What Differentiates Free Amino Acids and Aminoacyl Residues? An Exploration of Conformational and Lipophilicity Spaces

by **Bernard Testa***, **Isabelle Raynaud**, and **Lemont B. Kier**[1])

Institut de Chimie thérapeutique, BEP, Université de Lausanne, CH-1015 Lausanne-Dorigny, Switzerland

The objective of this study was to unravel the changes in property space resulting from the amino-acid-to-residue transformation. Conformation-dependent lipophilicity was chosen as the metric to assess changes in property spaces. Phe, Ala-Phe-Ala, Gln, and Ala-Gln-Ala were first submitted to a conformational search strategy using quenched molecular dynamics in order to obtain an efficient sampling of a conformational space. This search was performed for the four electrical forms of the compounds (cationic, zwitterionic, uncharged, and anionic). The virtual lipophilicity (logP) of each conformer was then calculated by the Molecular Lipophilicity Potential (MLP). Similarly, the lipophilicity increment of the Phe and Gln residues in all electrical states and conformers of Ala-Phe-Ala and Ala-Gln-Ala, respectively, were calculated by the MLP. As expected, the results showed a marked expansion in the property space of a tripeptide compared to an amino acid. However, they also revealed a marked reduction in property space resulting from the amino-acid-to-residue transformation.

Introduction. – *Complex Systems and Their Constituents.* The natural world is made of entities transacting or merging to form complex systems of higher order [1–4]. In a hierarchy of complexity ranging from the smallest to the largest objects, the domains of chemistry and biochemistry span atoms to molecules to macromolecules and molecular aggregates [4–8].

An essential characteristic of complex systems is that they display emergent properties, defined as properties that do not exist, and may even be meaningless in isolated systems of lower hierarchical order [1–4]. Thus, it is obvious to (bio)chemists that molecules can have a number of properties (*e.g.*, molecular topology, stereoisomerism, and conformational freedom) which are not recognizable in atoms. Similarly, macromolecules can have properties that are not found in their monomers. This is particularly true of biological macromolecules such as proteins and nucleic acids, which have evolved to possess an extraordinary array of functions (recognition of exquisite specificity, transport of chemicals and electrons, enzymatic catalysis, formation of molecular machines, *etc.*).

One aspect in the emergence of complex systems that has received little attention is the changes in the properties of constituting entities when they become integrated into a higher system. Indeed, some authors have called attention to the constraints and hierarchical controls imposed on their constituents by complex systems [9–12]. Chemical systems are no exception to this rule. Thus, the formation of covalent bonds between atoms to produce molecules is accompanied by the loss of many properties

[1]) Permanent address: Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298, USA

characteristic of isolated atoms. This is the case of the properties associated with the external layers of electrons, the valence electrons. These electrons no longer belong to the atom, but become delocalized over bonds or over the entire molecules to form the molecular orbitals which account for so many emergent molecular properties [1][4].

The question we address here is similar to that of atoms-in-molecules, but at the higher level of monomeric units incorporated into macromolecules [7]. It is trivial to recall that monomers become residues by undergoing chemical alterations such as loss of H and OH (in macromolecules formed by a reaction of dehydration, *e.g.*, proteins and polysaccharides), or change of a double bond to a single bond (in polymers formed by addition, *e.g.*, polyethylene). But it is intuitively understood that there is more than mere chemical alterations in the monomer-to-residue transformation. Physicochemical properties are also expected to be altered, a phenomenon we begin exploring here.

Exploring Property Spaces. All chemical compounds are characterized by their form (*i.e.*, the structure in a narrow sense), their function (the physicochemical properties), and their fluctuation (*e.g.*, flexibility, tautomerism) [2][7][13]. Form, function, and fluctuation are mutually interdependent, and together they define the molecular states in which a chemical compound can exist. In turn, the ensemble of all chemically possible molecular states delineates the property space of a chemical compound. A physically realistic representation of the property space is afforded by an energy landscape [14], namely a hypersurface whose dimensions are the energy of the system, plus all its other variables.

The subject of this paper is to examine how the property space of amino acids is altered upon their incorporation into peptides, *i.e.*, upon their transformation into aminoacyl residues [7]. The critical issue in such a comparison is to find an adequate metric. Because molecular fields contain information on form, function, and fluctuation, they provide a global and particularly informative approach to the diversity of molecular states and to property spaces. Molecular fields are, for example, molecular electrostatic potentials (MEPs) [15] and the molecular lipophilicity potential (MLP) [16–19]. The latter was chosen as the metric of this study for the following reasons.

First, lipophilicity expresses many intermolecular recognition forces and is, in fact, one of the biochemically and pharmacologically most informative physicochemical property. Second, the MLP is highly sensitive to 3D effects, conformational factors, and changes in ionization, meaning that it also encodes information on the 3D structure and dynamics of a molecule. Third, the MLP can be used to back-calculate a partition coefficient (*i.e.*, the lipophilicity) not only of many compounds, but of many of their molecular states defined by ionization and conformation. In other words, the MLP allows computation of the (experimentally unmeasurable) lipophilicity of individual conformers, called their *virtual lipophilicity*. The point of importance here is that, by calculating a virtual lipophilicity for all recognizable molecular states (*e.g.*, conformational and electrical) of a compound, one can survey the space of all lipophilicity values accessible to the various electrical states of that compound, when their conformation is allowed to fluctuate within energetically realistic values. The lowest and highest values of virtual lipophilicity define the lipophilicity range of a compound.

The final reason for using the MLP is the possibility to compute not only the virtual lipophilicity of all conformers of a given compound, but also the *virtual lipophilicity*

increment of a given residue in all conformations of an oligomer or polymer. In other words, the MLP can explore the property space of a residue and define its lipophilicity range.

A comparison of lipophilicity ranges is the metric used here to assess the differences in property space between amino acids and aminoacyl residues. To this end, we selected two amino acids, namely phenylalanine (which has a hydrophobic side chain with few degrees of conformational freedom) and glutamine (which has a flexible polar side-chain). Their conformational ranges were explored for each possible ionic state (cationic, anionic, zwitterionic, and uncharged), and a virtual logP was calculated for each retained conformer. The entire exercise was repeated for the two tripeptides Ala-Phe-Ala and Ala-Gln-Ala, examining both the entire peptide and the central residue. The results, indeed, revealed a marked reduction in property space resulting from the amino-acid-to-residue transformation.

Results and Discussion. – *Conformational Space.* Amino acids and peptides lacking a ionizable side-chain can exist in four distinct ionic states: cationic in acidic media, anionic in alkaline media, and as zwitterionic and uncharged tautomers in a broad pH range around neutrality [20]. Such states represent one of the dimensions in the property space of amino acids and peptides. This was taken into account here by carrying out a molecular-dynamics exploration of the conformational space of the four ionic states of Phe, Gln, Ala-Phe-Ala, and Ala-Gln-Ala.

Our search strategy [21–25] has been shown to explore efficiently a conformational space. In a first step, 200 out of 2000 generated conformers are randomly selected and their energy minimized. All pairs of conformers are compared, and, when two conformers show geometric similarity plus comparable energy (difference <3 kcal/mol), the conformer of higher energy is eliminated. This strategy ultimately yields a well distributed sampling of conformers, and the number of conformers thus obtained is related to the flexibility of the compound. As shown in the *Table*, Phe and Gln gave a mean of 11 and 23 conformers, respectively, for their four ionic forms. This result is in line with the difference in the flexibility of their side chains. The two peptides Ala-Phe-Ala and Ala-Gln-Ala gave a mean of 30 and 36 conformers, respectively, again a coherent result. The differences in the number of conformers for the various ionic states of the compounds are difficult to interpret and are not of relevance here. We simply note, as a trend, that the zwitterions appear to be slightly restricted respective to the other ionic forms.

Lipophilicity Ranges. The MLP was computed for each conformer of each ionic state of each compound [16][17][26][27]. When integrated over the solvent-accessible surface area (SASA) of a molecule or a molecular fragment, the MLP allows back-calculation of the lipophilicity (expressed as logP, log of octanol/H_2O partition coefficient) of that compound, or of the lipophilicity increment of that fragment [24]. The conformers of lowest and highest virtual logP determine the lipophilicity range accessible to a compound in a given electrical state.

The *Table* reports the ranges in lipophilicity for Phe, Gln, Ala-Phe-Ala, and Ala-Gln-Ala for each of their four ionic states. The two amino acids span lipophilicity values from -3.19 to 0.05 (range 3.24) for Phe, and from -5.61 to -2.06 (range 3.55) for Gln, with two forbidden zones in between, one rather narrow and the second broad. The two

Table. *Conformational and Lipophilicity Space of Phe, Ala-Phe-Ala, the Phe Residue in Ala-Phe-Ala, Gln, Ala-Gln-Ala, and the Gln Residue in Ala-Gln-Ala*

Compound	Number of conformers	Lipophilicity range	Lipophilicity range of central residue
Phenylalanine			
uncharged	17	−0.06 to 0.05 (0.11)	–
cation	13	−2.30 to −1.87 (0.43)	–
anion	10	−2.19 to −2.07 (0.12)	–
zwitterion	4	−3.19 to −2.87 (0.32)	–
Glutamine			
uncharged	27	−2.51 to −2.06 (0.45)	–
cation	27	−4.71 to −4.02 (0.69)	–
anion	20	−4.59 to −4.27 (0.32)	–
zwitterion	18	−5.61 to −4.96 (0.65)	–
Ala-Phe-Ala			
uncharged	29	−1.04 to −0.38 (0.66)	0.34 to 1.18 (0.84)
cation	35	−3.37 to −1.65 (1.72)	0.03 to 1.34 (1.31)
anion	28	−3.17 to −2.24 (0.93)	0.08 to 1.16 (1.08)
zwitterion	27	−5.02 to −3.21 (1.81)	−0.06 to 1.16 (1.22)
Ala-Gln-Ala			
uncharged	32	−3.35 to −2.44 (0.91)	−1.61 to −0.71 (0.90)
cation	38	−5.61 to −3.79 (1.82)	−1.64 to −0.58 (1.06)
anion	38	−5.53 to −4.43 (1.10)	−1.97 to −0.96 (1.01)
zwitterion	36	−6.33 to −5.49 (0.84)	−1.92 to −0.88 (1.04)

peptides span lipophilicity values from −5.02 to −0.38 (range 4.64) for Ala-Phe-Ala, and from −6.33 to −2.44 (range 3.89) for Ala-Gln-Ala, with just one narrow forbidden zone in between.

The primary objective of this study was a comparison between the property spaces of amino acids and aminoacyl residues. The data to be compared are also reported in the *Table*. Indeed, the virtual increments span the values −0.06 to 1.34 (range 1.40) for the phenylalanyl residue, and −1.97 to −0.58 (range 1.39) for the glutaminyl residue, with no forbidden zone.

The lipophilicity space of these compounds and residues can be better understood in pH-lipophilicity plots (*Figs. 1* and *2*). Amino acids have pK_a values of around 2.3 (pK_{a1}, COOH) and 9.3 (pK_{a2}, NH_2), the corresponding values for oligopeptides being *ca.* 3 and 8 [28][29]. In *Figs. 1* and *2*, the lipophilicity range of the cationic forms is represented in the acidic pH range below pK_{a1} (where they predominate), the lipophilicity range of the anionic forms in the alkaline pH region above pK_{a2}, and the lipophilicity range of the zwitterionic and uncharged tautomers in the pH range between pK_{a1} and pK_{a2}. In this pH range, the zwitterionic form of amino acids vastly outnumbers its uncharged tautomer, yet the latter also belongs to the property space and was, therefore, taken into consideration. The plots in *Figs. 1* and *2* are merely indicative as far as the vertical axis is concerned, yet they afford an easily grasped visualization of the lipophilicity space of the free amino acids, the peptides, and the residues.

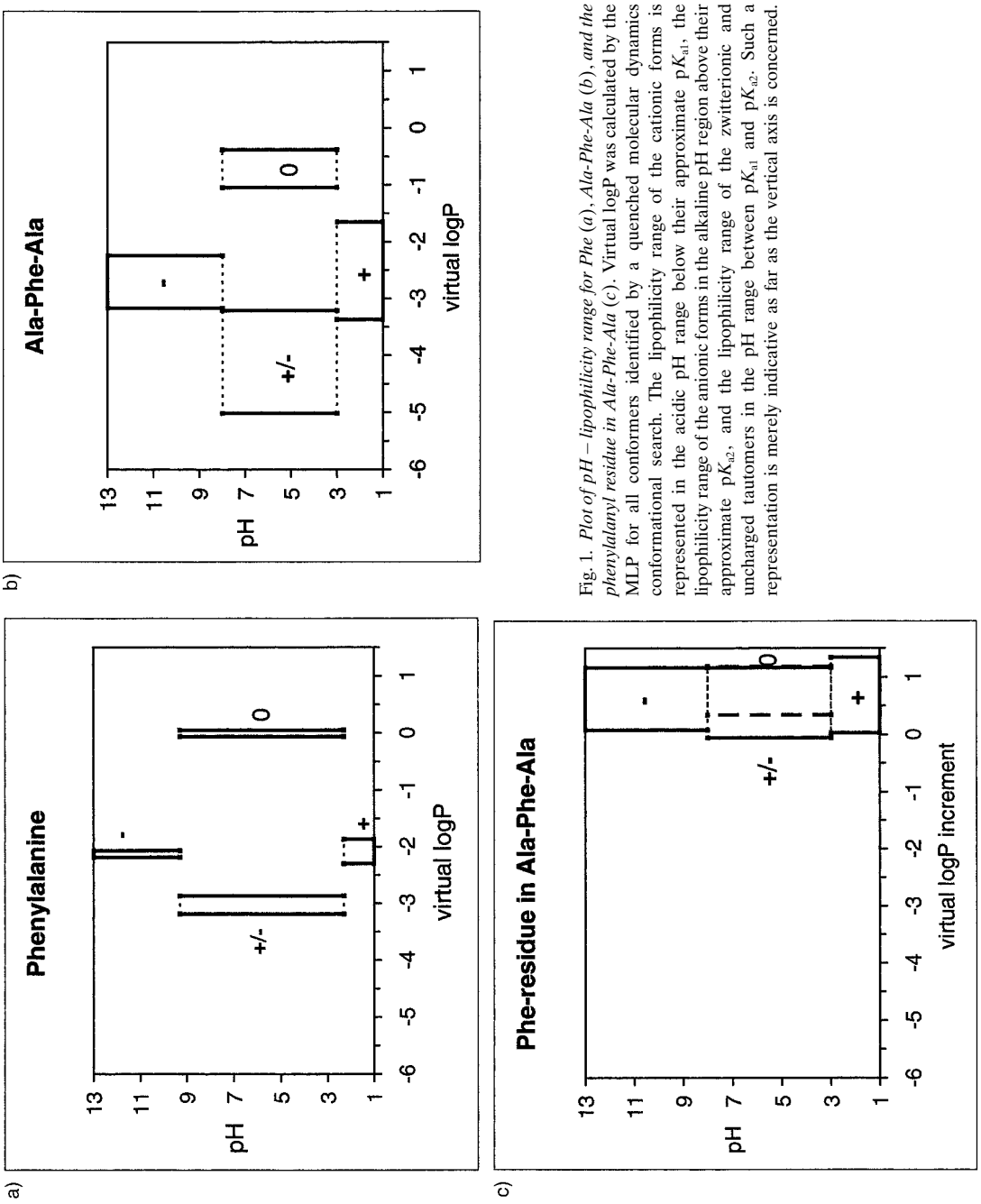

Fig. 1. *Plot of pH – lipophilicity range for Phe* (*a*), *Ala-Phe-Ala* (*b*), *and the phenylalanyl residue in Ala-Phe-Ala* (*c*). Virtual logP was calculated by the MLP for all conformers identified by a quenched molecular dynamics conformational search. The lipophilicity range of the cationic forms is represented in the acidic pH range below their approximate pK_{a1}, the lipophilicity range of the anionic forms in the alkaline pH region above their approximate pK_{a2}, and the lipophilicity range of the zwitterionic and uncharged tautomers in the pH range between pK_{a1} and pK_{a2}. Such a representation is merely indicative as far as the vertical axis is concerned.

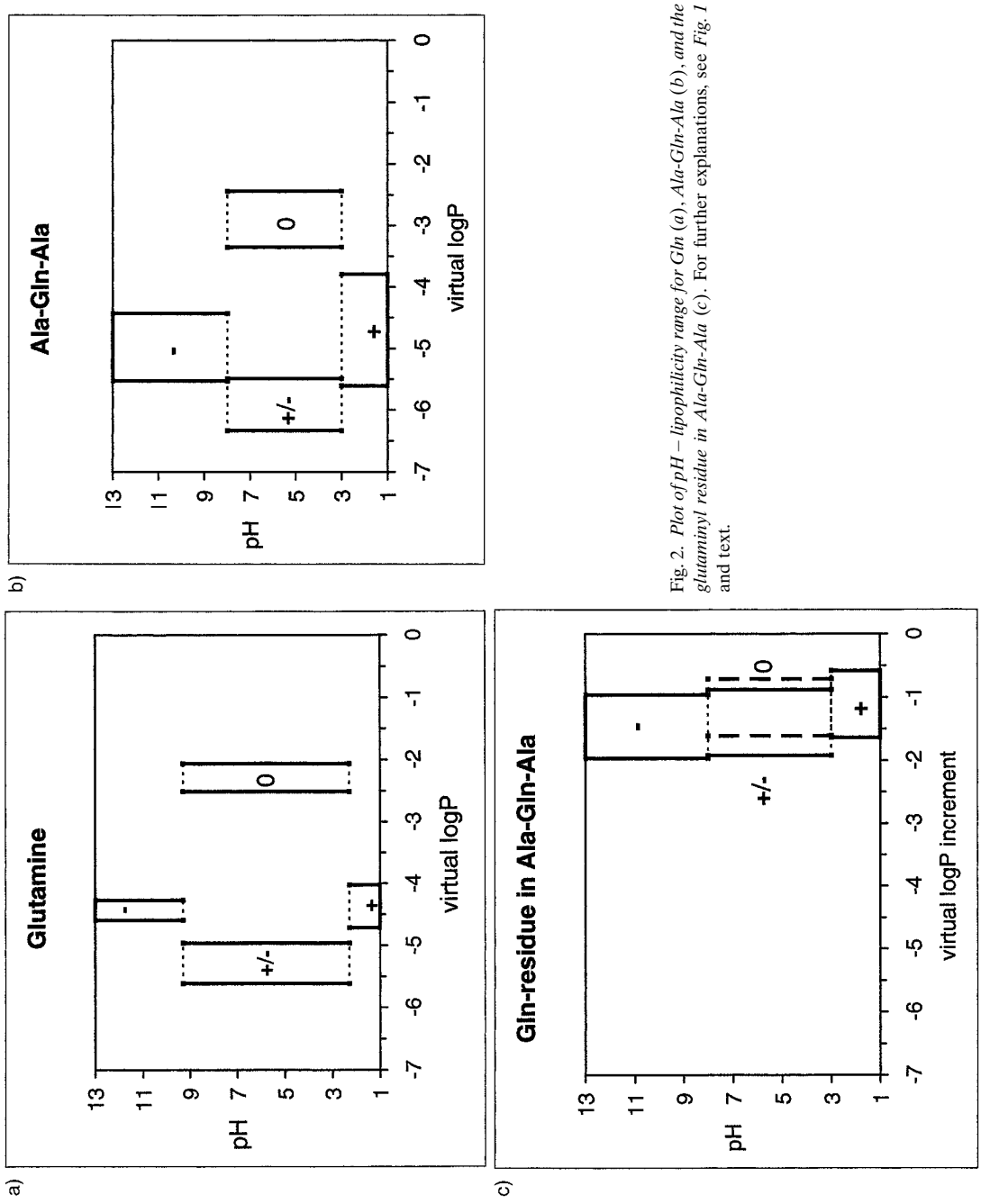

Fig. 2. *Plot of pH – lipophilicity range for Gln* (*a*), *Ala-Gln-Ala* (*b*), *and the glutaminyl residue in Ala-Gln-Ala* (*c*). For further explanations, see *Fig. 1* and text.

What appears clearly is that the two amino acids span rather narrow zones within a broad range of virtual lipophilicity (*Fig. 1 a*, and *Fig. 2 a*), whereas the two peptides span broad zones within a very broad range of virtual lipophilicity (*Fig. 1 b* and *Fig. 2 b*). These plots thus afford an informative insight into the enlarged property space of tripeptides compared to individual amino acids.

The comparison between amino acids (*Fig. 1 a*, and *Fig. 2 a*) and aminoacyl residues (*Fig. 1 c*, and *Fig. 2 c*) reveals the considerable alteration in property space resulting from this chemical transformation. Our computational exploration of property spaces, and the resulting *Figs. 1* and *2*, show that, just like atoms in molecules, amino acids in peptides are profoundly constrained by the higher system of which they are constituents. Indeed, the residues are imbedded in an extended and flexible molecular environment, including distal ionizable groups. That such a molecular environment should influence the property space of residues is to be expected. However, the nature and intensity of these influences is anybody's guess in the absence of relevant information. The present study offers a first quantitative approach in unveiling such information.

Thus, the profound changes seen in the lipophilicity space of the phenylalanyl and glutaminyl residues compared to free phenylalanine and glutamine result from electrical and conformational influences. It is a trivial to note that the loss of the capacity to ionize (excepting the case of some functional side chains) restricts the property space of aminoacyl residues. In contrast, we see here that the distal ionizable groups, expanded surface area, and great flexibility of the peptides markedly influence the lipophilicity space of aminoacyl residues. Untangling and quantifying these various influences would call for much broader investigations.

Conclusion. – Decades of progress in experimental and computational methods have shown the complex nature and behavior of (bio)chemical compounds, whose global description can be gainfully based on form, function, and fluctuation, as defining a property space. Remarkable emergent properties displayed by complex systems as the (bio)chemical levels include changes in form and function which allow mutual adaptability within the confines of the respective property spaces [8]. In turn, mutual adaptability amplifies incommensurably the scope and efficiency of molecular complementarity on which molecular recognition is based [30], ultimately playing an essential role in most, if not all, chemical and biochemical processes.

A better grasp of property spaces and their evolution as an object, integrated into a higher system, may, therefore, be of some interest in understanding molecular complementarity. In this study, we have investigated the property space of amino acids and its alteration as these become integrated into peptides. What has been confirmed here is that the number of formal and functional states of amino acids decreases as they become residues. This phenomenon we have called dissolvence [7][31], and we have looked at it as constraints (top-down causation) imposed on its constituents by a complex system [12][32]. The value of such a concept will depend on its potential to open new directions of research.

Methods. – Molecular modeling was performed with the SYBYL package version 6.2 (*Tripos Inc.*, St Louis, MO, USA). All calculations were performed on *Origin 2000* (*R10000*) and *O2* (*R5000*) *Silicon Graphics* workstations.

Conformational Studies by Quenched Molecular Dynamics (QMD). The conformational behavior of each amino acid or peptide was explored by a simplified conformational search strategy [21–25] able to describe efficiently a conformational space. The structures were built using the SYBYL/Biopolymer module. Various starting geometries (4 to 6) were used and energy-optimized using the Tripos force field [33] with *Gasteiger-Marsili* formal atomic charges [34] in order to remove initial high-energy interactions. High-temperature molecular dynamics (MD) calculations were carried out at 2000 K. Each simulation was run for 100 ps with steps of 1.0 fs. The frame data were stored every 0.05 ps, giving 2000 frames. The starting velocities were calculated from a *Boltzmann* distribution. Finally, 10% of all conformers were randomly selected and saved in a database thus containing about 200 conformers.

All conformers in the database were then subjected to energy minimization with the same force field as for the MD calculations. The *Powell* minimization method was applied with the gradient value of 0.001 kcal/mol · Å to test for convergence. The maximum number of iterations was set at 3000. The energy-minimized conformers were then classified according to increasing energy. The conformational similarity of the 200 energy-minimized conformers was investigated by comparing all pairs of conformers. The two criteria of comparison were the force-field energy and the RMS distance difference calculated by the option MATCH of SYBYL over all heavy atoms and polar hydrogens. An *ad hoc* Fortran program then calculated the mean and standard deviations of the RMS values. Two conformers were considered identical when their energy difference was ≤ 3 kcal/mol and their RMS distance difference less than or equal to the RMS mean minus the standard deviation. When this was the case, one of the two conformers was eliminated from the database, and it was always the one of higher energy.

Calculation of the MLP. The SASA of each selected conformer of Phe, Gln, Ala-Phe-Ala, and Ala-Gln-Ala was used as the space for integrating the MLP [16][17][26] calculated with the CLIP 1.0 software [27]. Similarly for the phenylalanyl or glutaminyl residue, the SASA of that residue in each selected conformer of the corresponding tripeptide was used as the space for integrating the MLP. The integrated MLP was transformed into a virtual log P_{oct} or a virtual lipophilicity increment by back-calculation using *Eqn. 1* [17][24]:

$$\log P_{oct} = 2.86 \cdot 10^{-3} (\pm 0.24 \cdot 10^{-3}) \sum MLP^{+} + 1.52 \cdot 10^{-3} (\pm 0.22 \cdot 10^{-3}) \sum MLP^{-} - 0.10 (\pm 0.23) \quad (1)$$

$$n = 114,\ r^2 = 0.94;\ s = 0.37;\ F = 926$$

where $\sum MLP^{+}$ and $\sum MLP^{-}$ represent the hydrophobic and polar parts of the molecule or residue, respectively.

The most lipophilic and hydrophilic conformers of each amino acid, peptide, or residue were retained and the difference between their virtual lipophilicity defined the lipophilicity range accessible to that compound or residue.

L.B.K. is grateful to the University of Lausanne for a Visiting Professorship and to the *Herbette Foundation* (University of Lausanne) for a grant.

REFERENCES

[1] B. Testa, L. B. Kier, *Complexity* **1996**, *1*, 29.
[2] F. Cramer, 'Chaos and Order. The Complex Structure of Living Systems', VCH, New York, 1993.
[3] K. Mainzer, 'Thinking in Complexity', Springer-Verlag, Berlin, 1994.
[4] H. Primas, 'Chemistry, Quantum Mechanics and Reductionism', Springer, Berlin, 1981.
[5] H. Rabitz, *Science* **1989**, *246*, 221.
[6] B. Testa, L. B. Kier, *Med. Res. Rev.* **1991**, *11*, 35.
[7] B. Testa, L. B. Kier, P. A. Carrupt, *Med. Res. Rev.* **1997**, *17*, 303.
[8] B. Testa, *Med. Chem. Res.* **1997**, *7*, 340.
[9] H. H. Pattee, in 'Foundations of Mathematical Biology', Ed. R. Rosen, Vol. 1, Academic Press, New York, 1972, pp. 1–22.
[10] H. A. Simon, in 'Hierarchy Theory', Ed. H. H. Pattee, Braziller, New York, 1973, pp. 1–27.
[11] C. Grobstein, in 'Hierarchy Theory', Ed. H. H. Pattee, Braziller, New York, 1973, pp. 29–47.
[12] D. T. Campbell, in 'Studies in the Philosophy of Biology', Eds. F. J. Ayala and T. Dobzhansky, Macmillan, London, 1974, pp. 179–186.
[13] I. Prigogine, *Angew. Chem.* **1978**, *90*, 704.
[14] T. Kenakin, *Trends Pharmacol. Sci.* **1996**, *17*, 190; and ref. cit. therein.
[15] P. A. Carrupt, N. El Tayar, A. Karlén, B. Testa, *Meth. Enzymol.* **1991**, *203*, 638.

[16] P. A. Carrupt, B. Testa, P. Gaillard, *Rev. Comput. Chem.* **1997**, *11*, 241.
[17] P. Gaillard, P. A. Carrupt, B. Testa, A. Boudon, *J. Comput.-Aided Mol. Design* **1994**, *8*, 83.
[18] B. Testa, P. A. Carrupt, P. Gaillard, R. S. Tsai, in 'Lipophilicity in Drug Research and Toxicology', Eds. V. Pliska, B. Testa and H. van de Waterbeemd, VCH Publishers, Weinheim, 1996, p. 49–71.
[19] B. Testa, P. A. Carrupt, P. Gaillard, F. Billois, P. Weber, *Pharm. Res.* **1996**, *13*, 335.
[20] A. Pagliara, P. A. Carrupt, G. Caron, P. Gaillard, B. Testa, *Chem. Rev.* **1997**, *97*, 3385.
[21] P. Gaillard, P. A. Carrupt, B. Testa, *Bioorg. Med. Chem. Lett.* **1994**, *4*, 737.
[22] C. Altomare, S. Cellamare, A. Carotti, G. Casini, M. Ferappi, E. Gavuzzo, F. Mazza, P. A. Carrupt, P. Gaillard, B. Testa, *J. Med. Chem.* **1995**, *38*, 170.
[23] D. Meyer, M. H. Fouchet, M. Petta, P. A. Carrupt, P. Gaillard, B. Testa, *Pharm. Res.* **1995**, *12*, 1583.
[24] Y. Giroud, P. A. Carrupt, A. Pagliara, B. Testa, R. G. Dickinson, *Helv. Chim. Acta* **1998**, *81*, 330.
[25] A. Pagliara, B. Testa, P. A. Carrupt, P. Jolliet, C. Morin, D. Morin, S. Urien, J. P. Tillement, J. P. Rihoux, *J. Med. Chem.* **1998**, *41*, 853.
[26] P. A. Carrupt, P. Gaillard, F. Billois, P. Weber, B. Testa, C. Meyer, S. Pérez, in 'Lipophilicity in Drug Research and Toxicology', Eds. V. Pliska, B. Testa and H. van de Waterbeemd, VCH Publishers, Weinheim, 1996, p. 195–217.
[27] CLIP1.0, Institute of Medicinal Chemistry, University of Lausanne, 1996.
[28] R. S. Tsai, B. Testa, N. El Tayar, P. A. Carrupt, *J. Chem. Soc. Perkin Trans. II*, **1991**, 1797.
[29] H. R. Horton, L. A. Moran, R. S. Ochs, J. D. Rawn, K. G. Scrimgeour, Principles of Biochemistry, Prentice Hall, Englewood Cliffs, N.J., 1993, Sections 3.3 and 6.3.
[30] R. S. Root-Bernstein, P. F. Dillon, *J. Theor. Biol.* **1997**, *188*, 447.
[31] B. Testa, L. B. Kier, *Adv. Drug Res.* **1997**, *30*, 1.
[32] T. F. H. Allen, T. W. Hoekstra, 'Toward a Unified Ecology', Columbia University Press, New York, 1992, Chapter 1.
[33] M. Clark, R. D. Cramer III, N. Van Opdenbosch, *J. Comput. Chem.* **1989**, *10,* 982.
[34] J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, *36*, 3219.

Received February 18, 1999