



**HAL**  
open science

# Temperature, pressure, and electrochemical constraints on protein speciation: Group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins

J. M. Dick, D. E. Larowe, H. C. Helgeson

## ► To cite this version:

J. M. Dick, D. E. Larowe, H. C. Helgeson. Temperature, pressure, and electrochemical constraints on protein speciation: Group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins. *Biogeosciences*, 2006, 3 (3), pp.311-336. hal-00297564

**HAL Id: hal-00297564**

**<https://hal.science/hal-00297564>**

Submitted on 18 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Temperature, pressure, and electrochemical constraints on protein speciation: Group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins

J. M. Dick, D. E. LaRowe, and H. C. Helgeson

Department of Earth and Planetary Science, University of California, Berkeley, CA 94720-4767, USA

Received: 5 September 2005 – Published in Biogeosciences Discuss.: 11 October 2005

Revised: 27 April 2006 – Accepted: 2 June 2006 – Published: 21 July 2006

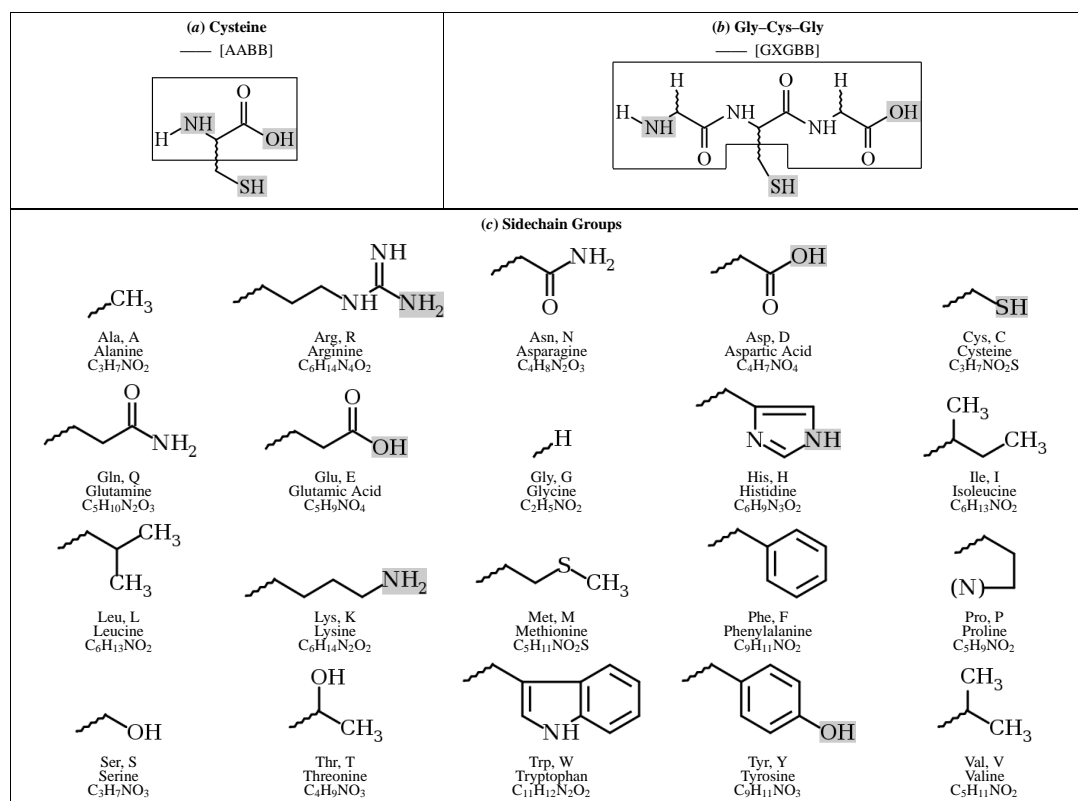
**Abstract.** Thermodynamic calculations can be used to quantify environmental constraints on the speciation of proteins, such as the pH and temperature dependence of ionization state, and the relative chemical stabilities of proteins in different biogeochemical settings. These calculations depend in part on values of the standard molal Gibbs energies of proteins and their ionization reactions as a function of temperature and pressure. Because these values are not generally available, we calculated values of the standard molal thermodynamic properties at 25°C and 1 bar as well as the revised Helgeson-Kirkham-Flowers equations of state parameters of neutral and charged zwitterionic reference model compounds including aqueous amino acids, polypeptides, and unfolded proteins. The experimental calorimetric and volumetric data for these species taken from the literature were combined with group additivity algorithms to calculate the properties and parameters of neutral and ionized sidechain and backbone groups in unfolded proteins. The resulting set of group contributions enables the calculation of the standard molal Gibbs energy, enthalpy, entropy, isobaric heat capacity, volume, and isothermal compressibility of unfolded proteins in a range of proton ionization states to temperatures and pressures exceeding 100°C and 1000 bar. This approach provides a useful frame of reference for thermodynamic studies of protein folding and complexation reactions. It can also be used to assign provisional values of the net charge and Gibbs energy of ionized proteins as a function of temperature and pH. Using these values, an Eh-pH diagram for a reaction representing the speciation of extracellular proteins from *Pyrococcus furiosus* and *Bacillus subtilis* was generated. The predicted predominance limits of these proteins correspond with the different electrochemical conditions of hydrothermal vents and soils. More comprehensive calculations of this kind may reveal pervasive chemical potential constraints on the interactions of microbes with their environment.

Correspondence to: J. M. Dick  
(jedick@berkeley.edu)

## 1 Introduction

The speciation of proteins encompasses ionization reactions associated with complexation with protons and other ions, as well as evolutionary processes that lead to differences in the composition of proteomes in organisms adapted to different environments. Patterns in the compositions of proteomes that support life in different parts of the biosphere have received considerable attention in recent years (Elser et al., 2000; Gasch et al., 2000; Fukuchi and Nishikawa, 2001; Kreil and Ouzounis, 2001; Kato et al., 2004; Tyson et al., 2004; Schulze, 2005; Boonyaratanakornkit et al., 2005), but their connections to gradients of oxidation state and pH both inside and outside the cell (e.g., Schafer and Buettner, 2001; Ding et al., 2001) are only beginning to be recognized (see Brett et al., 2006). A thermodynamic investigation of these processes, which is based on Gibbs energy calculations, affords a quantitative understanding of the speciation of proteins as a function of temperature, pressure, and electrochemical constraints such as Eh and pH.

Generating Eh-pH diagrams depends on calculations of the thermodynamic properties of reactions between proteins, which in turn depend on knowing the ionization states and standard molal Gibbs energies of proteins with different amino acid compositions as a function of temperature, pressure, and ionization state. These properties are generally unavailable in the literature. However, the calculations can be facilitated by adopting a reference state corresponding to aqueous unfolded proteins, the properties of which can be assessed using ionization-specific group additivity algorithms. The group contributions reported in the present study were generated using experimental data reported in the recent literature in conjunction with the revised Helgeson-Kirkham-Flowers (HKF) equation of state (Helgeson et al., 1981; Tanger and Helgeson, 1988). The ionization constants and other thermodynamic properties of sidechain and backbone groups in proteins permit the calculation of the net charge and Gibbs energy ( $Z_{\text{net}}$  and  $\Delta G^\circ$ ) of unfolded proteins, which vary not only as a function of temperature



**Fig. 1.** Representative structures of (a) an amino acid (cysteine), (b) a Gly-X-Gly tripeptide (Gly-Cys-Gly), and (c) the 20 sidechain groups commonly present in proteins. The solid outlines represent either the amino acid or Gly-X-Gly backbone group ([AABB] and [GXGBB], respectively), as indicated. The shaded boxes represent groups that are ionizable in the pH range 0 to 14 at 25°C. The wavy lines denote bonds between sidechain groups and the alpha carbon of the backbone group; the second bond of the proline sidechain group is to the nitrogen of the backbone group. Note that the formulas given in (c) are those of the corresponding amino acid.

and pressure, but also as a function of ionization state. The purposes of this paper are to summarize the results of these group additivity and equations of state calculations and to illustrate their application in calculations of the chemical and biological speciation of proteins.

Group additivity algorithms have been used to supplement experimental thermodynamic data of proteins and other organic species, both at 25°C and 1 bar (Cohn and Edsall, 1943; Cabani et al., 1981; Lepori and Gianni, 2000), and as a function of temperature and pressure (Makhatadze and Privalov, 1990; Privalov and Makhatadze, 1990; Makhatadze et al., 1990; Amend and Helgeson, 1997b; Marriott et al., 1998). The thermodynamic properties of the protein sidechain groups are often modeled from those of amino acids (Amend and Helgeson, 1997a, 2000) or other peptide molecules such as Gly-X-Gly tripeptides (Kharakoz, 1997; Hakin and Hedwig, 2001a; Hedwig and Hinz, 2003). These calculations generally have been restricted to non-ionized (or completely neutral) proteins, or to proteins in discrete ionization states (Amend and Helgeson, 2000). Kharakoz (1997) used group additivity for the volumes and compressibilities of ionized unfolded proteins at 25°C.

The reference model compounds used in these calculations include amino acids and Gly-X-Gly tripeptides. Representative structures of these molecules are shown in Figs. 1a and b. Among the twenty standard protein sidechain groups shown in Fig. 1c, seven are ionizable at pHs of 0 to 14. Because of the availability of experimental data for amino acids, we have used them as the primary model compounds for the sidechain groups. The amino acid backbone group, denoted by [AABB], is a zwitterionic structure that itself contains two ionizable groups. The repeating unit of polypeptides and proteins is denoted by [PBB], or the protein backbone group. Adjacent [PBB] groups in polypeptides and proteins are linked by the peptide bond. Proteins are distinguished from polypeptides by their size, often greater than one hundred backbone groups, and their occurrence in organisms.

Owing to specific interactions among the ionized groups in unfolded proteins (Whitten and García-Moreno E., 2000), the equilibrium constants of the ionization reactions of sidechain and backbone groups in unfolded proteins differ from those of the corresponding amino acid ionization reactions, in some cases by more than a log unit (Nozaki and Tanford, 1967; Steinhardt and Reynolds, 1969; Tan et al., 1995;

Elcock, 1999; Kundrotas and Karshikoff, 2002; Tollinger et al., 2003). However, the enthalpies, heat capacities and volumes of the the different ionization reactions in proteins generally are not experimentally known. Therefore, estimates of the temperature dependence of the ionization equilibrium constants of sidechain groups in proteins can be calculated in a first approximation using the sidechain group contributions derived from amino acids. Refinements to these provisions will become possible when more comprehensive calorimetric data are available for the ionization reactions of proteins.

## 2 Thermodynamic conventions

The revised HKF equations of state permit calculation of the standard molal thermodynamic properties of aqueous species as continuous functions of temperature and pressure (Helgeson et al., 1981; Tanger and Helgeson, 1988). They have been used to represent experimental values of  $C_p^\circ$ ,  $V^\circ$ , and  $\kappa_T^\circ$  of a wide variety of organic and inorganic aqueous species, including amino acids (Amend and Helgeson, 1997a; Marriott et al., 1998). The equations of state and the thermodynamic conventions adopted in the present study are summarized below.

### 2.1 Standard state conventions

The standard state convention adopted for H<sub>2</sub>O is unit activity of the pure component at any temperature and pressure. For other aqueous species, the convention corresponds to unit activity of a hypothetical one molal solution referenced to infinite dilution at any temperature and pressure. The conventional standard molal thermodynamic properties of charged aqueous species are given by

$$\Xi = \Xi^{\text{abs}} - Z \Xi_{\text{H}^+}^{\text{abs}}, \quad (1)$$

where  $\Xi$  and  $\Xi^{\text{abs}}$ , respectively, stand for any conventional or absolute standard molal thermodynamic property or equation of state parameter of the aqueous species of interest,  $\Xi_{\text{H}^+}^{\text{abs}}$  denotes the corresponding absolute standard molal thermodynamic property or equation of state parameter of the hydrogen ion, and  $Z$  represents the formal charge of the aqueous species of interest. The conventional standard molal thermodynamic properties and equations of state parameters of H<sup>+</sup> are therefore zero at all temperatures and pressures. Calculations taking account of the biochemical standard state, which is referenced to neutral pH, can be referenced to the standard state adopted here by taking account of the appropriate equations (LaRowe and Helgeson, 2006).

The standard molal enthalpies ( $\Delta H^\circ$ ) and Gibbs energies ( $\Delta G^\circ$ ) of aqueous species given below are expressed as apparent standard partial molal thermodynamic properties, which are defined by

$$\Delta H^\circ \equiv \Delta H_f^\circ + (H_{P,T}^\circ - H_{P_r,T_r}^\circ) \quad (2)$$

and

$$\Delta G^\circ \equiv \Delta G_f^\circ + (G_{P,T}^\circ - G_{P_r,T_r}^\circ), \quad (3)$$

where  $\Delta H_f^\circ$  and  $\Delta G_f^\circ$  represent the standard molal enthalpy and Gibbs energy of formation of the species from the elements at the reference temperature ( $T_r$ ) of 298.15 K (25°C) and pressure ( $P_r$ ) of 1 bar, and  $H_{P,T}^\circ - H_{P_r,T_r}^\circ$  and  $G_{P,T}^\circ - G_{P_r,T_r}^\circ$  denote the differences between the standard molal enthalpy and Gibbs energy, respectively, at the temperature ( $T$ ) and pressure ( $P$ ) of interest, and those at  $T_r$  and  $P_r$ .

The values of  $\Delta G_f^\circ$ ,  $\Delta H_f^\circ$ , and the standard molal entropy at 25°C and 1 bar ( $S_{P_r,T_r}^\circ$ ) are related by

$$\Delta G_f^\circ = \Delta H_f^\circ - T_r (S_{P_r,T_r}^\circ - S_{P_r,T_r,\text{elements}}^\circ), \quad (4)$$

where  $S_{P_r,T_r,\text{elements}}^\circ$  represents the total standard molal entropy at 25°C and 1 bar of the elements making up the species of interest. The values of the standard molal entropies of the elements at 25°C and 1 bar used in the present study were taken from Cox et al. (1989).

### 2.2 Summary of the revised HKF equations of state

The revised HKF equations of state are consistent with the separation of variables represented by

$$\Xi = \Delta \Xi_n + \Delta \Xi_s, \quad (5)$$

where  $\Xi$  stands for any standard molal thermodynamic property or equation of state parameter of an aqueous species, and  $\Delta \Xi_n$  and  $\Delta \Xi_s$  refer, respectively, to the nonsolvation and solvation contributions to that property or parameter.

The nonsolvation contributions to the standard partial molal isobaric heat capacity, volume, isothermal compressibility, and expansibility ( $\Delta C_{P,n}^\circ$ ,  $\Delta V_n^\circ$ ,  $\Delta \kappa_{T,n}^\circ$ , and  $\Delta E_n^\circ$ , respectively) of an aqueous species can be expressed as (Tanger and Helgeson, 1988)

$$\Delta C_{P,n}^\circ = c_1 + \frac{c_2}{(T - \Theta)^2} - \left( \frac{2T}{(T - \Theta)^3} \right) \left( a_3 (P - P_r) + a_4 \ln \left( \frac{\Psi + P}{\Psi + P_r} \right) \right), \quad (6)$$

$$\Delta V_n^\circ = a_1 + \frac{a_2}{\Psi + P} + \left( a_3 + \frac{a_4}{\Psi + P} \right) \left( \frac{1}{T - \Theta} \right), \quad (7)$$

$$\Delta \kappa_{T,n}^\circ = \left( a_2 + \frac{a_4}{T - \Theta} \right) \left( \frac{1}{\Psi + P} \right)^2, \quad (8)$$

and

$$\Delta E_n^\circ = - \left( a_3 + \frac{a_4}{\Psi + P} \right) \left( \frac{1}{T - \Theta} \right)^2, \quad (9)$$

where  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$ ,  $c_1$ , and  $c_2$  represent temperature- and pressure-independent parameters of the species of interest, and  $\Theta$  and  $\Psi$  represent solvent parameters equal to 228 K and 2600 bar, respectively. The solvation contributions to  $C_p^\circ$ ,  $V^\circ$ ,  $\kappa_T^\circ$ , and  $E^\circ$  are given by

$$\Delta C_{p,s}^\circ = \omega T X + 2TY \left( \frac{\partial \omega}{\partial T} \right)_P - T \left( \frac{1}{\epsilon} - 1 \right) \left( \frac{\partial^2 \omega}{\partial T^2} \right)_P, \quad (10)$$

$$\Delta V_s^\circ = -\omega Q + \left( \frac{1}{\epsilon} - 1 \right) \left( \frac{\partial \omega}{\partial P} \right)_T, \quad (11)$$

$$\Delta \kappa_{T,s}^\circ = \omega N + 2Q \left( \frac{\partial \omega}{\partial P} \right)_T - \left( \frac{1}{\epsilon} - 1 \right) \left( \frac{\partial^2 \omega}{\partial P^2} \right)_T, \quad (12)$$

and

$$\Delta E_s^\circ = -\omega U - Q \left( \frac{\partial \omega}{\partial T} \right)_P - Y \left( \frac{\partial \omega}{\partial P} \right)_T + \left( \frac{1}{\epsilon} - 1 \right) \left( \partial \left( \frac{\partial \omega}{\partial P} \right)_T / \partial T \right)_P, \quad (13)$$

where  $\omega$  denotes the solvation parameter of the species of interest,  $\epsilon$  stands for the dielectric constant of H<sub>2</sub>O, and  $Q$ ,  $N$ ,  $Y$ ,  $X$ , and  $U$  represent partial derivatives of the reciprocal dielectric constant of H<sub>2</sub>O equal to  $-\left(\frac{\partial(1/\epsilon)}{\partial P}\right)_T$ ,  $\left(\frac{\partial Q}{\partial P}\right)_T$ ,  $-\left(\frac{\partial(1/\epsilon)}{\partial T}\right)_P$ ,  $\left(\frac{\partial Y}{\partial T}\right)_P$ , and  $\left(\frac{\partial Q}{\partial T}\right)_P$ , respectively. The values of  $Q$ ,  $N$ ,  $Y$ ,  $X$ , and  $U$  used in the present study were taken from Tanger and Helgeson (1988) and Shock et al. (1992).

The solvation parameter  $\omega$  and its first and second partial derivatives with respect to temperature and pressure appear in Eqs. (10–13). The partial derivatives of  $\omega$  generally are calculated from charge-dependent correlations (Tanger and Helgeson, 1988). The values of  $\omega$  of neutral species are in fact effective values, and their partial derivatives with respect to temperature and pressure commonly are taken to be equal to zero (Shock et al., 1989). Hence, group additivity algorithms for zwitterionic species such as neutral amino acids can be facilitated by taking account of

$$\left( \frac{\partial \omega}{\partial T} \right)_P = \left( \frac{\partial \omega}{\partial P} \right)_T = \left( \frac{\partial^2 \omega}{\partial T^2} \right)_P = \left( \frac{\partial^2 \omega}{\partial P^2} \right)_P = 0, \quad (14)$$

which we take as an approximation for all the neutral and charged groups and species considered below.

### 2.3 Revised HKF equations of state for $\Delta G^\circ$ , $\Delta H^\circ$ , and $S^\circ$

The revised HKF equations of state for  $\Delta G^\circ$ ,  $\Delta H^\circ$ , and  $S^\circ$  can be written as (Tanger and Helgeson, 1988)

$$\begin{aligned} \Delta G^\circ = & \Delta G_f^\circ - S_{P_r, T_r}^\circ (T - T_r) \\ & - c_1 \left[ T \ln \left( \frac{T}{T_r} \right) - T + T_r \right] \\ & - c_2 \left\{ \left[ \left( \frac{1}{T - \Theta} \right) - \left( \frac{1}{T_r - \Theta} \right) \right] \left( \frac{\Theta - T}{\Theta} \right) \right\} \\ & + c_2 \left\{ \frac{T}{\Theta^2} \ln \left[ \frac{T_r (T - \Theta)}{T (T_r - \Theta)} \right] \right\} \\ & + a_1 (P - P_r) + a_2 \ln \left( \frac{\Psi + P}{\Psi + P_r} \right) \\ & + \left( \frac{1}{T + \Theta} \right) \left[ a_3 (P - P_r) + a_4 \ln \left( \frac{\Psi + P}{\Psi + P_r} \right) \right] \\ & \omega \left( \frac{1}{\epsilon} - 1 \right) - \omega_{P_r, T_r} \left[ Y_{P_r, T_r} (T_r - T) + \frac{1}{\epsilon_{P_r, T_r}} - 1 \right], \end{aligned} \quad (15)$$

$$\begin{aligned} \Delta H^\circ = & \Delta H_f^\circ + c_1 (T - T_r) \\ & - c_2 \left[ \left( \frac{1}{T - \Theta} \right) - \left( \frac{1}{T_r - \Theta} \right) \right] \\ & + a_1 (P - P_r) + a_2 \ln \left( \frac{\Psi + P}{\Psi + P_r} \right) \\ & + \left( \frac{2T - \Theta}{(T - \Theta)^2} \right) \left[ a_3 (P - P_r) + a_4 \ln \left( \frac{\Psi + P}{\Psi + P_r} \right) \right] \\ & + \omega \left( TY + \frac{1}{\epsilon} - 1 \right) - T \left( \frac{1}{\epsilon} - 1 \right) \left( \frac{\partial \omega}{\partial T} \right)_P \\ & - \omega_{P_r, T_r} \left( T_r Y_{P_r, T_r} + \frac{1}{\epsilon_{P_r, T_r}} - 1 \right), \end{aligned} \quad (16)$$

and

$$\begin{aligned} S^\circ = & S_{P_r, T_r}^\circ + c_1 \ln \left( \frac{T}{T_r} \right) \\ & - \frac{c_2}{\Theta} \left\{ \left( \frac{1}{T - \Theta} \right) - \left( \frac{1}{T_r - \Theta} \right) \right\} \\ & - c_2 \left\{ \frac{1}{\Theta^2} \ln \left[ \frac{T_r (T - \Theta)}{T (T_r - \Theta)} \right] \right\} \\ & + \left( \frac{1}{T - \Theta} \right)^2 \left[ a_3 (P - P_r) + a_4 \ln \left( \frac{\Psi + P}{\Psi + P_r} \right) \right] \\ & + \omega Y - \left( \frac{1}{\epsilon} - 1 \right) \left( \frac{\partial \omega}{\partial T} \right)_P - \omega_{P_r, T_r} Y_{P_r, T_r}, \end{aligned} \quad (17)$$

respectively. The calculations of  $\Delta G^\circ$ ,  $\Delta H^\circ$ , and  $S^\circ$  used in the present study take account of Eqs. (15)–(17) together with Eq. (14) and, hence,  $\omega = \omega_{P_r, T_r}$ .

## 2.4 Regression equations

Values of the nonsolvation revised HKF parameters can be retrieved by combining experimental or estimated values of the standard molal isobaric heat capacity, volume, and isothermal compressibility ( $C_p^\circ$ ,  $V^\circ$ , and  $\kappa_T^\circ$ , respectively) with values of  $\omega$  to calculate the corresponding nonsolvation properties as a function of temperature. These values can then be expressed as a linear function of two nonsolvation revised HKF parameters. At  $P \approx P_r$ , Eq. (6) reduces to

$$\Delta C_{P,n}^\circ = c_1 + \frac{c_2}{(T - \Theta)^2}, \quad (18)$$

which can be used to retrieve values of  $c_1$  and  $c_2$ . The contributions by the pressure-dependent terms are negligible for pressures less than a few hundred bars. A regression equation for volumetric data can be derived by defining the parameters  $\sigma$  and  $\xi$  as

$$\sigma \equiv a_1 + \frac{a_2}{\Psi + P} \quad (19)$$

and

$$\xi \equiv a_3 + \frac{a_4}{\Psi + P}. \quad (20)$$

The conversion between the volumetric units of  $\sigma$  and  $\xi$  and the energetic units of  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$  is given by  $1 \text{ cal} = 41.84 \text{ cm}^3 \text{ bar}$ . Combining Eqs. (19) and (20) with Eq. (7) yields

$$\Delta V_n^\circ = \sigma + \frac{\xi}{T - \Theta}, \quad (21)$$

which can be used to retrieve values of  $\sigma$  and  $\xi$ . The partial derivatives with respect to pressure of Eqs. (19) and (20) can be expressed as

$$\left( \frac{\partial \sigma}{\partial P} \right)_T = \frac{-a_2}{(\Psi + P)^2} \quad (22)$$

and

$$\left( \frac{\partial \xi}{\partial P} \right)_T = \frac{-a_4}{(\Psi + P)^2}, \quad (23)$$

which can be combined with Eq. (8) to give

$$-\Delta \kappa_{T,n}^\circ = \left( \frac{\partial \sigma}{\partial P} \right)_T + \left( \frac{\partial \xi}{\partial P} \right)_T \left( \frac{1}{T - \Theta} \right). \quad (24)$$

Equation (24) can be used to retrieve values of  $\left( \frac{\partial \sigma}{\partial P} \right)_T$  and  $\left( \frac{\partial \xi}{\partial P} \right)_T$ .

Regression equations that involve the solvation parameter can be written by taking account of Eqs. (5) and (14) together with Eqs. (10) and (18) for  $\Delta C_{P,n}^\circ$  or Eqs. (11) and (21) for  $\Delta V_n^\circ$ . Thus, it is possible to write

$$C_p^\circ = c_1 + \frac{c_2}{(T - \Theta)^2} + \omega T X \quad (25)$$

and

$$V^\circ = \sigma + \frac{\xi}{T - \Theta} - \omega Q, \quad (26)$$

which can be used to retrieve simultaneously values of  $c_1$ ,  $c_2$ , and  $\omega$ , or  $\sigma$ ,  $\xi$ , and  $\omega$  where sufficient high-temperature experimental data (above 100°C) are available.

## 3 Revised HKF equations of state parameters of amino acids and Gly-X-Gly tripeptides

Values of the revised HKF equations of state parameters of amino acids and Gly-X-Gly tripeptides summarized below were derived from experimental measurements reported in the literature of the standard molal heat capacities, volumes, and isothermal compressibilities as a function of temperature and pressure. Calorimetric data above 100°C are not available for most of the amino acids considered here, so accurate independent retrieval of the solvation parameter,  $\omega$ , in the revised HKF equations of state generally is not possible. Hence, a correlation algorithm was developed to estimate values of  $\omega$ . Earlier correlations between  $\omega$  and  $S_{P_r, T_r}^\circ$ , or between  $\omega$  and the standard molal Gibbs energy of hydration at 25°C and 1 bar ( $\Delta G_{\text{hyd}}^\circ$ ) of neutral species were used by Shock et al. (1989) and Plyasunov and Shock (2001). Exploratory calculations revealed that neither of these correlations could be used to represent values of  $\omega$  of the neutral amino acids that are consistent with high-temperature heat capacity data. Instead, we found a correlation between  $\omega$  and the standard molal entropy of hydration, denoted by  $\Delta S_{\text{hyd}}^\circ$ . The standard molal thermodynamic properties of both gaseous and aqueous species at 25 °C and 1 bar are required for the calculation of the hydration property,  $\Delta S_{\text{hyd}}^\circ$ , and are summarized below.

### 3.1 Standard molal thermodynamic properties of gaseous and aqueous amino acids

The values of  $\Delta G_f^\circ$ ,  $\Delta H_f^\circ$ , and  $S_{P_r, T_r}^\circ$  of the gaseous amino acids given in Table 1 were calculated by first taking account of the sidechain group additivity algorithm represented by

$$\Xi_{[SC]} = \sum_i n_i \Xi_i, \quad (27)$$

where  $\Xi_{[SC]}$  and  $\Xi_i$  stand for any thermodynamic property or parameter of the sidechain group and the  $i$ th organic group, respectively, and  $n_i$  stands for the number of the  $i$ th organic groups present in the sidechain group. Here, *organic groups* refers to groups whose properties can be summed to model the properties of sidechain groups. We used the group

**Table 1.** Group additivity algorithms for gaseous amino acid sidechain and backbone groups, and corresponding values of  $\Delta G_f^\circ$ ,  $\Delta H_f^\circ$ , and  $S_{P_r, T_r}^\circ$ . Unless otherwise noted, values of  $\Delta H_f^\circ$  and  $S_{P_r, T_r}^\circ$  of the groups were taken from Domalski and Hearing (1993).<sup>a</sup> Square brackets here and in the text are used to denote groups.

Species or Group	Sidechain group additivity algorithm	$\Delta G_f^{\circ b}$	$\Delta H_f^{\circ b}$	$S_{P_r, T_r}^{\circ c}$
Ala <sup>d</sup>	[C-(H) <sub>3</sub> (C)]	-71.91	-99.10	94.17
Arg	2[C-(H) <sub>2</sub> (C) <sub>2</sub> ] + [C-(H) <sub>2</sub> (C)(N)] + [N-(H)(C) <sub>2</sub> ] + [C-(N) <sub>2</sub> (N <sub>A</sub> )] + [N-(H) <sub>2</sub> (C)] <sub>second.amino acids</sub> + [N <sub>A</sub> -(H)(C)]	2.14	-69.05	128.73
Asn	[C-(H) <sub>2</sub> (CO)(C)] + [CO-(C)(N)] + [N-(H) <sub>2</sub> (CO)] <sub>amides, ureas</sub>	-98.81	-141.13	107.84
Asp	[C-(H) <sub>2</sub> (CO)(C)] + [CO-(C)(O)] + [O-(H)(CO)]	-151.04	-187.80	112.47
Cys	[C-(H) <sub>2</sub> (C)(S)] + [S-(C)(H)]	-64.33	-90.08	106.65
Gln	[C-(H) <sub>2</sub> (C) <sub>2</sub> ] + [C-(H) <sub>2</sub> (CO)(C)] + [CO-(C)(N)] + [N-(H) <sub>2</sub> (CO)] <sub>amides, ureas</sub>	-96.81	-146.06	117.20
Glu	[C-(H) <sub>2</sub> (C) <sub>2</sub> ] + [C-(H) <sub>2</sub> (CO)(C)] + [CO-(C)(O)] + [O-(H)(CO)]	-149.04	-192.73	121.83
Gly <sup>d</sup>		-73.35	-93.30	85.84
His	S <sup>o</sup> : [C-(H) <sub>2</sub> (C)(C <sub>B</sub> )] + [C <sub>B</sub> -(N)(C <sub>B</sub> )(C)] + [C <sub>B</sub> -(H)(C <sub>B</sub> )(N <sub>A</sub> )] + [N <sub>I</sub> -(C <sub>B</sub> )] + [C <sub>B</sub> -(H)(N <sub>I</sub> )(N)] + [N-(H)(C <sub>B</sub> ) <sub>2</sub> ] $\Delta H_f^\circ$ : 2-ethylimidazole - [C-(H) <sub>3</sub> (C)]	-14.31	-62.60	104.53
Ile	[C-(H) <sub>2</sub> (C) <sub>2</sub> ] + 2[C-(H) <sub>3</sub> (C)] + [C-(H)(C) <sub>3</sub> ]	-66.10	-114.41	121.15
Leu	[C-(H) <sub>2</sub> (C) <sub>2</sub> ] + 2[C-(H) <sub>3</sub> (C)] + [C-(H)(C) <sub>3</sub> ]	-66.10	-114.41	121.15
Lys	3[C-(H) <sub>2</sub> (C) <sub>2</sub> ] + [C-(H) <sub>2</sub> (C)(N)] + [N-(H) <sub>2</sub> (C)] <sub>second.amino acids</sub>	-49.47	-105.95	132.25
Met	[C-(H) <sub>2</sub> (C) <sub>2</sub> ] + [C-(H) <sub>2</sub> (C)(S)] + [S-(C) <sub>2</sub> ] + [C-(H) <sub>3</sub> (S)]	-59.13	-98.34	126.73
Phe	[C-(H) <sub>2</sub> (C)(C <sub>B</sub> )] + [C <sub>B</sub> -(C)(C <sub>B</sub> ) <sub>2</sub> ] + 5[C <sub>B</sub> -(H)(C <sub>B</sub> ) <sub>2</sub> ]	-32.33	-71.95	123.16
Pro <sup>d</sup>		-55.66	-87.52	112.48
Ser	[C-(H) <sub>2</sub> (O)(C)] + [O-(H)(C)]	-103.12	-134.94	103.16
Thr	[C-(H)(O)(C) <sub>2</sub> ] <sub>alcohols, peroxides</sub> + [C-(H) <sub>3</sub> (C)] + [O-(H)(C)]	-104.79	-143.42	112.92
Trp	[C-(H) <sub>2</sub> (C)(C <sub>B</sub> )] + [C <sub>B</sub> -(C)(C <sub>B</sub> ) <sub>2</sub> ] + [C <sub>B</sub> -(H)(C <sub>B</sub> )(N)] + [N-(H)(C <sub>B</sub> ) <sub>2</sub> ] + [C <sub>B</sub> -(N)(C <sub>B</sub> ) <sub>2</sub> ] + 4[C <sub>B</sub> -(H)(C <sub>B</sub> ) <sub>2</sub> ] + [C <sub>BF</sub> -(C <sub>BF</sub> )(C <sub>B</sub> ) <sub>2</sub> ]	1.48	-51.13	120.88
Tyr	[C-(H) <sub>2</sub> (C)(C <sub>B</sub> )] + [C <sub>B</sub> -(C)(C <sub>B</sub> ) <sub>2</sub> ] + 4[C <sub>B</sub> -(H)(C <sub>B</sub> ) <sub>2</sub> ] + [C <sub>B</sub> -(O)(C <sub>B</sub> ) <sub>2</sub> ] + [O-(H)(C <sub>B</sub> ) <sub>2</sub> ]	-69.87	-114.70	130.20
Val	2[C-(H) <sub>3</sub> (C)] + [C-(H)(C) <sub>3</sub> ]	-68.10	-109.48	111.79
[AABB] <sup>e</sup>		-67.12	-89.00	63.74

<sup>a</sup>:  $\Delta H_f^\circ$  and  $S_{P_r, T_r}^\circ$  of [C<sub>B</sub>-(N)(C<sub>B</sub>)(C)], and [C-(N)<sub>2</sub>(N<sub>A</sub>)] were taken to be equivalent to those of [C<sub>B</sub>-(N)(C<sub>B</sub>)<sub>2</sub>].  $\Delta H_f^\circ$  and  $S^\circ$  of [C<sub>B</sub>-(H)(N<sub>I</sub>)(N)], [N<sub>A</sub>-(H)(C)] and [C<sub>B</sub>-(H)(C<sub>B</sub>)(N<sub>A</sub>)] were taken to be equivalent to those of [C-(H)<sub>2</sub>(C)(N)], [N-(H)(C)<sub>2</sub>], and [C<sub>B</sub>-(H)(C<sub>B</sub>)<sub>2</sub>], respectively. We used  $S_{[N-(H)(C<sub>B</sub>)<sub>2</sub>]}^\circ = S_{[N-(H)(C)<sub>2</sub>]}^\circ$ ,  $\Delta H_f^\circ$ , [C<sub>B</sub>-(H)(C<sub>B</sub>)(N)] =  $\frac{1}{2} \Delta H_f^\circ$ , pyrrole -  $\Delta H_f^\circ$ , [C<sub>B</sub>-(H)(C<sub>B</sub>)<sub>2</sub>] -  $\frac{1}{2} \Delta H_f^\circ$ , [N-(H)(C<sub>B</sub>)<sub>2</sub>], and  $S_{[C<sub>B</sub>-(H)(C<sub>B</sub>)(N)]}^\circ = S_{[C<sub>B</sub>-(H)(C<sub>B</sub>)<sub>2</sub>]}^\circ$ . The value of  $\Delta H_f^\circ$  of 2-ethylimidazole used in the calculations was 16.30 kcal mol<sup>-1</sup> (Jiménez et al., 1992); that of pyrrole was 25.88 kcal mol<sup>-1</sup> (Scott et al., 1967). <sup>b</sup>: kcal mol<sup>-1</sup>. <sup>c</sup>: cal mol<sup>-1</sup> K<sup>-1</sup>. <sup>d</sup>: Values of  $\Delta G_f^\circ$  of these amino acids were calculated from the corresponding hydration property ( $\Delta G_{hyd}^\circ$ ) (Plyasunov and Shock, 2001) and  $\Delta G_f^\circ$  of the corresponding aqueous amino acids. Values of  $\Delta H_f^\circ$  were taken from Ngauv et al. (1977) and Sabbah and Laffitte (1978). <sup>e</sup>: See text.

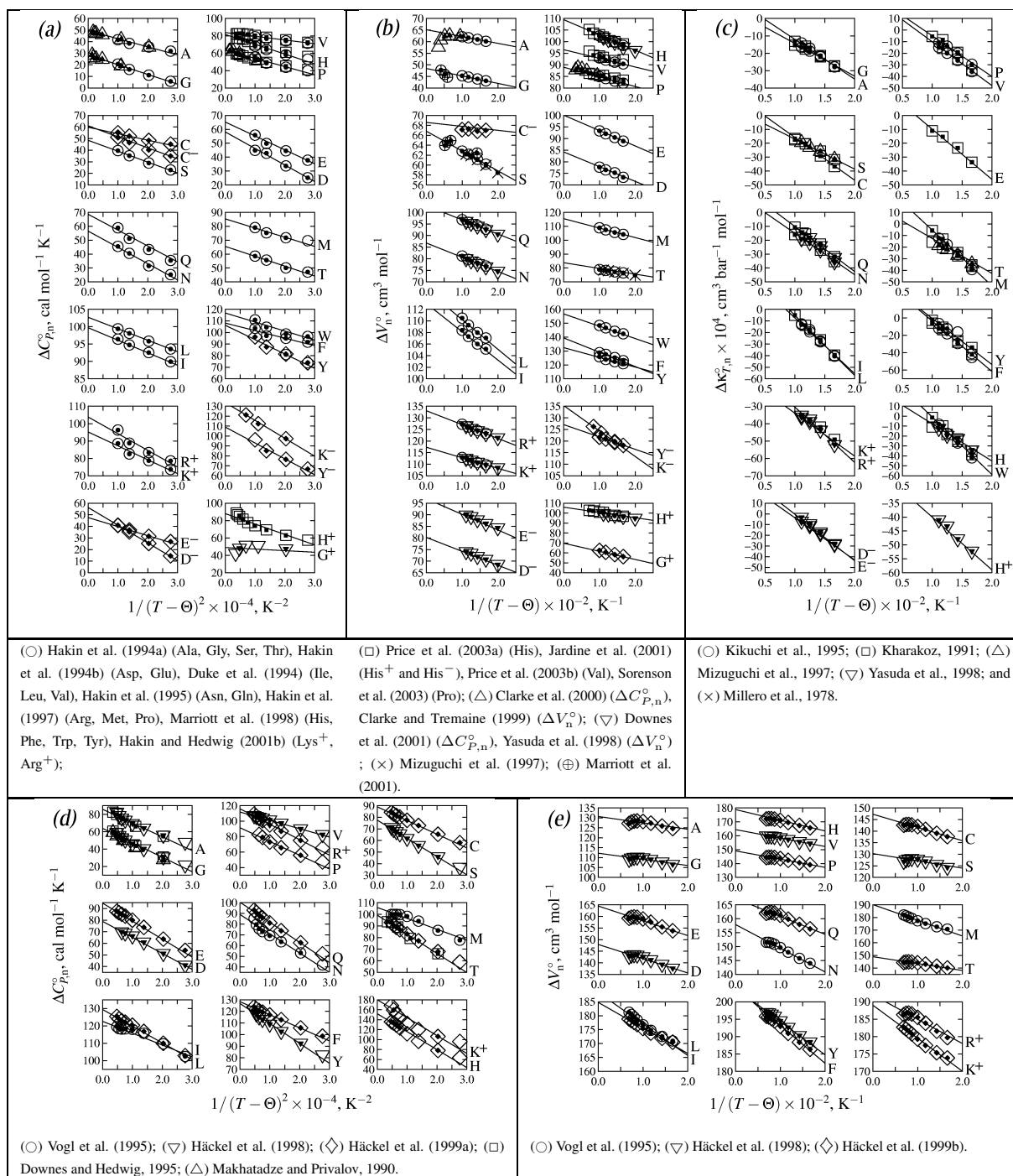
additivity schemes developed by Benson and Buss (1958) and Domalski and Hearing (1993) to generate the specific statements of Eq. (27) for gaseous amino acids shown in Table 1. Because of their structures, the sidechain groups of Gly and Pro are not readily modeled by the additivity algorithm. Instead,  $\Delta H_f^\circ$  and  $S_{P_r, T_r}^\circ$  of Gly and Pro were calculated from available measurements.

The values of  $\Delta H_f^\circ$  and  $S_{P_r, T_r}^\circ$  of gaseous [AABB] given in Table 1 were calculated from those of [C-(H)<sub>3</sub>(C)] and the corresponding experimental properties of Ala, together with the amino acid group additivity algorithm represented by

$$\Xi_{AA} = \Xi_{[AABB]} + \Xi_{[SC]}, \quad (28)$$

where  $\Xi_{[AABB]}$ ,  $\Xi_{AA}$ , and  $\Xi_{[SC]}$  stand for any standard molal thermodynamic property or equation of state parameter of the amino acid backbone group, an amino acid, and its sidechain group. The values of  $\Delta H_f^\circ$  and  $S_{P_r, T_r}^\circ$  of the other amino acids shown in Table 1 were calculated from those of [AABB] together with Eq. (28) and the values for the sidechain groups obtained using Eq. (27).

The values of  $\Delta G_f^\circ$ ,  $\Delta H_f^\circ$ , and  $S_{P_r, T_r}^\circ$  of the aqueous amino acids given in Table 2 were taken from Amend and Helgeson (1997a), except for those of Cys<sup>-</sup> and His<sup>+</sup>. The latter values were calculated from the standard molal Gibbs energies and enthalpies of ionization ( $\Delta G_{ion}^\circ$  and  $\Delta H_{ion}^\circ$ ) of these charged species taken from Goldberg et al. (2002)



**Fig. 2.** Regression plots for the revised HKF parameters of amino acids (a–c) and Gly–X–Gly tripeptides (d) and (e). The identities of the amino acids and Gly–X–Gly tripeptides are indicated by one-letter abbreviations to the right of the plots. The intercepts and slopes of the regression lines for  $\Delta C_{P,n}^{\circ}$ ,  $\Delta V_n^{\circ}$  and  $\Delta \kappa_{T,n}^{\circ}$  correspond to  $c_1$  and  $c_2$ ,  $\sigma$  and  $\xi$ , and  $-(\partial\sigma/\partial P)_T$  and  $-(\partial\xi/\partial P)_T$ , respectively, in the revised HKF equations of state. The values were calculated from experimental values of  $C_P^{\circ}$ ,  $V^{\circ}$  and  $\kappa_S^{\circ}$  taken from the sources cited in the caption of each subfigure. The values estimated in the present study are represented by open diamonds (◇) in (a) and (b). The sizes of the symbols were chosen to aid visual identification and do not necessarily indicate the experimental uncertainty; however, values that lie within the uncertainties of  $C_P^{\circ}$ ,  $V^{\circ}$  and  $\kappa_T^{\circ}$  given in Sect. 6 are represented by a small filled center point.



**Table 2.** Standard molal thermodynamic properties at 25°C and 1 bar and revised HKF equations of state parameters of aqueous amino acids.

Species	$\Delta G_f^\circ$ <sup>a</sup>	$\Delta H_f^\circ$ <sup>a</sup>	$S_{Pr,T_r}^\circ$ <sup>b</sup>	$\omega^c$ $\times 10^{-5}$	$c_1^b$	$c_2^d$ $\times 10^{-4}$	$a_1^e$ $\times 10$	$a_2^c$ $\times 10^{-2}$	$a_3^f$	$a_4^d$ $\times 10^{-4}$	$C_p^\circ$ <sup>b</sup>	$V^\circ$ <sup>g</sup>	$\kappa_T^\circ$ <sup>c</sup> $\times 10^4$
Ala	-88.81	-132.50	38.83	0.18	49.5	-7.00	14.90	1.74	7.16	-3.69	33.6	60.4	-23.5
Arg	-57.36	-143.06	80.06	0.22	94.8	-12.50	28.83	8.21	7.20	-5.95	67.4	123.9	-3.7
Arg <sup>+</sup>	-69.93	-154.88	82.57	0.73	103.8	-9.60	31.72	0.22	5.46	-5.15	77.6	122.5	-50.9
Arg <sup>-</sup>	-40.34	-129.42	68.72	2.30	141.9	-19.60	31.28	13.76	32.12	-9.94	81.1	143.0	-24.1
Asn	-125.49	-182.70	57.88	0.21	56.5	-11.70	19.83	2.37	3.76	-4.81	30.8	77.4	-29.7
Asp	-172.51	-226.34	55.22	0.17	58.1	-11.90	18.96	3.05	-6.62	-2.23	32.4	74.7	-2.4
Asp <sup>-</sup>	-167.17	-224.54	43.34	2.47	56.3	-15.30	16.96	5.77	10.11	-6.37	2.7	64.8	-43.6
Cys	-80.33	-124.07	46.32	0.15	59.8	-5.60	18.14	2.40	9.31	-4.97	47.1	73.5	-30.4
Cys <sup>-</sup>	-68.92	-115.44	37.01	2.59	61.0	-9.50	14.44	5.08	12.24	-3.68	18.1	60.6	-24.3
Gln	-126.28	-191.86	62.41	0.18	68.6	-11.40	23.22	3.58	5.44	-5.23	43.8	93.7	-25.7
Glu	-173.05	-234.82	61.20	0.15	65.2	-10.00	22.30	4.23	6.55	-5.84	43.5	90.2	-26.7
Glu <sup>-</sup>	-167.21	-234.15	43.86	2.65	47.6	-7.80	20.49	6.96	10.86	-7.02	7.6	80.0	-43.7
Gly	-90.95	-124.78	39.29	0.23	28.5	-8.40	11.30	0.71	3.99	-3.04	9.3	43.2	-24.6
Gly <sup>+</sup>	-94.16	-125.72	46.91	0.59	49.0	-1.80	19.57	-7.58	-40.35	5.41	40.0	56.4	-4.7
His	-48.42	-108.20	66.00	0.27	81.6	-11.30	24.32	4.80	7.52	-6.05	56.2	99.3	-26.2
His <sup>+</sup>	-56.70	-115.25	70.13	0.50	88.0	-12.00	26.64	-3.40	-0.88	-3.07	59.1	97.0	-52.8
Ile	-81.99	-151.60	49.70	0.09	99.7	-3.60	24.49	6.55	18.16	-7.76	91.6	105.8	-28.8
Leu	-84.20	-153.60	50.41	0.09	102.7	-3.30	24.68	7.51	19.93	-8.37	95.2	107.8	-28.2
Lys	-80.68	-159.33	57.91	0.07	86.6	-10.90	24.56	8.30	24.43	-9.53	63.8	108.6	-33.3
Lys <sup>+</sup>	-92.89	-170.37	61.82	1.21	95.2	-8.00	28.62	-1.51	4.47	-3.99	67.9	107.5	-55.9
Lys <sup>-</sup>	-66.31	-146.34	53.27	2.67	133.7	-18.00	28.29	10.53	2.97	-7.61	72.8	112.2	-27.0
Met	-120.12	-178.52	62.36	0.13	85.3	-6.60	24.95	6.90	13.59	-7.77	70.7	105.4	-27.1
Phe	-49.43	-108.90	56.60	0.12	108.1	-6.50	28.27	8.80	19.34	-9.38	93.8	122.1	-29.4
Pro	-73.56	-124.10	49.83	0.14	63.6	-9.50	19.39	4.87	11.88	-5.68	43.0	82.6	-21.3
Ser	-123.93	-172.42	47.24	0.18	48.5	-9.40	15.69	0.73	3.87	-3.49	27.8	60.6	-27.9
Thr	-119.83	-178.94	44.23	0.11	65.5	-7.10	18.94	2.83	8.87	-4.87	50.1	77.6	-26.5
Trp	-26.82	-97.59	59.96	0.15	116.8	-7.80	33.91	9.21	14.78	-9.36	99.6	143.6	-27.0
Tyr	-91.80	-157.74	59.41	0.09	106.2	-12.40	30.06	8.49	8.36	-8.58	80.2	124.5	-24.0
Tyr <sup>-</sup>	-79.38	-149.42	45.66	2.78	109.0	-16.00	26.34	10.49	16.42	-7.59	51.2	111.8	-28.1
Val	-85.33	-146.42	45.68	0.12	83.8	-4.30	21.35	4.48	14.58	-6.12	74.0	90.9	-27.4

a: kcal mol<sup>-1</sup>, b: cal mol<sup>-1</sup> K<sup>-1</sup>, c: cal mol<sup>-1</sup>, d: cal mol<sup>-1</sup> K, e: cal mol<sup>-1</sup> bar<sup>-1</sup>, f: cal K mol<sup>-1</sup> bar<sup>-1</sup>, g: cm<sup>3</sup> mol<sup>-1</sup>, h: cm<sup>3</sup> mol<sup>-1</sup> bar<sup>-1</sup>.

together with the properties of the corresponding neutral species given in Table 2.

### 3.2 Retrieval of the solvation parameter

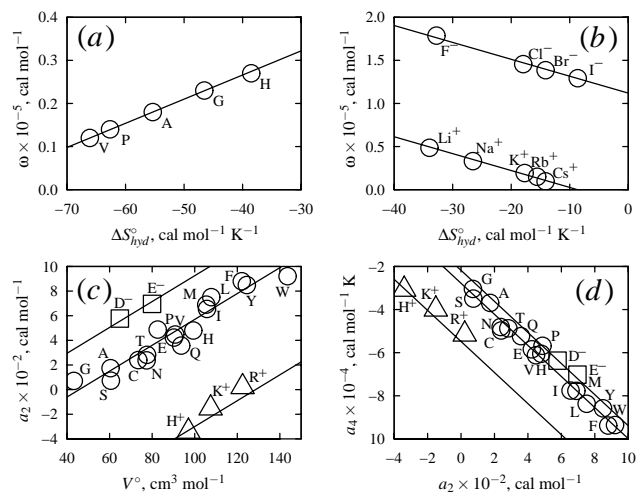
The equations for the correlations for  $\omega$  given in Table 3 were generated by first considering values of  $\omega$  of Gly, Ala, Pro, His, and Val that were independently regressed using Eq. (25) and experimental  $C_p^\circ$  data at temperatures >100°C (Clarke et al., 2000; Price et al., 2003a,b; Sorenson et al., 2003). The corresponding regression plots are shown in Fig. 2a. Values of  $\Delta S_{\text{hyd}}^\circ$  of the amino acids were calculated from those of  $S_{Pr,T_r}^\circ$  of the gaseous and aqueous amino acids taken from Tables 1 and 2. The equation for the correlation line shown in Fig. 3a was used to predict  $\omega$  of the other neutral amino acids. Sufficient high-temperature calorimetric data are not available to generate a similar regression for

charged amino acids. However, provisional estimates of  $\omega$  of charged amino acids were calculated using the correlations between  $\Delta S_{\text{hyd}}^\circ$  and  $\omega$  of alkali- and fluoride- group metal ions represented in Fig. 3b.

Experimental heat capacities to 125°C of Gly–Gly–Gly, Gly–Ala–Gly and Gly–Thr–Gly have been reported in the literature (Makhatadze and Privalov, 1990; Downes and Hedwig, 1995). The regression plots shown in Fig. 2d are consistent with values of  $\omega$  of Gly–X–Gly given by

$$\omega_{\text{Gly-X-Gly}} \times 10^{-5} = \omega_{AA} \times 10^{-5} - 1.77, \quad (29)$$

where the value of  $\omega_{AA}$ , in cal mol<sup>-1</sup>, is that of the corresponding amino acid. The values of  $\omega$  of the Gly–X–Gly tripeptides given by Eq. (29) are negative, but those of the amino acids are positive. It should be emphasized that the values of  $\omega$  of neutral species are effective values;



**Fig. 3.** Correlations between  $\Delta S_{\text{hyd}}^{\circ}$  at 25°C and 1 bar and  $\omega$  for neutral amino acids (a) and metal anions and cations (b), and correlations between  $V^{\circ}$  and  $a_2$  (c), and between  $a_2$  and  $a_4$  (d) for neutral (○), positively charged (△), and negatively charged (□) amino acids. The values of  $\omega$  and  $S_{P_r, T_r}^{\circ}$  of the aqueous metal ions shown in (b) were taken from Shock and Helgeson (1988), and those of  $\Delta S_{P_r, T_r}^{\circ}$  of the gaseous metal ions were taken from Marcus (1994). See text for the sources of the other values.

consequently, this difference in sign may not be indicative of divergent behavior of amino acids and Gly–X–Gly tripeptides at temperatures near the critical point of water.

Experimental values of  $V^{\circ}$  of Gly, Pro, and Ser at temperatures  $>100^{\circ}\text{C}$  (Clarke and Tremaine, 1999; Marriott et al., 2001) were combined with correlated values of  $\omega$  to calculate values of  $\Delta V_{\text{n}}^{\circ}$ . There is an overall linear trend apparent in the regression plots for  $\Delta V_{\text{n}}^{\circ}$  shown in Fig. 2b, which supports the use of the correlation equations for estimating values of  $\omega$  of these amino acids. This is also true for many of the regression plots of  $\Delta V_{\text{n}}^{\circ}$  of the Gly–X–Gly tripeptides shown in Fig. 2e. However, the values of  $\Delta V_{\text{n}}^{\circ}$  of Ala are considerably less than the predicted values at elevated temperatures. Although this represents a significant error in volumetric units ( $\sim 10\text{ cm}^3\text{ mol}^{-1}$ ), it corresponds to an energetic error (only  $\sim 0.24\text{ cal mol}^{-1}\text{ bar}^{-1}$ ) that is small compared to the uncertainties in  $C_p^{\circ}$ . A magnification of the error in  $V^{\circ}$  at high temperature calculated using calorimetric values of  $\omega$  has been observed for other neutral aqueous species (see, e.g., Schulte et al., 2001). Hence, we have more confidence in the values of  $\omega$  of amino acids retrieved from high-temperature experimental  $C_p^{\circ}$  data.

### 3.3 Retrieval of non-solvation parameters

The experimental or calculated values of  $C_p^{\circ}$ ,  $V^{\circ}$ , and  $\kappa_T^{\circ}$  were combined with values of  $\omega$  to calculate values of  $\Delta C_{P, \text{n}}^{\circ}$ ,  $\Delta V_{\text{n}}^{\circ}$ , and  $\Delta \kappa_{T, \text{n}}^{\circ}$  using the revised HKF equations of state. The intercepts and slopes of regression lines on plots

**Table 3.** Correlation equations for the properties of amino acids (see Fig. 3).

Equation	Z
$\omega \times 10^{-5} = 0.49 + 0.0056 \times \Delta S_{\text{hyd}}^{\circ}$	0
$\omega \times 10^{-5} = 1.12 - 0.0196 \times \Delta S_{\text{hyd}}^{\circ}$	-1
$\omega \times 10^{-5} = -0.17 - 0.0196 \times \Delta S_{\text{hyd}}^{\circ}$	1
$a_2 \times 10^{-2} = -4.80 + 0.105 \times V^{\circ}$	0
$a_2 \times 10^{-2} = -1.25 + 0.105 \times V^{\circ}$	-1
$a_2 \times 10^{-2} = -13.50 + 0.105 \times V^{\circ}$	1
$a_4 \times 10^{-4} = -2.78 - 0.721a_2 \times 10^{-2}$	0
$a_4 \times 10^{-4} = -2.18 - 0.721a_2 \times 10^{-2}$	-1
$a_4 \times 10^{-4} = -5.48 - 0.721a_2 \times 10^{-2}$	1

of  $\Delta C_{P, \text{n}}^{\circ}$  vs.  $1/(T-\Theta)^2$  (Figs. 2a and d) correspond to  $c_1$  and  $c_2$ , respectively. Regression lines of  $\Delta V_{\text{n}}^{\circ}$  as a function of  $1/(T-\Theta)$  (Figs. 2b and e) have intercepts of  $\sigma$  and slopes of  $\xi$ , and those of  $\Delta \kappa_{T, \text{n}}^{\circ}$  (Fig. 2c) correspond to  $-(\partial\sigma/\partial P)_T$  and  $-(\partial\xi/\partial P)_T$ , respectively. The values of  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$  can be retrieved from those of  $\sigma$ ,  $\xi$ ,  $(\partial\sigma/\partial P)_T$ , and  $(\partial\xi/\partial P)_T$  using Eqs. (19), (20), (22), and (23). The values of these parameters are given in Table 2 for amino acids and in Table 5 for Gly–X–Gly tripeptides.

The properties of the charged amino acids or Gly–X–Gly tripeptides corresponding to an Na- or Cl- salt were calculated by subtracting from the measured property that of  $\text{Na}^+$  or  $\text{Cl}^-$  at each experimental temperature using the revised HKF parameters taken from Shock et al. (1997). The polynomial equations representing the experimental measurements of  $C_p^{\circ}$  and  $V^{\circ}$  of the Gly–X–Gly tripeptides given by Vogl et al. (1995), Häckel et al. (1998), and Häckel et al. (1999a) were used to generate the values at 15, 25, 40, 55, 70, 80, 90, and  $100^{\circ}\text{C}$  shown in Figs. 2d and e.

Experimental compressibility data for amino acids reported in the literature generally refer to isentropic compressibilities ( $\kappa_S^{\circ}$ ). Accordingly, values of the isothermal compressibilities ( $\kappa_T^{\circ}$ ) of the amino acids were calculated from (Desnoyers and Philip, 1972; Amend and Helgeson, 1997b)

$$\kappa_T^{\circ} = \kappa_S^{\circ} + \frac{TV_{\text{H}_2\text{O}}^{\circ}\alpha_{\text{H}_2\text{O}}^{\circ}\left(2E^{\circ} - V_{\text{H}_2\text{O}}^{\circ}\alpha_{\text{H}_2\text{O}}^{\circ}C_p^{\circ}/C_{p, \text{H}_2\text{O}}^{\circ}\right)}{C_{p, \text{H}_2\text{O}}^{\circ}}, \quad (30)$$

where  $E^{\circ}$  represents the standard molal isobaric expansibility of the amino acid, and  $C_{p, \text{H}_2\text{O}}^{\circ}$ ,  $V_{\text{H}_2\text{O}}^{\circ}$  and  $\alpha_{\text{H}_2\text{O}}^{\circ}$  represent the standard molal heat capacity, volume and coefficient of isobaric thermal expansion of  $\text{H}_2\text{O}$ , respectively. Values of  $C_p^{\circ}$ ,  $V^{\circ}$ , and  $E^{\circ}$  of the amino acids were calculated using Eqs. (9) and (13) together with the regressed or calculated values of  $\omega$ ,  $c_1$ ,  $c_2$ ,  $\sigma$ , and  $\xi$ . The values of  $C_{\text{H}_2\text{O}}^{\circ}$ ,  $V_{\text{H}_2\text{O}}^{\circ}$  and

**Table 4.** Reference model compound algorithms for estimating  $C_p^\circ$  and  $V^\circ$  of amino acids as a function of temperature.

Amino Acid	Property or Parameter	Reference Model Compound Algorithm
Asp <sup>-</sup>	$C_p^\circ$	Asp <sup>a</sup> + propanoate <sup>b</sup> – propanoic acid <sup>b</sup>
Cys	$C_p^\circ$	Ser <sup>a</sup> + Gly–Cys–Gly <sup>c</sup> – Gly–Ser–Gly <sup>c</sup>
Cys <sup>-</sup>	$C_p^\circ, V^\circ$	Cys + propanoate <sup>b</sup> – propanoic acid <sup>b</sup>
Glu <sup>-</sup>	$C_p^\circ$	Glu <sup>a</sup> + butanoate <sup>b</sup> – butanoic acid <sup>b</sup>
Gly <sup>+</sup>	$V^\circ$	Gly + propanoic acid <sup>b</sup> – propanoate <sup>b</sup>
Lys <sup>-</sup>	$C_p^\circ, V^\circ$	Lys + <i>n</i> -butanamine <sup>d</sup> – <i>n</i> -butanamine <sup>+</sup> <sup>d</sup>
Tyr	$C_p^\circ$	Phe <sup>e</sup> + Gly–Tyr–Gly <sup>f</sup> – Gly–Phe–Gly <sup>f</sup>
Tyr <sup>-</sup>	$C_p^\circ, V^\circ$	Tyr + propanoate <sup>b</sup> – propanoic acid <sup>b</sup>

<sup>a</sup>: Hakin et al. (1994a), <sup>b</sup>: Shock (1995), <sup>c</sup>: Häckel et al. (1998), <sup>d</sup>:  $C_p^\circ$ : Makhatadze and Privalov (1990);  $V^\circ$ : Makhatadze et al. (1990), <sup>e</sup>: Marriott et al. (1998), <sup>f</sup>: Häckel et al. (1999a).

$\alpha_{\text{H}_2\text{O}}$  in Eq. (30) were calculated using SUPCRT92 (Johnson et al., 1992).

### 3.4 Estimation algorithms for non-solvation parameters

Where experimental data were insufficient to regress the revised HKF parameters of amino acids, the model compound algorithms shown in Table 4 were used to estimate values of  $C_p^\circ$  and  $V^\circ$  as a function of temperature. This approach was used to generate values of  $C_p^\circ$  of Asp<sup>-</sup>, Glu<sup>-</sup>, Cys, Tyr, Cys<sup>-</sup>, and Tyr<sup>-</sup> and of  $V^\circ$  of Cys<sup>-</sup>, Gly<sup>+</sup>, and Tyr<sup>-</sup> at 15, 25, 40, and 55°C, as well as values of  $C_p^\circ$  and  $V^\circ$  of Lys<sup>-</sup> at 25, 50, and 75°C. Estimated values of  $C_p^\circ$  of Gly<sup>+</sup> were calculated by combining values of  $\Delta C_{p,\text{ion}}^\circ$  taken from Wang et al. (1996) with those of  $C_p^\circ$  of Gly calculated using the revised HKF parameters given in Table 2.

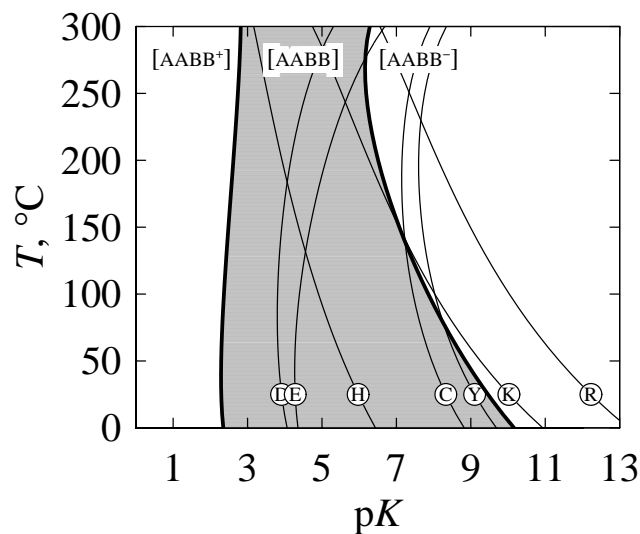
Experimentally unconstrained equations of state parameters of Arg<sup>-</sup> and Lys can be estimated by taking account of the corresponding ionizations of other amino acids. For example,  $c_2$  and  $\xi$  of Lys were calculated from

$$\Xi_{\text{Lys}} = \Xi_{\text{Lys}^+} + \Xi_{\text{Arg}} - \Xi_{\text{Arg}^+}. \quad (31)$$

The values of  $c_1$  and  $\sigma$  of Lys were then retrieved using Eqs. (18) and (21) and the measured values of  $C_p^\circ$  and  $V^\circ$  of Lys at 25°C and 1 bar (Jolicoeur et al., 1986). However, because no experimental values of  $C_p^\circ$  and  $V^\circ$  of Arg<sup>-</sup> can be found in the literature, the values of  $c_1$ ,  $c_2$ ,  $\sigma$ , and  $\xi$  of Arg<sup>-</sup> were estimated from

$$\Xi_{\text{Arg}^-} = \Xi_{\text{Arg}} + \Xi_{\text{Lys}^-} - \Xi_{\text{Lys}}. \quad (32)$$

Correlations between values of  $V^\circ$ ,  $a_2$ , and  $a_4$  of amino acids were used to estimate values of  $a_2$  and  $a_4$  of Arg, Arg<sup>-</sup>, Asp, Cys<sup>-</sup>, Gly<sup>+</sup>, Lys, Lys<sup>-</sup>, and Tyr<sup>-</sup>. This approach has been used for amino acids (Amend and Helgeson, 1997a) and for other organic species (Plyasunov and Shock, 2001). Updated correlations for the amino acids are depicted in Figs. 3c



**Fig. 4.** pKs of ionizable amino acid sidechain and backbone groups as a function of temperature at a pressure of 1 bar at temperatures below 100°C and the saturation vapor pressure of H<sub>2</sub>O at higher temperatures. Values for the ionization of the sidechain groups are represented by the thin lines. Values for [AABB<sup>+</sup>] and [AABB<sup>-</sup>] are denoted by the left and right bold lines, respectively, which delimit the predominance field of [AABB] relative to either of its charged counterparts.

and d. The intercepts of these correlation lines are charge-dependent, but the slopes in a first approximation can be considered to be independent of charge. This interpretation is consistent with the provisional correlations adopted by Amend and Helgeson (1997a).

### 3.5 Ionization as a function of temperature

The values of pK of the ionization of groups at 25°C and 1 bar can be calculated from the values of the standard molal Gibbs energy of formation ( $\Delta G_f^\circ$ ) of amino acids taken from the literature (see Table 2). These values are shown by the locations of the symbols in Fig. 4. It is apparent from this figure that, except for [Lys] and [Arg], the ionization of sidechain groups occurs in conditions where [AABB] is stable relative to either of its charged counterparts. This observation bears on the group additivity algorithms used to generate the parameters of the neutral and charged sidechain groups. The lines in Fig. 4 represent values of pK calculated as a function of temperature using the group contributions summarized below. Values of pK as a function of temperature can be calculated from

$$pK = -\log K = \Delta G^\circ / 2.303RT, \quad (33)$$

where  $K$  stands for the equilibrium constant for the ionization reaction and  $\Delta G^\circ$  is equal to the difference between the apparent Gibbs energies of the ionized and neutral forms of

**Table 5.** Revised HKF equations of state parameters and  $C_P^\circ$  and  $V^\circ$  at 25°C and 1 bar of Gly–X–Gly tripeptides and backbone group.

Species	$\omega^a$ $\times 10^{-5}$	$c_1^b$	$c_2^c$ $\times 10^{-4}$	$\sigma^d$	$\xi^e$ $\times 10^{-2}$	$C_P^\circ$ <sup>b</sup>	$V^\circ$
Gly–Ala–Gly	–1.59	85.8	–15.8	130.7	–3.20	68.2	130.5
Gly–Arg <sup>+</sup> –Gly	–1.04	116.0	–20.0	192.0	–7.00	84.8	184.9
Gly–Asn–Gly	–1.56	89.0	–18.0	157.9	–8.52	66.6	150.1
Gly–Asp–Gly	–1.60	78.9	–13.8	147.9	–6.15	65.4	143.6
Gly–Cys–Gly	–1.62	89.0	–11.8	147.3	–5.71	79.8	143.7
Gly–Gln–Gly	–1.59	100.5	–18.8	167.3	–6.45	76.8	162.5
Gly–Glu–Gly	–1.62	94.6	–15.5	164.5	–6.39	77.9	159.9
Gly–Gly–Gly	–1.54	63.9	–16.6	112.2	–3.10	44.2	112.1
Gly–His–Gly	–1.50	151.9	–36.1	178.8	–7.69	92.2	172.0
Gly–Ile–Gly	–1.68	122.6	–6.8	188.4	–11.34	124.1	176.9
Gly–Leu–Gly	–1.68	129.4	–9.6	184.8	–9.18	125.2	176.4
Gly–Lys <sup>+</sup> –Gly	–0.56	180.0	–36.0	189.0	–9.50	112.0	177.0
Gly–Met–Gly	–1.64	105.9	–9.4	190.2	–12.43	101.7	177.0
Gly–Phe–Gly	–1.65	128.4	–11.3	204.1	–10.90	120.5	193.1
Gly–Pro–Gly	–1.63	91.5	–17.8	149.1	–5.93	70.2	145.2
Gly–Ser–Gly	–1.59	77.5	–15.7	130.3	–3.14	60.1	130.2
Gly–Thr–Gly	–1.66	99.3	–16.5	149.0	–5.45	80.9	145.8
Gly–Tyr–Gly	–1.68	126.8	–17.2	203.9	–9.58	107.2	194.9
Gly–Val–Gly	–1.65	111.8	–10.8	164.4	–6.08	104.9	160.3
[GXGBB]	–1.54	52.6	–17.6	101.2	–1.59	30.9	103.2

<sup>a</sup>: cal mol<sup>–1</sup>, <sup>b</sup>: cal mol<sup>–1</sup> K<sup>–1</sup>, <sup>c</sup>: cal K mol<sup>–1</sup>, <sup>d</sup>: cm<sup>3</sup> mol<sup>–1</sup>, <sup>e</sup>: cm<sup>3</sup> K mol<sup>–1</sup>.

the group, which can be calculated using Eq. (15) together with the group contributions given in Table 6.

#### 4 Sidechain group contributions

The standard molal thermodynamic properties and equations of state parameters of amino acids and Gly–X–Gly tripeptides can be assessed by combining additivity algorithms for sidechain groups with the properties of the amino acid and Gly–X–Gly tripeptide backbone groups. Values of  $\Xi_{[AABB]}$  and  $\Xi_{[GXGBB]}$  can be computed from Eq. (28) for amino acids, or its analog for Gly–X–Gly tripeptides, which can be written as

$$\Xi_{GXG} = \Xi_{[GXGBB]} + \Xi_{[SC]}, \quad (34)$$

where [GXGBB] represents the backbone group of Gly–X–Gly tripeptides, and GXG and [SC] represent, respectively, any Gly–X–Gly tripeptide and its sidechain group. The values of  $\Xi_{[GXGBB]}$  and  $\Xi_{[AABB]}$  were calculated using the regression plots shown in Figs. 5 and 7.

Using retrieved values of  $\Xi_{[GXGBB]}$  and  $\Xi_{[AABB]}$ , Eq. (28) can be used to calculate the sidechain group contributions of the neutral amino acids considered here except Lys and Arg. Because the pKs of their sidechain groups are within the predominance limits of [AABB<sup>–</sup>] (see Fig. 4), the properties

of [Lys] and [Arg] were retrieved from

$$\Xi_{Lys^-} = \Xi_{[AABB^-]} + \Xi_{[Lys]} \quad (35)$$

and

$$\Xi_{Arg^-} = \Xi_{[AABB^-]} + \Xi_{[Arg]}. \quad (36)$$

The thermodynamic properties of the charged sidechain groups were retrieved from charge-explicit versions of Eq. (28), which can be written as

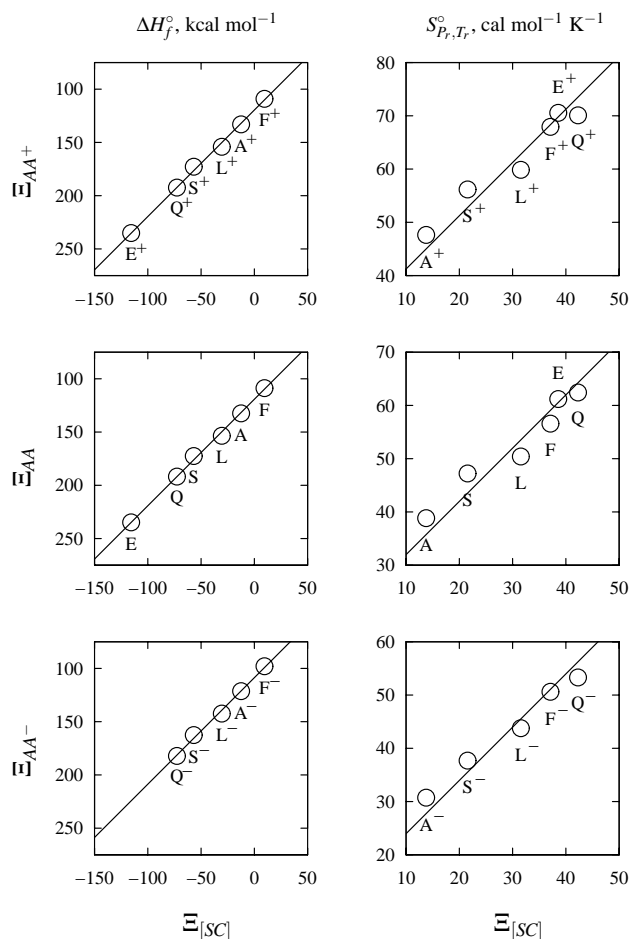
$$\Xi_{AA^-} = \Xi_{[AABB]} + \Xi_{[SC^-]} \quad (37)$$

for [Glu<sup>–</sup>], [Asp<sup>–</sup>], [Cys<sup>–</sup>], and [Tyr<sup>–</sup>], and

$$\Xi_{AA^+} = \Xi_{[AABB]} + \Xi_{[SC^+]} \quad (38)$$

for [His<sup>+</sup>], [Lys<sup>+</sup>], and [Arg<sup>+</sup>].

It follows from Eqs. (28), (37), and (38) that the intercepts of the regression lines shown in Fig. 5 correspond to the values of  $\Delta H_f^\circ$  and  $S_{P,T}^\circ$  of [AABB], [AABB<sup>+</sup>], and [AABB<sup>–</sup>]. The model sidechain groups considered for calculation of the properties of [AABB] and [GXGBB] include [Ala], [Ser], [Glu], [Gln], [Leu], and [Phe]. They contain the terminal and branched organic groups denoted by [–CH<sub>3</sub>], [–CH<sub>2</sub>OH], [–COOH], [–CONH<sub>2</sub>], [–CHCH<sub>3</sub>–], and [–C<sub>6</sub>H<sub>5</sub>] (see Amend and Helgeson (1997b)). Glu<sup>–</sup> was not used as a model compound for [AABB<sup>–</sup>] because the negative charge of Glu<sup>–</sup> is associated with its sidechain group.



**Fig. 5.**  $\Delta H_f^\circ$  and  $S_{Pr,T_r}^\circ$  of model amino acids as a function of the corresponding property of the sidechain groups. The values of  $\Xi_{[SC]}$  were calculated using Eq. (27) and the group contributions given in Table 6. The values of  $\Xi_{AA+}$ ,  $\Xi_{AA}$ , and  $\Xi_{AA-}$  were taken from Table 2. The regression lines have slopes of unity and intercepts that correspond to  $\Xi_{[AABB^+]}$ ,  $\Xi_{[AABB]}$ , or  $\Xi_{[AABB^-]}$ .

The equations of state parameters of  $[AABB^+]$  were calculated using Gly<sup>+</sup> as a reference by evaluating

$$\Xi_{[AABB^+]} = \Xi_{[AABB]} + \Xi_{\text{Gly}^+} - \Xi_{\text{Gly}}, \quad (39)$$

where  $\Xi$  stands for  $c_1$ ,  $c_2$ , or  $\omega$ . In contrast, the other equations of state parameters of  $[AABB^+]$ , and all the parameters of  $[AABB^-]$ , were calculated in a first approximation from those of Asp and Lys according to

$$\Xi_{[AABB^+]} = \Xi_{[AABB]} + \Xi_{\text{Asp}} - \Xi_{\text{Asp}^-} \quad (40)$$

and

$$\Xi_{[AABB^-]} = \Xi_{[AABB]} + \Xi_{\text{Lys}^-} - \Xi_{\text{Lys}}, \quad (41)$$

where  $\Xi$  stands for  $a_1$ ,  $a_2$ ,  $a_3$ , or  $a_4$  of  $[AABB^+]$  or  $c_1$ ,  $c_2$ ,  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$ , or  $\omega$  of  $[AABB^-]$ . The reaction between

$[AABB]$  and  $[AABB^+]$  involves ionization of the carboxylic acid group, which is why Asp and Asp<sup>-</sup> were used to calculate its parameters. A converse argument holds for the calculation of the parameters of  $[AABB^-]$  using those of Lys and Lys<sup>+</sup>.

#### 4.1 Revised HKF parameters of organic groups

It is apparent from Fig. 5 that the group contributions to  $\Delta H_f^\circ$  and  $S_{Pr,T_r}^\circ$  given by Amend and Helgeson (1997b) consistently represent the properties of the sidechain groups in amino acids. However, their group contributions do not accurately represent other properties of the sidechain groups. For example, trial values of  $C_{P,[AABB]}^\circ$  calculated using Eq. (28) and the values of  $C_P^\circ$  of Ala and Ser given in Table 6, together with those of  $[-\text{CH}_3]$  and  $[-\text{CH}_2\text{OH}]$  taken from Amend and Helgeson (1997b), are equal to  $-13.4$  and  $10.5 \text{ cal mol}^{-1} \text{ K}^{-1}$ , respectively, which differ by more than  $20 \text{ cal mol}^{-1} \text{ K}^{-1}$ . The differences between Amend and Helgeson's (2000) values for  $C_P^\circ$  of alanine and serine sidechain groups and those of other authors noted by Hakin et al. (2001) are likely a result of this uncertainty.

Amend and Helgeson (1997b) were limited in their studies by the dearth of calorimetric data for aqueous alkanes, diols and diamines at high temperatures available at the time. Owing to the availability of recent experimental data reported in the literature, these uncertainties can be reduced considerably. We take account of the recent data for diols and diamines by introducing a correction term, [di.corr], which can be used to modify Amend and Helgeson's (1997b) group additivity algorithms. This term appears in modified group additivity algorithms for model compounds which are given in Table 7. The values of  $\Delta G_f^\circ$ ,  $\Delta H_f^\circ$ , and  $S_{Pr,T_r}^\circ$  of [di.corr] were taken to be zero, because the group contributions to these properties given by Amend and Helgeson (1997b) consistently represent the sidechain groups in amino acids (see Fig. 5). We also considered the non-solvation equations of state parameters of  $[-\text{CH}_2-]$  given by Amend and Helgeson (1997b), which were derived from homologous series relationships for a number of reference model compounds, to be unaffected by the current revision.

The equations for propane and propane-1,4-diol given in Table 7 can be combined in order to obtain algorithms for modifying the group contributions according to

$$\Xi_{[-\text{CH}_3]} = \Xi_{[-\text{CH}_3]^{\text{AH97b}}} + \frac{1}{2} \Xi_{[\text{di.corr}]} \quad (42)$$

and

$$\Xi_{[-\text{CH}_2\text{OH}]} = \Xi_{[-\text{CH}_2\text{OH}]^{\text{AH97b}}} - \frac{1}{2} \Xi_{[\text{di.corr}]}, \quad (43)$$

where the superscripted AH97b refers to the group contributions given by Amend and Helgeson (1997b). Although Eq. (43) is written specifically for  $[-\text{CH}_2\text{OH}]$ , analogous statements apply to the other terminal organic

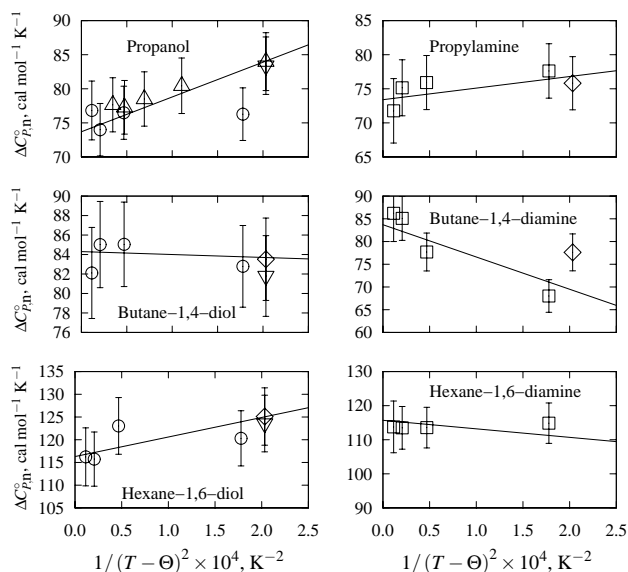
**Table 6.** Revised HKF equations of state parameters and standard molal thermodynamic properties at 25°C and 1 bar of aqueous sidechain, backbone, and organic groups.

Group	$\Delta G_f^{\circ a}$	$\Delta H_f^{\circ a}$	$S_{Pr,Tr}^{\circ b}$	$\omega^c$ $\times 10^{-5}$	$c_1^b$	$c_2^d$ $\times 10^{-4}$	$a_1^e$ $\times 10$	$a_2^c$ $\times 10^{-2}$	$a_3^f$	$a_4^d$ $\times 10^{-4}$	$C_P^{\circ b}$	$V^{\circ g}$	$\kappa_T^{\circ c}$ $\times 10^4$
Sidechain Groups													
[Ala]	-3.94	-13.29	16.85	-0.05	27.9	3.6	5.43	3.60	4.39	-1.92	35.7	26.9	5.8
[Arg]	31.62	-20.73	54.75	-0.53	73.2	-1.9	18.08	13.39	50.81	-10.09	74.2	105.8	-1.2
[Arg <sup>+</sup> ]	14.94	-35.67	60.59	0.50	82.2	1.0	22.25	2.08	2.69	-3.38	79.7	88.9	-21.6
[Asn]	-40.61	-63.49	35.90	-0.02	34.9	-1.1	10.36	4.23	0.99	-3.04	32.8	43.8	-0.5
[Asp]	-87.64	-107.13	33.24	-0.06	36.5	-1.3	9.49	4.91	-9.39	-0.46	34.4	41.1	26.9
[Asp <sup>-</sup> ]	-82.29	-105.33	21.36	2.24	34.7	-4.7	7.49	7.63	7.34	-4.60	4.7	31.2	-14.4
[Cys]	4.54	-4.86	24.34	-0.08	38.2	5.0	8.67	4.26	6.54	-3.20	49.1	39.9	-1.1
[Cys <sup>-</sup> ]	15.95	3.77	15.03	2.25	39.4	1.1	4.96	6.97	9.58	-1.94	21.1	27.0	4.9
[Gln]	-41.40	-72.65	40.43	-0.05	47.0	-0.8	13.75	5.44	2.67	-3.46	45.8	60.1	3.6
[Glu]	-88.18	-115.61	39.22	-0.08	43.6	0.6	12.83	6.09	3.78	-4.07	45.5	56.6	2.5
[Glu <sup>-</sup> ]	-82.34	-114.94	21.88	2.42	26.0	2.8	11.02	8.82	8.09	-5.25	9.6	46.4	-14.4
[Gly]	-6.07	-5.57	17.31	0.00	6.9	2.2	1.83	2.57	1.22	-1.27	11.4	9.6	4.7
[His]	36.46	11.01	44.02	0.04	60.0	-0.7	14.85	6.66	4.75	-4.28	58.2	65.8	3.1
[His <sup>+</sup> ]	28.17	3.96	48.15	0.27	66.4	-1.4	17.17	-1.54	-3.65	-1.30	61.1	63.5	-23.5
[Ile]	2.89	-32.39	27.72	-0.14	78.1	7.0	15.02	8.41	15.39	-5.99	93.6	72.2	0.5
[Leu]	0.67	-34.39	28.43	-0.14	81.1	7.3	15.21	9.37	17.16	-6.60	97.2	74.2	1.1
[Lys]	5.66	-37.65	39.30	-0.16	65.0	-0.3	15.09	10.16	21.66	-7.76	65.8	75.0	-4.1
[Lys <sup>+</sup> ]	-8.01	-51.16	39.84	0.98	73.6	2.6	19.15	0.35	1.70	-2.22	70.0	73.9	-26.6
[Met]	-35.25	-59.31	40.38	-0.10	63.7	4.0	15.48	8.76	10.82	-6.00	72.7	71.8	2.2
[Phe]	35.44	10.31	34.62	-0.11	86.5	4.1	18.80	10.66	16.57	-7.61	95.8	88.5	-0.1
[Pro]	11.31	-4.89	27.85	-0.09	42.0	1.1	9.92	6.73	9.11	-3.91	45.1	49.0	8.0
[Ser]	-39.05	-53.21	25.26	-0.05	26.9	1.2	6.22	2.59	1.10	-1.72	29.8	27.0	1.3
[Thr]	-34.96	-59.73	22.25	-0.12	43.9	3.5	9.47	4.69	6.10	-3.10	52.1	44.0	2.8
[Trp]	58.05	21.62	37.98	-0.08	95.2	2.8	24.44	11.07	12.01	-7.59	101.6	110.0	2.3
[Tyr]	-6.92	-38.53	37.43	-0.14	84.6	-1.8	20.59	10.35	5.59	-6.81	82.2	90.9	5.3
[Tyr <sup>-</sup> ]	5.50	-30.21	23.68	2.55	87.4	-5.4	16.87	12.35	13.65	-5.82	53.2	78.2	1.2
[Val]	-0.46	-27.21	23.70	-0.11	62.2	6.3	11.88	6.34	11.81	-4.35	76.0	57.3	1.9
Amino Acid and Protein Backbone Groups													
[AABB]	-84.87	-119.21	21.98	0.23	21.6	-10.6	9.47	-1.86	2.77	-1.77	-2.0	33.6	-29.3
[AABB <sup>+</sup> ]	-88.00	-119.58	31.23	0.59	42.1	-4.0	11.50	-4.58	-13.96	2.37	28.6	36.2	-13.0
[AABB <sup>-</sup> ]	-71.97	-108.69	13.97	2.83	68.7	-17.7	13.20	0.37	-18.69	0.15	6.9	37.2	-22.9
[PBB]	-21.44	-45.22	1.62	0.05	11.2	-7.5	8.10	-3.75	-6.73	1.13	-4.5	26.3	-13.7
Organic Groups													
[-CH <sub>2</sub> -]	2.24	-5.67	6.07	0.00	16.0	2.3	3.49	2.15	3.09	-1.70	20.7	16.0	-1.7
[-CH <sub>3</sub> ]	-2.19	-12.46	13.78	-0.05	31.5	1.9	5.55	3.60	3.96	-1.92	35.8	27.1	5.8
[-CHCH <sub>3</sub> -]	3.47	-12.48	11.71	-0.09	30.1	4.4	6.20	3.62	10.12	-2.98	39.9	31.2	-3.0
[-CH <sub>2</sub> NH <sub>2</sub> ]	7.04	-12.43	21.44	-0.19	25.9	-2.5	6.98	0.00	-0.74	0.00	22.6	29.3	1.8
[-CH <sub>2</sub> OH]	-41.53	-56.79	21.57	-0.05	26.2	0.9	5.86	2.59	3.03	-1.72	28.5	26.7	1.3
[-CONH <sub>2</sub> ]	-46.55	-61.42	30.14	-0.05	14.5	-4.9	7.02	1.14	-5.03	-0.06	5.0	28.2	7.0
[-COOH]	-92.57	-104.38	26.41	-0.08	12.9	-4.1	5.82	1.79	-2.40	-0.67	5.3	24.5	5.9
[-C <sub>6</sub> H <sub>5</sub> ]	31.66	15.17	31.02	-0.11	70.4	1.8	15.33	8.51	13.47	-5.91	75.1	72.6	1.6
[di.corr]	0.00	0.00	0.00	-0.02	-0.1	-6.7	0.06	-1.48	-0.05	2.32	-13.5	3.2	11.5

a: kcal mol<sup>-1</sup>, b: cal mol<sup>-1</sup> K<sup>-1</sup>, c: cal mol<sup>-1</sup>, d: cal mol<sup>-1</sup> K, e: cal mol<sup>-1</sup> bar<sup>-1</sup>, f: cal K mol<sup>-1</sup> bar<sup>-1</sup>, g: cm<sup>3</sup> mol<sup>-1</sup>, h: cm<sup>3</sup> mol<sup>-1</sup> bar<sup>-1</sup>.

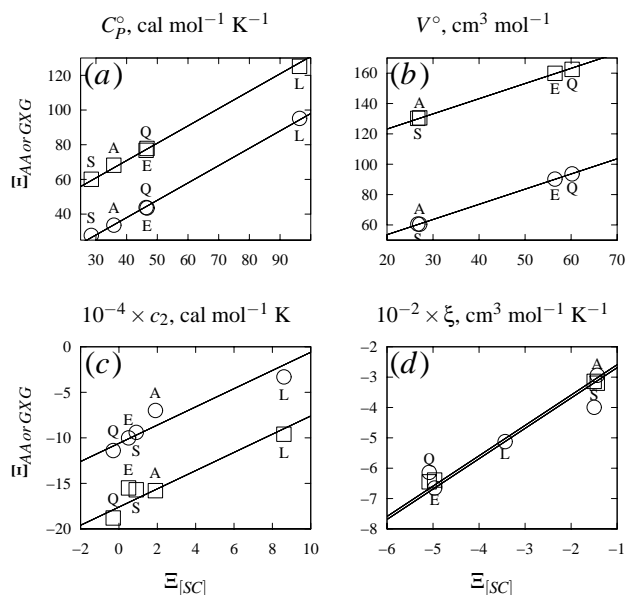
**Table 7.** Group additivity algorithms for the alcohols and amines shown in Fig. 6. [di.corr] represents a correction term used in the present study to modify the group additivity algorithms used by Amend and Helgeson (1997b).

Species	Amend and Helgeson (1997b)	Present Study
propanol	$[-\text{CH}_3] + [-\text{CH}_2-] + [-\text{CH}_2\text{OH}]$	$[-\text{CH}_3] + [-\text{CH}_2-] + [-\text{CH}_2\text{OH}]$
butane-1,4-diol	$2[-\text{CH}_2-] + 2[-\text{CH}_2\text{OH}]$	$2[-\text{CH}_2-] + 2[-\text{CH}_2\text{OH}] + [\text{di.corr}]$
butane-1,6-diol	$4[-\text{CH}_2-] + 2[-\text{CH}_2\text{OH}]$	$4[-\text{CH}_2-] + 2[-\text{CH}_2\text{OH}] + [\text{di.corr}]$
propylamine	$[-\text{CH}_3] + [-\text{CH}_2-] + [-\text{CH}_2\text{NH}_2]$	$[-\text{CH}_3] + [-\text{CH}_2-] + [-\text{CH}_2\text{NH}_2]$
butane-1,4-diamine	$2[-\text{CH}_2-] + 2[-\text{CH}_2\text{NH}_2]$	$2[-\text{CH}_2-] + 2[-\text{CH}_2\text{NH}_2] + [\text{di.corr}]$
hexane-1,6-diamine	$4[-\text{CH}_2-] + 2[-\text{CH}_2\text{NH}_2]$	$4[-\text{CH}_2-] + 2[-\text{CH}_2\text{NH}_2] + [\text{di.corr}]$

**Fig. 6.**  $\Delta C_{P,n}^\circ$  as a function of  $1/(T-\Theta)^2$  of aqueous propanol, butane-1,4-diol, hexane-1,6-diol, propylamine, butane-1,4-diamine, and hexane-1,6-diamine. Experimental values of  $C_P^\circ$  were taken from Inglesse and Wood, 1996 (○), Inglesse et al., 1997 (□), Makhatazde and Privalov, 1989 (△), Jolicoeur and Lacroix, 1976 (▽), Nichols et al., 1976 (◇; alcohols), and Cabani et al., 1981 (◇; amines). The error bars represent an estimated experimental uncertainty of  $\pm 5\%$ . The values of  $c_1$  and  $c_2$ , which correspond to the intercepts and slopes of the regression lines, were calculated using the equations given in Table 7 together with the organic group contributions given in Table 6.

groups, represented by  $[-\text{COOH}]$ ,  $[-\text{CONH}_2]$ ,  $[-\text{C}_6\text{H}_5]$ , and  $[-\text{CH}_2\text{NH}_2]$ . The constraints represented by Eqs. (42) and (43) ensure that the revised group contributions are consistent with the properties of the compounds from which they were originally derived.

Experimental values of  $C_P^\circ$  at 280 bar and temperatures to  $\sim 250^\circ\text{C}$  of propanol, butane-1,4-diol, hexane-1,6-diol, propylamine, butane-1,4-diamine and hexane-1,6-diamine have been reported by Inglesse and Wood (1996) and Inglesse

**Fig. 7.**  $C_P^\circ$ ,  $V^\circ$ ,  $c_2$ , and  $\xi$  of model amino acids (□) and Gly-X-Gly tripeptides (○) as a function of the corresponding property of the model sidechain groups. The values of  $\Xi_{[SC]}$  were calculated using Eq. (27) and the group contributions given in Table 6. The values of  $\Xi_{AA}$  and  $\Xi_{GXG}$  were taken from Table 2. The regression lines have slopes of unity and intercepts that correspond to  $\Xi_{[AABB]}$  or  $\Xi_{[GXGBB]}$ .

et al. (1997). These data were used in a simultaneous regression analysis taking account of Eqs. (42) and (43) to calculate values of  $\omega$ ,  $c_1$ , and  $c_2$  of [di.corr], as well as updated values of  $\omega$  of  $[-\text{CH}_3]$ ,  $[-\text{CH}_2-]$ ,  $[-\text{CH}_2\text{OH}]$ ,  $[-\text{CH}_2\text{NH}_2]$ , and the other terminal organic groups. The corresponding regression lines are depicted in Fig. 6, where it can be seen that the revised group contributions are consistent within experimental uncertainty with the high-temperature data for the mono- and di- alcohols and amines.

The values of  $C_P^\circ$  and  $c_2$  of the terminal organic groups computed in this manner closely represent those of the model sidechain groups shown in Figs. 7a and b. It can be seen in Fig. 7a, for example, that the differences in the calculation of

$C_p^\circ$  of [AABB] and [GXGBB] from different model amino acids is of the order of  $\pm 1 \text{ cal mol}^{-1} \text{ K}^{-1}$ , which is much smaller than the uncertainty associated with the original group contributions (see above).

Values of  $V^\circ$  and  $\xi$  of these backbone groups are depicted in Figs. 7b and d. The revised group contributions to these properties were calculated using

$$\Xi_{[\text{AABB}]} = \Xi_{\text{Ala}} - (\Xi_{[-\text{CH}_3]^{\text{AH97b}}} + \frac{1}{2} \Xi_{[\text{di.corr}]}) \quad (44)$$

and

$$\Xi_{[\text{AABB}]} = \Xi_{\text{Ser}} - (\Xi_{[-\text{CH}_2\text{OH}]^{\text{AH97b}}} - \frac{1}{2} \Xi_{[\text{di.corr}]}) \quad (45)$$

which represent statements of Eq. (28) combined with Eqs. (42) and (43). Equations (44) and (45) were also used to calculate values of  $a_2$  and  $a_4$  of [di.corr] and [AABB] using the values of  $a_2$  and  $a_4$  of Ala and Ser given in Table 2. The values of  $V^\circ$  and  $\xi$  of the other organic groups were then computed from those of the corresponding amino acids together with those of [AABB] using Eq. (28). Finally, the revised organic group contributions to  $a_1$  and  $a_3$  were computed from those of  $V^\circ$ ,  $\xi$ ,  $a_2$ , and  $a_4$  using Eqs. (7) and (20).

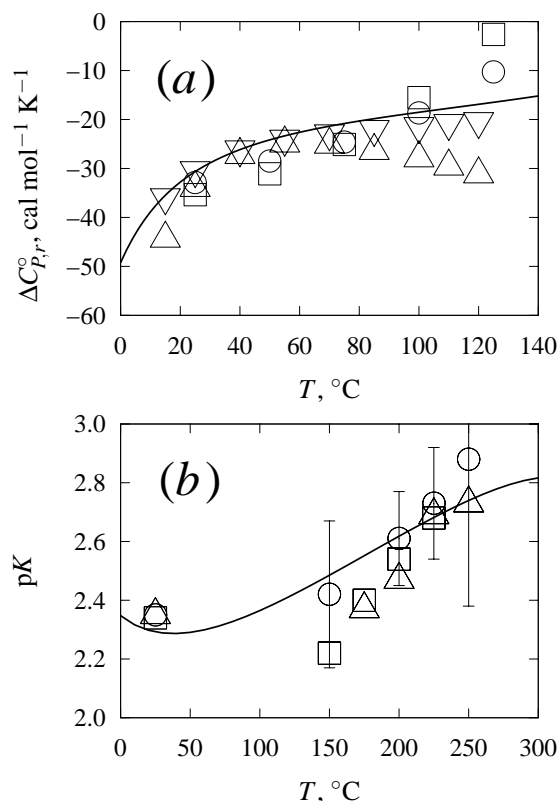
#### 4.2 $C_p^\circ$ and pK of amino acid backbone group ionization at high temperature

The ionization reaction of the carboxylic acid group of Gly, Ala, Pro, and Val can be represented as



This reaction corresponds to the low pH boundary of the predominance field of [AABB] shown in Fig. 4. Experimental values of  $\Delta C_{P,(R1)}^\circ$  from  $\sim 15$  to  $120^\circ\text{C}$  taken from Wang et al. (1996), Sorenson et al. (2003), and Price et al. (2003b) are shown in Fig. 8a, along with values calculated from the revised HKF equations of state parameters of [AABB] and [AABB<sup>+</sup>] given in Table 6. At low temperatures, the experimental values of  $\Delta C_{P,(R1)}^\circ$  of Ala, Pro, and Val are generally consistent with one another. However, at temperatures  $> 80^\circ\text{C}$ , the experimental values diverge from one another by up to  $\sim 20 \text{ cal mol}^{-1} \text{ K}^{-1}$ . The calculated values approximately split equally the differences between the experimental values. The large differences in  $\Delta C_{P,(R1)}^\circ$  at high temperatures may be attributed to interactions with specific sidechain groups, or perhaps to a high uncertainty in the measurements owing to the degradation of amino acids under the experimental conditions (Amend and Helgeson, 1997b; Andersson and Holm, 2000; Li et al., 2002).

Experimental and calculated values of  $\text{pK}_{(R1)}$  from  $\sim 25$  to  $250^\circ\text{C}$  are shown in Fig. 8b. The measured values for Gly, Ala, and Pro at  $25^\circ\text{C}$  and at higher temperatures do not differ by more than the stated experimental uncertainties. The calculated values of  $\text{pK}_{(R1)}$  at high temperatures most closely follow the trend for Gly, from which the parameters of [AABB<sup>+</sup>] were retrieved.



**Fig. 8.** Comparison of experimental and calculated values of  $\Delta C_{P,r}^\circ$  (a) and  $\text{pK}$  (b) of the ionization of the carboxylic acid group in [AABB] (Reaction (R1)) as a function of temperature. Experimental values shown in (a) are taken from Wang et al. (1996) for Gly ( $\circ$ ) and Ala ( $\square$ ), from Sorenson et al. (2003) for Pro ( $\triangle$ ), and from Price et al. (2003b) for Val ( $\nabla$ ). In (b), the same symbols are used to designate the amino acids, but the experimental values are taken from Amend and Helgeson (1997a) at  $25^\circ\text{C}$  and from Clarke et al. (2005) at higher temperatures. The estimated experimental errors for Gly are represented by the error bars; the error bars for Ala and Pro are omitted for clarity.

## 5 Protein backbone group contributions

The backbone group contributions to the standard molal thermodynamic properties and equations of state parameters of unfolded proteins are represented below by  $\Xi_{[\text{PBB}]}$ . The calculation of  $\Xi_{[\text{PBB}]}$  takes account of a reference state corresponding to a hypothetical non-ionized unfolded protein, represented by  $UP^0$ , the properties of which are given by the group additivity scheme represented by (Amend and Helgeson, 2000)

$$\Xi_{UP^0} = \Xi_{[\text{AABB}]} + (n-1) \Xi_{[\text{PBB}]} + \sum_{i=1}^{\hat{i}} n_{[\text{SC}]_i} \Xi_{[\text{SC}]_i} \quad (47)$$

where [PBB] and  $[\text{SC}]_i$  represent the protein backbone group and the  $i$ th sidechain group, respectively,  $n$  stands for the



**Table 8.** Model proteins considered in the present study. Masses and formulas are for hypothetical non-ionized, disulfide-free and metal-free proteins.

Name <sup>a</sup>	Accession No.	Protein	Organism	$n_{[SC]}$	$n_{[-SS-]}^b$	Mass <sup>c</sup>	Formula
AMY_PYRFU	O08452	Alpha amylase (extracellular)	<i>Pyrococcus furiosus</i>	434	2	50 113	C <sub>2338</sub> H <sub>3307</sub> N <sub>573</sub> O <sub>649</sub> S <sub>9</sub>
AMY_BACSU	P00691	Alpha amylase (extracellular)	<i>Bacillus subtilis</i>	619	0	68 423	C <sub>3006</sub> H <sub>4584</sub> N <sub>850</sub> O <sub>965</sub> S <sub>11</sub>
AMYA_PYRFU	P49067	Alpha amylase (intracellular)	<i>Pyrococcus furiosus</i>	648	0	76 178	C <sub>3526</sub> H <sub>5325</sub> N <sub>885</sub> O <sub>976</sub> S <sub>14</sub>
BPT1_BOVIN	P00974	Pancreatic trypsin inhibitor	<i>Bos taurus</i>	58	3	6518	C <sub>284</sub> H <sub>438</sub> N <sub>84</sub> O <sub>79</sub> S <sub>7</sub>
CYC_BOVIN	P62894	Cytochrome C	<i>Bos taurus</i>	104	0	11 572	C <sub>517</sub> H <sub>825</sub> N <sub>143</sub> O <sub>150</sub> S <sub>4</sub>
LYSC_CHICK	P00698	Lysozyme C	<i>Gallus gallus</i>	129	4	14 313	C <sub>613</sub> H <sub>959</sub> N <sub>193</sub> O <sub>185</sub> S <sub>10</sub>
MYG_PHYCA	P02185	Myoglobin	<i>Physeter catodon</i>	153	0	17 200	C <sub>783</sub> H <sub>1240</sub> N <sub>216</sub> O <sub>216</sub> S <sub>2</sub>
NUC_STAAU	P00644	Staphylococcal nuclease	<i>Staphylococcus aureus</i>	149	0	16 811	C <sub>747</sub> H <sub>1204</sub> N <sub>208</sub> O <sub>224</sub> S <sub>4</sub>
RNAS1_BOVIN	P61823	Pancreatic ribonuclease	<i>Bos taurus</i>	124	4	13 690	C <sub>575</sub> H <sub>909</sub> N <sub>171</sub> O <sub>193</sub> S <sub>12</sub>
RNBR_BACAM	P00648	Ribonuclease	<i>Bacillus amyloliquefaciens</i>	110	0	12 383	C <sub>555</sub> H <sub>847</sub> N <sub>153</sub> O <sub>170</sub>
RNH_ECOLI	P0A7Y4	Ribonuclease H	<i>Escherichia coli</i>	155	0	17 597	C <sub>776</sub> H <sub>1215</sub> N <sub>227</sub> O <sub>228</sub> S <sub>7</sub>
RNH_THET8	P29253	Ribonuclease H	<i>Thermus thermophilus</i>	166	0	18 728	C <sub>829</sub> H <sub>1297</sub> N <sub>253</sub> O <sub>231</sub> S <sub>7</sub>
RNT1_ASPOR	P00651	Ribonuclease T1	<i>Aspergillus oryzae</i>	104	2	11 089	C <sub>479</sub> H <sub>687</sub> N <sub>127</sub> O <sub>171</sub> S <sub>4</sub>

<sup>a</sup>: The names shown for the proteins are those used in the Swiss-Prot database (Boeckmann et al., 2003), except for AMY\_PYRFU, which is the name adopted in the present study. The sequence of O08452\_PYRFU was taken from the TrEMBL database (Boeckmann et al., 2003), and was combined with the sequence of the signal peptide given by Dong et al. (1997) to calculate the composition of secreted AMY\_PYRFU. Signal peptides and propeptides, where annotated in Swiss-Prot, were excluded from the calculations. <sup>b</sup>: The numbers of disulfide bonds are taken from Swiss-Prot, except for those of AMY\_BACSU and AMY\_PYRFU (Fitter, 2005). <sup>c</sup>: g mol<sup>-1</sup>.

number of residues in, or length of, the protein,  $n_{[SC]_i}$  designates the number of the  $i$ th sidechain groups in the protein, and  $i=1, 2, \dots, \hat{i}$ , where  $\hat{i}$  denotes the number of different kinds of sidechain groups in the protein. [AABB] in Eq. (47) is used to represent the terminal groups of the protein molecule. Although it is written for unfolded proteins, analogous statements of Eq. (47) for polypeptides are used to generate values of  $\Xi_{[PBB]}$  where experimental data for proteins are lacking.

### 5.1 $\Delta G_f^\circ$ , $\Delta H_f^\circ$ and $S_{P_r, T_r}^\circ$ of [PBB]

Experimental values of  $\Delta G_f^\circ$  of proteins are not commonly available in the literature. As a provisional approximation, they can be calculated using values of  $\Delta H_f^\circ$  and  $S_{P_r, T_r}^\circ$  of model dipeptides. This approach was taken by Amend and Helgeson (2000), who took into account the properties of diketopiperazine given by Shock (1992) together with the group additivity algorithm represented by

$$\Xi_{[PBB]} = \Xi_{\text{diketopiperazine}/2} - \Xi_{[\text{Gly}]} . \quad (48)$$

Because it is a circular dipeptide, diketopiperazine may not be a desirable model compound for the backbone group in unfolded proteins, which more closely resemble extended polypeptide chains. Therefore, linear dipeptides for which  $\Delta H_f^\circ$  and  $S_{P_r, T_r}^\circ$  are known were also included in the present study by considering

$$\Xi_{[PBB]} = \Xi_{[AABB]} + \Xi_{\text{GlyGly}} - (\Xi_{\text{Gly}} + \Xi_{\text{Gly}}) , \quad (49)$$

$$\Xi_{[PBB]} = \Xi_{[AABB]} + \Xi_{\text{AlaGly}} - (\Xi_{\text{Ala}} + \Xi_{\text{Gly}}) , \quad (50)$$

and

$$\Xi_{[PBB]} = \Xi_{[AABB]} + \Xi_{\text{LeuGly}} - (\Xi_{\text{Leu}} + \Xi_{\text{Gly}}) . \quad (51)$$

Using Shock's (1992) values for the dipeptides, we calculated values of  $\Delta H_{f, [PBB]}^\circ$  and  $S_{P_r, T_r, [PBB]}^\circ$  from each of Eqs. (48)–(51). The averages of the results are given in Table 6, as are the corresponding value of  $\Delta G_{f, [PBB]}^\circ$  calculated from Eq. (4). It should be noted that the value of  $\Delta H_{f, [PBB]}^\circ$  calculated from Eq. (48) ( $-44.08$  kcal mol<sup>-1</sup>) differs from our adopted value by only  $\sim 1$  kcal mol<sup>-1</sup>. However, the value of  $S_{P_r, T_r, [PBB]}^\circ$  calculated from Eq. (48) ( $9.44$  cal mol<sup>-1</sup> K<sup>-1</sup>) is different by  $\sim 8$  cal mol<sup>-1</sup> K<sup>-1</sup> from our adopted value. Hence, the value of the standard molal Gibbs energy of [PBB] may be subject to considerable uncertainty.

### 5.2 $C_p^\circ$ and $V^\circ$ of [PBB]

Values of  $C_p^\circ$  and  $V^\circ$  of [PBB] as a function of temperature were calculated from Eq. (47) and experimental values for the proteins and polypeptides. A value of  $\omega$  of [PBB] was chosen that linearized the trend of data points in the regression plots shown in Fig. 9. The values of  $\Delta C_{p, n}^\circ$  represented in this figure generally are consistent with the trend of the regression lines despite variations in experimental solution pH for the proteins from 2 to 6 (Privalov et al., 1989). Hence, it appears that the denaturing buffer used in the experiments may have had a compensating effect on the ionization properties of the proteins. Consequently, the symbols shown in Figs. 9b and d can be regarded in a first approximation as representative of non-ionized unfolded proteins.

There is a greater scatter in the values of  $\Delta V_n^\circ$  of unfolded proteins shown in Fig. 9d, but this does not appear to be associated with a systematic difference between  $\Delta V_n^\circ$  of the proteins and the polypeptides. In contrast, there is an almost constant difference between  $\Delta C_{p,n}^\circ$  of the proteins and polypeptides. The intercepts of the correlation lines shown in Figs. 9a and b, representing values of  $c_1$  of the backbone group in unfolded proteins and polypeptides, can be expressed as

$$c_{1,[PPBB]} = c_{1,[PBB]} + 10, \quad (52)$$

where [PPBB] denotes the backbone group contributions in polypeptides; the units of  $c_1$  are  $\text{cal mol}^{-1} \text{K}^{-1}$ . Because of limited independent experimental data for the other thermodynamic properties of the two types of molecules, this represents the only instance where we are presently able to resolve differences in the backbone group contributions of polypeptides and unfolded proteins.

### 5.3 $\kappa_T^\circ$ , $a_2$ and $a_4$ of [PBB]

The isothermal compressibility ( $\kappa_T^\circ$ ) at 25°C and 1 bar of [PBB] can be calculated from the measured compressibilities of poly-*d,l*-alanine and poly-glutamic acid, together with those of  $\kappa_T^\circ$  of [Ala] and [Glu] using the relations

$$\kappa_{T,[PBB]}^\circ = \kappa_{T,\text{poly-}d,l\text{-alanine}}^\circ - \kappa_{T,[Ala]}^\circ \quad (53)$$

and

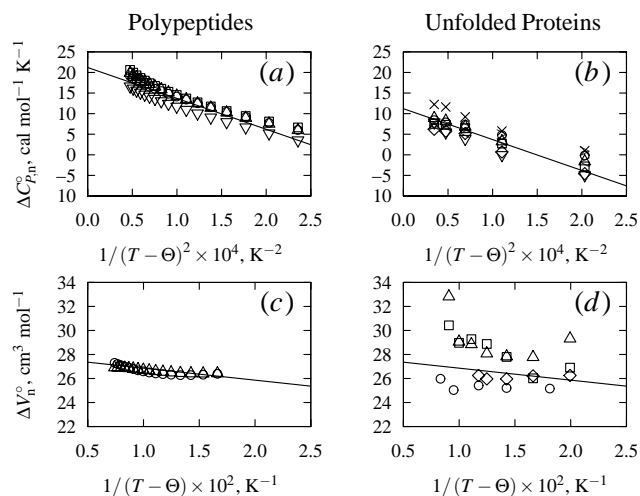
$$\kappa_{T,[PBB]}^\circ = \kappa_{T,\text{poly-glutamic acid}}^\circ - \kappa_{T,[Glu]}^\circ, \quad (54)$$

which yield values of  $\kappa_T^\circ$  of [PBB] of  $-13.40 \times 10^{-4}$  and  $-13.98 \times 10^{-4} \text{ cm}^3 \text{ bar}^{-1} \text{ mol}^{-1}$ , respectively. The isothermal compressibilities of the polypeptides were calculated from Eq. (30) using the values of  $\kappa_\zeta^\circ$  measured by Kharakoz (1997). The values of  $C_p^\circ$ ,  $V^\circ$  and  $E^\circ$  of the polypeptides in Eq. (30) were calculated using the revised HKF equation of state and the properties and parameters of the sidechain and backbone groups taken from Table 6, together with the values of  $\sigma$  and  $\xi$  of [PBB] retrieved above.

The group additivity algorithm used by Amend and Helgeson (2000) to calculate values of  $a_2$  and  $a_4$  of [PBB] is given by

$$\Xi_{[PBB]} = \Xi_{[AABB]} + 2\Xi_{[-CH_2]} - \Xi_{[-CH_2OH]} - \Xi_{[-CH_3]}, \quad (55)$$

where  $\Xi$  represents  $a_2$  or  $a_4$  of the subscripted group. Using this equation and the group contributions given in Table 6 results in  $a_{2,[PBB]} \times 10^{-2} = -3.75 \text{ cal mol}^{-1}$  and  $a_{4,[PBB]} \times 10^{-4} = -1.53 \text{ cal mol}^{-1} \text{ K}$ . These values, when taken together with the value of  $\omega_{[PBB]}$  given in Table 6, are consistent with  $\kappa_{T,[PBB]}^\circ = -37.15 \times 10^{-4} \text{ cm}^3 \text{ bar}^{-1} \text{ mol}^{-1}$ . This value is considerably lower than that computed above for the polypeptides. Accordingly, the value of  $a_2$  of [PBB] was calculated using Eq. (55), but that of  $a_4$  of [PBB] was chosen to give a value  $\kappa_T^\circ$  at 25°C and 1 bar that matches the average of the results of Eqs. (53) and (54).



**Fig. 9.**  $\Delta C_{p,n}^\circ$  as a function of  $1/(T-\Theta)^2$  (a) and (b) and  $\Delta V_n^\circ$  as a function of  $1/(T-\Theta)$  (c) and (d) of the protein backbone group ([PBB]). The intercepts and slopes of the regression lines correspond to the values of  $c_1$ ,  $c_2$ ,  $\sigma$  and  $\xi$  of [PBB] given in Table 6, except for (a), where the intercept of the regression line corresponds to the value of  $c_1$  of the polypeptide backbone group ([PPBB]) that is given by Eq. (52). Values are shown for the proteins CYC\_BOVIN ( $\circ$ ), MYG\_PHYCA ( $\diamond$ ), LYSC\_CHICK ( $\square$ ), RNAS1\_BOVIN ( $\triangle$ ), BPT1\_BOVIN ( $>$ ), IL1B\_HUMAN ( $<$ ), RNBR\_BACAM ( $\nabla$ ), and RNT1\_ASPOR ( $\times$ ) and for the polypeptides GGLGG ( $\circ$ ), GGSAG ( $\triangle$ ), GSGG ( $\square$ ), and GFVGG ( $\nabla$ ). The values of  $C_p^\circ$  of the unfolded proteins used in the calculations were taken from Privalov and Makhatadze (1990) ( $\circ$ ,  $\diamond$ ,  $\square$ ,  $\triangle$ ), Makhatadze et al. (1993) ( $>$ ), Yu et al. (1994) ( $\times$ ), Makhatadze et al. (1994) ( $<$ ) and Griko et al. (1994) ( $\nabla$ ); those for the polypeptides were taken from Häckel et al. (2000a). The values of  $V^\circ$  of the unfolded proteins and polypeptides were taken from Makhatadze et al. (1990) and Häckel et al. (2000b), respectively.

## 6 Representative uncertainties

The representative uncertainties of amino acids and [PBB] summarized below can be used to assess the reliability of biogeochemical calculations that depend on values of  $\Delta G^\circ$  as a function of temperature and pressure. These uncertainties exceed by far those arising from the number of significant figures used in the present study. The uncertainties in the additive properties of proteins can be estimated by multiplying the uncertainty in the corresponding property of [PBB] by the length (number of backbone groups) of the protein.

Benson and Buss' (1958) estimates of the uncertainty in the additive prediction of  $\Delta H_f^\circ$  and  $S_{p_r,T_r}^\circ$  of gaseous species are  $\pm 0.60 \text{ kcal mol}^{-1}$  and  $\pm 0.50 \text{ cal mol}^{-1} \text{ K}^{-1}$ , respectively. We adopt similar values for the estimated uncertainties in the values of  $\Delta G_f^\circ$ ,  $\Delta H_f^\circ$ , and  $S_{p_r,T_r}^\circ$  of the aqueous amino acids, which are, respectively,  $\pm 0.50 \text{ kcal mol}^{-1}$ ,  $\pm 0.50 \text{ kcal mol}^{-1}$ , and  $\pm 0.50 \text{ cal mol}^{-1} \text{ K}^{-1}$ . These values were taken from Amend and Helgeson (1997a), who assigned them on the basis of experimental error apparent in

**Table 9.** Estimated uncertainties in the standard molal thermodynamic properties of amino acids and the protein backbone group as a function of temperature and pressure.  $P_{SAT}$  refers to the saturation vapor pressure of  $H_2O$  which corresponds to 1.013 bar at 100°C and 15.536 bar at 200°C.

Property	100°C, $P_{SAT}$		200°C, $P_{SAT}$	
	AA	[PBB]	AA	[PBB]
$\Delta G^{\circ a}$	$\pm 0.56$	$\pm 1.10$	$\pm 0.70$	$\pm 1.28$
$\Delta H^{\circ a}$	$\pm 0.69$	$\pm 1.19$	$\pm 0.94$	$\pm 1.44$
$S^{\circ b}$	$\pm 1.06$	$\pm 1.56$	$\pm 1.65$	$\pm 2.15$

Property	200°C, 5000 bar		300°C, 5000 bar	
	AA	[PBB]	AA	[PBB]
$\Delta G^{\circ a}$	$\pm 0.89$	$\pm 1.48$	$\pm 1.08$	$\pm 1.72$
$\Delta H^{\circ a}$	$\pm 1.13$	$\pm 1.64$	$\pm 1.38$	$\pm 1.89$
$S^{\circ b}$	$\pm 1.65$	$\pm 2.15$	$\pm 2.13$	$\pm 2.63$

a: kcal mol<sup>-1</sup>. b: cal mol<sup>-1</sup> K<sup>-1</sup>.

the measurements of the solubilities and enthalpies of solution. Owing to the availability of experimental data for only a few short peptides (see above), the values of  $\Delta G_f^\circ$ ,  $\Delta H_f^\circ$ , and  $S_{P_r, T_r}^\circ$  of [PBB] are more uncertain; we assign them estimated values of  $\pm 1.00$  kcal mol<sup>-1</sup>,  $\pm 1.00$  kcal mol<sup>-1</sup>, and  $\pm 1.00$  cal mol<sup>-1</sup> K<sup>-1</sup>, respectively.

Judging from the scatter of the data points shown in the regression plots in Fig. 2, representative uncertainties for amino acids and Gly-X-Gly tripeptides were estimated as  $\pm 2.5$  cal mol<sup>-1</sup> K<sup>-1</sup> for  $C_p^\circ$ ,  $\pm 2$  cm<sup>3</sup> mol<sup>-1</sup> for  $V^\circ$ , and  $\pm 3.0 \times 10^{-4}$  cm<sup>3</sup> bar<sup>-1</sup> mol<sup>-1</sup> for  $\kappa_T^\circ$ . We provisionally assign these values also to the uncertainties in the properties of [PBB]. These values can be used to generate estimates of the temperature and pressure dependence of the uncertainties in  $\Delta G^\circ$ ,  $\Delta H^\circ$ , and  $S^\circ$  using the absolute finite difference counterpart of the corresponding revised HKF equation of state. For example, Eq. (15) can be used to write

$$\begin{aligned} \delta \Delta G^\circ = & \left| \delta \Delta G_f^\circ \right| + \left| \delta S_{P_r, T_r}^\circ (T - T_r) \right| \\ & + \left| \delta c_1 \left[ T \ln \left( \frac{T}{T_r} \right) - T + T_r \right] \right| \\ & + \left| \delta a_1 (P - P_r) \right| + \left| \delta a_2 \ln \left( \frac{\Psi + P}{\Psi + P_r} \right) \right|, \quad (56) \end{aligned}$$

where the Greek letter  $\delta$  denotes uncertainty. Corresponding equations for  $\delta \Delta H^\circ$  and  $\delta S^\circ$  can be written with the aid of Eqs. (16) and (17). Because we take the pressure and temperature dependence of the uncertainties in  $C_p^\circ$ ,  $V^\circ$ , and  $\kappa_T^\circ$  to be zero, the values of  $\delta c_1$ ,  $\delta a_1$ , and  $\delta a_2$  used in Eq. (54) were calculated by taking account of the first term only in the right-hand side of Eqs. (18), (21), and (24). Uncertainties in  $\Delta G^\circ$ ,  $\Delta H^\circ$  and  $S^\circ$  as a function of temperature and pressure calculated using this method are given in Table 9. Up

to  $\sim 300^\circ\text{C}$  and  $\sim 5000$  bar, the estimated uncertainties in the calculation of  $\Delta G^\circ$  of amino acids and sidechain and protein backbone groups increase only moderately with temperature and pressure. Hence, it appears that the largest source of error in the calculation of  $\Delta G^\circ$  over this temperature range is not uncertainty in the revised HKF equations of state parameters, but that inherent in  $\Delta G_f^\circ$  at 25°C and 1 bar.

It should be noted that these representative uncertainties, as such, do not account for large deviations at high temperature apparent in the volumes of Ala and some of the model proteins. However, the representative uncertainty in  $V^\circ$  is comparable to that observed in other group additivity studies (Cabani et al., 1981; Lepori and Gianni, 2000; Hnědkovský and Cibulka, 2004). The estimated uncertainties in  $\Delta G^\circ$  of [PBB] are considerably greater than  $\pm 0.1$  kcal mol<sup>-1</sup>, which is the value recommended by Dill (1997) for group additivity calculations of the energetics of conformational interactions of proteins such as unfolding and non-covalent structural changes. Nevertheless, the current method affords an approximation of the standard molal thermodynamic properties of proteins of different amino acid composition, which can be used a frame of reference for subsequent calculations of the properties of folded proteins and their complexes.

## 7 Calculations of protein speciation

The group contributions given in Table 6 can be used to calculate the standard molal thermodynamic properties of unfolded proteins in specific ionization states. These properties are useful for calculations of the environmental constraints on the speciation of proteins. Examples of these types of calculations are given below for the ionization of proteins and the evolution of proteins in organisms adapted to different environments. Although they proceed on quite different time scales, both of these processes may be constrained by geochemical conditions such as temperature and electrochemistry.

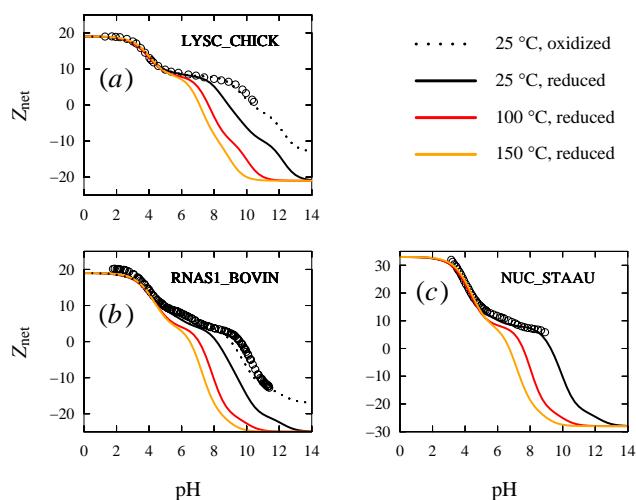
The two electrochemical variables considered here, pH and pe, are equal to the negative logarithms of the activities of the proton and electron, respectively. Together, they represent the electrochemical or oxidation-reduction properties of an aqueous solution. Many geochemical and biochemical studies use values of Eh, which are related to pe by (Drever, 1997)

$$pe = \frac{F}{2.303RT} Eh, \quad (57)$$

where  $F$  and  $R$  denote the Faraday constant and the gas constant. At 25°C, Eq. (57) gives Eh (volts) = 0.05916pe.

### 7.1 Ionization of proteins as a function of temperature and pH

The net charge of a protein,  $Z_{net}$ , arises from complexation with charged species, including metal ions, charged organic



**Fig. 10.** Calculated and experimental values of the net charge ( $Z_{\text{net}}$ ) of unfolded LYSC\_CHICK (a), RNAS1\_BOVIN (b), and NUC\_STAAU (c) as a function of pH and temperature. Open circles represent values of  $Z_{\text{net}}$  reported from titration experiments in 6.0 M GuHCl at 25°C (Nozaki and Tanford, 1967; Roxby and Tanford, 1971; Whitten and García-Moreno E., 2000), and solid lines represent values of  $Z_{\text{net}}$  and  $\Delta G^\circ$  calculated using Eqs. (59) and (58) at 25, 100 and 150°C. Dotted lines represent calculations for oxidizing conditions favoring formation of disulfide bonds. In these cases, the corresponding [Cys] groups are not available for proton ionization.

species, and the proton. The contributions by proton ionization to  $Z_{\text{net}}$  can be expressed as the sum of the charges of the ionizable sidechain and backbone groups in accord with

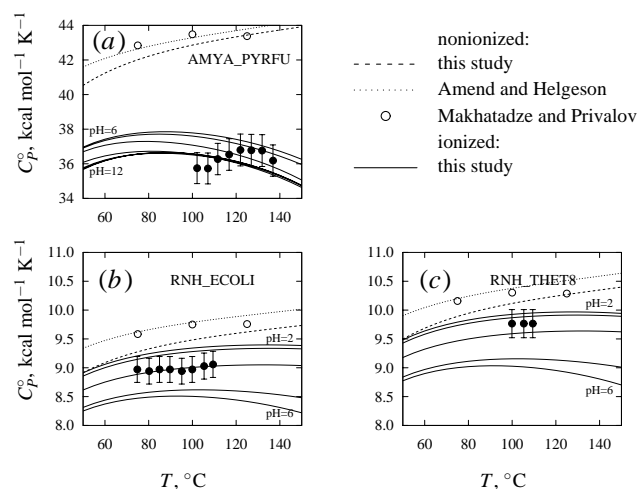
$$Z_{\text{net}} = \sum_i n_i \alpha_i Z_i, \quad (58)$$

where  $n_i$  stands for the number of the  $i$ th ionizable group in the protein,  $Z_i$  denotes the formal charge of the ionized form of that group, and  $\alpha_i$  stands for the degree of formation of the ionized form of that group (i.e., the molality of the ionized form of the group divided by the total molality of the group). Eq. (58) provides a frame of reference for calculating titration curves as a function of temperature. If we take activity coefficients to be unity, it can be shown that

$$\alpha_i = \frac{1}{1 + 10^{Z_i(\text{pH} - \text{p}K_i)}}, \quad (59)$$

where  $\text{p}K_i$  is equal to the negative logarithm of the equilibrium constant of the deprotonation reaction for the  $i$ th group.

Experimental and calculated titration curves as a function of temperature are represented in Fig. 10. Protein charge as a function of pH was first computed using Eq. (58), taking account of seven ionizable sidechain groups and the terminal groups of the protein. The trend of decreasing protein net charge with increasing temperature at high pHs apparent in Fig. 10 can also be noted in titrations of crystalline

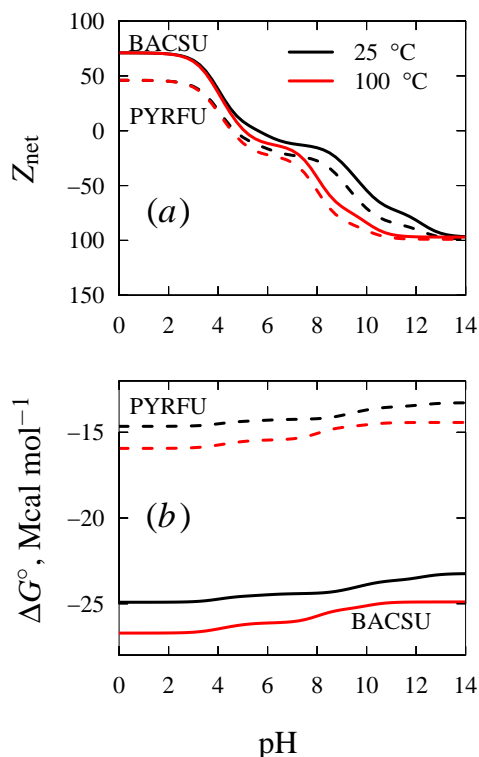


**Fig. 11.**  $C_p^\circ$  of heat-denatured proteins to temperatures  $>100^\circ\text{C}$  at pressures corresponding to 1 bar below  $100^\circ\text{C}$  and to the saturation vapor pressure of  $\text{H}_2\text{O}$  at higher temperatures. Experimental values are denoted by filled circles. Values calculated from group additivity are indicated by the solid and dashed lines (present study), dotted lines (Amend and Helgeson, 2000) and open circles (Makhatadze and Privalov, 1990). Experimental values for unfolded AMYA\_PYRFU (a) are taken from the curve for  $\text{pH}=10.3$  in Fig. 4 of Laderman et al. (1993). In (b) and (c), experimental values taken from Fig. 6 of Guzman-Casado et al. (2003) are shown for RNH\_ECOLI at pHs of 2.0–3.5 and RNH\_THET8 at pHs of 4.0–5.0.

horse serum albumin, even over the relatively narrow temperature range of 5 to  $25^\circ\text{C}$  (Cohn and Edsall, 1943). Another factor which may influence protein net charge includes the formation of disulfide bonds, which may be stabilized in unfolded proteins in oxidizing experimental conditions (Hu and Zou, 1992, 1993). Their formation destabilizes the ionized [Cys<sup>-</sup>] group, effectively increasing the minimum limit of protein charge at high pHs by the number of [Cys] groups that can form disulfide bonds (see Table 8). Hence, the consequences of disulfide bond formation can be represented by the titration curves for oxidized proteins shown in Fig. 10. These titration curves more closely represent the experimental data at high pHs for the proteins in which disulfide bonds can form.

The standard molal thermodynamic properties of proteins can be treated as continuous functions of pH in a manner analogous to the calculations of net charge. Note that the contributions by complexation with species other than the proton are taken to be negligible in the present study. For any standard molal thermodynamic property of an ionized protein ( $\Xi_{UPZ_{\text{net}}}$ ), we can write

$$\Xi_{UPZ_{\text{net}}} = \Xi_{UP^0} + \Delta \Xi_{\text{ion}}, \quad (60)$$



**Fig. 12.**  $Z_{\text{net}}$  and  $\Delta G^\circ$  of the extracellular  $\alpha$ -amylases from *B. subtilis* and *P. furiosus* as a function of pH at 25 °C and 1 bar, and at 100 °C and the saturation vapor pressure of  $\text{H}_2\text{O}$  at this temperature. Note that the units for  $\Delta G^\circ$  are  $\text{Mcal}$  ( $10^6$  cal)  $\text{mol}^{-1}$ .

where  $\Delta \Xi_{\text{ion}}$  denotes the contributions to that property due to the sum of ionization reactions. This value can be calculated by generalizing Eq. (58) to write

$$\Delta \Xi_{\text{ion}} = \sum_i n_i \alpha_i \Delta \Xi_{\text{ion},i}, \quad (61)$$

where  $\Delta \Xi_{\text{ion},i}$  represents the corresponding ionization property of any sidechain or backbone group. Values of  $C_p^\circ$  as a function of temperature calculated using Eqs. (47) and (60) for non-ionized and ionized proteins, respectively, are shown in Fig. 11. The predictions for non-ionized proteins are greater than the experimental values of  $C_p^\circ$  shown in this figure, which were obtained under controlled pH conditions (Laderman et al., 1993; Guzman-Casado et al., 2003). This is true also of the results of the group additivity algorithms for non-ionized proteins used in two other studies (Makhatadze and Privalov, 1990; Amend and Helgeson, 2000). The values of  $C_p^\circ$  calculated for ionized proteins using Eq. (60) much more closely reproduce the experimental measurements. Measured heat capacities in unfolded proteins that are lower than model compound predictions have been used to infer residual structure in unfolded proteins (Georgescu et al., 2001; Guzman-Casado et al., 2003). However, the differences between additivity and experiment may

be explained in part by the ionization of the groups in proteins.

## 7.2 Eh-pH predominance diagram for proteins

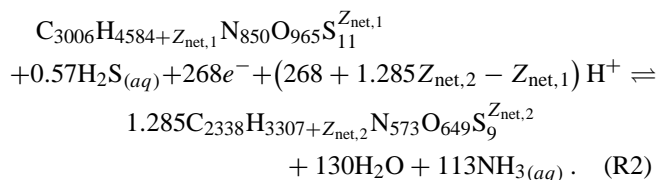
Eh-pH predominance diagrams are widely used to characterize the electrochemistry of inorganic systems (Pourbaix, 1949; Garrels and Christ, 1965; Buvet, 1976; Uchida et al., 1996). The chemical potential constraints on reactions involving  $e^-$ ,  $\text{H}^+$ ,  $\text{O}_2$ , or  $\text{H}_2$  are shown by the locations of the predominance field boundaries on Eh-pH diagrams. The chemical affinity ( $A$ ) of the corresponding reaction is given by (Kondepudi and Prigogine, 1998)

$$A = 2.303RT \log (K/Q), \quad (62)$$

where  $K$  and  $Q$ , respectively, represent the equilibrium constant and the reaction quotient. Along a predominance field boundary,  $K=Q$  and  $A=0$ .

Gradients of chemical potentials between environments are also apparent on Eh-pH diagrams (Garrels, 1960; Baas Becking et al., 1960). If the speciation of proteins among organisms adapted to different environments is characterized by reactions that tend to minimize the Gibbs energy, the location of the predominance field boundaries calculated for model proteins might tend to follow gradients in the electrochemical conditions of the environments. For example, electrochemical data are shown in Fig. 13 for characteristic soil and hot spring environments. The values shown are experimental ones, except for a value of Eh for hydrothermal fluid at 100 °C calculated by combining values of  $\log a_{\text{H}_2}$  and  $\log a_{\text{H}^+}$  taken from Amend and Shock (1998) with Eq. (57) and the logarithmic analog of the law of mass action for  $\text{H}_{2(\text{aq})} \rightleftharpoons 2\text{H}^+ + 2e^-$ .

An oxidation-reduction reaction representing the overall speciation of the extracellular  $\alpha$ -amylases from two organisms adapted to soils and hot springs, *Bacillus subtilis* and *Pyrococcus furiosus*, can be written as



The extracellular  $\alpha$ -amylase from *Bacillus subtilis* is represented with net charge  $Z_{\text{net},1}$ , and that from *Pyrococcus furiosus* is assigned a net charge of  $Z_{\text{net},2}$ . Extracellular  $\alpha$ -amylases were chosen for this study because they represent proteins commonly secreted by microbes into extracellular aqueous solution. Reaction (R2) is written to conserve carbon; note in this regard that the choice of basis species ( $\text{H}_2\text{S}_{(\text{aq})}$ ,  $\text{H}_2\text{O}$ ,  $\text{NH}_{3(\text{aq})}$ ,  $e^-$ , and  $\text{H}^+$ ) and balancing scheme influences the outcome of the calculations. The combination

selected here represents an approximation that yields realistic results for this example.

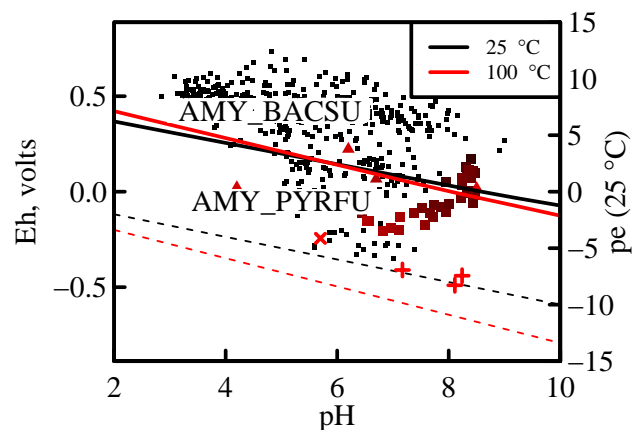
The logarithmic analog of the law of mass action for Reaction (R2) follows from Eq. (62) when  $A=0$ , and can be written as

$$\begin{aligned} \log K_{(R2)} = & \log \frac{a_{\text{AMY\_PYRFU}}^{1.285 Z_{\text{net},2}}}{a_{\text{AMY\_BACSU}}^{Z_{\text{net},1}}} + 130 \log a_{\text{H}_2\text{O}} \\ & + 113 \log a_{\text{NH}_3(\text{aq})} - 0.57 \log a_{\text{H}_2\text{S}(\text{aq})} \\ & + 268 \text{pe} + (268 + 1.285 Z_{\text{net},2} - Z_{\text{net},1}) \text{pH}, \quad (64) \end{aligned}$$

where the summation of terms on the right-hand side corresponds to  $\log Q$ . This equation reveals that decreasing values of pe and pH favor predominance of AMY\_PYRFU over AMY\_BACSU, and vice versa. This statement can be made quantitative by first considering the values of  $Z_{\text{net}}$  and  $\Delta G^\circ$  of the proteins, calculated using Eqs. (47) and (58–61), which are shown in Fig. 12 at two temperatures (25 and 100°C). The curved trends in the values of  $Z_{\text{net}}$  and  $\Delta G^\circ$  shown in this figure reflect the changing ionization states of the proteins as a function of pH. Besides temperature, pressure, and electrochemical variables, the environmental factors affecting the predominance limits of the proteins are represented by the activities of the basis species, which are given nominal values corresponding to  $a_{\text{H}_2\text{O}} = 1$ ,  $a_{\text{NH}_3(\text{aq})} = 10^{-6}$ , and  $a_{\text{H}_2\text{S}(\text{aq})} = 10^{-3}$ . These values represent estimates of the properties of hydrothermal solutions (Amend and Shock, 1998). The activities of the proteins were taken as  $a_{\text{AMY\_PYRFU}} = a_{\text{AMY\_BACSU}} = 10^{-3}$ , which is a nominal value for the activities of proteins in the aqueous solution near a cell surface. Note that changing the protein activities by three orders of magnitude in either direction has an effect of approximately  $\pm 0.1$  on the calculated value of pe.

The values of  $\log K_{(R2)}$  were computed using the values of  $\Delta G^\circ$  shown in Fig. 12 together with the values of  $\Delta G^\circ$  of the basis species calculated using SUPCRT92. The values of  $\log K_{(R2)}$  were then combined with Eq. (64) to obtain values of pe for a range of values of pH. pe was converted to Eh using Eq. (57), the values of which are plotted as the protein predominance field boundaries in Fig. 13. It can be seen in this figure that the calculated predominance field boundaries between the extracellular proteins correspond to the overall gradients of electrochemical potentials between the environments, and that the effect of increasing temperature on the location of the boundary is comparatively small.

These calculations of  $\log K_{(R2)}$  are for the unfolded protein reference state. The maximum uncertainty due to folding properties may be assessed by considering the largest observed values of  $\Delta G^\circ$  of the unfolding reaction of  $\alpha$ -amylases reported by Fitter (2005) ( $\sim 50 \text{ kcal mol}^{-1}$ ). Combining this value with Eq. (64) and the finite difference derivative of Eq. (33) gives a corresponding uncertainty in pe



**Fig. 13.** Eh-pH diagram showing the electrochemical and temperature constraints on the speciation of the extracellular  $\alpha$ -amylases from *B. subtilis* and *P. furiosus*, represented by Reaction (R2). The solid lines and dashed lines represent, respectively, the calculated predominance field boundaries of the proteins, and the reduction stability limits of water. The degree of shading is used to denote temperature. Values are shown for soils ( $\bullet$ ; Baas Becking et al., 1960), solfataric hot springs ( $\blacksquare$ ; Kato et al., 2004), Yellowstone hot springs ( $+$ ; Spear et al., 2005), Icelandic geothermal fluids ( $\blacktriangle$ ; Stefánsson and Arnórsson, 2002), and a nominal hydrothermal fluid ( $\times$ ; McCollom and Shock, 1997 and Amend and Shock, 1998).

of  $\pm 0.02$ . In contrast, the estimated group additivity uncertainty for  $\Delta G^\circ$  of a single protein in Reaction (R2) is of the order of  $\pm 600 \text{ kcal mol}^{-1}$  (see Sect. 6), which corresponds with a maximum uncertainty in pe of  $\pm 0.3$ . Hence, our confidence in the calculation of chemical potential constraints on protein speciation reactions appears to be limited not by the availability of thermodynamic data for protein folding, but instead by the uncertainty inherent in the group additivity calculations of the Gibbs energies of unfolded proteins.

## 8 Conclusions

The group contributions to the standard molal thermodynamic properties at 25°C and 1 bar and the revised HKF equations of state parameters of neutral and charged aqueous sidechain and backbone groups generated in the present study permit calculation of the thermodynamic properties of ionized unfolded proteins with any amino acid sequence as a function of temperature and pressure. The standard molal Gibbs energies of groups in proteins, which can be assessed using the equations and parameters described above, are critical for calculations of the chemical and biological speciation of proteins. These types of calculations include those of the ionization state of proteins, and of the thermodynamic properties of reactions between different proteins. The latter type of calculation may be used to assess the

temperature, pressure, and chemical potential constraints on proteins among organisms adapted to different environments. The outcome of a comparative calculation for two model extracellular proteins from different organisms indicates that the gradients of environmental conditions correspond with the calculated predominance limits of the proteins as a function of oxidation-reduction potential.

A thermodynamic approach to protein speciation confers a greater understanding of the changes of composition in the proteome that occur on the length and time scales of the cell. These reactions can be addressed through a global Gibbs free energy minimization for systems with many proteins, which will require the incorporation of protein thermodynamic data into software tools that traditionally have been used in geochemical equilibrium modeling (Connolly, 1990; Shvarov and Bastrakov, 1999; Karpov et al., 2001; Kulik, 2004). Such advances may make clear the relationship between gradients of oxidation state and pH among subcellular compartments and the spatial distribution of proteins in organisms (Al-Habori, 1995; Conour et al., 2004; Brett et al., 2006). Hence, a thermodynamic approach may ultimately lead to a quantitative framework for assessing the causes and consequences of oxidative stress, disease, and aging.

*Acknowledgements.* Funding for this project was provided by the U.S. National Science Foundation (NSF Grants EAR-96-13753 and EAR-03-09829), the U.S. Department of Energy (DOE Continuing Grants FG03-85ER-13419 and FG02-03ER-15418), the American Chemical Society Petroleum Research Fund (Grant 36467-AC2), and the University of California Committee on Research. We thank J. P. Amend, C. D. Cappa, L. L. Glaser, J. Middelburg, and an anonymous reviewer for providing useful feedback at different stages of writing.

Edited by: J. Middelburg

## References

- Al-Habori, M.: Microcompartmentation, metabolic channelling and carbohydrate metabolism, *Int. J. Biochem. Cell Biol.*, 27, 123–132, 1995.
- Amend, J. P. and Helgeson, H. C.: Calculation of the standard molal thermodynamic properties of aqueous biomolecules at elevated temperatures and pressures. Part I. L- $\alpha$ -amino acids, *J. Chem. Soc., Faraday Trans.*, 93, 1927–1941, 1997a.
- Amend, J. P. and Helgeson, H. C.: Group additivity equations of state for calculating the standard molal thermodynamic properties of aqueous organic species at elevated temperatures and pressures, *Geochim. Cosmochim. Acta*, 61, 11–46, 1997b.
- Amend, J. P. and Helgeson, H. C.: Calculation of the standard molal thermodynamic properties of aqueous biomolecules at elevated temperatures and pressures. II. Unfolded proteins, *Biophys. Chem.*, 84, 105–136, 2000.
- Amend, J. P. and Shock, E. L.: Energetics of amino acid synthesis in hydrothermal ecosystems, *Science*, 281, 1659–1662, 1998.
- Andersson, E. and Holm, N. G.: The stability of some selected amino acids under attempted redox constrained hydrothermal conditions, *Orig. Life Evol. Biosph.*, 30, 9–23, 2000.
- Baas Becking, L. G. M., Kaplan, I. R., and Moore, D.: Limits of the natural environment in terms of pH and oxidation-reduction potentials, *J. Geol.*, 68, 243–284, 1960.
- Benson, S. W. and Buss, J. H.: Additivity rules for the estimation of molecular properties. Thermodynamic properties, *J. Chem. Phys.*, 29, 546–572, 1958.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M.: The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, *Nucleic Acids Res.*, 31, 365–370, 2003.
- Boonyaratanakornkit, B. B., Simpson, A. J., Whitehead, T. A., Fraser, C. M., El-Sayed, N. M. A., and Clark, D. S.: Transcriptional profiling of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* in response to lethal heat and non-lethal cold shock, *Environ. Microbiol.*, 7, 789–797, 2005.
- Brett, C. L., Donowitz, M., and Rao, R.: Does the proteome encode organellar pH?, *FEBS Lett.*, 580, 717–719, 2006.
- Buvet, R.: Topics in Electrochemistry and Bioenergetics, chap. Energetic structure of metabolism, John Wiley and Sons, New York, vol. 1, pp. 105–177, 1976.
- Cabani, S., Gianni, P., Mollica, V., and Lepori, L.: Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution, *J. Solution Chem.*, 10, 563–595, 1981.
- Clarke, R. G. and Tremaine, P. R.: Amino acids under hydrothermal conditions: Apparent molar volumes of aqueous  $\alpha$ -alanine,  $\beta$ -alanine, and proline at temperatures from 298 to 523 K and pressures up to 20.0 MPa, *J. Phys. Chem. B*, 103, 5131–5144, 1999.
- Clarke, R. G., Hnědkovský, L., Tremaine, P. R., and Majer, V.: Amino acids under hydrothermal conditions: Apparent molar heat capacities of aqueous  $\alpha$ -alanine,  $\beta$ -alanine, glycine, and proline at temperatures from 298 to 500 K and pressures up to 30.0 MPa, *J. Phys. Chem. B*, 104, 11 781–11 793, 2000.
- Clarke, R. G. F., Collins, C. M., Roberts, J. C., Trevani, L. N., Bartholomew, R. J., and Tremaine, P. R.: Ionization constants of aqueous amino acids at temperatures up to 250°C using hydrothermal pH indicators and UV-visible spectroscopy: Glycine,  $\alpha$ -alanine, and proline, *Geochim. Cosmochim. Acta*, 69, 3029–3043, 2005.
- Cohn, E. J. and Edsall, J. T.: Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, Reinhold Publishing Corporation, New York, 1943.
- Connolly, J. A. D.: Multivariable phase diagrams: An algorithm based on generalized thermodynamics, *Am. J. Sci.*, 290, 666–718, 1990.
- Conour, J. E., Graham, W. V., and Gaskins, H. R.: A combined in vitro/bioinformatic investigation of redox regulatory mechanisms governing cell cycle progression, *Physiol. Genomics*, 18, 196–205, 2004.
- Cox, J. D., Wagman, D. D., and Medvedev, V. A. (Eds.): CO-DATA Key Values for Thermodynamics, Hemisphere Publishing Corporation, New York, 1989.
- Desnoyers, J. E. and Philip, P. R.: Isothermal compressibilities of aqueous solutions of tetraalkylammonium bromides, *Can. J.*

- Chem., 50, 1094–1096, 1972.
- Dill, K. A.: Additivity principles in biochemistry, *J. Biol. Chem.*, 272, 701–704, 1997.
- Ding, K., Seyfried, W. E., Tivey, M. K., and Bradley, A. M.: In situ measurement of dissolved H<sub>2</sub> and H<sub>2</sub>S in high-temperature hydrothermal vent fluids at the Main Endeavour Field, Juan de Fuca Ridge, *Earth Planet. Sci. Lett.*, 186, 417–425, 2001.
- Domalski, E. S. and Hearing, E. D.: Estimation of the thermodynamic properties of C-H-N-O-S-Halogen compounds at 298.15 K, *J. Phys. Chem. Ref. Data*, 22, 805–1159, 1993.
- Dong, G. Q., Vieille, C., Savchenko, A., and Zeikus, J. G.: Cloning, sequencing, and expression of the gene encoding extracellular  $\alpha$ -amylase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme, *Appl. Environ. Microbiol.*, 63, 3569–3576, 1997.
- Downes, C. J. and Hedwig, G. R.: Partial molar heat capacities of the peptides glycylglycylglycine, glycyl-L-alanylglycine and glycyl-DL-threonylglycine in aqueous solution over the temperature range 50 to 125°C, *Biophys. Chem.*, 55, 279–288, 1995.
- Downes, C. J., Hakin, A. W., and Hedwig, G. R.: The partial molar heat capacities of glycine and glycylglycine in aqueous solution at elevated temperatures and at  $p=10.0$  MPa, *J. Chem. Thermodyn.*, 33, 873–890, 2001.
- Drever, J. I.: *The Geochemistry of Natural Waters*, Prentice Hall, Upper Saddle River, New Jersey, 3rd edn., 436, 1997.
- Duke, M. M., Hakin, A. W., McKay, R. M., and Preuss, K. E.: The volumetric and thermochemical properties of aqueous solutions of L-valine, L-leucine, and L-isoleucine at 288.15, 298.15, and 328.15 K, *Can. J. Chem.*, 72, 1489–1494, 1994.
- Elcock, A. H.: Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability, *J. Mol. Biol.*, 294, 1051–1062, 1999.
- Elser, J. J., Sterner, R. W., Gorokhova, E., Fagan, W. F., Markow, T. A., Cotner, J. B., Harrison, J. F., Hobbie, S. E., Odell, G. M., and Weider, L. J.: Biological stoichiometry from genes to ecosystems, *Ecol. Lett.*, 3, 540–550, 2000.
- Fitter, J.: Structural and dynamical features contributing to thermostability in  $\alpha$ -amylases, *Cell. Mol. Life Sci.*, 62, 1925–1937, 2005.
- Fukuchi, S. and Nishikawa, K.: Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria, *J. Mol. Biol.*, 309, 835–843, 2001.
- Garrels, R. M.: *Mineral Equilibria*, Harper & Brothers, New York, 254, 1960.
- Garrels, R. M. and Christ, C. L.: *Solutions, Minerals, and Equilibria*, Harper & Row, New York, 450, 1965.
- Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., and Brown, P. O.: Genomic expression programs in the response of yeast cells to environmental changes, *Mol. Biol. Cell*, 11, 4241–4257, 2000.
- Georgescu, R. E., Garcia-Mira, M. M., Tasayco, M. L., and Sanchez-Ruiz, J. M.: Heat capacity analysis of oxidized *Escherichia coli* thioredoxin fragments (1–73, 74–108) and their noncovalent complex. Evidence for the burial of apolar surface in protein unfolded states, *Eur. J. Biochem.*, 268, 1477–1485, 2001.
- Goldberg, R. N., Kishore, N., and Lennen, R. M.: Thermodynamic quantities for the ionization reactions of buffers, *J. Phys. Chem. Ref. Data*, 31, 231–370, 2002.
- Griko, Yu. V., Makhatadze, G. I., Privalov, P. L., and Hartley, R. W.: Thermodynamics of barnase unfolding, *Protein Sci.*, 3, 669–676, 1994.
- Guzman-Casado, M., Parody-Morreale, A., Robic, S., Marqusee, S., and Sanchez-Ruiz, J. M.: Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H, *J. Mol. Biol.*, 329, 731–743, 2003.
- Häckel, M., Hinz, H.-J., and Hedwig, G. R.: Tripeptides in aqueous solution: Model compounds for the evaluation of the partial molar heat capacities of amino acid side-chains in proteins, *Thermochim. Acta*, 308, 23–34, 1998.
- Häckel, M., Hinz, H.-J., and Hedwig, G. R.: A new set of peptide-based group heat capacities for use in protein stability calculations, *J. Mol. Biol.*, 291, 197–213, 1999a.
- Häckel, M., Hinz, H.-J., and Hedwig, G. R.: Partial molar volumes of proteins: amino acid side-chain contributions derived from the partial molar volumes of some tripeptides over the temperature range 10–90°C, *Biophys. Chem.*, 82, 35–50, 1999b.
- Häckel, M., Hinz, H.-J., and Hedwig, G. R.: Additivity of the partial molar heat capacities of the amino acid side-chains of small peptides: Implications for unfolded proteins, *Phys. Chem. Chem. Phys.*, 2, 5463–5468, 2000a.
- Häckel, M., Hinz, H.-J., and Hedwig, G. R.: The partial molar volumes of some tetra- and pentapeptides in aqueous solution: a test of amino acid side-chain group additivity for unfolded proteins, *Phys. Chem. Chem. Phys.*, 2, 4843–4849, 2000b.
- Hakin, A. W. and Hedwig, G. R.: Group additivity calculations of the thermodynamic properties of unfolded proteins in aqueous solution: a critical comparison of peptide-based and HKF models, *Biophys. Chem.*, 89, 253–264, 2001a.
- Hakin, A. W. and Hedwig, G. R.: Apparent and partial molar heat capacities and volumes of the amino acids L-lysine monohydrochloride and L-arginine monohydrochloride in aqueous solution at temperatures from  $T=288.15$  K to  $T=328.15$  K, *J. Chem. Thermodyn.*, 33, 1709–1723, 2001b.
- Hakin, A. W., Duke, M. M., Klassen, S. A., McKay, R. M., and Preuss, K. E.: Apparent molar heat capacities and volumes of some aqueous solutions of aliphatic amino acids at 288.15, 298.15, 313.15, and 328.15 K, *Can. J. Chem.*, 72, 362–368, 1994a.
- Hakin, A. W., Duke, M. M., Marty, J. L., and Preuss, K. E.: Some thermodynamic properties of aqueous amino acid systems at 288.15, 298.15, 313.15 and 328.15 K: Group additivity analyses of standard-state volumes and heat capacities, *J. Chem. Soc., Faraday Trans.*, 90, 2027–2035, 1994b.
- Hakin, A. W., Duke, M. M., Groft, L. L., Marty, J. L., and Rushfeldt, M. L.: Calorimetric investigations of aqueous amino acid and dipeptide systems from 288.15 to 328.15 K, *Can. J. Chem.*, 73, 725–734, 1995.
- Hakin, A. W., Copeland, A. K., Liu, J. L., Marriott, R. A., and Preuss, K. E.: Densities, apparent molar volumes, and apparent molar heat capacities of *l*-arginine, *l*-proline and *d,l*-methionine in water at 288.15, 298.15, 313.15, and 328.15 K, *J. Chem. Eng. Data*, 42, 84–89, 1997.
- Hakin, A. W., Cavilla, B., Liu, J. L., and Zorzetti, B.: Thermodynamics of protein model compounds: An investigation of the apparent and partial molar heat capacities and volumes of aqueous solutions of alanyl and seryl side-chain containing cyclic dipep-



- tides, *Phys. Chem. Chem. Phys.*, 3, 3805–3810, 2001.
- Hedwig, G. R. and Hinz, H.-J.: Group additivity schemes for the calculation of the partial molar heat capacities and volumes of unfolded proteins in aqueous solution, *Biophys. Chem.*, 100, 239–260, 2003.
- Helgeson, H. C., Kirkham, D. H., and Flowers, G. C.: Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures, IV. Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600°C and 5 Kb, *Am. J. Sci.*, 281, 1249–1516, 1981.
- Hnědkovský, L. and Cibulka, I.: Group contributions for an estimation of partial molar volumes at infinite dilution for aqueous organic solutes at extended ranges of temperature and pressure, *Int. J. Thermophys.*, 25, 387–395, 2004.
- Hu, C.-H. and Zou, C.-L.: Disulfide containing proteins denatured in 6 mol/L guanidinium chloride are not completely unfolded, *Sci. China, Ser. B*, 35, 1214–1221, 1992.
- Hu, C.-H. and Zou, C.-L.: Effect of native disulfide bonds on the denaturation of proteins in 6 mol/L guanidinium chloride, *Sci. China, Ser. B*, 36, 568–574, 1993.
- Inglese, A. and Wood, R. H.: Apparent molar heat capacities of aqueous solutions of 1-propanol, butane-1,4-diol, and hexane-1,6-diol at temperatures from 300 K to 525 K and a pressure of 28 MPa, *J. Chem. Thermodyn.*, 28, 1059–1070, 1996.
- Inglese, A., Sedlbauer, J., Yezdimer, E. M., and Wood, R. H.: Apparent molar heat capacities of aqueous solutions of propylamine, 1,4-butanediamine, 1,6-hexanediamine, propylamine hydrochloride, propionamide, pyridine, and sodium benzenesulfonate at temperatures from 300 K to 525 K and a pressure of 28 MPa, *J. Chem. Thermodyn.*, 29, 517–531, 1997.
- Jardine, J. J., Call, T. G., Patterson, B. A., Origlia-Luster, M. L., and Woolley, E. M.: Thermodynamics for proton dissociations from aqueous L-histidine at temperatures from 278.15 K to 393.15 K and at the pressure 0.35 MPa: apparent molar volumes and apparent molar heat capacities of the protonated cationic, neutral zwitterionic, and deprotonated anionic forms, *J. Chem. Thermodyn.*, 33, 1419–1440, 2001.
- Jiménez, P., Roux, M. V., and Turrión, C.: Thermochemical properties of *N*-heterocyclic compounds, IV. Enthalpies of combustion, vapour pressures and enthalpies of sublimation, and enthalpies of formation of 2-methylimidazole and 2-ethylimidazole, *J. Chem. Thermodyn.*, 24, 1145–1149, 1992.
- Johnson, J. W., Oelkers, E. H., and Helgeson, H. C.: SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C, *Comp. Geosci.*, 18, 899–947, 1992.
- Jolicœur, C. and Lacroix, G.: Thermodynamic properties of aqueous organic solutes in relation to their structure, 3. Apparent molal volumes and heat capacities of low molecular weight alcohols and polyols at 25°C, *Can. J. Chem.*, 54, 624–631, 1976.
- Jolicœur, C., Riedl, B., Desrochers, D., Lemelin, L. L., Zamojska, R., and Enea, O.: Solvation of amino acid residues in water and urea-water mixtures: Volumes and heat capacities of 20 amino acids in water and in 8 molar urea at 25°C, *J. Solution Chem.*, 15, 109–128, 1986.
- Karpov, I. K., Chudnenko, K. V., Kulik, D. A., Avchenko, O. V., and Bychinskii, V. A.: Minimization of Gibbs free energy in geochemical systems by convex programming, *Geochem. Int.*, 39, 1108–1119, 2001.
- Kato, K., Kobayashi, T., Yamamoto, H., Nakagawa, T., Maki, Y., and Hoaki, T.: Microbial mat boundaries between chemolithotrophs and phototrophs in geothermal hot spring effluents, *Geomicrobiol. J.*, 21, 91–98, 2004.
- Kharakoz, D. P.: Volumetric properties of proteins and their analogues in diluted water solutions. 2. Partial adiabatic compressibilities of amino acids at 15–70°C, *J. Phys. Chem.*, 95, 5634–5642, 1991.
- Kharakoz, D. P.: Partial volumes and compressibilities of extended polypeptide chains in aqueous solution: Additivity scheme and implication of protein unfolding at normal and high pressure, *Biochemistry*, 36, 10276–10285, 1997.
- Kikuchi, M., Sakurai, M., and Nitta, K.: Partial molar volumes and adiabatic compressibilities of amino acids in dilute aqueous solutions at 5, 15, 25, 35, and 45°C, *J. Chem. Eng. Data*, 40, 935–942, 1995.
- Kondepudi, D. K. and Prigogine, I.: *Modern Thermodynamics*, John Wiley & Sons, New York, 486, 1998.
- Kreil, D. P. and Ouzounis, C. A.: Identification of thermophilic species by the amino acid compositions deduced from their genomes, *Nuc. Acids Res.*, 29, 1608–1615, 2001.
- Kulik, D. A.: GEM-Selektor version 2-PSI, <http://les.web.psi.ch/Software/GEMS-PSI/index.html>, 2004.
- Kundrotas, P. J. and Karshikoff, A.: Model for calculation of electrostatic interactions in unfolded proteins, *Phys. Rev. E*, 65, 011901, 2002.
- Laderman, K. A., Davis, B. R., Krutzsch, H. C., Lewis, M. S., Griko, Y. V., Privalov, P. L., and Anfinsen, C. B.: The purification and characterization of an extremely thermostable  $\alpha$ -amylase from the hyperthermophilic archaeobacterium *Pyrococcus furiosus*, *J. Biol. Chem.*, 268, 24394–24401, 1993.
- LaRowe, D. E. and Helgeson, H. C.: Biomolecules in hydrothermal systems: Calculation of the standard molal thermodynamic properties of nucleic-acid bases, nucleosides, and nucleotides at elevated temperatures and pressures, *Geochim. Cosmochim. Acta*, accepted, 2006.
- Lepori, L. and Gianni, P.: Partial molar volumes of ionic and non-ionic organic solutes in water: A simple additivity scheme based on the intrinsic volume approach, *J. Solution Chem.*, 29, 405–447, 2000.
- Li, J., Wang, X., Klein, M. T., and Brill, T. B.: Spectroscopy of hydrothermal reactions, 19: pH and salt dependence of decarboxylation of  $\alpha$ -alanine at 280–330°C in an FT-IR spectroscopy flow reactor, *Int. J. Chem. Kinet.*, 34, 271–277, 2002.
- Makhatadze, G. I. and Privalov, P. L.: Heat capacity of alcohols in aqueous solutions in the temperature range from 5 to 125°C, *J. Solution Chem.*, 18, 927–936, 1989.
- Makhatadze, G. I. and Privalov, P. L.: Heat capacity of proteins, 1. Partial molar heat capacity of individual amino acid residues in aqueous solution: Hydration effect, *J. Mol. Biol.*, 213, 375–384, 1990.
- Makhatadze, G. I., Medvedkin, V. N., and Privalov, P. L.: Partial molar volumes of polypeptides and their constituent groups in aqueous solution over a broad temperature range, *Biopolymers*, 30, 1001–1010, 1990.
- Makhatadze, G. I., Kim, K.-S., Woodward, C., and Privalov, P. L.: Thermodynamics of BPTI folding, *Protein Sci.*, 2, 2028–2036,

- 1993.
- Makhatadze, G. I., Clore, G. M., Gronenborn, A. M., and Privalov, P. L.: Thermodynamics of unfolding of the all  $\beta$ -sheet protein interleukin-1 $\beta$ , *Biochemistry*, 33, 9327–9332, 1994.
- Marcus, Y.: A simple empirical model describing the thermodynamics of hydration of ions of widely varying charges, sizes, and shapes, *Biophys. Chem.*, 51, 111–127, 1994.
- Marriott, R. A., Hakin, A. W., and Liu, J. L.: Modeling of thermodynamic properties of amino acids and peptides using additivity and HKF theory, *J. Solution Chem.*, 27, 771–802, 1998.
- Marriott, R. A., Hakin, A. W., Liu, J. L., and Lutter, E.: The volumetric properties of aqueous solutions of glycylglycine and L-serine at elevated temperatures and pressures, *J. Chem. Thermodyn.*, 33, 959–982, 2001.
- McCollom, T. M. and Shock, E. L.: Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems, *Geochim. Cosmochim. Acta*, 61, 4375–4391, 1997.
- Millero, F. J., Lo Surdo, A., and Shin, C.: The apparent molal volumes and adiabatic compressibilities of aqueous amino acids at 25 °C, *J. Phys. Chem.*, 82, 784–792, 1978.
- Mizuguchi, M., Sakurai, M., and Nitta, K.: Partial molar volumes and adiabatic compressibilities of *N*-acetyl-DL-serinamide and *N*-acetyl-L-threoninamide in dilute aqueous solutions, *J. Solution Chem.*, 26, 579–594, 1997.
- Ngauv, S. N., Sabbah, R., and Laffitte, M.: Thermodynamique de composés azotés. III. Etude thermochemique de la glycine et de la L- $\alpha$ -alanine, *Thermochim. Acta*, 20, 371–380, 1977.
- Nichols, N., Sköld, R., Spink, C., and Wadsö, I.: Thermochemistry of solutions of biochemical model compounds. 6.  $\alpha,\omega$ -dicarboxylic acids, -diamines, and -diols in aqueous solution, *J. Chem. Thermodyn.*, 8, 993–999, 1976.
- Nozaki, Y. and Tanford, C.: Proteins as random coils. II. Hydrogen ion titration curve of ribonuclease in 6M guanidine hydrochloride, *J. Am. Chem. Soc.*, 89, 742–749, 1967.
- Plyasunov, A. V. and Shock, E. L.: Correlation strategy for determining the parameters of the revised Helgeson-Kirkham-Flowers model for aqueous nonelectrolytes, *Geochim. Cosmochim. Acta*, 65, 3879–3900, 2001.
- Pourbaix, M. J. N.: *Thermodynamics of Dilute Aqueous Solutions*, Edward Arnold & Co., London, 644, 1949.
- Price, J. L., Jardine, J. J., Call, T. G., Patterson, B. A., Origliaster, M. L., and Woolley, E. M.: Thermodynamics for proton dissociations from aqueous L-histidine at temperatures from 278.15 to 393.15 K and at the pressure 0.35 MPa: apparent molar volumes and apparent molar heat capacities of the protonated cationic, neutral zwitterionic, and deprotonated anionic forms, *J. Chem. Thermodyn.*, 35, 195–198, 2003a.
- Price, J. L., Sorenson, E. C., Merkley, E. D., McRae, B. R., and Woolley, E. M.: Thermodynamics of proton dissociations from aqueous L-valine and L-2-amino-*n*-butanoic acid: apparent molar volumes and apparent molar heat capacities of the protonated cationic, neutral zwitterionic, and deprotonated anionic species at temperatures from  $278.15 \leq T/K \leq 393.15$ , at molalities  $0.015 \leq m/\text{mol}\cdot\text{kg}^{-1} \leq 0.67$ , and pressure  $p=0.35$  MPa, *J. Chem. Thermodyn.*, 35, 1425–1467, 2003b.
- Privalov, P. L. and Makhatadze, G. I.: Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: Protein unfolding effects, *J. Mol. Biol.*, 213, 385–391, 1990.
- Privalov, P. L., Tiktopulo, E. I., Venyaminov, S. Yu., Griko, Yu. V., Makhatadze, G. I., and Khechinashvili, N. N.: Heat capacity and conformation of proteins in the denatured state, *J. Mol. Biol.*, 205, 737–750, 1989.
- Roxby, R. and Tanford, C.: Hydrogen ion titration curve of lysozyme in 6 M guanidine hydrochloride, *Biochemistry*, 10, 3348–3352, 1971.
- Sabbah, R. and Laffitte, M.: Thermodynamique de composés azotés. IV. Étude thermochemique de la sarcosine et de la L-proline, *Bull. Soc. Chim. Fr.*, 1, 50–52, 1978.
- Schafer, F. Q. and Buettner, G. R.: Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple, *Free Radic. Biol. Med.*, 30, 1191–1212, 2001.
- Schulte, M. D., Shock, E. L., and Wood, R. H.: The temperature dependence of the standard-state thermodynamic properties of aqueous nonelectrolytes, *Geochim. Cosmochim. Acta*, 65, 3919–3930, 2001.
- Schulze, W. X.: Protein analysis in dissolved organic matter: What proteins from organic debris, soil leachate and surface water can tell us – a perspective, *Biogeosciences*, 2, 75–86, 2005, <http://www.biogeosciences.net/2/75/2005/>.
- Scott, D. W., Berg, W. T., Hossenlopp, I. A., Hubbard, W. N., Messerly, J. F., Todd, S. S., Douslin, D. R., McCullough, J. P., and Waddington, G.: Pyrrole: Chemical thermodynamic properties, *J. Phys. Chem.*, 71, 2263–2270, 1967.
- Shock, E. L.: Stability of peptides in high-temperature aqueous solutions, *Geochim. Cosmochim. Acta*, 56, 3481–3491, 1992.
- Shock, E. L.: Organic acids in hydrothermal solutions: Standard molal thermodynamic properties of carboxylic acids and estimates of dissociation constants at high temperatures and pressures, *Am. J. Sci.*, 295, 496–580, 1995.
- Shock, E. L. and Helgeson, H. C.: Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Correlation algorithms for ionic species and equation of state predictions to 5 kb and 1000°C, *Geochim. Cosmochim. Acta*, 52, 2009–2036, 1988.
- Shock, E. L., Helgeson, H. C., and Sverjensky, D. A.: Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of inorganic neutral species, *Geochim. Cosmochim. Acta*, 53, 2157–2183, 1989.
- Shock, E. L., Oelkers, E. H., Johnson, J. W., Sverjensky, D. A., and Helgeson, H. C.: Calculation of the thermodynamic properties of aqueous species at high pressures and temperatures: Effective electrostatic radii, dissociation constants, and standard partial molal properties to 1000°C and 5 kbar, *J. Chem. Soc., Faraday Trans.*, 88, 803–826, 1992.
- Shock, E. L., Sassani, D. C., Willis, M., and Sverjensky, D. A.: Inorganic species in geologic fluids: Correlations among standard molal thermodynamic properties of aqueous ions and hydroxide complexes, *Geochim. Cosmochim. Acta*, 61, 907–950, 1997.
- Shvarov, Yu. V. and Bastrakov, E. N.: HCh: A software package for geochemical equilibrium modelling, Australian Geological Survey Organisation, record 1999/25, 1999.
- Sorenson, E. C., Price, J. L., McRae, B. R., and Woolley, E. M.: Thermodynamics of proton dissociations from aqueous L-proline: apparent molar volumes and apparent molar heat capacities

- ities of the protonated cationic, zwitterionic, and deprotonated anionic forms at temperatures from 278.15 K to 393.15 K and at the pressure 0.35 MPa, *J. Chem. Thermodyn.*, 35, 529–553, 2003.
- Spear, J. R., Walker, J. J., McCollom, T. M., and Pace, N. R.: Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem, *Proc. Natl. Acad. Sci. USA*, 102, 2555–2560, 2005.
- Stefánsson, A. and Arnórsson, S.: Gas pressures and redox reactions in geothermal fluids in Iceland, *Chem. Geol.*, 190, 251–271, 2002.
- Steinhardt, J. and Reynolds, J. A.: *Multiple Equilibria in Proteins*, Academic Press, New York, 1969.
- Tan, Y.-J., Oliveberg, M., Davis, B., and Fersht, A. R.: Perturbed  $pK_A$ -values in the denatured states of proteins, *J. Mol. Biol.*, 254, 980–992, 1995.
- Tanger, J. C. IV. and Helgeson, H. C.: Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Revised equations of state for the standard partial molal properties of ions and electrolytes, *Am. J. Sci.*, 288, 19–98, 1988.
- Tollinger, M., Crowhurst, K. A., Kay, L. E., and Forman-Kay, J. D.: Site-specific contributions to the pH dependence of protein stability, *Proc. Natl. Acad. Sci. USA*, 100, 4545–4550, 2003.
- Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F.: Community structure and metabolism through reconstruction of microbial genomes from the environment, *Nature*, 428, 37–43, 2004.
- Uchida, S., Uchida, M., Okuwaki, A., and Umetsu, Y.: The Eh-pH diagram for N-H<sub>2</sub>O system at 623 K calculated by the revised HKF equation, *Nippon Kagaku Kaishi*, pp. 734–741, 1996.
- Vogl, T., Hinz, H.-J., and Hedwig, G. R.: Partial molar heat capacities and volumes of Gly-X-Gly tripeptides in aqueous solution: model studies for the rationalization of thermodynamic parameters of proteins, *Biophys. Chem.*, 54, 261–269, 1995.
- Wang, P., Oscarson, J. L., Gillespie, S. E., Izatt, R. M., and Cao, H.: Thermodynamics of protonation of amino acid carboxylate groups from 50 to 125°C, *J. Solution Chem.*, 25, 243–266, 1996.
- Whitten, S. T. and García-Moreno E., B.: pH dependence of stability of staphylococcal nuclease: Evidence of substantial electrostatic interactions in the denatured state, *Biochemistry*, 39, 14 292–14 304, 2000.
- Yasuda, Y., Tochio, N., Sakurai, M., and Nitta, K.: Partial molar volumes and isentropic compressibilities of amino acids in dilute aqueous solutions, *J. Chem. Eng. Data*, 43, 205–214, 1998.
- Yu, Y., Makhatazde, G. I., Pace, C. N., and Privalov, P. L.: Energetics of ribonuclease T1 structure, *Biochemistry*, 33, 3312–3319, 1994.