Journal of Chromatography A, 1411 (2015) 50–62

Contents lists available at ScienceDirect

Journal of Chromatography A

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

A two level hierarchical model of protein retention in ion exchange chromatography

Matteo Salvalaglio[a,1], Matteo Paloni[a], Bertrand Guelat[b,c], Massimo Morbidelli[b], Carlo Cavallotti[a,*]

[a] Department Chimica, Materiali e Ingegneria Chimica, via Mancinelli 7 20131 Milano, Italy
[b] Institute for Chemical & Bioengineering, ETH Zurich, 8093 Zurich, Switzerland
[c] Novartis Pharma AG, Postfach, 4002 Basel, Switzerland

ARTICLE INFO

Article history:
Received 21 May 2015
Received in revised form 23 July 2015
Accepted 27 July 2015
Available online 29 July 2015

Keywords:
Ion exchange chromatography
Molecular dynamics
Poisson–Boltzmann equation
Adsorption free energy
Protein retention factor

ABSTRACT

Predicting protein retention in ion exchange chromatography (IEX) from first principles is a fascinating perspective. In this work a two level hierarchical modeling strategy is proposed in order to calculate protein retention factors. Model predictions are tested against experimental data measured for Lysozyme and Chymotrypsinogen A in IEX columns as a function of ionic strength and pH. At the highest level of accuracy Molecular Dynamics (MD) simulations in explicit water are used to determine the interaction free energy between each of the two proteins and the IEX stationary phase for a reference pH and ionic strength. At a lower level of accuracy a linear response model based on an implicit treatment of solvation and adopting a static protein structure is used to calculate interaction free energies for the full range of pHs and ionic strengths considered. A scaling coefficient, determined comparing MD and implicit solvent simulations, is then introduced in order to correct the linear response model for errors induced by the adoption of a static protein structure. The calculated free energies are then used to compute protein retention factors, which can be directly compared with experimental data. The possibility to introduce a third level of accuracy is explored testing the predictions of a semiempirical model. A quantitative agreement between the predicted and measured protein retention factors is obtained using the coupled MD-linear response models, supporting the reliability of the proposed approach. The model allows quantifying the electrostatic, van der Waals, and conformational contributions to the interaction free energies. A good agreement between experiments and model is obtained also using the semiempirical model that, although requiring parameterization over higher level models or experimental data, proves to be useful in order to rapidly determine protein retention factors across wide pH and ionic strength ranges as it is computationally inexpensive.

1. Introduction

Although the purification of therapeutic proteins is usually performed using at least an affinity chromatography step [1–5], it has been proposed that the level of purity required for pharmaceutical applications may be obtained also with non-affinity techniques such as Ion Exchange (IEX) Chromatography, which exhibits a large optimization margin compared to conventional processes [6–9]. Even if IEX technology is often used as a complement to traditional affinity purification steps, processes entirely based on IEX principles are rapidly becoming commercially available [6,10,11]. The rational optimization of IEX chromatography can significantly benefit from the comprehension of the fundamental phenomena occurring at the solid–liquid interface during protein adsorption [12–15]. To fulfill this goal, the analytical and numerical solutions of the Poisson–Boltzmann equation (PBE) have been often used to determine the energy of interaction of proteins with a charged surface. The use of PBE based methods for modeling interactions in IEX chromatography has been pioneered mainly by Stahlberg and co-workers [6,16–19] and has led to the development of analytical methods for the description of electrostatics-based retention [20,21].

The rapidly increasing availability of computational resources has recently attracted much interest towards the possibility of

* Corresponding author at: Dipartimento di Chimica, Materiali e Ingegneria Chimica "Giulio Natta", Politecnico di Milano, via Mancinelli 7, 20131 Milano, Italy. Tel.: +39 02 23993176; fax: +39 02 23993180.
E-mail address: carlo.cavallotti@polimi.it (C. Cavallotti).
[1] Present address: Department of Chemical Engineering, University College London, London WC1E 7JE, UK.

http://dx.doi.org/10.1016/j.chroma.2015.07.101
0021-9673/

exploiting molecular modeling techniques to investigate complex systems at the molecular scale. Molecular dynamics simulations are in fact gradually becoming part of the arsenal of techniques applicable to address fundamental problems in protein purification [22–34]. The investigation of IEX adsorption processes at the atomistic scale has been performed with different approaches in the literature. Riccardi and Zhang focused their research on both the atomistic coarse grain simulations of charged proteins adsorption on IEX surfaces and their lateral diffusion [35] and on the molecular level modeling of the porous structure of the chromatographic supports [36–38]. Chung [39,40] and co-workers have proposed a structure-activity relationship study based on the calculation of the Poisson–Boltzmann interaction energy of a charged ligand with several mutants of the Cold Shock protein. The PBE was used to determine the influence of structural mutations on the most probable protein adsorption orientations. Similarly, Kubiak and Mulheran have investigated the dynamics of Lysozyme adsorbed on a charged surface by MD simulations to get a molecular insight on the most favorable binding orientations, finding that no tertiary structure depletion occurs when the protein adsorption takes place on a flat surface [41]. Dismer and Hubbuch have used a MD simulation protocol in order to determine the orientation of Lysozyme when adsorbed on a IEX surface, which allowed to determine an empirical correlation between the electrostatic component of the force field and the measured chromatographic retention parameters [42]. It was also found that, within typical simulation timescales, the protein does not macroscopically move its center of mass nor reorient itself with respect to the charged surface.

In principle, considering the systematic increase of computational resources, it may be conceivable to expect that within few years it will be possible to simulate protein adsorption on IEX surfaces using MD simulations in explicit solvent models. However this may not be so simple. A fully atomistic MD simulation in fact requires quite long simulation times in order to sample adequately the conformational space of the protein. In addition, using classical MD techniques, protonation states of all aminoacidic residues must be specified a priori, which may be an issue for pHs at which residues have similar probabilities of existing in the protonated and non-protonated states. Addressing properly this issue may require performing multiple MD simulations for the same system, which would increase considerably computational costs. In the present work we propose an alternative approach that combines the accuracy attainable through explicit MD simulations with the computational efficiency of models describing implicitly through the PBE the effect of the environment and, eventually, with semiempirical models. This paper is structured as follows: in Section 2 the hierarchical modeling strategy and experimental details are described in detail, while in Section 3 results are reported and discussed.

2. Method and theoretical background

The methodology here adopted to estimate protein retention coefficients is based on the calculation of free energies of interactions between the protein and a model surface. This is a problem well known to the scientific literature where many computational approaches, differing for level of complexity and accuracy, have been reported [12,19,21,24,41–45]. In the present work we investigate whether it is possible to determine the free energy of interaction of a protein with a surface using a computational approach formulated so that it retains a full molecular description of the system but it is also sufficiently simple to be applied to large and complex proteins on a routinely basis. In the following we describe the model, the adopted computational approaches, and report details of the experiments used for validation.

2.1. Free energy calculations

Two different synergic atomistic computational approaches were used to estimate the free energy of interaction of the proteins with the IEX surface, the first exploiting Umbrella Sampling (US) simulations and the second belonging to the family of the linear response approximation approaches. The two approaches differ significantly in their respective level of accuracy and computational cost. In addition, an analytical model is also used in order to test its capability to predict protein retention factors and to evaluate whether it may be used to complement and ease the computational cost of the other two models. The details of the semiempirical model are described by Guelat et al. [21], while the implementation of the MD and linear response model is described in the following two sections.

2.1.1. MD simulations

The first methodology used to determine interaction free energies between proteins and the IEX surface exploits free energy perturbation (FEP) theory. FEP was carried out through an US calculation protocol, in which the weighted histogram analysis method (WHAM) was applied to extrapolate the potential of mean force (PMF). Molecular dynamics simulations were performed in explicit water in the NPT ensemble. All simulations were performed with periodic boundary conditions and long-range interactions were computed with the Particle Mesh Ewald (PME) method. The simulation box for Lysozyme has dimensions of 145 Å × 145 Å × 100 Å and contains the IEX surface, the protein, 67,679 water molecules, and 450 Na^+ and Cl^- ions for a total of 20,8450 atoms. 162 Na^+ atoms have been added to obtain global electroneutrality. The surface was modeled by an ensemble of 169 propyl sulfonate groups organized in 13 rows of 13 elements equally spaced to give a global surface charge density of 1.57×10^{-6} mol m^{-2}. The simulations for Chymotrypsinogen A were performed using a similar system setup and parameters. The force field parameters of the IEX ligands were determined using the GAFF amber force field [45] with structure and charges calculated at the B3LYP/6-31+G(d,p) level and fitted using the RESP model. The 2VB1 (Lysozyme) [46] and 2CGA (Chymotrypsinogen A) [47] protein crystal structures were used for all the calculations. The AMBER ff03 force field [48] was used to model Lysozyme and Chymotrypsinogen A, while the TIP3P force field was used for water [49]. Throughout the simulations the IEX ligands were restrained to their position by applying a harmonic restraint of 1000 kcal mol^{-1} $Å^{-2}$ on the first and second C atoms of the propyl chain. A double restraint was applied to each ligand in order to limit its tendency to flip upside down in the course of the simulations. The cut-off for the non-bonded interactions was set to 10 Å.

The computational protocol adopted for the MD simulations was the following: first, a 2000-cycle energy minimization was carried out in order to remove bad contacts between the solute and the random-placed solvent molecules. In this step, solute molecules were restrained with a harmonic potential $k(\Delta x)^2$ with k set to 500 kcal $mol^{-1}Å^{-2}$. A second 3500-cycle energy minimization was performed for the entire system, removing the previous restraints. Then, the temperature was raised from 0 to 300 K by a simulated annealing of 20 ps at constant volume; a weak harmonic restraint (where k is now equal to 10 kcal mol^{-1} $Å^{-2}$) was applied on the solute with the purpose of avoiding wild fluctuations. This was followed by a 100 ps run at constant pressure, in order to allow the system relaxation and thus to reach the correct density. The system was finally relaxed through a 2 ns simulation performed at constant temperature and pressure. Temperature and pressure control was performed through Langevin dynamics with a collision frequency of 1 ps^{-1} and isotropic position scaling, respectively. The SHAKE algorithm was used for covalent bonds involving hydrogen atoms, allowing to adopt a time step of 2 fs. All MD simulations