Advanced spectra analysis to determine complex structure and chirality

Koenis, M.A.J.

Link to publication

Creative Commons License (see https://creativecommons.org/use-remix/cc-licenses):
Other

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Analytical Chemistry on Many-Center Chiral Compounds Based on Vibrational Circular Dichroism: Absolute Configuration Assignments and Determination of Contaminant Levels*

Abstract

Concurrently, it is very challenging to determine the levels and identity of potential stereochemical contaminants in a given sample with one and the same technique, leading in practice to extensive and laborious efforts employing multiple analytical techniques. In this chapter, we describe experimental and theoretical studies based on Vibrational Circular Dichroism (VCD) of dydrogesterone, a synthetic drug employed in reproductive medicine that is a prototypical example of such a multi-center chiral compound. We show that our approach allows us to distinguish and assign its absolute configuration without prior knowledge to one of the 64 possible stereoisomers associated with the six chiral centres. Studies on mixtures of dydrogesterone and 6-dehydroprogesterone, one of the diastereomers of dydrogesterone and generally the dominant impurity of dehydroosterone, show that we can identify the presence of both compounds from one single VCD spectrum. Moreover, we find that we can determine diastereomeric contamination levels as low as 5% from the experimental VCD spectra.

7.1 Introduction

Dydrogesterone is a synthetic derivative of progesterone (6-dehydroretro-progesterone) that has been developed and patented by Westerhof et al. in the early sixties\textsuperscript{222–224}. Since then it has been commercially available as a medicine against progesterone deficiency. The structure of dydrogesterone differs from natural progesterone by the inversion of two of its chiral centers (C\textsubscript{9} and C\textsubscript{10}) and the double bond between C\textsubscript{6} and C\textsubscript{7} (see figure 7.1a). This gives dydrogesterone a more rigid and bent shape than progesterone, and causes it to be much more selective towards the progesterone receptor\textsuperscript{7–9}. Unlike other forms of progesterone, dydrogesterone therefore has no or very limited affinity towards other hormone receptors, does not induce an increase in core temperature, and also is not contraceptive. The main medicinal usages of dydrogesterone are to counter infertility due to luteal insufficiency\textsuperscript{10}, menstrual disorders\textsuperscript{11} and miscarriages\textsuperscript{12–14}.

Figure 7.1: a) Structure of dydrogesterone ((8S,9R,10S,13S,14S,17S)-17-acetyl-10,13-dimethyl-1,2,8,9,10,11,12,13,14,15,16,17-dodecahydro-3H-cyclopenta[a]phenanthren-3-one). b) Structure of 6-dehydroprogesterone ((8S,9S,10R,13S,14S,17S)-17-acetyl-10,13-dimethyl-1,2,8,9,10,11,12,13,14,15,16,17-dodecahydro-3H-cyclopenta[a]phenanthren-3-one).
Dydrogesterone has 6 chiral centers (C₈, C₉, C₁₀, C₁₃, C₁₄, and C₁₇) causing it to have $2^6 = 64$ stereoisomers. Because of its different spatial structure, each of these stereoisomers has a different binding strength to hormone receptors, and will thus affect the human health differently. For pharmaceutical applications it is therefore of key importance (i) to ensure that the correct stereoisomer has been synthesized and (ii) to be able to quantify levels of possible diastereomeric impurities. Typically, diastereomers are distinguished using the torsional angle dependence of coupling constants in $^1$H NMR, which is then followed up with a chiral spectroscopic technique to distinguish the enantiomers. However, this approach fails for dydrogesterone as the molecule has several quartenary carbon centres that cannot be distinguished with $^1$H NMR alone.

Ideally, one would like to be able to assign the diastereomers of a chiral compound using a single spectroscopic technique, but this has so far proven to be far from trivial. One of the very few techniques that promises to accomplish such a task is Vibrational Circular Dichroism (VCD), whose differential nature leads to positive and negative bands with frequencies, signs, and strengths that are very sensitive to the finer details of the conformational and enantiomeric structures. Indeed, in literature there are examples where VCD has been used to distinguish between small numbers of diastereomers. However, for molecules with multiple chiral centres this is not straightforward and often requires the use of more than one chiroptical technique. Moreover, since the assignment of the absolute configuration is based on a comparison between experimentally measured and theoretically predicted spectra, a careful evaluation of the agreement between experiment and theory is needed for molecules with many stereoisomers.

Besides knowing the dominant absolute configuration, industry is also required to guarantee that there are no significant amounts of isomeric contaminations. Typically, this is achieved by chromatographic methods like chiral high-performance liquid chromatography. However, considering the large number of stereoisomers of dydrogesterone, a full separation is quite labor-intensive, if possible at all. It is therefore important to be able to test the purity of dydrogesterone independently by other analytical techniques. Because of its high stereoisomeric sensitivity, VCD could in principle be used as such a method, but as yet examples in literature where VCD has been employed to this purpose on diastereomeric mixtures are scarce. Based on the known pharmaceutical synthetic routes, the diastereomer 6-dehydroprogesterone (see figure 7.1b) is amongst the most likely impurity candidates in the case of dydrogesterone. It is therefore of particular interest to determine how accurately the presence of this diastereomer can be determined.

In this work we present extensive VCD studies of dydrogesterone and diastereomeric mixtures. We show that we can unambiguously distinguish dydrogesterone amongst all possible stereoisomers on the basis of the overlap between the experimental and theoretically predicted VCD spectra for the various stereoisomers. In addition, we demonstrate that concentrations as low as 5% of a diastereomeric impurity can still be established using VCD spectroscopy. These studies argue for
the further exploration, development and use of VCD not only to characterize the stereochemistry of compounds but for quantitative analytical applications as well.

7.2 Results and Discussion

Figure 7.2a displays the experimental and theoretically predicted vibrational absorption (VA) and VCD spectra of dydrogesterone. Overall the spectra show a very good agreement in the region from 1050 to 1500 cm$^{-1}$. The C=C and C=O stretch region of the VCD spectrum turned out to be much more difficult to measure accurately due to a low dissimilarity factor $g^{218}$ and this region has therefore been omitted from the comparison with theory (see figure 7.5a in appendix 7.5 for the full VA and VCD spectra). The calculations show that at room temperature three conformations of dydrogesterone contribute dominantly to the experimental spectra. Using Boltzmann weights as derived from the calculated conformational energies at the TZP/BP86 level of theory we find SimIR and SimVCD values between the experimental and computed spectra of 0.83 and 0.65 for the VA and VCD spectra, respectively (see equations 2.25 to 2.27).

![Figure 7.2](image)

**Figure 7.2**: Comparison between the experimental (black) and computed (red) VA (bottom) and VCD (top) spectra of (a) dydrogesterone and (b) 6-dehydroprogesterone.

Figure 7.2b displays the experimental and theoretically predicted VA and VCD spectra in the 1050 to 1500 cm$^{-1}$ region of one of the other diastereomers, 6-dehydroprogesterone, with the full spectrum up to 1800 cm$^{-1}$ being reported in figure 7.5b in appendix 7.5. Similar to the case of dydrogesterone, the calculations of 6-dehydroprogesterone indicate that only three conformations are predominantly present in experiment. Importantly, its VCD spectrum is strikingly
7.2 Results and Discussion

different from that of dydrogesterone, illustrating the resolving power of the technique. In this case comparison between experimental and with theoretical VA and VCD spectra leads to overlap values of 0.86 and 0.68, respectively. Both diastereomers thus show an excellent agreement between theory and experiment. The SimVCD values are much higher than the lower limit of 0.4 for enantiomeric discrimination\textsuperscript{139} indicating that we can assign the absolute configuration of the two compounds with high confidence.

To determine to what extent VCD allows one to distinguish not only these two diastereomers, but also yield a confidential assignment when all 64 stereoisomers are considered, we simulated the VA and VCD spectra of all these stereoisomers in the same way as was done for dydrogesterone and 6-dehydroprogesterone, that is, frequencies were optimally scaled to the experimental frequencies of each stereoisomer individually\textsuperscript{138} (see section 2.3.1 for further details). The resulting overlaps are shown in figure 7.3. As expected the VA spectra of the stereoisomers show only minor variations in the overlaps thus making it impossible to come to an unambiguous assignment of the absolute configuration of the stereocenters. The VCD overlaps, on the other hand, show that dydrogesterone has a significantly higher overlap than the other stereoisomers. In fact, the next-best performing diastereomer of dydrogesterone (the (8\text{R},9\text{S},10\text{S},13\text{R},14\text{R},17\text{S}) configuration) has a overlap of 0.34, which is almost two times lower than dydrogesterone (see figure 7.6 in appendix 7.5 for a comparison of the calculated VCD spectrum of the (8\text{R},9\text{S},10\text{S},13\text{R},14\text{R},17\text{S}) diastereomer with the experimental spectrum of dydrogesterone). The experimental VCD spectrum thus allows for an unambiguous assignment of the stereochemical configuration of the compound amongst the 64 possible stereoisomers. A similar analysis performed for samples of 6-dehydroprogesterone leads to the same conclusions. Also here the 6-dehydroprogesterone isomer has a much larger VCD overlap than the other diastereomers (see figure 7.7 in appendix 7.5). Interestingly, for 6-dehydroprogesterone we find that the next-best performing diastereomer has an overlap of 0.42 indicating that the lower limit of 0.40 taken for enantiomeric discrimination does not always suffice to discriminate between diastereomers\textsuperscript{139}.

A further assessment on how well VCD is able to distinguish between diastereomers has been made by studies of mixtures of 6-dehydrogesterone with 0\%, 5\%, 15\%, 25\%, 47\%, and 100\% dydrogesterone. From figure 7.4, in which the experimental VA and VCD spectra for these mixtures are shown, it can be concluded that there is a clear correlation between the spectra and the dydrogesterone:6-dehydroprogesterone ratio. It is clear therefore that VCD can distinguish between the different mixtures on a qualitative level.

From an industrial quality control point of view it is key to see to what extent VCD is able to determine diastereomeric impurity levels at a quantitative level. To this purpose we have first tested how well VCD can identify the compounds in the mixtures without any prior knowledge, that is, we start with the assumption that in principle all 64 stereoisomers might be present in the mixture. The contribution of each stereoisomer was then determined using a genetic algorithm (see section 2.3.2) that fits the experimental spectra to the 64 computed VCD spectra\textsuperscript{141}. Since there are only about 15 main bands in the experimen-
Figure 7.3: Overlap between computed spectra for all stereoisomers of dydrogesterone and the experimental spectra of dydrogesterone with (a) showing SimIR overlaps for VA spectra and (b) SimVCD overlaps for VCD spectra. Overlaps are only shown for enantiomers with a positive SimVCD, the opposite enantiomer giving the same SimVA and the same but negative SimVCD. The different stereoisomers are indicated by the R or S configuration of their chiral centres (respectively C₈, C₉, C₁₀, C₁₃, C₁₄ and C₁₇, see figure 7.1). The red line indicates the lower limit usually taken for enantiomeric discrimination¹³⁹.
Figure 7.4: Experimental VA and VCD spectra of mixtures of dydrogesterone and 6-dehydrogesterone. The colors purple –, blue –, green –, yellow –, orange – and red – correspond to dydrogesterone:6-dehydrogesterone ratios of 0:100, 5:95, 15:85, 25:75, 47:53 and 100:0 respectively.
tal spectrum, it is clear that for over-fitting reasons one should be cautious in using the fitted weights as representing the actual stereoisomeric weights. Nevertheless, the fitted weights can be used as an indication for the stereoisomeric composition associated with a measured spectrum. For example, when the spectra of pure dydrogesterone and pure 6-dehydroprogesterone are fitted, main contributions of 74 and 86% respectively, are indeed found for dydrogesterone and 6-dehydroprogesterone. Thus, without any prior knowledge on the configuration of the six stereocenters the fit to the experimental VCD spectrum identifies the correct stereoisomers. Even more impressively, the fit finds for the 47:53 mixture dydrogesterone and 6-dehydroprogesterone as the two main components with contributions of 32 and 40%, respectively. Interestingly, the fits find as the third component the all (S) configuration, which happens to be an epimer of both dydrogesterone and 6-dehydroprogesterone. Because of the previously indicated overfitting, the fit did not convincingly identify the presence of dydrogesterone in the mixtures with lower dydrogesterone concentrations. We therefore conclude that the VCD spectra are diastereomeric specific enough to be able to distinguish and identify two stereoisomers in one spectrum as long as the two stereoisomers have about the same concentration.

To obtain a quantitative assessment of the diastereomeric differentiation, we have fitted the experimental VCD spectra of the mixtures using only the computed spectra of pure dydrogesterone and 6-dehydroprogesterone and not those of the other 62 stereoisomers. As a further check the experimental spectra of the mixtures have also been fitted using the experimental spectra of pure dydrogesterone and 6-dehydroprogesterone. The results of these fits are given in table 7.1 while the corresponding spectra are given in figures 7.8 to 7.11 in appendix 7.5. The table also reports a reference overlap which is the overlap between the experimentally recorded VCD spectrum of a particular mixture and the spectrum constructed by adding the experimental or computed spectra of the two pure compounds using the ratio of the pertaining mixture.

The results given in table 7.1 are in first instance disappointing since even at a 15% dydrogesterone concentration no contribution of dydrogesterone is found, while at a 25% concentration only a 9% contribution is found. The table shows at the same time, however, that the fitted overlap increases for mixing ratios from 0 to 15% dydrogesterone even though in all cases the fits do not find a contribution from dydrogesterone. This observation indicates that adding some component of the experimental VCD spectrum of dydrogesterone increases the overlap between theory and experiment for 6-dehydroprogesterone. Further inspection of the experimental and computed spectra in figure 7.2 gives further insight into these result and indicates better ways to perform the fits.

Figure 7.2 shows that the highest-intensity band in the VCD spectrum of 6-dehydroprogesterone occurring at 1330 cm\(^{-1}\) is severely underestimated by the calculations. The VCD spectrum of dydrogesterone, on the other hand, features a band with the opposite sign at this frequency, and adding a small amount of the dydrogesterone spectrum will thus improve the agreement between the experimental and predicted spectra of 6-dehydroprogesterone. Accepting that the intensity
Table 7.1: Results of fits of experimental VCD spectra of mixtures of dydrogesterone and 6-dehydrogesterone. The Computational-I columns refer to a fit using the full computed spectra of dydrogesterone and 6-dehydrogesterone, the Computational-II columns to similar fits but excluding the 1330-1350 cm$^{-1}$ region, while the Experimental columns report fits using the experimental spectra of the pure compounds. Reference overlap refers to the overlap between the experimental spectrum and a spectrum constructed from the calculated (or experimental) spectra of the pure compounds with the experimental mixing ratio. Ratios are given as dydrogesterone:6-dehydroprogesterone.
of the 1330 cm\(^{-1}\) band is poorly predicted argues for fits in which this band is not taken into account. As table 7.1 shows this indeed leads to significantly better results. We now find fitted ratios that are closer to the experimental ones albeit that for the 5\% mixture still no contribution of dydrogesterone is predicted.

As it is in principle possible to use algorithms in which the sensitivity of the fits to particular spectral regions is tested and to exclude on that basis suspect regions, one could use such an approach to determine the ‘robust’ spectral regions for cases where it is not known which contamination(s) might be present. However, it is often the case that one -either because of the synthetic route or because of data from other analyses- has a fairly good idea on possible contaminations. In that case it is much more straightforward to use the experimental spectra of the components to fit to the spectra of an unknown mixture. This eliminates errors in the theoretical spectra, leaving just the experimental noise and artefacts as potential sources of error. The last three columns in table 7.1 indeed show that such fits result in very high overlaps (up to 0.97) and an excellent prediction of stereoisomeric ratios. Even for the mixture with only 5\% dydrogesterone, an 8\% contribution of dydrogesterone is predicted, indicating that VCD is capable to analyse diastereomeric compositions with an error of about 3\%.

In the present studies we have focused on identifying 6-dehydroprogesterone as a contaminant in dydrogesterone samples. Regarding the other diastereomeric contaminants, it is clear that when the experimental VCD spectrum of a contaminant is known, the statistical power to identify this contaminant and its concentration will primarily depend on how different the spectrum of the contaminant is from the spectrum of dydrogesterone. In that case the mixtures of dydrogesterone and 6-dehydroprogesterone represent an average situation as their overlap is close to zero (SimVCD = -0.08), indicating that the two spectra are neither similar nor the opposite of each other. When experimental spectra are unknown, the only possibility is to use the theoretically predicted spectra. In that case the diastereomeric discrimination also depends on how well the spectrum of the contaminant improves the differences between theory and calculation of the main compound. This may vary significantly for different stereoisomers and can be found easily by fitting the experimental spectrum with the computed spectra of the two diastereomers of interest. For dydrogesterone, for example, the overlap between the computed and experimental VCD spectrum increases considerably when 16\% of the spectrum of the (8\(R\),9\(R\),10\(S\),13\(R\),14\(S\),17\(R\)) diastereomer is mixed in (SimVCD increases from 0.65 to 0.69), while for about half of the diastereomers a potential mixing of the spectra does not lead to any improvement at all.

7.3 Conclusions

Using dydrogesterone and 6-dehydroprogesterone as prototypical examples, the present studies have shown that it is possible to identify without prior knowledge the absolute configuration of a molecule with six chiral centres on the basis of its VCD spectrum. They have moreover shown that VCD can identify the two
diastereomers correctly from the experimental spectrum of an equimolar mixture of the two. Further, without explicit knowledge of the experimental VCD spectrum of dydrogesterone and 6-dehydroprogesterone and merely using the theoretically calculated spectra of the compounds, mixing ratios down to 25% could still be detected. A much better quantification can be obtained when the experimental spectra of both the contaminant and the compound of interest are known. Using the experimental spectra to identify regions in the spectrum that are less well described by theory can then be used to increase significantly the effectiveness of diastereomeric detection from the theoretical spectra, although it still falls short of an analysis using the experimental spectra. In the latter case it has been shown that contamination levels of 5% with an uncertainty of 3% can easily be detected. Finally, it should be noted that the ability to detect stereoisomeric contaminants varies to a large extent from one stereoisomer to another, and depends significantly on the associated VCD spectra. For pharmaceutical and other synthetic industrial applications prior knowledge on which diastereomer(s) could be formed will thus be quite useful for VCD analyses of purity levels and quality control.

7.4 Methods

Vibrational absorption (VA) and VCD spectra were measured using a Bruker Vertex 70 spectrometer in combination with a PMA 50 module for polarization modulation measurements. Dydrogesterone was purchased from Sigma-Aldrich with European Pharmacopoeia (EP) Reference Standard (Y0-001004), while 6-dehydroprogesterone was ordered from TCI and had a purity higher than 98%. The samples were dissolved in deuterated chloroform at a concentration of 0.4 M and measured in a 0.56 µm CaF transmission cell for 6-8 hours. Six different samples have been measured: pure dydrogesterone, pure 6-dehydroprogesterone and four mixtures of dydrogesterone and 6-dehydro-progesterone with ratios of 5:95, 15:85, 25:75 and 47:53.

A conformational search was performed on 32 stereoisomers of dydrogesterone using the RDkit module incorporated into the Amsterdam Density Functional (ADF) software package (2015, r48476)\textsuperscript{105–107,117}. The 32 enantiomeric pairs of these stereoisomers yield identical VA and mirror-imaged VCD spectra. For each stereoisomer 1000 conformers were generated and subsequently optimized with UFF\textsuperscript{205}. Conformers were accepted within a energy range of 10 kcal/mol and with a minimum root mean square difference of 0.1. Subsequently, geometry optimizations and VCD calculations were done using Density Functional Theory (DFT) with ADF at the TZP/BP86 level of theory\textsuperscript{111,114,115,191}. For comparison with the experimental spectra the computed VA and VCD intensities were convoluted with a Lorenzian function using a full half-width at half-maximum of 8 cm\textsuperscript{-1} and the computed frequencies were scaled using the method developed by Shen et al.\textsuperscript{138} (see section 2.3.1 for further details). The resulting spectra were Boltzmann-weighted using the bonding energies. Spectral overlaps were computed with the SimIR and SimVCD measures\textsuperscript{138,140} given by equations 2.25 to 2.27.
7.5 Appendix to Chapter 7

Figure 7.5: Comparison between the experimental (black) and computed (red) vibrational absorption (bottom) and vibrational circular dichroism (top) spectra in the 1000-1800 cm$^{-1}$ region of (a) dydrogesterone and (b) 6-dehydroprogesterone.
Figure 7.6: Comparison between experimental (black) and computed (red) VA (bottom) and VCD (top) spectra. Panel (a) displays the experimental spectra of dydrogesterone and compares them with the computed spectrum for the \((8R,9S,10S,13R,14R,17S)\) which has the next-highest SimVCD value (0.34). Panel (b) makes a similar comparison for 6-dehydroprogesterone and the \((8S,9R,10R,13R,14R,17S)\) diastereomer with a next-highest SimVCD value (0.42)
Figure 7.7: Overlap between computed spectra for all stereoisomers of 6-dehydroprogesterone and the experimental spectrum of 6-dehydroprogesterone with (a) showing SimIR overlaps for VA spectra and (b) SimVCD overlaps for VCD spectra. Overlaps are only shown for enantiomers with a positive SimVCD, the opposite enantiomer giving the same SimVA and the same but negative SimVCD. The different stereoisomers are indicated by the R or S configuration of their chiral centres (respectively C₈, C₉, C₁₀, C₁₃, C₁₄ and C₁₇, see figure 7.1). The red line indicates the lower limit usually taken for enantiomeric discrimination\textsuperscript{139}.
Figure 7.8: Experimental, computed and fitted VCD spectra of a mixture of 5% dydrogesterone with 95% 6-dehydroprogesterone. (a) Experimental spectra of the mixture (black) compared with spectra constructed from the experimental spectra of the pure compounds (red). In the top panel the experimental mixing ratio has been used, while in the lower panel the ratio was fitted to mixed experimental spectra. (b) Experimental spectra of the mixture (black) compared with spectra constructed from the computed spectra of the pure compounds (red). In the top panel the experimental mixing ratio has been used, while in the lower panel the ratio was fitted to mixed theoretical spectra.
Figure 7.9: Experimental, computed and fitted VCD spectra of a mixture of 15% dydrogesterone with 85% 6-dehydroprogesterone. (a) Experimental spectra of the mixture (black) compared with spectra constructed from the experimental spectra of the pure compounds (red). In the top panel the experimental mixing ratio has been used, while in the lower panel the ratio was fitted to mixed experimental spectra. (b) Experimental spectra of the mixture (black) compared with spectra constructed from the computed spectra of the pure compounds (red). In the top panel the experimental mixing ratio has been used, while in the lower panel the ratio was fitted to mixed theoretical spectra.
Figure 7.10: Experimental, computed and fitted VCD spectra of a mixture of 25% dydrogesterone with 75% 6-dehydroprogesterone. (a) Experimental spectra of the mixture (black) compared with spectra constructed from the experimental spectra of the pure compounds (red). In the top panel the experimental mixing ratio has been used, while in the lower panel the ratio was fitted to mixed experimental spectra. (b) Experimental spectra of the mixture (black) compared with spectra constructed from the computed spectra of the pure compounds (red). In the top panel the experimental mixing ratio has been used, while in the lower panel the ratio was fitted to mixed theoretical spectra.
Figure 7.11: Experimental, computed and fitted VCD spectra of a mixture of 47% dydrogesterone with 53% 6-dehydroprogesterone. (a) Experimental spectra of the mixture (black) compared with spectra constructed from the experimental spectra of the pure compounds (red). In the top panel the experimental mixing ratio has been used, while in the lower panel the ratio was fitted to mixed experimental spectra. (b) Experimental spectra of the mixture (black) compared with spectra constructed from the computed spectra of the pure compounds (red). In the top panel the experimental mixing ratio has been used, while in the lower panel the ratio was fitted to mixed theoretical spectra.