# A Multilevel Strategy for the Exploration of the Conformational Flexibility of Small Molecules

Flavio Forti,[†] Claudio N. Cavasotto,[‡,⊥] Modesto Orozco,[§] Xavier Barril,*[,†,‖] and F. Javier Luque*[,†]

[†]Departament de Fisicoquímica and Institut de Biomedicina (IBUB), Facultat de Farmàcia, Universitat de Barcelona, Av. Diagonal 643, E-08028, Barcelona, Spain

[‡]School of Biomedical Informatics, University of Texas Health Science Center at Houston, 7000 Fannin Street Ste. 690, Houston, Texas 77030, United States

[§]Joint IRB-BSC Program in Computational Biology, Institute of Research in Biomedicine Barcelona, Barcelona Scientific Park, Josep Samitier 1-6, 08028 barcelona, Spain

[‖]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

**S** *Supporting Information*

**ABSTRACT:** Predicting the conformational preferences of flexible compounds is still a challenging problem with important implications in areas such as molecular recognition and drug design. In this work, we describe a multilevel strategy to explore the conformational preferences of molecules. The method relies on the predominant-state approximation, which partitions the conformational space into distinct conformational wells. Moreover, it combines low-level (LL) methods for sampling the conformational minima and high-level (HL) techniques for calibrating their relative stability. In the implementation used in this study, the LL sampling is performed with the semiempirical RM1 Hamiltonian, and solvent effects are included using the RM1 version of the MST continuum solvation model. The HL refinement of the conformational wells is performed by combining geometry optimizations of the minima at the B3LYP (gas phase) or MST-B3LYP (solution) level, followed by single point MP2 computations using Dunning's augmented basis sets. Then, the effective free energy of a conformational well is estimated by combining the MP2 energy, supplemented with the MST-B3LYP solvation free energy for a conformational search in solution, with the local curvature of the well sampled at the semiempirical level. Applications of this strategy involve the exploration of the conformational preferences of 1,2-dichloroethane and neutral histamine in both the gas phase and water solution. Finally, the multilevel strategy is used to estimate the reorganization cost required for selecting the bioactive conformation of HIV reverse transcriptase inhibitors, which is estimated to be at most 1.3 kcal/mol.

## INTRODUCTION

The binding affinity between a ligand and its macromolecular target is determined by a subtle balance of enthalpic and entropic contributions.[1,2] The chemical complementarity between the functional groups present in the ligand and the residues that delineate the binding site modulates the binding affinity through a variety of intermolecular interactions.[3,4] This stabilizing term compensates for unfavorable contributions to the binding, such as the desolvation of the ligand and receptor, as well as the entropic loss upon formation of the ligand–receptor complex. In addition, this process can be accompanied by conformational changes in both the ligand and receptor. In the latter case, those changes generally affect the orientation of side chains in the binding cavity, but often alterations in secondary structural elements and even large-scale rearrangements between domains can also occur.[5–7] With regard to the ligand, those conformational changes are associated with the adoption of the bioactive conformation found in the bound state. The conformational changes triggered upon ligand binding contribute to the binding free energy ($\Delta G_{bind}$; see Figure 1), which can be expressed as the addition of the free energy contribution due to the recognition between the ligand and receptor in the bound (i.e., bioactive) conformation ($\Delta G_{int}$) and the cost associated with the structural deformation
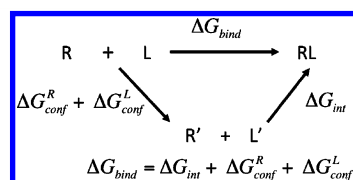


**Figure 1.** Thermodynamic cycle for the formation of a ligand–receptor complex.

of both the ligand and receptor in solution ($\Delta G_{conf}^{L}$ and $\Delta G_{conf}^{R}$, respectively).

The bioactive conformation of the ligand is just one of the many possible conformations sampled at room temperature in the physiological media, perhaps not the most prevalent one. The identification of the bioactive species is a challenging requisite in drug design, as the outcome of pharmacophore searches and docking largely depends on the ability to generate conformers that mimic the ligand in the bound state. Not surprisingly, great effort has been devoted to developing computational tools for predicting the bioactive conformation of drug-like compounds.[8–17] One can intuitively expect that a

good binder will be recognized by the receptor in a low energy conformation, though as noted above it might not correspond to the global minimum of the free ligand, which raises the question about the energetic cost required for adopting the bioactive conformation. If one assumes a direct relationship between the biological activity and the binding free energy, the uncertainty in the conformational penalty required for selecting the bioactive species can lead to a significant error in the binding affinity and hence in the potency of the ligand.[18]

Several studies have attempted to estimate the cost needed for selection of the bioactive conformation of ligands.[19−21] Perola and Charifson considered a set of 150 compounds with a known X-ray structure for their ligand−protein complexes and determined the ligand conformational reorganization cost by combining classical force field calculations with either a distance-dependent dielectric constant or a GBSA continuum model.[19] The results supported the common assumption that ligands rarely bind in their lowest energy conformation. However, the authors reported that around 60% of the ligands bind with strain energies lower than 5 kcal/mol, though strain energies over 9 kcal/mol were calculated in at least 10% of the cases. These large values seem counterintuitive for high affinity ligands, as it is reasonable to expect that the bioactive conformation should generally be either the most stable species in aqueous solution or slightly destabilized compared to the most stable conformation. Rather, one should also convey that the large conformational penalty could be attributed to the uncertainties of the computational scheme employed in the study. This latter point is supported by a systematic study by Tirado-Rives and Jorgensen, who examined the uncertainty in the calculation of $\Delta G_{conf}^{L}$.[20] They concluded that the uncertainty estimated when the energetics for key conformers is determined at the *ab initio* or DFT levels amounts to around 5 kcal/mol but can increase up to 10 kcal/mol when current force fields are used.

Recently, Butler et al. revisited a set of 99 drug-like molecules curated from the data set considered by Perola and Charifson.[21] They found a conformational penalty of only ∼1 kcal/mol between the bioactive conformation and the closest local minimum, it being even lower than 0.5 kcal/mol in around 50% of the cases, when relative energies determined at the B3LYP/6-31G(d) level[22] were combined with hydration free energies estimated from IEF-MST calculations.[23,24] Even though this study provides a reasonable threshold to the difference in stability between the bioactive conformation and the closest local minimum, the conformational reorganization cost of the ligand was not estimated, as this would have required the identification of the conformational ensemble in aqueous solution. It is noteworthy that a precise answer to this question is crucial for the efficient selection of biologically relevant conformations of flexible ligands, which in turn would have a direct impact on reducing the computational cost needed for virtual screening of compound libraries.

The aim of this work is to explore the suitability of a *multilevel* strategy conceived as an efficient methodology for the conformational search of drug-like compounds in solution and the estimation of the relative stability of the most populated conformations. To this end, the multilevel strategy combines a conformational search of the ligand using a low-level computational method and *a posteriori* calibration of the relative stabilities between conformational wells using high-level calculations. The reliability of the strategy is discussed by examining the conformational preferences of 1,2-dichloro-

ethane and neutral histamine as test molecules and by providing estimates of the energetic cost of selecting the bioactive conformation for a small set of HIV reverse transcriptase inhibitors.[20]

## ■ METHODS

**Theoretical Background.** In the context of the predominant-states approximation adopted by Gilson and co-workers,[25,26] one can assume that the conformational space can be partitioned into different $M$ local energy wells $j$ and that the sum of the configurational integral within each well equals the full configurational integral. In this context, the free energy of a flexible molecule can be expressed as the sum of the free energies from the set of $M$ local energy wells (eq 1).

$$A = \sum_{j=1}^{M} A_j = -RT \sum_{j=1}^{M} \ln \left[ \frac{\nu_j}{N_j} \sum_{\substack{k=1 \\ k \in j}}^{N_j} e^{-E_k/RT} \right] \quad (1)$$

where $\nu_j$ and $N_j$ denote the volume and number of conformational states for well $j$.

With regard to eq 1, we are also assuming that the dependence on the soft torsional degrees of freedom of the determinants of the mass tensor and of the Hessian restricted to the subspace of the hard variables is negligible, which holds for small molecules.[27]

If in a first approximation we limit our attention to the relative stability between two conformations $i$ and $j$ (i.e., the bioactive conformation and the global minimum), the free energy difference between those conformations ($\Delta A_i^j = A_j - A_i$) can then be determined from eq 2.

$$\Delta A_i^j = -RT \ln \frac{\frac{\nu_j}{N_j} \sum_{k=1}^{N_j} e^{-E_k/RT}}{\frac{\nu_i}{N_i} \sum_{l=1}^{N_i} e^{-E_l/RT}} \quad (2)$$

For a given conformational well $i$, it is convenient to express the conformational free energy by separating the contribution due to the minimum energy conformer in the well, as noted in eq 3.

$$A_i = -RT \ln \frac{\nu_i}{N_i} \left( e^{-E_{i,\min}/RT} \sum_{k=1}^{N_i} e^{-\Delta E_k/RT} \right)$$

$$= E_{i,\min} - RT \ln \sum_{k=1}^{N_i} e^{-\Delta E_k/RT} - RT \ln \frac{\nu_i}{N_i} \quad (3)$$

where $E_{i\_\min}$ stands for the energy of the minimum energy conformer in well $i$ and $\Delta E_k$ denotes the relative energy of conformation $k$ in the well $i$ relative to the minimum energy conformer ($\Delta E_k = E_k - E_{i\_\min}$), and thus the second term on the right-hand side accounts for the local curvature of the well $i$ (denoted $A_i^{local}$ hereafter).

By adopting the preceding expression, the free energy difference between two conformational wells $i$ and $j$ can then be rewritten in a more compact expression, as noted in eq 4.

$$\Delta A_i^j = \Delta E_{i,\min,j,\min} + \Delta A_{i,j}^{local} - RT \ln \frac{\nu_j N_i}{\nu_i N_j} \quad (4)$$

where the first term on the right-hand side denotes the relative energy between the minimum energy conformations in wells $i$ and $j$ (eq 5) and the second term stands for the free energy

contribution due to the distribution of conformational states in those wells (eq 6).

$$\Delta E_{i,\min,j,\min} = E_{j,\min} - E_{i,\min} \tag{5}$$

$$\Delta A_{i,j}^{\text{local}} = -RT \ln \frac{\sum_{k=1}^{N_j} e^{-\Delta E_k/RT}}{\sum_{l=1}^{N_i} e^{-\Delta E_l/RT}} \tag{6}$$

Finally, let us recall that if a uniform sampling of the conformational volume associated with wells $i$ and $j$ is performed, then the third term in eq 4 will cancel (as noted in Computational Details, in the present implementation, a uniformly distributed grid is generated after the Monte Carlo sampling, which results in the above-mentioned cancellation).

**The Multilevel Strategy.** Equation 4 points out that the relative stability between two conformational families depends on (i) the accuracy in estimating the relative energy between the minimum energy conformations of those wells and (ii) a reliable description of the curvature around those minima. Recently, an exhaustive comparative analysis of different quantum mechanical (QM) methods, including Hartree−Fock, second-order Möller−Plesset, and a variety of density functionals, supports the performance of MP2/aug-cc-pVDZ calculations for predicting conformational energies. In fact, this method was found to yield uncertainties below 7%,[28] it being substantially lower than the error predicted for DFT functionals and much lower than the uncertainty of HF calculations. However, this level of theory is impractical for routine applications dealing with the conformational sampling of drug-like compounds, especially keeping in mind that oral bioavailability can be fulfilled by drugs containing up to 10 rotatable bonds.[29] Though DFT methods are computationally more efficient, they are still too expensive for a complete sampling of highly flexible compounds, which adds to the finding by Riley et al. that no specific class of DFT functionals was found to be significantly more accurate than another for predicting conformational energies.[28]

In order to provide a practical solution to the preceding conflict, the multilevel strategy presented here pursues to reconcile the need for (i) chemical accuracy in predicting conformational energies and (ii) computational efficiency required for the exhaustive sampling of the conformational wells of (bio)organic ligands in solution. To this end, it combines a low-level (LL) method to explore the conformational landscape of a flexible compound and a high-level (HL) method to refine the relative stability of the structures corresponding to the minima of the conformational wells. Thus, our approach mimics to some extent the strategy adopted in other computational frameworks in order to obtain a chemically accurate description of reactive processes.[30]

The LL sampling for a molecule in solution is guided by an effective energy, $E_{\text{LL}}^{\text{eff}}(r_{\text{int}})$, that depends on the internal coordinates of the ligand, $r_{\text{int}}$, and that is expressed as the addition of the potential energy for the internal energy of the molecule, $U_{\text{LL}}(r_{\text{int}})$, and the solvation free energy of the molecule, $W_{\text{LL}}(r_{\text{int}})$ (eq 7; this latter term is obviously not considered if the conformational sampling is limited to the gas phase). Let us note that similar approaches regarding the use of effective potentials have been adopted for the analysis of free-energy surfaces for liquid phase reactions.[31]

$$E_{\text{LL}}^{\text{eff}}(r_{\text{int}}) = U_{\text{LL}}(r_{\text{int}}) + W_{\text{LL}}(r_{\text{int}}) \tag{7}$$

On the other hand, the refinement of the energy of conformational minima through HL computations allows us to correct the ordering of stability of the minimum energy structures using eq 8 as an approximate expression for estimating the free energy of a conformational well. This expression assumes that the local curvature of the wells is properly described through the LL sampling, whereas rescaling of the conformational wells is accomplished on the basis of the HL computations.

$$A_i \approx E_{i,\min}(\text{HL}) + A_i^{\text{local}}(\text{LL}) \tag{8}$$

**Computational Details.** In this section, we explain the details of the specific implementation adopted for the multilevel strategy in this study, particularly regarding the choice of computational methods used for HL and LL levels.

Even though the effective energy used in LL computations might a priori be determined using a classical force field, which would warrant a less expensive exploration, the current implementation of the multilevel strategy uses the semi-empirical RM1 Hamiltonian[32] implemented in a locally modified version of MOPAC 6.0. This choice is motivated by the fact that the RM1 method retains the formalism of AM1, PM3, and PM5 Hamiltonians but has been reparametrized using data of 1736 compounds relevant in organic and biochemical areas, leading to an improved accuracy compared to the former semiempirical methods. More importantly, this choice allows us to take advantage of the recent implementation of the MST continuum model in the RM1 framework for the calculation of solvation free energies,[33] thus leading to a consistent evaluation of the two contributions to $E^{\text{eff}}(r_{\text{int}})$ (eq 7). Finally, a practical advantage of this choice with regard to standard force fields is that it eliminates the substantial effort required for a precise parametrization of the ligands, which should a priori encompass a large variety of chemical scaffolds in their molecular structure.

The number of internal rotations can be a crucial factor in choosing the most suitable approach to identify the conformational wells, especially for ligands with rugged conformational landscapes. Thus, whereas a systematic search can be done for compounds with few rotatable bonds, enhanced sampling methods can be more effective for highly flexible drug-like compounds.[34] For our purposes here, the conformational minima have been located by using a Metropolis Monte Carlo (MC) simulation, where the conformational space is sampled exploring a set of active torsions, which includes the rotatable bonds that contribute to the conformational flexibility of the molecule. A restrained energy minimization is performed for each structure of the ligand, so that bond distances and the set of active torsions randomly selected in the MC search are fixed, whereas the other internal degrees of freedom are energy minimized. The restraint of bond lengths is motivated by the relatively frequent proton transfers found in gas phase geometry optimizations of compounds containing charged groups of opposite sign,[21] which would then lead to artifactual geometries of no biological relevance for flexible, charged compounds. Acceptance or rejection of the optimized structure is then performed on the basis of the RM1 energy of the molecule for the gas phase sampling or the effective functional that combines the gas phase RM1 potential and the hydration contribution determined at the MST-RM1 level (eq 7).

To determine the extent of a conformational well, the results from the simulations are projected onto a regularly spaced $n$-

dimensional grid (*n* being the number of active torsions), so that each conformation (characterized by a set of active torsions) is assigned to a given element of the grid. When two or more structures are assigned to the same grid element, then only the structure with the lowest energy is retained as a representative conformation for the corresponding set of conformational variables. Then, the lowest energy structure in the grid is identified and used as a starting point for a systematic search performed along each of the *n*-dimensional torsions around the minimum energy structure in the well. When the search leads to a structure with higher energy in the neighboring grid element, it is assumed that it pertains to the same conformational well of the minimum energy conformer. The search is ceased when the change along a given dimension leads to a new grid element where the energy of the representative structure is lower compared to the energy of the preceding grid element. Alternatively, the search is also stopped when it leads to a grid element that contains no representative structure. This situation might happen when the MC simulation does not sample high-energy conformations, which should have a negligible contribution to the conformational distribution of the compound. Finally, the extension of the conformational well is estimated on the basis of the values of the active torsion corresponding to the outermost grid elements of the curvature of the well in the torsional space. The whole process is then repeated by searching for the lowest-energy conformation upon exclusion of the conformational space already assigned to the previously found energy minima. For our purposes here, calculations were performed using a grid step of 10°, which was chosen as a compromise between computational efficiency in the analysis and accuracy in the population of the conformational families.

The HL refinement of the structures corresponding to the minimum energy conformers in solution is performed by geometry optimizations carried out using the IEF-MST continuum model parametrized at the B3LYP/6-31G(d) level[22] (for the exploration in the gas phase, geometry optimizations were carried out at the B3LYP level). In all cases, the minimum energy nature of the stationary points was verified by analysis of the vibrational frequencies.

Choice of the B3LYP method was motivated by its widespread application to the study of bioorganic molecules but especially due to the availability of the parametrized version of the IEF-MST continuum model, which should thus afford an internally consistent treatment in the gas phase and in solution. Nevertheless, it is known that this functional provides a poor description of dispersion,[35,36] and refinement of the internal energy might be convenient for a proper energetic balance of the conformational minima. This might be achieved by the addition of empirical correction terms as implemented in B3LYP-D[37] or by using other functionals that have addressed specifically the treatment of dispersion energy, as the suite of M05 and M06 methods.[38,39] For our purposes here, however, the internal energy of the molecule was refined by single-point calculations at the MP2/aug-cc-pVDZ level, which was considered to be a reliable level for the prediction of conformational energies based on the aforementioned study by Riley et al.[28] Additional calculations were also performed at the MP2/aug-cc-pVTZ level to check the convergence of the results.

On the basis of the preceding considerations, the implementation of the multilevel strategy adopted in this study to estimate the free energy of a conformational well in the gas phase or in solution is given by eqs 9 and 10, respectively. In these equations, the term $E_{i,\min}(HL)$ shown in eq 8 includes the contributions $E_i^{MP2}$ and $ZPE_i^{B3LYP}$ in eq 9 and $E_i^{MP2}$, $ZPE_i^{IEF-MST//B3LYP}$, and $\Delta G_{hyd,i}^{IEF-MST}$ in eq 10, whereas the last term corresponds to the local conformational contribution $(A_i^{local}(LL))$.

$$G_i^{MULTI} = E_i^{MP2} + ZPE_i^{B3LYP} + G_i^{RM1,local} \qquad (9)$$

where the terms on the right-hand side stand for the energy of the minimum determined at the MP2/aug-cc-pVDZ (or MP2/aug-cc-pVTZ) level, the zero-point energy correction at the B3LYP/6-31G(d) level, and the local free energy contribution of well *i* determined from the RM1 sampling.

$$G_i^{MULTI} = E_i^{MP2} + ZPE_i^{IEF-MST//B3LYP} + \Delta G_{hyd,i}^{IEF-MST}$$
$$+ G_i^{MST-RM1,local} \qquad (10)$$

where the terms on the right-hand side stand for the energy of the minimum determined at the MP2/aug-cc-pVDZ (or MP2/aug-cc-pVTZ) level, the zero-point energy correction at the IEF-MST/B3LYP/6-31G(d) level, the hydration free energy derived from IEF-MST computations, and the local free energy contribution of well *i* determined from the MST-RM1 sampling.

## ■ RESULTS

The suitability of the multilevel strategy outlined above for determining the conformational preferences of molecules has been calibrated considering two distinct areas of potential application. First, the exploration of the conformational landscape of flexible compounds, and second the estimation of the energy cost associated with selection of the bioactive conformation. In the former case, the conformational preferences of 1,2-dichloroethane and neutral histamine have been determined. Whereas 1,2-dichloroethane represents a simple case, our results can be compared with a large number of experimental and computational data reported in the literature (see below). In contrast, neutral histamine represents a more challenging test due to the presence of polar groups that can form intramolecular interactions, as well as to the influence exerted by tautomerism of the imidazole ring on the conformational distribution. With regard to the second scenario, our aim is to check the suitability of the multilevel strategy for predicting the conformational penalty required for selection of the bioactive conformation of ligands. To this end, a set of four HIV reverse transcriptase inhibitors that include a different range of conformational flexibility will be considered.

**1,2-Dichloroethane.** The *trans/gauche* equilibrium of 1,2-dichloroethane has been well established in different studies performed in the gas phase and in a wide range of solvents.[40−47] It is clear from all of these studies that the *trans* conformer is favored in the gas phase and in apolar solvents, whereas the population of the *gauche* conformer is enhanced as the polarity of the solvent increases, as a consequence of its larger dipole moment.

The balance between *trans* and *gauche* conformers has been determined with the multilevel strategy in both the gas phase and in aqueous solution. To this end, MC simulations ($3 \times 10^5$ configurations) were run by defining a single active torsion (Cl−C−C−Cl), and the conformational landscapes in gas phase (RM1) and in water (MST-RM1) are represented in Figure 2. As expected, there is a change in the relative
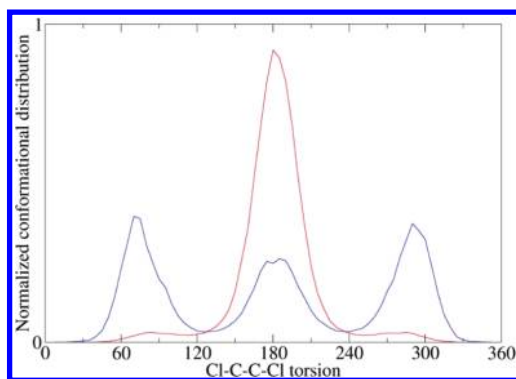
**Figure 2.** Normalized conformational distribution of 1,2-dichloro-ethane in the gas phase (red) and in water (blue) derived from LL sampling (MC simulations at the RM1 and MST-RM1 levels).

population of the two conformers in going from the gas phase to solution, leading to a predominance of the *gauche* conformer in water. Likewise, solvation gives rise to a slight reduction in the dihedral value of the *gauche* conformer, which varies from ~80° for the gas phase to ~70° in water.

From a quantitative point of view (see Table 1), the relative stability between conformers *gauche* and *trans* is predicted to be

**Table 1. Relative Stabilities (kcal/mol) between Conformations *gauche* and *trans* of 1,2-Dichloroethane at 298 K Determined Experimentally, Following the Multilevel Strategy Outlined in This Study and Taken from Previous Computations in the Literature**

| relative stability | gas phase | water |
|---|---|---|
| experimental[a] | 1.3 | 0.0 |
| *multilevel* | | |
| RM1 | 2.0 | 0.0 |
| B3LYP/6-31G(d) | 1.7 | 0.4 |
| MP2/aug-cc-PVDZ | 1.5 | 0.3 |
| MP2/aug-cc-pVTZ | 1.3 | 0.1 |
| *other theoretical studies* | | |
| G2(MP2)[b] | 1.2 | |
| CCSD(T)/cc-pVTZ[c] | 1.5 | |
| B3LYP/6-311+G(d,p), PCM[d] | 1.5 | 0.0 |
| MC, AMBER/OPLS[e] | 1.2 | |
| MC, polarizable force field[f] | | −0.3 |

[a]Refs 40−47. [b]Ref 40. [c]Ref 45. [d]Ref 42. [e]Ref 43. [f]Ref 47.

2.0 and 0.0 kcal/mol in the gas phase and water at the RM1 level, whereas the experimental free energy difference amounts to 1.3 and 0.0 kcal/mol, respectively. Following the *Multilevel* strategy, HL refinement is performed for the minimum energy conformers of *trans* and *gauche* wells, including geometry reoptimization at the IEFMST-B3LYP/6-31G(d) level for water (B3LYP/6-31G(d) in the gas phase) and single-point calculations at the MP2/aug-cc-pVDZ level. The deviation from the experimental values at the B3LYP level is reduced to 0.4 kcal/mol in the gas phase but increases to 0.4 kcal/mol in water (see Table 1). Extension of HL calculations to the MP2/aug-cc-pVDZ level improves the agreement with experimental data, and calculations at the MP2/aug-cc-pVTZ level afford estimates close to the experimental *trans/gauche* stability in the gas phase and water, which compare well with or are even slightly better than the results derived from other QM computations or classical force field simulations (see Table 1). From a practical

point of view, however, the use of MP2/aug-cc-pVTZ calculations for (bio)organic ligands can be very expensive, as the CPU time required for a single-point MP2/aug-cc-pVTZ calculation for 1,2-dichloroethane is ~30-fold larger compared to the MP2/aug-cc-pVDZ cost. Accordingly, the use of the MP2/aug-cc-pVDZ method as a HL reference seems a suitable compromise between chemical accuracy and computational cost, as noted by previous studies.[28]

**Histamine: LL Sampling in the Gas Phase and Aqueous Solution.** Histamine represents an interesting test case, as its chemical structure involves the tautomerism of the imidazole ring (N3H, N1H), as well as the conformational distribution of the ethylamino side chain, which is mainly characterized by the *trans/gauche* balance of the dihedral angle $\phi_2$ (Figure 3). It is noteworthy that such a balance is mediated
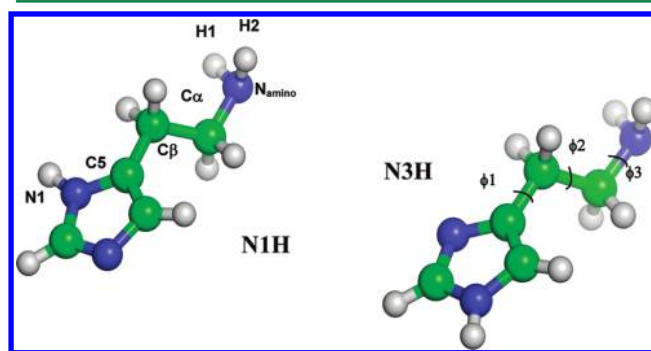


**Figure 3.** Representation of the two tautomeric forms and numbering of the torsional dihedrals that define the conformational space of histamine ($\phi_1$, $N_1$–$C_5$–$C_\beta$–$C_\alpha$; $\phi_2$, $C_5$–$C_\beta$–$C_\alpha$–$N_{amino}$; $\phi_3$, $C_\beta$–$C_\alpha$–$N_{amino}$–$H_1$).

by the formation of an intramolecular hydrogen bond between the amino group and the imidazole ring, as well as by interactions with solvent (water) molecules. Finally, from a computational viewpoint, the correct description of the tautomeric/conformational species of neutral histamine is challenged by the limited accuracy of the RM1 method to predict the tautomerism of the imidazole ring, as the N1H tautomer of 4(5)-methylimidazole is predicted to be favored by 0.5 kcal/mol at the RM1 level, whereas the most stable species at the MP2/aug-cc-pVDZ level is the N3H form (by 0.8 kcal/mol), in agreement with previous theoretical calculations.[48] Therefore, this molecule represents a challenging system to calibrate the robutness of the multilevel strategy to characterize the tautomeric/conformational preferences of flexible compounds.

The conformational exploration of neutral histamine was performed by means of MC simulations ($4 \times 10^5$ configurations) for the two tautomers in the gas phase and in aqueous solution. Inspection of the conformational distribution for the three dihedral angles shows that the *trans/gauche* ratio is larger in water compared to the gas phase (see Figure S1 in Supporting Information), which can be ascribed to the larger hydration of the polar groups present in histamine and to the corresponding reduction of species characterized by the presence of an intramolecular hydrogen bond. This trend is clearly seen in the lengthening of the separation between the N atom of the ethylamino side chain and the N1 atom of the ring (see Figure 4).

The effect of hydration on the conformational landscape of neutral histamine can be observed upon inspection of Figures 5
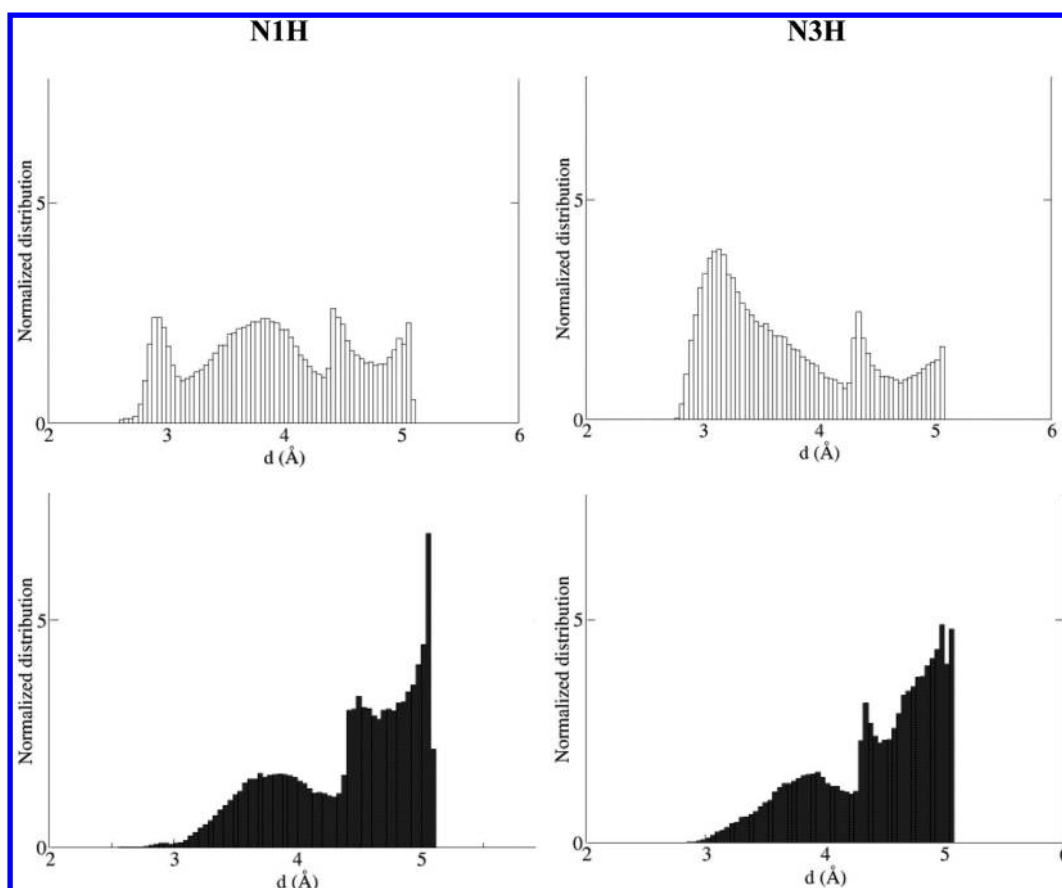
**Figure 4.** Normalized distribution (%) of the distance ($d$; in Å) between the N atom of the ethylamino side chain and the N1 atom of the ring in the gas phase (top) and water (bottom) for N1H and N3H tautomers.
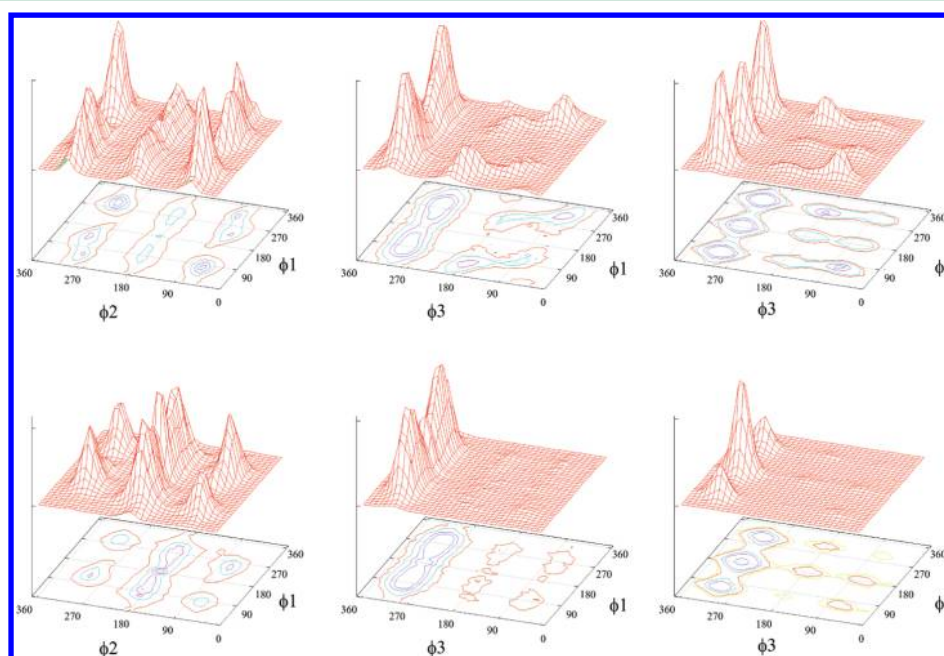


**Figure 5.** Conformational landscape at a low level (RM1 and RM1 in the gas phase and aqueous solution, respectively) of the N1H tautomer of neutral histamine in the gas phase (top) and water (bottom).

(N1H tautomer) and 6 (N3H tautomer). For the N1H tautomer in the gas phase (Figure 5, top panel), the ethylamino group is mainly oriented normal to the plane of the imidazole ring ($\phi_1$ centered around 90° and 270°), whereas the side chain

is primarily found in a *gauche* conformation ($\phi_2$ around 60° and 300°). Hydration triggers a small shift in the torsional angle $\phi_1$, which tends to populate values around 100 and 260° (Figure 5, bottom panel), and a slight enhancement in the population of
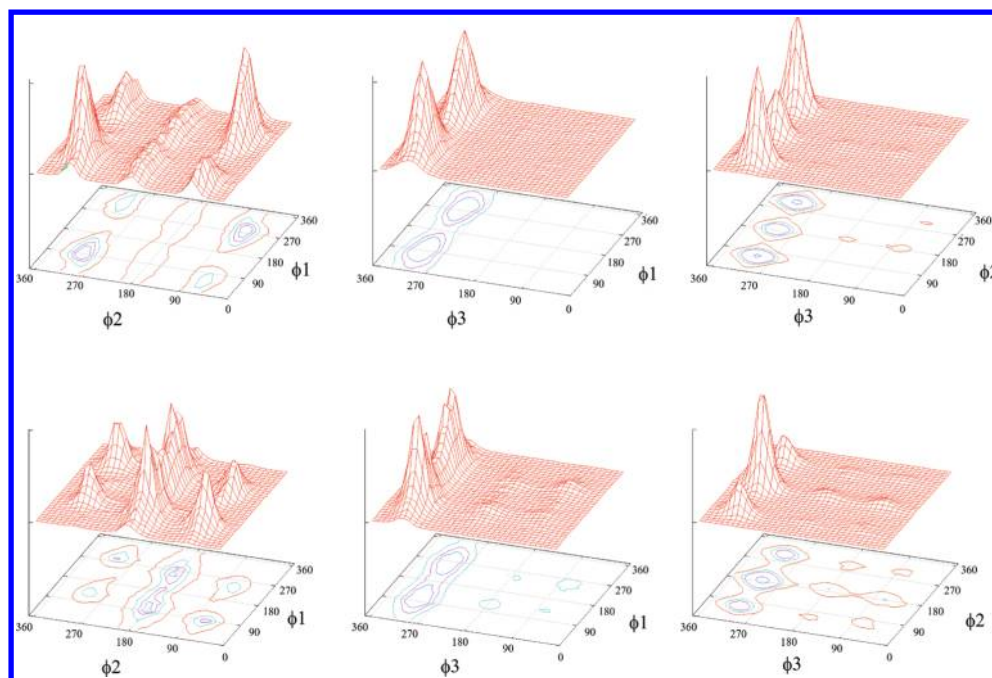
**Figure 6.** Conformational landscape at la ow level (RM1 and RM1 in the gas phase and aqueous solution, respectively) of the N3H tautomer of neutral histamine in the gas phase (top) and water (bottom).

conformers with the side chain lying close to the ring plane ($\phi_1$ around 180°). However, the most relevant trend is the increased populations of species that adopt an extended conformation of the side chain ($\phi_2$ around 180°). Finally, minor peaks found for the dihedral angle $\phi_3$ in the gas phase (at around 60° and 180°; see Figure 5, top panel) are almost annihilated in water, where this torsion mainly adopts values around 300° (i.e, the lone pair is oriented *trans* to $C_\beta$).

With regard to the N3H tautomer (Figure 6, top pannel), in the gas phase, the side chain is mainly oriented normal to the imidazole ring ($\phi_1$ around 90° and 270°) and folded ($\phi_2$ around 60° and 300°). Moreover, there is a correlation between the population of these two torsions, as noted in the peaks centered around $\phi_1 \sim 90°$ and $\phi_2 \sim 300°$, as well as between $\phi_1 \sim 270°$ and $\phi_2 \sim 60°$, which reflects the predominance of conformers with an internal hydrogen bond. These conformers are largely penalized by hydration, which renders the *trans* conformer the most populated species of the N3H tautomer.

**Conformational/Tautomeric Preferences of Neutral Histamine in the Gas Phase.** The conformational preferences predicted for histamine in the gas phase can be compared with the experimental data determined by free-expansion jet spectroscopy.[49,50] These experiments showed that histamine populates four main *gauche* conformers: one of the N1H tautomer and three of the N3H one (see Figure S2 in Supporting Information). The most populated conformers, N1H G-Iva and N3H G-Ib, have similar populations, whereas the abundances of N3H G-Vc and N3H G-Ic were estimated to be around 40% and 30%, respectively, of N1H G-Iva or N3H G-Ib (here, we follow the nomenclature used by Godfrey and Brown, though torsional angles of these species are given in Table 2).[49]

Multilevel calculations in the gas phase were performed by combining the LL RM1 sampling with HL calculations at the B3LYP/6-31G(d) and MP2/aug-cc-pVDZ levels. The results reveal the inadequacy of the RM1 method to provide even a qualitative description of the relative abundances between

**Table 2. Population (%) of the Main Conformations Found for Neutral Histamine in the Gas Phase from Free Expansion Jet Experiments and Multilevel Calculations**

| tautomer | N1H | N3H | | | |
|---|---|---|---|---|---|
| conformer (%)[a] | G-Iva | G-Ib | G-Vc | G-Ic | total[b] |
| experimental[c] | 37 | 37 | 15 | 11 | 100 |
| | | multilevel[d] | | | |
| RM1 | 14.2 | 0.2 | 8.1 | 2.7 | 25.0 |
| B3LYP/6-31G(d) | 84.3 | 0.9 | 5.0 | 1.3 | 91.5 |
| MP2/aug-cc-pVDZ | 62.8 | 10.8 | 3.2 | 5.2 | 82.0 |

[a]Torsion angles $\phi_1$, $\phi_2$, and $\phi_3$ (degrees) for the different conformations at the B3LYP/6-31G(d) level: G-Iva (43, 295, 174), G-Ib (71, 66, 181), G-Vc (59, 294, 297), and G-Ic (65, 63, 294). Mirror images are not indicated. [b]Contribution of the four conformers to the total conformation space. [c]Determined in a free-expansion jet spectrometer at 403 K. [d]Determined at 298 K.

conformers. Thus, though the four experimental conformers are sampled, their population only amounts to 25% of the total conformers in the gas phase. In fact, the most abundant conformers are predicted to be the species N1-H G-Ic (defined by torsions $\phi_1$, $\phi_2$, and $\phi_3$ centered around 70, 50, and 290°) and N1-H G-IIIc (torsions around 220, 60, and 290°), which are populated at 25% and 16%, respectively. The *trans* conformer T-b (torsions around 70, 170, and 290°) is also significantly populated (17%). Overall, these findings point out the shortcoming of the RM1 Hamiltonian for exploring the conformational preferences of a challenging system that involves both conformational and tautomeric changes in the chemical structure.

Compared to the RM1 results, a significant improvement is found upon rescaling of the conformational minima using B3LYP/6-31G(d) calculations as the HL method. Thus, the four experimental conformers cover around 90% of the conformational population. However, the conformer N3-H G-Ib, which is experimentally found to be the most populated

species for the N3-H tautomer, has a very minor population (less than 1%), while the G-Vc species is predicted to be the most abundant conformer of this tautomer. This trend is corrected when HL calculations are performed at the MP2/aug-cc-pVDZ level, as the conformer G-Ib becomes the most abundant species for the N3H tautomer (near 11%), in agreement with the experimental evidence.

At the MP2/aug-cc-pVDZ level, the four experimentally major conformers account for 82% of the total population of neutral histamine, which indicates the existence of minor fractions of other species. The population of these minor conformations is generally less than 1%, but a non-negligible population (5.6%) is found for a conformational well of the N1H tautomer characterized by torsion angles $\phi_1$, $\phi_2$, and $\phi_3$ of 67, 305, and 90° (see Figure S2 in the Supporting Information). Since the abundance of this conformer compares with the population estimated for N3H G-Vc and N3H G-Ic, one might ask why it was not observed in the experimental measurements. It might be due to the different temperature at which simulation (298 K) and experiments (403 K) were carried out, but we also hypothesize that this conformer relaxes to the conformer N1H G-Iva in the expanding jet, as the energetic barrier for such conformational conversion, which implies a rotation around $\phi_3$, is estimated to be around 0.4 kcal/mol (see Figure S3 in Supporting Information). At this point, let us note that relaxations from N3-H G-Vb to N3H G-Ib, and from N3H G-Vc to N3H G-Ic, in the jet expansion were proposed by Godfrey and Brown to rationalize the population of conformers for the N3H tautomer.[50]

It is worth noting that simulations suggest that the population of the N1H G-Iva conformer at 298 K is 1.7-fold larger compared to the experimental abundance measured. In turn, the population of the N3H G-Ib conformer is lower by a factor of 2.5. By taking into account all the conformational wells sampled for the two tautomers, the N1-H tautomer turns out to be favored by around 0.4 kcal/mol, whereas the experimental data indicate that the N3-H species is preferred by 0.3 kcal/mol. To shed light onto the origin of this tautomeric change, we have compared the multilevel results with the population distribution estimated by adopting the harmonic oscilator-rigid rotor model for a subset of predefined conformations (see Table 3). On the basis of these calculations, the population of the most abundant conformer (N1H G-Iva) varies only marginally from 62.8% to 77.2%, which would correspond to a free energy change of 0.1 kcal/mol (at 298 K). This finding suggests that the contribution of high-energy conformational wells can be neglected, thus supporting the predominant-state approximation adopted in the multilevel strategy. On the other hand, extension of the HL calculations to the MP2/aug-cc-pVTZ level does not introduce notable changes in the abundance of the conformers, which reinforces the notion that the MP2/aug-cc-pVDZ method offers a good compromise between computational cost and chemical accuracy.

The population predicted by using the harmonic oscillator-rigid rotor model has been investigated by combining single-point MP2/aug-cc-pVDZ with free energy corrections determined for geometries optimized at the B3LYP/6-31G(d) level (compare rows 4 and 7 in Table 3). The harmonic approximation leads to conformer abundances that reflect the trends derived from multilevel calculations. Thus, the population of the four conformers differs by factors ranging from 0.8 to 2.0 (i.e., changes in the free energy varying from 0.1 to −0.4 kcal/mol at 298 K). On the other hand, the abundances

**Table 3. Population (%) of the Main Conformations Found for Neutral Histamine in the Gas Phase Determined from Multilevel Calculations and by Using the Harmonic-Oscilator-Rigid Rotor Model[a]**

| conformer | N1H G-Iva | N3H G-Ib | N3H G-Vc | N3H G-Ic |
|---|---|---|---|---|
| multilevel | | | | |
| MP2/aug-cc-pVDZ (all sampled conformers)[b] | 62.8 | 10.8 | 3.2 | 5.2 |
| MP2/aug-cc-pVDZ | 77.2 | 13.0 | 3.5 | 6.3 |
| MP2/aug-cc-pVTZ | 71.1 | 15.9 | 5.1 | 7.9 |
| harmonic oscilator-rigid rotor model | | | | |
| MP2/aug-cc-pVDZ//B3LYP/6-31G(d) | 63.7 | 20.2 | 3.5 | 12.6 |
| MP2/aug-cc-pVDZ//MP2/aug-cc-pVDZ | 67.1 | 17.3 | 5.7 | 9.9 |
| MP2/6-311++G(d,p)//MP2/6-31G(d) | 62.5 | 16.8 | 2.7 | 12.1 |
| MP2/6-311++G(d,p)//MP2/6-31G(d) (403 K)[c] | 45.0 | 28.1 | 5.2 | 21.7 |
| MP2/6-311++G(d,p)//MP2/6-31G(d) (403 K)[d] | 37.2 | 23.3 | 4.3 | 18.0 |
| MP2/6-311++G(d,p)//MP2/6-31G(d) (403 K)[e] | 37.2 | 27.7 | 13.0 | 9.0 |

[a]Unless explicitly noted, all of the values are determined considering a conformational space limited to the main four conformations of neutral histamine and a temperature of 298 K. [b]Population determined taking into account all of the conformational wells sampled in multilevel conformational search. [c]Thermal (free energy) corrections determined using a scaling factor of 0.991 (as noted in ref 50) for vibrational frequencies and at a temperature of 403 K. [d]Population determined for a set of 10 conformations at 403 K without the contribution due to conformational relaxation of N3H G-Vb to N3H G-Ib and of N3H G-Vc to N3H-G-Ic (see ref 50). [e]Population determined by including the conformational relaxation of N3H G-Vb to N3H G-Ib and of N3H G-Vc to N3H-G-Ic (see ref 50 for details).

obtained from thermal corrections determined at B3LYP/6-31G(d) or MP2/aug-cc-pVDZ levels are very similar (compare rows 7 and 8 in Table 3).

Table 3 also reports the abundances predicted at the MP2/6-311++G(d,p)//MP2-6-31G(d) level (rows 9−12 in Table 3), which was used by Godfrey and Brown in their conformational analysis of histamine.[49] At 298 K, the population of the four conformers agrees with the MP2/aug-cc-pVDZ results, but increasing the temperature to 403 K has a significant influence on reducing the population of N1H G-Iva and enhancing the population of N3H G-Ib. In fact, increasing the simulation temperature to 403 K leads to a conformer distribution closer to the experimental values (see Table 2). Though the magnitude of this effect is moderate (equivalent to free energy changes of 0.3 kcal/mol), it suffices to revert the tautomeric balance, which favors the N1−H species by 0.3 kcal/mol at 298 K and the N3-H tautomer by 0.2 kcal/mol at 403 K. If one takes into account the contribution due to seven minor conformers examined in ref 50 (row 11 in Table 3), and the allowance of conformational relaxation between N3H G-Vb and N3H G-Ib and that between N3H G-Vc and N3H-G-Ic, a nice agreement is found with the experimental abundances (see ref 50 for details).

**Conformational/Tautomeric Preferences of Neutral Histamine in Aqueous Solution.** Unfortunately, the experimental information about the tautomeric/conformational preferences of neutral histamine in solution is less detailed than in the gas phase. Kraszni et al.[51] have reported that *trans*

conformers populate around 41% of neutral histamine in water, which is similar to the population (ranging from 38% to 47%)[51−53] found for the monocationic histamine and slightly lower than the value reported for the dication (between 48% and 55%),[52,53] thus indicating a small dependence of the conformational distribution on pH. With regard to tautomerism, several studies have indicated that the N3H tautomer is favored in aqueous solution at basic pH.[54−56] The ratio between N3H and N1H tautomers has been estimated to be around 80:20,[54,55] which is close to the tautomeric ratio determined for monocationic histamine,[56] as well as for histidine in basic solution (N3H/N1H ratio of 65:35).[57,58]

In contrast with the relative simplicity of the conformational preferences of neutral histamine in the gas phase, the multilevel analysis in aqueous solution reveals a much larger complexity for the conformational distribution of the ethylamino side chain of histamine. Thus, there is a plethora of wells in a narrow range of stability (between 1.9 and 3.1 kcal/mol when one excludes the less stable conformers), which indicates that the intrinsic conformational preferences of histamine cannot be simply described by a reduced number of conformers. Thus, the largest population of a conformational well amounts to 11%, and there are eight conformers with a population ranging from 7% to 4% (see Table S1 in the Supporting Information). This finding highlights the difficulty in choosing a priori a reduced set of representative conformations for flexible ligands in solution.

The multilevel results also show that there is a delicate balance among the different conformational and tautomeric species in water (see Table 4). Thus, when HL computations

**Table 4. Relative Population of Conformational/Tautomeric Species of Neutral Histamine in Water Determined from Multilevel Computations**

| tautomer/conformer | | B3LYP/6-31G(d) | MP2/aug-cc-pVDZ | MP2/aug-cc-pVTZ |
|---|---|---|---|---|
| N1H | gauche | 24 | 32 | 31 |
| | trans | 24 | 16 | 17 |
| N3H | gauche | 12 | 21 | 21 |
| | trans | 40 | 31 | 31 |
| total ratio trans/gauche | | 64/36 | 47/53 | 48/52 |
| total ratio N1H/N3H | | 48/52 | 48/52 | 48/52 |

are performed at both the MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ levels, a similar population is predicted for both trans and gauche conformers, which is in contrast with the situation found in the gas phase, but in agreement with the behavior of the ethylamino side chain in water.[51−53] This finding reflects the large effect of hydration on the conformational preferences of histamine, which increases substantially the stability of the trans conformers.

With regard to the tautomeric preferences, calculations predict a similar population for the two tautomers. As noted above, experimental studies suggest the predominance of the N3H tautomer. If one takes into account the N3H/N1H ratio estimated by Reynolds and Tzeng,[54] the population of the N3H species is underestimated by 0.3 kcal/mol, whereas the stability of the N1H tautomer is overestimated by 0.5 kcal/mol. This error reflects the difficulty of theoretical calculations in accounting for the subtle balance of effects that modulate the tautomeric/conformational preferences of histamine in sol-

ution, as small uncertainties in the estimation of the different contributions to the free energy can affect the "chemical" accuracy in predicted conformational/tautomeric preferences. Nevertheless, the suitability and robustness of the multilevel strategy can be assessed by comparison with previous theoretical studies. Thus, Nagy et al.[59] reported that the tautomeric equilibrium of neutral histamine favored the N3H form (95%), though their results pointed out that histamine primarily populates trans conformations (trans, 83%; gauche, 12%), whereas Ramírez and co-workers[60] reported a predominance of the N3H-gauche conformations by around 2 kcal/mol.

**Selection of Bioactive Conformation.** The characterization of the conformational landscape of flexible ligands allows the estimation of the free energy cost required for the selection of the bioactive conformation implicated in ligand binding. To this end, we have revisited the set of four HIV reverse transcriptase inhibitors examined in a previous study: MKC-442, UC781, trovirdine, and nevirapine.[20]

The MC MST-RM1 exploration (simulations comprised $10^5$−$10^6$ MC steps) reflected the complexity of the conformational space of these compounds, as we found a total of 57, 120, 53, and 3 conformational wells for MKC-442, UC781, trovirdine, and nevirapine, respectively. It should be remarked that the conformational space determined from these computations is smaller than considering it proportional to rotamer counting ($N_{conf} \alpha 3^{N_{rot}}$, where $N_{conf}$ and $N_{rot}$ denote the expected number of conformers and the number of rotatable bonds in the molecule, and one assumes three conformational states per rotatable bond), in agreement with what has been found in other systems.[61,62]

For nevirapine, the multilevel strategy indicated that the most stable conformation corresponds to the bioactive conformation (the RMSD between the global minimum and the X-ray structures is 0.07 Å), whereas the other two conformers are destabilized by around 5 kcal/mol. For the rest of the compounds, Figure 7 shows representative structures corresponding to the global minima of MKC-442, UC781, and trovirdine in water, as well as the conformer most similar to the bioactive conformation found in the X-ray complexes. The global minima found at the MP2/aug-cc-pVDZ and B3LYP levels for MKC-442 are very similar, as they only differ in the orientation of the isopropyl chain attached to the heterocyclic ring. However, there is less resemblance for the global minima found in water at the B3LYP/6-31G(d) and MP2/aug-cc-pVDZ levels for UC781 and trovirdine (see Figure 7). The different geometrical arrangement can likely be attributed to the poor description of dispersion by B3LYP compared to the MP2, as the fomer tend to favor "open" conformations.

At the MP2/aug-cc-pVDZ level, the global minimum of MKC-442 is found to be the bioactive conformation (the RMSD between the global minimum and the X-ray structure is only 0.27 Å), and accordingly there should be no significant reorganization penalty upon binding to the HIV reverse transcriptase (see Table 5). In the case of trovirdine, the bioactive conformation is achieved by a simple rotation around the side chain that links the piridine ring to the thiourea moiety, and the free energy difference between the bioactive and global minima is estimated to be 0.3 kcal/mol (the RMSD between the bioactive conformation and the X-ray structure is 0.22 Å). Finally, adoption of the bioactive conformation of UC781 implies a larger conformational rearrangement of the furan ring attached to the thioamide moiety, even though in this case the
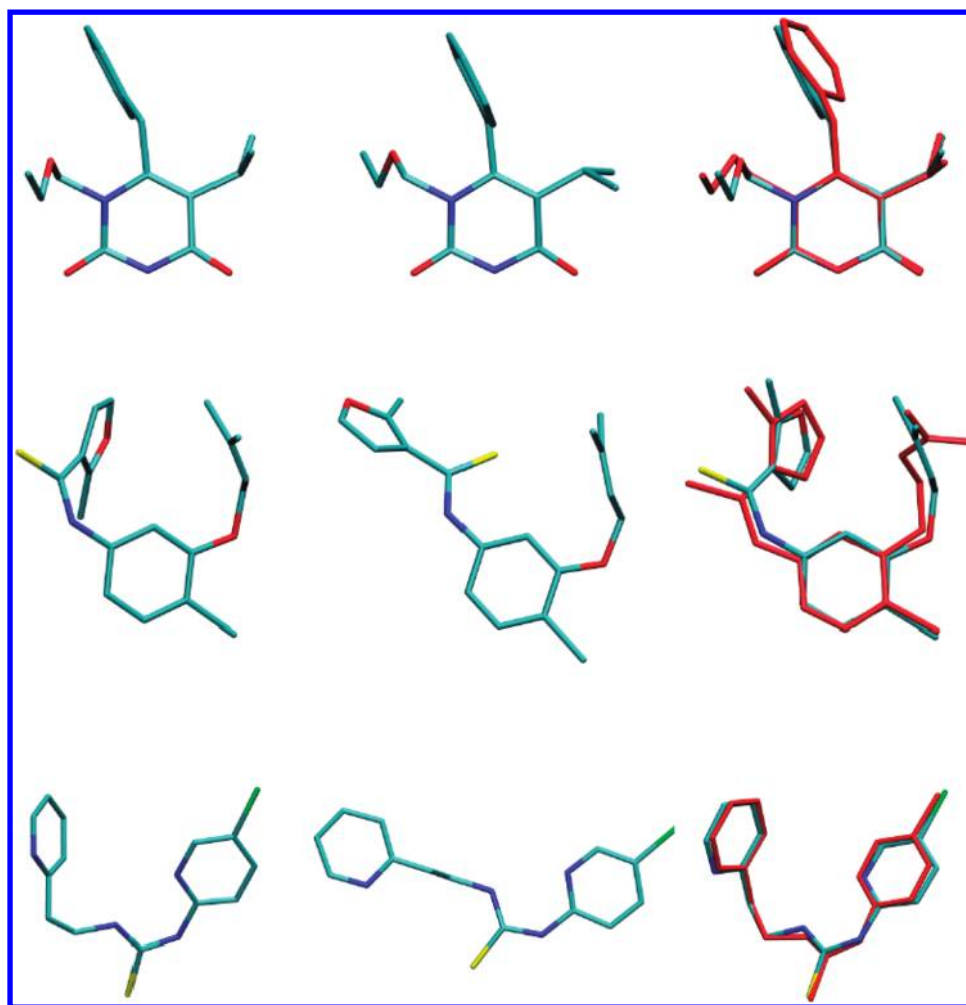
**Figure 7.** Representative conformation of MKC-442 (*top*), UC781 (*middle*), and trovirdine (*bottom*). For each compound, the global minima determined in water with the multilevel strategy at the MP2/aug-cc-pVDZ and B3LYP levels are shown on the left and central columns, respectively, whereas the superposition of the conformation most structurally similar to the X-ray (bioactive) one is shown in the right column.

free energy difference between bioactive and global minima is estimated to be 1.3 kcal/mol (the RMSD between the bioactive conformation and the X-ray structure is 0.8 Å). The preceding results indicate that the reorganization cost required for

**Table 5. Conformational Penalty (kcal/mol) for Selecting the Bioactive Conformation of MKC-442, UC781, Trovirdine, and Nevirapine[a]**

|  | MKC-442 | UC781 | trovirdine | nevirapine |
|---|---|---|---|---|
| reorganization free energy ($\Delta G_{reorg}$) | | | | |
| RM1 | 0.7 | 3.3 | 1.4 | 0.0 |
| B3LYP/6-31G(d) | 0.7 | 1.4 | 1.5 | 0.0 |
| MP2/aug-cc-pVDZ | 0.0 | 1.3 | 0.3 | 0.0 |
| loss of conformational states ($\Delta G_{cl}$) | | | | |
| MP2/aug-cc-pVDZ | 1.0 | 0.9 | 1.3 | 0.0 |
| conformational cost ($\Delta G_{reorg} + \Delta G_{cl}$) | | | | |
| MP2/aug-cc-pVDZ | 1.0 | 2.2 | 1.6 | 0.0 |

[a]The reorganization energy is estimated relative to the global minimum determined from the sampling in water at the RM1, B3LYP/6-31G(d), and MP2/aug-cc-pVDZ levels. A combination of reorganization energy and loss of conformational states upon binding (estimated by using the MP2/aug-cc-pVDZ energies of the minima) affords the conformational penalty.

selection of the bioactive conformation for the set of HIV reverse transcriptase inhibitors is generally small, as the global minimum agrees with the bioactive conformation in two cases and the conformational penalty is estimated to be 0.3 kcal/mol in a third case.

The conformational cost associated with the binding of the bioactive conformation must also take into account the free energy penalty due to the loss of conformational states upon ligand transfer from the unbound state (in solution) to the target complex. If we assume that the bound state is described by a single conformation (i.e., the bioactive conformer) and that it pertains to the set of conformations sampled in solution, this contribution can be estimated by using the expression $\Delta G_{cl} = -RT \ln \sum_j e^{-\Delta E_j / RT}$, where $\Delta E_j$ stands for the relative energy between the local minimum conformer $j$ and the lowest energy conformer. By using the MP2/aug-cc-pVDZ energies for the minimum energy structures of the conformational wells, the $\Delta E_{cl}$ term is estimated to vary from 1.0 to 1.3 kcal/mol for MKC-442, UC781, and trovirdine, while it is estimated to be negligible for nevirapine (see Table 5). Let us note that these values should be regarded as an upper estimate, as an accurate estimate of the contribution due to the loss of conformational states would require knowledge of the whole conformational space available for the ligand in the bound state. In addition, with the exception of nevirapine, which only has one rotatable

bond, the magnitude of $\Delta E_{cl}$ compares with or is larger than the reorganization cost, which suggests that restricting the conformational space of flexible ligands can be beneficial for improving the biological activity by reducing the entropic cost due to conformer focusing.

Overall, the selection of the bioactive conformation would be associated with a free energy cost less than 1.6 kcal/mol for nevirapine, MKC-442, and trovirdine and around 2.2 kcal/mol for UC781. This finding is consistently closer to chemical intuition, as present results suggest that selection of the bioactive conformation should not be generally a major contribution to the binding affinity.

## CONCLUSIONS

In this work, we have presented the methodology and validation of a conformational search strategy that relies on the use of different levels of theory for the sampling of conformers and rescaling of the minima for predicting the population distribution. Choice of a LL method for the exploration and identification of the conformational wells is justified by the need to keep a reduced cost, which should permit the conformational analysis of drug-like flexible molecules, whereas the combination of a HL method for minima rescaling should correct the deficiencies of the LL method and permit an accurate estimate of the conformational preferences.

The implementation of the multilevel strategy adopted in this study is supported by the analysis of the conformational preferences of 1,2-dichloroetane and neutral histamine in both the gas phase and solution. This strategy has proved successful in reflecting the change introduced by solvation on the conformational preferences of these compounds, which is especially notable in the case of histamine, where the distribution of conformers in the ethylamino side chain reflects the complex scenario defined by tautomerism, intramolecular hydrogen bonding, and solvation. Such a complexity is illustrated in the enhanced diversity and richness of the conformational landscape in solution compared to the relative simplicity of the conformational distribution in the gas phase. In turn, this finding has implications for the challenging task of selecting the bioactive conformations of drug-like compounds. In particular, the results reported here for a small set of HIV reverse transcriptase inhibitors confirm chemical intuition and illustrate that even in the case of flexible ligands the conformational penalty associated with the selection of the bioactive species is not dramatic for highly potent drugs.

From a methodological point of view, these results pave the way for the exploration of novel improvements that afford a reduction in computational cost and an increase in chemical accuracy. For instance, though the RM1 Hamiltonian was chosen as a LL method, the conformational sampling could benefit from the use of classical force fields, which would reduce substantially the computational burden needed for identification of conformational wells for very flexible molecules. In turn, this savings in computational cost should permit the use of HL QM methods that provide a more balanced description of interactions between molecular fragments. These aspects will be considered in future studies. On the other hand, even though the application of the multilevel methodology implemented here to high-throughput virtual screening of compound libraries is not feasible due to substantial computational cost, the identification of the most populated conformers for a representative set of flexible

molecules and the determination of the reorganization cost can be valuable to refining less sophisticated but faster approaches developed for the searching of bioactive conformations.

## ASSOCIATED CONTENT

### ⓈSupporting Information

Population distribution and representation of selected conformational minima are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author
*E-mail: fjluque@ub.edu, xbarril@ub.edu.

### Present Address
[⊥]Instituto de Biomedicina de Buenos Aires—Max Planck Society Partner (IBioBA-MPSP) Godoy Cruz 2201 C1425FQA, Buenos Aires, Argentina

### Notes
The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

## REFERENCES

(1) Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644–2676.

(2) *Protein-Ligand Interactions. From Molecular Recognition to Drug Design*; Böhm, H. J., Schneider, G., Eds.; Wiley-VCH: Weinheim, Germany, 2003.

(3) Bissantz, C.; Kuhn, B.; Stahl, M. *J. Med. Chem.* **2010**, *53*, 5061–5084.

(4) Hunter, C. A. *Angew. Chem., Int. Ed.* **2004**, *43*, 5310–5324.

(5) Spyrakis, F.; Bidon-Chanal, A.; Barril, X.; Luque, F. J. *Curr. Top. Med. Chem.* **2011**, *11*, 192–210.

(6) Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213–218.

(7) Cavasotto, C. N.; Singh, N. *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 221–234.

(8) Boström, J. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1067–1086.

(9) Boström, J.; Greenwood, J. R.; Gottfries, J. *J. Mol. Graphics Modell.* **2003**, *21*, 449–462.

(10) Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. *J. Chem. Inf. Model.* **2006**, *46*, 1848–1861.

(11) Li, J.; Ehlers, T.; Sutter, J.; Varma-O'Brien, S.; Kirchmair, J. *J. Chem. Inf. Model.* **2007**, *47*, 1923–1932.

(12) Loferer, M.; Kolossváry, I.; Aszódi, A. *J. Mol. Graphics Modell.* **2007**, *25*, 700–710.

(13) Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E. *J. Chem. Inf. Model.* **2007**, *47*, 1067–86.

(14) Brameld, K. A.; Kuhn, B.; Reuter, D. C.; Stahl, M. *J. Chem. Inf. Model.* **2008**, *48*, 1–24.

(15) Liu, X.; Bai, F.; Ouyang, S.; Wang, X.; Li, H.; Jiang, H. *BMC Bioinformatics* **2009**, *10*, 101.

(16) Takagi, T.; Amano, M.; Tomimoto, M. *J. Chem. Inf. Model.* **2009**, *49*, 1377–1388.

(17) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. *J. Chem. Inf. Model.* **2010**, *50*, 534–546.

(18) Chung, S.; Parker, J. B.; Bianchet, M.; Amzel, L. M.; Stivers, J. T. *Nat. Chem. Biol.* **2009**, *5*, 407−413.

(19) Perola, E.; Charifson, P. S. *J. Med. Chem.* **2004**, *47*, 2499−2510.

(20) Tirado-Rives, J.; Jorgensen, W. L. *J. Med. Chem.* **2006**, *49*, 5880−5584.

(21) Butler, K. T.; Luque, F. J.; Barril, X. *J. Comput. Chem.* **2009**, *30*, 601−610.

(22) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(23) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Orozco, M.; Luque, F. J. *THEOCHEM* **2005**, *727*, 29−40.

(24) Soteras, I.; Orozco, M.; Luque, F. J. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 281−291.

(25) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem. A* **1997**, *101*, 1609−1618.

(26) Chen, W.; Chang, C.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 3035−3049.

(27) Echenique, P.; Cavasotto, C. N.; García-Risueño, P. *Eur. Phys. J.* **2011**, *200*, 5−54.

(28) Riley, K. E.; Op't Holl, B. T.; Merz, K. M. *J. Chem. Theory Comput.* **2007**, *3*, 407−433.

(29) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615−2623.

(30) Vreven, T.; Morokuma, K. *Annu. Rep. Comp. Chem.* **2006**, *2*, 35−51.

(31) Kim, Y.; Mohrig, J. R.; Truhlar, D. G. *J. Am. Chem. Soc.* **2010**, *132*, 11071−11082.

(32) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101−1111.

(33) Forti, F.; Barril, X.; Luque, F. J.; Orozco, M. *J. Comput. Chem.* **2008**, *29*, 578−587.

(34) Okumura, H.; Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2010**, *31*, 1357−1367.

(35) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

(36) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787−1799.

(37) Civalleri, B.; Zicovich-Wilson, C. M.; Valenza, L.; Ugliengo, P. *CrystEngComm* **2008**, *10*, 405−410.

(38) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

(39) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(40) Wiberg, K. B.; Keith, T. A.; Frisch, M. J.; Murcko, M. *J. Phys. Chem.* **1995**, *99*, 9072−9079.

(41) Kato, M.; Abe, I.; Taniguchi, Y. *J. Chem. Phys.* **1999**, *110*, 11982−11986.

(42) Capelli, C.; Corni, S.; Tomasi, J. *J. Phys. Chem. A* **2001**, *105*, 10807−10815.

(43) Jorgensen, W. L.; McDonald, N. A.; Selmi, M.; Rablen, P. R. *J. Am. Chem. Soc.* **1995**, *117*, 11809−11810.

(44) Vilaseca, E. *J. Chem. Phys.* **1996**, *104*, 4243−4257.

(45) Christiansen, O.; Mikkelsen, K. V. *J. Chem. Phys.* **1999**, *110*, 1365−1375.

(46) Hernandez, B.; Curutchet, C.; Colominas, C.; Orozco, M.; Luque, F. J. *Mol. Simul.* **2001**, *28*, 153−171.

(47) Madurga, S.; Vilaseca, E. *J. Phys. Chem. A* **2004**, *108*, 8439−8447.

(48) Luque, F. J.; López-Bes, J. M.; Cemeli, J.; Aroztegui, M.; Orozco, M. *Theor. Chem. Acc.* **1997**, *96*, 105−113.

(49) Vogelsanger, B.; Godfrey, P. D.; Brown, R. D. *J. Am. Chem. Soc.* **1991**, *113*, 7864−7869.

(50) Godfrey, P. D.; Brown, R. D. *J. Am. Chem. Soc.* **1998**, *12*, 10724−10732.

(51) Kraszni, M.; Kökösi, J.; Noszál, B. *J. Chem. Soc., Perkin Trans. 2* **2002**, 914−917.

(52) Ganellin, C. R.; Pepper, E. S.; Port, G. N. J.; Richards, W. G. *J. Med. Chem.* **1973**, *16*, 610−616.

(53) Ham, N. S.; Casy, A. F.; Ison, R. R. *J. Med. Chem.* **1973**, *16*, 470−475.

(54) Reynolds, W. F.; Tzeng, C. W. *Can. J. Biochem.* **1977**, *55*, 576−578.

(55) Roberts, J. D.; Yu, C.; Flanagan, C.; Birdseye, T. R. *J. Am. Chem. Soc.* **1982**, *104*, 3945−3949.

(56) Ganellin, C. R. *J. Pharm. Pharmacol.* **1973**, *25*, 787−792.

(57) Farr-Jones, S.; Wong, W. Y. L.; Gutheil, W. G.; Bachovchin, W. W. *J. Am. Chem. Soc.* **1993**, *115*, 6813−6819.

(58) Munowitz, M.; Bachovchin, W. W.; Herzfeld, J.; Dobson, C. M.; Griffin, R. G. *J. Am. Chem. Soc.* **1982**, *104*, 1192−1196.

(59) Nagy, P. I.; Durant, G. J.; Hoss, W. P.; Smith, D. A. *J. Am. Chem. Soc.* **1994**, *116*, 4898−4909.

(60) Ramírez, F. J.; Tuñón, I.; Collado, J. A.; Silla, E. *J. Am. Chem. Soc.* **2003**, *125*, 2328−2340.

(61) Chang, C. A.; Chen, W.; Gilson, M. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *105*, 1534−1539.

(62) Anisimov, V. M.; Cavasotto, C. N. *J. Comput. Chem.* **2011**, *32*, 2254−2263.