Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 138 (2015) 700–710

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Chemometric analysis of correlations between electronic absorption characteristics and structural and/or physicochemical parameters for ampholytic substances of biological and pharmaceutical relevance

U. Judycka-Proma [a,b], L. Bober [b], A. Gajewicz [a], T. Puzyn [a,*], J. Błażejowski [a,*]

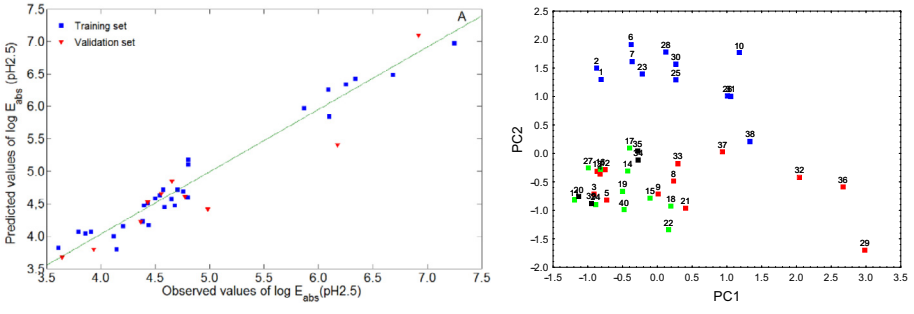
[a] Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland
[b] Research Laboratory, Polpharma S.A., Pelplińska 19, 80-200 Starogard Gdański, Poland

HIGHLIGHTS

- Spectral versus physicochemical parameters for ampholytic components of drugs.
- Chemometric analysis of relations between spectroscopic and molecular characteristics.
- An attempt to extract pharmaceutically valuable information from similarity analysis.

GRAPHICAL ABSTRACT

Data for the compounds: UV–Vis/predicted → chemometric analysis → correlations.



ARTICLE INFO

Article history:
Received 6 May 2014
Received in revised form 5 November 2014
Accepted 22 November 2014
Available online 9 December 2014

Keywords:
Ampholytes of biological and pharmaceutical relevance
Spectral (experimental) versus physicochemical (predicted) molecular characteristics
Chemometric analysis
Similarity analysis
Quantitative Structure–Property Relationships

ABSTRACT

Forty ampholytic compounds of biological and pharmaceutical relevance were subjected to chemometric analysis based on unsupervised and supervised learning algorithms. This enabled relations to be found between empirical spectral characteristics derived from electronic absorption data and structural and physicochemical parameters predicted by quantum chemistry methods or phenomenological relationships based on additivity rules. It was found that the energies of long wavelength absorption bands are correlated through multiparametric linear relationships with parameters reflecting the bulkiness features of the absorbing molecules as well as their nucleophilicity and electrophilicity. These dependences enable the quantitative analysis of spectral features of the compounds, as well as a comparison of their similarities and certain pharmaceutical and biological features. Three QSPR models to predict the energies of long-wavelength absorption in buffers with pH = 2.5 and pH = 7.0, as well as in methanol, were developed and validated in this study. These models can be further used to predict the long-wavelength absorption energies of untested substances (if they are structurally similar to the training compounds).

Introduction

The search for relations between empirical and non-empirical parameters characterising chemical substances is a subject generating great interest in many different natural sciences [1–5]. Such

* Corresponding authors. Tel.: +48 58 523 54 48; fax: +48 58 523 52 49 (T. Puzyn). Tel.: +48 58 523 51 11; fax: +48 58 523 50 12 (J. Błażejowski).
E-mail addresses: tomasz.puzyn@ug.edu.pl (T. Puzyn), jerzy.blazejowski@ug.edu.pl (J. Błażejowski).

http://dx.doi.org/10.1016/j.saa.2014.11.067
1386-1425/

relations facilitate the classification of compounds in relation to their features and the modelling of molecules with a desired set of properties [6,7]. In this publication we focus on 40 compounds of various levels of biological and pharmacological significance that exhibit ampholytic features owing to the presence in the molecules of carboxylic, sulphonyl, sulphonamide, hydroxy, thio, amino and imino groups or heterocyclic nitrogen atom(s) (Table 1S in the Supplementary material). The target compounds include natural and synthetic amino acids, pseudo-amino acids (containing an amino group in the chain or in a ring), sulphanilic acid and its derivatives (sulphonamides), and derivatives of fluoroquinone and purine. Some of the compounds are precursors of biologically important peptides or proteins, while others are significant as vitamins or broad-spectrum pharmaceuticals. The compounds selected for these investigations have one property in common: their ability to interact with both acids and bases. This provides an opportunity for finding general relations between experimental and computational characteristics, and between physicochemical features and biological or pharmacological activity, as well as for carrying out similarity analyses [5–10]. The aim of these investigations was to determine experimentally the basic spectral characteristics of the target compounds and to correlate them using chemometric methods of analysis with parameters predicted computationally, by quantum chemistry methods or empirical relationships [5–7,11–13]. Chemometric methods have already been successfully applied in spectroscopic investigations [2,3,14–17].

Materials and methods

Compounds

The numbers, canonical structures, chemical/customary names and selected features of the target compounds are given in Table 1S (Supplementary material). The compounds were obtained from: Kyowa Hakko Kogyo (Tokyo, Japan) (**9**, **11**, **14**, **15**, **16**, **17**, **18**, **19**, **22**, **24**, **27** and **40**), Alfa Wassermann (Bologna, Italy) (**33**), ACF Chemiefarm (Maarssen, Netherlands) (**8**), Merck (Darmstadt, Germany) (**2**, **4**, **12**, **13**, **20** and **39**), Aldrich (Gillingham, England) (**3**), Polpharma S.A. (Starogard Gdański, Poland) (**1**, **6**, **7**, **10**, **21**, **23**, **26**, **28**, **30** and **31**), Sigma–Aldrich (Deisenhofen, Germany) (**25**), Merck (West Point, USA) (**29**), Madex Pharmaceuticals (Lugano, Switzerland) (**32**), Ranbaxy Laboratories (New Delhi, India) (**36**), Biochefarm (Chiasso, Switzerland) (**37**), Sigma–Aldrich (St. Louis, USA) (**34**, **35** and **38**), Fluka BioChemica (Buchs, Switzerland) (**5**). They were all of analytical or pharmaceutical purity and were used as supplied.

Spectroscopic characteristics

Absorption spectra were recorded on a Perkin-Elmer Lambda 40 UV–Vis spectrophotometer. Fluorescence excitation and fluorescence spectra were measured with a Perkin-Elmer LS 50B spectrofluorimeter. The concentrations of the compounds, ranging from 1×10^{-4} to 1×10^{-5} M were such as to ensure absorbance at the long wavelength maximum at the 0.5 level. All measurements were carried out in quartz cuvettes with an optical path of 1 cm. Buffer solutions of pH 2.5, 7.0 and 11.5, as well as solvents (dielectric constant (ε), in parentheses) – 2-propanol (18.3), methanol (32.63) and water (78.64) – were of HPLC purity and were supplied by Merck (Darmstadt, Germany). All the solvents used exhibited proton donating ability (susceptibility to hydrogen bond formation) and were selected so as to cover a broad range of polarity. Buffer solutions were selected such as to cover the pH range characteristic of digestion.

The positions of long wavelength absorption (the average from absorption and fluorescence excitation spectra) and emission transitions were extracted directly from the spectra. The wavelengths relevant to these positions were converted to energies of long wavelength absorption (E_{abs}) and emission (E_{em}) transitions, expressed in eV, in a given solvent or buffer solution (E_{abs}(pH 2.5), E_{abs}(pH 7.0), E_{abs}(pH 11.5)). Stokes shifts ($\Delta\nu_{St}$), in eV, were calculated as $E_{abs} - E_{em}$. The other parameter used in the chemometric analysis was E_0, obtained from the relationship (1) [18]:

$$E_{abs} = a\varepsilon + b \quad (1)$$

$E_0 = E_{abs}$ for $\varepsilon = 1$ (a and b respectively represent the directional coefficient and the free term of a linear relation). The values of the spectroscopic characteristics determined in the above manner are given in Table 2S (Supplementary material).

Structural descriptors

Four groups of descriptors were determined. Parameters belonging to the first group were obtained by carrying out semi-empirical calculations at the AM1and AM1/CI level of theory [19] using the HyperChem version 7.5 program package [20]. Initially, the geometries of the molecules (Table 1S in the Supplementary material) were predicted by AM1 and the following characteristics were extracted directly from data files after geometry optimisations: total energy (TE in eV), binding energy (BE in eV), electronic energy (EE in eV), heat of formation (HF in kcal mol^{-1}), energy of the highest occupied molecular orbital (EHOMO in eV), energy of the lowest unoccupied molecular orbital (ELUMO in eV), dipole moment (DM in D) and polarisability (POL in bohr^3). The ionisation potential (IP in eV) was assumed to be the difference between the heats of formation of the molecular positive ion and the neutral molecule, while the electron affinity (EA in eV) as the difference between the heats of formation of the molecular negative ion and the neutral molecule [21]. Electronegativity (EN in eV) was taken to be the arithmetic mean of IP and EA [21]. Hardness (HARD in eV) was assumed to be half the difference between IP and EA [21]. The difference (DELTA) between the highest values of the relative positive and negative Mulliken atomic partial charges was also taken to be a parameter [22]. Using AM1 geometries, the following characteristics were predicted by single point AM1/CI calculations: the energy of the first long wavelength electronic transition for which the oscillator strength differed from zero (EL in eV), the energy of the strongest electronic transition (EMAX in eV) and the oscillator strength corresponding to this transition (OSMAX). Parameters belonging to the second group were obtained by carrying out *ab initio* HF geometry optimisations [23] with the 6-31G basis sets [23,24]. The following parameters were extracted directly from data files: total energy (TEAI in eV), energy of the highest occupied molecular orbital (EHAI in eV), energy of the lowest unoccupied molecular orbital (ELAI in eV), dipole moment (DMAI in D) and polarisability (POLAI in bohr^3). Further parameters were the difference between the highest values of the relative positive and negative Mulliken atomic partial charges [22] (DAI) and the electronic correlation energy (ECE in Hartree) obtained for HF optimised geometries by carrying out single point MP2 calculations [25]. The parameters belonging to the third group were predicted using the QSAR option of the HyperChem 7.5 program package [20]. In these calculations geometries optimised by the AM1 method were applied. The parameters calculated were the molecular surface accessible to solvent (SA in Å^2) [26], molecular volume (V in Å^3) [27], molecular polarisability (P in Å^3) [28], molecular refractivity (R in Å^3) [29], hydration energy (HE in kcal mol^{-1}) [30] and logarithm of octanol/water partition coefficient (LOG PC) [31]. Finally, topological descriptors, i.e. the first-order valence binding index (VCI) [32,33] and the zero-order index