), the fourth component shows resemblance to a class C-type ECD spectrum [41]. Adhering to Woody's terminology [41], the first component looks as class B type ECD spectrum, commonly assigned to a type II β-turn. (This spectrum has a negative maximum at around 223 nm with a positive peak around 205 nm.) The second base curve, composed from a single negative band around 204 nm, is usually associated with non-typical backbone conformational types. The interpretation of the third 'base curve' is the least conventional. In a previous analysis of linear peptides, in the presence of a high calcium/peptide ratio a similar ECD spectrum was recorded [41] and called a class 'K' spectrum [42]. However, no explicit secondary structure type was associated with the latter type of base curve. By analyzing their conformational weights (Tables 2 and 3) two following general conclusions can be drown.

In the case of the above, carefully selected cyclopeptide models the presence of Ca^{2+} or Al^{3+} ions does not induce any dramatic ECD spectral changes. This is because they do not have any specific Ca^{2+} or Al^{3+} ion binding sites (neither Glu nor Asp residues are incorporated). However, in some cases, we can pick up considerable spec-

Table 2. The analysis of 32 ECD spectra of six different cyclopeptides recorded at different Ca^{2+} ion per peptide (IPP) ratios

Peptide	Ca^{2+} IPP[a]	Comp.1 class B type II β-turn %	Comp.2 non-typ.%	Comp.3 class K %	Comp.4 class C-type I β-turn %
cyclo[(δ)Ava-Gly-Pro-Ser(OH)-Gly]	0	*11.1*	*4.8*	*52.2*	*32*
	0.33	*9.6*	*3.2*	*51.8*	*35.4*
	0.66	*9.1*	*2.9*	*51.5*	*36.5*
	1	*9.4*	*3.2*	*48.6*	*38.8*
	1.5	*9.7*	*3.1*	*45.2*	*42*
	2	*9.7*	*3*	*43.9*	*43.5*
cyclo[(δ)Ava-Gly-Pro-Ser(O^{t}Bu)-Gly]	0	*15.8*	*8.9*	*16.1*	*59.2*
	0.5	*13.7*	*15.3*	*17.6*	*53.5*
	1	*9.7*	*19.5*	*15.1*	*55.8*
	2	*3.4*	*37.3*	*24.7*	*34.7*
cyclo[(δ)Ava-Gly-Pro-Thr(OH)-Gly]	0	*11.5*	*5.2*	*51.1*	*32.2*
	0.5	*9.8*	*3.7*	*51.2*	*35.3*
	1	*8.9*	*4.4*	*49.7*	*37*
	1.5	*9.3*	*3.5*	*44.8*	*42.5*
	2	*9.4*	*3.4*	*43.6*	*43.6*
	3	*9.3*	*3.3*	*43.3*	*44.1*
cyclo[(δ)Ava-Gly-Pro-Thr(O^{t}Bu)-Gly]	0	*11*	*4.8*	*53.8*	*30.4*
	0.33	*10.2*	*4*	*53.4*	*32.4*
	0.66	*10*	*4*	*53.1*	*32.8*
	1	*9.7*	*3.1*	*47.9*	*39.3*
	2	*9.3*	*2.3*	*44.3*	*44.1*
	3	*9.4*	*2.9*	*43.8*	*43.8*
cyclo[(ε)Aca-Pro-Ala]	0	*30.2*	*0.1*	*1.1*	*68.7*
	0.5	*62.9*	*4.2*	*5.3*	*27.6*
	1	*83.5*	*1.6*	*9.5*	*5.4*
	1.5	*88.8*	*0.2*	*11*	*0.1*
	2	*89*	*0*	*10.9*	*0*
cyclo[(ε)Aca-Pro-Ala-(ε)Aca-Pro-Ala]	0	*10.6*	*5.1*	*44.4*	*39.9*
	0.5	*6.3*	*9.8*	*41.4*	*42.5*
	1	*1.2*	*16.3*	*39.2*	*43.2*
	1.5	*0*	*18.2*	*41.3*	*40.5*
	2	*0.1*	*18.4*	*41.5*	*40*

Deconvolution was completed for a total of 73 ECD curves (Ca^{2+} plus Al^{3+} data/Table 3/) resulting in pure ECD curves as follows: First component: class B spectrum; second component: non-typical; third component: class K spectrum [42]; fourth component: class C spectrum [41] (Figure 8).
[a] Cation-binding study using Ca^{2+} concentration in the range of 0–3 ion/peptide ratio.

Journal of PeptideScience

Table 3. The joint analysis of 41 ECD spectra recorded at different Al^{3+} ion per peptide (IPP) ratios

Peptide	Al^{3+} IPP[a]	Comp.1 class B type II β-turn %	Comp.2 non-typ.%	Comp.3 class K %	Comp.4 class C-type I β-turn %
cyclo[(δ)Ava-Gly-Pro-Ser(OH)-Gly]	0	*11.1*	*3.3*	*53.5*	*32.1*
	0.33	*9.9*	*5.4*	*54.5*	*30.3*
	0.66	*8.8*	*6.9*	*58.8*	*25.5*
	1	*9.9*	*4.5*	*59.2*	*26.4*
	1.5	*10.8*	*3.9*	*59.2*	*26*
	2	*11.7*	*3.5*	*59.6*	*25.2*
	2.5	*13.1*	*3.1*	*60.4*	*23.4*
	3	*13.9*	*3*	*61.4*	*21.6*
	3.5	*13.9*	*2.2*	*62.4*	*21.5*
	4	*14.1*	*2*	*62.4*	*21.5*
cyclo[(δ)Ava-Gly-Pro-Ser(O^tBu)-Gly]	0	*4.5*	*36.9*	*29*	*29.6*
	0.5	*0*	*61.9*	*0*	*38.1*
	1	*0*	*66.4*	*15.9*	*17.7*
	2	*10.5*	*62*	*27.5*	*0*
	4	*15.2*	*43.7*	*39*	*2*
	6	*16.5*	*28.9*	*54.7*	*0*
cyclo[(δ)Ava-Gly-Pro-Thr(OH)-Gly]	0	*11.5*	*4.5*	*51.4*	*32.7*
	0.5	*11.2*	*2.1*	*56.8*	*29.9*
	1	*12*	*2.7*	*61.3*	*24*
	1.5	*13.6*	*1.4*	*63.3*	*21.7*
	2	*14.7*	*0*	*66.8*	*18.6*
	3	*13.6*	*0.2*	*66.2*	*20*
cyclo[(δ)Ava-Gly-Pro-Thr(O^tBu)-Gly]	0	*11*	*4.2*	*54.2*	*30.6*
	0.33	*9.8*	*4.8*	*53.9*	*31.5*
	0.66	*10*	*3.7*	*54.5*	*31.8*
	1	*10.4*	*4.6*	*59.7*	*25.2*
	1.5	*11.7*	*3.3*	*61.5*	*23.5*
	2	*12.7*	*1.1*	*64.2*	*22*
	2.5	*13.3*	*0.4*	*66.1*	*20.2*
	3	*13.4*	*0.6*	*66.5*	*19.6*
cyclo[(ε)Aca-Pro-Ala]	0	*25.8*	*0*	*0*	*74.2*
	0.5	*47.9*	*10.1*	*1.4*	*40.6*
	1	*53.3*	*16.4*	*0*	*30.4*
	1.5	*55*	*11.9*	*9.3*	*23.8*
	2	*53.3*	*11.3*	*12.5*	*22.9*
cyclo[(ε)Aca-Pro-Ala-(ε)Aca-Pro-Ala]	0	*10.3*	*5*	*43.6*	*41.1*
	0.5	*9.5*	*11.3*	*36.7*	*42.5*
	1	*7.8*	*18.6*	*29.4*	*44.3*
	1.5	*7.3*	*21.1*	*27.5*	*44.1*
	2	*7.2*	*22.7*	*27.3*	*42.7*
	2.5	*8.1*	*23.8*	*26.9*	*41.2*

Deconvolution was completed for a total of 73 ECD curves (Ca^{2+}/Table 2/plus Al^{3+} data) resulting in pure ECD curves as follows: First component: class B spectrum; second component: non-typical; third component: class K spectrum [30]; fourth component: class C spectrum [29] (Figure 8).
[a] Cation-binding study using Al^{3+} concentration in the range of 0–6 ion/peptide ratio.

tral shifts as discussed below. In general, ignoring the few extremes, excess of Ca^{2+} ion has limited influence on the conformation equilibrium of these models as shown by the percentage changes. The weight of the first component (type II β-turn) is around 10% while that of the second pure component (non-typical) remains below 6%. Furthermore, the relative ratio of the so-called 'class K' spectrum varies around 45 $\pm$ 5%, while that of the fourth component (type I (III) β-turn) is around 35 $\pm$ 5%. If we perform a similar type of analysis for Al^{3+} ions, unlike for Ca^{2+} a gentle but significant ion concentration-induced spectral shift is monitored. In fact, the contribution of the first component (type II β-turn) increases slightly, while that of the second component (atypical) remains constant. Furthermore, the content of the third pure component (class K) increases from 50 up to 65%. Note that Al^{3+} ions have an opposite effect than Ca^{2+}, as picked up by the conformational weight changes of 'class K' pure ECD component curve.

More interestingly, by looking at the extremes, there are two types of models, which show an unexpectedly large ion-induced conformational-equilibrium shift. The tetrapeptide (cyclo[(ε)Aca-Pro-Ala]) shows a dramatic structural change when we increase

the Ca^{2+} ion concentration (Figure 9 and Table 2). Note that the type II β-turn spectral component will be dominant at <1.0 IPP Ca^{2+} ion ratio. In fact, at IPP $= 2.0$ the contribution of the first pure component is as high as 90%. However, for the very same model Al^{3+} ion (Figure 10 and Table 3) induces a similar but markedly smaller change. The spectral contribution of the type II β-turn remains around 50–55%. For both types of ion, as the type II β-turn content increases with ion strength, the type I β-turn ratio decreases. This ion type-induced β-turn type switching is perhaps coupled to the difference between their ion radii.

With respect to ion-induced conformational shift, another interesting model is that of cyclo[(δ)Ava-Gly-Pro-Ser(O^tBu)-Gly]. In the presence of Al^{3+} ions (Figure 11) the second (non-typical) spectral component contribution has a maximum at 1 IPP with a high value of 66.4%. However, the same spectral feature is not observed for the Ser(OH) containing model! Thus, the unusual conformational behavior is attached here to the presence of the bulky tert-butyloxi (O^tBu) side-chain protecting group (Figure 12). Furthermore, the very same model having Thr instead of Ser behaves also differently. In summary, changes of ion type and concentration can have a higher impact on the CD of a peptide, then the effect induced by changing the solvents. In addition, spectral features strongly change with the ring size of the model, with its constitution, and with the presence of the protecting groups.

Conclusion

The peptide conformational studies presented here have shown the versatile nature of the CCA+ program. The real power of this

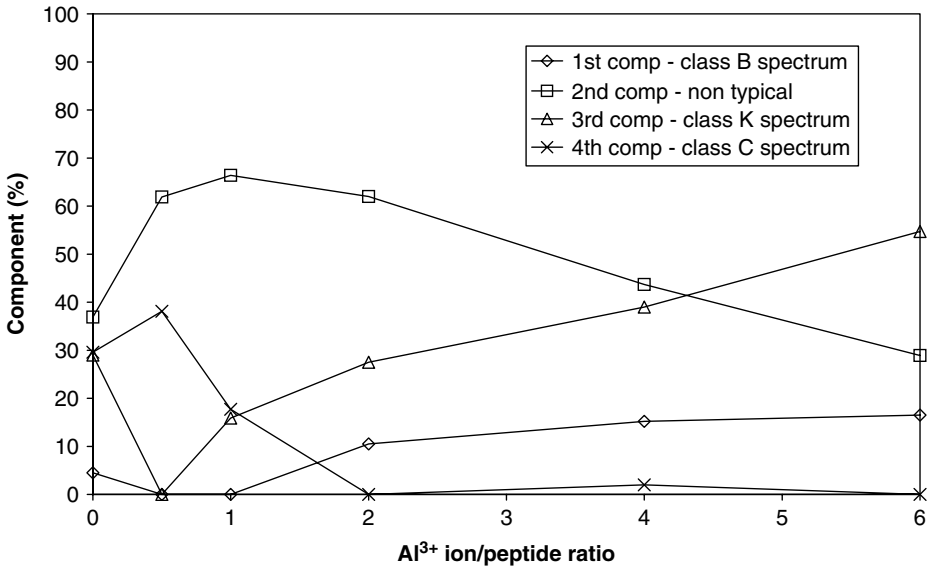

Figure 11. Base curve coefficients obtained from the Al^{3+} ion titration of cyclo[(δ)Ava-Gly-Pro-Ser(O^tBu)-Gly]. (Legend: ◊-first component: class B spectrum [41], □-second component: non-typical, Δ-third component: class K spectrum [42], ×-fourth component: class C spectrum [41]).

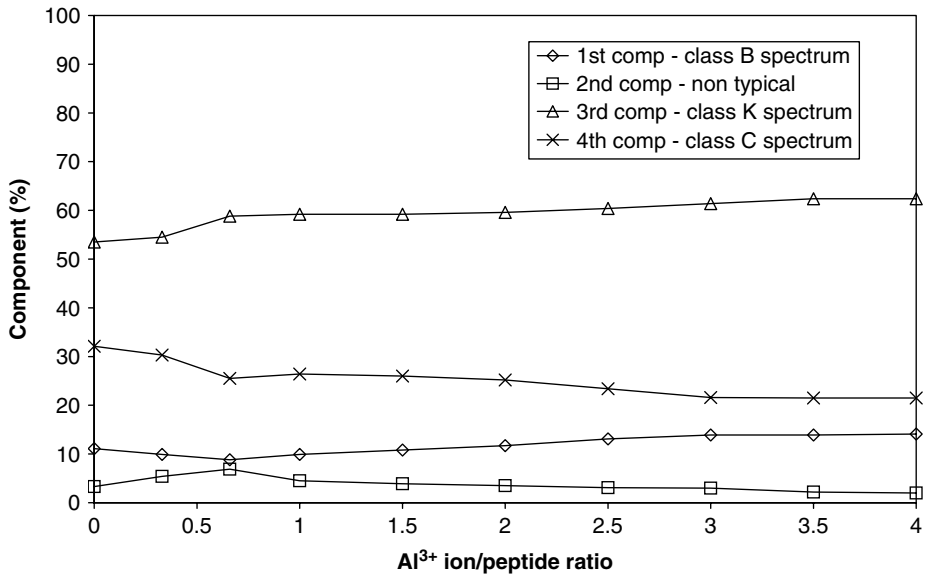

Figure 12. Base curve coefficients obtained from the Al^{3+} ion titration of cyclo[(δ)Ava-Gly-Pro-Ser(OH)-Gly]. (Legend: ◊-first component: class B spectrum [41], □-second component: non-typical, Δ-third component: class K spectrum [42], ×-fourth component: class C spectrum [41]).

algorithm was presented here, when temperature-, pH-, and/or solvent dependent ECD data sets were quantitatively analyzed. As the algorithm is independent from the spectroscopic methods in use for data acquisition, it can readily be applied for VCD, IR, or UV-VIS data analysis. Undoubtedly, ECD spectroscopy can detect and the applied deconvolution method can express in a quantitative manner the inherent conformational properties of oligo- and polypeptides.

Acknowledgements

We would like to thank Gábor Tóth for the synthesis of Tc5b-Aad, Viktor Farkas for the ECD measurements and Gábor Tusnády for helping the development of the original CCA+ algorithm. This research was supported by the National Science Fund of Hungary (OTKA K72973, NI-68466 and ICGEB 09-03).

References

1 Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. *Nat. Struct. Biol.* 2003; **9**: 425–430.
2 Hudaky P, Straner P, Farkas V, Varadi G, Toth G, Perczel A. Cooperation between a salt bridge and the hydrophobic core triggers fold stabilization in a Trp-cage miniprotein. *Biochemistry* 2008; **47**: 1007–1016.
3 Blanco FJ, Rivas G, Serrano L. A short linear peptide that folds into a native stable β-hairpin in aqueous solution. *Nat Struct Biol.* 1994; **1**: 584–590.
4 Reed J, Reed TA. A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism. *Anal. Biochem.* 1997; **254**: 36–40.
5 Brahms S, Brahms J. Determination of protein secondary structure in solution by vacuum ultraviolet circular-dichroism. *J. Mol. Biol.* 1980; **138**: 149–178.
6 Chang CT, Wu CSC, Yang JT. Circular dichroic analysis of protein conformation – inclusion of beta-turns. *Anal. Biochem.* 1978; **91**: 13–31.
7 Crisma M, Fasman GD, Balaram H, Balaram P. Peptide models for beta-turns – a circular dichroism study. *Int. J. Peptide Protein Res.* 1984; **23**: 411–419.
8 Perczel A, Hollosi M, Sandor P, Fasman GD. The evaluation of type-I and type-II beta-turn mixtures – circular-dichroism, NMR and molecular-dynamics studies. *Int. J. Pept. Protein Res.* 1993; **41**: 223–236.
9 Perczel A, Fasman GD. Quantitative analysis of cyclic beta-turn models. *Protein Sci.* 1992; **1**: 378–395.
10 Perczel A, Jákli I, Foxman BM, Fasman GD. A search for the ideal type I beta-turn. *Biopolymers* 1996; **38**: 723–732.
11 Perczel A, Park K, Fasman GD. Analysis of the circular-dichroism spectrum of proteins using the convex constraint algorithm – A practical guide. *Anal. Biochem.* 1992; **203**: 83–93.
12 Compton LA, Johnson WC Jr. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* 1986; **155**: 155–167.
13 Deleage G, Geourjon C. An interactive graphic program for calculating the secondary structure content of proteins from circular dichroism spectrum. *Comput Appl Biosci.* 1993; **2**: 197–199.
14 Bohm G, Muhr R, Jaenicke R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* 1992; **5**: 191–195.
15 Lees JG, Smith BR, Wien F, Miles AJ, Wallace BA. *CDTool* – An integrated software package for circular dichroism spectroscopic data processing, analysis and archiving. *Anal. Biochem.* 2004; **332**: 285–289.
16 Andrade MA, Chacón P, Merelo JJ, Morán F. Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Protein Eng.* 1993; **6**: 383–390.
17 Perez-Iratxeta C, Andrade-Navarro MA. Estimation of protein secondary structure from circular dichroism spectra. *BMC Struct. Biol.* 2008; **8**: 25.
18 Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* 2004; **32**: 668–673.
19 Hoffmann A. *ACDP – A JAVA APPLICATION FOR DATA PROCESSING AND ANALYSIS OF PROTEIN CIRCULAR DICHROISM SPECTRA. J. Appl. Cryst.* 2009; **42**: 137–139.
20 Provencher SW, Glockner J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 1981; **20**: 33–37.
21 Hennessey JP, Johnson WC. Information-content in the circular-dichroism of proteins. *Biochemistry* 1981; **20**: 1085–1094.
22 Manavalan P, Johnson WC Jr. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* 1987; **167**: 76–85.
23 Van Stokkum IHM, Spoelder HJW, Bloemendal M, Van Grondelle R, Groen FCA. Estimation of protein secondary structure and error analysis from CD spectra. *Anal. Biochem.* 1990; **191**: 110–118.
24 Pancoska P, Keiderling TA. Systematic comparison of statistical-analyses of electronic and vibrational circular-dichroism for secondary structure prediction of selected proteins. *Biochemistry* 1991; **30**: 6885–6895.
25 Sreerema N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* 1993; **209**: 32–44.
26 Sreerema N, Venyaminov SY, Woody RW. Estimation of the number of helical and strand segments in proteins using CD spectroscopy. *Protein Sci.* 1999; **8**: 370–380.
27 Sreerama N, Woody RW. Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 2000; **287**: 252–260.
28 Pancoska P, Yasui SC, Keiderling TA. Statistical analysis of vibrational circular dichroism of selected proteins and relationship with secondary structures. *Biochemistry* 1991; **30**: 5089–5103.
29 Pancoska P, Blazek M, Keiderling TA. Relationships between secondary structure fractions for globular-proteins – neural network analyses of crystallographic data sets. *Biochemistry* 1992; **31**: 10250–10257.
30 Dalmas B, Hunter G, Bannister WH. Prediction of protein secondary structure from circular-dichroism spectra using artificial neural-network techniques. *Biochem. And Mol. Biol. Int.* 1994; **34**: 17–26.
31 Pancoska P, Janota V, Kiederling TA. Novel matrix descriptor for secondary structure segments in proteins: Demonstration of prediction from circular dichroism spectra. *Anal. Biochem.* 1999; **267**: 72–83.
32 Santiveri CM, Rico M, Jiménez MA. Position effect of cross-strand side-chain interactions on β-hairpin formation. *Protein Sci.* 2000; **9**: 2151–2160.
33 Santiveri CM, Santoro J, Rico M, Jiménez MA. Thermodynamic analysis of β-hairpin-forming peptides from the thermal dependence of ^{1}H NMR chemical shifts. *J. Am. Chem. Soc.* 2002; **124**: 14903–14909.
34 Santiveri CM, Pantoja-Uceda D, Rico M, Jiménez MA. b-hairpin formation in aqueous solution and in the presence of trifluoroethanol: A ^{1}H and ^{13}C nuclear magnetic resonance conformational study of designed peptides. *Biopolymers* 2005; **79**: 150–162.
35 Perczel A, Hollósi M, Foxman BM, Fasman GD. Conformational-analysis of pseudocyclic hexapepetides based on quantitative circular-dichroism (CD), NOE, and X-ray data – the pure CD spectra of type-I and type-II beta-turns. *J. Am. Chem. Soc.* 1991; **113**: 9772–9784.
36 Perczel A, Hollósi M, Sándor P, Fasman GD. The evaluation of type-I and type-II beta-turn mixtures – circular dichroism, NMR and molecular dynamics studies. *Int. J. Pept. Protein Res.* 1993; **41**: 223–236.
37 Savory J, Herman MM, Ghribi O. Mechanisms of aluminum-induced neurodegeneration in animals: Implications for Alzheimer's disease. *J. Alzheimers Dis.* 2006; **10**: 135–144.
38 Poschner BC, Reed J, Langosch D, Hofmann MW. An automated application for deconvolution of circular dichroism spectra of small peptides. *Anal. Biochem.* 2007; **363**: 306–308.
39 Perczel A, FasmanGD. Effect of spectral window size on circular-dichroism spectra deconvolution of proteins. *Biophys. Chem.* 1993; **48**: 19–29.
40 Torres-Roldán RL, García-Casco A, García-Sánchez PA. CSpace: An integrated workplace for the graphical and algebraic analysis of

phase assemblages on 32-bit Wintel platforms. *Comput Geosci* 2000; **26**: 779–793.

41 Woody R. Udenfriend S. Meienhofer J. HrubyV. J. (eds). *The Peptides 7*. Acad. Press: New York, 1985; 16–115.

42 Hollósi M, Kövér KE, Holly S, Radics L, Fasman GD. Beta-turns in serine-containing linear and cyclic models. *Biopolymers* 1987; **26**: 1527–1533.

