Chapter 1

Introduction

Amino acids are the building blocks for all life on earth. Albeit the almost infinite variety of species the number of amino acids forming proteins is limited to about 20. The majority of them are rather simple organic compounds with an amino group and an aliphatic, aromatic, or heterocyclic side chain (R) attached α to the carboxylic group. With the exception of the simplest amino acid glycine, all other amino acids are chiral compounds. In nature the L-form prevails. However, D-alanine, for example, is found in the cell membrane of bacteria.

Amino acids and their derivatives are of particular economical importance since they are utilized on a large scale as nutrition factors and flavor enhancers (e. g. glutamate). Additionally, many pharmaceuticals such as antibiotics, heart, cancer, and anti-HIV drugs are produced starting from amino acids. An excellent overview about production methods and the application of several amino acids is given by Kleemann et al.[1]. Most production processes are based on white biotechnology – in other words, on fermentation and enzymatic technologies. The purification of the target product is achieved by crystallization from the filtered fermentation broth often containing also electrolytes and other by-products. For the design of separation units knowledge of thermodynamic data is indispensable. Therefore, since the beginning of the past century phase behavior in amino acid solutions has attracted the interest of many research groups especially with regard to its modeling. Activity coefficients as well as solubilities in aqueous amino acid solutions also containing electrolytes have been measured and described with models of varying complexity.

In addition to amino acids, peptides and proteins are economically highly valuable bioproducts. Precipitation by addition of salts, non-ionic polymers, polyelectrolytes, or organic solvents is commonly used to isolate target proteins from solutions. Often two metastable liquid phases – one protein-rich and one protein-lean – form instead of the thermodynamically stable solid precipitate. The phase behavior is very complex and strongly depending on solution conditions such as ionic strength, salt type and pH. Although still far away from a rigorous quantitative modeling of the thermodynamics of complex protein solutions, qualitative agreement between model and experiment can already be achieved.

1.1 Aim of the Thesis

The scope of this thesis is twofold: Because solutions in biotechnology often contain salts, a model that can deal with charged species is essential. The Perturbed-Chain Statistical Association Theory (PC-SAFT) developed by Groß et al.[2] is extended by an electrolyte term in order to describe solution densities, vapor pressure depression, and mean ionic activity coefficients of aqueous electrolyte solutions. In a second step, phase behavior of binary water-amino acid (peptide) mixtures is calculated with the same model. Herewith, the foundation is laid for modeling the influence of electrolyte type and concentration on activity coefficient and solubility of amino acids and peptides in aqueous solutions.
Another aim is to model phase behavior in protein systems with an equation of state. The applicability of a rigorous approach such as PC-SAFT is investigated. Further, a model based on the McMillan-Mayer framework is applied to describe second osmotic virial coefficients of hen egg-white lysozyme in aqueous salt solutions as well as the meta-stable liquid-liquid demixing and the solubility.

1.2 Structure of the Thesis

After this introduction Chapter 2 provides the thermodynamic basics of phase equilibria calculations. Besides some fundamentals, Chapter 3 contains a detailed description of how equations of state, especially the PC-SAFT equation of state, are developed from statistical mechanics and perturbation theory. The concept of the radial distribution function and its relation to other thermodynamic functions is elucidated. Chapter 4 is dedicated to the potential of mean force ansatz and its theoretical application to colloidal dispersions. The extension of the PC-SAFT model to electrolyte systems follows in Chapter 5. Further, the method of modeling amino acid (peptide) solutions with electrolyte PC-SAFT (ePC-SAFT) is explained. Modeling results for electrolyte solutions, amino acid (peptide) solutions, and electrolyte/amino acid solutions are presented in Chapter 6 whereas Chapter 7 summarises the results for the protein systems. A summary of the thesis, conclusions, and an outlook for future work follow in Chapter 8.
The Appendix mainly comprises component model parameters and supplementary result figures.

1.3 Investigated Systems

The electrolyte PC-SAFT equation of state is used to model solution densities, vapor pressures, (mean ionic) activity coefficients and solubilities of following aqueous electrolyte, amino acid, and peptide solutions:

- salts containing

  cations: Li^+, Na^+, K^+, NH_4^+, Mg^{2+}, Ca^{2+}

  anions : F^-, Cl^-, Br^-, I^-, OH^-, NO_3^-, SO_4^{2-}

- glycine, DL-alanine, DL-serine, L-serine, L-proline, L-valine, L-arginine, L-lysine, DL-threonine, L-threonine, L-histidine
- α-ABA, β-ABA, γ-ABA*
- α-AVA, γ-AVA†
- diglycine, triglycine, dialanine, alanylglycine, glycylalanine

As far as experimental data was available solubilities of amino acids and peptides were modeled.
The influence of amino acids on the mean ionic activity coefficient of salts was investigated for 17 electrolyte/amino acid systems (see Appendix H, pp. 130).
As an example of a protein system hen egg-white lysozyme/NaCl solutions were chosen.

*ABA: aminobutyric acid
†AVA: aminovaleric acid

Chapter 2

Thermodynamical Background

2.1 Basics

A mixture of chemical compounds can distribute to multiple phases (solid, liquid, gas). There will be an interchange of components between the coexisting phases until equilibrium is reached and all intensive properties do not change anymore. Intensive properties are those that do not depend on the size, mass, or shape of the phase, such as temperature, pressure, density, and composition. The equilibrium thermodynamics provides an abstract mathematical framework which quantitatively relates the variables describing the state of the system.
A system is completely characterized by the following function which contains the information of the first and second law of thermodynamics:

$$U = U(S, V, \mathbf{n}) \tag{2.1}$$

Eq. 2.1 is called thermodynamic potential or fundamental function of the internal energy with the entropy S and the volume V as the respective fundamental variables. The total differential of this function is

$$dU = \underbrace{\left(\frac{\partial U}{\partial S}\right)_{V,\mathbf{n}}}_{T} dS + \underbrace{\left(\frac{\partial U}{\partial V}\right)_{S,\mathbf{n}}}_{-P} dV + \sum_{i=1}^{N} \underbrace{\left(\frac{\partial U}{\partial n_i}\right)_{S,V,n_{j\neq i}}}_{\mu_i} dn_i \tag{2.2}$$

Here, the dependence of the internal energy on the amount of each component i is given by the partial differential quotient $\left(\frac{\partial U}{\partial n_i}\right)_{S,V,n_{j\neq i}}$ and is called chemical potential μ_i.
Equilibrium is reached when the entropy is maximized or in other words the internal energy is at its minimum, i. e.

$$dS = 0 \quad , \quad d^2S < 0 \tag{2.3}$$

$$dU = 0 \quad , \quad d^2U > 0 \tag{2.4}$$

For a system consisting of C components distributed in π phases this can be translated

into the three well-known phase equilibrium conditions:

$$T^{(1)} = T^{(2)} = ... = T^{(\pi)} = T \tag{2.5}$$

$$P^{(1)} = P^{(2)} = ... = P^{(\pi)} = P \tag{2.6}$$

$$\mu_i^{(1)}(T, P, \mathbf{n}^{(1)}) = \mu_i^{(2)}(T, P, \mathbf{n}^{(2)}) = ... = \mu_i^{(\pi)}(T, P, \mathbf{n}^{(\pi)}) \ \ \forall\ i \in C \tag{2.7}$$

Applying Legendre transformation three other fundamental equations can be derived from Eq. 2.2:

$$dH = d(U + PV) = TdS + VdP + \sum_{i=1}^{N} \mu_i dn_i \tag{2.8}$$

$$dG = d(H - TS) = -SdT + VdP + \sum_{i=1}^{N} \mu_i dn_i \tag{2.9}$$

$$dA = d(U - TS) = -SdT - PdV + \sum_{i=1}^{N} \mu_i dn_i \tag{2.10}$$

The fundamental variables for the enthalpy H are S and P. The Gibbs (free) energy (or free enthalpy) plays an important role in equilibrium thermodynamics since its fundamental variables T and P are easily accessible experimentally in contrast to the entropy, for example. Also the Helmholtz (free) energy A, with its fundamental variables T and V, is of major interest because many thermodynamic models (especially those derived from statistical mechanics) are written in terms of the Helmholtz energy.
A comparison of the coefficients of the total differential equations (Eqs. 2.2, 2.9-2.10) and their respective fundamental equations (not shown here) yields following differential quotients:

$$\left(\frac{\partial U}{\partial S}\right)_{V,n_i} = \left(\frac{\partial H}{\partial S}\right)_{P,n_i} = T \tag{2.11}$$

$$\left(\frac{\partial U}{\partial V}\right)_{S,n_i} = \left(\frac{\partial A}{\partial V}\right)_{T,n_i} = -P \tag{2.12}$$

$$\left(\frac{\partial A}{\partial T}\right)_{V,n_i} = \left(\frac{\partial G}{\partial T}\right)_{P,n_i} = -S \tag{2.13}$$

$$\left(\frac{\partial H}{\partial P}\right)_{S,n_i} = \left(\frac{\partial G}{\partial P}\right)_{T,n_i} = V \tag{2.14}$$

and

$$\left(\frac{\partial U}{\partial n_i}\right)_{S,V,n_{j\neq i}} = \left(\frac{\partial H}{\partial n_i}\right)_{S,P,n_{j\neq i}} = \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_{j\neq i}} = \left(\frac{\partial A}{\partial n_i}\right)_{T,V,n_{j\neq i}} = \mu_i \tag{2.15}$$

The equations shown above prove the fact that the knowledge of one thermodynamic potential suffices to derive all other properties of a system. Hence, without loss of generality we confine ourselves to dealing only with the Helmholtz energy A and its derivatives throughout this work.

2.2 Calculation of Phase Equilibria

2.2.1 Vapor-Liquid and Liquid-Liquid Equilibria

The equality of the chemical potential of each component in every phase (Eq. 2.7) can be reformulated as the isofugacity criterium

$$f_i^{(1)} = f_i^{(2)} = ... = f_i^{(\pi)} \quad \forall\ i \in C \tag{2.16}$$

Within the scope of this work the number of phases is restricted to two. There are three concepts for the calculation of phase equilibria: $\varphi-\varphi$, $\gamma-\varphi$, and $\gamma-\gamma$. These are explained below.

$\varphi - \varphi$ Concept

Using the definition of the fugacity coefficient φ_i

$$\varphi_i \equiv \frac{f_i}{x_i P} \tag{2.17}$$

with x_i being the mole fraction of component i one obtains

$$(x_i\varphi_i)^{(1)} = (x_i\varphi_i)^{(2)} \quad \forall\ i \in C \tag{2.18}$$

The fugacity coefficients are complex functions of system temperature, volume or density, and composition: $\varphi_i = \varphi_i(T, v, \mathbf{x})$. Therefore, the compositions in each phase must be calculated iteratively. As an example, the iteration algorithm for an isobaric-isothermal flash VLE or LLE calculation is illustrated in Fig. 2.1. However, this concept is generally applicable (also in the critical phase region where liquid and gas phase have similar properties) as long as there are equations of state providing the fugacity coefficients for each phase.

$\gamma - \varphi$ Concept

When calculating VLE the fugacity of the liquid phase can also be described by an activity coefficient γ_i while the behavior of the vapor phase is still captured by the fugacity coefficient:

$$f_i^L = x_i^L \gamma_i P_{0i}^{LV} \varphi_{0i}^{LV} Poy = x_i^V \varphi_i^V P = f_i^V \tag{2.19}$$

P_{0i}^{LV} and φ_{0i}^{LV} are the vapor pressure and the fugacity coefficient of pure component i, respectively. The Poynting factor Poy captures the pressure dependence of the fugacity coefficient and is negligible (≈ 1) for pressures below 10 bar. Further, the ratio $\varphi_{0i}^{LV}/\varphi_i^V \approx 1$ at low pressures. Hence, Eq. 2.19 simplifies to

$$x_i^L \gamma_i P_{0i}^{LV} = x_i^V P \tag{2.20}$$

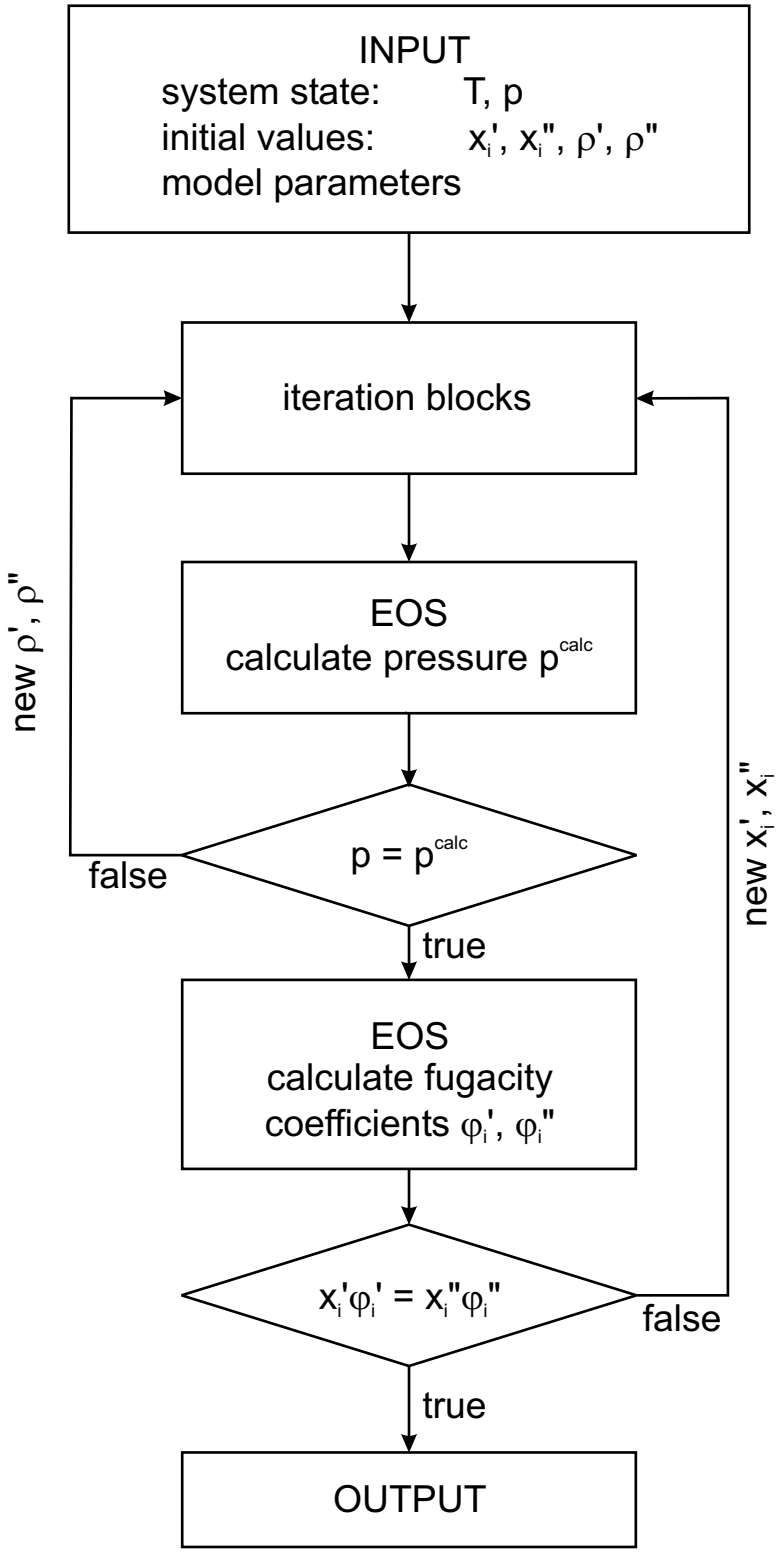

Fig. 2.1: Flowchart of an isobaric-isothermal VLE or LLE flash calculation for two phases ' and ".

and at low concentration, i. e. in the limit of an ideal solution where $\gamma_i \approx 1$, to the well-known Raoult's law

$$x_i^L P_{0i}^{LV} = x_i^V P \tag{2.21}$$

The activity coefficient is obtained from g^E-models* or can as well be calculated with the help of an equation of state by the following definition

$$\gamma_i \equiv \frac{\varphi_i}{\varphi_{0i}} \tag{2.22}$$

One advantage of the $\gamma - \varphi$ concept is that for isothermal calculations and when the vapor pressure is given as a function of temperature (e. g. Antoine equation) the phase compositions are numerically easy to obtain. That is not the case for isobaric calculations. Phase equilibrium calculations at higher pressures with the $\gamma - \varphi$ concept are seldom performed because an equation of state is needed to evaluate the Poynting factor and the fugacities for Eq. 2.19. Hence, in this case it makes more sense to directly use the $\varphi - \varphi$ concept. Another shortcoming of the $\gamma - \varphi$ concept is that – based on different assumptions and simplifications for the liquid and the gas phase – the critical phase region is not well described.

$\gamma - \gamma$ Concept

For the description of LLE both liquid phases can be described with activity coefficients. This method is only applicable for pressure and density independent phase equilibria because g^E models generally are only able to capture the temperature and concentration dependence of the excess free energy. Nevertheless, due to the incompressibility of liquids, many binary mixtures reveal a negligible influence of the pressure on the phase equilibrium. Density effects due to temperature changes are also often neglected assuming similar thermal expansion coefficients for all components.

2.2.2 Solid-Liquid Equilibria

The fugacities of the solid (S) and the liquid (L) phase are formulated using activity coefficients:

$$f_i^S = x_i^S \gamma_i^S f_{0i}^S \tag{2.23}$$
$$f_i^L = x_i^L \gamma_i^L f_{0i}^L \tag{2.24}$$

where f_{0i}^S and f_{0i}^L are the (pure component) standard fugacities of the solid and the liquid, respectively. Rearranging Eqs. 2.23-2.24 one obtains the solubility of component i in the liquid phase

$$x_i^L = \frac{x_i^S \gamma_i^S f_{0i}^S}{\gamma_i^L f_{0i}^L} \tag{2.25}$$

*Remember: $g^E = k_B T \sum_i^C x_i \ln \gamma_i$.

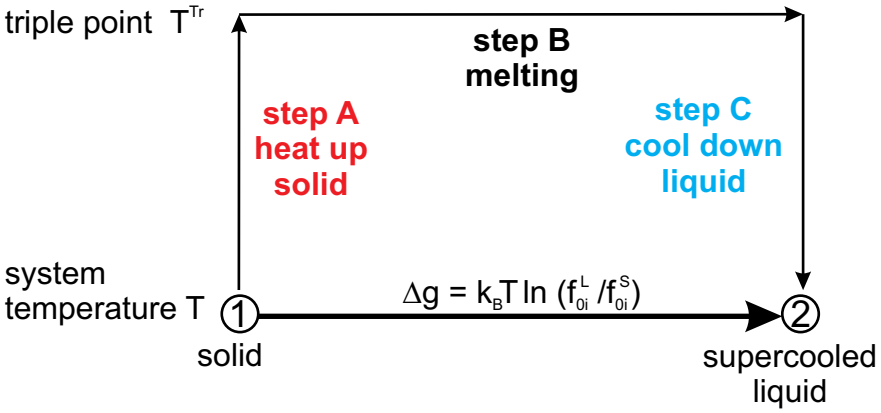

Fig. 2.2: Thermodynamic cycle for the calculation of the ratio f_{0i}^S/f_{0i}^L.

Unfortunately, the standard fugacities are not readily available, especially when the system temperature lies above the triple point of the crystallizing component and below the one of the other component. In this case 'pure solid' and 'pure liquid' represent hypothetical states. However, for the calculation of the solubility only the ratio f_{0i}^S/f_{0i}^L is of importance. One harnesses a thermodynamic cycle as illustrated in Fig. 2.2.

The specific Gibbs enthalpy change for the transition from pure solid ① to supercooled liquid ② at temperature T is given by

$$\Delta g = k_B T \ln \frac{f_{0i}^L}{f_{0i}^S} \tag{2.26}$$

In analogy to Eq. 2.10 in a closed system ($dn = 0$)

$$\Delta g = \Delta h - T\Delta s \tag{2.27}$$

where the change in enthalpy and entropy can be expressed as the sum of the changes for step A, step B, and step C:

$$\begin{aligned} \Delta h &= \int_T^{T_{0i}^{Tr}} c_{p,i}^S \, dT + \Delta h_{0i}^{Tr} + \int_{T_{0i}^{Tr}}^{T} c_{p,i}^L \, dT \\ &= \Delta h_{0i}^{Tr} + \int_{T_{0i}^{Tr}}^{T} \Delta c_{p,i} \, dT \end{aligned} \tag{2.28}$$

$$\begin{aligned} \Delta s &= \int_T^{T_{0i}^{Tr}} \frac{c_{p,i}^S}{T} \, dT + \frac{\Delta h_{0i}^{Tr}}{T_{0i}^{Tr}} + \int_{T_{0i}^{Tr}}^{T} \frac{c_{p,i}^L}{T} \, dT \\ &= \frac{\Delta h_{0i}^{Tr}}{T_{0i}^{Tr}} + \int_{T_{0i}^{Tr}}^{T} \frac{\Delta c_{p,i}}{T} \, dT \end{aligned} \tag{2.29}$$