Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 37 (2002) 865–872
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
www.elsevier.com/locate/ejmech

Original article

Hydration free energy a fragmental model and drug design

Gérard Pèpe [a,*], Gaël Guiliani [a], Sophie Loustalet [a], Philippe Halfon [b]

[a] *Groupe de Chimie Organique et Physique de Luminy, Faculté des Sciences de Luminy, Université de la Méditerranée, UMR-CNRS 6114, case 901, F-13288 Marseilles, Cédex 09, France*
[b] *Genoscience and Alphabio Laboratories, Marseilles, France*

Received 14 January 2002; accepted 14 March 2002

Abstract

The aim of this work was to provide a simple model to determine hydration free energies of molecular compounds at ambient temperature in order to help drug design. Since the free energy of interaction between molecules and water can be approximated by an additive function of their constituent groups, a model based on a fragmental analysis was developed with all the hydration free energies determined from experimental values. The model gives accurate hydration free energies on small molecules by simply adding the hydration energies of the fragments constituting the molecule. The correlation factor r^2 is 0.990 on the 125 molecules used to establish the model parameters. As experimental values are determined at ambient temperature (25 °C), the values obtained from the model will correspond to this temperature. For mean size molecules or big molecules as proteins, it was not possible to calibrate the model owing to the absence of experimental values. However, the model makes it possible to estimate the hydration free energy of such molecules. The fragmental representation of any molecule with coded colours, depending on the fragment nature (hydrophobic, hydrophilic or neutral), can be useful to understand molecule behaviour and help drug design.

Keywords: Hydration energy; Drug design; Hydrophobic

1. Introduction

One of our field of research is the design of protein inhibitors, in order to find new drugs that act, for example on HIV. When such molecules are synthesised, the problem is to know if they are able to reach their targets in the biological medium.

The hydration energy corresponding to the number of atoms, in order to take into account molecule size, is a good predictive index for molecule availability in the biologic medium.

These compounds are generally defined in order to give the best energy association with the enzyme active site. Unfortunately, there is no direct relation between this energy and drug efficiency. In the complex mechanism driving a molecule from water to the protein active site, the effects of hydration play an important role. The driving force that brings the molecule in the active site is directly correlated to the difference between the hydration energy of the drug and the energy association of the drug with the protein. The molecular system must gain energy or the molecule will 'prefer' to stay in water. The efficiency of a drug depends on the energy difference, the lifetime of the complex will be proportional to this difference energy. Thermodynamic dissection of the binding energies of KNI-272, a potent HIV-1 protease inhibitor, with water and with the non-mutated protease (wild type) and in a mutated form confirms this hypothesis [1].

Approximately 150 papers on solvation studies are published every year. However, most of the models used to determine solvation or hydration energies are not easy to apply.

The models are either derived from molecular mechanics in which the solvent is explicitly taken into account, or generally based upon empirical methods where the solvent is treated as a continuum.

Most of the theoretical solvation approaches have been analysed in the review articles of Tomasi and Persico [2] and Cramer and Truhlar [3,4].

* Correspondence and reprints
E-mail address: pepe@luminy.univ-mrs.fr (G. Pèpe).

0223-5234/02/$ - see front matter
PII: S0223-5234(02)01371-5

The very simple model we developed is applied to HIV anti-protease drugs in a comparison with other models.

2. Model and results

The model considers that the free energy of interaction between small molecules and water can be approximated by an additive function of their constituent groups [5].

For small molecules, of which all the fragments accessible to the solvent, the hydration free energy is:

$$\Delta G_{\mathrm{HYD}} = \sum_{i=1}^{N} \Delta G_{\mathrm{fr}_i}$$

where ΔG_{fr_i} is the hydration free energy of the fragment i and N is the total number of fragments in the molecule.

In the model, the aromatic systems are treated in a Kekule representation for the fragmental analysis, but the energy obtained with the C=C groups must be corrected as follow:

$$\Delta G_{\mathrm{HYDaro}} = n \times \Delta G_{\mathrm{C=C}} - m \times 1.8$$

where n is the number of double bonds in a Kekule representation, and m the number of aromatic centres of the system; for example:

in benzene, $n = 3$ and $m = 1$;
in naphthalene, $n = 5$ and $m = 2$;

which gives -1.05 and -2.1 kcal mol^{-1} for the hydration free energies of these molecules, respectively.

Remark 1. This approximation is not of a great importance because hydration free energy of such hydrophobic groups is close to zero, their contribution to the hydration free energy of a water soluble molecule, will be weak. The volume of these fragments will be significant parameter for the hydrophobic forces that act on the molecule. The distribution of the hydrophobic fragments is more important than their number in a water-soluble molecule. Few hydrophilic fragments can compensate the presence of a great number of hydrophobic fragments in order to keep the molecule water soluble, like in soap molecules for example.

Molecules are decomposed into fragments that often correspond to the usual chemical functions. The analysis is performed in such way to identify the peripheral fragments, and if impossible, not taking into account the buried atoms. In neo-pentane for instance, the molecule is considered to be made up of four groups, the central carbon is ignored.

The energy obtained from the model corresponds to the transfer of the molecule from the vapour phase to water. For groups as COOH and NH_2, the model does not consider ionisation, but only the hydration of the groups (however, the ratio between the charged and the uncharged forms of these groups at a given pH can be calculated from the molecule pK_a—and therefore, the hydration free energy of the molecule at any pH).

All the fragment hydration free energies used in the model were deduced from experimental values [6–8]. As these values were established at ambient temperature (25 °C), the model will give hydration energies at this temperature. Fragments with their hydration energies are reported in Table 1.

The constituting fragments of a molecule must be identified in an order corresponding to the hierarchy that exists between the fragments. In most cases, this hierarchy is related to the fragment's atomic size (generally from the biggest to the smallest).

Analysis order is very important. For example: the hydration energy of C=O and OH groups are -5.6 and -8.0 kcal mol^{-1}, respectively, while the hydration energy of the COOH fragment is -7.6 kcal mol^{-1} (not -13.6 kcal mol^{-1}), the COOH fragment must be identified first.

Note 1. The program treats other fragments built around S and P atoms. As their parameters could not be calibrated on experiment values but only on calculated values. For reasons of homogeneity, they are not provided in this paper.

In Table 2, we report the molecules used to establish the hydration free energies of the fragments, the hydration free energy values observed and determined from the model. The correlation coefficient r^2 between the values is 0.990.

It has been shown that 'AM1-SM2 full' [9] gives, on a chosen sample of small molecules, the hydration free energy that is the closest to the experimental ones. Our model was applied to the same molecules, the results indicate that our values are globally closer to the experimental ones, as displayed in Table 3. The correlation coefficients r^2 obtained with our model and 'AM1-SM2 full' are 0.964 and 0.938, respectively.

3. Hydration free energy and accessible surface area

Although the model gives provides good correlation coefficients between the calculated and the observed hydration free energy values, it can be only applied to small molecules whose size corresponds to the mean size of the molecules used to establish it. In order to extend the model to larger molecules, we tried to generalise it in taking into account the surface area value of the fragments in contact water, to determine both the fragment hydration energies and to calculate the hydration energies of the molecules.

For calculation of fragment surface areas, these atom groups were assimilated to spheres whose radii Table 1 corresponds to the limits of the van der Waals spheres of their constituting atoms. Thus the spheres were placed on the centroids of these atoms. The Connolly [10] algorithm was used to represent the molecule surface defined by these spheres and to calculate the contact surface area with the solvent. The molecules were generated by GENMOL ([11]) (http://chimir1.univ-mrs.fr/serveur). The selected applications of this software can be found in the following papers [12–14].

This approximation led to a lower value of the correlation factor $r^2 = 0.903$ between the calculated and the experimental values. This result, therefore, indicated that the effects of hydration were not only related to the contact surface area but also to the nature of the fragment. This observation was confirmed by Ulmschneider and Pénigault [15] who proposed, at the atomic level, a correction taking into account the σ or the π area of the atom surface in contact with water. The same approach in a fragmental representation could be of interest for larger molecules. Unfortunately the lack of experimental values for such molecules does not make it possible to improve the model in this way.

Table 1
Fragments, hydration free energy (kcal mol^{-1}), mean van der Waals diameter (Å) and fragment nature

Fragment	ΔG_{fr}	Mean van der Waals diameter	Nature
CF_3	1.50	3.96	HB
CCl_3	−1.15	4.29	HB
CBr_3	−1.50	4.38	HB
CI_3	−2.05	4.52	HB
COO^-	−78	3.70	HL
COOH	−7.60	3.90	HL
COO	−5.02	3.73	HL
$CONH_2$	−10.60	4.18	HL
$CSNH_2$	−8.80	4.32	HL
CONH	−10.60	4.03	HL
CF_2	−1.00	3.75	HB
CCl_2	−1.10	3.99	HB
CBr_2	−1.40	4.07	HB
CI_2	−1.80	4.18	HB
NO_2	−4.70	3.83	NE
N–C=O	−4.60	3.73	NE
C=NH	−4.00	3.68	NE
C=N	−3.70	3.54	NE
OCH_3	−2.90	4.02	HB
C–O–C	−3.60	3.76	NE
SCH_3	−2.30	4.16	HB
C–S–C	−3.10	3.88	HB
NH_3^+	−71.00	3.74	HL
CN	−3.60	3.39	NE
CO	−5.60	3.40	HL
CHO	−4.52	3.67	NE
C=S	−2.80	3.56	HB
NH_2	−5.60	3.54	HL
NH	−5.40	3.34	HL
NCCC	−4.80	3.99	NE
CF	−1.00	3.60	HB
CCl	−1.50	3.65	HB
CBr	−1.55	3.72	HB
Cl	−1.65	3.82	HB
CH_2	0.18	3.61	HB
NCH_3	−4.00	4.04	NE
OH	−8.00	3.31	HL
SH	−2.25	3.45	NE
CH_3	0.90	3.78	HB
C=C	0.25	3.44	HB
C-3-C	−1.20	3.40	HB
CH	0.15	3.37	HB

HB, hydrophobic; HL, hydrophilic and NE, neutral (the frontiers hydrophobic/neutral and hydrophilic/neutral are arbitrary; C–O–C, ether junction, carbon atoms are sp^3; NCCC, tertiary nitrogen atom linked to sp^3 carbon atoms; C-3-C, a triple bond between two carbon atoms.

4. Model and mean size molecules

In order to have an idea on the hydration free energy accuracy deduced from the model on mean-sized molecules (where all the fragments are not necessarily contacting water), the calculations were performed on six HIV protease inhibitors. Their geometry came from the Brookhaven Protein Data Bank [16], where these drugs are associated with the HIV protease. The structure codes are, respectively: 1HXW, 1OHR, 1HXB, 2BPX, 1HPV, for complexes with ritonavir, nelfinavir, saquinavir, indinavir and amprenavir. The geometry of lopinavir is obtained by an energy fit of the molecule geometry to the protease active site with the help of GENMOL. In Table 4, we report for each inhibitor the number of atoms NAT (to characterise its size) and different values of the hydration free energies: from the model:

- E1 = the sum of the hydration free energies of all the fragments constituting the molecule,
- E2 = the sum of the hydration free energies of only the fragment contacting water whatever the surface area in contact with water,
- E3 = the sum of the hydration free energies of the fragments with more than 50% of their surface area contacting water,

from other models considering the solvent as a continuum:

- COSMO [17] that values are very close to ours.
- AM1-SM2 with geometry optimised in vacuum, the values appeared to be generally underestimated, if compared with all other values, the sum of the energies is −149 instead of −226 kcal mol^{-1}, the average value given by the other methods.
- AM1-SM2 with geometry optimised in water, for ritonavir the energy value −85.8 was not coherent

Table 2
Observed and calculated hydration free energies (kcal mol^{-1}) of a sample of molecules used to define the hydration energies of the fragments

ΔG_{HYD}		
	Experimental	Calculated
Halogenated compounds		
CHF_2CH_3	−0.1	−0.1
CF_3CH_2OH	−4.3	−4.3
CH_3Cl	−0.6	−0.78
CH_2Cly	−1.4	−1.1
$CHCl_3$	−1.1	−1.15
CF_3CH_2Cl	0.1	0
CH_3CH_2Cl	−0.6	−0.6
$ClCH_2CH_2Cl$	−1.7	−3
$(CH_3)_2CHCl$	−0.3	0.3
$ClCH_2CHCl_2$	−2	−2.6
CH_3CCl_3	−0.3	−0.25
$CH_2=CHCl$	−0.6	−1.3
Z-ClCH=CHCl	−1.2	−3
E-ClCH=CHCl	−0.8	−3
$ClCH=CCl_2$	−0.4	−2.6
$Cl_2C=CCl_2$	0.1	−2.2
$BrCH_2CH_2Cl$	−2	−3.2
CH_3Br	−0.8	−0.83
CH_2Br_2	−2.1	−1.4
CH_3CH_2Br	−0.7	−0.65
$BrCH_2CH_2Br$	−2.1	−3.1
CH_3I	−0.9	−0.93
C_2H_5I	−0.7	−0.75
Heterocycles		
Pyridine	−4.7	−5
4-Methylpyridine	−4.9	−4.1
Piperazine	−7.4	−10.1
4-Dimethyl piperazine	−7.6	−8.3
2-methyl pyrazine	−5.5	−8.95
Ethyl-2-methoxy pyrazine	−4.4	−10.77
Piperidine	−2.7	−4.48
Imidazole	−5.9	−8.85
Pyrolidine	−3.1	−4.66
Dioxane	−5.1	−7.2
THF	−3.5	−3.24
Alcohols		
CH_3OH	−5.1	−5.1
C_2H_5OH	−5	−4.92
n-C_3H_7OH	−4.8	−4.74
i-C_3H_7OH	−4.8	−4.74
$CH_2=CHCH_2OH$	−5	−5.57
Hydrocarbons		
CH_3CH_3	1.8	1.8
C_3H_8	2	1.98
n-C_4H_{10}	2.1	2.16
t-C_4H_{10}	2.3	2.85
$C(CH_3)_4$	2.5	3.6
n-C_6H_{14}	2.6	2.52
n-C_8H_{18}	2.9	2.88
Cyclopropane	0.8	0.56
Cyclopentane	1.2	0.9
Cyclohexane	1.2	1.08
C_2H_4	1.3	0.25
$C_2H_3CH_3$	1.3	1.15
$(C_2H_3)_2$	0.6	0.5
$CH_2=C(CH_3)_2$	1.2	2.05
E-$C_2H_5CH=CHCH_3$	1.3	2.23
Cyclopentene	0.6	0.79
CH_3CCH	−0.3	−0.3
CH_3CH_2CCH	−0.2	−0.12
C_4H_9CCH	0.3	0.24
Aromatics		
Benzene	−0.9	−1.05
Toluene	−0.9	−0.15
p-Xylene	−0.8	0.75
Naphtalene	−2.4	−2.35
C_6H_5OH	−6.6	−7.05
C_6H_5SH	−2.7	−3.3
$C_6H_5NH_2$	−4.9	−6.65
$C_6H_5NO_2$	−4.1	−5.75
$C_6H_5OOCCH_3$	−4.6	−5.25
C_6H_5Cl	−1.1	−2.6
C_6H_5Br	−1.5	−2.65
p-BrC_6H_4OH	−7.1	−8.65
o-$CH_3C_6H_4NO_2$	−3.6	−4.85
C_6H_5CHO	−4	−5.57
$C_6H_5SCH_3$	−2.7	−2.7
$C_6H_5COCH_3$	−4.6	−5.75
Thiols		
$(CH_3)_2S$	−1.4	−1.4
$(C2H_5)_2S$	−1.3	−1.3
CH_3SH	−1.2	−1.35
C_2H_5SH	−1.3	−1.17
Ethers		
CH_3OCH_3	−1.9	−2
$CH_3OC_3H_7$	−1.7	−1.64
$C_2H_5OC_2H_5$	−1.8	−1.8
$C_6H_5OCH_3$	−2.4	−3.95
$(CH_3OCH_2)_2$	−4.8	−5.44
Esters/acids		
CH_3COOCH_3	−3.3	−3.22
$CH_3COOC_2H_5$	−3.1	−3.04
$C_2H_5COOCH_3$	−2.9	−3.04
$C_3H_7COOCH_3$	−2.8	−2.86
CH_3COOH	−6.7	−6.7
C_2H_5COOH	−6.5	−6.52
C_3H_7COOH	−6.4	−6.34
Cetones/aldehydes		
$(CH_3)_2CO$	−3.8	−3.8
$CH_3COC_2H_5$	−3.6	−3.62
$(C_2H_{5)2}CO$	−3.4	−3.44
$(C_3H_7)_2CO$	−2.9	−3.08
C_3H_7CHO	−3.2	−3.24
Charged molecules		
$CH_3CONH_3^+$	−66	−75.7
$C_6H_5NH_3^+$	−68	−72
$CH_3NH_3^+$	−70	−70
$C_6H_5COO^-$	−76	−79
CH_3COO^-	−77	−77
Amines		
CH_3NH_2	−4.5	−4.7
$C_2H_5NH_2$	−4.5	−4.52
n-$C_3H_7NH_2$	−4.4	−4.34
n-$C_4H_9NH_2$	−4.3	−4.16
$(CH_3)_2NH$	−4.3	−3.6
$(CII_3)_2N$	−32	−3
$(C_2H_5)_3N$	−3	−2.46
$CH_3CH(OCH_3)NH_2$	−6.6	−7.45
Aniline	−4.9	−6.65
Amides		
CH_3CONH_2	−9.7	−9.7
$CH_3CONHCH_3$	−10	−8.8
Nitro		
CH_3NO_2	−3.7	−3.8
$C_2H_5NO_2$	−3.7	−3.62
Nitriles		
C_4H_9CN	−2	−2.14
Acetonitrile	−2.8	−2.7

The correlation factor r^2 between observed and calculated values is 0.99.

with the value obtained with the geometry optimised in vacuum, the other values generally correspond to an average decrease of -2 kcal mol^{-1}.

- Note: for amprenavir, because the absence of an experiment value (or experiment values) for the SO_2 fragment, and because the analogy between our results and those of COSMO, we attributed the COSMO calculated value $= -4.2$ kcal mol^{-1} to this fragment.

Table 3
Comparison of hydration free energies(kcal mol^{-1}) with the values obtained from our model and calculated with AM1-SM2 full [10] method giving the most reliable values for hydration energies on small molecule), the value of the corresponding correlation factors r^2 are: 0.964 and 0.938, respectively

	Hydration free energies		
	Experimental	Our model	AM1-SM2 full
CH_3Cl	−0.6	−0.8	−0.7
$CHCl_3$	−1.1	−1.2	−1.2
$ClCH_2CH_2Cl$	−1.7	−3	−1
$(CH_3)_2CHCl$	−0.3	0.3	−0.3
CH_3OH	−5.1	−5.1	−5.8
C_2H_5OH	−5	−4.9	−4.9
n-C_3H_7OH	−4.8	−4.7	−4.6
i-C_3H_7OH	−4.8	−4.7	−4.1
CH_3SH	−1.2	−1.4	−0.8
C_2H_5SH	−1.3	−1.2	−0.6
C_6H_5SH	−2.6	−3.9	−3.2
$(CH_3)_2CO$	−3.8	−3.8	−4.1
CH_3OCH_3	−1.9	−2	−1.4
CH_3COOCH_3	−3.3	−3.2	−4
CH_3COOH	−6.7	−6.7	−7.7
C_2H_5COOH	−6.5	−6.5	−6.7
C_3H_7COOH	−6.4	−6.3	−6.3
Benzene	−0.9	−1.1	−0.5
Toluene	−0.9	−0.2	−0.3
C_6H_5OH	−6.6	−7.1	−5.8
C_6H_5SH	−2.7	−3.3	−3.2
Pyridine	−4.7	−5	−4.4
Dioxane	−5.1	−7.2	−3.4
THF	−3.5	−3.2	−1.5
CH_3CH_3	1.8	1.8	1.2
C_3H_8	2	2.	1.4
n-C_4H_{10}	2.1	2.2	1.7
t-C_4H_{10}	2.3	2.9	1.6
n-C_6H_{14}	2.6	2.5	2.2
n-C_8H_{18}	2.2	2.9	2.7
Cyclopropane	0.8	0.6	1.1
Cyclohexane	1.2	1.1	1.9
C_2H_4	1.3	0.3	0.8
CH_3CCH	−0.3	−0.3	−0.9
CH_3NH_2	−4.5	−4.7	−6.2
$C_2H_5NH_2$	−4.5	−4.5	−5.2
n-$C_3H_7NH_2$	−4.4	−4.3	−5
n-$C_4H_9NH_2$	−4.3	−4.2	−4.7
$(CH_3)_2NH$	−4.3	−3.6	−2.6
$(CH_3)NH_2$	−3.2	−3	−3.3
$C_6H_5NH_2$	−4.9	−6.7	−5.8
CH_3CONH_2	−9.7	−9.7	−11.5
$CH_3CONHCH_3$	−10	−8.8	−9.9
Acetonitrile	−2.8	−2.7	−4.3

From the model, the greater difference between the energy values, obtained from the model depending on the approximations, was 6%. This result made it possible to estimate the hydration free energy error on such large molecules (most drug compounds), when calculated by addition of the hydration free energies of all the fragments. It should be noted that such a calculation may be performed by hand, without the knowledge of the molecule geometry. The values given by the model were very close to COSMO values and particularly when only fragments, with more of 50% of their surface area contacting water, were considered. Values provided by AM1-SM2 were generally lower. Consequently, only experimental values on such large molecules will be able to demonstrate the most accurate model.

5. Model and proteins

It is well known, and obvious for the protein solubility, that in such molecules, hydrophilic residues are distributed on the external shell of the molecule, while the hydrophobic ones are buried. As the fragmental analysis is located between the atomic and the residue analysis, we thought that it would be of interest to study proteins in this representation.

For that purpose we chose 84 enzymes from the Brookhaven Protein Data Bank ranging in size from 50 to 456 residues corresponding to the following codes: 1PTQ 1PGA 1PGB 1ISU 1OPS 1HPI 1CEI 1POH 1MZL 1MZM 1SHA 1SHB 2FKE 1BOW 5CYT 1AV5 1FIT 3BP2 1MKS 1RIE 1BM9 1RFS 1A2W 1DKJ 1IFB 2HMB 1KXX 1ICM 2IFB 1OPA 1CRB 1CBQ 1CBS 2AAK 1NCL 1U9B 1BUU 1WPK 1MMB 1AA7 1A6K 1AWR 1AWQ 1VMO 5P21 6Q21 2SCP 1GOX 4SGB 1TIP 2SFA 2TMK 5GST 1YER 3GST 1AK2 2AK2 6GSU 6GSX 6GSW 6GST 4GST 1ORO 1YET 1DST 1TAW 1AMF 1ENY 1HVQ 1AOG 8PRK 1FNC 1FND 1FNB 1MLA 1GFI 4ECA 2ER9 1ATP 1FMO 1YDS 1A99 1AXE 8GEP.

In order to correctly identify the fragments, only complete molecules (with all the atom positions solved) at a resolution better or equal to 2.2 Å, were chosen. The hydrogen atoms were added to the molecules using GENMOL software.

The molecules were considered in a physiological medium at ambient temperature (25 °C) and neutral pH, with all their acidic and basic functions in the charged form.

The results indicated that the most hydrophilic groups: NH_3^+, COO^- and $CONH_2$ representing 6(1)% of the total number of fragments were located on the protein surface (89(11)%), while the most hydrophobic ones: CH, CH_2, CH_3 and C=C (generally corresponding to aromatic systems in a Kekule representation) (67(2)%

Table 4
Hydration free energies of HIV protease inhibitor in kcal mol^{-1}

Inhibitor	Nat	E1	E2	E3	COSMO	(AM1-SM2)V	(AM1-SM2)W
Ritonavir	98	−48.2	−44.6	−42	−43.2	−40.8	−85.8
Nelfinavir	85	−34.6	−34.7	−34.7	−34.1	−13	−15.6
Saquinavir	99	−50.4	−50.6	−46.9	−42	−20.9	−23.3
Indinavir	92	−40.7	−40.8	−40.8	−37.8	−21	−24.4
Amprenavir	70	−25.2	−25.2	−25.2	−32.1	−33.4	−35.4
Lopinavir	94	−38.1	−38.4	−38.7	−39.1	−20	−19.9
SUM		−237.2	−234.2	−228.3	−228.3	−149.1	−204.4

NAT, number of molecule atoms; E1, Sum of hydration energies of all the fragments; E2, Sum of hydration energies of the fragments contacting water; E3, Sum of hydration energies of the fragments with at least 50% of their surface contacting water. The other values come from COSMO, AM1-SM2 with molecules optimised in vacuum and in water, respectively.

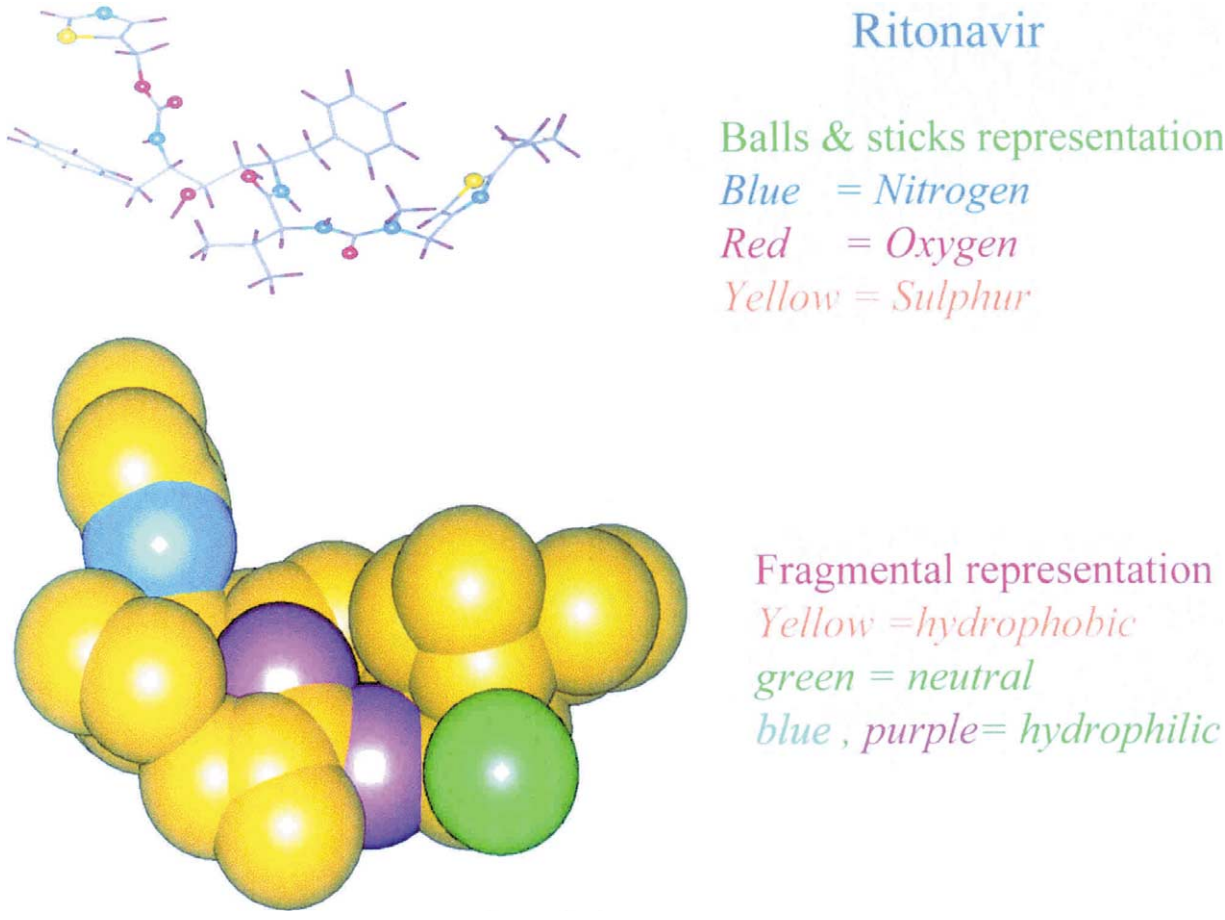

Fig. 1. Ritonavir in 'Ball and sticks' and 'fragmental' representations.

of the total number of fragments) were buried, only 32(5)% in the shell contacting water.

The very hydrophilic fragments CONH (20(1)% of the total number of fragments) belonging to the protein main chains are buried with only 25(6)% in the external shell of fragments, inducing a huge loss of hydration energy, about half of the total value.

Remark 2. This analysis provided a more accurate representation than a residue analysis-in terms of hydrophilic or hydrophobic protein regions, and made it possible to point out this last very important feature along with the existence of compensatory energies to keep the protein folded. We will not develop this point since it is not part of our subject.

The fragment was considered on the molecule surface, according to the correspondence with the COSMO results, when at least 50% of its surface area in a Connolly representation is contacting solvent.

The model made it possible to make an estimation, in the approximation established in the comparison with other models given above-of the hydration free energy

of a protein at ambient temperature in a neutral medium:

$(-1.14\text{Nat}-492)$ kcal mol^{-1} ($r^2 = 0.87$),
Nat = the number of atoms in the molecule.

Note 2. The standard deviations given in the text and in the tables correspond to the distribution widths of the different values and must not be taken as confident domains.

6. The fragmental representation of molecules

Independently of the energy values, the model may be used to represent molecules with coded colours depending on their nature (hydrophilic, hydrophobic or neutral). The HIV protease in association with ritonavir is given as an example in Fig. 2, (Fig. 1 displays ritonavir in 'Balls and sticks' and fragmental representation). In order to show the advantage of the fragmental analysis if compared with the residue analysis, the protease is represented by Levitt [18] spheres with coded colours corresponding to the classical qualification of the residues: hydrophilic, hydrophobic or neutral,

- yellow = hydrophobic (A, C, T, V, M, P, Y, L, I, F),
- green = neutral (H, G, N, S, Q),
- red and blue = hydrophilic, respectively, charged + (K, R) and charged − (D, E)

and by fragments in association with ritonavir (darker colours) with the following coded colours:

- yellow = hydrophobic
- green = neutral
- others = hydrophilic: dark blue = COO−, red = NH_3^+, blue = hydrophilic, purple = CONH, white = $CONH_2$.

This representation points out some interesting observations that do not appear in the residue representation: a chain of hydrophilic fragments are located on the

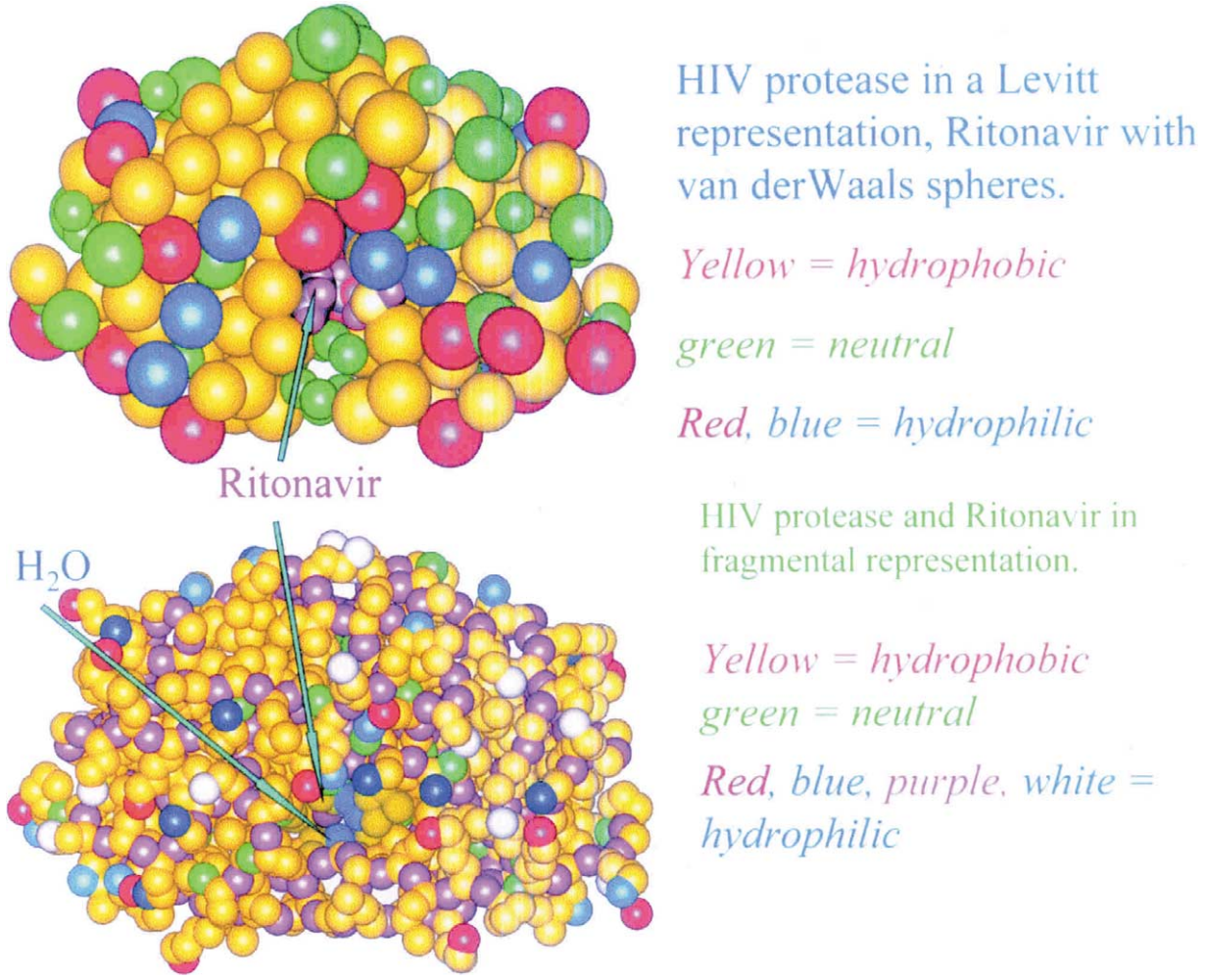

Fig. 2. HIV protease with ritonavir in 'Levitt' and 'fragmental' representations.

binary pseudo-symmetry of the molecule, above the active site, the central part of the inhibitor assumes the continuity of this chain. This representation demonstrates without calculation the necessity of one water molecule to fill the hydrophilic hole located at the bottom of the inhibitor. The other parts of the inhibitor are hydrophobic and are in contact with the hydrophobic pockets of the protease active site. This example shows how the fragmental representation may help the design of drug molecules.

7. Conclusion

By simply adding the fragments constituting a molecule, the fragmental model developed in this work makes it easy to obtain accurate hydration free energies of small molecules at ambient temperature (25 °C).

For mean size molecules such as drug ones, the hydration free energy is obtained in the same way, and with a good level of accuracy if compared with other methods, without the need of the knowledge of the molecule's geometry. Hydration free energy can be used to predict if a designed drug molecule is able to reach its biologic target.

For bigger molecules such as proteins, because of the lack of experimental data, it is not possible to test the model's validity, but the model makes it possible to obtain an approximate value of the hydration free energy of these molecules by considering only the fragments contacting water. However, the fragmental representation can be useful to analyse molecule features in order to rapidly define modifications improving their association with a biologic target, for example.

Programs to break down molecules into fragments, to calculate hydration free energies and to display molecules in the fragmental representation in an interactive way, are available on Silicon Graphic stations. They can be obtained by sending an e-mail to pepe@luminy.univ-mrs.fr.

References

[1] A. Velazquez-Campoy, I. Luque, M.J. Todd, M. Milutinovitch, Y. Kiso, E. Freire, Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor, Protein Sci. 9 (2000) 1801–1809.

[2] J. Tomasi, M. Persico, Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent, Chem. Rev. (1994) 2027–2094.

[3] C.J. Cramer, D.G. Truhlar, Continuum solvation models classical and quantum mechanical implementations, in: B.Lipkowitz Kenny, B.Boyd Donald (Eds.), Reviews in Computational Chemistry, vol. 6, VCH Publishers, New York, 1995, pp. 1–72.

[4] C.J. Cramer, D.G. Truhlar, in: Tapia, Bertràn (Eds.), Solvent Effects and Chemical Reactivity, Kluver Academic Publishers, Netherlands, 1996, pp. 1–80.

[5] J. Hine, P.K. Mookerjee, The intrinsic hydrophilic character of organic compounds. Correlations in terms of structural contributions, J. Organomet. Chem. 40 (3) (1975) 292–298.

[6] S. Cabani, P. Gianni, V. Mollica, L. Lepori, Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution, J. Solution Chem. 10 (1981) 563–595.

[7] A. Ben-Naim, Y. Marcus, Solvation thermodynamics of non-ionic solutes, J. Chem. Phys. 81 (1984) 2016–2027.

[8] R.G. Pearson, Ionization potentials and electron affinities in aqueous solution, J. Am. Chem. Soc. 81 (1986) 6109–6114.

[9] C.J. Cramer, D.G. Truhlar, PM3–SM3: a general parameterisation for including aqueous solvation effects in the PM3 molecular orbital model, J. Comput. Chem. 13 (9) (1992) 1089–1097.

[10] M.C. Connolly, Solvent-accessible surfaces of proteins and nucleic acids, Science 221 (1983) 709–713.

[11] G. Pèpe, D. Siri, GENMOL: a fast program for molecular modelling. Application to the determination of the psychotonic or sedative effects of tricyclic antidepressant drugs, in: J.L. Rivail (Ed.), Modelling of Molecular Structure and Properties, Elsevier, Amtersdam, 1990, pp. 93–101.

[12] F. Cavelier-Frontin, G. Pèpe, J. Verducci, D. Siri, R. Jacquier, Prediction of the best linear precursor in the synthesis of cyclotetrapeptides by molecular mechanics calculations, J Am. Chem. Soc. 114 (1992) 8885–8890.

[13] G. Pèpe, M. Meyer, P. Faury, J.C. Graciet, J.C. Chermann, J.L. Kraus, A model allowing the design of modified nucleosides as HIV-RT inhibitors, Eur. J. Med. Chem. 31 (1996) 775–786.

[14] C.H. Lin, N. Gabas, J.P. Canselier, G. Pèpe, Prediction of the growth morphology of amino acid crystals in solution I. oc-Glycine, J. Crystal Growth 191 (1998) 791–802.

[15] M. Ulmschneider, E. Pénigault, An approximate procedure for the calculation of van der Waals and solvent-accessible surface areas, computing free energies of hydration, J. Chem. Phys. 96 (1999) 566–590.

[16] E. Abola, F.C. Bernstein, S.H. Bryant, T.F. Koetzle, Weng, in: F.H. Allen, G. Bergerhoff, R. Sievers (Eds.), *J. Crystallographic* Databases—Information Content Software Systems. Scientific Applications, Data commission on the International Union of Crystallography, Bonn, 1987, pp. 107–132.

[17] A. Klampt, COSMO: conductor like screening model for real solvent, J. Phys. Chem. 102 (1998) 5074–5079.

[18] M. Levitt, A simplified representation of protein conformations for rapid simulations of protein folding, J. Mol. Biol. 104 (1976) 59–107.