Theor Chem Acc (2013) 132:1343
DOI 10.1007/s00214-013-1343-y

REGULAR ARTICLE

On the transferability of fractional contributions to the hydration free energy of amino acids

Josep M. Campanera · Xavier Barril · F. Javier Luque

Received: 15 December 2012 / Accepted: 15 January 2013 / Published online: 2 February 2013

Abstract This study reports the application of the quantum mechanical self-consistent reaction field MST method to compute the solvation profile in water of the twenty natural amino acids. The aim is to derive intrinsic fractional contributions to the hydration free energy and to examine their transferability to peptides. To this end, IEF-MST calculations have been performed at the B3LYP/6-31G(d) level for the series of acetyl amino acid amides, which were chosen as model compounds. In order to account for the flexibility of both the backbone and the side chain in deriving the hydration fractional contributions, calculations have been performed for representative conformers taken from the Dunbrack's backbone-dependent conformational library. The results allow us to dissect the hydration free energy into backbone and side chain contributions and examine the conformational dependence of the fragmental contributions to hydration. For the backbone, different hydration contributions are found for α-helical and β-sheet conformations, which mainly reflect differences in the electrostatic contribution to hydration of the carbonyl group. In contrast, the conformational flexibility of the side chain is found to have little impact on the fractional contribution to hydration. These findings should be valuable to refine semiempirical methods for predicting solvation properties of peptides and proteins in large-scale genomic studies.

Keywords Hydration free energy · Continuum solvation models · Atomic solvation profile · Amino acids · Peptides

Published as part of the special collection of articles derived from the 8th Congress on Electronic Structure: Principles and Applications (ESPA 2012).

Electronic supplementary material The online version of this article (doi:10.1007/s00214-013-1343-y) contains supplementary material, which is available to authorized users.

J. M. Campanera (✉) · F. J. Luque (✉)
Department of Physical Chemistry, Faculty of Pharmacy and Institute of Biomedicine (IBUB), University of Barcelona, Campus de l'Alimentació Torribera, Avgda. Prat de la Riba 171, 08921 Santa Coloma de Gramenet, Spain
e-mail: campanera@ub.edu

F. J. Luque
e-mail: fjluque@ub.edu

X. Barril
Department of Physical Chemistry, Faculty of Pharmacy and Institute of Biomedicine (IBUB), University of Barcelona, Avgda. Diagonal 643, 08028 Barcelona, Spain

X. Barril
Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

1 Introduction

Solvation plays a crucial role in modulating the structure and flexibility of proteins, as well as in mediating their interaction with small ligands and other macromolecular entities. The study of these challenging biological processes and their functional implications is associated with an understanding of the intrinsic solvation properties of amino acid residues. In this context, it is not surprising that many experimental studies have examined the solvation preferences of amino acids in different environments, paying particular attention to solvation in aqueous solution and to the transfer free energies from water to a nonpolar medium [1–9]. Alternatively, computational studies, generally based on the use of discrete treatments of solvation coupled to free energy calculations, have also been used to determine the solvation properties of amino acids [10–13].

The solvation free energy (ΔG_{solv}) measures the reversible work needed to transfer the solute from the gas phase to solution at constant pressure, temperature and concentration [14, 15]. It is convenient to decompose the solvation of a solute into three steps: (1) creation of a solute-shaped cavity in the solvent, (2) switching on the steric properties of the solute inside the cavity and (3) building up of the charge distribution of the solute. This latter term gives rise to the electrostatic component of ΔG_{solv}, whereas the two former contributions are generally grouped into the "steric" or non-electrostatic term. A plethora of theoretical methods relying on different formalisms has been developed to estimate the solvation free energy [16–19]. Among them, quantum mechanical self-consistent reaction field (QM-SCRF) continuum methods have found widespread acceptance for the study of small compounds in solution [20, 21]. The success of these methods can be attributed to the rigorous physical basis of their formalisms, the small increase in computational cost compared to *in vacuo* calculations and the possibility to examine the effect of solvation on the solute properties from the wave function in solution. Finally, it is also worth noting that the most elaborate QM-SCRF methods are capable of estimating the solvation free energy of small (bio)organic compounds with chemical accuracy, as noted in the SAMPL challenges for prediction of hydration free energies [22–29].

The application of these methods to larger systems such as peptides and even proteins is limited by their large size and the complexity of their conformational space. Although the solvation free energy is a property of the entire solute, it is however, useful to partition this quantity into group contributions, since not all the fragments of the solute interact with the same strength with the solvent molecules. Thus, such a partitioning would be valuable to identify chemical groups in the solute that display a prominent contribution to the solvation, to explore the dependence of the solvation free energy on conformational changes that affect the spatial arrangement of fragments or to examine their sensitivity to changes in the local environment [30–32]. This partitioning scheme is implicitly assumed for peptides and proteins, as the total solvation free energy can be expressed as the sum of the solvation contributions from the amino acid residues, which are typically weighted by the ratio of the solvent accessible surface area (SAS) of the residue in the protein and in the fully solvent-exposed free residue [33–42]. In the framework of QM-SCRF methods, partitioning schemes have been developed for the Generalized Born-based SMx model [43, 44], the COSMO-RS method [45, 46] and within the MST model [47–50].

The aim of this study is to determine the fractional contributions of backbone and side chain to the hydration free energy of the twenty amino acid residues using the partitioning strategy implemented in the IEF-MST model [51]. Keeping in mind the wide range of physicochemical properties covered by the natural amino acids and the conformational flexibility of both backbone and side chains, the transferability of these fragmental contributions to the hydration is not fully guaranteed. Accordingly, our first goal is to quantify the contribution of both backbone and side chain to the hydration free energy of amino acids. In addition, the influence of conformation flexibility on the transferability of the fragmental contributions to hydration will also be examined. Finally, the potential application of the fragmental hydration contributions for predicting the total hydration free energy of secondary structural elements will be determined.

2 Methods

2.1 Molecular models and geometries

N-acetyl-L-amino acid amides (CH_3–CO–NH–CHR–CONH–CH_3) were chosen as molecular models to study the hydration contribution of the backbone and the side chain for the twenty natural amino acids. Inclusion of acetyl and methylamine capping groups allow us to build models of amino acids in a context that mimics the local environment of residues in a polypeptide. For the set of acidic (Asp, Glu) and basic (Lys, Arg) amino acids, calculations were performed for both the neutral and ionized species.

In order to examine the conformational dependence on the hydration free energy, calculations were performed for the set of most relevant conformations found in proteins following the backbone-dependent conformational library reported by Dunbrack and Karplus [52, 53]. This library contains the probabilities of side chain conformations (defined by a set of χ angles) as a function of the backbone dihedrals (φ and ψ, based on a grid of $20 \times 20°$) in 132 protein chains taken from the Brookhaven Protein Database (PDB). For our purposes here, calculations have been performed for all rotamers with a probability contribution higher to 5 % over the total conformational space of a specific amino acid. In the case of Gly and Ala, which are the simplest models of the peptide backbone, we have chosen eight conformations that encompass the torsional angles defined by the regions of the Ramachandran plot associated with α-helical and β-sheet conformations. Within these premises, up to 408 conformations were selected for the whole set of natural amino acids. By using this set of conformations, more than 90 % of the conformational space is covered for 13 amino acids (Gly, Ala, Leu, Ile, Val, Pro, Phe, Tyr, Ser, Thr, Cys, Asn and Asp). For Met, Trp, His, Glu, Lys and Gln, it covers between 50

and 90 % of the rotamers found in proteins. Finally, only for Arg the coverage of the conformational space is lower than 50 %, which reflects the higher flexibility of the side chain for this residue. On average, more than 15 rotamers are used to describe the conformational space of each residue covering either α-helical or β-sheet structures (see Table 1).

The initial structures of *N*-acetyl-L-amino acid amides were generated automatically by using the obrotate program from OpenBabel suite [54], which permits to rotate the dihedral angle of a specified bond to fit a target value. Then, all the structures were geometrically optimized while keeping the backbone dihedrals fixed to the torsional values of the Dunbrack's library. The geometrical optimization was performed at the B3LYP/6-31G(d) level.

In addition, four pentapeptides that adopt α-helical and β-sheet structures were extracted from selected crystallographic or NMR structures of polypeptides taken from the Protein Data Bank (PDB) [55] and used as models to examine the transferability of the intrinsic hydration free energy contributions of amino acid residues. The residue segments 27–31 of PDB entry 1SPF (Ile-Val-Gly-Ala-Leu; from NMR [56]) and 38–42 of PDB entry 2P5K (Gln-Ala-Thr-Val-Ser; from X-ray at 1.0 Å resolution [57]) were chosen as α-helical model structures, while segments 265–269 and 37–41 of PDB entries 1R6J (Val-Thr-Ile-Thr-Ile; from X-ray at 0.73 Å resolution [58]) and 3PUC (Thr-Ala-Ile-Trp-Thr; from X-ray at 0.96 Å resolution [59]), respectively, were taken for β-sheet structures. In all cases, the pentapeptides were capped with acetyl and methylamine groups at both N- and C-terminus, and the hydration free energy was determined from IEF-MST B3LYP/6-31G(d) calculations.

Table 1 Number of conformers considered for each residue (with a population larger than or equal to 5 %) and distribution between α-helical and β-sheet conformations

Residue	Total no. conformers	α-helix conformers	β-sheet conformers	% Conformational space covered
Gly	8	4	4	99.9
Ala	8	4	4	99.9
Leu	10	5	5	94.3
Ile	9	5	4	96.0
Met	28	12	16	80.3
Val	7	3	4	98.9
Pro	8	4	4	97.9
Phe	12	6	6	91.5
Tyr	14	6	8	96.2
Trp	24	10	14	84.1
His[a]	18	8	10	89.6
Ser	9	4	5	99.0
Thr	7	3	4	97.5
Cys	8	4	4	99.9
Asn	21	8	13	93.9
Gln	23	12	11	72.3
Asp	17	8	9	97.2
Glu	28	16	12	81.7
Arg	25	13	12	42.0
Lys	27	15	12	52.5
Total	408	201	207	–

The percentage of conformational space covered by the selected conformers is also given

[a] HIE, histidine residue protonated in the epsilon position

2.2 Solvation calculations

The hydration free energy of molecular systems was determined by using the B3LYP/6-31G(d) IEF-MST version [60, 61], which relies on the IEF version of the PCM model [62]. The reader is addressed to Ref. [51] for details about the formalism of the MST model and the parametrization of the electrostatic, cavitation and van der Waals components of ΔG_{solv}. Here, we limit ourselves to describe the essential features of the partitioning scheme implemented in the MST method [38, 39], where the hydration free energy (ΔG_{hyd}) is partitioned into atomic contributions as noted in Eq. 1.

$$\Delta G_{\mathrm{hyd}} = \sum_{i=1}^{N} \Delta G_{\mathrm{hyd},i} = \sum_{i=1}^{N} \left(\Delta G_{\mathrm{ele},i} + \Delta G_{\mathrm{cav},i} + \Delta G_{\mathrm{vW},i} \right) \tag{1}$$

where N is the number of atoms.

Following a perturbative treatment of the solvent response [47], the fractional electrostatic contribution of a given atom is determined from the interaction energy between the whole charge distribution of the molecule and the apparent charges located at the surface elements pertaining to the portion of the cavity generated from that atom (Eq. 2).

$$\Delta G_{\mathrm{ele},i} = \sum_{\substack{j=1 \\ j \in i}}^{M} \left\langle \Psi^{o} \middle| \frac{1}{2} \frac{q_j^{sol}}{|r_j - r|} \middle| \Psi^{o} \right\rangle \tag{2}$$

where Ψ^{o} denotes the wave function of the solute in the gas phase, and q_j^{sol} stands for the apparent charge created on the surface element j (located at r_j) in response to the fully polarized solute in solution.

The cavitation (ΔG_{cav}) and van der Waals (ΔG_{vW}) terms are easily decomposed into atomic components taking advantage of the relationship with the atomic contribution to the solvent-exposed surface of the solute, as noted in Eqs. 3 and 4.

$$\Delta G_{\mathrm{cav},i} = \frac{S_i}{S_{\mathrm{tot},i}} \Delta G_{P,i} \quad (3)$$

where $\Delta G_{P,i}$ stands for the cavitation free energy of the isolated atom i in Pierotti's formalism [63], S_i is the solvent-exposed surface of such an atom and $S_{\mathrm{tot},i}$ denotes the total surface of the atom.

$$\Delta G_{\mathrm{vW},i} = S_i \, \xi_i \quad (4)$$

where ξ_i is the atomic surface tension determined by fitting experimental values.

On the basis of Eqs. 2–4, one can easily partition the hydration free energy between contributions due to the atoms pertaining to the backbone and to the side chain, excluding the atoms in the capping groups (acetyl and methylamine moieties), as well as to separate the contribution of apolar and polar fragments of the side chain.

2.3 Computational details

Following the standard parameterization of the IEF-MST model, molecular geometries of the amino acid residues were optimized at the B3LYP/6-31G(d) level, but for the torsional angles that define a given conformation in the backbone-dependent conformational library (see above). The optimized geometries were then kept frozen in MST calculations in water. A similar strategy was adopted for the pentapeptide systems, though the backbone dihedrals were fixed to the values found in the crystallographic structure. All calculations were performed using Gaussian03 package [64]. In addition, besides MST calculations, the hydration free energy of pentapeptides was also determined at both MM-GBSA and MM-PBSA levels for the sake of comparison. These calculations were carried out by mmpbsa.pl script under the AMBER11 suite [65]. Finally, the multivariate partial least squares (PLS [66]) regression technique was used to extract the relevant trends between the hydration free energy and the fragmental contributions following the module implemented in the R statistics package [67].

Due to the large number of calculations required to examine the conformational dependence of hydration properties and the analysis of the huge amount of data derived from atomic contributions to ΔG_{hyd}, Kepler [68, 69] was used to set up a workflow environment to extract the relevant conformers from Dunbrack's library, to build automatically the molecular geometries with fixed torsional angles, to perform QM and QM-SCRF calculations and to store all data in a Mysql database. This workflow is available upon request to the authors. Finally, the library of hydration fractional contributions derived from the partitioning scheme is also available upon request.

3 Results and discussion

3.1 Hydration free energies: comparison with experimental data and other theoretical values

Table 2 reports the side chain contributions to the hydration free energy determined from IEF-MST calculations (weighted by population of each conformer) for the amino acid residues, as well as the experimental data determined from dynamic vapour pressure measurements for a set of molecules that can be considered analogues of amino acid side chains [70]. The involvement of different molecular structures in both IEF-MST calculations and experimental measurements (i.e., the side chain bound to the peptide backbone versus a small molecule that mimics the side chain) makes it necessary to impose some caution regarding the strict comparison of the corresponding hydration values. As an example, let us note that the side chains of small hydrophobic residues (Ala, Val, Leu and Ile) are predicted to be less hydrophobic than the molecular analogues. This difference, however, can be attributed to the partial burial of the side chain atoms by the backbone and, hence, less exposure to water compared to the analogue compounds, reflecting the solvent-excluding effect of the backbone on the fractional hydration of the side chain [38]. In spite of these considerations, it is worth noting that the two sets of data exhibit a nice correlation ($r = 0.96$; Fig. 1). The mean absolute error is close to 1 kcal/mol for a range of hydration free energies of 13.2 kcal/mol (experimental data varying from +2.3 kcal/mol for isobutane to −10.9 kcal/mol for *N*-propylguanidine as mimics of Leu and Arg, respectively). Overall, these findings suggest that the IEF-MST fragmental contributions reflect the influence of the distinctive physicochemical properties of side chain on the hydration of amino acid residues.

Table 2 also reports the hydration free energies determined for analogues of the amino acid side chains by using other theoretical approaches, including either a solvent interaction potential based on quantum mechanically derived charges [71] or free energy calculations coupled to Monte Carlo [38] and molecular dynamics simulations [72]. The agreement between IEF-MST results and the other theoretical values is significant (Fig. 1), as noted in correlation coefficients ranging between 0.92 and 0.98 and average absolute errors varying from 1.2 to 1.9 kcal/mol.

3.2 Backbone/side chain partitioning of the hydration free energy

Table 3 shows the hydration free energy of the 20 *N*-acetyl-L-amino-acid amides decomposed into the considered backbone and side chain moieties, including the contributions to apolar/polar fragments in the side chain as

well. For the sake of consistency with the molecular model, Table 3 also reports the contributions of the capping (acetyl, methylamine) groups added to mimic the environment of the residue in a peptide chain. For neutral residues, the hydration free energy of the *N*-acetyl-L-amino-acid amides varies from −10.4 (Pro) and −24.7 (Arg) kcal/mol, whereas the ΔG_{hyd} values of the residues with charged side chains range from −68.0 (Asp) to −83.1 (Lys) kcal/mol.

The partitioning of the hydration free energy shows that the contribution of the capping groups is almost constant for the neutral residues. Their average values amount to −3.6 (acetyl; ΔG_{acetyl}) and −2.5 (methylamine; $\Delta G_{methylamine}$) kcal/mol for methylamine, accounting for 40–50 % of ΔG_{hyd}. This indicates that the contribution of the capping groups is rather independent from the nature of the neutral residue and thus can be considered to be additive. However, this is not valid for ionic residues, as the capping group contribution varies from −7.9 to 1.4 kcal/mol. In fact, the fractional hydration can be realized from the net charge of the residue and the length of the side chain. Thus, the reaction field created by a positive charge tends to favour the hydration of the methylamine group, while a negative charge favours hydration of the acetyl group, this effect being larger for residues with shorter side chains.

By excluding the contribution of the capping groups, the hydration of the neutral residues ranges between −7.3 and −18.5 kcal/mol. The backbone ($\Delta G_{backbone}$) exerts a favourable contribution to the hydration free energy, which on average amounts to −5.1 kcal/mol (see Table 3; Fig. 2). For ionic residues, however, the contribution becomes more negative for Asp (−11.9 kcal/mol) and Glu (−10.0 kcal/mol), but more positive for Arg (−3.7 kcal/mol) and Lys (−3.4 kcal/mol). As noted above, these trends reflect the perturbing effect due to the net charge of the side chain. Moreover, as will be discussed later, these trends suggest that the hydration of the backbone is primarily determined by the contribution of the CO group compared to the NH unit. On the other hand, the contribution of the

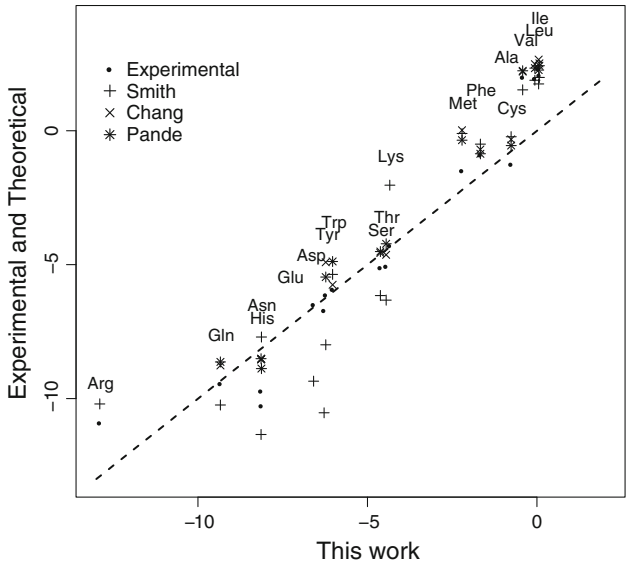

Fig. 1 Correlation between the IEF-MST hydration free energy (*x* axis) for the side chain of neutral amino acids and experimental [70] or theoretical (Smith [71]; Chang [38]; Pande [72]) values determined for side chain analogues. The *dashed line* represents a perfect regression line with slope unity. Values in kcal/mol

Table 2 Comparison between the IEF-MST hydration free energy of the side chain and the values determined for side chain analogues from computational and experimental studies

Residue	This work	Side chain analogue	Smith[a]	Chang[b]	Pande[c]	Exp.[d]
Ala	−0.42	Methane	1.53	2.21	2.24	2.01
Leu	0.05	Isobutane	1.75	2.66	2.27	2.32
Ile	0.07	Butane	1.99	2.52	2.43	2.08
Met	−2.21	Methyl ethyl sulphide	−0.10	0.02	−0.35	−1.48
Val	−0.06	Propane	1.89	2.49	2.34	1.96
Phe	−1.67	Toluene	−0.50	−0.70	−0.86	−0.89
Tyr	−6.23	p-cresol	−7.99	−4.91	−5.46	−6.12
Trp	−6.03	Methylindole	−5.36	−5.75	−4.88	−5.91
His	−8.13	Methylimidazole	−7.70	−8.52	−8.88	−10.26
Ser	−4.62	Methanol	−6.15	−4.56	−4.51	−5.10
Thr	−4.45	Ethanol	−6.32	−4.63	−4.22	−5.05
Cys	−0.76	Methanethiol	−0.22	−0.32	−0.55	−1.24
Asn	−8.14	Acetamide	−11.34	−8.54	−8.51	−9.71
Gln	−9.34	Propionamide	−10.24	−8.76	−8.63	−9.43
Asp	−6.28	Acetic acid	−10.53	–	–	−6.70
Glu	−6.59	Propionic acid	−9.35	–	–	−6.48
Lys	−4.34	*N*-butylamine	−2.03	–	–	−4.28
Arg	−12.90	*N*-propylguanidine	−10.20	–	–	−10.90

Values in kcal/mol

[a] Ref. [71]; [b] Ref. [38]; [c] Ref. [72]; [d] Ref. [70]

Table 3 Decomposition of the conformation-weighted hydration free energy determined from IEF-MST calculations for the set of *N*-acetyl-L-amino-acid amides

Residues	ΔG_{hyd}	$\Delta G_{backbone}$	$\Delta G_{side\ apolar}$	$\Delta G_{side\ polar}$	ΔG_{acetyl}	$\Delta G_{methylamine}$
Neutral						
Gly	−14.7 (1.7)	−6.7 (0.6)	−0.6 (0.3)	0.0 (0.0)	−4.8 (1.0)	−2.7 (0.9)
Ala	−11.8 (1.5)	−6.0 (1.0)	−0.2 (0.2)	0.0 (0.0)	−3.5 (0.6)	−2.0 (1.1)
Leu	−11.5 (1.2)	−5.3 (1.3)	0.1 (0.2)	0.0 (0.0)	−3.7 (0.3)	−2.6 (0.5)
Ile	−11.1 (1.0)	−4.8 (1.3)	0.1 (0.2)	0.0 (0.0)	−3.8 (0.2)	−2.7 (0.7)
Met	−13.9 (1.4)	−5.3 (1.4)	−0.5 (0.3)	−1.7 (0.2)	−3.7 (0.4)	−2.7 (0.6)
alpha	−15.0 (0.4)	−6.4 (0.1)	−0.7 (0.2)	−1.7 (0.2)	−4.0 (0.1)	−2.2 (0.1)
beta	−12.6 (0.4)	−3.9 (0.6)	−0.2 (0.1)	−1.9 (0.2)	−3.4 (0.3)	−3.2 (0.4)
Val	−11.2 (1.0)	−4.6 (1.3)	−0.1 (0.3)	0.0 (0.0)	−3.8 (0.2)	−2.7 (0.6)
Pro	−10.4 (0.5)	−3.9 (0.5)	−0.7 (0.2)	0.0 (0.0)	−3.6 (0.3)	−2.3 (0.4)
Phe	−13.1 (1.4)	−5.0 (1.3)	−1.7 (0.4)	0.0 (0.0)	−3.8 (0.3)	−2.7 (0.7)
Tyr	−17.5 (1.2)	−4.9 (1.3)	−1.5 (1.7)	−4.7 (0.1)	−3.7 (0.3)	−2.7 (0.6)
alpha	−18.6 (0.6)	−6.4 (0.5)	−1.7 (0.5)	−4.6 (0.2)	−3.9 (0.2)	−2.0 (0.1)
beta	−16.7 (0.5)	−4.0 (0.6)	−1.3 (0.6)	−4.7 (0.1)	−3.6 (0.3)	−3.1 (0.4)
Trp	−17.4 (1.4)	−5.2 (1.4)	−2.4 (0.4)	−3.7 (0.2)	−3.7 (0.4)	−2.4 (0.6)
His	−18.6 (0.9)	−4.7 (0.6)	−0.6 (0.1)	−7.6 (0.2)	−3.5 (0.2)	−2.2 (0.1)
Ser	−16.2 (1.2)	−5.4 (1.0)	−0.4 (0.1)	−4.2 (0.7)	−3.6 (0.4)	−2.6 (0.4)
Thr	−15.5 (2.1)	−4.7 (1.2)	−0.4 (0.1)	−4.1 (0.3)	−3.8 (0.4)	−2.5 (0.8)
Cys	−11.7 (1.4)	−4.6 (1.2)	−0.5 (0.1)	−0.3 (0.2)	−3.4 (0.3)	−2.9 (0.8)
Asn	−19.7 (1.8)	−5.6 (1.6)	−0.4 (0.2)	−7.8 (0.9)	−3.3 (0.3)	−2.6 (0.8)
Gln	−21.1 (1.7)	−5.6 (1.4)	−0.9 (0.3)	−8.5 (0.6)	−3.6 (0.3)	−2.6 (0.7)
alpha	−22.0 (1.2)	−6.4 (0.5)	−1.5 (0.2)	−8.5 (0.7)	−3.8 (0.2)	−2.3 (0.1)
beta	−19.7 (0.8)	−4.3 (0.8)	−0.6 (0.2)	−8.5 (0.5)	−3.3 (0.2)	−3.0 (0.7)
Asp	−17.6 (1.6)	−5.4 (1.2)	−0.6 (0.2)	−5.7 (0.3)	−3.5 (0.3)	−2.4 (0.5)
Glu	−18.5 (1.6)	−5.8 (1.2)	−0.7 (0.3)	−5.9 (0.4)	−3.7 (0.4)	−2.5 (0.5)
Arg	−24.7 (1.5)	−5.6 (1.2)	−0.7 (0.4)	−12.2 (0.5)	−3.7 (0.4)	−2.5 (0.5)
Lys	−16.2 (1.2)	−5.7 (1.2)	−0.1 (0.2)	−4.3 (0.2)	−3.7 (0.4)	−2.4 (0.5)
Charged						
Asp	−68.0 (4.3)	−11.9 (1.6)	−0.3 (0.6)	−48.7 (3.9)	−7.9 (1.3)	0.9 (0.8)
Glu	−73.5 (4.6)	−10.0 (1.7)	−1.6 (1.1)	−56.0 (5.3)	−7.4 (1.1)	1.4 (0.5)
alpha	−73.4 (1.5)	−10.6 (0.9)	−1.5 (0.2)	−55.5 (0.4)	−7.0 (0.2)	1.2 (0.2)
beta	−73.8 (0.6)	−8.8 (0.7)	−1.8 (0.1)	−56.9 (0.3)	−8.0 (0.3)	1.8 (0.5)
Arg	−76.1 (4.5)	−3.7 (1.6)	−9.8 (1.3)	−54.5 (3.1)	−1.8 (0.9)	−6.3 (1.4)
Lys	−83.1 (5.3)	−3.4 (1.4)	−18.8 (1.7)	−53.2 (4.0)	−1.6 (0.9)	−6.1 (1.2)

The conformation-weighted average and the standard deviation (in parenthesis) are shown for each hydration component. For selected residues, the values determined for α-helical and β-sheet conformations are also tabulated. Values in kcal/mol

side chain exhibits the largest variance between residues, thus reflecting the distinct chemical nature of the groups present in the side chain (see Table 3; Fig. 2). In fact, there is an excellent correlation between the fractional contributions of the side chain ($\Delta G_{sidechain}$) and the total hydration free energy of the capped amino acids, as noted in Eq. 5 (see also Fig. 3), where the independent term nicely fits the addition of the average contributions due to the backbone ($\overline{\Delta G_{backbone}}$; −5.1 kcal/mol) and capping acetyl ($\overline{\Delta G_{acetyl}}$; −3.6 kcal/mol) and methylamine ($\overline{\Delta G_{methylamine}}$; −2.5 kcal/mol) groups. Overall, these findings strongly support the separability of contributions due to both backbone and side chains for neutral residues.

$$\Delta G_{hyd} = 1.00\Delta G_{side\_chain} - 11.5 (r = 0.98) \quad (5)$$

The separability between backbone and side chain contributions is also supported by the results determined for ionized residues (see Fig. 3). Compared to neutral residues, however, one must take into account the different

hydration contribution of the backbone, which is reflected in the shift of the values determined for Asp and Glu from the ideal regression line, as well as in the larger unsigned mean deviation from the dashed line for charged residues (3.3 kcal/mol) compared to the neutral ones (1.2 kcal/mol). This trend is less apparent for Lys and Arg due to the larger length of the side chain.

The effect of the conformational flexibility on the hydration free energy is encoded in the standard deviation of the conformation-weighted fractional contributions. For neutral residues, the standard deviation of the side chain component is generally lower than 0.7 kcal/mol, but for the backbone, it is generally comprised in the range 1.2–1.6 kcal/mol, which indicates a larger conformational dependence of the backbone hydration. In fact, the PLS multivariate analysis of the distinct fractional contributions reveals a distinctive trend between the conformation of the backbone and its hydration, so that the α-helical conformation is found to be better hydrated (by around 2 kcal/mol) than the β-sheet structure (see Fig. 4). This effect is illustrated for Met, Tyr and Gln in Table 3, which reports the values determined for the structures pertaining to α-helical and β-sheet conformations. It is worth noting the drastic reduction in the standard deviation within both subgroups compared with the whole set of conformations for those residues. As noted in previous studies [73], this trend can be realized by the fact that the parallel dipoles of the adjacent peptide groups in α-helices reinforce the interaction with water molecules compared to the antiparallel dipole arrangement found in β-sheets. The other fractional components, particularly the side chain, have a minor effect in modulating the hydration free energy in α-helix and β-sheet structures.

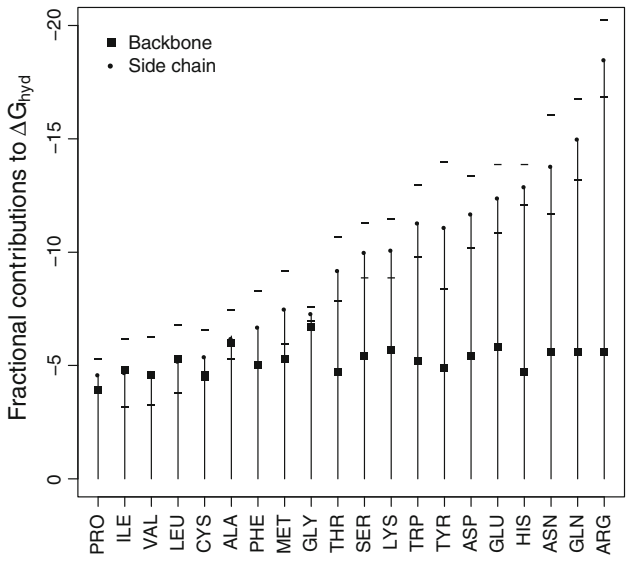

Fig. 2 Representation of the fractional contributions to the hydration free energy due to backbone and side chain (polar/apolar) fragments for the set of neutral amino acids. The results are ordered according to the total hydration free energy of the residue. The plot also includes the standard deviation around the total hydration free energy (*tick marks*). Values in kcal/mol

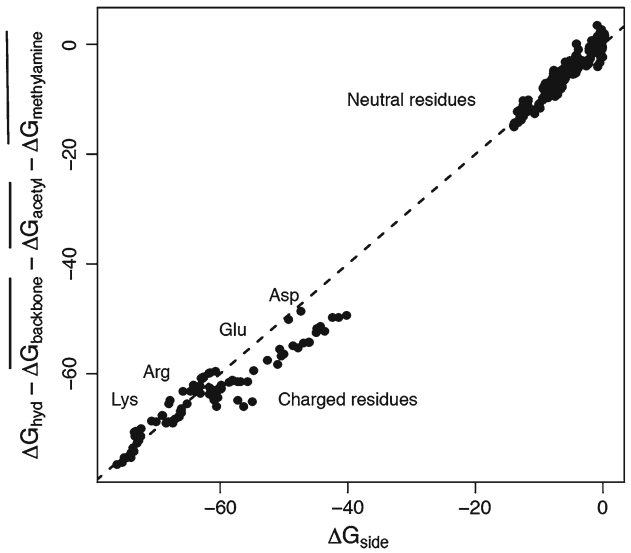

Fig. 3 Representation of the total hydration free energy of residues (corrected by subtracting the average contributions due to backbone and capping groups) versus the fractional contribution of the side chain for the whole set of conformations chosen for neutral and ionized residues. The *dashed line* represents the hypothetical behaviour where backbone, side chain and capping groups show perfect additivity. Values in kcal/mol

To gain further insight into the backbone hydration, we have extended the PLS analysis to the contributions due to the backbone atoms. The results point out that the dominant component to the hydration, hence, to the conformational dependence of the backbone, comes from the carbonyl oxygen (see Fig. 5). Thus, the fractional hydration of this atom amounts to −3.5 kcal/mol in the α-helical conformation and to −2.5 kcal/mol in the β-sheet arrangement. Then, this single atom accounts for up to 60–70 % of the backbone contribution to the hydration free energy. For the sake of comparison, the average contribution of selected side chain heteroatoms amounts to −1.5 for S in Met, −2.3 for O in Ser, −3.9 for O in Gln, −2.0 for N in Gln and −2.0 for N in His.

In summary, the analysis of the fractional hydration contributions supports the additivity of backbone and side chain contributions for neutral and charged residues, even though it seems necessary to assign different hydration contributions to the backbone of negatively charged residues (Asp, Glu). On the other hand, while the conformational preferences of the side chain are found to have little impact on the hydration contribution, the backbone exhibits a larger sensitivity to the peptide conformation, suggesting the convenience to assign distinct hydration contributions for predicting solvation patterns of α-helical and β-sheet conformations.

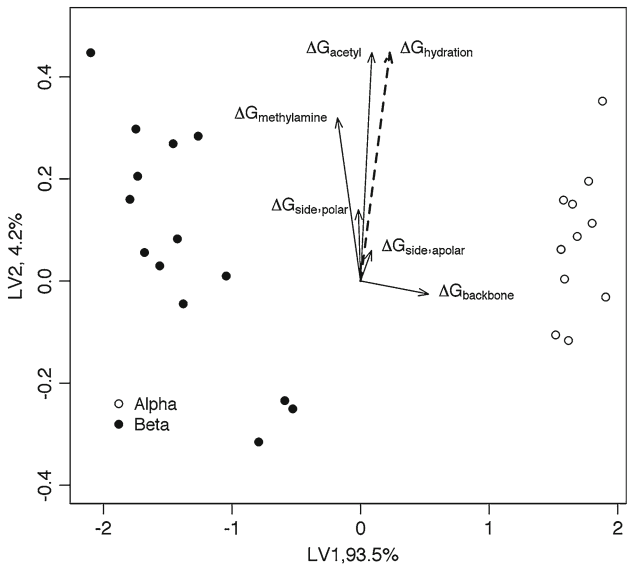

Fig. 4 PLS multivariate analysis between the total hydration free energy (ΔG_{hyd}) and its backbone ($\Delta G_{backbone}$), side chain (including polar, $\Delta G_{side,polar}$ and apolar, $\Delta G_{side,apolar}$, moieties) and capping group (acetyl: ΔG_{acetyl}; methylamine: $\Delta G_{methylamine}$) components for the 28 conformations chosen for *N*-acetyl methionine amide. The plot shows the distribution of the conformers according to the two main latent variables (LV1 and LV2) derived from the PLS analysis, which account for 93.5 and 4.2 % of the variance in hydration free energies. *Dots* denote the conformations and *arrows* correspond to fractional contributions

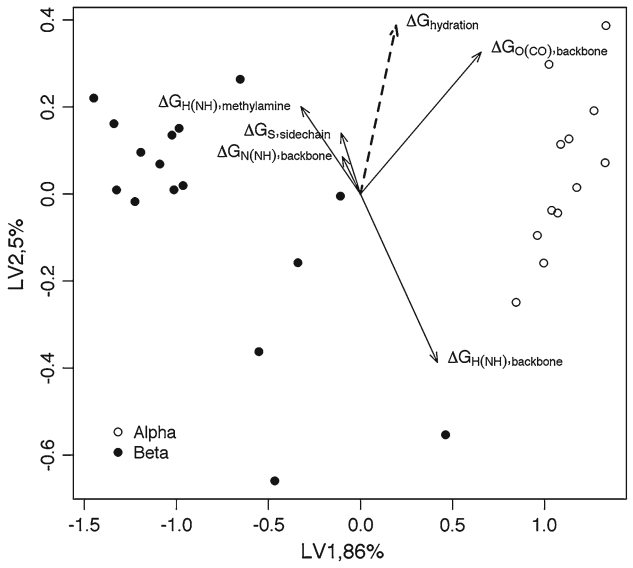

Fig. 5 PLS multivariate analysis between the total hydration free energy (ΔG_{hyd}) and the atomic contributions for the 28 conformations chosen for *N*-acetyl methionine amide. The plot shows the distribution of the conformers according to the two main latent variables (LV1 and LV2) derived from the PLS analysis, which account for 86.0 and 2.5 % of the variance in hydration free energies. *Dots* denote the rotamers and *arrows* correspond to the fractional contributions

3.3 Free energy-type decomposition of the hydration free energy

For computational purposes, it is widely accepted that the solvation free energy can be decomposed into electrostatic and non-electrostatic components. Accordingly, it is convenient to examine the conformation-weighted contributions of electrostatic, cavitation and van der Waals components to the hydration of both backbone and side chain (see Table 4).

For the backbone, the cavitation free energy has a constant contribution of 6.6 kcal/mol, which is compensated by the similar contribution of the van der Waals and electrostatic terms, which have average values of −5.9 and 6.3 kcal/mol, respectively. The electrostatic component accounts for the conformation-dependent hydration of the backbone, as it is found to favour α-helices by 1.8 kcal/mol compared to β-sheets. The different size of the side chain gives rise to a sizable difference in both van der Waals and cavitation contributions. For instance, this latter term varies from 5.2 kcal/mol for Ala to 18.8 kcal/mol for Trp. It is also worth noting that even for polar residues the van der Waals contribution is similar or even larger than the electrostatic one. Furthermore, even though the presence of a net charge makes the electrostatic term to be the dominant component to the side chain hydration, the van der Waals term still has a significant contribution to the hydration (ranging from 16 % for Asp to 31 % for Arg).

For neutral residues, the conformational dependence of the hydration free energy is associated with the electrostatic component of the backbone ($\Delta G_{back,ele}$). This is illustrated for the distinct conformations of Met in Fig. 6, which shows that the discrimination between α-helical and β-sheet conformers is dictated by the $\Delta G_{back,ele}$ component. With regard to ionic residues, the variability introduced by the electrostatic term of the backbone is overcome by the electrostatic contribution of the side chain, which is the main source for the conformational variability of the hydration free energy. In fact, the PLS analysis of the free energy components at the atomic level (data not shown) reveals that the electrostatic contribution of the carbonyl oxygen of the backbone is the main responsible for the distinction between α-helical and β-sheet conformations. Minor contributions to the conformational dependence of hydration are also due to the hydrogen atom of the backbone NH unit and to the heteroatoms in the side chain. The rest of atomic fractional contributions only play a marginal role. Overall, these findings point out that the electrostatic term is implicated in the conformational dependence of hydration, in agreement with previous studies [74, 75] that highlighted the importance of electrostatics in explaining the hydration of peptides.

Table 4 Decomposition of the conformation-weighted hydration contribution of backbone and side chain into electrostatic, cavitation and van der Waals components for the set of *N*-acetyl-L-amino-acid amides

Residues	ΔG_{hyd}	Backbone			Side chain		
		ΔG_{ele}	ΔG^{a}_{vW}	ΔG^{a}_{cav}	ΔG_{ele}	ΔG^{a}_{vW}	ΔG^{a}_{cav}
Neutral							
Gly	−7.3 (0.3)	−7.1 (0.6)	−6.7	7.0	−0.6 (0.3)	−1.4	1.5
Ala	−6.2 (1.1)	−6.3 (1.0)	−6.4	6.7	−0.4 (0.1)	−5.0	5.2
Leu	−5.2 (1.5)	−5.6 (1.3)	−6.3	6.6	−0.5 (0.2)	−15.0	15.6
Ile	−4.7 (1.5)	−5.1 (1.3)	−6.2	6.5	−0.5 (0.2)	−14.9	15.5
Met	−7.5 (1.6)	−5.6 (1.4)	−6.3	6.7	−2.7 (0.3)	−14.6	15.0
alpha	−8.8 (0.3)	−6.8 (0.1)	−6.4	6.7	−2.08 (0.2)	−14.6	15.0
beta	−5.9 (0.5)	−4.3 (0.6)	−6.3	6.6	−2.5 (0.1)	−14.6	14.8
Val	−4.7 (1.5)	−5.0 (1.3)	−6.3	6.6	−0.5 (0.3)	−11.7	12.2
Pro	−4.5 (0.7)	−4.6 (0.5)	−4.8	5.5	−1.1 (0.2)	−10.3	10.7
Phe	−6.7 (1.6)	−5.4 (1.3)	−6.3	6.7	−2.0 (0.4)	−16.6	17.0
Tyr	−11.1 (1.5)	−5.2 (1.3)	−6.3	6.7	−6.5 (0.4)	−17.9	18.1
alpha	−12.8 (1.2)	−6.7 (1.0)	−6.4	6.7	−6.6 (0.3)	−17.9	18.1
beta	−10.1 (0.7)	−4.3 (0.6)	−6.3	6.6	−6.4 (0.3)	−17.9	18.1
Trp	−11.3 (1.6)	−5.6 (1.4)	−6.4	6.7	−6.6 (0.3)	−20.4	20.9
His	−12.8 (0.9)	−5.0 (0.6)	−6.3	6.6	−8.6 (0.3)	−12.6	13.1
Ser	−10.0 (1.2)	−5.7 (1.0)	−6.3	6.7	−4.5 (0.8)	−6.5	6.4
Thr	−9.2 (1.4)	−5.1 (1.3)	−6.2	6.6	−4.4 (0.3)	−9.9	9.8
Cys	−5.4 (1.2)	−4.9 (1.2)	−6.3	6.6	−2.5 (0.2)	−6.0	7.8
Asn	−13.8 (2.2)	−6.0 (1.6)	−6.3	6.6	−8.4 (0.9)	−9.6	9.8
Gln	−14.9 (1.8)	−5.9 (1.4)	−6.3	6.7	−9.7 (0.7)	−12.9	13.4
alpha	−16.0 (1.1)	−6.8 (0.5)	−6.4	6.7	−9.9 (0.6)	−12.9	13.4
beta	−13.3 (1.7)	−4.6 (0.8)	−6.2	6.6	−9.4 (0.6)	−12.9	13.4
Arg	−18.5 (1.7)	−5.9 (1.2)	−6.4	6.7	−12.5 (0.8)	−19.8	19.4
Glu	−12.4 (1.5)	−6.1 (1.2)	−6.4	6.7	−7.1 (0.4)	−12.3	12.9
Lys	−10.1 (1.3)	−6.1 (1.2)	−6.4	6.7	−4.5 (0.2)	−17.1	17.4
Asp	−11.6 (1.6)	−5.6 (1.3)	−5.0	5.2	−6.7 (0.5)	−9.0	9.4
Charged							
Asp	−60.9 (5.0)	−12.2 (1.6)	−6.5	6.8	−50.0 (3.9)	−7.9	8.9
Glu	−67.5 (5.2)	−10.3 (1.7)	−6.3	6.7	−58.7 (5.4)	−11.2	12.3
alpha	−67.5 (1.3)	−10.9 (0.9)	−6.3	6.9	−58.2 (0.4)	−11.2	12.3
beta	−67.5 (0.8)	−9.2 (0.7)	−6.3	6.7	−59.8 (0.3)	−11.2	12.4
Arg	−68.0 (5.0)	−4.0 (1.6)	−6.3	6.7	−64.1 (3.3)	−19.8	19.8
Lys	−75.4 (5.8)	−3.7 (1.4)	−6.3	6.7	−72.4 (4.2)	−17.3	17.7

The conformation-weighted average and the standard deviation (in parenthesis) are shown for each hydration component

For selected residues, the values determined for α-helical and β-sheet conformations are also tabulated. Values in kcal/mol

[a] Standard deviation values <0.1 kcal/mol

3.4 Hydration free energy of peptide models

A direct application of the intrinsic *per residue* hydration of amino acids is the prediction of the solvation properties of peptides, since it can be assumed that the solvation of a given residue can be determined by weighting the intrinsic hydration by the fraction of the residue surface that remains exposed to the solvent [41]. This affords a simple but computationally fast approach that might be useful for estimating the solvation properties in structural genomics and other large-scale studies. In this context, the suitability of the IEF-MST fractional hydration contributions has been calibrated by predicting the hydration free energy of four pentapeptides extracted from PDB entries 1R6J, 3PUC, 1SPF and 2P5K (see Sect. 2). They were chosen to encompass α-helical and β-sheet arrangements and

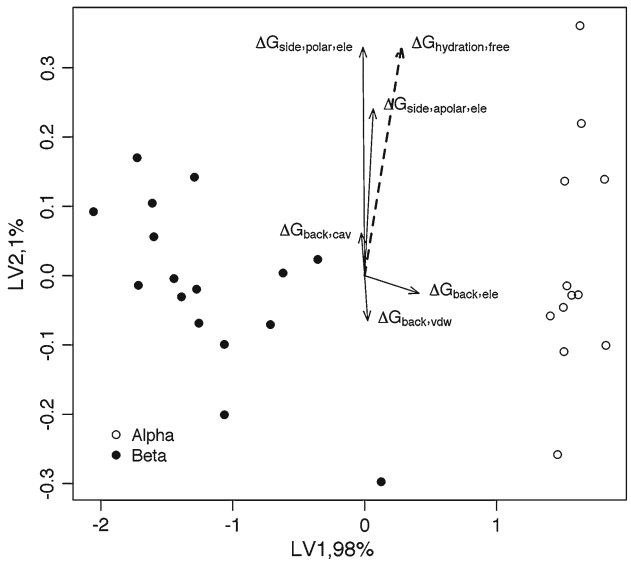

Fig. 6 PLS multivariate analysis between the total hydration free energy (ΔG_{hyd}) and the free energy components (electrostatic, cavitation and van der Waals) for the 28 conformations chosen for *N*-acetyl methionine amide. The plot shows the distribution of the conformers according to the two main latent variables (LV1 and LV2) derived from the PLS analysis, which account for 98.0 and 1.0 % of the variance in hydration free energies. *Dots* denote the rotamers and *arrows* correspond to the fractional contributions

sequences with distinct polar/apolar character. The hydration free energy determined for the four pentapeptides from IEF-MST B3LYP/6-31G(d) calculations amounts to −28.5, −32.6, −28.4 and −43.1 kcal/mol, respectively.

Two additive schemes have been used to estimate the hydration free energy from fractional hydration contributions. In the first case, the hydration free energy was determined by combining the conformation-weighted fractional hydration contributions of both backbone and side chain of amino acid residues by the fraction of the solvent-exposed surface of the fragments in the peptide and the free residue (as determined for the *N*-acetyl-L-amino acid amide; Eq. 6).where N_{res} is the number of residues in the peptide, $\Delta G_X^{i,\mathrm{free}}$ stands for the fractional contribution of the fragment (*X*: backbone, apolar side chain, polar side chain), and λ_X^i denotes the fraction of solvent-exposed surface of the fragment (Eq. 7).

$$\Delta G_{\mathrm{hyd}} = \sum_{i=1}^{N_{\mathrm{res}}} \left(\lambda^i_{\mathrm{backbone}} \Delta G^{i,\mathrm{free}}_{\mathrm{backbone}} + \lambda^i_{\substack{\mathrm{sidechain}\\ \mathrm{apolar}}} \Delta G^{i,\mathrm{free}}_{\substack{\mathrm{sidechain}\\ \mathrm{apolar}}} + \lambda^i_{\substack{\mathrm{sidechain}\\ \mathrm{polar}}} \Delta G^{i,\mathrm{free}}_{\substack{\mathrm{sidechain}\\ \mathrm{polar}}} \right) \tag{6}$$

$$\lambda_X^i = \frac{S_X^i}{S_X^{i,\mathrm{free}}} \tag{7}$$

The second approach consists of an atom-based scheme, which exploits the conformation-weighted atomic fractional contributions derived from the set of conformation chosen for residue *i* (Eq. 8).

$$\Delta G_{\mathrm{hyd}} = \sum_{i=1}^{N_{\mathrm{res}}} \sum_{j=1}^{N_{\mathrm{at}}} \zeta_j^i \Delta G_j^{i,\mathrm{free}} \tag{8}$$

where N_{at} is the number of atoms in the peptide, and ζ_j^i is the fraction of solvent-exposed surface of atom *j* in residue *i*.

The results point out that both fragment-based and atom-based partitioning schemes yield very similar hydration free energies (see Table 5). Compared to the IEF-MST B3LYP/6-31G(d) values, the two fractional schemes slightly underestimate the hydration free energy of the two β-sheet peptides (by 1–3 kcal/mol). For the two α-helix peptides, however, the fractional schemes underestimate the hydration by 9 and 6 kcal/mol for 1SPF and 2P5K, respectively.

In order to realize the different trends found for α-helical and β-sheet peptides, we have compared the free energy contributions to the hydration free energy determined from the IEF-MST partitioning scheme and from the average contributions derived for the *N*-acetyl amino acid amide compounds (see Table 6). For the sake of simplicity, cavitation and van der Waals components are given together, whereas the electrostatic component of both backbone and side chain are given separately. The results show that there is a close agreement for the non-electrostatic (cavitation + van der Waals) components upon scaling of the conformation-weighted atomic contributions of residues by the solvent-exposed surface. In fact, the RMSD determined for the separate residues in the peptide models is only 0.2 kcal/mol. It is also worth noting the resemblance between the electrostatic contributions of the side chain to the hydration of the peptides. Thus, comparison of the IEF-MST and fractional contributions for the separate residues in the peptides yields a RMSD of 0.7 kcal/mol. Nevertheless, a distinct trend is found for the backbone electrostatic term:

Table 5 Per residue decomposition of the hydration free energy of selected pentapeptide models

Peptide	IEF-MST	Fragment-based[a]	Atom-based[b]
β-sheet			
1R6J			
Val	−4.0	−3.2	−3.3
Thr	−8.8	−8.0	−8.1
Ile	−3.6	−2.9	−3.3
Thr	−8.4	−7.7	−7.6
Ile	−3.7	−2.9	−3.4
Total	−28.5	−24.6	−25.7
3PUC			
Thr	−8.7	−8.0	−8.0
Ala	−4.1	−3.8	−3.3
Ile	−4.3	−2.6	−2.9
Trp	−9.5	−8.7	−9.6
Thr	−6.0	−7.9	−7.7
Total	−32.6	−31.0	−31.5
α-helix			
1SPF			
Ile	−4.7	−3.3	−2.9
Val	−3.3	−2.4	−2.3
Gly	−6.8	−5.5	−5.3
Ala	−6.9	−5.8	−5.3
Leu	−6.7	−4.7	−4.0
Total	−28.4	−21.6	−19.8
2P5K			
Gln	−16.0	−12.7	−12.4
Ala	−5.3	−4.2	−4.6
Thr	−6.0	−7.2	−6.7
Val	−5.0	−4.1	−4.0
Ser	−10.9	−9.1	−9.4
Total	−43.1	−37.2	−37.0

Values in kcal/mol

[a] Eq. 6; [b] Eq. 8

while this component nicely fits the IEF-MST values for the β-sheet peptides, it overestimates the contribution for the α-helical ones.

It is worth noting that this effect cannot be relieved by correcting the backbone electrostatic component for the fraction of the solvent-exposed surface in the peptide and in the free N-acetyl amino acid amide, as this approach leads to an underestimation of the scaled values (−20.5 and −20.4 kcal/mol for peptides 1SPF and 2P5K, respectively) with regard to the IEF-MST ones (−27.9 and −25.8 kcal/mol; see Table 6). In fact, this effect reflects the different geometrical arrangement of the backbone in α-helical and β-sheet peptides, and specifically the occlusion of the backbone by the formation of hydrogen bonds in α-helices. At this point, note that the carbonyl oxygen of residues 1 and 2 is hydrogen-bonded to the NH unit of residue 5 and capping methylamine, respectively (Fig. 7). For these residues, the deviation between IEF-MST and fractional values of the backbone electrostatic term is, on average, 2.7 kcal/mol, which should be attributed to the shielding of the carbonyl oxygen due to intrahelical hydrogen bonds. In contrast, for residues 3–5, where the carbonyl oxygen is not directly involved in hydrogen bonds, the deviation between IEF-MST and fractional values amounts to only 0.4 kcal/mol. Overall, whereas the scaling by the solvent-exposed surface is well suited for the non-electrostatic term, the inclusion of correction factors that account for the shielding of polar groups through hydrogen bonding might be more advisable for the electrostatic component. In fact, the addition of the surface-scaled non-electrostatic terms, the hydrogen-bonded corrected backbone electrostatic term and the unscaled side chain electrostatic component yield hydration free energies very close to the IEF-MST values, as noted in Table 7.

Table 6 Decomposition of the hydration free energy of selected pentapeptide models

Peptide	Cavitation + van der Waals		Electrostatic (backbone)		Electrostatic (side chain)	
	IEF-MST	Atom-based[a]	IEF-MST	Atom-based	IEF-MST	Atom-based
β-sheet						
1R6J	3.0	2.4	−20.9	−20.6	−10.7	−10.4
3PUC	2.9	2.7	−21.2	−22.8	−14.2	−16.3
α-helix						
1SPF	3.5	3.0	−27.9	−34.1	−4.0	−2.5
2P5K	2.6	2.0	−25.8	−32.8	−20.0	−19.6

Values in kcal/mol

[a] Scaled by the fraction of solvent-exposed surface

Finally, let us note that even though MM/PBSA and MM/GBSA calculations lead to hydration free energies that overestimate the IEF-MST values obtained for the pentapeptides (see Table 7), there is a good correspondence between the hydration contributions of residues determined from the conformation-weighted fractional contributions and from MM/PB(GB)SA computations, as noted in the regression equations $y = 1.26x$ ($r = 0.92$) and $y = 1.22x$ ($r = 0.89$), where y stands for the *per residue* MM/PB(GB)SA hydration free energies and x for the fractional ones (see Table S1 and Figure S1 in Supporting Material). Keeping in mind the distinct nature of QM-SCRF and classical calculations, the correspondence found between residue contributions is remarkable and opens the way to further calibrations of the fractional hydration scheme for the widespread application to more complex polypeptides, including the analysis of less populated

Table 7 Hydration free energy of selected pentapeptide models determined from IEF-MST B3LYP/6-31G(d) calculations, conformation-weighted fractional contributions, MM/GBSA and MM/PBSA methods

Peptide	IEF-MST	Atom-based[a]	MM/GBSA	MM/PBSA
β-sheet				
1R6J	−28.5	−28.6	−37.1	−37.0
3PUC	−32.6	−36.5	−48.4	−47.3
α-helix				
1SPF	−28.4	−28.2	−30.8	−27.4
2P5K	−43.1	−45.1	−56.0	−53.5

Values in kcal/mol

[a] Derived by adding the surface-scaled non-electrostatic terms, the hydrogen-bonded corrected backbone electrostatic term and the unscaled side chain electrostatic component

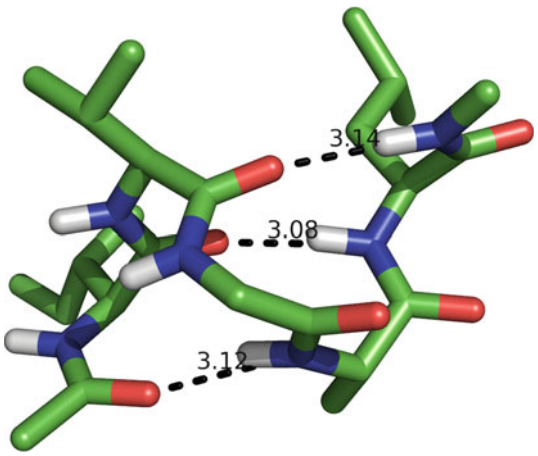

Fig. 7 Representation of the model peptide taken from PDB entry 1SPF. The *dashed lines* represent the hydrogen bonds formed between (1) the carbonyl oxygen of the capping acetyl unit with the NH of residue 4, (2) the carbonyl oxygen of residue 1 with the NH unit of residue 5 and (3) the carbonyl oxygen of residue 2 and the NH unit of capping methylamine

main-chain conformations [76, 77] and nonadditivity effects due to neighbouring residues [78, 79].

4 Conclusions

The IEF-MST formalism has been applied to investigate hydration using a dataset that comprises the most representative backbone-dependent conformational preferences of amino acids and to dissect the hydration free energy into fractional contributions. The results support the additivity of fragmental contributions determined for backbone and side chains for neutral residues. The hydration free energy of the backbone has a dominant contribution to the total hydration energy with an average contribution of −5.1 kcal/mol, though it also has a significant dependence on conformation. Thus, hydration of the backbone in α-helix is found to be 1 kcal/mol more favourable than in β-sheets, making it convenient to assign distinct fractional contributions to these structures. On the other hand, the different physicochemical properties of the side chain is reflected in hydration free energies of the side chain ranging from 0.1 kcal/mol in Leu or Ile to −12.9 in the neutral Arg. For ionic residues, however, the permanent charge in the side chain leads to nonadditive effects in the fragmental contributions, which makes it necessary to assign specific charge-dependent contributions to the backbone.

The availability of conformational-weighted fragmental contributions at the residue level permits to devise fast strategies for estimating the solvation properties of peptides and proteins in large-scale genomic studies. The approach examined here relies on three main features: (1) the partitioning of the hydration free energy between backbone and side chain, (2) the distinction between fractional contributions for the backbone representative of distinct secondary structure conformations and (3) the correction of the intrinsic hydration components by the fraction of solvent accessibility for the non-electrostatic term and by specific (hydrogen bond) factors for the electrostatic term. The results obtained for a set of pentapeptides are encouraging and support the computational efficiency of this simple strategy compared to MM/PB(GB)SA calculations. Future studies will be performed to evaluate the implementation of additional refinements in the formalism, mainly targeting the electrostatic response in the backbone, and to calibrate the range of applicability of the fractional-based solvation formalism.

Acknowledgments Dr. Carles Curutchet is acknowledged for his assistance. This work was supported by the Spanish Ministerio de Innovación y Ciencia (SAF2011-27642 and "Juan de la Cierva" contract granted to JMC) and the Generalitat de Catalunya (2009-SGR00298 and XRQTC). Computational facilities provided by the Centre de Supercomputació de Catalunya are acknowledged.

References

1. Pliska V, Schmidt M, Fauchere JL (1981) Partition coefficients of amino acids and hydrophobic parameters π of their side-chains as measured by thin-layer chromatography. J Chromatogr 216: 79–92
2. Fauchere JL, Pliska V (1983) Hydrophobic parameters-π of amino-acid side-chains from the partitioning of N-acetyl-amino-acid amides. Eur J Med Chem 18:369–375
3. Kim A, Szoka FC (1992) Amino-acid side-chain contributions to free energy of transfer of tripeptides from water to Octanol. Pharm Res 9:504–514
4. Radzicka A, Wolfenden R (1988) Comparing the polarities of the amino acids—side-chain distribution coefficients between the vapor phase cyclohexane, 1-octanol and neutral aqueous solution. Biochemistry 27:1664–1670

5. Radzicka A, Pedersen L, Wolfenden R (1988) Influences of solvent water on protein folding—free energies of solvation of cis and trans peptides are nearly identical. Biochemistry 27: 4538–4541
6. Wimley WC, Creamer TP, White SH (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry 35:5109–5124
7. Auton M, Bolen DW (2004) Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. Biochemistry 43:1329–1342
8. Shaytan AK, Shaitan KV, Khokhlov AR (2009) Solvent accessible surface area of amino acid residues in globular proteins: correlation of apparent transfer free energies with experimental hydrophobicity scales. Biomacromolecules 10:1224–1237
9. Moon CP, Fleming KG (2011) Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. Proc Natl Acad Sci 108:10174–10177
10. Gu W, Rahi SJ, Helms V (2004) Solvation free energies and transfer free energies for amino acids from hydrophobic solution to water solution from a very simple residue model. J Phys Chem B 108:5806–5814
11. Villa A, Mark AE (2002) Calculation of the free energy of solvation for neutral analogs of amino acid side chains. J Comp Chem 23:548–553
12. Shirts MR, Pitera JW, Swope WC, Pande VS (2003) Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins. J Chem Phys 119:5740–5761
13. Deng YQ, Roux B (2004) Hydration of amino acid side chains: nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. J Phys Chem B 108:16567–16576
14. Ben-Naim A, Marcus Y (1984) Solvation thermodynamics of non-ionic solutes. J Chem Phys 81:2016–2027
15. Tomasi J, Persico M (1994) Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. Chem Rev 94:2027–2094
16. Gao J (1996) Hybrid quantum and molecular mechanical simulations: an alternative avenue to solvent effects in organic chemistry. Acc Chem Res 29:298–305
17. Bashford D, Case DA (2000) Generalized born models of macromolecular solvation effects. Ann Rev Phys Chem 51:129–152
18. Orozco M, Luque FJ (2000) Theoretical methods for the description of the solvent effect in biomolecular systems. Chem Rev 100:4187–4226
19. Prabhu N, Sharp K (2006) Protein-solvent interactions. Chem Rev 106:1616–1623
20. Cramer CJ, Truhlar DG (1999) Implicit solvation models: equilibria structure spectra and dynamics. Chem Rev 99:2161–2200
21. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105:2999–3093
22. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. J Phys Chem B 113:4501–4507
23. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) The SAMPL2 blind prediction challenge: introduction and overview. J Comput Aided Mol Des 24:259–279
24. Soteras I, Forti F, Orozco M, Luque FJ (2009) Performance of the IEF-MST solvation continuum model in a blind test prediction of hydration free energies. J Phys Chem B 113: 9330–9334
25. Marenich AV, Cramer CJ, Truhlar DG (2009) Performance of SM6, SM8, and SMD on the SAMPL1 test set for the prediction of small-molecule solvation free energies. J Phys Chem B 113:4538–4543
26. Klamt A, Eckert F, Diedenhofen M (2009) Prediction of the free energy of hydration of a challenging set of pesticide-like compounds. J Phys Chem B 113:4508–4510
27. Soteras I, Orozco M, Luque FJ (2010) Performance of the IEF-MST solvation continuum model in the SAMPL2 blind test prediction of hydration and tautomerization free energies. J Comput-Aided Mol Des 24:281–291
28. Ribeiro RF, Marenich AV, Cramer CJ, Truhlar DG (2010) Prediction of SAMPL2 aqueous solvation free energies and tautomeric ratios using the SM8, SM8AD, and SMD solvation models. J Comput-Aided Mol Des 24:317–333
29. Klamt A, Diedenhofen M (2010) Blind prediction test of free energies of hydration with COSMO-RS. J Comput Aided Mol Des 24:357–360
30. Jones S, Thornton JM (1997) Analysis of protein–protein interaction sites using surface patches. J Mol Biol 272:121–132
31. Vonheijne G (1994) Membrane-proteins—from sequence to structure. Annu Rev Biophys Biomol Struct 23:167–192
32. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. J Mol Biol 324:105–121
33. Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding and binding. Nature 319:199–203
34. Eisenberg D, Wilcox W, McLachlan AD (1986) Hydrophobicity and amphiphilicity in protein structure. J Cell Biochem 31:11–17
35. Sippl MJ (1993) Boltzmann principle knowledge-based mean fields and protein folding—an approach to the computational determination of protein structures. J Comput Aided Mol Des 7:473–501
36. Jackson RM, Sternberg MJE (1995) A continuum model for protein–protein interactions: application to the docking problem. J Mol Biol 250:258–275
37. Lazaridis T, Karplus M (1999) Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J Mol Biol 288:477–487
38. Chang J, Lenhoff AM, Sandler SI (2007) Solvation free energy of amino acids and side-chain analogues. J Phys Chem B 111:2098–2106
39. Morreale A, de la Cruz X, Meyer T, Gelpi JL, Luque FJ, Orozco M (2005) Partition of protein solvation into group contributions from molecular dynamics simulations. Proteins-Struct Funct Bioinf 58:101–109
40. Morreale A, Gelpi JL, Luque FJ, Orozco M (2003) Continuum and discrete calculation of fractional contributions to solvation free energy. J Comput Chem 24:1610–1623
41. Talavera D, Morreale A, Meyer T, Hospital A, Ferrer-Costa C, Gelpí JL, de la Cruz X, Soliva R, Luque FJ, Orozco M (2006) A fast method for the determination of fractional contributions to solvation in proteins. Prot Sci 15:2525–2533
42. Arab S, Sadeghi M, Eslahchi C, Pezeshk H, Sheari A (2010) A pairwise residue contact area-based mean force potential for discrimination of native protein structure. BMC Bioinf 11:16
43. Giesen DJ, Chamber CC, Cramer CJ, Truhlar DG (1997) What controls partitioning of the nucleic acid bases between chloroform and water? J Phys Chem B 101:5084–5088
44. Hawkins GD, Cramer CJ, Truhlar DG (1998) Universal quantum mechanical model for solvation free energies based on gas-phase geometries. J Phys Chem B 102:3257–3271
45. Hornig M, Klamt A (2005) COSMOfrag: a novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry. J Chem Inf Model 45:1169–1177
46. Thormann M, Klamt A, Hornig M, Almstetter M (2006) Cosmosim: bioisosteric similarity based on cosmo-RS σ profiles. J Chem Inf Model 46:1040–1053
47. Luque FJ, Bofill JM, Orozco M (1995) New strategies to incorporate the solvent polarization in self-consistent reaction field and

free-energy perturbation simulations. J Chem Phys 103: 10183–10191
48. Luque FJ, Barril X, Orozco M (1999) Fractional description of free energies of solvation. J Comput Aided Mol Des 13:139–152
49. Muñoz J, Barril X, Hernandez B, Orozco M, Luque FJ (2002) Hydrophobic similarity between molecules: a MST-based hydrophobic similarity index. J Comput Chem 23:554–563
50. Muñoz-Muriedas J, Perspicace S, Bech N, Guccione S, Orozco M, Luque FJ (2005) Hydrophobic molecular similarity from MSDT fractional contributions to the octanol/water partition coefficient. J Comput-Aided Mol Des 19:401–419
51. Luque FJ, Curutchet C, Munoz-Muriedas J, Bidon-Chanal A, Soteras I, Morreale A, Gelpi JL, Orozco M (2003) Continuum solvation models: dissecting the free energy of solvation. J Comput Chem 5:3827–3836
52. Dunbrack JRL, Karplus M (1994) Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Nat Struct Biol 1:334–340
53. Dunbrack JRL, Karplus M (1993) Backbone-dependent rotamer library for proteins application to side-chain prediction. J Mol Biol 230:543–574
54. Guha R, Howard MT, Hutchison GR, Murray-Rust P, Rzepa H, Steinbeck C, Wegner J, Willighagen EL (2006) The blue obelisk—interoperability in chemical informatics. J Chem Inf Model 46:991–998
55. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
56. Johansson J, Szyperski T, Curstedt T, Wuthrich K (1994) The NMR structure of the pulmonary surfactant-associated polypeptide SP-C in an apolar solvent contains a valyl-rich alpha-helix. Biochemistry 33:6015–6023
57. Garnett JA, Baumberg S, Stockley PG, Phillips SE (2007) A high-resolution structure of the DNA-binding domain or AhrC, the arginine repressor/activator protein fro *Bacillus subtilis*. Acta Crystallogr Sect F 63:914–917
58. Kang BS, Devedjiev Y, Derewenda U, Derewenda ZS (2004) The PDZ2 domain of syntenin at ultra-high resolution: bridging the gap between macromolecular and small molecule crystallography. J Mol Biol 338:483–493
59. Sauer F, Wilmanns M. doi:10.2210/pdb3puc/pdb
60. Curutchet C, Orozco M, Luque FJ (2001) Solvation in octanol: parametrization of the continuum MST model. J Comput Chem 22:1180–1193
61. Soteras I, Curutchet C, Bidon-Chanal A, Orozco M, Luque FJ (2005) Extension of the MST model to the IEF formalism: HF and B3LYP parametrizations. J Mol Struct Theochem 727:29–40
62. Cancès E, Mennucci B, Tomasi J (1997) A new integral equation formalism for the polarizable continuum model: theoretical background and applications to isotropic and anisotropic dielectrics. J Chem Phys 107:3032–3041
63. Pierotti RA (1976) Scaled particle theory of aqueous and nonaqueous solutions. Chem Rev 76:717–726
64. Gaussian 03 Revision C02 Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JJA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian Inc, Wallingford CT
65. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts BP, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh HJ, Cui G, Roe DR, Mathews DH, Seetin MG, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2010) AMBER 11
66. Massart DL, Vandeginste BGM, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J (1998) In handbook of chemometrics and qualimetrics. Elsevier Science, Amsterdam
67. R-Development Core Team (2011) R: a language and environment for statistical computing. http://www.r-project.org (accessed 2011)
68. Deelman E, Gannon D, Shields M, Taylor I (2009) Workflows and e-science: an overview of workflow system features and capabilities. Future Gener Comput Syst 25:528–540
69. Altintas I, Berkley C, Jaeger E, Jones M, Ludäscher B, Mock S (2004) In Kepler: an extensible system for design and execution of scientific workflows. In: Proceedings of the international conference on scientific and statistical database management SSDBM, vol 16, pp 423–424
70. Wolfenden R, Andersson L, Cullis PM, Southgate CCB (1981) Affinities of amino acid side chains for solvent water. Biochemistry 20:849–855
71. Smith BJ (1999) Solvation parameters for amino acids. J Comput Chem 20:428–442
72. Shirts MR, Pande VS (2005) Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. J Chem Phys 122:134508
73. Avbelj F (2000) Amino acid conformational preferences and solvation of polar backbone atoms in peptides and proteins. J Mol Biol 300:1335–1359
74. Lin B, Pettitt BM (2011) Electrostatic solvation free energy of amino acid side chain analogs: implications for the validity of electrostatic linear response in water. J Comput Chem 32: 878–885
75. Baldwin RL (2002) Relation between peptide backbone solvation and the energetics of peptide hydrogen bonds. Biophys Chem 101:203–210
76. Gould IR, Cornell WD, Hillier IH (1994) A quantum mechanical investigation of the conformational energetics of the alanine and glycine dipeptides in the gas phase and in aqueous solution. J Am Chem Soc 116:9250–9256
77. Beachy MD, Chasman D, Murphy RB, Halgren TA, Friesner RA (1997) Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. J Am Chem Soc 119:5908–5920
78. Staritzbichler R, Gu W, Helms V (2005) Are solvation free energies of homogeneous helical peptides additive? J Phys Chem B 109:19000–19007
79. Avbelj F, Baldwin RL (2009) Origin of the change in solvation enthalpy of the peptide group when neighboring peptide groups are added. Proc Natl Acad Sci USA 106:3137–3141