Predicting the Physicochemical Profile of Diastereoisomeric Histidine-containing Dipeptides by Property Space Analysis

GIULIO VISTOLI,[1]* VALENTINA STRANIERO,[1] ALESSANDRO PEDRETTI,[1] LAURA FUMAGALLI,[1] CRISTIANO BOLCHI,[1] MARCO PALLAVICINI,[1] ERMANNO VALOTI,[1] AND BERNARD TESTA[2]

[1]*Dipartimento di Scienze Farmaceutiche "Pietro Pratesi", Facoltà di Farmacia, Università degli Studi di Milano, Milano, Italy*
[2]*Dept of Pharmacy, Lausanne University Hospital (CHUV), Lausanne, Switzerland*

ABSTRACT ***Objectives*** This study aimed at measuring the lipophilicity and ionization constants of diastereoisomeric dipeptides, interpreting them in terms of conformational behavior, and developing statistical models to predict them.

Methods A series of 20 dipeptides of general structure NH_2-*L*-X-(*L* or *D*)-His-OMe was designed and synthetized. Their experimental ionization constants (pK_1, pK_2 and pK_3) and lipophilicity parameters ($\log P^N$ and $\log D^{7.4}$) were measured by potentiometry. Molecular modeling in three media (vacuum, water, and chloroform) was used to explore and sample their conformational space, and for each stored conformer to calculate their radius of gyration, virtual log P (preferably written as $\log P^{MLP}$, meaning obtained by the molecular lipophilicity potential (MLP) method) and polar surface area (PSA). Means and ranges were calculated for these properties, as was their sensitivity (i.e., the ratio between property range and number of rotatable bonds).

Results Marked differences between diastereoisomers were seen in their experimental ionization constants and lipophilicity parameters. These differences are explained by molecular flexibility, configuration-dependent differences in intramolecular interactions, and accessibility of functional groups. Multiple linear equations correlated experimental lipophilicity parameters and ionization constants with PSA range and other calculated parameters.

Conclusion This study documents the differences in lipophilicity and ionization constants between diastereoisomeric dipeptides. Such configuration-dependent differences are shown to depend markedly on differences in conformational behavior and to be amenable to multiple linear regression. *Chirality 00:000–000, 2012.*

KEY WORDS: lipophilicity; ionization constants; potentiometry; diastereoisomers; conformational space; molecular modeling; molecular lipophilicity potential; polar surface area; molecular sensitivity

INTRODUCTION

Lipophilicity is a key physicochemical property in drug design. It is usually measured and expressed as the partition coefficient (log P), a parameter describing the equilibrium of a given solute between water and an immiscible organic solvent (commonly *n*-octanol, $\log P_{oct}$).[1] Besides its ability to account for hydrophobic contacts in molecular recognition and binding,[2] lipophilicity strongly influences the absorption, distribution, metabolism, and excretion processes that depend on the ability of a molecule to cross biological membranes.[3] Indeed, it proved successful to predict, for example, oral bioavailability, central nervous system penetration, disposition, and volume of distribution.[4–6] Thus, the Lipinski's well-known rule of 5 warns of unsuitable oral bioavailability when the log P of a molecule exceeds 5.[7] Other relevant physicochemical properties, such as aqueous solubility,[8] are closely related to lipophilicity and can be deduced from log P values that indeed encode different recognition forces ranging from ionic contacts to dispersion forces, as evidenced by solvatochromic analyses.[9]

By considering the manifold biological processes in which lipophilicity can find fertile applications, it comes as no surprise that countless approaches to predict log P values have been hitherto proposed in literature.[10,11] Overall, these approaches can be subdivided into (a) bidimensional (2D) or fragmental approaches based on the additive–constitutive character of lipophilicity,[12] and (b) three-dimensional 3D approaches that involve the projection of atomic increments on the molecular surface (or on the space around the molecule, the so-called molecular fields).[13] The main difference between these approaches is that 3D methods can calculate conformer-dependent, virtual log P values, whose reliability has received experimental validation by kinetic nuclear magnetic resonance (NMR) studies,[14] whereas 2D methods rely only on molecular connectivity and reflect neither conformational nor configurational factors.

The possibility of calculating conformer-dependent property values, which indeed characterize all properties computable by molecular interaction fields,[15] allows property variability to be analyzed because flexible molecules obey a one-to-one relation between a given conformer and the resulting property value. The ensemble of all values a

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Giulio Vistoli, Dipartimento di Scienze Farmaceutiche "Pietro Pratesi", Facoltà di Farmacia, Università degli Studi di Milano, Via Mangiagalli, 25, I-20133 Milano, Italy. E-mail: Giulio.Vistoli@unimi.it

Received for publication 17 November 2011; Accepted 19 March 2012
DOI: 10.1002/chir.22056
Published online in Wiley Online Library (wileyonlinelibrary.com).

conformer-dependent property can assume defines the corresponding property space, which can be described by parameters quantifying the variability of property values as well as by descriptors relating the variability of a given property with that of other geometric or physicochemical properties. These last relations lead to the concept of molecular sensitivity, which encodes the ability of a molecule to modulate its properties by varying its conformational profile. The concept (and the descriptors) of property space found noteworthy applications to analyze the dynamic behavior of both small ligands and large biomacromolecules as recently reviewed.[16,17]

Benchmarking analyses revealed that, on average, 2D methods for log P prediction are largely more accurate than 3D ones.[11] On the one hand, this may confirm that the likelihood of errors usually increases with algorithm complexity.[18] On the other hand, this suggests that accounting for conformational factors by considering a single (albeit optimized) conformation is not enough to take the greater informational complexity into account. Thus, a study based on a database of 125 heterogeneous compounds showed that the 3D approaches can be slightly improved by including in the predictive relationships property space descriptors that account for the dynamic profile of flexible molecules.[19] Similarly, a recent study indicated that it is possible to derive truly accurate log P predictions for homogenous sets of molecules by combining their log P values computed by a 3D method with molecular flexibility descriptors. Specifically, the study developed a targeted correlation for steroidal derivatives and used it to evaluate the lipophilicity of diastereoisomeric metabolites of methylprednisolone.[20] Altogether, these studies emphasized the opportunity of accounting for molecular variability to enhance the 3D predictive methods, and more importantly, they shed light on the possibility of exploiting such enhanced equations to take the configurational effects into account.

The relevance of this second application is underlined by the consideration that currently the log P differences between diastereoisomers cannot be conveniently predicted. Indeed, 2D methods do not account for configurational effects, whereas 3D approaches are not sufficiently accurate to successfully account for them. The only attempt hitherto published was based on semi-empirical calculations even though the accuracy of the results was comparable with that obtainable by the best 3D approaches.[21] Furthermore, the very limited availability of experimental log P values of diastereoisomers may have contributed to the lack of truly accurate prediction approaches. Yet, configurational factors play a crucial role in all biological phenomena, and accurate predictions of physicochemical properties of diastereoisomers can find useful applications to better understand their differences in biological activity, chemical reactivity, and separation properties.

On these grounds, the study presents the synthesis and physicochemical profiling of a set of diastereoisomeric dipeptides. The physicochemical properties so obtained were used to investigate configurational effects on ionization and lipophilicity as well as to develop correlative models able to predict the diastereoisomeric differences exploiting property space descriptors whose ability to account for configurational factors was preliminarily investigated in a recent study.[22] *D*-Amino acids have been found in several peptides and can be produced by the post-translational isomerization of natural *L*-residues. Such an isomerization plays a key role in protein aging because *D*-residues alter the hydrophobicity profiles with effects on protein conformation and deposition.[23] Specifically, the present study analyzes the effects of peptide isomerization by investigating a set of histidine-containing diastereoisomeric dipeptides that could find future applications as quenching agents of reactive carbonyl species.[24,25]

MATERIALS AND METHODS

The study involved the synthesis and physicochemical profiling of a set of 10 diastereoisomeric pairs of histidine-containing dipeptides (see Table 1). The dipeptides were prepared by maintaining the *N*-terminal residue in its natural *L* configuration and alternating the configuration of the histidine residue to obtain the desired diastereoisomeric pairs of general formula NH_2-*L*-X-(*L* or *D*)-His-OMe. All dipeptides were synthetized as methyl esters to facilitate their physicochemical profiling. Indeed, esterification (1) reduces polarity, (2) simplifies the ionization scheme, and (3) abolishes the zwitterionic behavior that would prevent experimental log P determination.

Synthesis

Preparation of H-*L*-Hys-OMe 2HCl and H-*D*-Hys-OMe 2HCl. Thionyl chloride (0.42 ml, 5.72 mmol) was added dropwise to a suspension of *L*-histidine or *D*-histidine monochloride monohydrate (1.0 g, 4.77 mmol) in trimethyl orthoformate (0.51 ml, 4.77 mmol) and anhydrous methanol (10 ml) at 0 °C under nitrogen atmosphere. The mixture was refluxed for 16 h and, after cooling to room temperature, concentrated to give the methyl ester dihydrochloride as a white solid in almost quantitative yield.

General procedures for the dipeptides synthesis. *L*-His-OMe 2HCl and *D*-His-OMe 2HCl were coupled with the *L* isomer of the following suitably protected amino acids: Z-Hys-OH, Z-Val-OH, Z-Glu(γBn)-OH, Z-Tyr(Bn)-OH, Z-Lys(Z)-OH, Z-Asn-OH, Z-Trp-OH, Z-Ser(tBu)-OH, Z-Met-OH, Z-Dmt-COOH. Successively, the fully protected dipeptides were deprotected to give the target dipeptides, methyl esterified at the C-terminal hystidine.

Coupling. A solution of 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide, hydrochloride (EDAC) (1.2 eq) and diisopropylethylamine (DIPEA) (1.2 eq) in dry dimethylformamide (DMF) was added dropwise to a solution of Z-AA-OH (1 eq), hydroxybenzotriazole (HOBT) (1.2 eq) in dry DMF under nitrogen atmosphere. After stirring for 30 min at room temperature, a solution of *L*-His-OMe 2HCl or *D*-His-OMe 2HCl (1 eq) and DIPEA (2 eq) in dry DMF was added dropwise. The reaction mixture was stirred at room temperature overnight and then concentrated under vacuum. The residue was treated with ethyl acetate and a solution of $NaHCO_3$. The organic phase was separated, washed with brine, dried (Na_2SO_4), and evaporated. The resultant crude full-protected dipeptide was purified by column chromatography. The purification conditions and the yields are listed in Table 2. Each purified dipeptide was analyzed by high-performance liquid chromatography (HPLC) by using Elite LaChrom HPLC system with diode array detector (190–400 nm) and a Water XBridgeTM C-18 column (5 μm, 4.6 × 150 mm) at 1 ml/min flow rate (0.5 ml/min for the dipeptides containing asparagine or glutamic acid) with the elution gradient stated in Table 3. Under such conditions, proved to be effective in separating all the diastereoisomeric pairs of the dipeptides epimers at the hystidine stereocentre, each purified dipeptide was found exempt from its diastereisomers.

Deprotection. The fully protected dipeptides, with the exception of those containing serine, cysteine, and methionine, were deprotected by hydrogenolisis in dry methanol/1.25 M methanolic hydrogen chloride (2 eq) in the presence of 10% Pd/C. The concentration of the filtered reaction mixture gave, in quantitative yield, the target dipeptides, methyl esterified at the C-terminal amino acid hystidine, as trihydrochloride, dihydrochloride, or monohydrochloride methyl. Each dipeptide was analyzed by HPLC on a Water XBridgeTM C-18 column at 1 ml/min flow rate (0.5 ml/min for the dipeptides containing asparagine or glutamic acid)

TABLE 1. Determined physicochemical properties for the 20 analyzed dipeptides plus the corresponding diastereoisomeric differences and some general mean values

Dipeptide	pK_1	pK_2	pK_3	ΔpK_{mean}	$\log P^N$	$\log D^{7.4}$	Rotors
Carnosine[a]	2.76	6.72	9.32	—	—[a]	—[a]	5
Homocarnosine[a]	2.75	6.79	9.88	—	—[a]	—[a]	6
L-Asn-*D*-His-OMe	7.28	6.06	—	—	0.65	0.41	8
L-Asn-*L*-His-OMe	7.04	5.98	—	—	0.20	−0.14	8
Δ(Asn-His-OMe)	0.24	0.08	—	0.16	0.45	0.55	—
L-Cys-*D*-His-OMe	6.98	5.94	9.19	—	2.46	0.90	7
L-Cys-*L*-His-OMe	6.96	5.94	9.10	—	1.78	0.47	7
Δ(Cys-His-OMe)	0.02	0.00	0.09	0.04	0.68	0.43	—
L-Glu-*D*-His-OMe	7.97	6.30	3.71	—	—[a]	—[a]	8
L-Glu-*L*-His-OMe	7.99	6.44	3.79	—	—[a]	—[a]	8
Δ(Glu-His-OMe)	0.02	0.14	0.08	0.08	—[a]	—[a]	—
L-His-*D*-His-OMe	7.40	6.35	4.92	—	−0.77	−1.08	8
L-His-*L*-His-OMe	7.48	6.42	5.15	—	0.10	−0.05	8
Δ(His-His-OMe)	0.08	0.07	0.23	0.12	0.87	1.03	—
L-Lys-*D*-His-OMe	7.65	6.03	10.30	—	−0.79	−1.01	10
L-Lys-*L*-His-OMe	7.63	6.66	10.78	—	−0.47	−0.86	10
Δ(Lys-His-OMe)	0.02	0.63	0.48	0.38	0.32	0.15	—
L-Met-*D*-His-OMe	7.52	6.19	—	—	2.96	2.69	8
L-Met-*L*-His-OMe	7.30	6.10	—	—	1.94	1.73	8
Δ(Met-His-OMe)	0.22	0.09	—	0.16	1.02	0.96	—
L-Ser-*D*-His-OMe	7.32	6.26	—	—	0.36	0.20	7
L-Ser-*L*-His-OMe	7.23	6.24	—	—	0.47	0.32	7
Δ(Ser-His-OMe)	0.09	0.02	—	0.06	0.11	0.12	—
L-Trp-*D*-His-OMe	7.57	6.34	—	—	0.71	0.44	8
L-Trp-*L*-His-OMe	7.54	6.43	—	—	1.17	0.89	8
Δ(Trp-His-OMe)	0.03	0.09	—	0.06	0.46	0.45	7
L-Tyr-*D*-His-OMe	7.57	6.41	9.86	—	0.02	−0.13	8
L-Tyr-*L*-His-OMe	7.52	6.41	9.69	—	0.15	0.01	8
Δ(Tyr-His-OMe)	0.05	0.00	0.17	0.07	0.13	0.14	—
L-Val-*D*-His-OMe	8.09	6.30	—	—	−0.13	−0.56	—
L-Val-*L*-His-OMe	7.72	6.30	—	—	0.44	0.24	7
Δ(Val-His-OMe)	0.37	0.00	—	0.18	0.57	0.80	7
Means	7.52	6.27	7.65	—	0.66	0.15	—
(*L*, *L*) Means	7.48	6.31	7.69	—	0.74	0.28	—
(*L*, *D*) Means	7.57	6.23	7.62	—	0.59	0.02	—
ΔMean	0.12	0.06	0.14	0.11	0.61	0.65	—
Standard deviation	0.40	0.35	2.88	—	1.06	0.93	0.86

The number of rotatable bonds is also reported as an index of dipeptide flexibility.
[a]Because of their zwitterionic nature, lipophilic data for carnosine, homocarnosine and glutamate-containing dipeptides cannot be accurately determined.

with the elution gradient stated in Table 4. When purity was found lower than 98% (ultraviolet detector 220 nm), as in the case of H-*L*-Tyr-*D*-Hys-OMe and H-*L*-Asn-*L*-His-OMe, the dipeptide was purified and isolated as di-trifluoroacetate by reverse-phase HPLC under the same conditions reported previously for the analyses but using a preparative Water XBridgeTM C-18 column of 19 mm internal diameter at 14 ml/min flow.

In the case of serine, Z-*L*-Ser(tBu)-OH was coupled with H-*L*-Hys-OMe 2HCl and H-*D*-Hys-OMe 2HCl according to the general coupling procedure. The side chain protection was removed stirring Z-*L*-Ser(tBu)-*L*-His-OMe and Z-*L*-Ser(tBu)-*D*-His-OMe in 7:3 DCM/trifluoroacetic acid (TFA) for 3 h in the presence of anisole (2 eq). The solvent was removed and the residue treated with water and diethyl ether. The aqueous phase was separated and washed with diethyl ether twice. Successive removal of water under vacuum gave Z-*L*-Ser-*L*-His-OMe and Z-*L*-Ser-*D*-His-OMe in quantitative yield. The crude products underwent hydrogenolisis as described in the general procedure giving H-*L*-Ser-*L*-His-OMe and H-*L*-Ser-*D*-His-OMe as dihydrochlorides.

The synthesis of H-*L*-Cys-*L*-H-OMe and H-*L*-Cys-*D*-His-OMe started from *L*-Dmt-COOH that was quantitatively converted into Z-*L*-Dmt-COOH by treatment with 1.2 eq of benzyl chloroformate and 1 eq of DIPEA in acetonitrile under nitrogen for 3 h and, after isolation, coupled with H-*L*-His-OMe2HCl and H-*D*-His-OMe 2HCl following the general method. The coupling products, namely Z-*L*-Dmt-*L*-His-OMe and Z-*L*-Dmt-*D*-His-OMe, were stirred in TFA for 24 h in the presence of anisole (2 eq). The solvent was removed, and the residue was dissolved in water, washed with diethyl ether three times. Finally, water removal under vacuum gave H-*L*-Cys-*L*-His-OMe 2TFA and H-*L*-Cys-*D*-H-OMe 2TFA in quantitative yield.

H-*L*-Met-*L*-His-OMe and H-*L*-Met-*D*-His-OMe were obtained from Z-*L*-Met-*L*-His-OMe and Z-*L*-Met-*D*-His-OMe, respectively, by stirring in TFA for 24 h in the presence of 2 eq of anisole. The solvent was removed and the residue dissolved in water and washed three times with diethyl ether. Water removal under vacuum gave H-*L*-Met-*L*-His-OMe 2TFA and H-*L*-Met-*D*-His-OMe 2TFA in quantitative yield.

Physicochemical Profiling

The ionization constants and lipophilicity parameters of the dipeptides were determined at 25 °C by potentiometric titration with the GlpKa apparatus (Sirius Analytical Instruments Ltd., Forest, Row, East Sussex, UK). All experiments were carried out under a slow nitrogen flow to avoid CO_2 absorption. A weighted sample (5–10 mg) was supplied manually, whereas the diluent and all the other reagents were added automatically. In detail, the compounds were solubilized in 0.15 M KCl (to adjust ionic strength) and acidified with 0.1 M HCl to pH 1.8. The solutions were then

TABLE 2. Yields and elution conditions of full-protected dipeptides

Dipeptide	Yield (%)	Eluent
Z-*L*-Hys-*L*-Hys-OMe	47	AcOEt/MeOH (95:5)
Z-*L*-Hys-*D*-Hys-OMe	50	AcOEt/MeOH (95:5)
Z-*L*-Val-*L*-Hys-OMe	60	DCM/MeOH (90:10)
Z-*L*-Val-*D*-Hys-OMe	46	DCM/MeOH (90:10)
Z-*L*-Glu(γBn)-*L*-Hys-OMe	49	AcOEt/MeOH (95:5)
Z-*L*-Gly(γBn)-*D*-Hys-OMe	66	AcOEt/MeOH (95:5)
Z-*L*-Tyr(Bn)-*L*-Hys-OMe	51	DCM/MeOH (95:5)
Z-*L*-Tyr(Bn)-*D*-Hys-OMe	49	DCM/MeOH (95:5)
Z-*L*-LysZ-*L*-Hys-OMe	55	DCM/MeOH (90:10)
Z-*L*-LysZ-*D*-Hys-OMe	50	DCM/MeOH (90:10)
Z-*L*-Asn-*L*-Hys-OMe	36	DCM/MeOH (96:4)
Z-*L*-Asn-*D*-Hys-OMe	38	DCM/MeOH (96:4)
Z-*L*-Trp-*L*-Hys-OMe	92	AcOEt/MeOH (95:5)
Z-*L*-Trp-*D*-Hys-OMe	86	AcOEt/MeOH (95:5)
Z-*L*-Ser(tBu)-*L*-Hys-OMe	73	AcOEt/MeOH (95:5)
Z-*L*-Ser(tBu)-*D*-Hys-OMe	80	AcOEt/MeOH (95:5)
Z-*L*-Met-*L*-Hys-OMe	50	DCM/MeOH (90:10)
Z-*L*-Met-*D*-Hys-OMe	63	DCM/MeOH (90:10)
Z-*L*-Dmt-*L*-Hys-OMe	51	AcOEt
Z-*L*-Dmt-*D*-Hys-OMe	58	AcOEt

TABLE 3. Elution gradient for the analysis of full-protected dipeptides by high-performance liquid chromatography

Time	Water (%)	ACN (%)	1% aqueous TFA (%)
t_o	90	0	10
$t_3 = 25'$	75	15	10
$t_4 = 35'$	10	80	10
$t_5 = 40'$	90	0	10

ACN, acetonitrile; TFA, trifluoroacetic acid.

TABLE 4. Elution gradient for the analysis and the preparative purification of deprotected dipeptides by high-performance liquid chromatography

Time	Water (%)	ACN (%)	1% aqueous TFA (%)
t_o	90	0	10
$t_1 = 5'$	89	1	10
$t_2 = 10'$	85	5	10
$t_3 = 25'$	75	15	10
$t_4 = 35'$	10	80	10
$t_5 = 40'$	90	0	10

ACN, acetonitrile; TFA, trifluoroacetic acid.

titrated with standardized 0.1 M KOH up to a pH of 12.2. Bjerrum difference plots were deduced from each titration and used to calculate precise pK values.

To obtain lipophilicity parameters, separate titrations of the compounds were carried out using various volumes of *n*-octanol. In the presence of *n*-octanol, the pK value shifts as a consequence of the partitioning of the substance into the organic phase, allowing a new p_oK constant to be determined. These shifts in the pK values were used to determine $\log P^N$, the logarithm of the partition coefficient of the neutral form. Apparent $\log D^{7.4}$ values (i.e., distribution coefficients at pH 7.4) were then extrapolated from the titration analyses. All measurements were performed in triplicate. The detailed experimental procedures and data analyses as well as the recommended apparatus standardization can be found elsewhere.[26]

Computational Details

All investigated dipeptides were simulated in their electrical state favored at pH 7.4, as derived from their experimental ionization constants. After a preliminary optimization by PM6 semi-empirical calculations to calculate precise atomic charges, the conformational profiles were investigated by a clustered Monte Carlo analysis as implemented in the VEGA[27] suite of programs to produce 1000 minimized conformers. The lowest energy structure was then used in Molecular Dynamics (MD) simulations in a vacuum and in two solvents of very different polarity, that is, water and chloroform. For the simulations in chloroform and water, the peptide was placed in a cluster of 12 Å radius containing 192 water molecules described by the model TIP3S or 86 chloroform molecules as described in VEGA templates.[27] Before performing the MD simulations, the complexes were optimized for the relative position of the solvent molecules to eliminate any high-energy interaction.

All MD simulations lasted for 10 ns and had the following characteristics: (a) Newton's equation was integrated every femtosecond according to Verlet's algorithm; (b) the temperature was maintained at 300 ± 10 K by means of the Langevin's algorithm; (c) spherical boundary conditions (radius = 15 Å) were applied to stabilize the solvent clusters; (d) Lennard–Jones (L-J) interactions were calculated with a cutoff of 10 Å, and the pair list was updated every 20 iterations; and (e) a frame was stored every 10 ps, to yield 1000 frames. The simulations were carried out in two phases: an initial period of heating from 0 K to 300 K over 30,000 iterations (30 ps, that is, 10 K/ps) and the monitored phase of 10 ns. All MD runs were carried out using Namd2.7[28] with the force-field CHARMm v22 and Gasteiger's atomic charges on a 16-core Tyan VX-50 system.

RESULTS

Chemistry

The dipeptides, prepared as described in Section 2.1, were characterized by ^{1}H NMR spectroscopy at 300 MHz (see Supporting Information) and analyzed by reverse-phase HPLC under conditions effective in diastereomer separation proving their >98% purity and, what is more, the absence of any epimer, a mandatory requisite for the present investigation.

Physicochemical Profiles

The dipeptides included in this study were chosen to cover exhaustively the available chemical space (see Table 1) even though the zwitterionic character of the glutamate-containing dipeptides impeded their accurate lipophilicity evaluation. The attention was focused on the accurate measurement of $\log P^N$ only because the log P values for most ionized states were somewhat out of the range of the apparatus (i.e., $\log P^I << -1$). The physicochemical results are compiled in Table 1 that includes ionization constants (namely pK_1 for the N-terminus, pK_2 for the histidine imidazole ring, and eventually, pK_3 when the second side chain possessed an ionizable group), lipophilic data (i.e., $\log P^N$ and $\log D^{7.4}$) plus the corresponding diastereoisomeric differences and some overall averages. Physicochemical data for some well-known natural histidine-containing dipeptides (i.e., carnosine, β-Ala-His and homocarnosine, γ-aminobutyryl-His) were also included for easy comparison (taken from previous studies[25,29]).

pKa values. Comparing the pK_1 values of the dipeptides with those of the corresponding free amino acids confirms that the amino group is markedly less basic when belonging to a peptide.[30] Indeed, although the pK_1 values in free amino acids are almost always greater than 9, they rarely exceed 8 in the analyzed dipeptides because of the electron-withdrawing effect of the peptide bond. This electronic effect is particularly evident when considering the pK_1 values of the included natural histidine-containing dipeptides. Indeed, the shifting

of the amino group from the α to the β position (as seen in carnosine) induces a pK_1 increase of more than one logarithmic unit, whereas further shifting to the γ position (as seen in homocarnosine) has a more modest effect on pK_1, even though it should be noted that the electron-withdrawing effect of the peptide bond is still significant also in a γ position as evidenced by a comparison with the clearly greater basicity of alkylamines (e.g., butylamine pK = 10.77[31]). Moreover, the obtained pK_1 values are in line with those reported for unprotected dipeptides,[32] thus indicating that the esterification of the C-terminus does not induce pronounced effects on the pK_1 values, but it may influence differences between diastereoisomers, as discussed next.

The pK_2 values of the imidazole ring remain in a narrow range somewhat above that of free histidine ($pK_2 = 5.97$) and are rather similar to those of the reported endogenous dipeptides, thus being scarcely affected by the vicinal residue. The third ionization constant, when present, shows marked differences compared with the corresponding values for the free residues probably because of a mutual influence between adjacent ionizable side chains. For example, the low basicity of the second imidazole ring in the NH_2-*L*-His-(*L* or *D*)-His-OMe dipeptides indicates that the influence between contacting basic moieties destabilizes the electrical state with both protonated imidazole rings. Also, NH_2-*L*-Cys-(*L* or *D*)-His-OMe dipeptides show higher pK_3 values compared with that of the sulfhydryl group in free cysteine (pK = 8.33), whereas lysine-containing and tyrosine-containing dipeptides show pK_3 values similar to those of the corresponding free amino acids.

Before examining the differences between diastereoisomers (ΔpK_1, ΔpK_2, and ΔpK_3), it is worth remembering that our preliminary study of the lowest energy geometries for the four stereoisomers of the Ac-NH-His-Ser-OMe dipeptide revealed that the lowest energy conformation of the homochiral isomers is mainly characterized by intramolecular interactions between side chains, whereas that of the heterochiral isomers shows stabilizing contacts between terminal groups.[22] As a trend, this key difference is also confirmed here by our MD simulations (see next section) and suggests that the side chains are more accessible in the heterochiral stereoisomers. Consequently, the slightly greater basicity of the N-terminus of the (*L,D*)-isomers compared with that of the (*L,L*)-isomers can be rationalized in terms of polar interactions between terminal groups that tend to stabilize the ammonium head by a probable H-bond with the ester moiety. Similarly, the greater imidazole basicity in the homochiral isomers can be explained by stabilizing interactions between side chains, whereas the heterogeneous nature of third ionizable group does not allow a general behavior to be unveiled. The averages of diastereoisomeric differences are in line with the variability previously described. Indeed, the imidazole basicity shows, on average, the lowest differences, thus indicating that it is minimally influenced by both configurational effects and nearby moieties, whereas pK_3 values show the greatest diastereoisomeric differences because the discussed interactions between adjacent residues are clearly influenced by configuration.

When considering the mean difference in the pK values of each diastereoisomeric pair (ΔpK_{mean}), one may recognize three different behaviors that depend on conformational flexibility and intramolecular interactions. They can be schematized as follows:

1. When the dipeptide is characterized by significant intramolecular interactions (mainly, but not exclusively, between side chains), it shows low diastereoisomeric differences ($\Delta pK_{mean} < 0.1$) because the configurational effects are restrained by conformational rigidity.
2. When the dipeptide does not show relevant intramolecular interactions (apart from some weak hydrophobic contacts), it exhibits intermediate diastereoisomeric differences ($0.1 < \Delta pK_{mean} < 0.2$), thus indicating that the configurational effects are paralleled by a good conformational flexibility.
3. When the dipeptide conformation is affected by relevant intramolecular repulsions (as in the case of NH_2-*L*-Lys-(*L* or *D*)-His-OMe dipeptides between the ammonium heads), it exhibits large diastereoisomeric differences ($\Delta pK_{mean} > 0.2$), thus indicating that the configurational effects are amplified by an extreme conformational mobility.

The effect of intramolecular interactions in the diastereoisomeric differences is further corroborated by the analysis of a nonesterified dipeptide pair (i.e., NH_2-*L*-Val-(*L* or *D*)-His-OH) included in the study for an in-depth comparison. The ionization constants [(*L,L*): $pK_1 = 7.83$, $pK_2 = 6.67$, $pK_3 = 2.73$; (*L,D*): $pK_1 = 7.89$, $pK_2 = 6.64$, $pK_3 = 2.57$; $\Delta pK_1 = 0.06$ $\Delta pK_2 = 0.03$; $\Delta pK_3 = 0.16$; $\Delta pK_{mean} = 0.08$] reveal markedly reduced diastereoisomeric differences compared with those of the corresponding esterified dipeptide (as exemplified by ΔpK_{mean}, 0.08 vs. 0.18), allegedly because of the constraining intramolecular ion pair between charged termini. Again, the pK values of nonesterified dipeptides are quite similar to those of the corresponding esterified derivatives, and furthermore, the pK_1 values are almost superimposable to those of the reported natural dipeptides. This confirms the scarce influence between the ionizable groups, and accordingly, esterification does not markedly impact on ionization constants apart from the imidazole ring that appears slightly more basic in the nonesterified peptides presumably because of a stabilizing interaction with the carboxyl terminus.

Lipophilicity. Comparing the present lipophilicity data (Table 1) with literature values confirms the remarkable polarity of these dipeptides, and, despite the impossibility of a direct comparison, suggests that hydrophilicity is only slightly reduced by methyl esterification. Specifically, most $\log P^N$ values are found in the range $0 < \log P^N < 1$, and only the Cys-containing and Met-containing dipeptides show markedly higher $\log P^N$ values. The extrapolated $\log D^{7.4}$ values are well correlated with the $\log P^N$ values ($r^2 = 0.93$) and were found to be slightly lower than the latter. This may indicate that the neutral forms play a relevant role at pH 7.4, a result explainable by the weak basicity of the amino groups.

The diastereoisomeric differences in lipophilicity values ($\Delta\log P^N$ and $\Delta\log D^{7.4}$) are well understandable considering the discussed conformational properties. Table 1 reveals that homochiral isomers are, on average, more lipophilic than the heterochiral isomers as confirmed by both $\Delta\log P^N$ and $\Delta\log D^{7.4}$ values. This can be explained by the greater side chain accessibility of the heterochiral isomers that results in more polar derivatives and suggests that the hydrophilicity increase due to a greater imidazole exposition is rarely counterbalanced by the second side chain. Only the significant

apolarity of the Cys and Met side chains can overcome the polarity of the imidazole ring, and indeed, sulfur-containing dipeptides show an opposite trend, their (*L,D*)-isomers being more lipophilic than the (*L,L*)-isomers.

Conversely, the accessibility of the dipeptide termini shows a reduced influence as demonstrated by the higher lipophilicity of the (*L,L*)-isomers despite the greater accessibility of both termini. This may imply that the hydrophilic contributions of the termini, and in particular that of the ionizable N-termini, are so crucial to be almost independent on conformational and configurational effects. Collectively, lipophilicity data show significant diastereoisomeric differences that appear more pronounced than those observed for the ionization constants and also more homogeneously distributed. However, one can also recognize here a constraining effect of the intramolecular interactions as seen, for example, when comparing $\Delta\log P^N$ for NH_2-*L*-Met-(*L* or *D*)-His-OMe with that of NH_2-*L*-Ser-(*L* or *D*)-His-OMe (1.02 vs. 0.11).

Altogether, the lipophilicity data confirm that peptide isomerization can have a remarkable impact on lipophilicity profile, essentially because of a greater exposition of the side chains in the heterochiral peptides. Clearly, the resulting effect depends on the hydrophilicity of the side chains. In our histidine-containing dipeptides, isomerization almost always induced a log P decrease ascribable to the polarity of the imidazole moiety, whereas the heterochiral diastereoisomers of apolar peptides are expected to be more lipophilic (and thus more prone to amyloid deposition) as suggested by Munegumi[23] and confirmed here for the sulfur-containing dipeptides.

Property Space Analysis

The property space of the dipeptides was investigated by calculating the following properties for each conformer stored during the MD simulations in vacuo, in water, and in chloroform:

1. Its radius of gyration, a well-known descriptor encoding molecular shape and size,[33] and whose variations can be used to estimate molecular flexibility;[16]
2. Its conformer-dependent $\log P^{MLP}$ value (also known as virtual log P, as computed here by the Molecular Lipophilicity Potential (MLP) approach;[34] and
3. Its polar surface area (PSA), which parameterizes their H-bonding capacity.[35]

For each property, the resulting space is described by the mean plus the range (namely, the difference between maximum and minimum values) of the computed values, as well as the sensitivity computed as the ratio between property range and number of rotatable bonds.[17]

Tables S1, S2, and S3 (Supplementary Information) compile the property space descriptors of the aforementioned conformer-dependent properties for the 20 simulated dipeptides as derived from MD simulations in vacuo (Table S1), in water (Table S2), and in chloroform (Tables S3). Tables S1, S2, and S3 also report the diastereoisomeric differences for the analyzed properties plus some overall means. A bird's eye view of the compiled values shows a notable degree of similarity within each property, with very few dipeptides deviating significantly from the mean. Such a result can be explained considering that all computed properties are markedly influenced by the common moieties among these dipeptides (i.e., their backbone and imidazole ring). As suggested by the overall standard deviations, lipophilicity is the most variable property presumably because its values are strongly influenced by the polarity of the second side chain. Radius of gyration and PSA values are more homogeneously distributed with few dipeptides showing significantly different values, as in the case of the bulky tyrosine-containing and tryptophan-containing dipeptides (for the radius of gyration), and for the very apolar NH_2-*L*-Val-(*L* or *D*)-His-OMe peptides (for PSA). By comparing the simulated media, the most nonhomogeneous values are observed in chloroform probably because its incapacity to form strong polar interactions with the solute amplifies the structural diversity between the considered dipeptides.

Tables S4, S5, and S6 collect the interrelation matrices for the nine computed property space descriptors (as seen in Tables S1, S2, and S3) as derived from MD simulations in vacuo (Table S4), in water (Table S5), and in chloroform (Tables S6). In all media, the descriptors of the radius of gyration are significantly cross-related, indicating that all of them similarly encode molecular size and flexibility. Moreover, for all properties and in all simulated media, there are marked relations between property ranges and sensitivity values because of the almost constant number of rotatable bonds that characterize the modeled dipeptides (as compiled in Table 1). Apart from the mentioned high correlations, significant relationships ($r > 0.6$) are seen between log P range and radius of gyration, thus confirming the ability of this latter to account for the structural variability.

Similarly, Tables S7, S8, and S9 collect the correlation matrices for the diastereoisomeric differences in the property space descriptors as derived from MD simulations in vacuo (Table S7), in water (Table S8), and in chloroform (Tables S9). The resulting relations are reasonably in line with those reported for property values and reveal notable relations between the diastereoisomeric differences in log P and PSA values, thus suggesting that both depend on the different accessibility of the polar side chains.

Figures 1–3 compare the overall averages and ranges as obtained in the simulated media for the radius of gyration (Fig. 1), lipophilicity (Fig. 2), and the PSA (Fig. 3). With regard to the radius of gyration, Figure 1 shows that all simulated dipeptides assumed more folded conformations in vacuo because, essentially, of intramolecular polar interactions. In contrast, the monitored conformations appear largely more extended in water and chloroform probably because of strong solute–solvent interactions. Conceivably, the effect is more pronounced in water where the extended geometries are stabilized by strong polar contacts, whereas the effect in chloroform is mainly due to a greater molecular friction experienced by extended geometries. Conformational variability, as encoded by the ranges of values, is greater in solvents than in a vacuum, a result that depends on the abundance of extended geometries in the investigated solvents. The diastereoisomeric differences are quite marginal and without well-defined trends for both means and ranges, thus suggesting that the conformation profiles of all isomers are similarly affected by the simulated media.

The $\log P^{MLP}$ means (Fig. 2) are in line with the polarity of the simulated environments. Indeed, a vacuum represents the most apolar medium, water as the most polar

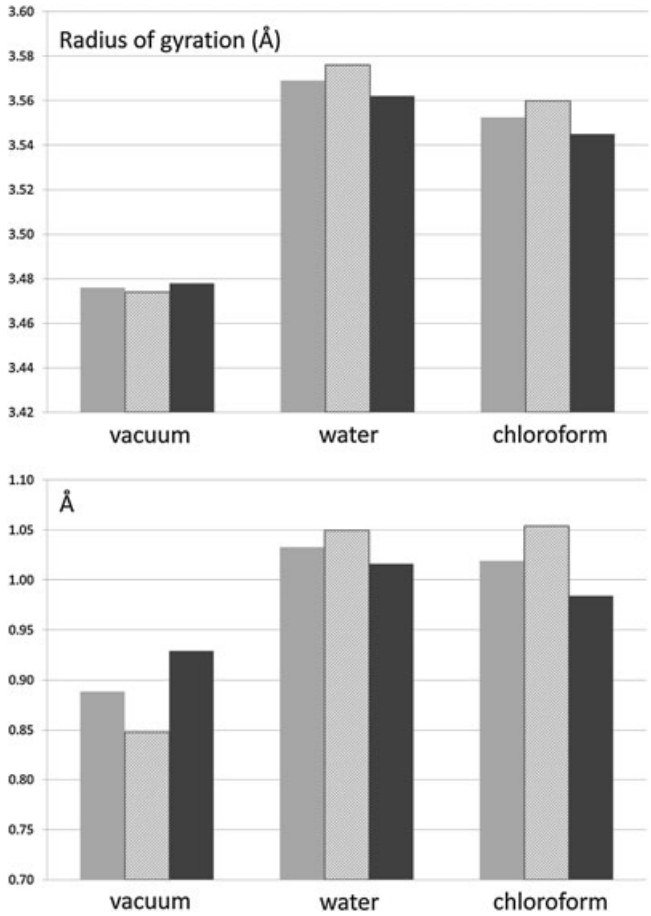

Fig. 1. Comparison of the overall means (upper plot) and ranges (lower plot) for the radius of gyration as calculated by molecular dynamics simulations in vacuo, in water, and in chloroform and computed for all dipeptides (light gray bars), homochiral peptides (dashed bars), and heterochiral peptides (dark gray bars).

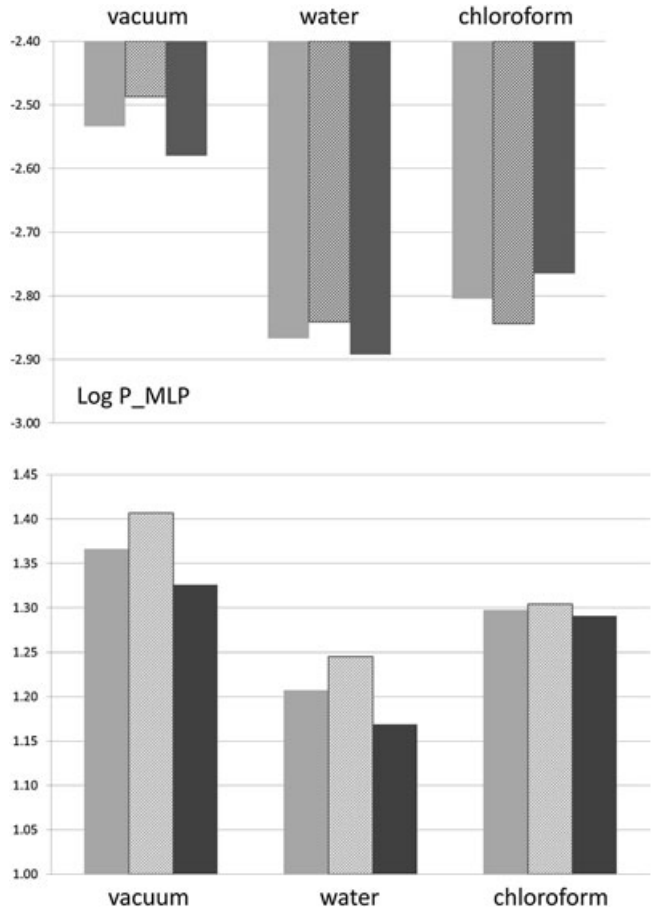

Fig. 2. Comparison of the overall means (upper plot) and ranges (lower plot) for the $\log P^{MLP}$ values as calculated by molecular dynamics simulations in vacuo, in water, and in chloroform and computed for all dipeptides (light gray bars), homochiral peptides (dashed bars), and heterochiral peptides (dark gray bars).

one, whereas chloroform properties are intermediate. Thus, Fig. 2 confirms what was already observed for acetylcholine,[36–38] namely, that molecules tend to adapt their physicochemical properties to those of the environment without restraining their conformational spaces (as evidenced by the ranges of radius of gyration). The global $\log P^{MLP}$ means are in encouraging agreement with the experimental data because the heterochiral dipeptides appear more hydrophilic than the homochiral ones in a vacuum and in water because of a greater accessibility of their side chains. The $\log P^{MLP}$ means show an opposite trend in chloroform probably because the exposed side chains tend to collapse in an apolar solvent.

Unlike the trends shown by the radius of gyration, the overall $\log P^{MLP}$ ranges clearly decrease when the dipeptides are simulated in a solvent. This implies that the property adaptability is achieved by narrowing the property space while expanding the corresponding conformational space because the medium is able to select in each conformational cluster those conformers whose polarity most resembles its own. This confirms that conformational space and property spaces are only partly related and that each cluster of conformers spans most of the property space. Albeit with minimal differences, the computed ranges suggest that the $\log P^{MLP}$ variability is greater in the homochiral peptides presumably because of a variable exposition of their termini. The range differences vanish in chloroform probably because here, the polar termini tend to minimize their exposure regardless of configuration.

The overall averages and ranges for PSA (Fig. 3) are in line with those of $\log P^{MLP}$, although the reported differences are far less pronounced allegedly because of the discussed homogeneity in PSA values (Tables S1–S3). Specifically, Fig. 3 confirms the adaptability of the dipeptides simulated in solvents of different polarity and the constraining effects that the solvents exert on property variability as seen in PSA ranges. Similar to what was observed for the radius of gyration, the diastereoisomeric differences are marginal and without defined trends.

Collectively, these results represent a compelling validation for the concept of property space, confirming what was shown in our first studies with acetylcholine.[35–38] In particular, the present study confirms that property space descriptors can properly define the ability of a given molecule to adapt itself to its environment; furthermore, it shows that such an adaptability occurs through a dynamic process where conformational and property spaces can be influenced independently. The reliability of the property space concept is corroborated here by the number of molecules investigated as well as by its ability to account for configurational effects. Lastly, the investigated property spaces suggest that genuine physicochemical properties (namely, $\log P^{MLP}$) are more sensible

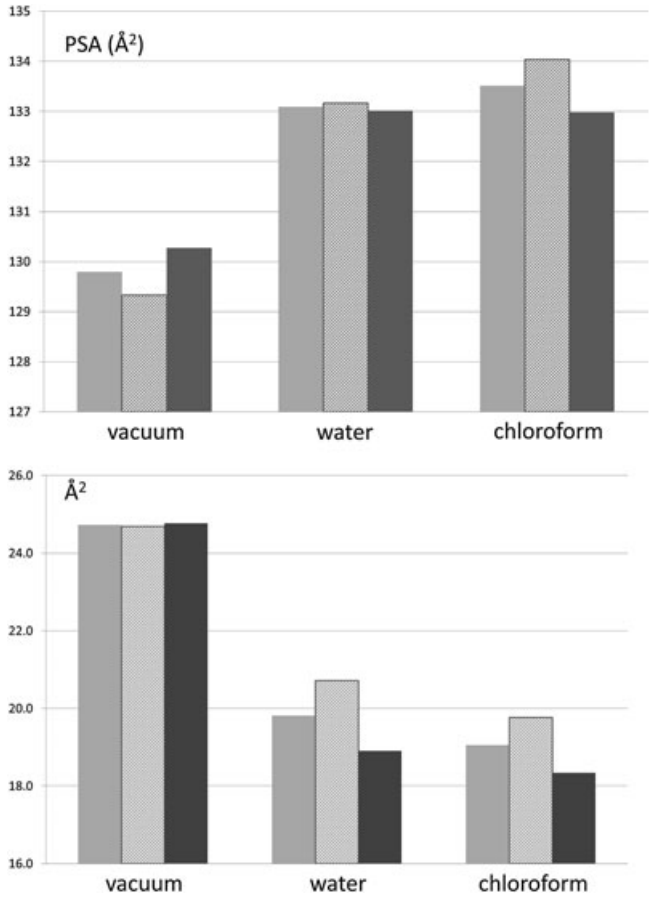

Fig. 3. Comparison of the overall means (upper plot) and ranges (lower plot) for the polar surface area (PSA) as calculated by molecular dynamics simulations in vacuo, in water, and in chloroform and computed for all dipeptides (light gray bars), homochiral peptides (dashed bars), and heterochiral peptides (dark gray bars).

to both environmental and configurational effects compared with geometrical (radius of gyration) or mixed (PSA) properties.

Predicting Configurational Effects

Table 5 reports the correlation matrices between the measured properties (i.e., pKs, $\log P^N$, $\log D^{7.4}$, Table 5A) as well as between the corresponding diastereoisomeric differences (Table 5B). Both matrices also include the number of rotatable bonds as a parameter describing molecular flexibility. Besides the aforementioned correlation between $\log P^N$ and $\log D^{7.4}$ values, which concerns also the corresponding diastereoisomeric differences, Table 5A evidences the remarkable correlation between the pK_1 and pK_2 constants, thus indicating that the imidazole ring and N-terminus finely influence each other, despite the quite homogeneous distribution of their values. Although the relations involving pK_3 and ΔpK_3 may be somewhat overestimated because of the small number of data included (only 10 dipeptides have a third ionizable group), Table 5B shows notable relations between the number of rotatable bonds and the diastereoisomeric differences of all ionization constants. This emphasizes the significant role of flexibility in determining physicochemical differences between diastereoisomers and invites to exploit property space descriptors to predict them. Lastly, the fair correlation between ΔpK_1 and $\Delta \log D^{7.4}$ confirms the role of the N-terminus ionization in determining $\Delta \log D^{7.4}$ variations.

TABLE 5. Correlation matrices between the (A) measured properties and the (B) corresponding diastereoisomeric differences

A

Property	Rotors	pK_1	pK_2	pK_3	$\log P^N$
$\log D^{7.4}$	−0.28	0.03	0.17	−0.32	0.93
$\log P^N$	−0.36	−0.11	0.05	−0.15	
pK_3	0.32	−0.52	−0.17		
pK_2	0.26	0.80			
pK_1	0.27				

B

Property	Rotors	ΔpK_1	ΔpK_2	ΔpK_3	ΔpK_{mean}	$\Delta \log P^N$
$\Delta \log D^{7.4}$	−0.34	0.54	−0.29	−0.25	−0.02	0.90
$\Delta \log P^N$	−0.24	0.29	−0.14	−0.28	0.02	
ΔpK_{mean}	0.72	0.19	0.86	0.97		
ΔpK_3	0.80	0.01	0.89			
ΔpK_2	0.81	−0.31				
ΔpK_1	−0.15					

The number of rotatable bonds (rotors) is also included in both matrices.

Tables S10–S12 report the correlation matrices between the experimental physicochemical data and the property space parameters as computed in a vacuum (Table S10A), in water (Table S11A), and in chloroform (Table S12A). The correlation matrices between the corresponding diastereoisomeric differences are also included (Table S10B–S12B). A comprehensive analysis of these matrices reveals that all relationships involving experimental physicochemical data (Table S10A–S12A) are constantly poor regardless of the simulated medium (r always <0.6). In particular, $\log P^{MLP}$ means appear to be unable to successfully predict the experimental $\log P^N$ values because the correlations between them are very low ($r \cong 0.35$). On the one hand, this result confirms the intrinsic difficulty to predict peptide lipophilicity as seen also with the best performing fragmental approaches.[39] On the other hand, this result underlines the need to develop innovative approaches to conveniently predict the lipophilicity for peptide molecules whose relevance is continuously increasing in medicinal chemistry. The correlations between computed $\log P^{MLP}$ means and experimental $\log D^{7.4}$ values are slightly better than the previous ones (0.41 vs. 0.36) probably because the MD simulations did not involve the neutral peptides but the favored ionized forms at physiological pH.

Apart from the good but possibly overvalued correlations involving ΔpK_3, Tables 10B–12B evidence other relevant correlations ($r > 0.6$), which appear more frequent in vacuo than in the simulated solvents. Specifically, significant correlations are seen between the radius of gyration and the diastereoisomeric differences in ionization constants as well as between the latter and diastereoisomeric differences in $\log P^{MLP}$ and PSA ranges. As discussed previously, when analyzing the experimental data (see Section 3.2), these results confirm that the ionization differences between diastereoisomers can be adequately rationalized in terms of flexibility (as encoded by the radius of gyration) and tethering intramolecular interactions (as encoded by PSA).

Although the computed $\log P^{MLP}$ values proved unsuccessful in predicting experimental $\log P^N$ data, the correlations

between the differences in $\log P^N$ (or $\log D^{7.4}$) and in virtual $\log P^{MLP}$ values appear rather significant when computed by in vacuo MD runs (Table S10B), thus emphasizing that $\log P^{MLP}$ values can account for the effects of a greater exposition of the side chains in the heterochiral peptides. This underlines the key role played by conformational factors in determining the differences between diastereoisomers, demonstrating the value of conformer-dependent approaches to account for such differences despite their limitations.

With all data in hand, the last part of our study endeavored to exploit the computed property space descriptors to predict the experimentally determined physicochemical properties accounting for configurational effects. Measured $\log P^N$ and $\log D^{7.4}$ values were used as dependent variables to develop correlative equations as iteratively generated by maximizing the r^2 value and including at most three independent variables.

The study involving all investigated dipeptides did not afford noteworthy relationships ($r^2 < 0.5$), a failure seemingly due mostly to an inaccurate calculation of the methionine-containing dipeptides that were constantly predicted to be markedly more hydrophilic than determined experimentally. This significant source of error can be due to an unsuitable parameterization of the significant lipophilic contribution of sulfur-containing functional groups. Indeed, besides the methionine-containing derivatives, the NH_2-*L*-Cys-(*L* or *D*)-His-OMe dipeptides were also poorly predicted even though here, the inaccuracy was markedly smoothened by the weaker apolarity of the thiol group. This emphasizes the relevance of a better parameterization of the lipophilic increments for all sulfur-containing functional groups. Consequently, the predictive studies discussed next were carried out initially discarding the two NH_2-*L*-Met-(*L* or *D*)-His-OMe dipeptides.

The best so developed correlative models are reported in Table 6. They reveal very remarkable statistics that are comparable with, if not better than, those of best performing 2D log P approaches when applied to peptides. Equations 1 and 2 similarly include the same three independent variables, a result well understandable considering the very high correlations between $\log P^N$ and $\log D^{7.4}$ values. Specifically, these relationships include $\log P^{MLP}$ means, $\log P^{MLP}$ ranges, and PSA ranges as computed from MD simulations in a vacuum (Table S1). They can be simultaneously used in a correlative equation as they are not significantly cross-correlated as reported by Table S4. On the one hand, the included independent variables suggest that the computed $\log P^{MLP}$ values can be useful for $\log P^N$ predictions despite their limitations. On the other hand, they further confirm the relevance of property space descriptors (as encoded by $\log P^{MLP}$ and PSA ranges) to account for conformational and configurational effects. Notably, both ranges appear in the equations with positive coefficients, thus suggesting that lipophilicity increases with extended property variability.

To confirm the outlier character of the methionine-containing dipeptides (focusing the attention on $\log D^{7.4}$ prediction), a third equation was generated by adding a Boolean variable (I_{SMe}) that is equal to 1 for NH_2-*L*-Met-(*L* or *D*)-His-OMe and 0 for the other peptides. As reported in Table 3, eq. 3 shows quite impressive statistics that can counterbalance the inclusion of a fourth independent variable as documented by the increased *F* value (51.5 vs. 44.9), thus providing a model able to account for all measured $\log D^{7.4}$ values. Moreover, eq. 3 indicates that the lipophilic

TABLE 6. Main predictive equations developed in the study

N	Equation	Statistics
1	$\log D^{7.4} = -3.04 + 0.35\ \{\log P^{MLP}\_mean\} + 1.54\ \{\log P^{MLP}\_range\} + 0.094\ \{PSA\_range\}$	$n = 16$; $r^2 = 0.92$; $r^2_{cv} = 0.90$ $SE = 0.20$; $F = 44.9$; $P < 0.0001$
2	$\log P^N = -3.78 + 0.32\ \{\log P^{MLP}\_mean\} + 1.30\ \{\log P^{MLP}\_range\} + 0.16\ \{PSA\_range\}$	$n = 16$; $r^2 = 0.88$; $r^2_{cv} = 0.85$ $SE = 0.33$; $F = 29.8$; $P < 0.0001$
3	$\log D^{7.4} = -3.03 + 0.34\ \{\log P^{MLP}\_mean\} + 1.47\ \{\log P^{MLP}\_range\} + 0.096\ \{PSA\_range\} + 3.02\ I_{SMe}$	$n = 18$; $r^2 = 0.94$; $r^2_{cv} = 0.92$ $SE = 0.26$; $F = 51.5$; $P < 0.0001$
4	$\log D^{7.4} = -3.44 + 0.37\ \{\log P^{MLP}\_mean\} + 1.42\ \{\log P^{MLP}\_range\} + 1.13\ \{PSA\_range\} + 3.00\ I_{SMe}$	$n = 12$; $r^2 = 0.96$; $r^2_{cv} = 0.94$ $SE = 0.30$; $F = 41.6$; $P < 0.0001$
5	Predicted $\log D^{7.4} = -0.18 + 0.89\ \exp \log D^{7.4}$	$n = 6$; $r^2 = 0.91$; $r^2_{cv} = 0.89$ $SE = 0.16$; $F = 51.1$; $P = 0.002$
6	$pK_1 = 9.16 + 0.14\ \{Rotors\} - 0.0020\ \{PSA\_mean\} - 0.0012\ \{PSA\_range\}$	$n = 20$; $r^2 = 0.83$; $r^2_{cv} = 0.79$ $SE = 0.13$; $F = 32.4$; $P < 0.0001$

MLP, molecular lipophilicity potential; SE, standard error; PSA, polar surface area.
All included property space descriptors are derived from MD runs in vacuo (Table S1).

contribution of a thioether function is globally underestimated by a value of 3.02, a result that further confirms the need of a better reparameterization of sulfur-containing functional groups.

To further assess the predictive power of eq. 3 and despite the low number of included data, the examined dipeptides were randomly subdivided into a training set ($n = 12$) and an external test set ($n = 6$). Equation 4 was developed considering only the training set and appears very similar to the previous one, thus confirming a substantial stability of the predictive model that is almost independent on the included data. The relation between experimental and predicted log $D^{7.4}$ values (eq. 5) for the external test set affords an encouraging validation for the predictive power of eq. 4. Equation 5 is indeed a satisfactory one given its statistics and its slope being close to 45°, whereas the analysis of all residuals (training and test sets) shows that there is no correlation with the log $D^{7.4}$ values suggesting that eq. 4 is equally predictive for polar and apolar dipeptides (plot not shown).

In general, similar correlative studies involving experimental pK values afforded worst relationships. Nevertheless, eq. 6 unveils the possibility to predict pK_1 values with satisfactory statistics. In detail, eq. 6 includes the number of rotors, PSA means, and PSA ranges, thus emphasizing that the N-terminus ionization depends on molecular flexibility (as encoded by rotors) and intermolecular interactions (as encoded by PSA parameters). Noticeably, both PSA parameters appear with negative coefficients indicating that the exposure of amino group destabilizes its protonated state.

CONCLUSIONS

Although peptides find ever increasing therapeutic applications, their physicochemical profiling has not received the attention it deserves. Similarly, the computational approaches to predict the main physicochemical properties have been mainly focused on small organic molecules, whereas the studies of peptides were performed sporadically and failed to yield encouraging results. Lipophilicity is not an exception, and indeed, the best fragmental approaches have intrinsic difficulties to successfully predict the log P of peptides.[39]

Besides the scarcity of computational studies purposely focused on peptides, such a failure can find two main justifications. First, peptides are quite complex molecules in which flexibility (of both backbone and side chains) and the simultaneous presence of more interacting functional groups concur to hamper a successful log P prediction. Second, the few available data concern peptides that differ largely in length, capping, and polarity, a heterogeneity that further complicates the development of accurate predictive methods. As stated by Thompson and coworkers,[39] such problems would be resolved by producing a proper dataset composed by homogenous peptides that exhaustively cover the available chemical space. Moreover, although several computational approaches and many chirality descriptors have been proposed in the last years to predict enantioselective binding, asymmetric reactivity, or chiral separations,[40,41] the physicochemical profiling of diastereoisomers has been scantly investigated and very few studies have tried to predict their differences.

Accordingly, this study was undertaken with a view to fill these noteworthy gaps by (1) preparing a homogeneous set of diastereoisomeric peptides that suitably cover the accessible chemical space, (2) rationalizing the experimentally determined physicochemical differences between diastereoisomers, and (3) developing correlative models exploiting the property space parameters to conveniently predict the lipophilicity of dipeptides also accounting for their configurational factors. The results presented herein confirm that the accuracy of conformer-dependent predictive methods can be enhanced by exploiting property space parameters that can take both property variability and diastereoisomeric differences into account. One doubts that versatile equations could be derived for heterogeneous molecules. In contrast, it now appears possible to develop targeted relationships for homogeneous classes of compounds accurate enough to satisfactorily deal with the challenge of diastereoisomeric peptides. Finally, the obtained results suggest that property space parameters can also support the prediction of ionization constants, always within homogeneous datasets, thus underlining a similar relevance of the property space concept in the whole physicochemical profiling.

LITERATURE CITED

1. Pliska V, Testa B, van de Waterbeemd H, In Lipophilicity in drug action and toxicology, VCH Weinheim, New York, 1996.
2. Testa B, Vistoli G, Pedretti A A fresh look at molecular structure and properties. In (Manhold R, Ed.) Molecular drug properties. Wiley-VCH, Weinheim: Germany, 2007, p 3–23.
3. Liu X, Testa B, Fahr A. Lipophilicity and its relationship with passive drug permeation. Pharm Res 2011;28:962–77.
4. Henchoz Y, Bard B, Guillarme D, Carrupt PA, Veuthey JL, Martel S. Analytical tools for the physicochemical profiling of drug candidates to predict absorption/distribution. Anal Bioanal Chem 2009;394:707–29.
5. Meanwell NA. Improving drug candidates by design: a focus on physicochemical properties as a means of improving compound disposition and safety. Chem Res Toxicol 2011;24:1420–56.
6. Wang J, Skolnik S. Recent advances in physicochemical and ADMET profiling in drug discovery. Chem Biodivers 2009;6:1887–99.
7. Ritchie TJ, Ertl P, Lewis R. The graphical representation of ADME-related molecule properties for medicinal chemists. Drug Discov Today 2011;16:65–72.
8. Kerns EH, Di L, Carter GT. In vitro solubility assays in drug discovery. Curr Drug Metab 2008;9:879–85.
9. Pagliara A,Caron G,Lisa G,Fan W,Gaillard P,Carrupt PA,Testa B, Abraham MH. Solvatochromic analysis of di-n-butyl ether/water partition coefficients as compared to other solvent systems. J Chem Soc, Perkin Trans 1997;2:639–43.
10. Carrupt PA, Testa B, Gaillard P Computational approaches to lipophilicity: methods and applications. Rev. Comput. Chem. 1997;11:241–315.
11. Mannhold R, Poda GI, Ostermann C, Tetko IV. Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. J Pharm Sci 2009;98:861–93.
12. Leo A. Calculating log Poct from structures. Chem Rev 1993;93:1281–1306
13. Efremov RG, Chugunov AO, Pyrkov TV, Priestle JP, Arseniev AS, Jacoby E Molecular lipophilicity in protein modeling and drug design. Curr Med Chem 2007;14:393–415.
14. Kraszni M, Bányai I, Noszál B Determination of conformer-specific partition coefficients in octanol/water systems. J Med Chem 2003;46:2241–2245.
15. Cruciani G, Mannhold R, Kubinyi H, Folkers G. Molecular interaction fields: applications in drug discovery and ADME prediction. Weinheim: Wiley, 2005.
16. Vistoli G, Pedretti A, Testa B. Chemodiversity and molecular plasticity: recognition processes as explored by property spaces. Future Med Chem. 2011;3:995–1010.
17. Vistoli G, Pedretti A, Testa B Assessing drug-likeness – what are we missing?. Drug Discov Today 2008;13:285–293.
18. Goldreich O. In Computational complexity: a conceptual perspective. Cambridge: Cambridge University Press, 2008.

19. Vistoli G, Pedretti A, Testa B. Partition coefficient and molecular flexibility: the concept of lipophilicity space. Chem Biodivers 2009;6:1152–69.
20. Panusa A, Aldini G, Orioli M, Vistoli G, Rossoni G, Carini M. A sensitive and specific precursor ion scanning approach in liquid chromatography/electrospray ionization tandem mass spectrometry to detect methylprednisolone acetate and its metabolites in rat urine. Rapid Commun Mass Spectrom 2010;24:1583–94.
21. Bodor N, Huang MJ. Predicting partition coefficients for isomeric diastereoisomers of some tripeptide analogs. J Comput Chem 1991;12:1182–1186.
22. Testa B, Pedretti A, Vistoli G. Organic stereochemistry: part 1: symmetry elements and operations, classification of stereoisomers. Helv Chim Acta 2012, in press.
23. Munegumi T. Hydrophobicity of peptides containing D-amino acids. Chem Biodivers 2010;7:1670–9.
24. Hipkiss AR. Carnosine and its possible roles in nutrition and health. Adv Food Nutr Res 2009;57:87–154.
25. Vistoli G, Carini M, Aldini G. Transforming dietary peptides in promising lead compounds: the case of bioavailable carnosine analogs. Amino Acids, 2012, Epub: Jan 28, (DOI) 10.1007/s00726-012-1224-z.
26. Avdeef A pH-metric LogP. II. Refinement of partition coefficients and ionization constants of multiprotic substances. J Pharm Sci 1993;82:183–190.
27. Pedretti A, Villa L, Vistoli G VEGA: a versatile program to convert, handle and visualize molecular structure on windows-based PCs. J Mol Graph 2002;21:47–49.
28. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;16:1781–802.
29. Vistoli G, Orioli M, Pedretti A, Regazzoni L, Canevotti R, Negrisoli G, Carini M, Aldini G. Design, synthesis, and evaluation of carnosine derivatives as selective and efficient sequestering agents of cytotoxic reactive carbonyl species. ChemMedChem 2009;4:967–75.
30. Dawson RMC; Elliott DC; Elliott WH. In Data for biochemical research. Clarendon Press: Oxford, 1959.
31. Perrin DD. In Dissociation constants of organic bases in aqueous solution, Butterworths: London, 1965.
32. Plasson R; Cottet H Determination and modeling of peptide pKa by capillary zone electrophoresis. Anal Chem 2006;78:5394–5402.
33. Vree C, Mayr SG. Dynamics and diffusive–conformational coupling in polymer bulk samples and surfaces: a molecular dynamics study New. J. Phys. 2010;12:023001.
34. Gaillard P, Carrupt PA, Testa B, Boudon A. Molecular lipophilicity potential, a tool in 3D-QSAR. Method and applications. J Comput Aided Mol Des 1994;8:83–96.
35. Clark DE. What has polar surface area ever done for drug discovery? Future Med Chem. 2011;3:469–84.
36. Vistoli G, Pedretti A, Villa L, Testa B Solvent constraints on the property space of acetylcholine. I. Isotropic solvents. J Med Chem 2005;48:1759–1767.
37. Vistoli G, Pedretti A, Villa L, Testa B Solvent constraints on the property space of acetylcholine. II. Ordered media. J Med Chem 2005;48:6926–6935.
38. Vistoli G, Pedretti A, Testa B, Matucci R The conformational and property space of acetylcholine bound to muscarinic receptors. Arch Biochem Biophys 2007;464:112–121.
39. Thompson SJ, Hattotuwagama CK, Holliday JD, Flower DR. On the hydrophobicity of peptides: comparing empirical predictions of peptide log P values. Bioinformation. 2006;1:237–41.
40. Del Rio A. Exploring enantioselective molecular recognition mechanisms with chemoinformatic techniques. J Sep Sci 2009;32:1566–84.
41. Feenstra KA, Hofstetter K, Bosch R, Schmid A, Commandeur JN, Vermeulen NP. Enantioselective substrate binding in a monooxygenase protein model by molecular dynamics and docking. Biophys J 2006;91:3206–16.