Analytical chemistry on many-center chiral compounds based on vibrational circular dichroism: Absolute configuration assignments and determination of contaminant levels


DOI
10.1016/j.aca.2019.09.021

Publication date
2019

Document Version
Final published version

Published in
Analytica Chimica Acta

License
Article 25fa Dutch Copyright Act

Citation for published version (APA):
Analytical chemistry on many-center chiral compounds based on vibrational circular dichroism: Absolute configuration assignments and determination of contaminant levels

Mark A.J. Koenis a, Eveline H. Tieink a, Davita M.E. van Raamsdonk a, Nadav U. Joosten a, Susanne A. Gooijer a, Valentin P. Nicub b, Lucas Visscher c, Wybren J. Buma a, d, *  

a Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, the Netherlands  
b Department of Environmental Science, Physics, Physical Education and Sport, Lucian Blaga University of Sibiu, Loan Ratu Street, Nr. 7-9, 550012, Sibiu, Romania  
c Amsterdam Center for Multiscale Modeling, Section Theoretical Chemistry, Faculty of Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV, Amsterdam, the Netherlands  
d Institute for Molecules and Materials, FELIX Laboratory, Radboud University, Toernooiveld 7c, 6525 ED, Nijmegen, the Netherlands

HIGHLIGHTS
• Stereochemistry of target and isomeric impurities determined with single technique.
• Assignment absolute configuration of 6 stereocenters from one VCD spectrum.
• Determination of stereomeric composition of mixtures.
• Determination of diastereoisomeric impurity levels down to 5%.

ARTICLE INFO
Article history:
Received 3 August 2019
Received in revised form 6 September 2019
Accepted 7 September 2019
Available online 14 September 2019

Keywords:
Stereochemistry  
Vibrational optical activity  
Density functional calculation  
Absolute configuration  
Diastereomeric impurity levels  
Pharmaceutics

Abstract

The absolute configuration of a chiral molecule is key to its biological activity. Being able to find out what this configuration is, is thus crucial for a wide range of applications. The difficulties associated with such a determination steeply rise as the number of chiral centers in a given compound becomes larger. Concurrently, it becomes increasingly more challenging to determine the levels and identity of potential stereoisomers in a given sample with one and the same technique, leading in practice to extensive and laborious efforts employing multiple analytical techniques. Here, experimental and theoretical studies based on Vibrational Circular Dichroism (VCD) are presented for 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 centers. Studies on mixtures of dydrogesterone and 6-dehydroprogesterone, one of the diastereomers of dydrogesterone and generally the dominant impurity of dehydrogesterone, 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.

© 2019 Elsevier B.V. All rights reserved.

* Corresponding author. Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, the Netherlands.  
E-mail address: w.j.buma@uva.nl (W.J. Buma).

https://doi.org/10.1016/j.aca.2019.09.021  
0003-2670/© 2019 Elsevier B.V. All rights reserved.
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 [1–3]. 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 (C6 and C7) (see Fig. 1a). This gives dydrogesterone a more rigid and bent shape than progesterone, and causes it to be much more selective towards the progesterone receptor [4–6]. 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 is not contraceptive. The main medicinal usages of dydrogesterone are to counter infertility due to luteal insufficiency [7], menstrual disorders [8] and miscarriages [9–11].

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 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 $^1H$ NMR [12], 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 centers that cannot be distinguished with $^1H$ 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) [13–16], 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 [17–24]. However, for molecules with multiple chiral centers this is not straightforward and often requires the use of more than one chiroptical technique [25,26]. Moreover, since the assignment of the absolute configuration is based on a comparison between experimentally measured and theoretically predicted spectra, a — more than usual — 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 [27]. 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 [28]. Based on the known pharmaceutical synthetic routes [3,29,30] the diastereomer 6-dehydroprogesterone (see Fig. 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.

2. Materials and 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 h. The measurements were performed with a resolution of 4 cm$^{-1}$ and the central frequency was set at 1400 cm$^{-1}$. 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] [31–34]. 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 [35]. 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 [36–39]. For comparison with the experimental spectra the computed VA and VCD intensities were convoluted with a Lorentzian function using a full half-width at half-maximum of 8 cm\(^{-1}\) and the computed frequencies were scaled using the method developed by Shen et al. [40] (see SI Section 1 for further details). The resulting spectra were Boltzmann-weighted using the relative bonding energies. Spectral overlaps were computed with the SimIR and SimVCD measures [40,41] given by

\[
\text{SimIR} = \frac{I_{ab}}{I_{aa} + I_{bb} - |I_{ab}|} \quad (1)
\]

and

\[
\text{SimVCD} = \frac{I_{ab}}{I_{aa} + I_{bb} - |I_{ab}|} \quad (2)
\]

where

\[
I_{ij} = \int f_i(v) f_j(v) \, dv \quad (3)
\]

in which \(f_i(v)\) and \(f_j(v)\) are the two spectra that are compared at frequencies \(v\). SimIR can have a value between 0 and 1 while the range of SimVCD is between \(-1\) and 1. Two enantiomers thus have computed VA and VCD spectra with a SimIR of 1 and a SimVCD of \(-1\), respectively.

### 3. Results and discussion

Fig. 2a displays the experimental and theoretically predicted 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\(_2\)-C\(_2\) and C\(_2\)-O stretch region of the VCD spectrum turned out to be much more difficult to measure accurately due to a low dissimilarity factor [42] and this region has therefore been omitted from the comparison with theory (see SI Fig. S2a 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 overlaps between the experimental and computed spectra of 0.83 and 0.65 for the VA and VCD spectra, respectively.

Fig. 2b displays the experimental and theoretically predicted VA and VCD spectra in the 1050–1500 cm\(^{-1}\) region of one of the other diastereomers, 6-dehydroprogesterone, with the full spectrum up to 1800 cm\(^{-1}\) being reported in SI Fig. S2b. 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 different from that of dydrogesterone, illustrating the resolving power of the technique. In this case the comparison between experimental and 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 [43] 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 an assignment with high confidence value 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 [40] (see SI section S1 for further details). The resulting overlaps are shown in Fig. 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 (8R,9S,10S,13R,14R,17S) configuration) has a overlap of 0.34, which is almost two times lower than dydrogesterone (see SI Fig. S3 for a comparison of the calculated VCD spectrum of the (8R,9S,10S,13R,14R,17S) diastereomer with the experimental spectrum of dydrogesterone). The experimental VCD spectrum thus allows for an unambiguous assignment of the stereocentric configuration of the compound amongst the 64 possible stereoisomers. A similar analysis

---

**Fig. 2.** Comparison between the experimental (black) and computed (red) VA (bottom) and VCD (top) spectra of (a) dydrogesterone and (b) 6-dehydroprogesterone. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
performed for samples of 6-dehydroprogesterone 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 centers (respectively C8, C9, C10, C13, C14 and C17, see Fig. 1). The red line indicates the lower limit usually taken for enantiomeric discrimination [43].

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 Fig. 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 that fits the experimental spectra with the 64 computed VCD spectra [44]. Since there are only about 15 main bands in the experimental 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 fit finds 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 over-fitting, 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 1 while the corresponding spectra are given in
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 regard 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 (8R,9R,10S,13R,14S,17R) 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.

4. 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 centers 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. For this purpose, the present studies have shown that it is 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 compounds to fit to the spectra of an unknown mixture. This eliminates errors in the theoretical spectra, leaving just the experimental noise and artifacts as potential sources of error. The last three columns in Table 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 dihydroprogesterone is predicted, indicating that VCD is capable to analyze diastereomeric compositions with an error of about 3%.

Table 1

<table>
<thead>
<tr>
<th>Mixing ratio</th>
<th>Computational-I</th>
<th></th>
<th>Computational-II</th>
<th></th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fitted ratio</td>
<td></td>
<td>Fitted overlap</td>
<td></td>
<td>Reference overlap</td>
</tr>
<tr>
<td>0:100</td>
<td>0.100</td>
<td>0.6808</td>
<td>0.6808</td>
<td>0:100</td>
<td>0.8451</td>
</tr>
<tr>
<td>5:95</td>
<td>0.100</td>
<td>0.7145</td>
<td>0.6992</td>
<td>0.100</td>
<td>0.8296</td>
</tr>
<tr>
<td>15:85</td>
<td>0.100</td>
<td>0.7461</td>
<td>0.7110</td>
<td>0.955</td>
<td>0.8463</td>
</tr>
<tr>
<td>25:75</td>
<td>0.901</td>
<td>0.7486</td>
<td>0.7184</td>
<td>20:79</td>
<td>0.8381</td>
</tr>
<tr>
<td>47:53</td>
<td>43:57</td>
<td>0.7010</td>
<td>0.6979</td>
<td>50:50</td>
<td>0.7700</td>
</tr>
<tr>
<td>100:0</td>
<td>0.100</td>
<td>0.6463</td>
<td>0.6463</td>
<td>0.100</td>
<td>0.6974</td>
</tr>
</tbody>
</table>

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.

SI Figs. S5–S8. This 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 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 Fig. 2 gives further insight into these results and indicates better ways to perform the fits.

Fig. 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 of the 1330 cm⁻¹ band is poorly predicted argues for fits in which this band is not taken into account. As Table 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 artifacts as potential sources of error. The last three columns in Table 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 dihydroprogesterone is predicted, indicating that VCD is capable to analyze 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
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement


Acknowledgements

VPN acknowledges funding from UEFISCDI (PN-III-P1-1.1-TE-2016-1049, contract Nr. 46/2018). MAJK, LV and WJB acknowledge financial support from NWO in the framework of the Fund New Chemical Innovations (NWO Project Nr. 731.014.209).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2019.09.021.

References