Unified Physicochemical Property Estimation Relationships (UPPER)

BO LIAN, SAMUEL H. YALKOWSKY

College of Pharmacy, The University of Arizona, Tucson, Arizona 85721

Received 8 March 2014; revised 26 April 2014; accepted 1 May 2014

Published online 6 June 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jps.24033

ABSTRACT: The knowledge of physicochemical properties of organic compounds becomes increasingly important in pharmaceutical sciences, chemical engineering, and other fields. In this study, we developed UPPER (Unified Physicochemical Property Estimation Relationships), a comprehensive model for the estimation of 20 physicochemical properties of organic compounds. UPPER is a system of thermodynamically sound relationships that relate the various phase-transition properties to one another, which includes transition heats, transition entropies, transition temperatures, molar volume, vapor pressure, solubilities and partition coefficients in different solvents, and so on. UPPER integrates group contributions with the molecular geometric factors that affect transition entropies. All of the predictions are directly based on molecular structure. As a result, the proposed model provides a simple and accurate prediction of the properties studied. UPPER is designed to predict industrially, pharmaceutically, and environmentally relevant physicochemical properties. It can be an aid for the efficient design and synthesis of compounds with optimal physicochemical properties. © 2014 Wiley Periodicals, Inc. and the American Pharmacists Association J Pharm Sci 103:2710–2723, 2014

Keywords: physicochemical properties; QSPR; thermodynamics; molecular modeling; drug-like properties; solubility; log *P*; multivariate analysis

INTRODUCTION

As the synthesis of new compounds becomes more efficient, the need for knowledge of their physicochemical properties becomes increasingly important in chemical engineering and other research fields. As experimental measurement can be costly and time-consuming and cannot be performed on compounds that have not yet been synthesized, good prediction methods can be of great value, especially for compounds yet to be synthesized.

A number of techniques are available for the estimation of partition coefficients, solubility, melting points, vapor pressure, and other relevant properties. Unfortunately, many of these methods are not compatible with each other, because they are based upon different models or assumptions for the molecule. It is frequently difficult for a user to work with a number of different estimation techniques, and the data obtained from the various methods are sometimes contradictory. Many of these physicochemical property estimation methods are summarized by Baum,[1] Boethling and Mackay,[2] Lyman et al.,[3] Reinhard et al.,[4] and Yaws.[5] One of the most important developments in the estimation of physicochemical properties of organic compounds is that of Bondi,[6] who proposed a number of schemes based upon a uniform molecular fragmentation pattern. Joback and Reid[7] developed a series of equations for the estimation of various physicochemical properties that are based upon Bondi's molecular fragmentation scheme. The recent work of Gani and coworkers is nicely illustrated by Hukkerikar et al.[8] ,where 18 pure component properties are calculated by a set of schemes based upon a single molecular fragmentation pattern. In each case, the calculation of a particular property is based upon a training set of data for that property.

Unfortunately, these fragmentation schemes or group contributions do not consider those entropic properties of molecules that affect boiling points and melting points, as well as other properties that depend upon transition temperatures, such as solubility and vapor pressure. In addition, these fragmentation schemes cannot distinguish between isomers and they do not consider the interdependencies among the various properties. Over the last 20 years, this laboratory has studied the role of entropy as a determinant of the melting point of organic compounds. Recent reports[9–23] have shown that entropic factors related to molecular geometry are important determinants of both boiling and melting temperatures.

In this study, we developed a model for the estimation of 20 important physicochemical properties of organic compounds, UPPER (Unified Physicochemical Property Estimation Relationships), which integrates group contribution with the molecular geometry factors that affect transition entropies. This paper is a continuation of previous work from this laboratory on the UPPER. The experimental dataset used has been extended to over 4600 data points. This enabled the generation of coefficients for a wide variety of aromatic and aliphatic hydrocarbon groups. It also enabled the refinement of the entropic descriptors and the definition of eccentricity as a new descriptor. UPPER is a system of thermodynamically sound relationships that relate the various phase-transition properties to one another. All of the predictions are directly based on molecular structure. UPPER is designed to facilitate predicting industrially, environmentally, and pharmaceutically relevant physicochemical properties of organic compounds. It also leads to the efficient design and synthesis of compounds with optimal physicochemical properties.

Correspondence to: Bo Lian (Telephone: +520-626-2014; Fax: +520-626-2466; E-mail: (lian@pharmacy.arizona.edu)

This article contains supplementary material available from the authors upon request or via the Internet at http://onlinelibrary.wiley.com/.

Journal of Pharmaceutical Sciences, Vol. 103, 2710–2723 (2014)

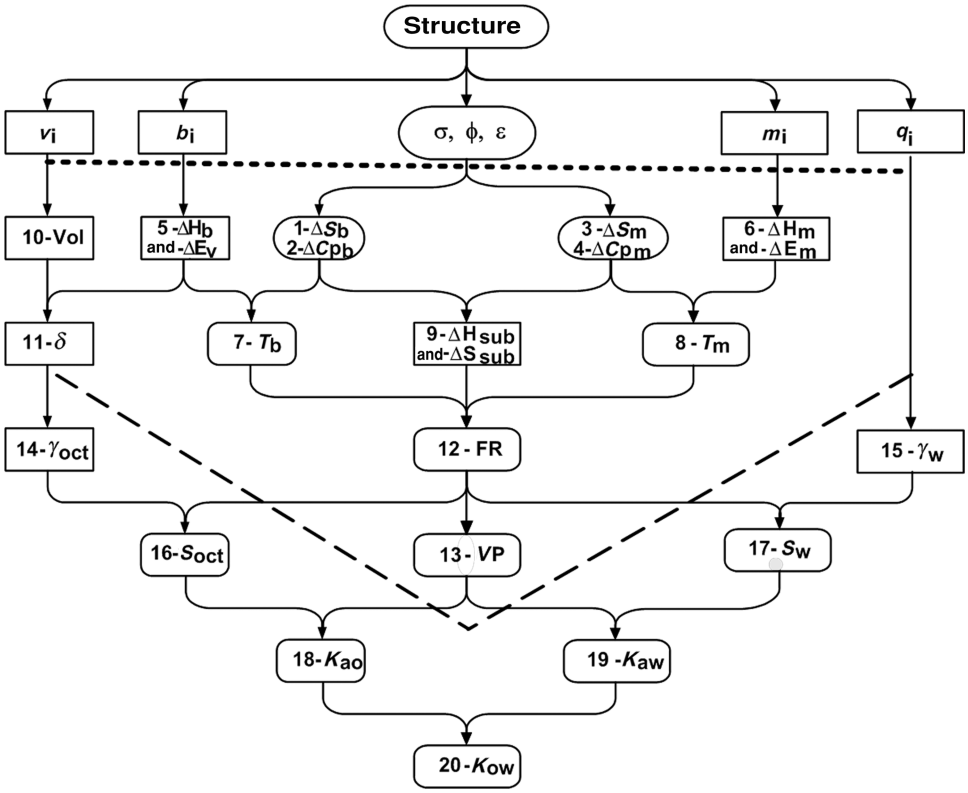

Figure 1. Unified Physicochemical Property Estimation Relationships Scheme.

THEORETICAL BACKGROUND

The UPPER Scheme

Equations are derived for each property in the UPPER scheme on the basis of the input descriptors, sound thermodynamic reasoning, and/or reasonable and well-accepted approximations. The 20 properties and their interdependencies are illustrated in the UPPER scheme in Figure 1. Each arrow represents an equation for a property in terms of properties or descriptors that appear above it in the figure.

The only input required for the UPPER scheme is the SMILES string for the compound of interest, from which the seven primary descriptors are calculated. These appear above the horizontal dotted line in Figure 1. Below the dotted line and above the dashed "V" (including boiling point, melting point, and vapor pressure) are properties of the pure compounds. The remaining seven properties are equilibria that involve the compound in a dispersed state. The properties on the left of the dashed "V" are for octanol as the solvent, and those on the right are for water as the solvent. The partition coefficient in the center describes equilibrium between solutions of octanol and water.

In Figure 1, rectangles represent additive-constitutive group contribution values and the properties calculated directly from them, and the ovals represent geometric descriptors and the properties that are solely dependent upon geometry. The remainder of the properties are calculated from a combination of additive and geometric descriptors and are shown in rectangles with rounded corners.

Enthalpic Descriptors

Unified Physicochemical Property Estimation Relationships treats the properties, which reflect enthalpic interactions as additive-constitutive properties that have the general form:

$$P_{\text{molec}} = \sum n_i p_i \tag{1}$$

where P_{molec} is the property of the whole molecule, the p_i are the contributions of molecular fragments or groups to the property, and n_i is the number of p_i fragments in the molecule. UPPER utilizes four sets of additive-constitutive parameters: heat of boiling (b_i), heat of melting (m_i), molar volume (v_i), and aqueous activity coefficient (q_i).

Each additive-constitutive molecular fragment in the UPPER scheme is defined as the smallest group of atoms (consisting of all carbons, hydrogens, and heteroatoms, including their nonbonded electrons) that are not separated by an isolating carbon. An isolating carbon was defined by Leo[24] as a carbon atom that is not doubly or triply bonded to a heteroatom.

Each fragment, i, is described by its chemical formula and by its environment. The chemical formula is written in the traditional manner with "=" and "≡" representing double and triple bonds, respectively. This is preceded by an environmental descriptor that is related to the hybrid state of the group's isolating carbon and whether or not it is in a ring or a bridgehead. The environmental descriptors are: X, the group that is bonded only to sp3 hybrid atoms; Y, the group is singly bonded to one sp2 hybrid (aromatic or vinyl) atom; YY, the group is bonded to two sp2 hybrid atoms; Z, the group is bonded to a sp atom; RG,