Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 849 (2007) 53–68
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

Current insights on protein behaviour in hydrophobic interaction chromatography☆

M. Elena Lienqueo[a,*], Andrea Mahn[a,b], J. Cristian Salgado[a], Juan A. Asenjo[a]

[a] *Centre for Biochemical Engineering and Biotechnology, Department of Chemical and Biotechnology Engineering, University of Chile, Beauchef 861, Santiago, Chile*
[b] *Medicine Faculty, Institute for Biomedical Sciences, University of Chile, Independencia 1027, Santiago, Chile*

Received 14 June 2006; accepted 13 November 2006
Available online 1 December 2006

Abstract

This paper gives a summary of different aspects for predicting protein behaviour in hydrophobic interaction chromatography (HIC). First, a brief description of HIC, hydrophobic interactions, amino acid and protein hydrophobicity is presented. After that, several factors affecting protein chromatographic behaviour in HIC are described. Finally, different approaches for predicting protein retention time in HIC are shown. Using all this information, it could be possible to carry out computational experiments by varying the different operating conditions for the purification of a target protein; and then selecting the best conditions in silico and designing a rational protein purification process involving an HIC step.

Keywords: Hydrophobic interaction chromatography; Retention time prediction; Retention mechanism; Hydrophobicity

Contents

☆ This paper is part of a special volume entitled "Analytical Tools for Proteomics", guest edited by Erich Heftmann.

* Corresponding author.
E-mail address: mlienque@ing.uchile.cl (M.E. Lienqueo).

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2006.11.019

1. Introduction

Today, modern optimized procedures for the purification of biological macromolecules typically consist of two chromatographic separation stages; first, a ionic exchange chromatography followed by a hydrophobic interaction chromatography [1]. Then, hydrophobic interaction chromatography (HIC) is an important method for the purification of biological macromolecules, especially therapeutic proteins [2–8], DNA vaccines [9] and hydrophobic tagged proteins [10], etc.

Hydrophobic interaction chromatography exploits the reversible interaction between the hydrophobic surface patch on a protein and the hydrophobic ligand of a chromatographic medium at moderately high concentrations of salt, especially antichaotropic salt (also called kosmotropic or lyotropic salt). This kind of salts has higher polarity and bind water strongly, which induces exclusion of water on the protein and ligand surface and promotes hydrophobic interactions and protein precipitation (salting-out effect). Additionally, the presence of this kind of salts has a stabilizing effect on protein structure. In contrast, chaotropic salts have less polarity and bind water loosely, which induces inclusion of water on the protein and ligand surface, and thus tend to decrease the strength of hydrophobic interactions (salting-in effect) [11].

The first reports about hydrophobic interaction chromatography were done by Shepard and Tiselius [12], using the term "salting-out chromatography". Next, Shepard and Tiselius [13] reported that proteins are bound to neutral solid support in presence of sulfate and phosphate solutions. Afterwards, several terms were used: "hydrophobic chromatography", "hydrophobic affinity chromatography" [14] or "hydrophobic adsorption chromatography" [15]. Finally, Hjertén in 1973 described a salt mediated separation of proteins on weakly hydrophobic gel matrices, and called the method "hydrophobic interaction chromatography" [16]. Additionally, Porath et al. [17] discovered that the hydrophobic adsorption was reinforced by adding salts, like sodium chloride or phosphate chloride, and proposed the name "salt-promoted adsorption" or "salt-promoted adsorption chromatography" (SPAC).

In this review, different aspects for predicting protein behaviour in hydrophobic interaction chromatography have been presented. First, the amino acid and protein hydrophobicity concepts are explained. Next, a brief description of hydrophobic interactions and the main factors affecting it are presented. Finally different approaches and models for predicting protein retention time in HIC are discussed along with their advantages and disadvantages.

2. Hydrophobic interactions and retention mechanisms in hydrophobic interaction chromatography

Hydrophobic interactions are the most important non-covalent forces that will cause processes, such as structure stabilization of proteins [18], binding of enzymes to substrates [19], and folding of proteins [20,21]. This kind of interaction appears when non-polar compounds are put into water. In this situation, an increase in entropy is observed ($\Delta S > 0$), resulting from a displacement of the ordered water molecules around the non-associated hydrophobic groups to more unstructured bulk water. The positive enthalpy, ΔH, is smaller than the entropy. Therefore, there is a negative change in free energy ($\Delta G < 0$), and then a thermodynamically favourable process, according to the Gibbs function [19]:

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

where ΔH and ΔS are the changes in enthalpy and entropy, respectively, and T is the absolute temperature.

In particular, in the case of hydrophobic interaction chromatography, the separation happens on the basis of hydrophobic interactions between immobilized hydrophobic ligands, like butyl, octyl and phenyl, and hydrophobic solvent-exposed regions on proteins. A protein frequently has hydrophobic patches on its surface and when these are in contact with an aqueous solvent, the water molecules close to the hydrophobic patches are arranged in an ordered mode. Then, the interactions between a hydrophobic matrix and hydrophobic areas on a protein can be explained based on the thermodynamic aspects previously discussed, i.e. hydrophobic interactions are a thermodynamically favourable process [19].

There are several studies on binding mechanisms in HIC [22–25], but none of them has had broad acceptance. The first and most widespread theoretical framework was developed by Melander and Horvath [26], based on Manning's counterion condensation theory for electrostatic interactions [27], and an adaptation of Sinanoglu's solvophobic theory [28] of the salting-out of proteins and their retention in HIC. This theory was adapted to a simpler form in order to account for salt effects in protein retention due to hydrophobic interactions [29]. The disadvantage of this model is that it is not valid for a wide range of salt concentrations; in particular, this model is only valid for low salt concentrations. More details will be presented in Section 5.1 below.

Staby and Mollerup [30] have proposed a model for solute retention behaviour of proteins on HIC perfusion media, based on the influence of the protein activity coefficient in the mobile

and stationary phase. Additionally, they modelled the activity coefficient in the mobile phase by a Debye-Hückel equation, and the activity coefficient in the stationary phase by a simple non-linear term. The main advantage of this model, in comparison with Melander and Horvath's model, was that it could be applied to the whole range of salt concentration, from zero to high ionic strength. More details in Section 5.1 below.

Oscarsson [31] suggested that the retention of protein in HIC could be described considering conformation change. He proposed that proteins change their conformation continually, and some of those conformations are improved by the specific ligand and operating condition (type and concentration of salt), promoting the exposure of interactive sites on the surface of the protein. Therefore, favourable interactions can happen if these surface sites are complementary to the groups on the stationary phase surface. On the other hand, several authors have reported that these hydrophobic interactions between proteins and some HIC media could produce loss of enzyme activity [32,33], low chromatographic recovery [34–36], and in the case of unstable proteins (α-lactoalbumin, lysozyme), partial or total unfolding may occur [37–39]. For instance, there are reports about two peaks in HIC of α-lactoalbumin; the less retained was identified as native, and the more retained as an "unfolded mixture of species." The magnitude of both peaks is highly dependent on the salt type in the mobile phase [37,38]. This result demonstrates that the unfolding of proteins upon adsorption could be an important factor that has not been included in prediction of protein behaviour in HIC.

Jennissen [23] proposed that the adsorption of a protein on a hydrophobic surface is a saturable process, where the adsorption is a multi-step reaction, which the rate-limiting step is a slow conformational change or reorientation step of the protein on its hydrophobic surface.

Chen et al. [40], Huang et al. [24], Lin et al. [22], and Tsai et al. [25] suggested that the mechanism has five sequential sub-processes: (a) the dehydration or de-ioning (removing the electrical double layer) process of the protein; (b) dehydration or de-ioning process of the gel; (c) Van der Waals forces between proteins and hydrophobic resin; (d) the structure of the protein is arranged upon adsorption; and (e) the excluded water or ion molecules in a bulk solution is rearranged.

Perkins et al. [41] applied a model based on the preferential interaction theory proposed by Timasheff and co-worker [42]. This theory is based on the interaction between protein and salt, and compares the number of water molecules released with salt and ions released on protein binding in HIC. Some applications of this model will be described in Section 5.1 below.

Recently, Ladiwala et al. [43] have proposed a quantitative structure retention relationship (QSRR) model, based on a support vector machine (SVM), for evaluating the effects of stationary phase resin chemistry and protein physicochemical properties on protein binding affinity in HIC. This QSRR model used molecular descriptors based on the three-dimensional structure of proteins, the primary structure information, and a set of new hydrophobicity descriptors. Their results have shown a good capacity to predict the protein retention and to interpret the physicochemical effects that contribute to the binding affinity of proteins under different operating conditions. Additional results will be presented in Section 5.2 below.

Recently, Jakobsson et al. [44] have described the hydrophobic interaction using a description of the interaction between the protein and solid phase, and a description of the dispersion in the column. The solid-phase interaction was modelled based on the solvophobic theory using an interaction model including kinetics. As a result of this model, they proposed a method of gaining process knowledge and assisting in the robustness analysis and optimization of an HIC step.

3. Amino acid and protein hydrophobicity

3.1. Amino acid hydrophobicity

The amino acid hydrophobicity could be estimated in different ways; therefore, there are several different scales that have been used to estimate it. The variation in the hydrophobicity ranking of individual amino acid is sometimes significant [45,46], as shown in Fig. 1. This discrepancy could be due to several factors; for example, the amphiphilic character of the aromatic amino acids (phenylalanine, trytophan and tyrosine) leads to different relative contributions to the hydrophobicity, depending on the method and solute chosen [45–47]. The assumption regarding the charged or uncharged state of histidine residues affects the hydrophobicity ranking, and also, cysteine residues can form disulfide bonds; the cysteine residues then appear more hydrophobic.

3.2. Protein hydrophobicity

In the case of protein hydrophobicity, it could be defined based on the hydrophobicities of the exposed and buried amino acids [48], called "degree of hydrophobicity" or only upon the hydrophobicities of the exposed amino acids, called "average surface hydrophobicity" [49].

The classical parameters used to characterize the "degree of hydrophobicity" were "average hydrophobicity", based on Tanford's free energies of transfer of amino acid side chains from an organic environment to an aqueous environment [50]; "non-polar chain frequency", NPS, calculated as the frequency of non-polar side chains [51]; "polarity ratio", p, calculated as the ratio between external and internal volumes of the protein [52]; or "net hydrophobicity" [19], amongst others.

In the case of "average surface hydrophobicity" there are several ways to quantify "protein surface hydrophobicity". The first one is by using fluorescent probe methods based on 1-anilinonaphthalene-8-sulfonic acid (ANS), *cis*-parinaric acid (CPA), and 6-propionyl-2-(*N*,*N*-dimethylamino)-naphthalene (PRODAN) [53]. Upon the binding of the probes to accessible hydrophobic regions of proteins, an increase in fluorescence is observed, which is used as a measurement of protein surface hydrophobicity. However, due to the possible contribution of both electrostatic and hydrophobic interactions to the binding of these anionic probes, the interpretation based on these probes has not been easy [53]. The second approach to estimate this value considers the protein's three-dimensional structure data

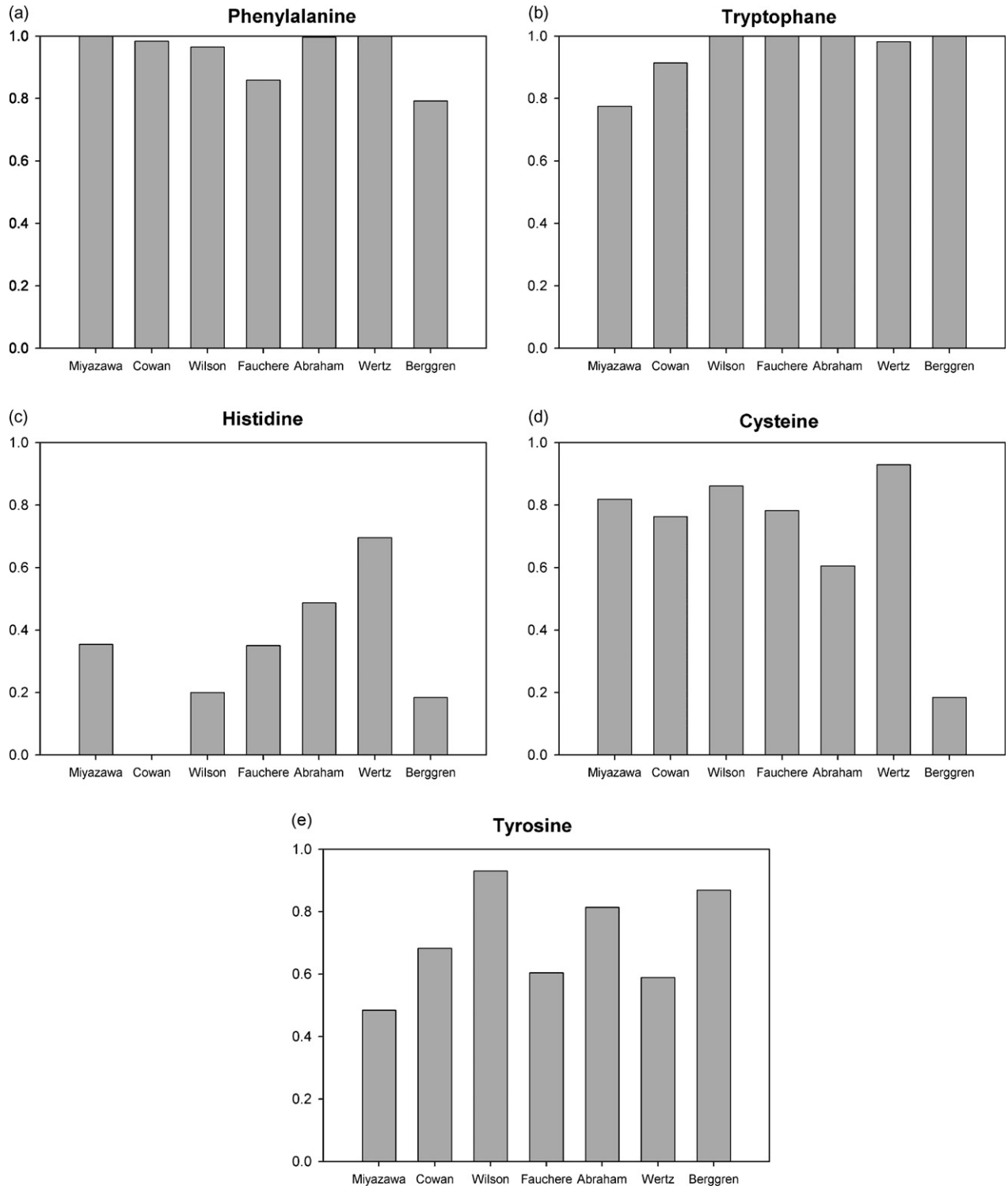

Fig. 1. Comparison of amino acid hydrophobicity using different hydrophobicity scales: Miyazawa and Jernigan [78], Cowan and Whittaker [94], Wilson et al. [95], Fauchere and Pliska [96], Abraham and Leo [97], Wertz and Scheraga [98], Berggren et al. [49]. (a) Phenylalanine, (b) tryptophane, (c) histidine, (d) cysteine, (e) tyrosine.

and assumes that each amino acid on the surface of a protein has a contribution proportional to its surface accessible area [49]. Then the average surface hydrophobicity, ASH, is calculated as:

$$\mathrm{ASH} = \frac{\sum s_{\mathrm{aa}i}\phi_{\mathrm{aa}i}}{s_{\mathrm{p}}} \quad (2)$$

where i ($i = 1,\ldots, 20$) indicates the different amino acids, $s_{\mathrm{aa}i}$ is the solvent accessible area occupied by amino acid "i", $\phi_{\mathrm{aa}i}$ is the hydrophobicity value assigned to amino acid "i", and finally, s_{p} is the total solvent accessible area of the protein. It should be noted that for proteins with a prosthetic group, s_{p} is bigger than $\sum s_{\mathrm{aa}i}$.

Recently, Mahn et al. [54] introduced the concept of "local hydrophobicity" (LH) as the average surface hydrophobicity of the interaction zone of the protein with the hydrophobic ligand. LH is very useful to characterize the surface hydrophobicity distribution. LH was defined as:

$$\mathrm{LH} = \frac{\sum(s_{\mathrm{aa}i}\phi_{\mathrm{aa}i})}{s_{\mathrm{IZ}}} \tag{3}$$

where $s_{\mathrm{aa}i}$ is the solvent accessible area of each residue in the interaction zone; s_{IZ} is the solvent accessible area of the interaction zone.

Unfortunately, the previous approach needs the three-dimensional structure of the protein to perform calculations, and sometimes this data is not readily available or is very difficult or expensive to obtain. However, Salgado et al. [55] have developed a new parameter based only on the amino acid composition of the protein, called Γ. The estimation of Γ was computed using the following equation:

$$\Gamma = c_0 + \sum_{i=1}^{20} c_i \hat{a}_i + c_{21}\hat{l} \tag{4}$$

where c_i (i from 0 to 21) corresponds to the parameters of the linear model obtained by the least squares procedure, $\hat{l}$ is the ratio between the length of the protein sequence and the maximum length observed in the working database. The $\hat{a}_i$ value was calculated by:

$$\hat{a}_i = \frac{n_i S_{\mathrm{max},i}\beta_i + \eta_i}{\sum_{j\in A}(n_j S_{\mathrm{max},j}\beta_j + \eta_j)} \tag{5}$$

where n_i is the number of amino acids of class i in the protein and $S_{\mathrm{max},i}$ is the maximun possible value of the accessible surface area (ASA) for amino acid i. Finally, β_i and η_i are the coefficients of the linear model between S_i (sum of the ASA for all the amino acids of class I) and $n_i \cdot S_{\mathrm{max},i}$ calculated in a collection of 1982 proteins with known three-dimensional structure using the least squares procedure.

4. Representation of protein chromatographic behaviour in HIC

The chromatographic behaviour could be represented by the elution curve. It consists of a retention volume (V_{r}) or time (t_{r}) of the protein and the shape of the curve.

In the case of isocratic elution, the chromatographic behaviour must be represented by capacity or retention factor, *k'*, defined as:

$$k' = \frac{V_{\mathrm{r}} - V_{\mathrm{o}}}{V_{\mathrm{o}}} \tag{6}$$

where V_{o} is the void volume in the column. This parameter could be determined using direct and indirect methods [56]. In the case of the indirect method, the column is filled and weighed with a light solvent (e.g. methanol) and a heavy solvent (e.g. ethyleneglycol or carbon tetrachloride). Then, the void volume in the column is calculated by $V_{\mathrm{o}} = \frac{m_2 - m_1}{\rho_2 - \rho_1}$; where m_1 and m_2 are the masses of the column filled with the light and the heavy solvents, respectively. ρ_1 and ρ_2 are the densities of the light solvent and the heavy solvent, respectively.

Additionally, the retention capacity could be related to the equilibrium constant for the distribution of solute between the bulk mobile phase and stationary phase, K_{HIC}. This relation is shown in the following equation:

$$k' = K_{\mathrm{HIC}}\varphi \tag{7}$$

where φ is the phase ratio of the column, i.e. ratio of the volume of the stationary phase to that of the mobile phase.

In the case of salt gradient elution, the chromatographic behaviour could be represented by dimensionless retention time, DRT.

$$\mathrm{DRT} = \frac{t_{\mathrm{r}} - t_{\mathrm{o}}}{t_{\mathrm{f}} - t_{\mathrm{o}}} \tag{8}$$

where t_{r} is the time corresponding to the retention time of the target protein, t_{o} is the time corresponding to the start of the elution gradient, and t_{f} is the time corresponding to the end of the salt gradient. If a protein is not retained by the resin, DRT is equal to 0, and if a protein elutes only after the gradient has been completed, its DRT is equal to 1.

In this review, we present several models proposed for predicting the capacity factor and/or dimensionless retention time. Previously, we described the main factor affecting the chromatographic behaviour of proteins in HIC.

5. Main factor affecting protein chromatographic behaviour in HIC

In HIC, protein retention is mainly affected by protein hydrophobicity [34,46,48], and especially its surface hydrophobicity distribution [54,57], which explains the main part of the selectivity of this purification technique. However, operating conditions obviously affect chromatography behaviour of protein in HIC [23,43,48] to a great extent. The operating conditions showing a significant effect on HIC performance are mobile phase properties (salt type, ionic strength, pH), stationary phase characteristics (chemical nature of the backbone, type of hydrophobic ligand, substitution level of the resin), and temperature of the chromatographic system [17,22,26,29,58–62]. The effects of these three factors on HIC performance are discussed below.

5.1. Effect of mobile phase

The factors characterizing the mobile phase are ionic strength, determined principally by ion strength, type of salt, and buffer pH.

5.1.1. Effect of ionic strength

Adsorption of proteins to a HIC media is favoured by a high salt concentration, but due to differences in the interaction strength between the adsorbent and different proteins, the concentration of salt needed for adsorption can vary considerably. However, the concentration of salt used should be below

the concentration precipitating different proteins. Concentration of salt is usually between 0.75 and 2.0 M with ammonium sulfate or 1.0 to 4.0 M with sodium chloride [19]. The interpretation of salt effects on protein retention has largely been investigated, because of their complex nature which can not be explained satisfactorily by classical electrostatic theories [26,29,34,58].

Melander and Horvath [26] were the first in describing the effect of salt on electrostatic and hydrophobic interactions using an adaptation of solvophobic theory [28]. Their theory proposed a linear relationship between $\ln k'$ and salt molality at sufficiently high salt concentration. This relationship is shown in the following equation:

$$\ln k' = \Delta A \sigma m + C \tag{9}$$

where ΔA is the difference in surface area of ligand and protein exposed to mobile phase between the bound and unbound states, called molecular contact area upon binding; m is the molal salt concentration; C represents salt independent terms, and σ is the molal surface tension increment.

The disadvantage of this model is that it is not valid for a wide range of salt concentrations, e.g. low salt concentrations.

Staby and Mollerup [30] evaluated the retention behaviour of a protein on HIC perfusion media, based on the thermodynamic theory. They proposed that $\ln k'$ is a function of the protein activity coefficient in the mobile phase and the protein activity coefficient on the stationary phase, and that could be calculated using the following equation:

$$\ln k = \ln k'_0 + \ln \frac{\gamma^{\mathrm{m}}}{\gamma_0^{\mathrm{m}}} - \ln \frac{\gamma^{\mathrm{s}}}{\gamma_0^{\mathrm{s}}} + \ln \frac{\upsilon^{\mathrm{m}}}{\upsilon_0^{\mathrm{m}}} \tag{10}$$

where k'_0 is the capacity factor at zero ionic strength. $(\gamma^{\mathrm{m}}/\gamma_0^{\mathrm{m}})$ is the ratio of the protein activity coefficient at finite ionic strength over the protein activity coefficient at zero ionic strength; and it was modelled by a Debye-Hückel term plus a linear term. $(\gamma^{\mathrm{s}}/\gamma_0^{\mathrm{s}})$ is the ratio of the activity coefficient of the protein on the stationary phase at finite ionic strength over the activity coefficient of the protein on the stationary phase at zero ionic strength, and it was calculated using an empirical expression. $(\upsilon^{\mathrm{m}}/\upsilon_0^{\mathrm{m}})$ is a ratio between molar volume of mobile phase at finite and zero ionic strength. The three terms have been modelled as a function of ionic strength, I; Eq. (10) could be represented by the following empirical equation:

$$\ln k' = \ln k'_0 + \left(\frac{1.5}{a} \frac{I}{1 + 1.6\sqrt{I}} + 0.15I \right) + (bI - cI^3) + 0.016I \tag{11}$$

where I is the ionic strength and a, b and c are constant according to mobile phase pH.

This model was tested with good results for lysozyme retention under various ammonium sulfate concentrations, ionic strength, and pH on four HIC perfusion media.

Additionally, Machold et al. [60] proposed to use simple polynomial functions to fit $\ln k'$ versus ionic strength, I. They suggested polynomial functions as:

$$\ln k' = a + bI + cI^2$$

$$\ln k' = a + bI + cI^2 + dI^3 \tag{12}$$

where a, b, c and d are parameters obtained under a variety of conditions.

The models were validated for a great number of proteins and sorbents with different hydrophobicities and ionic strength.

Perkins et al. [41] used the model of preferential interaction theory to the HIC system and developed a correlation between the capacity factor of a solute and salt concentration. This relationship is given by the following equation:

$$\ln k' = c + \frac{n\Delta v_1}{m_1 g} m + \frac{(\Delta v_+ + \Delta v_-)}{g} \ln(m) \tag{13}$$

where m_1 and m are the molal concentration of water and salt, respectively; n is the valence of salt ions; Δv_1 is the number of water molecules released during the binding process; Δv_+ and Δv_- are the number of cations and anions released during the binding process, respectively; $g = \left(\frac{\partial \ln m}{\partial \ln a_\pm} \right)_{T,P}$, a is the activity of ions. The constant g could be calculated from Debye-Hückel equation.

Eq. (13) could be simplified; then:

$$\ln k' = \alpha + \beta m + \gamma \ln(m) \tag{14}$$

where β and γ are called the preferential interaction parameter. The total number of water molecules and salt ions released during the binding process can be calculated by:

$$\Delta v_1 = \frac{\beta g m_1}{n} \tag{15}$$

$$\Delta v_+ + \Delta v_- = \gamma g \tag{16}$$

Eqs. (15) and (16) were also used to evaluate the effects of salt, pH, and stationary phase.

Finally, Tsai et al. [25] showed that adsorption enthalpies can be used to interpret the interaction mechanism in HIC by proposed sub-processes of adsorption. The data obtained also showed that the addition of salts not only enhanced the hydrophobic interactions between proteins and hydrophobic adsorbents but also reduced the heat required for dehydration.

5.1.2. Effect of type of salt

Effect of salt type on protein retention has largely been investigated, and it has been demonstrated that it follows the Hofmeister series for the precipitation of proteins from aqueous solutions [23,26,63]:

Anions : PO_4^{3-}, SO_4^{2-}, CH_3COO^-, Cl^-, Br^-, NO_3^-, ClO_4^-, I^-, SCN^-

Cations : NH_4^+, Rb^+, K^+, Na^+, Cs^+, Li^+, Mg^{2+}, Ca^{2+}, Ba^{2+}.

In Hofmeister lyotropic series, the chaotropic salts (magnesium sulfate and magnesium chloride) randomize the structure of the liquid water and thus tend to decrease the strength of hydrophobic interactions. In contrast, the kosmotropic salts (sodium, potassium or ammonium sulfates) promote hydrophobic interactions and protein precipitation, due to the higher 'salting-out' or molal surface tension increment effects [64].

Therefore, the selection of an adequate type of salt in the eluent results in significant alterations not only in overall protein retention, but also in separation selectivity [65].

Based on Melander and Horvath theory, Fausnaugh and Regnier [66] studied the effect of different types of salt on a set of related birds' lysozymes differing only in a few amino acids. They found that factors other than the surface tension increment suggested by Melander [26,29] affect protein retention, such as specific salt-protein interactions, and the hydration of the protein. Then the solvophobic theory [26,28,29] is not able to adequately explain protein retention differences in HIC due to the presence of a different salt type in the mobile phase. In order to explain this behaviour, the preferential interaction theory was developed based on the protein-salt interaction [11,41,67]. This theory has been used to investigate the effect of kosmotropic, chaotropic and neutral salts on protein binding and selectivity in HIC [11]. In that study, Xia et al. [11] compared the total number of released water molecules in the presence of different salt types, using the approach proposed by Perkins et al. [41], based on the fact that the number of water molecules released is a determining factor affecting protein retention in HIC. The authors demonstrated that selectivity reversals exist when different types of salt are used in the mobile phase, and the reversal concentration threshold varies among the different types of salts. For kosmotropic and neutral salts, such as $(NH_4)_2SO_4$ and NaCl, protein retention increases with an increase in salt concentration, but for chaotropic salts, such as NaSCN, a decrease in protein retention was observed with an increase in salt concentration [11,68]. The effect of salts on protein retention was explained by the number of released water molecules induced by different types of salt. Thus, selectivity of a certain salt type in HIC could be interpreted as differences in their ability to exclude water molecules from a protein surface and a resin surface.

5.1.3. Effect of mobile phase pH

The mobile phase pH is another factor affecting HIC performance. Although pH has been studied not as widely as salt type and salt concentration, some pH effects are relatively clear. Basic proteins, such as lysozyme (pI = 10.7) show high retention time when the mobile phase pH is close to its pI, while acidic proteins, such as human serum albumin (pI = 5.2) exhibit lower retention time with basic pH values [61]. When pH is close to a protein's pI, net charge is zero and hydrophobic interactions are maximum, due to the minimum electrostatic repulsion between the protein molecules allowing them to get closer. Fausnaugh and Regnier [66] studied the effect of mobile phase pH on different birds' lysozymes, and found that alterations in pH affected only the intercept of the $\log k'$ versus salt molality plot, indicating that the contact area between the protein and the matrix was not affected, but the charge of the amino acids that belong to the interaction zone would moderate the strength of the hydrophobic interaction. Another study showed that an optimal pH can be determined for a certain purification process based on HIC to obtain the highest purity and yield [69].

On the other hand, calorimetric measurements have been made in order to study the effects of mobile phase pH. A relationship based on Maxwell's equations was found by Alberty [70] to account for Gibbs energy changes due to pH, temperature and salt concentration. Given that the change in Gibbs energy is related to protein retention in HIC, which in turn is related to the number of water molecules released upon binding, the latter parameter was related to pH and temperature effects on HIC systems [71]. Recently, Xia et al. [71] compared the pH effect in the presence of different salt types (sodium sulfate, sodium chloride, and sodium thiocyanate), using different hydrophobic resins (phenyl and butyl Sepharose), based on Alberty's relationship. The authors found that the total number of released water molecules increased as the buffer pH became closer to the protein's pI, and decreased away from pI. They also found that the number of released water molecules for a pH change increased both with salt concentration and when changing from chaotropic to kosmotropic salts.

5.1.4. Other elution factors

Elution of proteins on HIC can be achieved in three different ways: (a) by changing the ion strength: this is the most widely used method; in this case the elution of proteins, regarding the order of increasing hydrophobicity, is carried out by decreasing the salt concentration; (b) by changing the polarity of the solvent: a decrease in the interaction is achieved by adding solvent as (poly) ethylene glycol, ethanol or (iso) propanol. The addition of polarity-decreasing agents can be made after salt has been removed from the column or concomitantly with the decrease of salt concentration; (c) by adding detergents or chaotropic agents: detergents or urea work as displacers of the proteins; they have been used mainly when proteins fail to elute at low salt concentration, which may lead to protein denaturation [72]. Only the simulations of changing the ion strength are described in the present review.

5.2. Stationary phase effect

Stationary phase consists of small non-polar groups (e.g. butyl, octyl or phenyl) attached to a hydrophilic polymer backbone. Therefore, the various types of stationary phases used in HIC can differ in chemical nature of the ligand, surface concentration of the ligand on the support, and chemical nature and particle size of the base support [48,63].

5.2.1. Ligand type effect

The most widely used ligands are linear chain alkanes (as butyl, octyl) and some aromatic groups (such as phenyl). An increase in the chain length of an alkyl ligand increases the strength of hydrophobic interaction between the protein and the resin; also, the specificity of the resin towards the adsorbed protein is changed [73]. However, resolution decreases when chain length is higher [48]. On the other hand, an increase in the substitution degree of the resin leads to an increase in the binding capacity of the stationary phase, due to the higher probability of forming multipoint attachment, and at times, it can be hard to elute the bound protein without denaturation [62]. Using the same type of ligand and the same type of base support, the selectivity of an HIC resin can be manipulated by changing the

Table 1
Prediction of protein retention times in HIC using $\phi_{aai\text{-Miyazawa-Jernigan}}$: quadratic model parameters ([79,80])

Resin	Salt type and concentration	Quadratic model coefficients		
		A	*B*	*C*
Phenyl sepharose	Ammonium sulfate (1 M)	11.79	−0.29	−0.35
	Ammonium sulfate (2 M)	−12.14	12.70	−1.74
	Sodium chloride (2 M)	−77.10	42.33	−5.13
	Sodium chloride (4 M)	−65.01	37.55	−4.71
Butyl sepharose	Ammonium sulfate (1 M)	36.76	−16.07	1.73
	Ammonium sulfate (2 M)	10.02	0.54	−0.38
	Sodium chloride (4 M)	−1.74	5.55	−1.01

The quadratic model is DRT $= A\phi^2 + B\phi + C$.

ligand density [22]. It has been shown that protein adsorption is a sigmoidal function of the surface concentration of immobilized alkyl ligands, then critical resin hydrophobicity exists and the threshold is determined by the ligand surface concentration [63].

Although it has become evident that stationary phase characteristics have an important role in HIC performance, only recently efforts have been made in order to investigate quantitatively the effect of the type of resin on chromatographic behaviour of proteins in HIC. Machold et al. [60] investigated protein retention in a variety of HIC resins of different hydrophobicity and different base support (a total number of 15 different commercially available resins). The retention data of seven model proteins was fitted using a polynomial function (Eq. (12)), and the area below the curve was considered as an apparent hydrophobicity value, characteristic of each protein-resin system. This approach was used to compare the performance of different HIC sorbents. The authors demonstrated that base support chemistry affects selectivity of an HIC system, which is also affected by type and density of the hydrophobic ligand, and to a lesser extent by resin particle size.

Ladiwala et al. [43] developed a quantitative structure property relationship (QSPR)-based model to investigate the role of ligand and backbone chemistry of stationary phase on protein retention in HIC systems. They evaluated a set of different HIC resins (a total number of four resins) with different chemical properties, and used 27 model proteins. Their results demonstrate that the selectivity of HIC systems can be significantly influenced by changing the ligand and/or the base support chemistry. The QSPR models provided good correlations between experimental and predicted data, and then these models were able to reflect differences in protein binding affinity on the different HIC resins investigated.

Based on the preferential interaction theory [41], Xia et al. [11] reported that the total number of released water molecules upon adsorption in an HIC process not only reflects a combined effect of salt on proteins but also on resin surfaces. The total

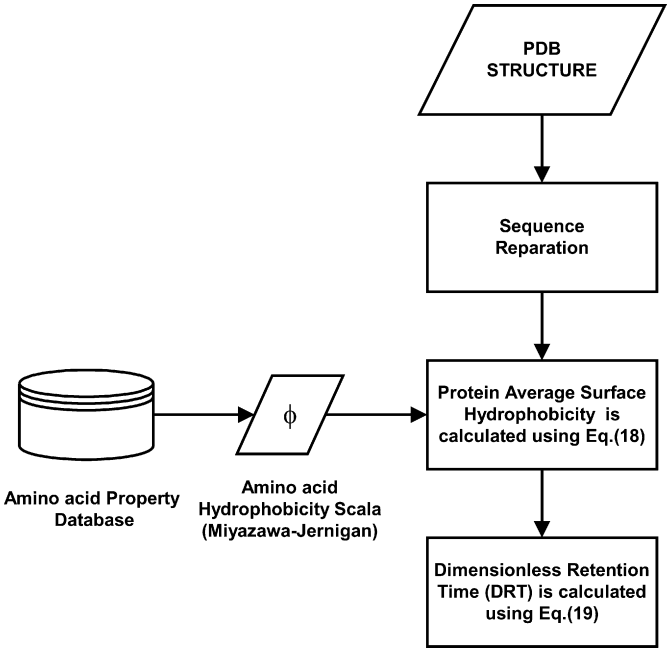

Fig. 2. Block diagram of the methodology based on average surface hydrophobicity using $\phi_{aai\text{-Miyazawa-Jernigan}}$ [46].

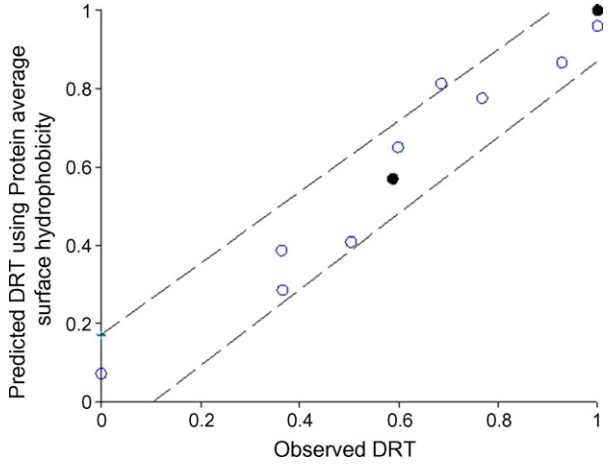

Fig. 3. Comparison between predicted and observed dimensionless retention times on phenyl sepharose 6FF 2 M ammonium sulfate using the methodology based on average surface hydrophobicity using $\phi_{aai\text{-Miyazawa-Jernigan}}$. (◯) Standard proteins (conalbumin, ribonuclease A, ovalbumin, chymotrypsinogen A, lysozyme, α-lactoalbumin, myoglobin, α-chymotrypsin, and thaumatin), (●) monomeric and multimeric proteins used for model validation (cytochrome C, β-lactoglobulin), (- - -) confidence intervals (95%).

number of released water molecules is proportional to the total hydrophobic area of the resin, which in turn reflects the chemical properties of the resin, i.e. hydrophobicity of the ligand, substitution degree, and hydrophobicity of the base support.

5.2.2. Support effect

The support effect is given by chemical nature and particle size. In the case of chemical nature of the support, it could be hydrophilic carbohydrates (e.g. cross-linked agarose), silica or

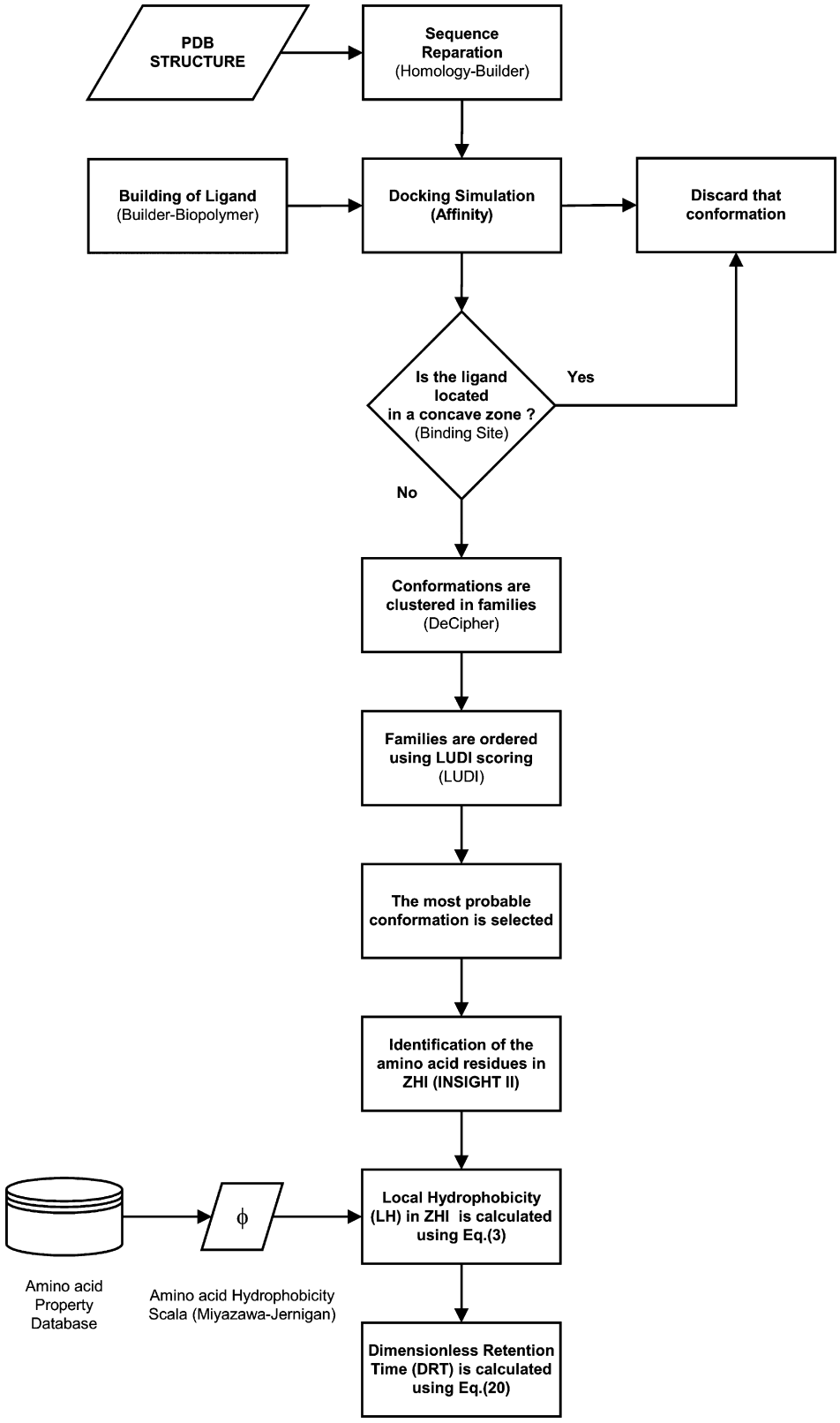

Fig. 4. Block diagram of the approach based on local hydrophobicity (LH) [82].

Table 2
Prediction of protein retention times in HIC using molecular docking: lineal model parameters [54,82]

Operating conditions	Lineal model coefficients			
	A[a]	B[a]	r^2	Observation
Phenyl sepharose	0.77	0.21	0.99	Model proposed by Mahn et al. [54], using RNAse T1, a variant of RNAse T1, and RNAse A
Ammonium sulfate (2 M)	2.80	−0.53	0.87	Model proposed by Lienqueo et al. [82], using α-lactalbumin, chymotrypsinogen A, lysozyme, RNAse T1, a variant of RNAse T1, and RNAse A

[a] The lineal model is DRT = ALH + B.

synthetic copolymer material [48]. Using the same type of ligand, the selectivity of the stationary phases can change according to the different type of supports [48].

5.3. Temperature effect

The temperature effect on HIC performance has been studied by many authors [61,62,71,74–76]. Vailaya and Horvath [76] established the existence of exothermodynamic relationships (enthalpy–entropy compensation) in the HIC process. Large and positive enthalpy and entropy changes were observed at low temperatures, which decrease with an increase in temperature. Hjertén et al. [62] established that an increase in temperature enhances protein retention, and a decrease in temperature usually promotes protein elution. This behaviour was explained by the fact that protein retention in HIC is an entropy-driven process, where the Gibbs energy is given by Eq. (1).

In HIC, ΔG is controlled by a positive entropy change; then it increases with an increase in temperature, as well as in the capacity factor k', as shown by Eq. (17) [75]:

$$\ln k' = \ln \varphi - \frac{\Delta G}{RT} \quad (17)$$

where R is the universal gas constant, φ is the phase ratio and T is the absolute temperature. However, temperature can also affect the conformational state of proteins and protein solubility. This phenomenon explains the inverse relation between protein retention and temperature observed in some cases [48].

Haidacher et al. [74] investigated the effect of temperature on the retention of amino acids derivatives in HIC, in the range from 5 to 50 °C. They found that the van't Hoff plots for different chromatographic systems were highly nonlinear and showed a maximum in the temperature interval used. This behaviour was attributed to a large negative heat capacity change associated with protein retention. The heat capacity change was found to increase with temperature, while enthalpy and entropy changes were positive at low temperatures, but negative at high temperatures. In other words, the authors demonstrated that retention in HIC is an entropy-driven process at low temperatures and an enthalpy-driven one at high temperatures.

Xia et al. [71] determined the capacity factor of lysozyme at different temperatures in butyl and phenyl sepharose resins. They found an increase in the capacity factor with temperature in the pH range from 5 to 7.5. Besides, they observed a combined effect of temperature and buffer pH on the capacity factor, while a similar behaviour was observed in both resins.

The state-of-the-art in this field may let us conclude that it would be possible to determine the optimum temperature, optimum pH and optimum chromatographic resin to reach the purification of a certain protein by HIC. Then, the next step in this field should be the simultaneous optimization of the main operational condition for a given protein in a certain HIC system.

6. Predictions of protein retention time in HIC

The prediction of protein retention in HIC could be very useful for selecting the optimal operating conditions. Following we summarize several proposed methodologies for predicting protein retention in HC using the protein 3-D structure and/or their amino acid composition.

6.1. Prediction of retention times of proteins in HIC using their 3-D structure

Lienqueo et al. [46] developed a quadratic model to predict the "dimensionless retention time" (DRT) in HIC with salt gradient elution based on "average surface hydrophobicity". This methodology has three steps. First, it is necessary to know or to estimate the 3-D structure of proteins, which means having the protein data bank (PDB) file [77] of the protein. Next, it is necessary to calculate the "average surface hydrophobicity"

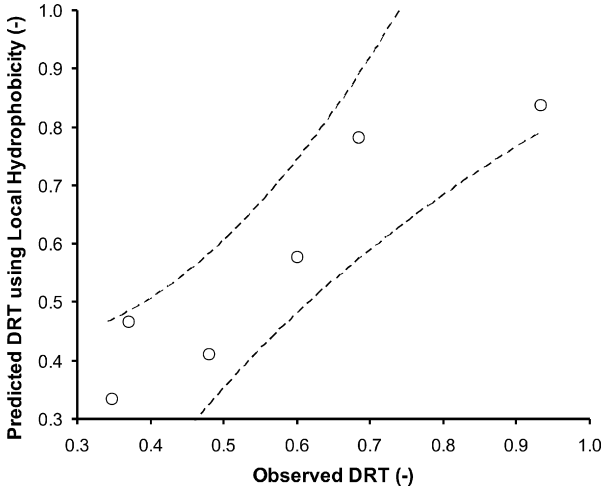

Fig. 5. Comparison between predicted and observed dimensionless retention times on phenyl sepharose 6FF 2 M ammonium sulfate using methodology based on local hydrophobicity (LH) (◯) standard proteins (α-lactalbumin, chymotrypsinogen A, lysozyme, RNAse T1, a variant of RNAse T1, and RNAse A), (- - -) confidence intervals (95%) [82].

of the protein, $\Phi_{surface}$, considering that each amino acid has a relative contribution to surface properties, as calculated in Eq. (2), and using $\phi_{aai\text{-Miyazawa-Jernigan}}$ as the amino acid hydrophobicity given by the normalized scale reported by Miyazawa and Jernigan [78].

$$\Phi_{surface} = \frac{\Sigma(s_{aai}\phi_{aai-miyazawa-jernigan})}{s_p} \quad (18)$$

Finally, by using a simple model it is possible to predict the chromatographic behaviour of proteins in HIC. The model can be written as follows:

$$DRT = A\Phi^2_{surface} + B\Phi_{surface} + C \quad (19)$$

where $\Phi_{surface}$ is the average surface hydrophobicity value calculated using Eq. (18). A, B and C are constants for each set of operating conditions. The values of A, B and C, for several operating conditions (phenyl sepharose-ammonium sulfate, phenyl sepharose-sodium chloride, butyl sepharose-ammonium sulfate and, butyl sepharose-sodium chloride), are summarized in Table 1 [79,80]. Additionally, a schematic diagram with the different steps of this methodology is shown in Fig. 2.

The model has been validated for several proteins with a relatively homogeneous surface hydrophobicity distribution, and results have always been adequate (see Fig. 3) [81]. The main disadvantage of this methodology is that it does not consider the effect of the distribution of the surface hydrophobicity and possible unfolding on protein retention [57]; however, it is a simple methodology, since it only needs the 3-D structure of the studied protein and the operating conditions.

On the other hand, Mahn et al. [57] studied the effect of surface hydrophobicity distribution of proteins on retention in HIC and developed a model based on "hydrophobic contact area" (HCA), which accounts for the contact area between the stationary phase and the protein when attached to the HIC resin. This model correlated particularly well with the DRT of different RNAses (RNAse S, RNAse T1, a variant of RNAse T1, and RNAse A) with similar average surface hydrophobicity, but with different and heterogeneous surface hydrophobicity distribution. Unfortunately, this methodology is experimentally exhaustive, and needs a large number of tedious experiments, which limits its utility.

After that, Mahn et al. [54] proposed a new methodology based on the parameter called "local hydrophobicity" (LH), defined in Eq. (3) below:

$$LH = \frac{\Sigma(s_{aai}\phi_{aai})}{s_{IZ}} \quad (3)$$

In this case, ϕ_{aai} is the amino acid hydrophobicity given by the normalized scale reported by Miyazawa and Jernigan [78].

This methodology used molecular docking simulations to identify the interaction zone of the protein with the hydrophobic ligand and computed LH [5]. The results showed that there is a linear correlation between the parameters LH and DRT for proteins with heterogeneous surface hydrophobicity distribution ($r^2 > 0.99$). This correlation is shown in the following equation:

$$DRT = A \cdot LH + B \quad (20)$$

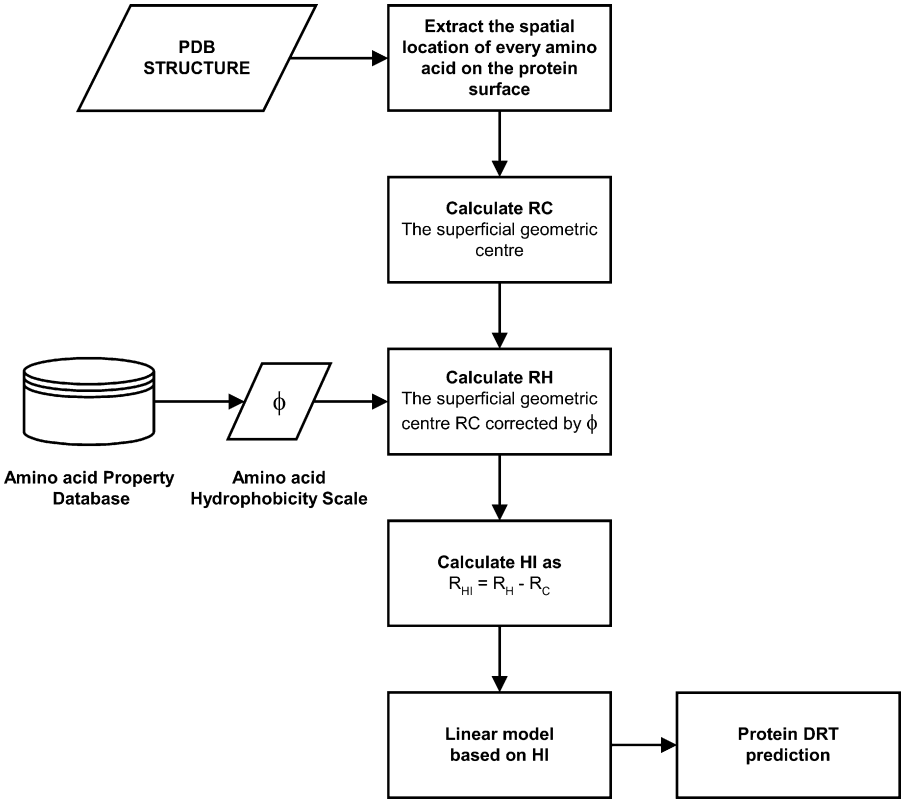

Fig. 6. Flowchart of the methodology based on hydrophobic imbalance (HI) [83].

A and B are constants for each set of operating conditions. The values of A and B, for phenyl sepharose-2 M ammonium sulfate, are summarized in Table 2.

The disadvantages of the methodology proposed by Mahn et al. [54] are that it was carried out only for a small set of homologous proteins and it is computationally expensive. Therefore, Lienqueo et al. [82] have recently extended and automated this methodology. A schematic diagram with the different steps of this methodology is shown in Fig. 4. They studied seven proteins (α-lactalbumin, chymotrypsinogen A, lysozyme, RNAse S, RNAse T1, a variant of RNAse T1, and RNAse A) with heterogeneous and homogeneous surface hydrophobicity distribution, and used phenyl sepharose-2 M ammonium sulfate at experimental conditions; the results showed a suitable correlation level, $r^2 > 0.87$ (see Fig. 5 and Table 2) [82]. Then, this methodology can be used to satisfactorily predict the retention time in HIC for proteins with heterogeneous and homogeneous surface hydrophobicity distribution and without a large number of tedious experiments, only using computational simulation. One disadvantage of this methodology was that it was carried out for a single HIC medium and process conditions (phenyl sepharose-2 M ammonium sulfate); then, it is necessary to evaluate this methodology under other operational conditions, e.g. phenyl sepharose-sodium chloride, butyl sepharose-ammonium sulfate and, butyl sepharose-sodium chloride. Another disadvantage of this methodology was that it requires a considerable amount of computational time to identify the most probable LH interaction zone by docking simulation. Finally, another disadvantage of this methodology was that it does not consider possible unfolding on retained proteins.

On the other hand, Salgado et al. proposed two mathematical tools to quantify the amino acid distribution on the protein surface: the hydrophobic imbalance (HI) [83] and the statistical description of the surface characteristic [84]. These models

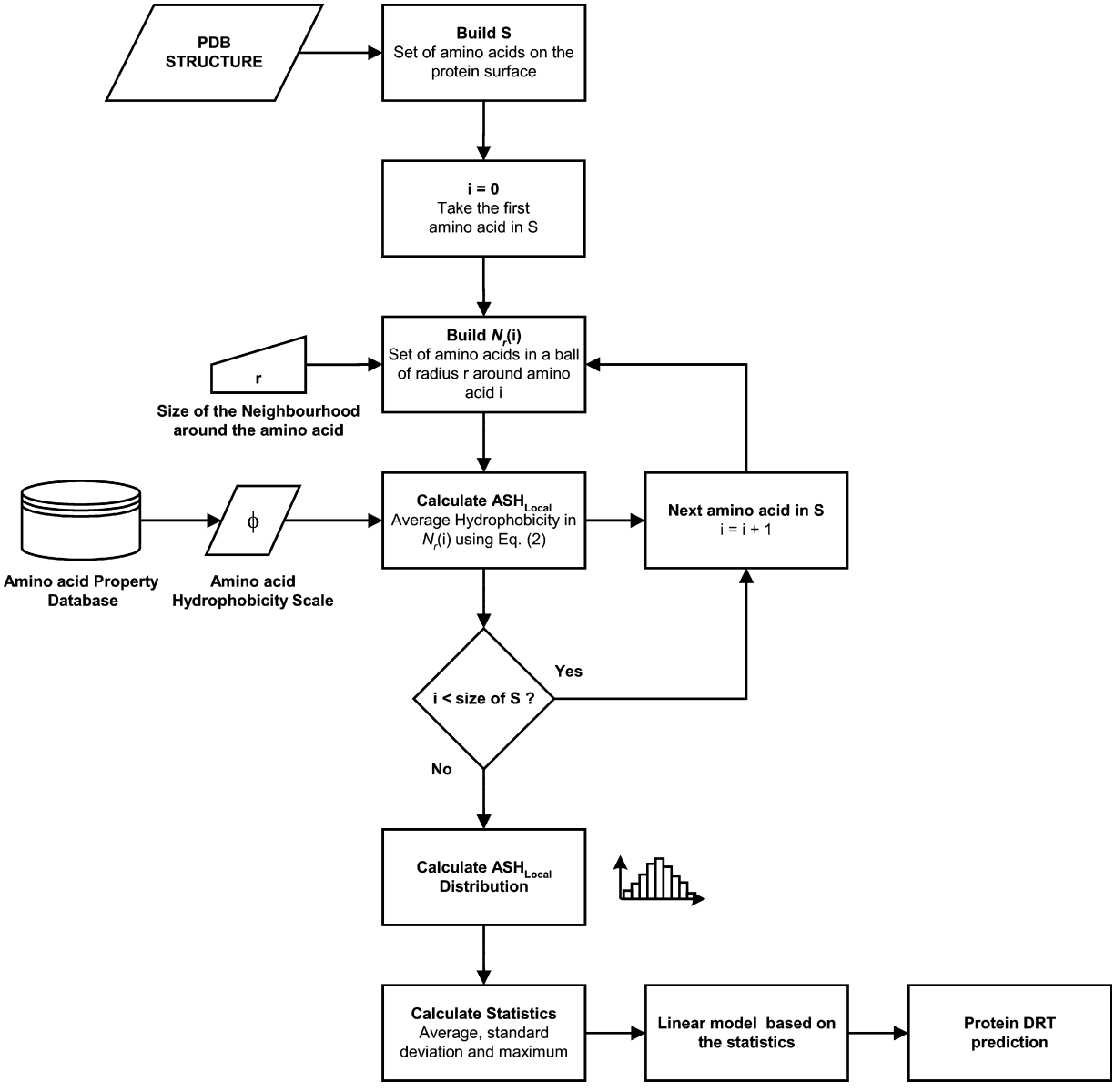

Fig. 7. Block diagram of the approach based on statistical description of their surface amino acid distribution [84].

are significantly simpler and computationally inexpensive compared to those reported by Mahn et al. [54] and Lienqueo et al. [82].

The first tool, the hydrophobic imbalance (HI) [83], is obtained from the characteristics of the protein surface and represents the displacement of the superficial geometric centre of the protein when the effect of the hydrophobicity of each amino acid is considered. A schematic diagram with the different steps of this methodology is shown in Fig. 6. The HI was calculated for a set of four RNAses reported in Mahn et al. [57], whose DRT is hard to predict using the average superficial hydrophobicity. The correlation coefficients obtained by the HI model were remarkably better (at least 67%) than those achieved by the models based on HCA [57] and LH [54]. Additionally, the HI model was tested with a set of 15 proteins, with molecular mass ranging from 11,000 to 76,000 Da, and pI from 4.0 to 11.8 (more details see Table 3), showing improvement in the predictive capacity displayed by models based on ASH by 9.1%. Also, a linear multivariable model based on characteristics determined from HI was studied. By using this multivariable model, a correlation coefficient of 0.899 was obtained. In addition, an improvement of 31.8% of the predictive characteristics shown by previous models based on ASH was achieved [83].

The second approach proposed by Salgado et al. [84] uses a statistical description of the surface amino acid distribution in order to predict the DRT. The characterization of the hydrophobicity distribution on the protein surface was carried out through the systematic calculation of a local ASH (ASH_{local}) in the neighbourhood of each of the amino acids located on the protein surface, on a one-by-one basis. The distribution of this local ASH was characterized using simple statistics as the average, maximum and standard deviation. This way, ASH_{avg} and ASH_{max} gave the average hydrophobicity and the hydrophobic content of the most hydrophobic cluster or hotspot, respectively. On the other hand, ASH_{std} quantified the heterogeneity of the hydrophobicity distribution on the protein surface. Again, those statistics were used to model the DRT in the same set of RNAses whose DRT is hard to predict using the ASH, and their results were very good. The correlation coefficients obtained by the models based on those statistics were almost twice as high as those obtained by the models based on HCA [57] and LH [54], and slightly lower than that obtained by HI [83]. When the DRT predictive capacity of linear models which use those statistics was analysed, it was found that their performance is in general superior to those reported previously. The best linear model is obtained with ASH_{max}, which shows a predictive capacity 26.9% better than that obtained by the best ASH model, and 19.5% better than that by the HI model. This result is significant since ASH_{max} quantifies directly the hydrophobicity of the high hydrophobicity clusters on the protein surface, and it has been reported that the presence of those clusters on the protein surface favours the interaction of the protein with the HIC stationary matrix [54,57,83,85]. A flowchart of this methodology is shown in Fig. 7.

Additionally, a linear multivariable model which combines the HI and ASH_{max} approaches was developed by Salgado et al. [84]. This model uses Aboderin's hydrophobicity scale [86], which is an index of the mobilities of amino acids in chromatography. The dispersion between the experimental values and the prediction carried out by this model is smaller than that observed in previous models showing a predictive capacity 8.7% better than that obtained by the multivariate model based on HI.

An obvious common weakness of DRT predictive models previously described [46,54,82] is that they require the three-dimensional structure of the protein to perform their calculations. Even though the number of solved protein structures on the PDB database [77] grows continuously (being near 36,000 structures at May 2006), this data is not readily available, or is very difficult or expensive to obtain. If the three-dimensional structure data is not available, an alternative is to use an estimate

Table 3
Molecular mass and isoelectric point of the different used proteins for testing the hydrophobic imbalance model

Protein	PDB ID[a]	Mr[b] [Da]	pI[c]
Cytochrome C	1HRC	11701	10.15
Myoglobin	1YMB	16951	8.13
Conalbumin	1OVT	75827	6.63
Ovalbumin	1OVA	42750	5.01
Lysozyme	2LYM	14313	10.76
Thaumatin	1THV	22204	9.69
Chymotrypsinogen A	2CHA	25651	9.69
β-Lactoglobulin	1CJ5	18375	4.54
α-Amylase	1BLI	55193	6.00
α-Chymotrypsin	4CHA	25207	9.69
α-lactalbumin	1A4V	14978	4.47
Ribonuclease S	1RBC	13196	8.64
Ribonuclease A	1AFU	13574	9.77
Ribonuclease T1 wild type	RGC	11072	4.27
Ribonuclease T1 variant Y45W/W59Y	1TRP	11072	4.27

[a] PDB ID is the file code given by the protein data bank.
[b] Mr is the molecular mass of the protein.
[c] pI is the isoelectric point of the protein.

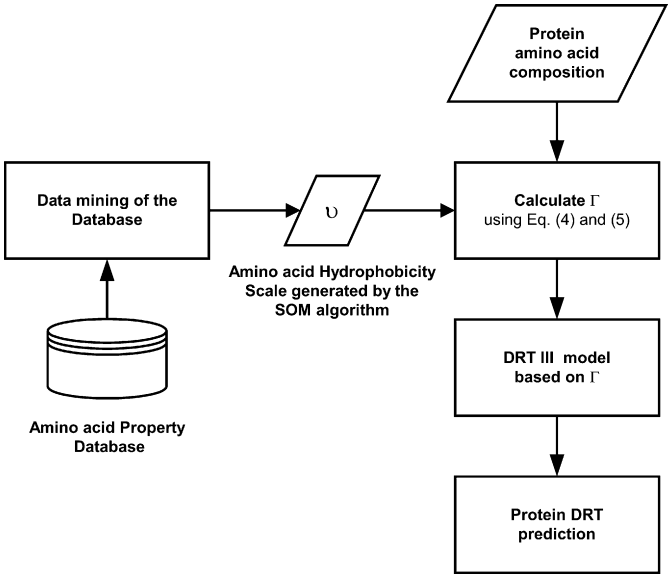

Fig. 8. Flowchart of the methodology based on the amino acidic composition of the target protein [55].

that is developed using comparative modelling or, in some cases, *ab initio* models. However, these methodologies are known to be quite complex and computationally expensive.

A natural approach is to use features associated to the amino acidic sequence of the protein, since the number of sequences in the UniProtKB/Swiss-Prot database [47] is almost six-fold the number of structures in PDB (being near 220,000 sequences in the release of May 2006). Below is a review of different methodologies for predicting protein retention in HIC using only their amino acid composition.

6.2. *Prediction of protein retention times in HIC using only their aminoacidic composition*

The first to use this kind of approach has been Salgado et al. [87]. They proposed three models for predicting the dimensionless retention time of proteins in HIC [87]. These models use only the amino acidic composition to perform their calculations and, therefore, they do not require the protein's amino acid sequence, its secondary structure, or its three-dimensional structure.

The core idea behind the models proposed by Salgado et al. [55] is to predict the average superficial hydrophobicity of a protein by using different assumptions about the amino acid's tendency to be exposed to the solvent: the first one states that all amino acids are fully exposed (called DRT I model), the second one uses a simple correction factor considering the general tendency of each amino acid to be exposed (called DRT II model), and the last one is based on a linear estimation of the amino acidic surface composition (called DRT III model). Moreover, these models use a collection of 74 amino acidic property vectors (APV) which cover a wide spectrum of physical, chemical and biological amino acidic characteristics plus a collection of 6388 vectors derived from these, using *k*-means [88] and self-organizing maps (SOM) [89] algorithms. A schematic diagram with the different steps of this methodology is shown in Fig. 8.

By using these mathematical models and data, Salgado et al. [87] achieved the same predictive performance given by the

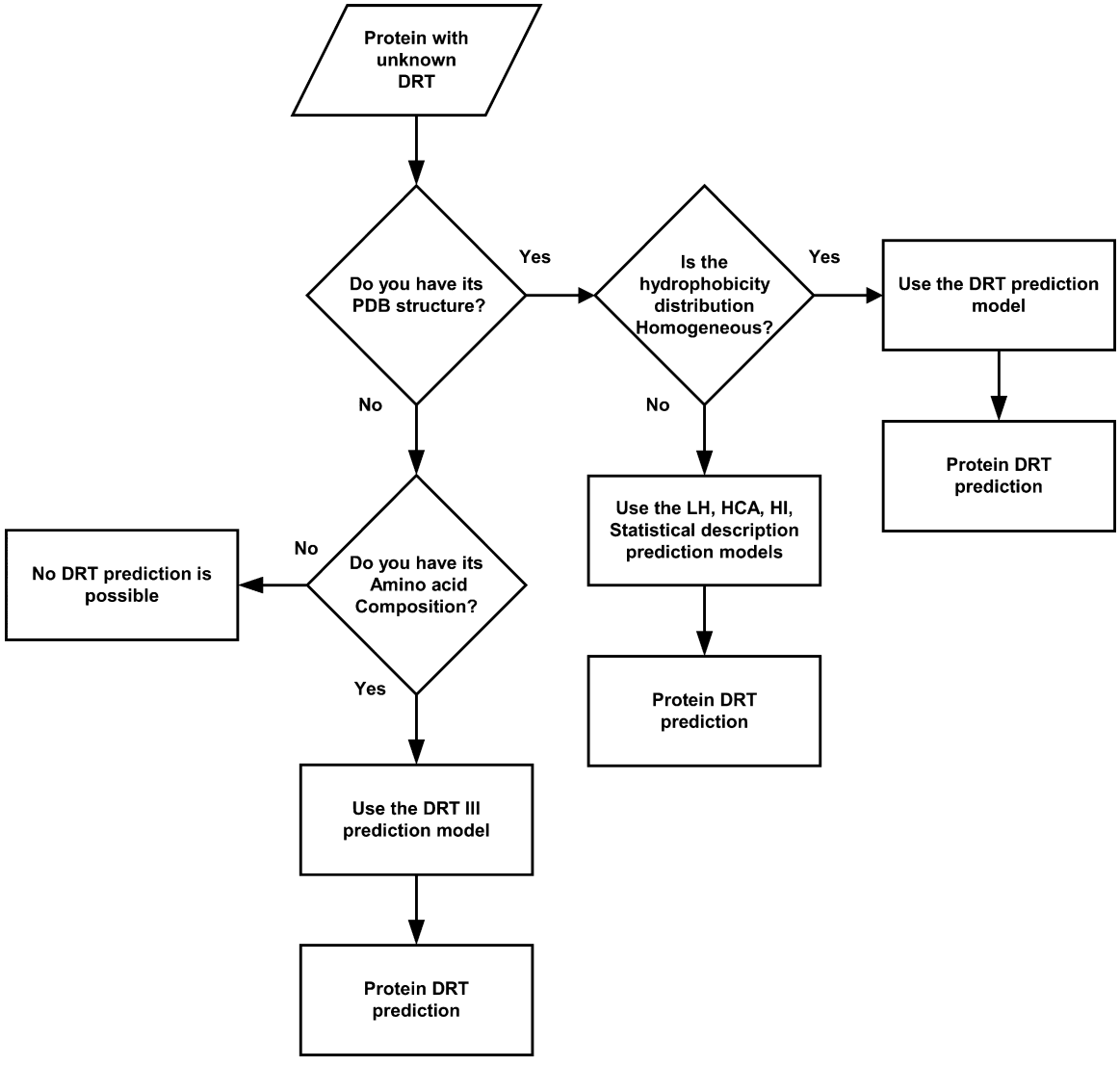

Fig. 9. Flowchart to explain how to implement the different methodologies to carry out computational experiment for determining suitable purification of a target protein.

models that use the three-dimensional protein structure. The predictive capacity of the model was estimated using the Jack Knife re-sampling method (leave-one-out), which is a widely known methodology [90,91] considered as an effective tool for the evaluation of predictor models [92,93]. The best results were obtained by the DRT III model with a vector generated by the SOM algorithm. This vector was interpreted as a hydrophobicity scale based to a certain extent on the tendency of the amino acids to be inside proteins. The performance of DRT III with that vector was even 5% better than the one observed with the model using the three-dimensional structure of proteins. The disadvantage of these models was that they do not consider the way in which amino acids are distributed on the protein surface. Consequently, their models do not consider how the hydrophobicity is distributed. This distributional effect is only possible to be evaluated if the 3-D structure of protein is well-known and using methodologies based on LH [54,82], HI [83], ASH_{avg}, ASH_{max} or ASH_{std} [84], described above.

On the other hand, Ladiwala et al. [43] have investigated the protein retention in HIC using quantitative structure retention relationship. This proposed methodology used over 20 descriptors based on 2-D and/or 3-D structure for predicting the protein retention under different stationary phases (ligand and backbone chemistry and ligand density). The QSRR models showed a good correlation level between experimental and predicted time ($0.96 > r^2 > 0.84$). Additionally, the QSRR models could be useful for understanding the physicochemical effects that contribute to protein retention on different HIC media; therefore, these methodologies facilitate the in-silico optimization of HIC processes.

6.3. Selecting the best in-silico operating condition in HIC

Finally, for selecting the best in-silico operating condition in HIC, it is necessary to predict the DRT of target and contaminant proteins at different operating conditions (i.e. type of salt, ionic strength, chemical nature of the backbone, type of hydrophobic ligand, and substitution level of the stationary phase), using methodologies shown in Fig. 9. After that, it is necessary to estimate the resolution of HIC for each operating condition, and finally select the operating condition that gives the maximum resolution. Additional methodologies could be useful for designing a rational protein purification process that involves an HIC step.

7. Conclusions

In our view, the main factors affecting protein chromatographic behaviour in HIC are salt type and ionic strength of the mobile phase and chemical nature of the backbone, type of hydrophobic ligand, and substitution level of the stationary phase. The effect on protein retention time in HIC of each of these factors could be evaluated by computational experiments, using the different methodologies described. For example, in the simplest case, when the target protein has a well-known 3-D structure and relatively homogeneous surface hydrophobic distribution, it is possible to use the methodology based on average surface hydrophobicity proposed by Lienqueo et al. [46]. In the case when proteins have a well-known 3-D structure but heterogeneous surface hydrophobic distribution, it is possible to use the methodology proposed by Mahn et al. [54] for determining the local surface hydrophobicity, or the methodologies proposed by Salgado, based on hydrophobic imbalance [83] or statistical descriptions of the surface amino acid distribution [84]. On the other hand, in the case of proteins with only the amino acidic composition is well-known, it is possible to predict the average surface hydrophobicity of the protein based on methodologies proposed by Salgado et al. [55], which use different assumptions about the tendency of amino acids to be exposed to the solvent. The main disadvantage of the latter methodology is that it does not consider the hydrophobic distributions of amino acids on the protein surface; however, this methodology needs basic information about the protein, i.e. its amino acidic composition.

Then, considering the different developed methodologies for predicting protein retention time in HIC, it could be possible to carry out computational experiments by varying the different operating conditions for the purification of a target protein, then select the best in-silico conditions, and last but not least, design a rational protein purification process that involves an HIC step.

Acknowledgements

This review was supported by Fondecyt project 1030668.

References

[1] J.A. Asenjo, B.A. Andrews, J. Mol. Recognit. 17 (2004) 236.
[2] B.S. Warren, P. Kusk, R.G. Wolford, G.L. Hager, J. Biol. Chem. 271 (1996) 11434.
[3] Z. Hrkal, J. Rejnkova, J. Chromatogr. 242 (1982) 385.
[4] D.E. Comings, A.G. Miguel, B.H. Lesser, Biochim. Biophys. Acta 563 (1979) 253.
[5] G. Morris, D. Goodsell, R. Halliday, R. Huey, W. Hart, R. Belew, A. Olson, J. Comput. Chem. 19 (1998) 1639.
[6] S. Lefort, P. Ferrara, J. Chromatogr. 361 (1986) 209.
[7] J.E. Seely, C.W. Richey, J. Chromatogr. A 908 (2001) 235.
[8] T.P. Hopp, K.R. Woods, Proc. Natl. Acad. Sci. USA 78 (1981) 3824.
[9] M.M. Diogo, J.A. Queiroz, G.A. Monteiro, S.A. Martins, G.N. Ferreira, D.M. Prazeres, Biotechnol. Bioeng. 68 (2000) 576.
[10] S. Fexby, L. Bulow, Trends Biotechnol. 22 (2004) 511.
[11] F. Xia, D. Nagrath, S. Garde, S.M. Cramer, Biotechnol. Bioeng. 87 (2004) 354.
[12] A. Tiselius, Arkiv för Kemi, Mineralogi Geologi 26B (1948) 1.
[13] C.C. Shepard, A. Tiselius, Discussions of the Faraday Society, 7, Hazell Watson and Winey, London, 1949.
[14] S. Shaltiel, Z. Er-El, Proc. Natl. Acad. Sci. USA 70 (1973) 778.
[15] B.H. Hofstee, Anal. Biochem. 52 (1973) 430.
[16] S. Hjertén, J. Chromatogr. 87 (1973) 325.
[17] J. Porath, L. Sundberg, N. Fornstedt, I. Olson, Nature 245 (1973) 465.
[18] P.L. Privalov, S.J. Gill, Adv. Protein Chem. 39 (1988) 191.
[19] K.O. Erikkson, in: J.C. Janson, L. Ryden (Eds.), Protein Purification: Principles, High-Resolution Methods, and Applications, Wiley-Liss, New York, 1998.
[20] W. Kauzmann, Adv. Protein Chem. 14 (1959) 1.
[21] K.A. Dill, Biochemistry 29 (1990) 7133.
[22] F.-Y. Lin, W.-Y. Chen, R.-C. Ruaan, H.-M. Huang, J. Chromatogr. A 872 (2000) 37.
[23] H. Jennissen, Methods Mol. Biol. 305 (2005) 81.

[24] H. Huang, F.Y. Lin, W.Y. Chen, R.C. Ruaan, J. Colloid Interface Sci. 229 (2000) 600.
[25] Y. Tsai, F. Lin, W. Chen, C. Lin, Colloids, Surf. A: Physicochem. Eng. Aspects 197 (2002) 111.
[26] W. Melander, C. Horvath, Arch. Biochem. Biophys. 183 (1977) 200.
[27] G.S. Manning, Q. Rev. Biophys. 11 (1978) 179.
[28] O. Sinanoglu, S. Abdulnur, Fed. Proc. 24 (1965) S12.
[29] W.R. Melander, D. Corradini, C. Horvath, J. Chromatogr. 317 (1984) 67.
[30] A. Staby, J. Mollerup, J. Chromatogr. A 734 (1996) 205.
[31] S. Oscarsson, J. Chromatogr. B Biomed. Appl. 666 (1995) 21.
[32] Y. Kato, T. Kitamura, T. Hashimoto, J. Chromatogr. 333 (1985) 202.
[33] D. Josic, W. Schutt, J. Van Renswoude, W. Reutter, J. Chromatogr. 353 (1986) 13.
[34] J.L. Fausnaugh, L.A. Kennedy, F.E. Regnier, J. Chromatogr. 317 (1984) 141.
[35] J.L. Fausnaugh, E. Pfannkoch, S. Gupta, F.E. Regnier, Anal. Biochem. 137 (1984) 464.
[36] Y. Kato, K. Nakamura, T. Kitamura, H. Moriyama, M. Hasegawa, H. Sasaki, J. Chromatogr. A 971 (2002) 143.
[37] S.L. Wu, A. Figueroa, B.L. Karger, J. Chromatogr. 371 (1986) 3.
[38] T.T. Jones, E.J. Fernandez, Biotechnol. Bioeng. 87 (2004) 388.
[39] J.L. McNay, E.J. Fernandez, J. Chromatogr. A 849 (1999) 135.
[40] W.-Y. Chen, J.-F. Lee, C.-F. Wu, H.-K. Tsao, J. Colloid Interface Sci. 190 (1997).
[41] T. Perkins, D. Mak, T.W. Root, L. EN, J. Chromatogr. A 766 (1997) 1.
[42] T. Arakawa, S.N. Timasheff, Biochemistry 23 (1984) 5924.
[43] A. Ladiwala, F. Xia, Q. Luo, C.M. Breneman, S.M. Cramer, Biotechnol. Bioeng. 93 (2006) 836.
[44] N. Jakobsson, M. Degerman, B. Nilsson, J. Chromatogr. A 1099 (2005) 157.
[45] K.M. Biswas, D.R. DeVido, J.G. Dorsey, J. Chromatogr. A 1000 (2003) 637.
[46] M.E. Lienqueo, A. Mahn, J.A. Asenjo, J. Chromatogr. A 978 (2002) 71.
[47] A. Bairoch, R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S. Yeh, Nucleic Acids Res. 33 (2005) D154.
[48] J.A. Queiroz, C.T. Tomaz, J.M. Cabral, J. Biotechnol. 87 (2001) 143.
[49] K. Berggren, A. Wolf, J.A. Asenjo, B.A. Andrews, F. Tjerneld, Biochim. Biophys. Acta 1596 (2002) 253.
[50] C.C. Bigelow, J. Theor. Biol. 16 (1967) 187.
[51] D.F. Waugh, Adv. Protein Chem. 9 (1954) 325.
[52] H.F. Fisher, Proc. Natl. Acad. Sci. USA 51 (1964) 1285.
[53] N. Alizadeh-Pasdar, E.C. Li-Chan, J. Agric. Food Chem. 48 (2000) 328.
[54] A. Mahn, G. Zapata-Torres, J.A. Asenjo, J. Chromatogr. A 1066 (2005) 81.
[55] J.C. Salgado, I. Rapaport, J.A. Asenjo, J. Chromatogr. A 1075 (2005) 133.
[56] J. García-Dominguez, J. Díez-Masa, Pure Appl. Chem. 3 (2001) 969.
[57] A. Mahn, M.E. Lienqueo, J.A. Asenjo, J. Chromatogr. A 1043 (2004) 47.
[58] T. Arakawa, S.N. Timasheff, Biochemistry 23 (1984) 5912.
[59] T. Arakawa, Arch. Biochem. Biophys. 248 (1986) 101.
[60] C. Machold, K. Deinhofer, R. Hahn, A. Jungbauer, J. Chromatogr. A 972 (2002) 3.
[61] S. Hjertén, K. Yao, K.-O. Eriksson, B. Johansson, J. Chromatogr. 359 (1986) 99.
[62] S. Hjertén, J. Rosengren, S. Pahlman, J. Chromatogr. 101 (1974) 281.
[63] H. Jennissen, J. Colloid Interface Sci. 111 (1986) 570.
[64] J. Porath, J. Chromatogr. 376 (1986) 331.
[65] G. Rippel, L. Szepesy, Chromatogr. A 664 (1994) 27.
[66] J.L. Fausnaugh, F.E. Regnier, J. Chromatogr. 359 (1986) 131.
[67] F. Xia, D. Nagrath, S.M. Cramer, J. Chromatogr. A 989 (2003) 47.
[68] B. Roettger, J. Myers, M. Ladisch, F. Regnier, Biotechnol. Prog. 5 (1989).
[69] Y. Sanz, F. Mulholland, F. Toldra, J. Agric. Food Chem. 46 (1998) 349.
[70] R.A. Alberty, J. Biol. Chem. 244 (1969) 3290.
[71] F. Xia, D. Nagrath, S.M. Cramer, J. Chromatogr. A 1079 (2005) 229.
[72] M. Sandberg, L. Eriksson, J. Jonsson, M. Sjostrom, S. Wold, J. Med. Chem. 41 (1998) 2481.
[73] Z. Er-el, Y. Zaidenzaig, S. Shaltiel, Biochem. Biophys. Res. Commun. 49 (1972) 383.
[74] D. Haidacher, C. Horvath, F. Dondi, G. Guichon (Eds.), Physicochemical Basis of Hydrophobic Interaction Chromagraphy, Kluwer, Dordrecht, 1992.
[75] Z. EL Rassi, J. Chromatogr. A 720 (1996) 93.
[76] A. Vailaya, C. Horvath, Biophys. Chem. 62 (1996) 81.
[77] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235.
[78] S. Miyazawa, R. Jernigan, Macromolecules 18 (1985) 534.
[79] A. Mahn, J.A. Asenjo, Biotechnol. Adv. 23 (2005) 359.
[80] M. Lienqueo, A. Mahn, Chem. Eng. Technol. 28 (2005) 1326.
[81] M.E. Lienqueo, A. Mahn, L. Vasquez, J.A. Asenjo, J. Chromatogr. A 1009 (2003) 189.
[82] M. Lienqueo, A. Mahn, G. Navarro, J. Salgado, T. Perez-Acle, I. Rapaport, J. Asenjo, J. Mol. Recognit. (2006) 260.
[83] J.C. Salgado, I. Rapaport, J.A. Asenjo, J. Chromatogr. A 1107 (2006) 110.
[84] J.C. Salgado, I. Rapaport, J.A. Asenjo, J. Chromatogr. A 1107 (2006) 120.
[85] M. Jonsson, M. Skepo, F. Tjerneld, P. Linse, J. Phys. Chem. B 107 (2003) 5511.
[86] A.A. Aboderin, Int. J. Biochem. 2 (1971) 537.
[87] J.C. Salgado, I. Rapaport, J.A. Asenjo, J. Chromatogr. A 1098 (2005) 44.
[88] G.A.F. Seber, Multivariate Observations, Wiley, New York, 1984.
[89] T. Kohonen, Self Organization and Associative Memory, Springer-Verlag, Berlin-Heidelberg-New York-Tokio, 1989.
[90] K.C. Chou, C.T. Zhang, Crit. Rev. Biochem. Mol. Biol. 30 (1995) 275.
[91] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, London, 1979.
[92] G.P. Zhou, J. Protein Chem. 17 (1998) 729.
[93] G.P. Zhou, N. Assa-Munt, Proteins 44 (2001) 57.
[94] R. Cowan, R.G. Whittaker, Pept. Res. 3 (1990) 75.
[95] K.J. Wilson, A. Honegger, R.P. Stotzel, G.J. Hughes, Biochem. J. 199 (1981) 31.
[96] J.L. Fauchere, V.E. Pliska, Eur. J. Med. Chem. 18 (1983) 369.
[97] D.J. Abraham, A.J. Leo, Proteins 2 (1987) 130.
[98] D.H. Wertz, H.A. Scheraga, Macromolecules 11 (1978) 9.