ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Chromatography A, 1115 (2006) 164–186

JOURNAL OF
CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Electrostatic calculations and quantitative protein retention models for ion exchange chromatography

Gunnar Malmquist, Ulrika Hjellström Nilsson, Mathias Norrman,
Ulrika Skarp, Maria Strömgren, Enrique Carredano*

GE Healthcare, Björkgatan 30, SE-751 84 Uppsala, Sweden

Received 18 October 2005; received in revised form 24 February 2006; accepted 28 February 2006

Abstract

A novel set of protein descriptors has been developed to increase the understanding of protein behavior on chromatographic media. The protein descriptors are pH-dependent and based on electrostatic and hydrophobic properties of mainly the surface of the proteins as revealed by their three-dimensional structure. Interpretable and predictive quantitative structure property relationship (QSPR) models were then obtained for protein retention in ion exchange chromatography at different pH values. In most cases the calculated average surface potential could be directly related to retention times. Moreover, the high retention of human lactoferrin observed in cation exchange even at high pH values could be modeled by adding descriptors of the charge asymmetry.

Keywords: Electrostatic potential; Lactoferrin; Ion exchange chromatography; Protein retention; PLS; PCA; QSPR; SASA; Surface potential; Surface hydrophobicity

1. Introduction

A number of studies describing "models of protein retention in ion exchange chromatography" have been presented in the literature. This is a broad theme that includes different generalizations of the simplest coulombic model describing the attraction of groups of opposite charge on the protein surface and the stationary phase, respectively. Some examples are models that explain the effect of the eluting salt concentration [1–3], models of retention of a "general protein" where proteins are modeled as spheres [4], mechanistic interpretations of parameters in the so-called stoichiometric displacement model (SDM) [5] as well as prediction of retention on different stationary-phase materials [6]. Recently a study was presented where retention parameters were shown to be correlated to the net positive charge corrected by surface accessibility for a series of analogous proteins (human monoclonal antibodies) [7].

In this study, we focus on retention of different proteins in ion exchange systems and in the mechanistic explanation of that retention.

Assuming that the pK_a values of the titrable amino acids can be approximated by those in solution, an estimate of the net charge of a protein at a certain pH can be obtained and a simple coulombic model can be used as a first approximation to understand retention differences among proteins. More accurate explanations have been proposed over the years, for instance in relation to charge asymmetry [1], calculated adsorption potential between the adsorbent and full atomic models of the protein [8] and molecular electrostatics computations performed on protein structures to determine the average potential over the molecular surface [9]. In the latter study a direct relation between the average surface potential and retention time for one set of previously reported cation exchange data [1] was reported. It is in the interest of the authors of this paper to investigate if the relation holds if the protein test set is expanded to include more proteins in the cation exchange case and also in the anion exchange case.

The comprehensive goal of this study is, however, to develop a set of pH-dependent and interpretable descriptors based on the three-dimensional structure of proteins describing electrostatic and hydrophobic properties of the proteins and to test their use in so-called quantitative structure property relationship (QSPR) modeling of chromatographic data. The experimental response in such QSPR modeling can, in principle, be any type of chromatographic response which might be depen-

---

* Corresponding author. Tel.: +46 18 612 00 00; fax: +46 18 612 18 44.
E-mail address: enrique.carredano@ge.com (E. Carredano).

0021-9673/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2006.02.097

dent on protein properties. In this study, retention data (retention times) were considered, however, other responses like parameters of an adsorption isotherm model or mass transfer models which are related to binding capacity can also be considered.

Recently QSPR models have been derived for protein retention in ion exchange chromatography by means of different numerical approaches that attempt to correlate retention to functions of descriptors derived from the three-dimensional structure of the proteins [10–12]. This is an important extension of the quantitative structure retention relationship (QSRR) models for small molecules, mainly in reversed phase chromatography, that successfully have been developed for more than a decade [13,14].

In the previously published work on protein retention, predictive models have been obtained although the interpretation of such models in terms of the properties of the proteins has been a challenge. This can be related to the nature of the descriptors used which are probably more appropriate for small molecules rather than macromolecules and have been implemented with prediction as the main objective. As compared to this, the interest of the authors of this paper is as much in interpretation of the models as it is in prediction and this guides the development of the descriptor set, keeping in mind that the models must be good enough at prediction to make it meaningful to interpret them. Therefore, QSPR descriptors that are both easy to interpret and relevant for the description of the chromatographic phenomena have been developed for and used in this study. For instance, as the chromatographic behavior of proteins is strongly pH dependent, the QSPR descriptors used in the current work were made fully pH-dependent to enhance their relevance.

The new descriptors are electrostatic descriptors based on both the charge and the electrostatic potential distributions mainly on the surface and hydrophobic descriptors based on pH-dependent hydrophobicity scales of the amino acids combined with solvent accessible area information. Furthermore, a few non-pH dependent size and shape descriptors where included to capture some effects related to mass transfer. The descriptors based on the surface charge properties developed in this study can be seen as a development of previously described charge asymmetry effects [1] and average surface potential electrostatic calculations [9]. The pH dependent hydrophobic descriptors, can on the other hand, be seen as a development of published non-pH dependent hydrophobic descriptors [15,16].

Two data sets have been used for QSPR in this study. One is a published data set of retention times in anion exchange columns for 11 different proteins at five different pH values [17] and the other is made up of newly produced retention times for 13 proteins at various pH values in cation exchange. The latter data, which was obtained using a Mono S® (GE Healthcare) column, has not been published before.

2. Materials and methods

2.1. Descriptor definition

A total of 58 descriptors were defined, developed and implemented in the proprietary program SCARP consisting of about 10,000 lines of C++ code. The majority of the descriptors, including 17 descriptors of the surface charge distribution, two descriptors of hydrogen bonding possibilities on the surface, 28 descriptors of the surface electrostatic potential dis-

Table 1
Protein descriptors

1	*pH–pI*	The pH minus the calculated isoelectric point
2	*Ap_Radius*	The apparent protein radius approximated from the protein volume, assuming a spherical shape. The protein volume is approximated by using the number of non-solvent accessible grid points multiplied by the cubic grid spacing
3	*Shape*	For each of the of surface grid points, the distance to the protein mass centre is calculated and the shape is defined as the average of these distances divided by the maximum distance. This definition would give a value of 1 for a spherical protein and 0.5 for a (hypothetical) rod-shaped protein
4	*NetCharge*	Net charge of the protein
5	*Charge_surf_den*	The net charge divided by the number of surface grid points
6	*SurfAverage*	The average surface potential (the average of the electrostatic potential at the grid points on the protein surface)
7	*StdDev*	Standard deviation of the surface potential
8	*FracOfPos*	Fraction of surface grid points with positive electrostatic potential
9	*AverPos*	Sum of the electrostatic potential of the positive surface grid points divided by the number of positive surface grid points
10	*WFracOfPos*	Sum of the electrostatic potential of the positive surface grid points divided by the total number of surface grid points
11	*FracOfNeg*	Fraction of surface grid points with negative electrostatic potential
12	*AverNeg*	Sum of the electrostatic potential of the negative surface grid points divided by the number of negative surface grid points
13	*WFracOfNeg*	Sum of the electrostatic potential of the negative surface grid points divided by the total number of surface grid points
14	*Number of surface points*	The number of surface grid points (or in a 4.5 Å thick shell around the protein)
15	*SumOfPos*	The sum of the potential at the surface points with positive potential
16	*NoOfPos*	The number of surface points with positive potential
17	*SumOfNeg*	The sum of the potential at the surface points with negative potential
18	*NoOfNeg*	The number of surface points with negative potential
19	*PosDenseArea*	The sum of the area in clusters with high positive charge density
20	*PosRegionCharge*	The sum of the charge in clusters with large positive charge
21	*ChrgInPosClusters*	The sum of the charge of positive clusters
22	*NegDenseArea*	The sum of the area in clusters with high negative charge density
23	*NegRegionCharge*	The sum of the charge in clusters with large negative charge
24	*ChrgInNegClusters*	The sum of the charge of negative clusters