Prediction of protein retention in hydrophobic interaction chromatography

Andrea Mahn[a,*], Juan A. Asenjo[b]

[a]*Center for Molecular Studies of Cell, Institute for Biomedical Sciences, University of Chile, Independencia 1027, Santiago, Chile*
[b]*Centre for Biochemical Engineering and Biotechnology, Chemical Engineering Department, University of Chile, Independencia 1027, Santiago, Chile*

Abstract

Hydrophobic interaction chromatography (HIC) is a powerful technique for protein separation. This review examines methodologies for predicting protein retention time in HIC involving elution with salt gradients. The methodologies discussed consider three-dimensional structure data of the protein and its surface hydrophobicity. Despite their limitations, the methods discussed are useful in designing purification processes for proteins and easing the tedious experimental work that is currently required for developing purification protocols.

Keywords: Hydrophobic interaction chromatography; Average surface hydrophobic; Hydrophobic contact area; Local hydrophobicity

Contents

1. Introduction 360
2. Methodology 1: average surface hydrophobicity ($\phi_{surface}$) 361
3. Methodology 2: hydrophobic contact area (HCA) 363

* Corresponding author. Fax: +56 2 7353510.
E-mail address: amahn_2000@yahoo.es (A. Mahn).

4. Methodology 3: local hydrophobicity–molecular docking 364
5. Concluding remarks 367
References. 367

1. Introduction

Hydrophobic interaction chromatography (HIC) is a key technique that is used in purifying proteins. The HIC process consists of injecting a protein sample in a hydrophobic column under conditions of high salt concentration. Elution is typically achieved by decreasing the ionic strength or the concentration of salt in the mobile phase. During HIC, a protein coming in contact with the hydrophobic ligands of the resin or chromatography matrix experiences a spatial reorientation. The hydrophobic ligands of the resin interact with the exposed hydrophobic zones of the protein, to reversibly bind the protein to the resin. The adsorption capacity of HIC resins and the resolving power of HIC are similar to that seen in ion exchange chromatography of proteins (Fausnaugh et al., 1984).

The main system characteristics affecting protein retention in HIC are concentration and type of salt (Melander and Horvath, 1977; Sofer and Hagel, 1998) and density and type of hydrophobic ligand attached to the matrix (Jennissen, 2000). The main physicochemical property of proteins that determines chromatographic behavior in HIC is hydrophobicity. At present, no universally agreed single measure exists for hydrophobicity of proteins. There is consensus that a protein's hydrophobicity is determined by the hydrophobic contributions of its amino acid residues (Tanford, 1962; Eriksson, 1998). Hydrophobicity has been estimated in several possible ways, including measures such as "average hydrophobicity" (Tanford, 1962), "non-polar chain frequency" (Waugh, 1954), "polarity ratio" (Fisher, 1964), and "net hydrophobicity" (Eriksson, 1998). In HIC protein retention occurs because of a surface adsorption phenomenon, therefore, the use of "average surface hydrophobicity" has been suggested for characterizing retention behavior (Lienqueo et al., 2002). Average surface hydrophobicity can be estimated from a knowledge of the protein's three-dimensional structure by taking into account the hydrophobic contribution of the amino acid residues that are exposed on the surface.

Hydrophobicity is of course not uniform over the entire surface of a large protein molecule. Therefore, the distribution of surface hydrophobicity can be important in HIC. Indeed, it has been reported that protein retention in HIC is considerably affected by the distribution of hydrophobic patches on a protein's surface (Fausnaugh and Regnier, 1986; Mahn et al., 2004).

As HIC is widely used in downstream processing of proteins, we have focused on developing methods for predicting protein retention time in HIC involving gradient elution. Ability to predict retention times will greatly ease the design of protein purification processes and reduce the need for numerous tedious trial runs. Here we review three methods that have shown promise in predicting protein–resin interactions in HIC.

2. Methodology 1: average surface hydrophobicity ($\phi_{surface}$)

A methodology has been developed based on the average surface hydrophobicity ($\phi_{surface}$) of proteins. The latter is estimated starting from the three-dimensional structural data and considering only the solvent accessible amino acid residues of the protein. This methodology uses $\phi_{surface}$ to estimate the protein retention time via simple quadratic models. The coefficients of the quadratic equations depend on the chromatographic conditions used in the HIC (initial salt concentration, type of salt, type of matrix) (Lienqueo et al., 2002).

To describe protein retention in HIC, a "dimensionless retention time" (DRT) is defined, as follows:

$$\mathrm{DRT} = \frac{RT - t_0}{t_f - t_0}. \tag{1}$$

Here RT is the time corresponding to the peak maximum in the chromatogram, t_0 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 the hydrophobic column does not retain a protein, DRT for that protein is 0. If a protein elutes only after the salt gradient has been completed, its DRT equals unity.

This method based on average surface hydrophobicity assumes that each amino acid on a protein's surface has a hydrophobic contribution proportional to its solvent accessible area; thus:

$$\phi_{surface} = \frac{\sum (s_{aai}\phi_{aai})}{s_p} \tag{2}$$

where $\phi_{surface}$ is the calculated value of the surface hydrophobicity for a given protein, i ($i=1,\ldots, 20$; different i-values indicate different standard amino acids), s_{aai} is the solvent accessible area occupied by the amino acid i, ϕ_{aai} is the hydrophobicity value assigned to amino acid i by any amino acid hydrophobicity scale, and s_p is the total solvent accessible area of the entire protein. It has to be noted that for proteins with a prosthetic group s_p is bigger than $\sum s_{aai}$ and for proteins without prosthetic group, these values are equal.

For estimating $\phi_{surface}$, it is necessary to choose an amino acid hydrophobicity scale that gives a certain hydrophobicity value to each standard amino acid. A study of the existing scales was carried out by Lienqueo et al. (2002), where the scales were normalized and compared based on their ability to describe proteins' hydrophobicity related to their behavior in HIC. The most appropriate hydrophobicity scale was found to be that developed by Miyazawa and Jernigan (1985). This scale must be used in a normalized form so that the ϕ_{aai} values range from 0 (most hydrophilic amino acid, lysine) to 1 (most hydrophobic amino acid, phenylalanine).

Fig. 1 explains Methodology 1. A PDB file ("The Protein Data Bank", Berman et al., 2000) is used as input to the program GRASP ("Graphical Representation and Analysis of Surface Properties", Nicholls et al., 1991), the total (s_p) and partial (s_{aai}) solvent accessible areas of the exposed amino acids are determined. Then, using the normalized amino acid hydrophobicity scale (Miyazawa and Jernigan, 1985) and Eq. (2), the average surface hydrophobicity of the protein is estimated. Finally, a simple quadratic equation is used to

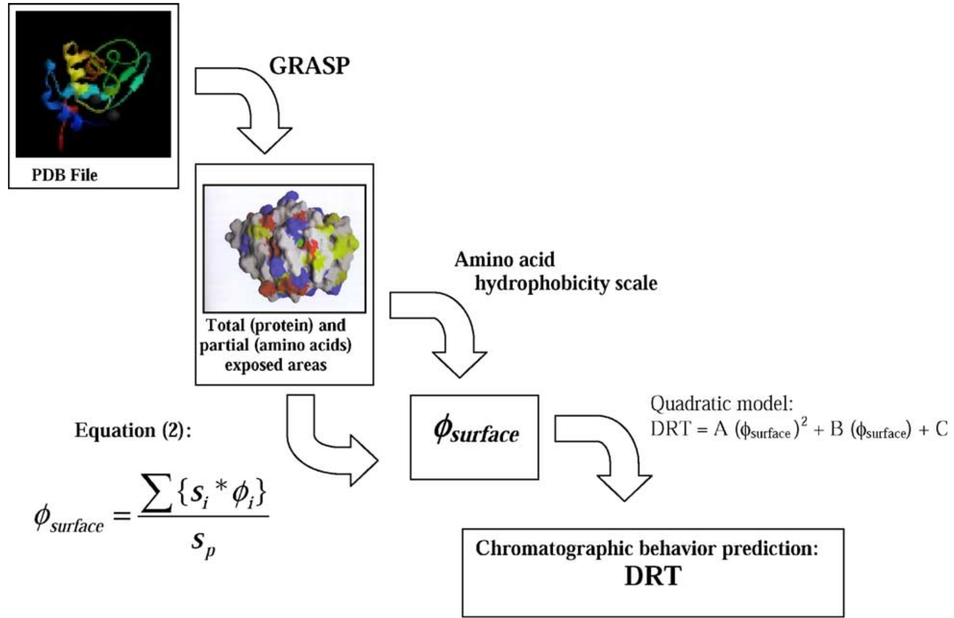

Fig. 1. Schematic representation of the methodology based on average surface hydrophobicity (ϕ_{surface}) (Methodology 1).

calculate the dimensionless retention time (DRT) of the protein. The quadratic models obtained for different chromatographic conditions are shown in Table 1 (Lienqueo et al., 2002; Mahn, 2004). The coefficients A, B and C were determined empirically through a simple linear regression analysis.

Table 1
Coefficients of the quadratic model used in Methodology 1 to estimate protein retention time (as DRT) under different conditions

Resin	Salt type and concentration[a]	Model coefficients[b]		
		A	B	C
Phenyl sepharose[a]	Ammonium sulfate (1 M)	11.79	−0.29	−0.35
	Ammonium sulfate (2 M)	−12.14	12.7	−1.14
	Sodium chloride (2 M)	−77.10	42.33	−5.13
	Sodium chloride (4 M)	−65.01	37.55	−4.71
Butyl sepharose[b]	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

[a] The chromatographic conditions used were: phenyl or butyl sepharose™ resin and ammonium sulfate or sodium chloride at specified initial concentrations to build the elution gradient. The final salt concentration was always zero.

[b] The general form of the quadratic model is $\text{DRT}=A\phi^2_{\text{surface}}+B_{\text{surface}}+C$. This equation applies to proteins with $0.185\leq\phi_{\text{surface}}\leq0.345$, as determined by the amino acid hydrophobicity scale. If $\phi_{\text{surface}}<0.185$, the protein is not retained by the resin. If $\phi_{\text{surface}}>0.345$, the protein remains attached to the resin.

This methodology has been validated for monomeric and multimeric proteins, with predictions resulting in low deviations (Lienqueo et al., 2002). In addition, it has been used reasonably successfully, to predict protein separation behavior from a “real” cell extract (Lienqueo et al., 2003). However, this methodology is not valid for proteins that have highly heterogeneous distributions of hydrophobic patches on their surfaces (Mahn et al., 2004). The methodology is useful for predicting retention times of stable proteins that have a relatively homogeneous distribution of surface hydrophobicity.

3. Methodology 2: hydrophobic contact area (HCA)

Some protein’s behavior during HIC with gradient elution cannot be explained by the methodology based on average surface hydrophobicity. For example, some proteins have very similar ϕ_{surface} values, but rather different retention times. This discrepancy has been attributed to differences in the surface distribution of the hydrophobic patches (Mahn, 2004).

The surface hydrophobicity distribution of proteins related to retention in HIC has been investigated by Mahn et al. (2004). Based on a classical thermodynamic model (Melander et al., 1989) that describes protein retention in terms of electrostatic and hydrophobic interactions, the contact area between the hydrophobic ligands of the HIC matrix and the protein when adsorbed to the resin (hydrophobic contact area, HCA) was experimentally determined. This variable was found to be linked to the surface hydrophobicity distribution of proteins. If the HCA of a protein is much lower than the total hydrophobic accessible area HAA (i.e. the solvent accessible area of a protein occupied only by hydrophobic residues), the interaction zone between the protein and the HIC resin would be a small fraction of the hydrophobic zones on the protein’s surface. Under these circumstances, it can be assumed that the protein has a relatively homogeneous surface hydrophobicity distribution (many hydrophobic patches, each one covering a small area). On the other hand, if the hydrophobic contact area and the hydrophobic accessible area of a protein have similar values, the protein is expected to have a heterogeneous surface hydrophobicity distribution. This is because the exposed hydrophobic residues would be clustered in only one extensive hydrophobic patch that is accessible to the hydrophobic ligand of the resin.

An alternative to Methodology 1 is the methodology based on hydrophobic contact area (HCA). This method, or Methodology 2, consists of estimating the HCA through empirical determinations and using this to estimate the protein’s retention time (as DRT). A linear equation is found to correlate well the HCA and DRT values for different proteins under certain chromatographic conditions (Mahn et al., 2004).

Estimation of HCA starts with the following equation (Melander et al., 1989):

$$\log k' = A + Cm_{\text{s}} \tag{3}$$

where k' is the isocratic retention factor and m_{s} is the molal salt concentration. Eq. (3) applies only when a high salt concentration is used and retention due to electrostatic interactions can be neglected. In Eq. (3) the constant A is determined by the system

properties and C is the hydrophobic interaction parameter. The latter is calculated using the following equation:

$$C = \frac{(\mathrm{HCA})\sigma_{\mathrm{s}}}{2.3(RT)}. \tag{4}$$

Here HCA is the hydrophobic contact area between the matrix and the protein when adsorbed to the resin, σ_{s} is a salt property measured as the surface tension increment due to the addition of a neutral salt, R is the universal gas constant, and T is the absolute temperature. The hydrophobic interaction parameter (C) can be obtained from the slope of a plot of log k' versus salt molality.

Determining the HCA value for a single protein requires at least eight HIC runs with isocratic elution, using different salt concentrations in the elution buffer. The isocratic retention factor is determined for each run and then plotted against the salt molality. The slope of the plot provides a value for the parameter C from which the HCA value can be determined. DRT is now predicted using a linear equation that correlates DRT and HCA for different proteins under the same experimental conditions.

The methodology based on HCA has been used successfully to describe protein retention in HIC with salt gradient elution (Mahn et al., 2004). The use of this methodology has allowed demonstrating experimentally that the surface hydrophobicity distribution, not only the average surface hydrophobicity, can be an important factor affecting protein retention in HIC. Methodology 2 can be applied to proteins that present no conformational changes during the HIC process. It is suitable for proteins with a heterogeneous surface hydrophobicity distribution. Methodology 2 has explained chromatographic behavior of proteins that could not be explained with Methodology 1. Nevertheless, the HCA methodology has a major disadvantage as it requires a large number of experiments to determine HCA. To overcome this difficulty, a third methodology has been proposed as discussed next.

4. Methodology 3: local hydrophobicity–molecular docking

This method is based on identifying the zones on a protein's surface that most probably interact with the hydrophobic matrices used in HIC (Mahn et al., 2005). Molecular modeling and computational tools are used for this purpose. Molecular docking simulations are required starting from a knowledge of the three-dimensional structural data. A parameter termed "local hydrophobicity" (LH) has been found to satisfactorily describe the most probable interaction zone between a protein and the HIC resin. LH correlates extremely well with both the hydrophobic accessible area HAA and dimensionless retention time (DRT) of different proteins.

The local hydrophobicity of a protein is given by the following equation:

$$\mathrm{LH} = \frac{\sum s_{\mathrm{aa}i}\phi_{\mathrm{aa}i}}{s_{\mathrm{IZ}}} \tag{5}$$

where LH is the average surface hydrophobicity of the interaction zone of the protein with the hydrophobic ligand (local hydrophobicity); $s_{\mathrm{aa}i}$ is the solvent accessible area of each

residue that belongs to the interaction zone; ϕ_{aai} is the amino acid hydrophobicity given by the normalized scale reported by Miyazawa and Jernigan (1985); and s_{IZ} is the solvent accessible area of the whole interaction zone.

Molecular docking simulations involve a conformational sampling procedure in which different protein–ligand conformations are examined to find the ones with the correct fit. The sampling procedure is normally based on methods such as genetic algorithms and Monte Carlo simulation. In addition, conformational sampling involves an energy function used to evaluate the fitness between the protein and the ligand (Wang et al., 2003). Molecular docking has three steps: identification of the binding sites, a search algorithm to efficiently perform the conformational sampling in the search space, and a score function (Mac Conkey et al., 2002).

Mahn et al. (2005) simulated the interactions between different ribonucleases of known three-dimensional structure and the hydrophobic ligand used in the phenyl sepharose™ resin. AutoDock™ 3.0.5 (Morris et al., 1998) was the software used in these simulations. Eight simulations were carried out for each protein. Each simulation consisted of ten grids to obtain eighty possible conformations of the protein–ligand complex for each protein. Based on qualitative (i.e. location of the interaction zone) and quantitative (i.e. free energy of the complex) considerations, the most probable protein–ligand conformations were selected. Once the interaction zone had been identified, the local hydrophobicity (LH) was determined, considering that the amino acid residues that belonged to that zone and their exposure levels (Lienqueo et al., 2002).

The specific steps of Methodology 3 are shown in Fig. 2. First, molecular docking simulations are carried out to identify different protein–ligand conformations. Molecular docking simulations give a high number of possible protein–ligand complex conformations and each of these has an associated free energy value. To establish the conformation that is closest to the real one (i.e. that occurs during HIC), it was necessary to develop a selection procedure. Thus, first, the protein–ligand conformations were ordered in a descending series based on their calculated free energy (i.e. configuration with the least negative value occupied the top position). The conformation with the highest free energy was chosen and analyzed qualitatively based on the position of the ligand on the protein's surface. If the ligand was set in a concave zone, that conformation was discarded and the conformation with the next free energy value was analyzed until a conformation in which the hydrophobic ligand was located in a convex zone on the protein surface, preferably opposite to the active site, was found. This selection criterion is based on the knowledge that hydrophobic patches located in convex zones of a protein are more accessible to the hydrophobic ligands of HIC resins (Fausnaugh and Regnier, 1986; Mahn et al., 2004). Once the most probable complex conformation has been chosen, the interaction zone is identified as the amino acid residues that are found within a radius of 5 Å from the center of the identified zone. Then, the solvent accessible area of the interaction zone (s_{IZ}) and the partial solvent accessible areas (s_{aai}) that correspond to each residue in that zone are determined. Finally, Eq. (5) is used to calculate the local hydrophobicity of the protein. Normalized Miyazawa and Jernigan (1985) amino acid hydrophobicity scale is used.

This methodology is able to explain the chromatographic behavior in HIC of proteins that have a heterogeneous surface hydrophobicity distribution. However, it does not give

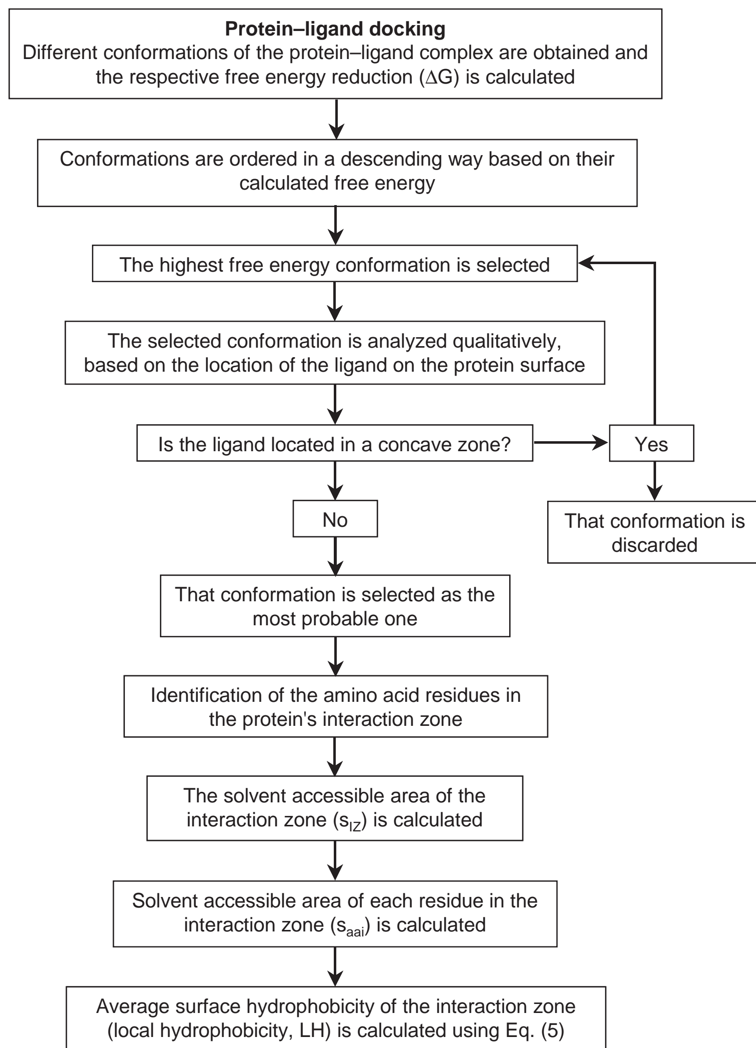

Fig. 2. Flowchart for the methodology based on local hydrophobicity (LH) (Methodology 3). Adapted from Mahn et al. (2005).

results as good as expected when applied to proteins with a homogeneous distribution of the hydrophobic patches, as the probable interaction zone in these proteins is difficult to identify.

The local hydrophobicity methodology is a first approach for estimating protein retention time using molecular docking tools and knowledge of the crystal structure of proteins. This methodology reduces experimental development effort considerably, but has been tested for only one chromatographic condition (phenyl sepharose resin and 2–0 M ammonium sulfate elution gradient) with a few proteins. This method requires further experimental validation.

5. Concluding remarks

The separation behavior of proteins during HIC is mainly determined by the average surface hydrophobicity of a protein (ϕ_{surface}) and by the distribution of the hydrophobic zones on a protein's surface. A protein's retention time in HIC can be predicted starting from its tertiary structure. This review presented three methodologies for predicting protein retention times. These methodologies have a significant potential for contributing to rational design of protein purification processes that involve a HIC step. Some limitations of the methods discussed include the need to know the three-dimensional structure of the protein and establish the empirical correlations between DRT and the hydrophobicity parameters under selected chromatographic conditions.

References

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res 2000;28:235–42.

Eriksson K. Hydrophobic interaction chromatography. In: Janson J-C, Ryden L, editors. Protein purification: principles, high-resolution methods, and applications. New York: Wiley-Liss; 1998. p. 283–309.

Fausnaugh JL, Regnier FE. Solute and mobile phase contributions to retention in hydrophobic interaction chromatography of proteins. J Chromatogr 1986;359:131–46.

Fausnaugh JL, Kennedy LA, Regnier FE. Comparison of hydrophobic interaction and reversed phase chromatography of proteins. J Chromatogr 1984;317:141–55.

Fisher HF. A limiting law relating the size and shape of protein molecules to their composition. Proc Natl Acad Sci U S A 1964;51:1285–91.

Jennissen HP. Hydrophobic interaction chromatography. Int J Bio-Chromatogr 2000;5:131–8.

Lienqueo ME, Mahn AV, Asenjo JA. Mathematical correlations for predicting protein retention time in hydrophobic interaction chromatography. J Chromatogr, A 2002;978:71–9.

Lienqueo ME, Mahn A, Vásquez L, Asenjo JA. Methodology for predicting the separation of proteins by hydrophobic interaction chromatography and its application to a cell extract. J Chromatogr, A 2003; 1009:189–96.

Mac Conkey B, Sobolev V, Edelman M. The performance of current methods in ligand–protein docking. Curr Sci 2002;7(83):845–56.

Mahn, A. Study of the main factors affecting protein behavior in hydrophobic interaction chromatography. PhD thesis, University of Chile, Santiago, Chile; 2004. p. 256.

Mahn A, Lienqueo ME, Asenjo JA. Effect of surface hydrophobicity distribution on protein retention in hydrophobic interaction chromatography. J Chromatogr, A 2004;1043:47–55.

Mahn A, Zapata-Torres G, Asenjo JA. A theory of protein–resin interaction in hydrophobic interaction chromatography. J Chromatogr, A 2005;1066:81–8.

Melander W, Horvath CS. Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series. Arch Biochem Biophys 1977;183:200–15.

Melander W, El Rassi Z, Horvath CS. Interplay of hydrophobic and electrostatic interactions in biopolymer chromatography: effect of salts on the retention of proteins. J Chromatogr 1989;469:3–27.

Miyazawa S, Jernigan R. Estimation of effective inter residue contact energies from protein crystal structures Quasi-chemical approximation. Macromolecules 1985;18:534–52.

Morris G, Goodsell D, Halliday R, Huey R, Hart W, Belew R, et al. Automated docking using a Lamarckian Genetic Algorithm and an empirical binding free energy function. J Comp Chem 1998;19:1639–62.

Nicholls A, Sharp K, Honing B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Struct Funct Gen 1991;11:281–96.

Sofer G, Hagel L, 1998. Handbook of process chromatography: a guide to optimization, scale-up, and validation. San Diego: Academic Press; 1998. p. 387.

Tanford C. Contribution of hydrophobic interactions to the stability of globular conformation of proteins. J Am Chem Soc 1962;84:4240–7.
Wang R, Lu Y, Wang S. Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 2003;46:2287–303.
Waugh DF. Protein–protein interactions. Adv Protein Chem 1954;9:325–437.