Published in final edited form as:

Biochemistry. 2018 April 17; 57(15): 2227–2237. doi:10.1021/acs.biochem.8b00087.

Quantifying Interactions of Nucleobase Atoms with Model Compounds for the Peptide Backbone and Glutamine, Asparagine Side Chains in Water

Xian Cheng^{1,2}, Irina A. Shkel^{2,3}, Cristen Molzahn^{2,#}, David Lambert^{2,@}, Rezwana Karim^{2,3}, and M. Thomas Record Jr.^{*,1,2,3}

¹Program in Biophysics, University of Wisconsin – Madison, Madison WI 53706

²Department of Biochemistry, University of Wisconsin – Madison, Madison WI 53706

³Department of Chemistry, University of Wisconsin – Madison, Madison WI 53706

Abstract

Alkylureas display hydrocarbon and amide groups, the primary functional groups of proteins. To obtain thermodynamic information needed to analyze interactions of amides and proteins with nucleobases and nucleic acids, we quantify preferential interactions of alkylureas with nucleobases differing in amount and composition of water-accessible surface area (ASA) by solubility assays. Using an established additive ASA-based analysis, we interpret these thermodynamic results to determine interactions of each alkylurea with five types of nucleobase unified atoms (carbonyl sp^2O , amino sp^3N , ring sp^2N , methyl sp^3C , and ring sp^2C). All alkylureas interact favorably with nucleobase sp^2 and sp^3C atoms; these interactions become more favorable with increasing alkylation of urea. Interactions with nucleobase sp^2O are most favorable for urea, less favorable for methylurea and ethylurea, and unfavorable for di-alkylated ureas. Contributions to overall alkylurea-nucleobase interactions from interactions with each nucleobase atom-type are proportional to ASA of that atom-type with proportionality constant (interaction strength) α , as observed previously for urea. α -Value trends for interactions of alkylureas with nucleobase atom-types parallel those for corresponding amide-compound atom-types, offset because nucleobase α -values are more favorable. Comparisons between ethylated and methylated ureas show interactions of amide-compound sp^3C with nucleobase sp^2/sp^3C and N atoms are favorable while amide sp^3C -nucleobase sp^2O interactions are unfavorable. Strongly-favorable interactions of urea with nucleobase sp^2O but weakly-favorable interactions with nucleobase sp^3N indicate that amide sp^2N -nucleobase sp^2O and nucleobase sp^3N -amide sp^2O hydrogen-bonding ($NH\cdots O=C$) interactions are favorable while amide sp^2N -nucleobase sp^3N interactions are unfavorable. These

*Corresponding Author: mrecord@wisc.edu.

#Current address: Department of Biochemistry, University of British Columbia, Canada

@Current address: Medical College of Wisconsin, Milwaukee, WI, 53213

Supporting Information

The supporting information is available free of charge on the ACS Publications website. It includes chemicals, experimental methods, ASA calculations, uncertainty determination and robustness tests of α -values, the supporting tables (Tables S1 – S7) and supporting figures (Figures S1 – S6).

Author Contributions

X.C., M.T.R. designed the research. X.C., C.M., D.L. and R.K. performed the experiments. X.C., M.T.R., I. S. analyzed the data. X.C. and M.T.R. wrote the paper.

favorable amide-nucleobase hydrogen bonding interactions are prevalent in specific protein-nucleotide complexes.

Graphical abstract

Introduction

A thermodynamic and molecular understanding of interactions between amide compounds and nucleobases and related aromatic compounds in water is needed in order to understand the origins of specificity and stability of binding of amide compounds and proteins to nucleic acid duplexes¹⁻⁸ and of nucleotide substrates and cofactors to enzymes and to the molecular machines of biology.⁹⁻¹¹ Insights into diverse topics including the effects of alkylureas on nucleic acid duplex formation and the nonspecific interactions that play a role in the action of amides and nucleotides as solubilizing and structure-destabilizing agents¹²⁻¹⁴ will also be obtained. Another fundamental motivation driving this research is that many of the same hybridization states of C, N and O unified atoms are presented in very different contexts on nucleobases and base analogs as compared with the amide compounds investigated previously.¹⁵⁻¹⁷ Comparison of the same interactions in these two contexts provides previously-unavailable information about the roles of C, N and O atom hybridization state and atom context in their interactions with amide molecules.

One route to obtaining this information has been through molecular dynamics simulations, which yield radial distributions of a solute and water in the vicinity of a biopolymer or another solute (model compound). Thermodynamic information about the solute-biopolymer or solute-model compound interaction is obtained by integrating these radial distributions to obtain the potential of mean force and Kirkwood-Buff integrals¹⁸⁻²⁴ or transfer free energies.²⁵ Interactions of urea with benzene²⁶ and with nucleobases and nucleic acids^{27, 28} have been simulated.

Another route, taken here, is to obtain experimental data on the thermodynamics of interactions of a solute of interest with a set of model compounds displaying subsets of the functional groups of proteins or nucleic acids, and interpret these data in terms of the interactions of the solute with the individual types of groups and unified atoms of the model compounds. Previously we used vapor pressure osmometry (VPO, for pairs of soluble

solutes) and solubility assays (where one solute is only sparingly soluble) to quantify preferential interactions (chemical potential derivatives $(\partial\mu_2/\partial m_3)_{P,T,m_2} = \mu_{23}$) of amide

compounds with one another and with aromatic hydrocarbons. These μ_{23} values can be interpreted as free energies of transfer of one solute from water to a 1 molal solution of the other solute.¹⁵ Values of μ_{23} were determined for interactions of urea and a series of alkylureas (methyl and ethylurea, dimethyl and diethyl ureas) with a series of amides and aromatic hydrocarbon compounds displaying the principal unified atoms of proteins (amide sp^2O , sp^2N , sp^2C ; aliphatic sp^3C , aromatic sp^2C).¹⁵ We observed that chemical potentials μ_2 of all amide and aromatic hydrocarbon compounds investigated are reduced by addition of any of these alkylureas (i.e. $\mu_{23} < 0$), indicating favorable interactions relative to interactions with water. Here solubility assays are used to quantify preferential interactions of alkylureas with nucleobases and base analogs.

An analysis of μ_{23} values for preferential interactions of a solute or salt with model compounds displaying functional groups of proteins or nucleic acids in water was developed and tested in recent papers.^{15-17, 29-33} Foundations of this approach are the use of water-accessible surface area (ASA) to analyze model compounds and obtain amounts of each different hybridization state of C, N, and O unified atoms, and the assumption of additivity of contributions of individual C, N and O atom interactions with the solute to μ_{23} . The additivity assumption is tested in the course of the analysis, and has been verified in all cases examined.

This analysis yields intrinsic strengths of preferential interaction of the solute or salt with unit surface area ($1 \text{ \AA}^2$) of each type of unified C, N and O atom on the model compound, relative to interactions with water.^{15-17, 29-33} These are designated as α -values; negative α -values indicate favorable preferential interactions. Justification for the use of ASA in this analysis is provided by statistical thermodynamic analysis of the solute partitioning model.^{16, 17, 29, 31-34} This approach is providing previously-unavailable quantitative information about the interactions of biochemical solutes with the types of unified atoms of proteins and nucleic acids (i.e. sp^2 and sp^3 hybridization states of C, N and O atoms in various contexts).

Guinn et al^{16, 17} determined preferential interactions of urea with model compounds displaying subsets of the principal unified atoms and contexts of proteins and nucleic acids in water by VPO and hexanol-water distribution assays. Cheng et al¹⁵ quantified preferential interactions of a series of alkylureas with amide compounds and aromatic hydrocarbons by VPO and solubility assays. ASA-based analysis of these data sets yielded α -values for interactions of urea and alkylureas with unit area of the types of unified atoms displayed on amide and aromatic hydrocarbon compounds (amide sp^2O , sp^2N , and sp^2C ; aliphatic sp^3C and aromatic sp^2C), as well as proteins (amide, carboxylate, and hydroxyl O; amide and cationic N; aromatic sp^2C ; and aliphatic sp^3C). Interactions of urea with nucleic acid atom types (heterocyclic aromatic ring $sp^2(C+N)$, methyl sp^3C , carbonyl sp^2O , amino sp^3N ; sugar sp^3C , sp^3O ; phosphate sp^2O) were also determined¹⁷. Lambert and Draper³⁵ independently obtained α -values for interactions of urea with nucleic acid base, sugar and phosphate groups. Urea is found to interact favorably with all these protein and nucleic acid atoms and groups (except cationic N of proteins), relative to interactions with water. This quantitative

information was used to interpret destabilizing effects of urea on secondary structure of DNA and RNA oligomers and to estimate the large amount of residual stacking of nucleobases in the separated strands.^{17, 35} Cheng et al¹⁵ found that interactions of amide sp^2O with alkylureas are unfavorable, becoming more unfavorable with increasing alkylation. Interactions of all ureas with sp^3C and amide sp^2N are favorable and increase in strength with increasing alkylation. Comparisons of α -values and of trends in α -values indicate favorable interactions of amide sp^2N with amide sp^2O (presumably the $N-H\cdots O=C$ hydrogen bond) and of aliphatic sp^3C with amide sp^2N , amide and aromatic sp^2C , and aliphatic sp^3C , as well as an unfavorable interaction of amide-context aliphatic sp^3C with amide sp^2O . Interaction potentials (α -values) of urea were successfully used to predict and interpret destabilizing effects of urea on globular proteins^{15, 16} and effects of urea on folding and unfolding rate constants.³⁶

Motivated by these previous studies, here we determine preferential interactions (μ_{23} values) of alkylureas with nucleobases and base analogs and dissect these data using the ASA-based analysis to obtain α -values that quantify interactions of each alkylurea with the five nucleobase atom types (carbonyl sp^2O , amino sp^3N , ring sp^2N , methyl sp^3C and ring sp^2C). We compare α -values for the interactions of these alkylureas with corresponding hybridization states of C, N and O atoms of nucleobases, aromatic hydrocarbons, and protein model compounds (amide compounds), and compare with results of molecular dynamics simulations of interactions of amide-containing amino acid side chains with nucleobases.^{23, 24} As representative applications, we use μ_{23} values to predict the baseline contribution of weak adenine-protein interactions (solute effects) to the observed effects of adenine nucleotides on protein processes (hydrotrope effect) at low concentrations of ATP (5-10 mM),¹⁴ and use α -values to discuss the solubilizing effects of urea and alkylureas on the aromatic compound nifedipine.^{12, 13} We discuss the noncovalent interactions responsible for stability and specificity of protein-nucleobase and protein-nucleic acid complexes,^{7, 9-11} many of which are the same as those observed here between amide compounds and nucleobases.

Experimental Methods

Solubility Assays for Interactions of Alkylureas with Nucleobases and Base Analogs

Chemical potential derivatives $(\partial\mu_2/\partial m_3)_{P,T,m_2} = \mu_{23}$ quantifying preferential interactions of urea and alkylureas (designated as component 3) with a set of sparingly soluble nucleobases and other aromatic compounds including base analogs (11-14 in all; designated component 2 and referred to here collectively as nucleobases) were determined from measurements of the effect of alkylurea concentration (up to 3 – 4 molal) on the nucleobase concentration in the saturated solution. All solutions were prepared gravimetrically as previously described,³² an excess of the nucleobase or base analog was added, and the solutions were shaken in tightly-capped vials for 9-14 days to achieve solubility equilibrium. Molar scale quantities proportional to the nucleobase solubility were determined by UV absorbance for accurately prepared dilutions and converted to the molal scale (m_2^{ss}) as previously described.³² The quantity $\ln(m_2^{ss}/m_{2,0}^{ss})$, where $m_{2,0}^{ss}$ is the extrapolated molal

solubility in the absence of solute, is plotted vs. alkylurea molal concentration m_3 and fit to a quadratic in m_3 . The preferential interaction of the alkylurea with the nucleobase μ_{23} is obtained from the initial slope of this plot according to Eq. 1.³² (See Supporting Information (SI)).

$$\left(\frac{\partial \ln m_2^{ss}}{\partial m_3}\right)_{T,P} = \left(\frac{\partial \ln m_2}{\partial m_3}\right)_{T,P,\mu_2} = \left(-\frac{\mu_{23}}{RT}\right) \left(1 + \left(\frac{\partial \ln \gamma_2^{(m)}}{\partial \ln m_2}\right)_{T,P,m_3}\right)^{-1} \cong -\frac{\mu_{23}}{RT} \quad \text{Eq. 1}$$

In Eq. 1, $\gamma_2^{(m)}$ is the activity coefficient of the nucleobase on the molal scale.

For the nucleobase -alkylurea interactions studied here, the alkylurea is in great excess and the nucleobase solubility is small enough that self-interactions of the nucleobase appear to be negligible even in the saturated solution (i.e. $(\partial \ln \gamma_2^{(m)} / \partial \ln m_2)_{P,T,m_3} \ll 1$ even near m_2^{ss}). In

particular, for caffeine (the most soluble of these nucleobases investigated) and the eight other nucleobases examined, no stacking interactions are detected even at m_2^{ss} as judged by the observation that Beer's Law behavior holds up to m_2^{ss} (see SI Figure S1). Stacking would reduce absorbance and cause deviations from Beer's Law.

Results

Preferential Interactions Between Alkylureas and Nucleobases

Negative logarithms of normalized 25 °C solubilities of six representative nucleobases and base analogs (3-methylindole, theobromine, adenine, 6-amino-1,3-dimethyluracil, thymine and uracil) are plotted in the panels of Figure 1 as a function of molality m_3 of each alkylurea investigated. Results for other nucleobases and analogs (caffeine, theophylline, hypoxanthine, 2,6-diaminopurine, cytosine, 5-aminouracil, 6-aminouracil, xanthine) are presented in SI Figure S2. Chemical structures of these nucleobases and analogs are given in Figure 2A and chemical structures of the alkylureas are in Figure 2B. All alkylureas investigated, as well as urea, increase the solubility of all nucleobases and analogs studied. Where these semi-log plots are linear, μ_{23} is obtained from the slope (Eq. 1). Over the wide alkylurea concentration range examined (up to 4 molal), curvature is observed for most combinations of nucleobases, and alkylureas in Figures 1 and S2. In these cases, the initial slope of a second order polynomial fit is used to obtain μ_{23} .

Several general trends are observed in the μ_{23} values of Table S1.

- i. All μ_{23} values are negative, indicating favorable interactions of all alkyl ureas investigated with all nucleobases and analogs investigated. For example, μ_{23} for the interaction of caffeine and 1,1-diethylurea (1,1-deu) is $-462 \pm 29 \text{ cal mol}^{-1} \text{ molal}^{-1}$, indicating that the chemical potential μ_2 of caffeine is reduced by $462 \pm 29 \text{ cal}$ per molal of 1,1-diethylurea added. Therefore, the free energy of transfer of caffeine from water to a 1 molal solution of 1,1-diethylurea is

favorable: $\Delta G^{\circ}_{\text{transfer}} = -462 \pm 29 \text{ cal mol}^{-1}$. These μ_{23} values are well-determined, as demonstrated by fitting uncertainties obtained from Igor 5.5 program (Table S1). Analysis of published solubility data³⁷ for three nucleobases (adenine, thymine and cytosine) as a function of concentration of four ureas (urea, methylurea (mu), ethylurea (eu) and 1,3-dimethylurea (1,3-dmu)) yields μ_{23} values (initial slopes of quadratic fits) which in general agree well with the observed and predicted values reported here (see SI Table S2).

- ii. Preferential interactions (μ_{23} values) of the series of alkylureas with each nucleobase or base analog investigated become more favorable (i.e. more negative μ_{23}) with increasing extent of alkylation of the urea: $0 > \mu_{2,\text{urea}} > \mu_{2,\text{mu}} > \mu_{2,\text{eu}} > \mu_{2,\text{dmu}} > \mu_{2,\text{deu}}$. In addition to increasing the water-accessible surface area (ASA) of aliphatic sp^3C , alkylation reduces the ASA of amide sp^2N and, to a lesser extent, amide sp^2O as summarized in Figure S3 and Table S3. For example, ethylation of urea to form 1,1-deu generates $209 \text{ \AA}^2$ of sp^3C ASA while reducing sp^2N ASA by $80 \text{ \AA}^2$ (62% reduction) and sp^2O ASA by $12 \text{ \AA}^2$ (25% reduction), and increases the strength of the favorable preferential interaction with caffeine by almost two-fold, from $-249 \pm 11 \text{ cal mol}^{-1} \text{ molal}^{-1}$ to $-462 \pm 29 \text{ cal mol}^{-1} \text{ molal}^{-1}$.
- iii. The effect of increased alkylation of the urea on the μ_{23} value is much smaller for nucleobases and analogs with two or three carbonyl sp^2O atoms than for nucleobases with zero or one carbonyl sp^2O atom. For uric acid (three carbonyl sp^2O) the increase in magnitude of μ_{23} from urea to 1,1-dmu is less than 50%. For thymine and uracil (both with two carbonyl sp^2O) the increase in magnitude of μ_{23} from urea to 1,1-deu is about 50%. For nucleobase analogs like theophylline, theobromine, caffeine and hypoxanthine (all with one carbonyl sp^2O) the increase in magnitude of μ_{23} from urea to 1,1-deu is 200–250%. The pyrimidine 1,3-dimethyl-6-amino uracil, with two carbonyl sp^2O , is an apparent exception to the above trend, but its greatly-reduced amount of carbonyl sp^2O ASA makes it similar to nucleobases with one carbonyl sp^2O , and explains why the increase in magnitude of its μ_{23} from urea to 1,1-deu is about 200-250% and not 50%. For aromatics and base analogs like naphthalene and 3-MI with no carbonyl sp^2O , the increase in magnitude of μ_{23} from urea to 1,1-deu is 400-500%.

These trends in μ_{23} values provide evidence that alkylurea – nucleobase interactions, like the alkylurea – amide and – hydrocarbon interactions quantified in previous work¹⁵, can be interpreted as sums of contributions from interactions of the alkyl urea with each type of unified atom on the set of nucleobases. This quantitative analysis is developed in the next section.

Analysis and Discussion

Structures, Water Accessible Surface Areas (ASA) and Compositions of Nucleobases, Base Analogs and Alkylureas

The five different types of unified atoms present on the nucleobases and analogs investigated here (carbonyl sp^2O ; amino sp^3N and ring sp^2N ; aliphatic (methyl) sp^3C and ring sp^2C) are shown with different colors in Figure 2A. (The same color scheme is used in Figure 2B for the four different unified atoms present on the alkylureas investigated here (amide sp^2O , sp^2N , and sp^2C ; aliphatic sp^3C).) Nucleobase ring sp^2N unified atoms with a covalently-bonded proton (ring $sp^2N(H)$) are potential hydrogen-bond donors while proton-less ring sp^2N is only a hydrogen bond acceptor. Approximately equal numbers of these two types of sp^2N unified atoms are present in the set of nucleobases and analogs studied here (Figure 2A). Likewise the two types of ring sp^2C unified atoms (with and without H) may differ in their capacity to interact with unified atoms of urea and alkylureas.

Water accessible surface areas (ASA) of the different types of nucleobase unified atoms, calculated from Cactus structures³⁸ using the Surface Racer program,³⁹ are listed in Table S4 in order of decreasing total hydrocarbon (sp^2C and sp^3C) ASA. ASA values of nucleobases, analogs and alkylureas¹³ calculated from PubChem⁴⁰ and (where available) BMRB⁴¹ structures do not differ significantly from those obtained from Cactus structures. Agreement is within 10%, even for small ASA contributions, comparable to the experimental % uncertainty (see below). BMRB structures are available for only 7 of the 18 nucleobases and analogs investigated, as noted in Table S4, and consequently were not used in these analyses.

Figure S3 and Table S3 show that the urea and the alkylureas form a series in which large increases in sp^3C ASA with increasing alkylation are accompanied by smaller, mostly correlated reductions in amide sp^2N ASA and still-smaller reductions in amide sp^2O ASA. 1,3-dmu displays less amide sp^2O and sp^2N ASA than the alkylureas (1,1-dmu and 1,1 deu) which bracket it in the alkylation series. Changes in ASA of all atom types in progressing from urea to the first alkyl urea (methylurea, mu) are larger than those between neighboring members of the series.

Dissection of μ_{23} values for Alkylurea-Nucleobase Interactions into Contributions from Interactions of Alkylureas with Individual Nucleobase Atom Types

Previous research^{15-17, 29-33} demonstrated that preferential interactions (μ_{23} values) of a variety of different solutes and Hofmeister salts with series of related model compounds are successfully dissected into contributions from interactions of the solute with each type of unified atom present on the model compounds using Eq. 2.

$$\mu_{23} = \sum_i \alpha_i ASA_i \quad \text{Eq. 2}$$

Each term in Eq. 2 is the product of an α_i -value (interaction potential) which quantifies the strength of interaction of the solute with 1 Å² of ASA_i of that atom type on the model

compound and the water accessible surface area (ASA_i) of the i -th type of atom on the model compounds. The additivity of contributions to μ_{23} and the ASA-based analysis used in Eq. 2 have been tested and verified in all previous applications^{15-17, 29-33} and are tested here as well. The chemical significance of the α_i -values has also been demonstrated in previous work^{15-17, 29-33}.

The specific form of Eq. 2 used here to analyze μ_{23} values for interactions of each alkylurea with the set of nucleobases and obtain interaction potentials (α_i -values) for the interactions of each alkylurea with each of the five types of nucleobase unified atom (carbonyl sp^2O , amino sp^3N , ring sp^2N , methyl sp^3C and ring sp^2C) is given in Eq. 3.

$$\mu_{23} = \alpha_{sp^2O} ASA_{sp^2O} + \alpha_{sp^3N} ASA_{sp^3N} + \alpha_{sp^2N} ASA_{sp^2N} + \alpha_{sp^3C} ASA_{sp^3C} + \alpha_{sp^2C} ASA_{sp^2C}$$

Eq. 3

For each alkylurea, a total of 13-16 μ_{23} values from Table S1 were fitted together to Eq. 3 to determine the five α -values. For urea, an additional two μ_{23} values obtained from literature data quantifying effects of urea on solubility of purine¹⁷ and uric acid⁴² are included in the analysis (18 μ_{23} values).

In principle, these sets of μ_{23} values could also be analyzed in terms of the interactions of each intact nucleobase with the four types of unified atom present on alkylureas (amide sp^2O , sp^2N , sp^2C ; aliphatic sp^3C). But, because the alkylurea-nucleobase data set (Table S1) includes 13-18 nucleobases and analogs but only 6 members of the alkylurea series, the dissection in Eq. 3 (13-18 equations to determine 5 α -values) is much more robust than the other direction (6 equations to determine 4 α -values).

α -Values Quantifying Interactions of Alkylureas with Individual Nucleobase Atom Types

α -Values quantifying interactions of the six members of the urea series with the five types of nucleobase unified atoms are plotted as bar graphs in Figure 3 and listed in Table 1. From left to right, the panels of Figure 3 compare α -values for interactions of the alkylurea series with a unit area ($1\text{\AA}^2$) of nucleobase carbonyl sp^2O , amino sp^3N , ring sp^2N , methyl sp^3C and ring sp^2C atoms. In each panel, alkylureas are arranged in order of increasing ASA of aliphatic sp^3C and concomitant decreasing ASA of amide sp^2N (see Figure S3 and Table S3).

Interaction potentials of urea and monoalkylated ureas (mu, eu) with carbonyl sp^2O (Figure 3, left panel) are favorable (negative α -values), but interactions with sp^2O become less favorable with increasing alkylation of the urea. Interactions of the larger alkylated ureas (1,1- and 1,3-dmu, 1,1-deu) with carbonyl sp^2O are unfavorable (positive α -values). Interactions of all alkylureas with both sp^2 and sp^3 hybridization states of nucleobase C and N atoms (Figure 3, center and right panels) are favorable (negative α -values), and in general become more favorable with increasing alkylation of the urea. (1,3-dmu is somewhat of an exception to these trends because of its disproportionately small amounts of sp^2O and sp^2N

ASA relative to its sp^3C ASA (Figure S3; Table S3). Very analogous trends in interactions of the alkylurea series with the different hybridization states of C, N and O atoms to those reported for nucleobases and analogs in Figure 3 were reported previously for interactions of these ureas with the corresponding atoms of amide compounds¹⁵. These trends and comparisons are developed further in Discussion.

Table 1 shows that interactions of the dialkylated ureas with nucleobase ring sp^2C are significantly more favorable than interactions with nucleobase ring sp^2N , while for the monoalkylated ureas these interactions are of comparable strength. In addition, interactions of the two monoalkylated and two dimethyl ureas in Table 1 with nucleobase ring sp^2C and sp^2N are more favorable than interactions with methyl sp^3C and amino sp^3N . Interactions of all five alkylated ureas (but not urea) with nucleobase carbonyl sp^2O are less favorable than interactions with sp^2 and sp^3 hybridization states of nucleobase C and N atoms. For the monoalkylated and two dimethyl ureas, the rank order of α -values is

$$\alpha_{sp^2C} < \alpha_{sp^2N} < \alpha_{sp^3C} < \alpha_{sp^3N} < \alpha_{sp^2O}$$

This same rank order of α -values is observed for the interactions of these alkylureas with the corresponding hybridization states of C, N and O atoms of amide compounds (amide sp^2C , sp^2N and sp^2O ; aliphatic sp^3C).¹⁵

An attempt was made to dissect interactions of urea and alkylureas with ring sp^2N into interactions with proton-less sp^2N and $sp^2N(H)$ unified atoms; this extension of Eq. 3 yielded α -values with acceptable uncertainties only for urea (the largest nucleobase data set). Interactions of urea with sp^2N and $sp^2N(H)$ unified atoms are found to be equal to one another and to the value reported for urea – sp^2N atoms in Table 1. Other urea α -values are not affected.

Comparison of Observed μ_{23} with Predicted Values from α -Values and ASA

Figure 4 and Table S1 compare experimental μ_{23} values for interactions of urea and alkylureas with nucleobases and base analogs with predicted μ_{23} values (Eq. 3) obtained from α -values (Table 1) and ASA information (Table S4). Table S1 shows good agreement between predicted and observed μ_{23} values for the aromatic hydrocarbons and the less polar nucleobases and base analogs investigated, but not as good agreement for the two most polar purines and four most polar pyrimidines. (See SI Figure S4 and accompanying text.)

These comparisons provide additional support for the ASA-based analysis and the hypothesis of additivity in Eq. 3. As another test, μ_{23} values for interactions of one nucleobase or analog with the alkylurea series were removed (Table S1) and the remaining dataset was analyzed to obtain α -values. Figures S5 (A and B) show that in most cases α -values are not significantly affected. An exception is uric acid, with three carbonyl sp^2O atoms; deletion of uric acid affects the carbonyl sp^2O α -value significantly. μ_{23} -Values predicted for the deleted nucleobase and for the remainder of the nucleobase data set from the set of α -values derived for the remaining nucleobases are compared with experimental μ_{23} -values in Table S5, and can be compared with the μ_{23} -values predicted for each nucleobase using α -values from analysis of the full dataset in Table S1. In general, predicted μ_{23} -values in Table S5 agree well with those in Table S1, and the level of agreement between predicted and observed μ_{23} -values in Table S5 is only moderately degraded from that in

Table S1. The systematic deviations between predicted and observed μ_{23} -values for the six most polar nucleobases observed in Table S1 are also present in Table S5.

Discussion

Comparison of Interactions of the Alkylurea Series with Corresponding Hybridization States of O, N and C Atoms of Nucleobases and Amides

Overview—Figure 5 compares α -values for the interactions of the urea-alkylurea series with sp^2O , sp^2N , sp^2C and sp^3C atoms of nucleobases and amides (see also Table 2). Two very significant general observations from Figure 5 are that: i) for all alkylureas, α -values for interactions with nucleobase atoms of a given type (either sp^2O , sp^2N , sp^2C , or sp^3C) are significantly more favorable than for interactions with the same type of atom of amide compounds, as previously observed for urea¹⁷, and ii) trends in α -values in the urea-alkylurea series are very similar for both aromatic and amide contexts of each of these atom types. Both nucleobase carbonyl and amide sp^2O α -values increase (become less favorable) with increasing alkylation, while α -values of all types of C and N (sp^2 , sp^3 C, N) atoms on both amide and nucleobase in general decrease (become more favorable) with increasing alkylation. Observation (i) means that context matters; these atoms in the context of an aromatic ring system interact more favorably with all members of the urea-alkyl urea series than do the same types of atoms on amide compounds. In general, the difference in α -values between aromatic and non-aromatic contexts increases as the extent of alkylation of the urea increases. Figure 5 illustrates how atom identity, hybridization state, and context together determine α -values (Tables 1, 2), which together with ASA information (Table S4) determine strengths of interaction (μ_{23} -values; Table S1) of the urea-alkylurea series with nucleobases and amide compounds.

Interactions with Nucleobase Carbonyl sp^2O and with Amide sp^2O — α -Values quantifying strengths of interactions of urea and alkylureas with nucleobase carbonyl sp^2O (Figure 3) and amide sp^2O atoms¹⁵ are compared in Figure 5A and Table 2. α -Values for interactions of urea with both nucleobase carbonyl and amide sp^2O are favorable (negative); the interaction of urea with carbonyl sp^2O is significantly more favorable than the interaction of urea with amide sp^2O (Figure 3; see also reference¹⁷). Because the ASA of urea is mostly hydrogen-bond-donor sp^2N (70% sp^2N , 26% sp^2O , only 4% sp^2C) and because sp^2O – sp^2O interactions in water are certainly unfavorable, the favorable interactions of urea with nucleobase carbonyl and amide sp^2O atoms almost certainly are $NH\cdots O=C$ hydrogen bonds, as discussed previously.^{15, 17} A hydrogen bond from urea amide sp^2N to nucleobase carbonyl sp^2O is expected to be intrinsically more favorable (i.e. make a more negative free energy contribution) than one from urea amide sp^2N to amide sp^2O because of the adjoining aromatic ring, the electron-withdrawing character of carbonyl sp^2O , and the resulting more negative charge on carbonyl sp^2O than on amide sp^2O . Our results are in agreement with and quantify this expectation. This is another example of the chemical significance of α -values, indicating that the ASA-based analysis of μ_{23} values (Eq. 3) is appropriate and that hydrogen bonding strengths increase with increasing overlap of donor and acceptor atoms.

Figure 5A shows that α -values for interactions of the alkyl urea series with nucleobase carbonyl sp^2O become progressively less favorable with increasing alkylation of the urea, as observed previously for amide sp^2O atom.¹⁵⁻¹⁷ Alkylation of urea increases aliphatic sp^3C ASA and concomitantly reduces amide sp^2N ASA, with a smaller reduction in amide sp^2O ASA (Table S3). Comparing sp^2O α -values for 1,1-diethyl vs 1,1-dimethyl urea (also ethylurea vs methylurea), amide compounds that differ primarily in aliphatic sp^3C ASA, indicates that interactions of aliphatic sp^3C with carbonyl sp^2O are unfavorable relative to interactions with water, as expected and as deduced previously for amide sp^2O atom.¹⁵ Hence interactions of the urea series with carbonyl and amide sp^2O atoms shift from favorable for urea to very unfavorable for the dialkylated ureas because of the increase in unfavorable interactions with aliphatic sp^3C and the concomitant reduction in favorable interactions with amide sp^2O . Figure 5A shows that the crossover point from favorable to unfavorable interactions with nucleobase carbonyl sp^2O occurs between the mono- and the dialkylated ureas, while both mono- and dialkylated ureas exhibit unfavorable interactions with amide sp^2O .

Interactions with Nucleobase Ring sp^2N (also Amino sp^3N) and with Amide sp^2N — α -Values for interactions of urea series members with nucleobase ring sp^2N are compared with previously-published α -values for interactions of these ureas with amide sp^2N , as well as with α -values for interactions of these ureas with nucleobase amino sp^3N , in Figure 5B and Table 2. For each alkylurea investigated, all three α -values are favorable. α -Values for interactions with nucleobase ring sp^2N are significantly more favorable than for amide sp^2N , and α -values for interaction with nucleobase amino sp^3N are intermediate in strength. For all three types of nucleobase and amide N, Figure 5B shows that interactions of the urea-alkylurea series become more favorable with increasing alkylation of the urea, being weakly favorable for urea, moderately favorable for the monoalkylated ureas, and strongly favorable for the dialkylated ureas, especially 1,1-diethyl urea.

Because α -values for interactions of urea with nucleobase amino sp^3N are only weakly favorable, as observed previously for amide sp^2N atoms¹⁵⁻¹⁷ while α -values for interactions of urea with both carbonyl and amide sp^2O are very favorable, we conclude that interactions of amide (urea) sp^2N with nucleobase amino sp^3N are unfavorable, relative to interactions with water. α -Values for interactions of urea with nucleobase ring sp^2N are also only weakly favorable. This may reflect a balance of favorable interactions (e.g. urea sp^2O – nucleobase ring $\text{sp}^2\text{N}(\text{H})$, urea sp^2N – nucleobase proton-less sp^2N) and unfavorable interactions (urea sp^2O – nucleobase protonless ring sp^2N ; urea sp^2N – nucleobase $\text{sp}^2\text{N}(\text{H})$). These N – N interactions are almost certainly hydrogen bonds.

Because α -values for interactions with nucleobase amino sp^3N are significantly more favorable for 1,1-diethyl than for 1,1-dimethyl urea, and also significantly more favorable for ethyl urea than for methyl urea, we infer that C – N interactions of amide-context aliphatic sp^3C with nucleobase amino sp^3N are favorable relative to interactions with water, as deduced previously for interactions of aliphatic sp^3C with amide sp^2N atom.¹⁵ This conclusion probably extends to interactions of amide-context aliphatic sp^3C with aromatic ring sp^2N as well, though the uncertainties in the α -values are large enough to prevent a quantitative comparison.

Interactions with Nucleobase Ring sp^2C and Amide sp^2C ; Also with Nucleobase Methyl sp^3C and Amide-Context Aliphatic sp^3C —

α -Values for interactions of the alkylurea series with nucleobase aromatic ring sp^2C and with nucleobase methyl sp^3C are compared with the corresponding α -values for amide sp^2C and aliphatic (amide context) sp^3C in Figure 5C and Table 2. All interactions are favorable, and in general all increase with increasing alkylation of the urea. Interactions of each alkyl urea with this set of sp^2C and sp^3C atoms exhibit the same rank order of α -values: most favorable with nucleobase ring sp^2C , followed by amide sp^2C , nucleobase methyl sp^3C and then aliphatic sp^3C atoms on amide compounds. Differences between these α -values are smallest for urea, and largest for 1,1-diethylurea. Interactions of urea with nucleobase sp^3C and sp^2C atoms are of similar intrinsic strength.

Because α -values for interactions with nucleobase ring sp^2C and methyl sp^3C are significantly more favorable for 1,1-diethyl than for 1,1-dimethyl urea, and also significantly more favorable for ethyl urea than for methyl urea, we infer that C – C interactions of aliphatic sp^3C with nucleobase ring sp^2C and methyl sp^3C are favorable relative to interactions with water, as expected and as deduced previously for amide-context sp^3C and homocyclic aromatic and amide sp^2C atom.¹⁵

Similar α -Values for Favorable Interactions of Alkylureas with Nucleobase sp^2C and sp^2N Atoms, But not Amide sp^2C and N Atoms—

Table 2 indicates that α -values for interactions of any alkylurea with nucleobase and base analog ring sp^2C and sp^2N are similarly favorable; ring sp^2C α -values are on average about 25% more favorable than ring sp^2N α -values. By contrast, α -values for favorable interactions of the urea-alkylurea series with amide sp^2C are 2-3 times larger on average than for interactions with amide sp^2N . α -Values for favorable interactions with nucleobase methyl sp^3C and nucleobase amino sp^3N are similar for ethyl urea and the dialkylated ureas, agreeing within about 20%. For urea and methyl urea, however, favorable sp^3C and sp^3N α -values differ by approximately twofold.

Comparison with Previous Analysis of Urea-Nucleobase Interactions—

Favorable interactions of urea with the different types of unified atoms of nucleobases were quantified previously.¹⁷ α -Values in Table 1 for interactions of urea with amino sp^3N and with ring sp^2C and sp^2N agree within the uncertainty (± 1 SD) with those previously reported. (Only a combined α -value for ring C and N atoms could be obtained previously.) The α -value in Table 1 for methyl sp^3C also agrees quite well with that previously reported (within ± 2 SD), while that for carbonyl sp^2O is approximately twice as favorable (with a much smaller uncertainty) compared to that obtained previously. The larger number and variety of sp^2O -containing nucleobases in the current dataset is responsible for this improved determination. Use of the current set of urea-nucleobase α -values does not significantly affect our previous conclusion from analysis of urea effects on stability of short RNA duplexes that the single RNA strands are extensively ($\sim 80\%$) stacked in the melted form.¹⁷

Application to Alkyl Ureas and Nucleotides as Hydrotropes

Urea and alkylureas have been studied as solutes (hydrotropes) that increase the solubility of slightly-soluble pharmaceuticals.^{12, 43} The μ_{23} values determined here quantify the solubilizing effects of urea and alkylureas on each of the aromatic compounds and nucleobases investigated. The α -values obtained from analysis of these μ_{23} values are useful to predict the solubilizing effect of these alkylureas on any other aromatic compound and to explain why the alkylurea is an effective solubilizer in terms of its independent interactions with the different types of unified C, N and O atoms of the homocyclic or heterocyclic aromatic compound. As an example, 1 M urea increases the solubility of the slightly-soluble aromatic nitro compound nifedipine (NF; see structure in Figure S6) by ~1.6-fold at 25°C,¹² from which we estimate $\mu_{23} \approx -3.0 \times 10^2 \text{ cal mol}^{-1} \text{ molal}^{-1}$. Approximating α -values for nifedipine aromatic-context ester and nitro sp^2O by the nucleobase carbonyl sp^2O α -value, we predict $\mu_{23} \approx -3.4 \times 10^2 \text{ cal mol}^{-1} \text{ molal}^{-1}$ at 25°C, in close agreement with the observed value. Methyl, ethyl and especially butyl urea are more effective solubilizers than urea; 1 M solutions of these alkylureas increase nifedipine solubility 2.8-fold, 4.3-fold and 34-fold, respectively.¹² These greatly exceed the effects predicted from Table 1 α -values at 1 M of these alkylureas (2.0-fold, 2.3-fold and 2.6-fold solubility increases, respectively), and the discrepancy increases strongly with increasing alkylation, indicating that the amide and hydrocarbon regions of these alkylated ureas likely interact concertedly (not independently) with nifedipine to form a weak stoichiometric complex and thereby increase its solubility more than predicted, as discussed for nicotinamide¹². Similarly, PEG 400 was found to solubilize smaller aromatics like naphthalene to the extent predicted by an α -value analysis, but had a much greater than expected effect on solubility of all steroids examined, explained as concerted interactions of multiple monomer units of PEG 400 with the extended steroid aromatic-aliphatic ring system.³²

A recent report provided evidence for an unusual role of adenine nucleotides as protein hydrotropes at low mM concentrations.¹⁴ Results in Table S1 for alkyl urea-adenine interactions indicate that these effects do not originate simply from preferential interactions of the adenine moiety of ATP with hydrocarbon and amide groups on protein surface exposed in the affected process. Instead something analogous to the enhanced interactions of PEG 400 with steroids and of butyl urea with nifedipine, perhaps involving the phosphate (and sugar) portions of ATP as well as adenine, must be involved. The analysis behind this conclusion is the following. Because $\mu_{23} = \mu_{32}$, values of μ_{23} quantifying effects of alkylureas on the chemical potential of adenine also quantify effects of adenine on the chemical potentials of the alkylureas. Ratios of C : (N + O) ASA of these alkylureas are similar to those observed for the ASA of protein processes (1:1 to 3:1). Values of μ_{32} , expressed per $10^3 \text{ \AA}^2$ of this ASA, are $-1.6 \pm 0.2 \text{ kcal mol}^{-1} (\text{molal adenine})^{-1}$ per $10^3 \text{ \AA}^2$ of ASA. Hence at 10 mM the nonspecific effect of adenine on the free energy change for a protein process exposing $10^3 \text{ \AA}^2$ of ASA is predicted to be only ~16 cal/mol. A protein process would therefore have to expose $>10^4 \text{ \AA}^2$ of ASA in order for these nonspecific adenine-protein atom interactions to exert a detectable hydrotrope effect at 10 mM of an adenine nucleotide. At an adenine nucleotide concentration of 100 mM, a small but significant effect ($\sim 160 \text{ cal mol}^{-1}$ per $10^3 \text{ \AA}^2$ of ASA) of adenine interactions is predicted. The significant hydrotrope effects of 5-10 mM ATP reported recently¹⁴, which are quite

specific for adenine, must therefore result from some concerted set of interactions (i.e. weak binding) involving ribose and/or phosphate moieties of ATP as well as adenine.

Comparison with Results of Molecular Dynamics Simulations

Standard free energies of interaction of nucleobases with amino acid side chains have been calculated from statistical thermodynamic analyses of radial distribution functions and potentials of mean force (PMF) of these side chains in the vicinity of nucleobases in water, obtained from molecular dynamics (MD) simulations.^{23, 24} The asparagine and glutamine side chains are very similar to methylurea and ethylurea, respectively, differing primarily in that the side chains have about 20 Å² (25%) less amide sp²N ASA than the ureas. Reported MD values of standard binding free energies are readily converted to μ_{23} values for the nucleobase-side chain interaction using the analysis of the solute partitioning model³⁴; these MD μ_{23} values are compared with observed and predicted μ_{23} values for interactions of methyl- and ethylurea with six nucleobases and base analog (HPA (9-methylhypoxanthine), A, G, T, U, C) in Table S6. (MD-predicted values of μ_{23} could be obtained directly from the PMF using Kirkwood-Buff integrals, but these were not reported.) Table S6 shows quite good agreement between MD-predicted μ_{23} values for asparagine-nucleobase interactions and μ_{23} values measured (and/or predicted from Table 1 α -values) for HPA, A, G, T, U and C nucleobases. Agreement between ethylurea-nucleobase μ_{23} values and those obtained from MD simulations of glutamine-nucleobase interactions is not quantitative, though still within a factor of two, and the trend in the MD predictions largely matches that observed experimentally.

Relevance for Analysis of Interactions in Nucleotide-Protein and Amide- and Protein-Nucleic Acid Complexes

Nucleobase contributions to specificity and stability of protein-nucleotide complexes and specific complexes of amide-based ligands and proteins with nucleic acids appear to involve the same noncovalent interactions as those responsible for the favorable interactions of the alkylureas investigated here with nucleobases and base analogs. Analyses of high-resolution structures of these complexes^{1, 4, 6, 9-11} reveal the primary role of hydrogen bonding interactions of backbone and side chain (Asn, Gln) amide sp²N (hydrogen bond donor) and sp²O (acceptor) atoms of the ligand or protein with nucleobase carbonyl sp²O (acceptor), amino sp³N (donor) and ring sp²N (acceptor) and sp²N(H) (donor). In complexes of imidazole-amide ligands with DNA, imidazole ring sp²N atoms are also hydrogen bond acceptors.⁶

Interactions of aliphatic sp³C hydrocarbon side chains of the protein with the faces of the nucleobase ring are identified in protein-nucleotide complexes with adenine and guanine.¹⁰ The interaction of amide sp²O of the protein with nucleobase ring sp²C has also been discussed.¹¹ All these interactions are identified as contributing to the favorable interactions of alkylureas with nucleobases reported here and are discussed in previous sections.

Interactions with the Nucleobase Edges—Each nucleobase is uniquely recognized by hydrogen bonding and other interactions of atoms contacted from the edges of the ring system (designated Watson-Crick, Hoogsteen, and sugar edges) rather than from above or

below the plane of the ring(s).^{9, 11} A large majority of protein-nucleobase⁹ and protein-nucleic acid⁴ structures analyzed utilize the Watson-Crick edge for base recognition.⁹ For example, recognition of the Watson-Crick edge of adenine involves three interactions: i) between adenine amino sp^3N (position N6) as hydrogen bond donor and a protein peptide backbone amide sp^2O as acceptor; ii) between an adenine ring sp^2N (position N1, proton-less) hydrogen bond acceptor and a peptide backbone amide sp^2N donor; and iii) between a ring sp^2C (position C2) and either a backbone or sidechain amide sp^2O . Of these three interactions providing specificity of recognition of the Watson-Crick edge of adenine, two are deduced to be quite favorable (amino sp^3N – amide sp^2O , ring sp^2C – amide sp^2O) from the trends in α -values reported here. A quantitative analysis of these α -value trends should establish a rank order of these various favorable hydrogen bonding interactions, present in both alkylurea-nucleobase interactions and nucleobase-protein interactions.

Interactions with the Hoogsteen edge of adenine¹¹ (less common than with the Watson-Crick edge) are very analogous to the above, in that they also involve the N6 amino group and two adjacent positions on the adenine ring. The three interactions are: i) amino sp^3N (position N6), hydrogen bond donor, in some cases to a protein peptide backbone or Asn and Gln side chain amide sp^2O ; ii) ring proton-less sp^2N (position N7), hydrogen bond acceptor, usually with a peptide backbone or Asn and Gln side chain amide sp^2N as donor; iii) ring sp^2C (position C8), in an interaction with either backbone or sidechain amide sp^2O . As with the Watson-Crick edge of adenine, our present and previous analysis¹⁵ indicates that Hoogsteen-edge interactions i) and iii) are favorable but does not yet provide information about interaction ii).

In GTP-protein complexes, recognition of guanine^{9, 11} using the Watson-Crick edge also involves three interactions: i) carbonyl sp^2O (position O6), hydrogen bond acceptor, usually provided by a protein peptide backbone amide sp^2N ; ii) and iii) ring $sp^2N(H)$ (position N1) and amino sp^3N (position N2), both hydrogen bond donors, usually to peptide backbone amide sp^2O as acceptor.

Recognition of adenine and thymine in minor grooves of DNA by drugs such as netropsin and distamycin^{1, 5} and recognition of the four nucleobases by pyrrole-imidazole polyamides⁶ involves amide-nucleobase interactions (hydrogen bonding, hydrophobic). The sugar edge is used in hydrogen bonding between amide sp^2N (proton donor) of the ligand and proton-less ring sp^2N acceptor at position N3 of adenine and guanine or carbonyl sp^2O acceptor at position O2 of thymine and cytosine. The Watson-Crick edge is used in hydrogen bonding between amino sp^3N donor (position 6 for adenine and position 2 for guanine) with ring sp^2N (proton-less, acceptor) of imidazoles of the ligand. Other important interactions between netropsin or distamycin and AT regions of DNA are C – C (hydrophobic) contacts between adenine sp^2C (position C2) and CH groups on the pyrrole rings of the drug molecules. Many of these interactions are present in the alkylurea-nucleobase series studied here.

Interactions with the Nucleobase Ring Faces—Many of the protein interactions with the nucleobase ring identified in complexes of bound adenine or guanine nucleotides involve aliphatic (sp^3C) hydrocarbon atoms of side chains¹⁰. From the comparisons of α -values for

nucleobase interactions with methylated and ethylated ureas presented here, we infer that sp^3C interactions with both ring sp^2C and ring sp^2N are favorable, with similar α -values. Additional analysis of the α -values determined here should allow more quantitative information about these sp^3C – ring interactions.

Additional analysis of the α -values determined here (in progress) will provide more quantitative information about the atom-atom interactions responsible for stability and specificity of these protein-nucleotide complexes, as well as those (like sp^2O – sp^2O) which are unfavorable and presumably do not occur. To predict relative contributions of different atom-atom interactions to stability of a complex or interface, it will be important to determine the molecular contact areas for the different unified atoms involved in these different interactions.

Conclusions

Analysis of protein-nucleotide (ATP, GTP) structures reveal that both peptide backbone amide N and O atoms and side chain amide N and O atoms of glutamine and asparagine residues make hydrogen bonds to carbonyl O, amino N, and ring N atoms on the edges of these nucleobases^{4, 9-11} Specific interactions of ligands and proteins with the edges of nucleobases, accessed from the grooves of double helical nucleic acids, also involves amide-nucleobase hydrogen bonds^{1, 2, 6-8} The analysis of the thermodynamic data reported her, quantifying interactions of amide compounds with nucleobases and base analogs, not only indicates the significance of these favorable amide-nucleobase hydrogen bonding interactions but also provides evidence for other favorable interactions (relative to interactions with water). These include interactions of aliphatic C atoms of amide compounds with ring and amino N atoms of nucleobases as well as of these amide aliphatic C atoms with ring and methyl C atoms of nucleobases. The latter (C – C) are presumably an example of the hydrophobic effect, while these favorable C – N interactions have not previously been recognized. Also of interest is the deduction that amide N – nucleobase amino N interactions (presumably hydrogen bonding) are unfavorable, relative to interactions (again presumably hydrogen bonding) with water. In addition to these qualitative conclusions, the results presented here allow one to predict the effect of any amide compound investigated here (urea, alkyl ureas) on any process that changes the exposure of any nucleobase or aromatic atom type investigated here to water, including dissolving nucleobases and aromatics as well as on nucleic acid helix formation and protein-nucleic acid binding. In ongoing research, we are quantifying these and other atom-atom interactions of amides with amides and amides with nucleobases in order to obtain the full spectrum of favorable and unfavorable C, N and O atom interactions for these key constituent atoms of proteins and nucleic acids in water.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Hunter Cochran, who performed some of the methylurea experiments in her undergraduate research. We thank the reviewers and the editor for their helpful comments on the manuscript. We gratefully acknowledge support of this research from National Institutes of Health Grant GM118100 (to M. T. R.) and from the University of Wisconsin - Madison.

References

1. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc Natl Acad Sci U S A*. 1985; 82:1376–1380. [PubMed: 2983343]
2. Nadassy K, Wodak SJ, Janin J. Structural features of protein-nucleic acid recognition sites. *Biochemistry*. 1999; 38:1999–2017. [PubMed: 10026283]
3. Mrksich M, Wade WS, Dwyer TJ, Geierstanger BH, Wemmer DE, Dervan PB. Antiparallel Side-by-Side Dimeric Motif for Sequence-Specific Recognition in the Minor Groove of DNA by the Designed Peptide 1-Methylimidazole-2-Carboxamide Netropsin. *Proc Natl Acad Sci U S A*. 1992; 89:7586–7590. [PubMed: 1323845]
4. Cheng AC, Chen WW, Fuhrmann CN, Frankel AD. Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J Mol Biol*. 2003; 327:781–796. [PubMed: 12654263]
5. Dolenc J, Oostenbrink C, Koller J, van Gunsteren WF. Molecular dynamics simulations and free energy calculations of netropsin and distamycin binding to an AAAAA DNA binding site. *Nucleic Acids Res*. 2005; 33:725–733. [PubMed: 15687382]
6. Dervan PB, Edelson BS. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol*. 2003; 13:284–299. [PubMed: 12831879]
7. Seeman NC, Rosenberg JM, Rich A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc Natl Acad Sci U S A*. 1976; 73:804–808. [PubMed: 1062791]
8. von Hippel PH. From “simple” DNA-protein interactions to the macromolecular machines of gene expression. *Annu Rev Biophys Biomol Struct*. 2007; 36:79–105. [PubMed: 17477836]
9. Kondo J, Westhof E. Classification of pseudo pairs between nucleotide bases and amino acids by analysis of nucleotide-protein complexes. *Nucleic Acids Res*. 2011; 39:8628–8637. [PubMed: 21737431]
10. Nobeli I, Laskowski RA, Valdar WSJ, Thornton JM. On the molecular discrimination between adenine and guanine by proteins. *Nucleic Acids Res*. 2001; 29:4294–4309. [PubMed: 11691917]
11. Denessiouk KA, Johnson MS. “Acceptor-donor-acceptor” motifs recognize the Watson-Crick, Hoogsteen and Sugar “Donor-acceptor-donor” edges of adenine and adenosine-containing ligands. *J Mol Biol*. 2003; 333:1025–1043. [PubMed: 14583197]
12. Suzuki H, Sunada H. Mechanistic studies on hydrotropic solubilization of nifedipine in nicotinamide solution. *Chem Pharm Bull (Tokyo)*. 1998; 46:125–130. [PubMed: 9468644]
13. Cui Y. Hydrotropic Solubilization by Urea Derivatives: A Molecular Dynamics Simulation Study. *J Pharm (Cairo)*. 2013; 2013:791370. [PubMed: 26555993]
14. Patel A, Malinowska L, Saha S, Wang J, Alberti S, Krishnan Y, Hyman AA. ATP as a biological hydrotrope. *Science*. 2017; 356:753–756. [PubMed: 28522535]
15. Cheng X, Shkel IA, Connor KO, Henrich J, Molzahn C, Lambert D, Record T Jr. Experimental Atom-by-Atom Dissection of Amide-Amide and Amide-Hydrocarbon Interactions in H₂O. *J Am Chem Soc*. 2017; 139:9885–9894. [PubMed: 28678492]
16. Guinn EJ, Pegram LM, Capp MW, Pollock MN, Record MT Jr. Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer. *Proc Natl Acad Sci U S A*. 2011; 108:16932–16937. [PubMed: 21930943]
17. Guinn EJ, Schwinefus JJ, Cha HK, McDevitt JL, Merker WE, Ritzer R, Muth GW, Engelskjerd SW, Mangold KE, Thompson PJ, Kerins MJ, Record MT Jr. Quantifying functional group interactions that determine urea effects on nucleic acid helix formation. *J Am Chem Soc*. 2013; 135:5828–5838. [PubMed: 23510511]

18. Weerasinghe S, Smith PE. A Kirkwood-Buff derived force field for mixtures of urea and water. *J Phys Chem B*. 2003; 107:3891–3898.
19. Smith PE. Chemical potential derivatives and preferential interaction parameters in biological systems from Kirkwood-Buff theory. *Biophys J*. 2006; 91:849–856. [PubMed: 16679363]
20. Canchi DR, Garcia AE. Cosolvent Effects on Protein Stability. *Annu Rev Phys Chem*. 2013; 64:273–293. [PubMed: 23298246]
21. Holehouse AS, Garai K, Lyle N, Vitalis A, Pappu RV. Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. *J Am Chem Soc*. 2015; 137:2984–2995. [PubMed: 25664638]
22. Sapir L, Harries D. Is the depletion force entropic? Molecular crowding beyond steric interactions. *Curr Opin Colloid In*. 2015; 20:3–10.
23. de Ruiter A, Zagrovic B. Absolute binding-free energies between standard RNA/DNA nucleobases and amino-acid sidechain analogs in different environments. *Nucleic Acids Res*. 2015; 43:708–718. [PubMed: 25550435]
24. Hajnic M, de Ruiter A, Polyansky AA, Zagrovic B. Inosine Nucleobase Acts as Guanine in Interactions with Protein Side Chains. *J Am Chem Soc*. 2016; 138:5519–5522. [PubMed: 27093234]
25. Horinek D, Netz RR. Can Simulations Quantitatively Predict Peptide Transfer Free Energies to Urea Solutions? Thermodynamic Concepts and Force Field Limitations. *J Phys Chem A*. 2011; 115:6125–6136. [PubMed: 21361327]
26. Ganguly P, van der Vegt NF. Representability and Transferability of Kirkwood-Buff Iterative Boltzmann Inversion Models for Multicomponent Aqueous Systems. *J Chem Theory Comput*. 2013; 9:5247–5256. [PubMed: 26592264]
27. Priyakumar UD, Hyeon C, Thirumalai D, Mackerell AD Jr. Urea destabilizes RNA by forming stacking interactions and multiple hydrogen bonds with nucleic acid bases. *J Am Chem Soc*. 2009; 131:17759–17761. [PubMed: 19919063]
28. Ueda Y, Zouzumi Y, Maruyama A, Nakano S, Sugimoto N, Miyoshi D. Effects of trimethylamine N-oxide and urea on DNA duplex and G-quadruplex. *Sci Technol Adv Mat*. 2016; 17:753–759.
29. Capp MW, Pegram LM, Saecker RM, Kratz M, Riccardi D, Wendorff T, Cannon JG, Record MT Jr. Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values. *Biochemistry*. 2009; 48:10372–10379. [PubMed: 19757837]
30. Pegram LM, Wendorff T, Erdmann R, Shkel I, Bellissimo D, Felitsky DJ, Record MT Jr. Why Hofmeister effects of many salts favor protein folding but not DNA helix formation. *Proc Natl Acad Sci U S A*. 2010; 107:7716–7721. [PubMed: 20385834]
31. Diehl RC, Guinn EJ, Capp MW, Tsodikov OV, Record MT Jr. Quantifying additive interactions of the osmolyte proline with individual functional groups of proteins: comparisons with urea and glycine betaine, interpretation of m-values. *Biochemistry*. 2013; 52:5997–6010. [PubMed: 23909383]
32. Knowles DB, Shkel IA, Phan NM, Sternke M, Lingeman E, Cheng X, Cheng LX, O'Connor K, Record MT. Chemical Interactions of Polyethylene Glycols (PEGs) and Glycerol with Protein Functional Groups: Applications to Effects of PEG and Glycerol on Protein Processes. *Biochemistry*. 2015; 54:3528–3542. [PubMed: 25962980]
33. Cheng X, Guinn EJ, Buechel E, Wong R, Sengupta R, Shkel IA, Record MT. Basis of Protein Stabilization by K Glutamate: Unfavorable Interactions with Carbon, Oxygen Groups. *Biophys J*. 2016; 111:1854–1865. [PubMed: 27806267]
34. Record MT Jr, Guinn E, Pegram L, Capp M. Introductory lecture: interpreting and predicting Hofmeister salt ion and solute effects on biopolymer and model processes using the solute partitioning model. *Faraday Discuss*. 2013; 160:9–44. discussion 103–120. [PubMed: 23795491]
35. Lambert D, Draper DE. Denaturation of RNA secondary and tertiary structure by urea: simple unfolded state models and free energy parameters account for measured m-values. *Biochemistry*. 2012; 51:9014–9026. [PubMed: 23088364]

36. Guinn EJ, Kontur WS, Tsodikov OV, Shkel I, Record MT Jr. Probing the protein-folding mechanism using denaturant and temperature effects on rate constants. *Proc Natl Acad Sci U S A*. 2013; 110:16784–16789. [PubMed: 24043778]
37. Herskovits TT, Bowen JJ. Solution studies of the nucleic acid bases and related model compounds. Solubility in aqueous urea and amide solutions. *Biochemistry*. 1974; 13:5474–5483. [PubMed: 4549016]
38. Sadowski J, Gasteiger J, Klebe G. Comparison of Automatic 3-Dimensional Model Builders Using 639 X-Ray Structures. *J Chem Inf Comput Sci*. 1994; 34:1000–1008.
39. Tsodikov OV, Record MT Jr, Sergeev YV. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J Comput Chem*. 2002; 23:600–609. [PubMed: 11939594]
40. Bolton EE, Chen J, Kim S, Han LY, He SQ, Shi WY, Simonyan V, Sun Y, Thiessen PA, Wang JY, Yu B, Zhang J, Bryant SH. PubChem3D: a new resource for scientists. *J Cheminform*. 2011; 3:32–46. [PubMed: 21933373]
41. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao HY, Markley JL. BioMagResBank. *Nucleic Acids Res*. 2008; 36:D402–D408. [PubMed: 17984079]
42. Roseman M, Jencks WP. Interactions of Urea and Other Polar Compounds in Water. *J Am Chem Soc*. 1975; 97:631–640.
43. Dhapte V, Mehta P. Advances in hydrotropic solutions: An updated review. *St Petersburg Polytechnical University Journal: Physics and Mathematics*. 2016; 1:424–435.

Figure 1.

Solubility Determinations of Preferential Interactions (μ_{23}) of the Urea-Alkylurea Series (component 3) with Six Nucleobases and Base Analogs (component 2) at 25°C. The negative logarithm of the solubility ratio $m_2^{ss}/m_{2,0}^{ss}$, where m_2^{ss} is the molal solubility of the nucleobase or base analog in the presence of solute and $m_{2,0}^{ss}$ is its extrapolated molal solubility in the absence of solute, is plotted versus molality m_3 of each solute. Initial slopes of quadratic fittings yield μ_{23}/RT at 25 °C (See Table S1). Solubility data quantifying interactions of these ureas with other nucleobases and analogs are plotted analogously in SI Figure S2.

Figure 2.

Structures of Nucleobases and Base Analogs (Panel A) and Urea and Alkylureas (Panel B) with Color-coded Surfaces. Red: Carbonyl or amide sp^2O . Purple: Amino sp^3N . Green: aromatic ring or amide sp^2N . Blue: methyl or amide context aliphatic sp^3C . Black: aromatic ring sp^2C .

Figure 3.

Trends in Strengths of Interaction of the Urea-Alkylurea Series with Unified Atoms of Nucleobases. Bar graphs compare interaction potentials (α -values; Table 1) quantifying interactions of ureas with a unit area of nucleobase carbonyl sp^2 O, amino sp^3 N, ring sp^2 N, methyl sp^3 C and ring sp^2 C at 25°C. Negative (positive) α -values indicate favorable (unfavorable) interactions. Abbreviations for alkyl ureas: mu and dmua, methyl and dimethyl urea; eu and deu, ethyl and diethyl urea.

Figure 4.

Comparison of Predicted and Observed μ_{23} Values for Interactions of Urea and Alkylureas with Nucleobases and Analogs at 25 °C. Predictions of μ_{23} use α -values (Table 1) and ASA information (Table S3) for nucleobase carbonyl sp^2O , amino sp^3N , ring sp^2N , methyl sp^3C and ring sp^2C . (See Table S1.) Abbreviations for alkyl ureas are defined in the caption to Figure 3. The line represents equality of predicted and observed values. Uncertainties of observed and predicted μ_{23} values are calculated as discussed in SI and here shown for the two large alkylated ureas (1,3-dmu and 1,1-deu).

Figure 5.

Comparison of Interaction Potentials (α -Values) for Interactions of Urea and Alkyl Ureas with Corresponding Types of Unified Atoms of Nucleobases and Amides. Panel A: sp^2O . Panel B: sp^2N and sp^3N . Panel C: sp^2C and sp^3C . α -Values for interactions of the urea-alkylurea series with nucleobase unified atoms are from Table 1, while α -values for interactions of these ureas with amide unified atoms are from Cheng et al.¹³ Abbreviations for alkyl ureas are defined in the caption to Figure 3.

Table 1
Strengths of Interaction (α -values) of Urea and Alkyl Ureas with Nucleobase Unified Atoms

Nucleobase Surface Type (<i>i</i>)	α -values (cal mol ⁻¹ molal ⁻¹ Å ⁻²) ^a					
	urea	mu	eu	1,1-dmu	1,3-dmu	1,1-deu
Carbonyl sp ² O	-0.80 ± 0.05	-0.23 ± 0.06	-0.12 ± 0.06	0.06 ± 0.09	0.45 ± 0.05	0.71 ± 0.12
Amino sp ³ N	-0.29 ± 0.04	-0.46 ± 0.05	-0.81 ± 0.05	-1.23 ± 0.07	-1.07 ± 0.05	-1.64 ± 0.1
Ring sp ² N	-0.70 ± 0.04	-1.15 ± 0.08	-1.36 ± 0.07	-1.7 ± 0.1	-1.13 ± 0.07	-2.01 ± 0.13
Methyl sp ³ C	-0.57 ± 0.03	-0.80 ± 0.04	-1.01 ± 0.05	-1.07 ± 0.06	-1.33 ± 0.03	-1.26 ± 0.08
Ring sp ² C ^b	-0.57 ± 0.02	-1.3 ± 0.07	-1.57 ± 0.07	-2.02 ± 0.07	-1.92 ± 0.04	-2.7 ± 0.1

^a α -values are obtained by fitting experimental μ_{23} values (Table S1) for interactions of each urea and alkylurea with the series of nucleobases and base analogs to Eq. 3. Propagated uncertainties in α -values are calculated as described in SI Methods and Ref. 32.

^b Calculated for combined heterocyclic and homocyclic ring sp²C.

Comparison of Interaction Potentials (α -Values; cal mol⁻¹ molal⁻¹ Å⁻²) of the Urea-Alkylurea Series with Unified Atoms of Proteins and Nucleobases

Table 2

Solute	sp ² O		sp ² and sp ³ N			
	Protein Context ^a	Nucleobase	Protein Context ^d		Nucleobase	
	Amide sp ² O	Carbonyl sp ² O	Amide sp ² N	Amino sp ³ N	Ring sp ² N	
urea	-0.52 ± 0.04	-0.80 ± 0.05	-0.09 ± 0.02	-0.29 ± 0.04	-0.70 ± 0.04	
mu	0.52 ± 0.08	-0.23 ± 0.06	-0.44 ± 0.03	-0.46 ± 0.05	-1.15 ± 0.08	
eu	0.79 ± 0.06	-0.12 ± 0.06	-0.55 ± 0.02	-0.81 ± 0.05	-1.36 ± 0.07	
1,1-dmu	1.09 ± 0.07	0.06 ± 0.09	-0.61 ± 0.03	-1.23 ± 0.07	-1.7 ± 0.1	
1,3-dmu	1.68 ± 0.09	0.45 ± 0.05	-0.77 ± 0.03	-1.07 ± 0.05	-1.13 ± 0.07	
1,1-deu	1.73 ± 0.09	0.71 ± 0.12	-0.84 ± 0.03	-1.64 ± 0.1	-2.01 ± 0.13	

Solute	sp^2 and sp^3C			
	Protein Context ^d	Nucleobase		Protein Context ^d
	Aliphatic sp^3C	Amide sp^2C	Ring sp^2C^b	Methyl sp^3C
urea	-0.07 ± 0.01	-0.69 ± 0.06	-0.57 ± 0.02	-0.57 ± 0.03
mu	-0.31 ± 0.01	-0.99 ± 0.07	-1.3 ± 0.07	-0.80 ± 0.04
eu	-0.43 ± 0.01	-1.2 ± 0.05	-1.57 ± 0.07	-1.01 ± 0.05
1,1-dmu	-0.35 ± 0.01	-1.53 ± 0.05	-2.02 ± 0.07	-1.07 ± 0.06
1,3-dmu	-0.56 ± 0.01	-1.87 ± 0.06	-1.92 ± 0.04	-1.33 ± 0.03
1,1-deu	-0.64 ± 0.01	-1.88 ± 0.07	-2.7 ± 0.1	-1.26 ± 0.08

^a α -values from ref. 15.

^b α -values for combined homocyclic and heterocyclic aromatic sp^2C , which agree well with those for homocyclic aromatic sp^2C in ref. 15.