2017

Theoretical pKa prediction of the α-phosphate moiety of uridine 5′-diphosphate-GlcNAc

Bhavani Prasad Vipperla
University of Wollongong, bpv974@uowmail.edu.au

Thomas M. Griffiths
University of Wollongong, tmg994@uowmail.edu.au

Xingyong Wang
University of Wollongong, xingyong@uow.edu.au

Haibo Yu
University of Wollongong, hyu@uow.edu.au

Publication Details

Vipperla, B., Griffiths, T. M., Wang, X. & Yu, H. (2017). Theoretical pK_a prediction of the α-phosphate moiety of uridine 5′-diphosphate-GlcNAc. Chemical Physics Letters, 667 220-225.

Theoretical pKa prediction of the α-phosphate moiety of uridine 5′-diphosphate-GlcNAc

Abstract
The pK_a value of the α-phosphate moiety of uridine 5'-diphosphate-GlcNAc (UDP-GlcNAc) has been successfully calculated using density functional theory methods in conjunction with the Polarizable Continuum Models. Theoretical methods were benchmarked over a dataset comprising of alkyl phosphates. B3LYP/6-31+G(d,p) calculations using SMD solvation model provide excellent agreement with the experimental data. The predicted pK_a for UDP-GlcNAc is consistent with most recent NMR studies but much higher than what it has long been thought to be. The importance of this study is evident that the predicted pK_a for UDP-GlcNAc supports its potential role as a catalytic base in the substrate-assisted biocatalysis.

Keywords
moiety, α-phosphate, uridine, prediction, 5′-diphosphate-glcnac, pka, theoretical

Disciplines
Medicine and Health Sciences

Publication Details
Vipperla, B., Griffiths, T. M., Wang, X. & Yu, H. (2017). Theoretical pK_a prediction of the α-phosphate moiety of uridine 5′-diphosphate-GlcNAc. Chemical Physics Letters, 667 220-225.

Theoretical pK_a prediction of the α-phosphate moiety of uridine 5'-diphosphate-GlcNAc

Bhavaniprasad Vipperla[a], Thomas M Griffiths[a], Xingyong Wang[a,b] and Haibo Yu*[a,b]

[a] School of Chemistry, University of Wollongong, NSW 2522, Australia
[b] Illawarra Health and Medical Research Institute, Wollongong, NSW 2522, Australia
* Corresponding author: hyu@uow.edu.au

Abstract

The pK_a value of the α-phosphate moiety of uridine 5'-diphosphate-GlcNAc (UDP-GlcNAc) has been successfully calculated using density functional theory methods in conjunction with the Polarizable Continuum Models. Theoretical methods were benchmarked over a dataset comprising of alkyl phosphates. B3LYP/6-31+G(d,p) calculations using SMD solvation model provide excellent agreement with the experimental data. The predicted pK_a for UDP-GlcNAc is consistent with most recent NMR studies but much higher than what it has long been thought to be. The importance of this study is evident that the predicted pK_a for UDP-GlcNAc supports its potential role as a catalytic base in the substrate-assisted biocatalysis.

Keywords

UDP-GlcNAc; OGT; Gaussian; pK_a; Density functional theory; SMD

1. Introduction

The modification of serines and threonines on nuclear and cytoplasmic proteins with *O*-linked β-D-N-acetylglucosamine (*O*-GlcNAc) is known as *O*-GlcNAcylation[1,2]. Two enzymes are mainly involved in the regulation of *O*-GlcNAcylation: a glycoside transferase called *O*-GlcNAc transferase (OGT) and a glycoside hydrolase called *O*-GlcNAcase (OGA)[3-5]. OGT installs *O*-GlcNAc from the donor UDP-GlcNAc at the sites of modification and OGA removes the modification[6]. *O*-GlcNAc modification is associated with various biological processes including transcription and translation. Apart from cell signalling, faulty regulation of *O*-GlcNAc may also be involved in diseases such as diabetes mellitus, neurodegenerative diseases and cancers[7,8]. Recent studies have also reported the importance of *O*-GlcNAc signalling in the immune system[9].

Considering the importance of UDP-GlcNAc and its key role in acting as the substrate for several enzymes including OGT[3-5,10], UDP-GlcNAc 2-epimerase[11] and UDP-GlcNAc enolpyruvyl transferase (MurA)[12], an accurate prediction of the structure and reactivities of UDP-GlcNAc and its analogues could prove to be vital in understanding many biochemical reactions. Particularly, the pK_a of the α-phosphate in UDP-GlcNAc has been debated in the literature[5,13]. It has long been thought to be around 2-3 in both UDP and UDP-GlcNAc[10,13]. Based on this, a substrate-assisted catalytic mechanism in which UDP-GlcNAc functions as a general base to deprotonate the Serine side chain in the substrate in OGT has been explicitly ruled out due to the expected unfavourable energetics[4,5,10] (Figure S1b). Only recently, Jancan and Macnaughtan have first reported a ^{31}P-NMR titration of UDP-GlcNAc and shown that the pK_a of the α-phosphate in UDP-GlcNAc is around 6.5, which makes it suitable as a general base in OGT[14]. In contrast, the popular empirical pK_a predictor Marvin[15] predicted its pK_a to be 3.3. In order to completely understand the catalysis in such reactions, a detailed analysis of its pK_a and potential pK_a shifts in the enzymatic active site is required.

To address this problem, we aimed at establishing a rational and well-calibrated method that could predict the absolute pK_a value for UDP, UDP-GlcNAc and their analogues. Systematic benchmarking calculations on a set of alkyl phosphates with known experimental pKa values were performed to validate the theoretical models. By comparing the predicted pK_a's from various methods to their experimental values, we try to find an optimal combination of methods for gas phase geometry, implicit solvation model for solvation free energies and the proton solvation free energy[16,17]. The predicted pK_a of UDP-GlcNAc was in accordance with the literature findings, supporting its general base nature.

2. Theoretical Calculation of pK_a

Continuum solvent pK_a calculations using direct method utilize a thermodynamic cycle (Figure 1), which combines gas-phase acidity with solvation free energies obtained from various models.

[Insert Figure 1]

The directly calculated pK_as may be obtained through Eq (1)

$$\mathrm{p}K_a = \frac{\Delta G^*_{aq}}{RT \ln(10)} \quad (1)$$

ΔG^*_{aq} is defined as the difference in the free energies in solution between the acid (HA) and the conjugate base (A^-) and the free proton (H^+). For computational efficiency, ΔG^*_{aq} can be obtained through the thermodynamic cycle defined in Figure 1 using Eq. 2-5,

$$\Delta G^*_{aq} = \Delta G^*_{gas} + \Delta\Delta G^*_{solv} \quad (2)$$

$$\Delta\Delta G^*_{solv} = \Delta G^*_{solv}(H^+) + \Delta G^*_{solv}(A^-) - \Delta G^*_{solv}(HA) \quad (3)$$

$$\Delta G^o_{gas} = G^o_g(H^+) + G^o_g(A^-) - G^o_g(HA) \quad (4)$$

$$\Delta G^*_{gas} = \Delta G^o_{gas} + \Delta n_{gas} RT \ln(\tilde{R}T) \quad (5)$$

where the symbol ΔG^* denotes Gibbs free energy referred to a standard state of 1 mol L^{-1} as opposed to ΔG^o referring to a standard state of 1 atm for each species. R and $\tilde{R}$ are the gas constant in units of J mol^{-1} K^{-1} and L atm mol^{-1} K^{-1}, respectively. Often, ΔG^*_{solv} (H^+) is taken from the experimental solvation free energy of the proton. Its value has been extensively discussed in the literature[16-18] and we have used ΔG^*_{solv} (H^+) of -1112.5 kJ/mol, recommended by Tissandier et.al[19] based on their cluster-pair-based method. The two standard states for the gas-phase reaction free energies are related by Eq. (5) where G^o_{gas} (H^+) is -26.4 kJ/mol.

3. Computational Methodology

To calculate the absolute pK_a value of UDP-GlcNAc, we benchmarked with a dataset of alkyl phosphates with known experimental pK_a (Figure 2, their optimised structures are provided in Supporting Material). All the gas and solvent phase *ab initio* calculations were performed using Gaussian03[20] or Gaussian09[21]. There is no standard methodology in the literature considered best for the calculation of pK_a for this class of molecules. Hence a rigorous benchmark against experimental data is highly desirable to make reliable predictions. The complete basis set method (CBS-QB3) developed by Petersson and co-workers[22] and density functional theory (DFT) methods were used to obtain accurate gas phase energies in Eq (4). The hybrid exchange-correlation functional of Becke, Lee, Yang, and Parr (B3LYP)[23] and the hybrid functional of Zhao and Truhlar (M06-2X)[24], were used for the DFT calculations, with two standard basis sets 6-31+G(d,p) and 6-311++G(d, p). The motivation for adopting DFT calculations is that CBS-QB3 is computationally prohibitive for UDP and UDP-GlcNAc in routine calculations. After validating the theoretical methods using the dataset of compounds at varying levels of theory, the estimation of pK_a for UDP-GlcNAc was performed.

[Insert Figure 2]

Comparatively speaking, it has been shown in previous studies that the major error source in calculating pK_as according to Figure 1 is from the solvation free energy calculations[25,26]. To identify an appropriate solvation model for our calculations, the solvation free energies were computed at

recommended levels of theory using various solvation models[16,18,27]. The conductor-polarizable continuum model (CPCM)[28] was applied at the B3LYP/6-31+G(d) and HF/6-31+G(d) levels of theory with the united atom (UA) cavity models, UAKS and UAHF using Gaussian03. We have also computed solvation free energies using SMD solvation model by Truhlar and co-workers[29], which uses the Polarizable Continuum Model (PCM) with the integral equation formalism variant using UFF radii (IEFPCM) using Gaussian 09. The SMD model is optimized with both B3LYP and M06-2X functionals. Using the thermodynamic cycle (Figure 1) and various combinations of ΔG^*_{gas} and $\Delta\Delta G^*_{solv}$ values, the pK_as for each compound in its gas-phase energy minima were calculated. For solvent phase calculations with the SMD model, single point energy calculations were performed on the gas-phase optimized geometries. For the UAKS and UAHF models, geometry optimisations were carried out starting from the gas phase optimised geometries. Takano and Houk[30] have earlier demonstrated based on their benchmarking studies of CPCM models on a dataset of 70 organic molecules (30 neutral, 21 anions and 19 cations) that geometries optimized in gas phase and in water were rather similar. Considering other competitive mechanisms proposed in the literature for OGT (Figure S1), the pK_as for other potential catalytic bases including Histidine 498 (His)[4] and Aspartic acid 554 (Asp) [5,10,31] were calculated as well. In these calculations, both His and Asp were capped with acetylated N-terminus (ACE) and N-Methylamide C-terminus (CT3).

Table 1 should be used as a key to classify the level of theory used to obtain ΔG^*_{gas} and $\Delta\Delta G^*_{solv}$ values in this paper. For example, G1/S1 notation indicates the gas-phase calculations evaluated at CBS-QB3 level of theory and the solvation free energies obtained at B3LYP/6-31+G (d) level of theory using SMD model.

[Insert Table 1]

There are also several empirical pK_a prediction tools available[32], which are found to predict rather accurate pK_a values including Epik (Schördinger, New York, USA), Marvin (Chemaxon, Budapest, Hungary), ACD pK_a DB (ACDLabs, Toronto, Canada). These methods are fast and cost-effective for the primary evaluation of ionization constants. Hence, as a next step we calculated the pK_a for our dataset using the pK_a prediction tool available in the Chemaxon's Marvin interface[15], which is based on the Hammett-Taft approach[33]. Marvin estimates the pK_a, based on the sum of the partial charge increment, structure specific and polarizable increments from the ionization site-specific regression equations.

4. Source of Error and Definition of Acceptable Margin

There are various factors contributing to error during the calculation of pK_a using the direct method. Aqueous reaction free energies of deprotonation as calculated by the cycle in Figure 1 are dependent on two components, $\Delta\Delta G^*_{solv}$ and ΔG^o_{gas}. Computing gas-phase acidities is reasonably straightforward due to the absence of environmental effects[17].

Based on the studies by Ho and Coote on a dataset comprising of carbon acids, gas-phase reaction energies were found to have an uncertainty of ~4.2 kJ/mol while errors in continuum solvent

calculations to be around 4.2 and 16.7 kJ/mol from various benchmarking studies[16,19,34-36]. Takano and Houk have suggested a mean absolute deviation (MAD) of ~10 kJ/mol by comparing the aqueous solvation free energies calculated by CPCM model on a dataset of 70 organic molecules to their experimental data[30]. Based on their parameterisation of the PCM model to calculate anionic solvation free energies in an aqueous solution, Pliego and Riveros[25] have reported an error of ~5 kJ/mol.

The use of an experimental solvation free energy of the proton, i.e., $\Delta G^{*}_{solv}(H^{+}) = -1112.5$ kJ/mol, also contributes an error of around ~10 kJ/mol[34]. However, pK_a calculations by direct method with this $\Delta G^{*}_{solv}(H^{+})$ were able to reproduce experimental values within an acceptable error margin[37-40]. In comparison with various protocols used by different groups in pK_a calculations in earlier studies, an acceptable error margin for a directly calculated pK_a was defined, which should be in the vicinity of 3.5 pK_a units, a relatively large range according to the benchmarking studies of Ho and Coote[16].

5. Results and Discussion

5.1. Gas-phase Acidities

In previous benchmarking studies, where the gas-phase experimental data was not available, the CBS-QB3 ΔG^{o}_{gas} values were used as the reference[41]. The CBS-QB3 was well established to predict accurate ΔG^{o}_{gas} values, which are used in various studies reporting the calculated pK_a for benchmarking[18,22]. The results are summarized in Table S1 and Figure S2. Overall comparison suggests that the gas-phase data obtained from DFT (G2, G3, G4, G5) methods had a MAD of 4.5 - 6.5 kJ/mol compared to the results from CBS-QB3, with G5 being the closest. Addition of an extra diffusion function in DFT based calculations (G2, G4 vs. G3, G5) didn't improve the computational ΔG^{o}_{gas} accuracy significantly. We have achieved higher accuracy when comparing with previous benchmarking studies on a dataset comprising of 46 acids and 30 bases by Burk et al. where, B3LYP/6-311++G(d,p) predicted gas-phase acidities and basicities with an average accuracy of <13 kJ/mol and that of < 20 kJ/mol for B3LYP/6-31+G(d)[42]. The gas-phase calculation for UDP and UDP-GlcNAc was prohibitive with larger basis sets (G1, G2, G4) due to unaffordable computational cost.

5.2. Solvation Free Energies

All the solvent-phase calculations with SMD model were carried out on the gas-phase optimized geometries while the calculations with CPCM (UAHF and UAKS radii) models used solution optimized geometries. From the solvent-phase calculations, it was observed that each of the PCM models used would predict the solvation free energies with a considerable variation and thus influence the pK_a values significantly. Both the CPCM models (UAHF and UAKS) contributed to the highest values obtained in each case, while the smallest values were with the SMD model (Table S2 and Figure S3).

In recent study by Ho and Ertem[36], a different value of ΔG^{*}_{solv} (H^{+}) = -1093.7 kJ/mol was used for the CPCM-UAHF model, as it was the parameterized value for this particular solvation model[28]. So to see if it would improve our results any further, we calculated the solvation free energies using this parameterized value for CPCM-UAHF calculations alone. Using this value didn't improve our results, when the pK_a was calculated using these solvation free energies, the MAD increased by 2.2 pK_a units compared to our earlier results using the CPCM-UAHF model with ΔG^{*}_{solv} (H^{+}) of -1112.5 kJ/mol recommended by Tissandier et.al[19]. From these calculations it was evident that having an accurate ΔG^{*}_{solv} (H^{+}) influences the predicted pK_a significantly.

5.3. Predicated pK_a Values

pK_a values have been calculated by combining theoretical gas-phase acidities with the solvation free energies obtained from three solvent models. The calculated pK_a values with SMD model using direct method for the dataset, along with the experimental data[43,44], are presented in Table 2. The remaining pK_a values obtained by CPCM (UAHF and UAKS radii) models are provided in Tables S1 and S2 in the supplementary material.

[Insert Table 2]

Analysis of the pK_a values from Tables 2, S1 and S2 reveal that more accurate pK_a values were obtained using the SMD model, than those with the CPCM models. In case of G1/S1 and G3/S1 using SMD model, the MADs are within 1 pK_a unit. All the pK_as predicted by G5/S2 indicates a systematic underestimation with a MAD of 1.2 units. The results from G1/S1 are overestimated by a MAD value of 0.7 units. Comparatively the best results were produced using G3/S1 with a MAD value of 0.4 units (Figure 3).

[Insert Figure 3]

It is well known that the performance of the CPCM models is limited to the restricted functional groups used in the parameterization[26,28]. Although the CPCM-UAHF/UAKS models have given accurate pK_a values previously, severe stability problems and larger deviations were also reported in earlier studies and they might not suit other systems like ours[16,45]. Previously Rayne and Forest[27] have demonstrated that CPCM-UAHF/UAKS models underestimated the solvation free energies by ~25 to 71 kJ/mol based on their parameterization of a dataset made up of perfluorinated alkyl compounds. Our results also show that these two models lead to relatively larger errors regardless of the gas-phase theories used. These larger deviations might be due to how these solvent models have been parameterized to a particular dataset.

From the benchmarking studies, it is evident that the SMD model gives best pK_a values of alkyl phosphates which might result from the use of an accurate atomic radius and the fact that most studied compounds have similar functional groups[26].

5.4. pK_a Calculations with Marvin and PM6 for the Dataset

Marvin was used to predict the pK_a for the dataset using the inbuilt pK_a prediction tool plugin. The predicted results were very accurate with a MAD of 0.5 (Table 2), which was better than a few of the

ab initio models we have used. Interestingly though major part of the MAD was contributed by phosphopyruvic acid, with an absolute deviation of 2.7 pK_a units. These results were also in agreement with the previous comparative studies by Balogh et al[46]. According to their comparative evaluation of various empirical pK_a prediction tools, Marvin was found to outperform the other similar tools in terms of accuracy, which was one of the main factors for choosing it for our benchmarking studies.

Recently it has been shown that PM6 provides a satisfactory prediction for a dataset comprising of pyridines, alcohols, phenols, benzoic acids, carboxylic acids, and phenols with the so-called isodesmic model[47] (Figure S4). We carried out additional pK_a calculations with the isodesmic and direct methods using PM6 level of theory. The results show that, though PM6/SMD provided a satisfactory prediction for the isodesmic method with a MAD of 0.7 pK_a units, it failed to provide a reasonable prediction with the direct method, which resulted in a MAD of 14.8 pK_a units (See Supplementary data).

5.5. pK_a Calculation of UDP and UDP-GlcNAc

As the computational method with acceptable accuracy has been validated, we used G3/S1 along with Marvin to predict the pK_a of UDP and UDP-GlcNAc (Table 3). The pK_a was calculated for the deprotonation of both α-phosphate and β-phosphate moieties of UDP. The predicted pK_a values using G3/S1 for UDP were found to be close to the experimental data[14], with each exhibiting a deviation of 0.6 and 1.0 units respectively, whereas Marvin significantly underestimated the pK_a for both with an absolute deviation of 3.3 and 4.7 respectively.

[Insert Table 3]

The G3/S1 combination predicted the pK_a of UDP-GlcNAc to be around 6.9, which is very close to the experimental value by Jancan and Macnaughtan[14] with an absolute deviation of 0.3 units. Marvin predicted it to be 3.0, with an absolute deviation of 3.6. The significant underestimation of Marvin might be due to the mid range effect reported by Balogh et al.[46] According to their study, empirical tools including Marvin have a limitation for compounds with a mid-range pK_a value (pK_a around 6) and for weak acids (pK_a around 12). They have reported a mean absolute error of 2 pK_a units for compounds at pK_a ~6 which is consistent with our findings. This also highlighted the necessity to use validated quantum chemistry methods to study the complex chemistry of phosphate.

5.6. pK_a Calculation of Amino Acids as the Potential Alternative Base in OGT

Next we tried to verify, if this method could be extended to provide an accurate pK_a for the side chains of amino acids that potentially serve as a general base in OGT. His and Asp were considered, as they have also been proposed to be the catalytic base in the catalysis of OGT[4,5,10,31] (Figure S1a and S1c). The reaction used for the calculation of pK_a of His and Asp can be seen in Figure 4.

[Insert Figure 4]

The pK_a of His calculated using G3/S1 model is comparable to the experimental value (6.0)[48] with an absolute deviation of 0.8 units. This is a better agreement comparing to the study by Sastre et al in

which an isodesmic reaction scheme with PM6 was used[49]. Whereas for Asp the calculated pK_a was not so accurate with a significant deviation of 3.0 units from the experimental value (3.7)[48], which is 1.1 units larger than that reported in Ref [49] (Table 4). While the pK_a calculation for His is encouraging, the Asp result shows that a specific model cannot be universally applied to different set of compounds. From the above benchmarking studies, we believe that despite a lack of universality, the G3/S1 (B3LYP/6-31+G(d,p)//B3LYP/6-31+G(d) (SMD)) model can predict accurate pK_a for phosphate related compounds having a similar functional group.

[Insert Table 4]

6. Conclusions

The present benchmarking study assesses the ability of various gas-phase and solvent-phase models to reproduce the experimental pK_a of a series of alkyl-phosphates. The CBS-QB3 and DFT methods were investigated along with various solvation models. The calculated gas-phase deprotonation energies and free energies of solvation were compared, and the errors associated with each method were estimated. It was found that B3LYP/6-31+G(d,p) together with the SMD solvation model is more accurate compared to CPCM-UAHF and CPCM-UAKS models for our dataset. This model was used to evaluate the absolute pK_a value of UDP-GlcNAc. It successfully reproduced the ^{31}P-NMR experimental data to within 0.3 pK_a units. Future work will be carried out to characterize the pK_a values of UPD-GlcNAc analogues and their apparent pK_a values (and thus pK_a shifts) within a protein environment[50-52], with the aim to establish the relevant catalytic mechanism in OGT.

Acknowledgements

H. Yu. is the recipient of an Australian Research Council Future Fellowship (Project number FT110100034). X. W is the recipient of the Vice-Chancellor's Postdoctoral Research Fellow at the University of Wollongong. This research was in part supported under Australian Research Council's Discovery Projects funding scheme (project number DP170101773). We also wish to acknowledge that this research was undertaken with the assistance of resources provided at the NCI National Facility systems at the Australian National University through the National Computational Merit Allocation Scheme supported by the Australian Government.

References

(1) Torres, C.-R.; Hart, G. W. *J. Biol. Chem.* **1984**, *259*, 3308.

(2) Wells, L.; Vosseller, K.; Hart, G. W. *Science.* **2001**, *291*, 2376.

(3) Schimpl, M.; Zheng, X.; Borodkin, V. S.; Blair, D. E.; Ferenbach, A. T.; Schüttelkopf, A. W.; Navratilova, I.; Aristotelous, T.; Albarbarawi, O.; Robinson, D. A.; Macnaughtan, M. A.; van Aalten, D. M. F. *Nat. Chem. Biol.* **2012**, *8*, 969.

(4) Lazarus, M. B.; Nam, Y.; Jiang, J.; Sliz, P.; Walker, S. *Nature.* **2011**, *469*, 564.

(5) Levine, Z. G.; Walker, S. *Annu. Rev. Biochem.* **2016**, *85*, 631.

(6) Hurtado-Guerrero, R.; Davies, G. J. *Curr. Opin. Chem. Biol.* **2012**, *16*, 479.

(7) Hart, G. W.; Housley, M. P.; Slawson, C. *Nature.* **2007**, *446*, 1017.

(8) Dias, W. B.; Hart, G. W. *Mol. Biosyst.* **2007**, *3*, 766.

(9) Golks, A.; Guerini, D. *EMBO Rep.* **2008**, *9*, 748.

(10) Lazarus, M. B.; Jiang, J.; Gloster, T. M.; Zandberg, W. F.; Whitworth, G. E.; Vocadlo, D. J.; Walker, S. *Nat. Chem. Biol.* **2012**, *8*, 966.

(11) Chen, S.-C.; Huang, C.-H.; Lai, S.-J.; Yang, C. S.; Hsiao, T.-H.; Lin, C.-H.; Fu, P.-K.; Ko, T.-P.; Chen, Y. *Sci. Rep.* **2016**, *6*, 23274.

(12) Skarzynski, T.; Mistry, A.; Wonacott, A.; Hutchinson, S. E.; Kelly, V. A.; Duncan, K. *Structure.* **1996**, *4*, 1465.

(13) Withers, S. G.; Davies, G. J. *Nat. Chem. Biol.* **2012**, *8*, 952.

(14) Jancan, I.; Macnaughtan, M. A. *Anal. Chim. Acta.* **2012**, *749*, 63.

(15) Szegezdi, J.; Csizmadia, F. In *Presented at the 233rd ACS NationalMeeting,* Chicago, IL, Mar 25-29, 2007; Vol. 233.

(16) Ho, J.; Coote, M. *Theor. Chem. Acc.* **2010**, *125*, 3.

(17) Ho, J. *Aust. J. Chem.* **2014**, *67*, 1441.

(18) Keith, J. A.; Carter, E. A. *J. Chem. Theory Comput.* **2012**, *8*, 3187.

(19) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys. Chem. A.* **1998**, *102*, 7787.

(20) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery Jr, J.; Vreven, T.; Kudin, K.; Burant, J. *Inc., Wallingford, CT* **2004**, *4*.

(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. *Gaussian Inc., Wallingford, CT* **2009**, *1*.

(22) Montgomery Jr, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822.

(23) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(24) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(25) Pliego, J. R.; Riveros, J. M. *Chem. Phys. Lett.* **2002**, *355*, 543.

(26) Pliego, J. R.; Riveros, J. M. *J. Phys. Chem. A.* **2002**, *106*, 7434.

(27) Rayne, S.; Forest, K. *J. Mol. Struct.: THEOCHEM.* **2010**, *949*, 60.

(28) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.

(29) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B.* **2009**, *113*, 6378.

(30) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70.

(31) Lazarus, M. B.; Jiang, J.; Kapuria, V.; Bhuiyan, T.; Janetzko, J.; Zandberg, W. F.; Vocadlo, D. J.; Herr, W.; Walker, S. *Science.* **2013**, *342*, 1235.

(32) Liao, C.; Nicklaus, M. C. *J. Chem. Inf. Model.* **2009**, *49*, 2801.

(33) Csizmadia, F.; Tsantili-Kakoulidou, A.; Panderi, I.; Darvas, F. *J. Pharm. Sci.* **1997**, *86*, 865.

(34) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B.* **2006**, *110*, 16066.

(35) Ho, J.; Coote, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 295.

(36) Ho, J.; Ertem, M. Z. *J. Phys. Chem. B.* **2016**, *120*, 1319.

(37) Namazian, M.; Zakery, M.; Noorbala, M. R.; Coote, M. L. *Chem. Phys. Lett.* **2008**, *451*, 163.

(38) Oliva, A.; Henry, B.; Ruiz-López, M. F. *Chem. Phys. Lett.* **2013**, *561–562*, 153.

(39) Psciuk, B. T.; Lord, R. L.; Munk, B. H.; Schlegel, H. B. *J. Chem. Theory Comput.* **2012**, *8*, 5107.

(40) Close, D. M.; Wardman, P. *J. Phys. Chem. A*. **2016**, *120*, 4043.

(41) Montgomery Jr, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys*. **2000**, *112*, 6532.

(42) Burk, P.; Koppel, I. A.; Koppel, I.; Leito, I.; Travnikova, O. *Chem. Phys. Lett.* **2000**, *323*, 482.

(43) Freedman, L. D.; Doak, G. O. *Chem. Rev.* **1957**, *57*, 479.

(44) McElroy, W. D.; Glass, B. *Phosphorus Metabolism*; Johns Hopkins University Press: Baltimore, 1951; Vol. 1.

(45) Śmiechowski, M. *J. Mol. Struct.* **2009**, *924–926*, 170.

(46) Balogh, G. T.; Gyarmati, B.; Nagy, B.; Molnár, L.; Keserű, G. M. *QSAR COMB SCI.* **2009**, *28*, 1148.

(47) Kromann, J. C.; Larsen, F.; Moustafa, H.; Jensen, J. H. *PeerJ Preprints*, *4*, e2075v1.

(48) Lange, N. A.; Dean, J. A. *Lange's Handbook of chemistry*; McGraw-Hill, 1979; Vol. 12.

(49) Sastre, S.; Casasnovas, R.; Munoz, F.; Frau, J. *Phys. Chem. Chem. Phys.* **2016**, *18*, 11202.

(50) Yu, H.; Griffiths, T. M. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5785.

(51) Yu, H.; Ratheal, I. M.; Artigas, P.; Roux, B. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1159.

(52) Xiao, K.; Yu, H. *Phys. Chem. Chem. Phys.* **2016**, 30305.

Table 1. Quantum chemical calculations used in computations of gas-phase and solvent-phase calculations given in shorthand notation used in the text

Gas-Phase Calculations			Solvent-Phase Calculations			
Label	Level of Theory	Basis Set	Label	Level of Theory	Basis Set	Solvation Model
G1	CBS-QB3		S1	DFT-B3LYP	6-31+G (d)	SMD
G2	DFT-B3LYP	6-311++G (d, p)	S2	DFT-M06-2X	6-31+G (d)	SMD
G3	DFT-B3LYP	6-31+G (d, p)	S3	RHF	6-31+G (d)	CPCM-UAHF
G4	DFT-M06-2X	6-311++G (d, p)	S4	DFT-B3LYP	6-31+G (d)	CPCM-UAKS
G5	DFT-M06-2X	6-31+G (d, p)				

Table 2. Calculated pK_a of the dataset with SMD model using different gas-phase energies and solvation energies

Alkyl Phosphate Name	Experi-mental pK_a	G1/S1	Abs. Dev.	G3/S1	Abs. Dev.	G5/S2	Abs. Dev.	Marvin	Abs. Dev.
Methyl Phosphate	1.5[a]	2.0	0.5	0.7	0.8	0.7	0.8	1.8	0.3
Ethyl Phosphate	1.6[a]	2.2	0.6	1.6	0.0	1.1	0.5	1.8	0.2
n-propyl Phosphate	1.8[a]	2.0	0.2	1.2	0.6	0.8	1.0	1.8	0.0
n-butyl Phosphate	1.8[a]	2.9	1.1	1.8	0.0	1.2	0.6	1.8	0.1
Dimethyl Phosphate	1.3[a]	1.7	0.4	1.0	0.3	0.3	1.0	2.0	0.7
Di-*n*-propyl Phosphate	1.6[a]	2.7	1.1	1.7	0.1	1.2	0.4	1.9	0.4
Di-*n*-Butyl Phosphate	1.7[a]	2.8	1.1	1.9	0.2	1.0	0.7	1.9	0.2
Glyceraldehyde Phosphate	2.1[b]	2.3	0.2	1.7	0.4	0.4	1.7	1.4	0.7
3-phosphoglyceric acid	1.4[b]	1.8	0.4	0.8	0.6	-0.1	1.5	1.3	0.1
Phosphopyruvic acid	3.5[b]	1.8	1.7	2.3	1.1	-0.5	4.0	0.8	2.7

[a] Ref. 43. [b] Ref. 44.

Table 3. Calculated pK_a of UDP and UDP-GlcNAc

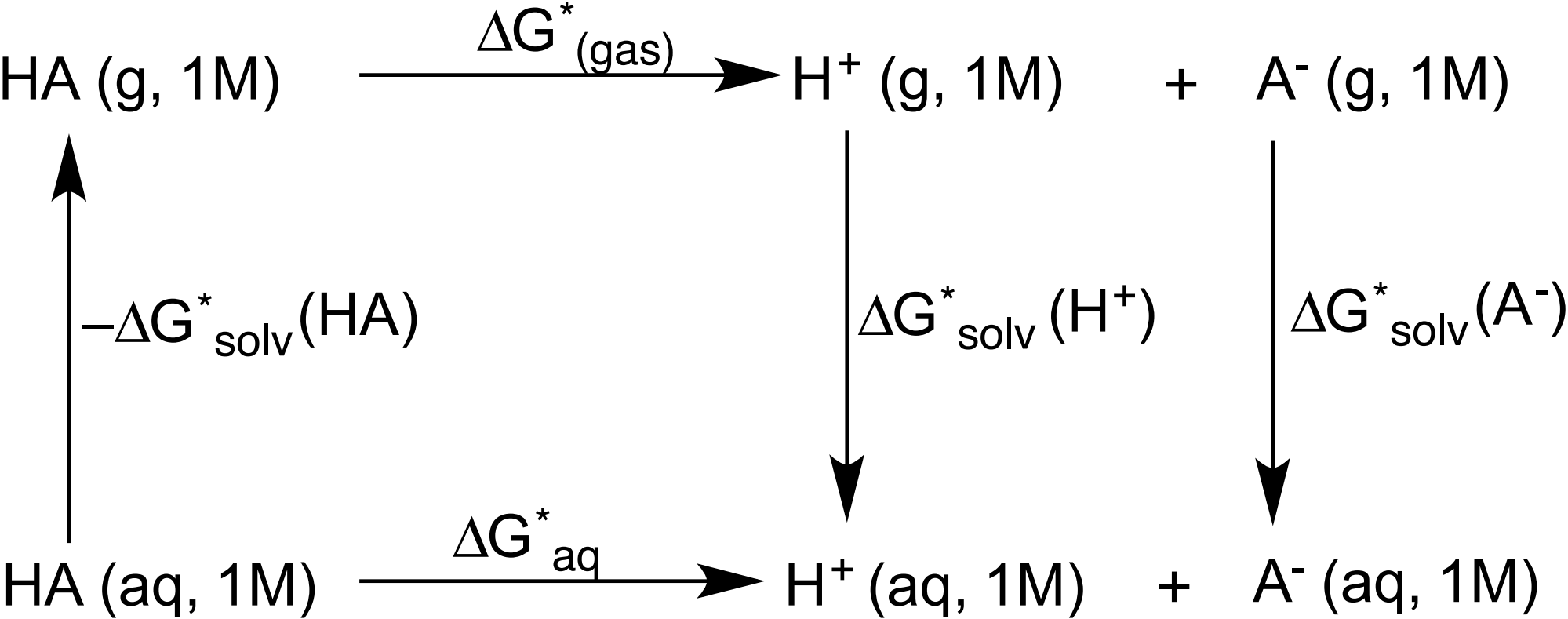
Analogue Name	Experimental pK_a	G3/S1	Abs. Dev.	Marvin	Abs. Dev.
UDP α-Phosphate	6.5[a]	5.9	0.6	3.2	3.3
UDP β-Phosphate	6.5[a]	5.5	1.0	1.8	4.7
UDP-GlcNAc	6.6[a]	6.9	0.3	3.0	3.6

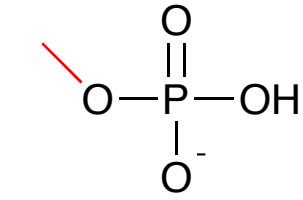
[a] Ref. 14.

Table 4. Calculated pK_a of Histidine and Aspartic Acid

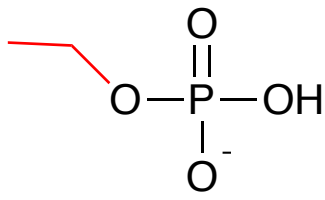
Amino acid	Experimental pK_a	G3/S1	Abs. Dev.	Theoretical pK_a reported by Sastre et al ΔpK_a [b]
Histidine	6.0[a]	6.8	0.8	0.9
Aspartic acid	3.7[a]	6.6	3.0	1.9

[a] Ref. 48. [b] Ref. 49.

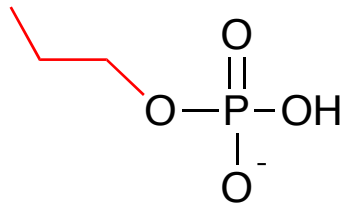

HA (g, 1M)
ΔG*(gas)
H+ (g, 1M)
+
A- (g, 1M)
–ΔG*solv(HA)
ΔG*solv (H+)
ΔG*solv(A-)
HA (aq, 1M)
ΔG*aq
H+ (aq, 1M)
+
A- (aq, 1M)



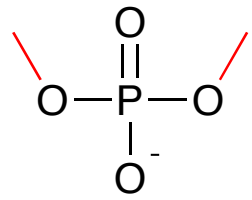
Methyl Phosphate



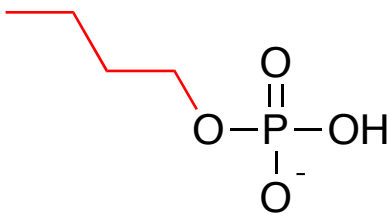
Ethyl Phosphate



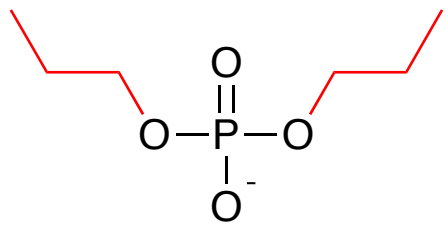
n-propyl Phosphate



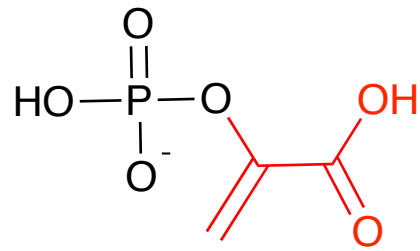
Dimethyl Phosphate



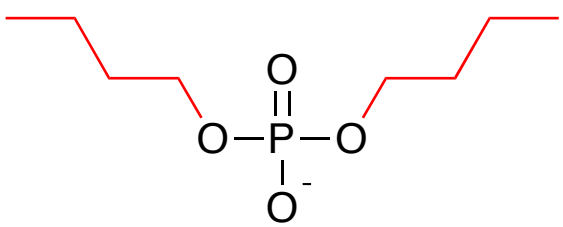
n-butyl Phosphate



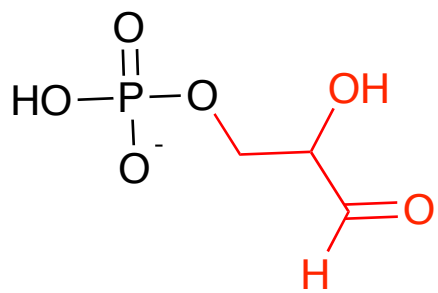
Di-n-propyl Phosphate



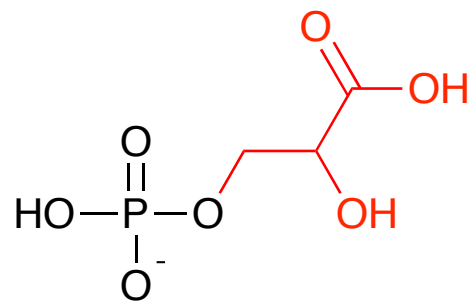
Phosphopyruvic Acid



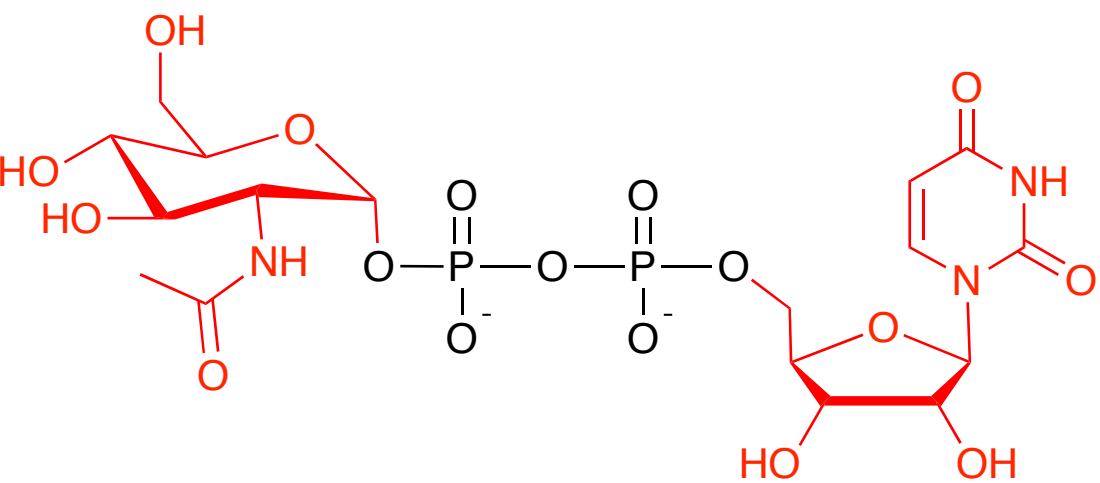
Di-n-butyl Phosphate



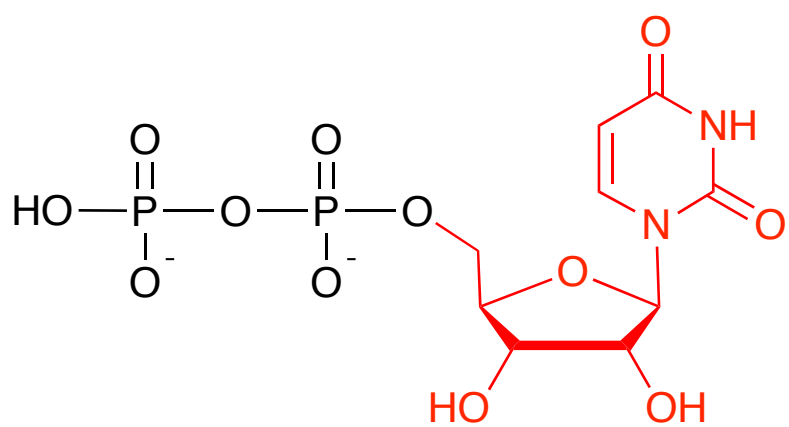
Glyceraldehyde Phosphate



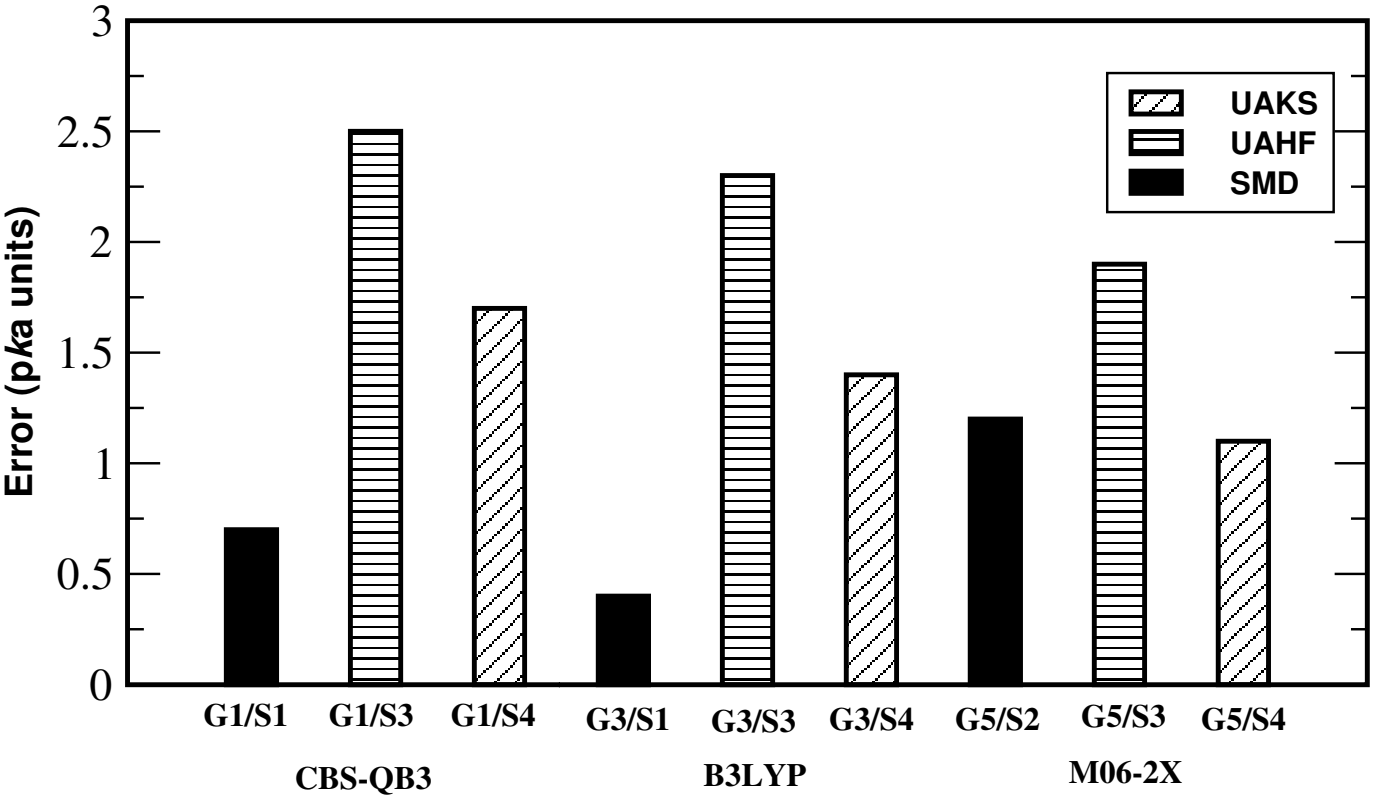
3-Phosphoglyceric Acid

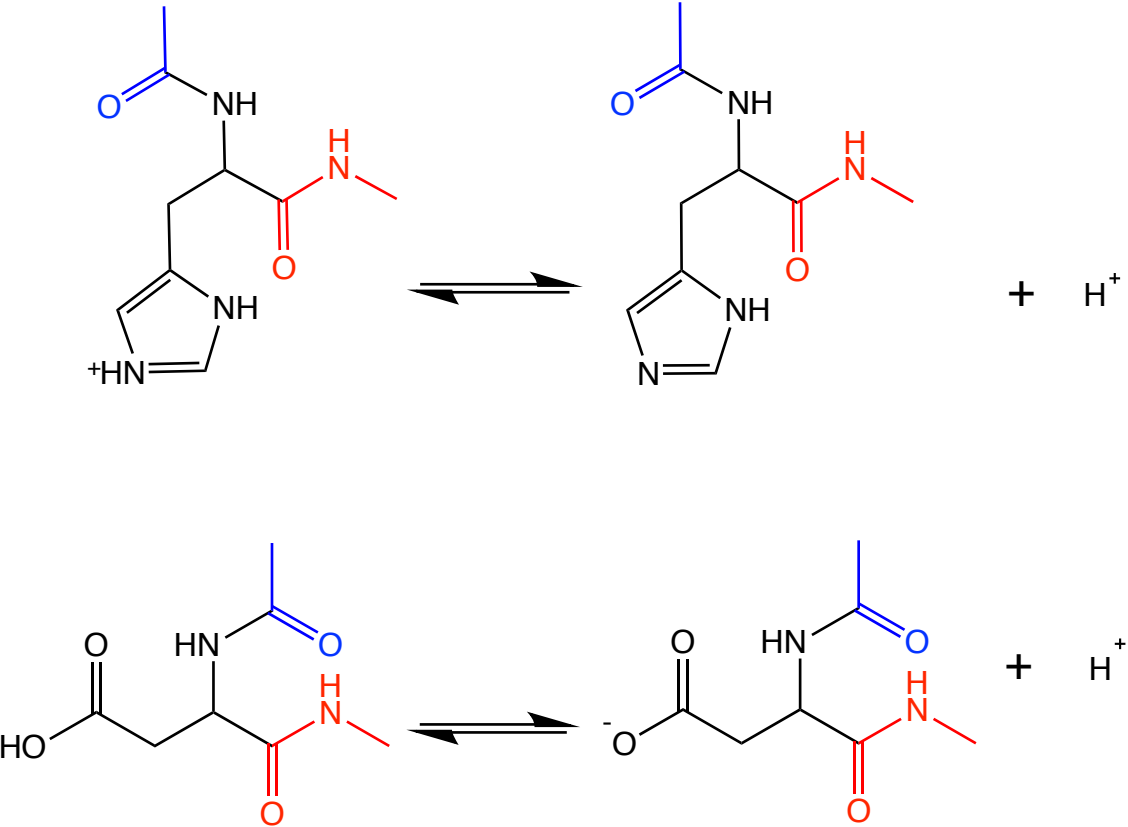


Uridine Diphosphate N-acetylglucosamine (UDP-GlcNAc)



Uridine Diphosphate (UDP)


Error (pKa units)
3
2.5
2
1.5
1
0.5
0
UAKS
UAHF
SMD
G1/S1
G1/S3
G1/S4
G3/S1
G3/S3
G3/S4
G5/S2
G5/S3
G5/S4
CBS-QB3
B3LYP
M06-2X


O
NH
H
N
O
NH
+HN
NH
N
+
H+
O
HN
O
HO
H
N
O
-O
+
H+