Journal of Computer-Aided Molecular Design, **16:** 1–10, 2002.
KLUWER/ESCOM

3D hydrogen bond thermodynamics (HYBOT) potentials in molecular modelling

Oleg A. Raevsky[a,*] & Vladlen S. Skvortsov[b]
[a]*Department of Computer-Aided Molecular Design, Institute of Physiologically Active Compounds of the Russian Academy of Sciences, 142432, Chernogolovka, Moscow region, Russia;* [b]*Laboratory of Molecular Graphics and Drug Design, Institute of Biomedical Chemistry of the Russian Medicinal Academy, 119932, Pogodinskaya Str., 10, Moscow, Russia*

Received 14 November 2001; accepted 4 February 2002

Key words: distance function, enthalpy, H-bonding, molecular modelling, optimum energy, optimum distance

Summary

A new approach is proposed to more accurately estimate the energies of H-bond interactions in three-dimensional (3D) molecular modelling. The approach is based on the use of H-bond acceptor and donor enthalpy factor values calculated by means of program HYBOT, the use of a sigmoid relationship to determine the optimum H-bond distances and established force-field methods to determine distance and angle dependencies. The base-pair interactions in a short A-form RNA double-helix are presented as an example of enthalpy calculations of hydrogen bonding for a model system.

Introduction

Interactions between hydrogen-bond donor and acceptor groups in different molecules result in the formation of many kinds of molecular or ionic complexes that are of great importance in chemical processes, including enzymatic catalysis. Moreover, the significance of H-bonds is especially manifest in drug-receptor systems where they are important structural elements, and are a determining force in molecular recognition. However, these interactions are complex, and have been difficult to quantify.

A reliable way to describe H-bonding quantitatively is to consider the thermodynamics of H-bond complex formation. Starting in 1972, Raevsky et al. [1] developed such an approach by applying the multiplicative principle to the enthalpy (ΔH) and free energy (ΔG) of H-bonding in different polar and non-polar solvents by means of enthalpy acceptor factors (E_a), enthalpy donor factors (E_d), free energy acceptor factors (C_a) and free energy donor factors (C_d):

$$\Delta H = k_1(\text{kcal/mol})E_{a^*}E_d, \quad (1)$$

$$\Delta G = k_2(\text{kcal/mol})C_{a^*}C_d + k_3(\text{kcal/mol}). \quad (2)$$

This allowed the creation of uniform enthalpy and free energy H-bond scales [1–4]. Here, the above-mentioned H-bond factors can characterize the acceptor or donor strength of any atom in any molecule. Large data bases for these factor values as well as a program to estimate such values for atoms in any normal organic molecule are contained in the program HYBOT (Hydrogen Bond Thermodynamics) for Windows 95/98/2000/NT or for UNIX [5, 6]. Applications of these 2D factors for QSPR estimation of physical properties and QSAR estimation for biological activity have been described in a series of publications [6–12].

Most modern procedures for three-dimensional (3D) computer-assisted drug design use force-field calculations to take H-bonding into account. These calculations are used in many research areas including conformational analysis, pharmacophore identification, ligand docking, *de novo* ligand design, comparative molecular field analysis (CoMFA), and the identification of favorable binding sites from molecular interaction fields [13–19].

*To whom correspondence should be addressed: E-mail: raevsky@ipac.ac.ru

On the basis of a sophisticated and carefully parameterized method called GRID, Goodford et al. [20–24] made efforts to improve the calculation of the H-bond contribution to the interaction energy of ligand-macromolecular complexes. The GRID method is parameterized by fitting experimental X-ray data from protein and small-molecule crystals, and is designed to calculate the interactions of a probe (a small molecule such as water or ammonia) and a macromolecular system. The 8–6 function that was adopted was found to give satisfactory result [21]:

$$E_r = C/r^8 - D/r^6 \quad (3)$$

where $C = -3E_m r^8$ Å^8/mol; $D = -4E_m r_m^6$ Å^6/mol; r is the distance between the acceptor *atom* and the donor heavy atom; E_m is optimum H-bond enthalpy in kcal/mole; r_m is the optimum H-bond length. Some workers found that 12–10 and 6–4 functions were also useful for estimating the distance dependence of H-bond energy [25, 26].

Thus, since the beginning of the 1980s, two different directions to quantify the H-bond contribution in complex formation developed simultaneously in QSAR and Drug Design at the 2D and 3D levels. The carefully parameterized methodology of HYBOT allows one to take into account the influence of substituents on H-bond acceptor and donor strengths. Modern force-field procedures based on X-ray data from ligand-macromolecular complexes carefully consider distance and angle dependencies.

It is possible to consider the recently proposed hydrogen-bonding potentials (MHBP) [27] as the first step to unite 2D and 3D methods of description of H-bonding:

$$MHBP_k = \sum_{i=1}^{N} f_i \cdot f_{ct} \cdot (D_{ik}) \cdot f(U),$$

where k indicates a given point in space, i is a given molecular fragment, N is the total number of fragments in the molecule, f_i is α or β free energy solvatochromic parameters of H-bonding [28] of atom i, f(U) is angular function, f_{ct} is distance function, d_{ik} is the distance between fragments i and point k. However, the application of this approach is limited by the use of free energy solvatochromic H-bonding parameters and a fixed parameter of distance function for optimal H-bond potentials.

The present report is devoted to extending the estimation of H-bond potentials to 3D situations using H-bond HYBOT factors calculated on the basis of direct (and thus more realistic) enthalpy experimental data and distance functions for optimal energies estimated in a set of 'ideal' hydrogen-bonded complexes.

Results

The analysis of H-bonding thermodynamics data

The HYBOT program package contains a database of thermodynamic H-bonding parameters for 13,688 complexes including 5,984 of the type 'O-H...O', 3,039 of the type 'O-H...N', 1,016 of the type 'N-H...O' and 305 of the type 'N-H...N' [5]. It is obvious from these data that H-bonding enthalpies depend on the nature of substituents on the atoms that participate in new bond formation. The ranges of enthalpy values for the above-mentioned types are 0.9–15.9 kcal/mol for 'O-H...O'; 0.1–19.1 kcal/mol for 'O-H...N'; 0.5–10.5 kcal/mol for 'N-H...O', and 0.5–11.5 kcal for 'N-H...N'. Table 1 contains, as an example, the data for 60 small molecular complexes that typify functional groups of biochemical interest involved in H-bond formation. Here, in almost every case, the data came from one group of researchers [29]. Thus, the data are consistent, coming from a single experimental method with no laboratory-to-laboratory variation. The ranges of enthalpy values for this set of complexes are also fairly broad: 2.26–8.20 kcal/mol for 'O-H...O'; 3.00–11.50 kcal/mol for 'O-H...N', and 3.73–6.31 kcal/mol for 'N-H...N'.

It is possible to suppose that an optimum arrangement of partners in H-bonding is realized in cases where small molecules do not have any bulky substituents. Thus, one can presume that the above-mentioned enthalpy interval values correspond to the optimum energy (E_m) of H-bonding. And those values depend on the types of functional groups participating in H-bonding and the nature of their substituents. This means that, up to now, the full complexity of H-bonding has not been fully taken into account in molecular modelling. For example, there are only three fixed H-bond potentials in the GRID framework: −4.00 kcal/mol as the optimum H-bond energy for 'O-H...O'; −2.8 kcal/mol as optimum energy for 'O-H...N' and −2.0 kcal/mol as optimum energy for 'N-H...N' [21].

The HYBOT program package provides a more precise estimate of the optimum H-bonding energy by calculating it on the basis of H-bond acceptor and donor factors. In the current version, there are twelve databases of different H-bond factors in framework

Table 1. Enthalpies and H-bond acceptor and donor enthalpy factors for few complexes of small molecules.

NN	H-bond acceptor		H-bond donor		−Ea∗Ed	−ΔHexp
	Name	E_a [5]	Name	E_d [5]		(kcal/mol)
1	Pyridine	2.41	C_6H_5OH	−2.50	6.03	8.00 [29]
2	Pyridine	2.41	mF-C_6H_5OH	−2.73	6.58	8.40 [29]
3	Pyridine	2.41	mCF_3-C_6H_5OH	−2.62	6.31	8.50 [29]
4	Pyridine	2.41	$(CH_3)_3COH$	−1.37	3.30	4.30 [29]
5	Pyridine	2.41	$(CF_3)_2CHOH$	−3.02	7.28	9.80 [29]
6	Pyridine	2.41	H_2O	−1.44	3.47	5.10 [29]
7	Pyridine	2.41	NH_3	−1.36	3.28	3.73 [29]
8	NH_3	2.29	C_6H_5OH	−2.50	5.73	7.80 [29]
9	CH_3NH_2	2.09	C_6H_5OH	−2.50	5.23	8.60 [29]
10	$(CH_3)_2NH$	2.50	C_6H_5OH	−2.50	6.25	8.60 [29]
11	$(CH_3)_3N$	2.24	C_6H_5OH	−2.50	5.60	8.80 [29]
12	$C_2H_5NH_2$	2.72	C_6H_5OH	−2.50	6.80	8.60 [29]
13	$(C_2H_5)_3N$	2.70	C_6H_5OH	−2.50	6.75	9.00 [29]
14	$(C_2H_5)_3N$	2.70	$(CF_3)_2CHOH$	−3.02	8.15	11.50 [29]
15	$(C_2H_5)_3N$	2.70	pyrrole	−1.34	3.62	5.90
16	Imidazole	2.14	Imidazole	−2.00	4.28	6.31
17	CH_3CN	1.16	C_6H_5OH	−2.50	2.90	4.60 [29]
18	CH_3CN	1.16	mF-C_6H_5OH	−2.73	3.17	4.90 [29]
19	CH_3CN	1.16	mCF_3-C_6H_5OH	−2.62	3.04	4.90 [29]
20	CH_3CN	1.16	$(CF_3)_2CHOH$	−3.02	3.50	5.90 [29]
21	CH_3CN	1.16	H_2O	−1.44	1.68	3.00 [29]
22	$HCON(CH_3)_2$	2.20	C_6H_5OH	−2.50	5.50	6.10 [29]
23	$HCON(CH_3)_2$	2.20	mF-C_6H_5OH	−2.73	6.01	7.00 [29]
24	$HCON(CH_3)_2$	2.20	H_2O	−1.44	3.17	3.50 [29]
25	$CH_3CON(CH_3)_2$	2.39	C_6H_5OH	−2.50	5.98	6.70 [29]
26	$CH_3CON(CH_3)_2$	2.39	mCF_3-C_6H_5OH	−2.62	6.26	7.30 [29]
27	$CH_3CON(CH_3)_2$	2.39	$(CF_3)_2CHOH$	−3.02	7.22	8.20 [29]
28	$CH_3CON(CH_3)_2$	2.39	H_2O	−1.44	3.44	4.30 [29]
29	$CH_3COOC_2H_5$	1.54	C_6H_5OH	−2.50	3.85	4.80 [29]
30	$CH_3COOC_2H_5$	1.54	mF-C_6H_5OH	−2.73	4.20	5.10 [29]
31	$CH_3COOC_2H_5$	1.54	mCF_3-C_6H_5OH	−2.62	4.03	5.20 [29]
32	$CH_3COOC_2H_5$	1.54	$(CF_3)_2CHOH$	−3.02	4.65	5.90 [29]
33	CH_3COOCH_3	1.57	C_6H_5OH	−2.50	3.93	4.80 [29]
34	CH_3COOCH_3	1.57	mCF_3-C_6H_5OH	−2.62	4.11	5.00 [29]
35	CH_3COOCH_3	1.57	H_2O	−1.44	2.26	3.00 [29]
36	H_2CO	1.50	C_6H_5OH	−2.50	3.75	4.80 [29]
37	$(CH_3)_2CO$	1.57	C_6H_5OH	−2.50	3.93	5.10 [29]
38	$(CH_3)_2CO$	1.57	mCF_3-C_6H_5OH	−2.62	4.11	5.90 [29]
39	$(CH_3)_2CO$	1.57	$(CF_3)_2CHOH$	−3.02	4.74	6.70 [29]
40	$(CH_3)_2CO$	1.57	H_2O	−1.44	2.26	3.20 [29]
41	$(CH_3)_2O$	1.37	C_6H_5OH	−2.50	3.43	5.80 [29]
42	$(C_2H_5)_2O$	1.77	C_6H_5OH	−2.50	4.43	6.00 [29]
43	$(C_2H_5)_2O$	1.77	mCF_3-C_6H_5OH	−2.62	4.93	6.70 [29]
44	$(C_2H_5)_2O$	1.77	$(CF_3)_2CHOH$	−3.02	5.35	7.20 [29]
45	$(C_2H_5)_2O$	1.77	H_2O	−1.44	2.55	4.00 [29]
46	$((CH_3)_2CH)_2O$	1.88	C_6H_5OH	−2.50	4.70	6.20 [29]
47	$((CH_3)_2CH)_2O$	1.88	mCF_3-C_6H_5OH	−2.62	4.93	6.70 [29]
48	p-dioxane	1.66	C_6H_5OH	−2.50	4.15	5.60 [29]
49	p-dioxane	1.66	mCF_3-C_6H_5OH	−2.62	4.35	6.00 [29]
50	p-dioxane	1.66	H_2O	−1.44	2.39	3.20 [29]

Table 1. Continued.

NN	H-bond acceptor		H-bond donor		−Ea∗Ed	−ΔHexp
	Name	E_a [5]	Name	E_d [5]		(kcal/mol)
51	Tetrahydrofurane	1.90	C_6H_5OH	−2.50	4.75	6.00 [29]
52	Tetrahydrofurane	1.90	mCF_3-C_6H_5OH	−2.62	4.98	6.50 [29]
53	Tetrahydrofurane	1.90	$(CF_3)_2CHOH$	−3.02	5.74	6.90 [29]
54	$(CH_2)_5O$	1.75	C_6H_5OH	−2.50	4.38	6.10 [29]
55	$(CH_2)_5O$	1.75	mCF_3-C_6H_5OH	−2.62	4.59	6.50 [29]
56	$(CH_3)_2S$	1.62	C_6H_5OH	−2.50	4.05	4.60 [29]
57	$(CH_3)_2S$	1.62	mCF_3-C_6H_5OH	−2.62	4.24	5.40 [29]
58	$(C_2H_5)_2S$	1.33	C_6H_5OH	−2.50	3.33	4.60 [29]
59	$(C_2H_5)_2S$	1.33	mF-C_6H_5OH	−2.73	4.42	5.20 [29]
60	$(C_2H_5)_2S$	1.33	mCF_3-C_6H_5OH	−2.62	3.48	5.40 [29]

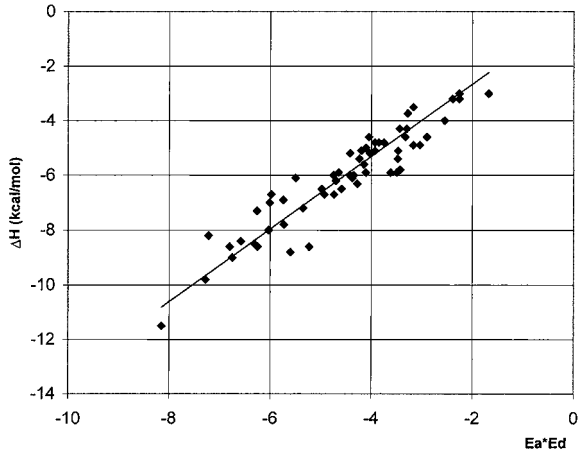

Figure 1. Graphic comparison between experimental hydrogen bond enthalpy values and calculated factor products.

of four H-bond scales (enthalpy, free energy, binding constant and 'overall' scales). Factor calculations are carried out by means of a procedure that searches for the nearest neighbor in a selected factor data base(s) [6]. Enthalpy H-bond scale and corresponding factors were used for this study. The efficiency of enthalpy values prediction by means of HYBOT program calculations of hydrogen bond enthalpy factors can be demonstrated for the 60 H-bond complexes listed in Table 1 [29]. The following regression equation (forced through the origin) was obtained:

$$\Delta H = E_m = 1.33(\pm 0.02)(\text{kcal/mol})E_{a*}E_d,$$
$$n = 60, \quad R = 0.930, \quad s = 0.65. \qquad (4)$$

A graphic comparison of H-bond donor and acceptor factor values products ($E_a * E_d$) with experimental enthalpy values for this set of 60 H-bond complexes is presented in Figure 1.

Nevertheless, despite the good H-bond enthalpy prediction ability, this thermodynamic approach is not enough for a complete description of H-bond complexes because no geometric and distance parameters of hydrogen bonding are considered.

X-ray and thermodynamics data analysis of 'ideal' H-bond complexes

There are only three fixed optimum H-bond lengths in the GRID framework: 2.8 Å for 'O...O', 3.0 Å for 'N...O' and 3.2 Å for 'N...N' [21]. The wide intervals of enthalpy values for each of these types of H-bonding complexes allow us to infer the existence of a dependence between energies and distances even in the case of an optimum arrangement of the atoms that participate in the H-bond. Hence, to create a realistic platform for the quantitative description of H-bonding, one must recognize that optimum H-bond en-

Table 2. X-ray data, H-bond parameters, optimum energies and distances for 58 'ideal' complexes.

Compound[a]	D_{ad}[b]	D_{ah}[c]	t (H)[d]	p (LP)[e]	E_a	E_d	Acceptor	Donor	$E_a * E_d$	E_m	r_m
CAMALH	2.819	1.814	179.4	173.44	1.50	−1.23	O	O	−1.84	−2.43	2.844
CUKJAT	2.831	1.83	177.41	175.73	1.38	−0.88	O	O	−1.21	−1.60	2.965
DFMLON	2.704	1.855	175.29	173.5	1.24	−2.93	O	O	−3.65	−4.82	2.618
DLASPA02	2.542	1.508	179.42	177.97	1.50	−2.63	O	O	−3.93	−5.19	2.590
GOSPOL10	2.643	1.864	175.41	177.25	1.18	−2.54	O	O	−2.99	−3.95	2.689
KAPVUS	2.688	2.024	175.55	177.73	1.50	−2.11	O	O	−3.16	−4.17	2.670
NEWXAO	2.649	1.745	178.53	174.45	1.50	−2.67	O	O	−3.99	−5.27	2.584
SIGBEP	2.667	1.719	176.7	175.13	1.24	−2.63	O	O	−3.26	−4.30	2.660
SIGBEP	2.667	1.719	176.7	175.12	1.24	−2.63	O	O	−3.26	−4.30	2.660
SIGBEP	2.667	1.719	176.71	175.12	1.24	−2.63	O	O	−3.26	−4.30	2.660
WAZLIS	2.934	1.826	178.42	173.51	1.44	−0.88	O	O	−1.27	−1.67	2.953
ZOTSIK	2.61	1.699	175.17	173	1.24	−2.93	O	O	−3.65	−4.82	2.618
ZZZPZE01	2.674	1.737	176.09	178.49	1.24	−2.63	O	O	−3.26	−4.30	2.660
BAFWIO	2.926	2.068	175.06	175.09	2.06	−1.29	O	N	−2.66	−3.51	2.995
BAFWIO	2.915	2.057	176.37	175.8	2.06	−1.29	O	N	−2.66	−3.51	2.995
BEQVUO	2.62	1.782	175.47	175.78	2.07	−2.74	N	O	−5.67	−7.49	2.578
CXMESX	2.938	1.859	177.7	175.03	1.78	−1.29	O	N	−2.30	−3.03	3.055
DEPGAG	2.897	1.951	178.25	176.93	1.78	−1.91	O	N	−3.40	−4.49	2.884
ICRFRD10	2.946	2.049	179.11	177.42	1.78	−1.83	O	N	−3.27	−4.31	2.903
JICVOG20	2.939	1.957	176.44	177.82	1.78	−1.29	O	N	−2.30	−3.03	3.055
JICWIB10	3.045	2.146	177.05	175.54	2.06	−1.29	O	N	−2.66	−3.51	2.995
KEMWUU	2.936	1.917	176.43	175.69	2.06	−1.29	O	N	−2.66	−3.51	2.995
KEMWUU	2.936	1.917	176.43	175.68	2.06	−1.29	O	N	−2.66	−3.51	2.995
KEMWUU	2.936	1.917	176.43	176.76	2.06	−1.29	O	N	−2.66	−3.51	2.995
NEDDOP	2.802	1.984	175.11	177.37	1.99	−2.11	N	O	−4.19	−5.54	2.776
NEFLUF	2.662	1.683	179.56	178.54	2.19	−2.94	N	O	−6.42	−8.47	2.468
PIPNUX	2.836	1.974	179.37	175.36	2.06	−1.83	O	N	−3.78	−4.99	2.832
PIYHEK	3.087	2.189	176.43	176	2.06	−1.83	O	N	−3.78	−4.99	2.832
REGKIX	2.941	2.074	177.32	174.46	1.84	−1.29	O	N	−2.37	−3.12	3.043
RIWJOW	2.94	2.021	177.68	176.12	2.06	−1.29	O	N	−2.66	−3.51	2.995
SIKXEP	2.878	1.934	179.57	176.78	1.78	−1.83	O	N	−3.26	−4.30	2.905
VOZVEL	3.257	2.379	177.15	175.45	1.51	−1.83	O	N	−2.78	−3.67	3.089
AMCLPY	3.058	2.246	178.86	170.29	1.99	−1.29	N	N	−2.57	−3.39	3.073
AMNTPY	3.009	2.044	175.92	173.46	1.99	−1.29	N	N	−2.57	−3.39	3.073
AZGUCM10	2.987	2.081	176.58	176.94	1.50	−1.29	N	N	−1.93	−2.55	3.238
BARIMZ10	2.781	1.759	175.71	177.46	2.14	−1.83	N	N	−3.92	−5.17	2.772
BEYHIW	2.954	1.988	177.59	172.88	1.19	−1.83	N	N	−2.19	−2.88	3.170
CIDWUH	3.037	2.164	173.92	170.3	1.99	−1.29	N	N	−2.57	−3.39	3.073
FUWVAU	3.04	2.061	172.28	174.22	1.99	−1.29	N	N	−2.57	−3.39	3.073
GIFZAW01	3.059	2.008	178.57	175.94	1.99	−1.29	N	N	−2.57	−3.39	3.073
GOJPEA	3.063	2.139	176.39	173.31	1.19	−1.83	N	N	−2.19	−2.88	3.170
GOMSIK	2.956	2.072	176.15	171.96	1.86	−1.83	N	N	−3.42	−4.51	2.880
HIBKAE	2.927	1.944	170.42	170.69	1.99	−1.83	N	N	−3.65	−4.82	2.829
HIBZIB	2.976	2.152	174.53	178.27	1.50	−1.83	N	N	−2.75	−3.62	3.031
JICVOG20	2.923	2.104	171.41	172.96	1.99	−1.83	N	N	−3.65	−4.82	2.829
JICWIB10	2.917	2.017	177.79	173.37	1.99	−1.83	N	N	−3.65	−4.82	2.829
KOKPOP	3.268	2.302	175.33	177.16	1.99	−1.29	N	N	−2.57	−3.39	3.073
NOVZON	2.771	1.824	177.75	173.12	1.99	−1.83	N	N	−3.65	−4.82	2.829
PIYDEG	2.914	2.014	177.05	171.25	1.99	−1.83	N	N	−3.65	−4.82	2.829
PYOCHP	2.903	1.986	173.27	170.41	1.99	−1.83	N	N	−3.66	−4.83	2.828

Table 2. Continued.

Compound[a]	D_{ad}[b]	D_{ah}[c]	t (H)[d]	p (LP)[e]	E_a	E_d	Acceptor	Donor	E_a*E_d	E_m	r_m
REGKIX	2.961	2.117	176.58	170.69	1.99	−1.83	N	N	−3.66	−4.83	2.828
REGKIX	2.96	2.117	176.58	170.71	1.99	−1.83	N	N	−3.66	−4.83	2.828
REMTUY	3.087	2.232	176.69	170.98	1.99	−1.29	N	N	−2.57	−3.39	3.073
RIWJIQ	2.896	2.084	173.98	176.16	1.99	−1.83	N	N	−3.66	−4.83	2.828
TUHJAH	3.006	2.084	174.3	172.76	1.99	−1.29	N	N	−2.57	−3.39	3.073
VABLUF	2.992	2.026	176.31	174.32	1.99	−1.87	N	N	−3.73	−4.92	2.813
WICWUA	3.051	2.078	179.19	175.41	1.99	−1.29	N	N	−2.57	−3.39	3.073
ZUFDEJ	2.745	1.902	178.31	173.48	2.17	−1.87	N	N	−4.06	−5.36	2.741

[a]Compound codes correspond to [31].
[b]D_{ad} is distance between donor heavy and acceptor atoms.
[c]D_{ah} is distance between acceptor atom and hydrogen.
[d,e]t (H) and p (LP) are angles in accordance with [21]; E_m values were calculated on the basis of Equation 4; r_m values were calculated on the basis of Equation 6 with next coefficient values: $k_4 = -15$, $k_7 = 0.58$, $k_8 = 2.43$ for 'OH...O' complexes; $k_4 = -12$, $k_7 = 0.69$, $k_8 = 2.73$ for 'OH...N' complexes; $k_4 = -10$, $k_7 = 0.94$, $k_8 = 2.80$ for 'NH...N' complexes.

Table 3. Statistical parameters equation $E_m = k_4/(1 + 10^{k_5+k_6 R_m})$ and $R_m = k_7 \log[(k_4 - E_m)/(E_m)] + k_8$.

	O...O			O...N			N...N		
k_4	−20.0	−15.0	−12.0	−20.0	−15.0	−12.0	−15.0	−12.0	−10.0
k_5	3.60(±0.30)	4.20(±0.81)	5.62(±0.95)	2.93(±0.64)	3.44(±0.77)	3.96(±0.94)	2.09(±0.51)	2.50(±0.57)	2.97(±0.65)
k_6	1.57(±0.28)	1.73(±0.30)	2.16(±0.36)	1.20(±0.22)	1.32(±0.27)	1.45(±0.32)	0.85(±0.17)	0.94(±0.19)	1.06(±0.22)
N[a]	13	13	13	19	19	19	26	26	26
R[b]	0.892	0.897	0.908	0.769	0.756	0.739	0.712	0.712	0.711
S[c]	0.58	0.57	0.54	0.98	1.00	1.03	0.61	0.62	0.62
F[d]	43.1	45.4	51.7	24.7	22.7	20.4	24.7	24.7	24.5
k_7	0.64	0.58	0.52	0.83	0.76	0.69	1.18	1.06	0.94
k_8	2.29	2.43	2.53	2.44	2.61	2.73	2.46	2.66	2.80

[a]N is number of complexes.
[b]R is correlation coefficient.
[c]S is standard deviation.
[d]F is Fisher criterium.

Table 4. Optimum distances 'X...Y' as the function of optimum H-bonding enthalpy.

$-E_m$ (kcal/mol)	R_m (Å)								
	O...O			**O...N**			**N...N**		
	$k_4 = -20.0$	$k_4 = -15.0$	$k_4 = -12.0$	$k_4 = -20.0$	$k_4 = -15.0$	$k_4 = -12.0$	$k_4 = -15.0$	$k_4 = -12.0$	$k_4 = -10.0$
0.25	3.50	3.46	3.40	4.02	3.96	3.88	4.55	4.43	4.30
0.50	3.31	3.28	3.24	3.76	3.72	3.67	4.19	4.10	4.00
1.00	3.11	3.09	3.07	3.50	3.48	3.45	3.81	3.76	3.70
2.00	2.90	2.90	2.89	3.23	3.23	3.21	3.42	3.40	3.37
3.00	2.77	2.78	2.78	3.07	3.07	3.06	3.17	3.17	3.15
4.00	2.68	2.68	2.69	2.94	2.94	2.94	2.98	2.98	2.97
5.00	2.60	2.60	2.61	2.84	2.84	2.83	2.82	2.81	2.80
6.00	2.53	2.53	2.53	2.75	2.74	2.73	2.67	2.66	2.63
7.00	2.46	2.46	2.45	2.66	2.65	2.63	2.53	2.51	2.45
8.00	2.40	2.40	2.37	2.59	2.57	2.52	2.39	2.34	2.23
9.00	2.35	2.33	2.28	2.51	2.48	2.40	2.25	2.15	1.90
10.00	2.29	2.26	2.17	2.44	2.38	2.25	–	–	–

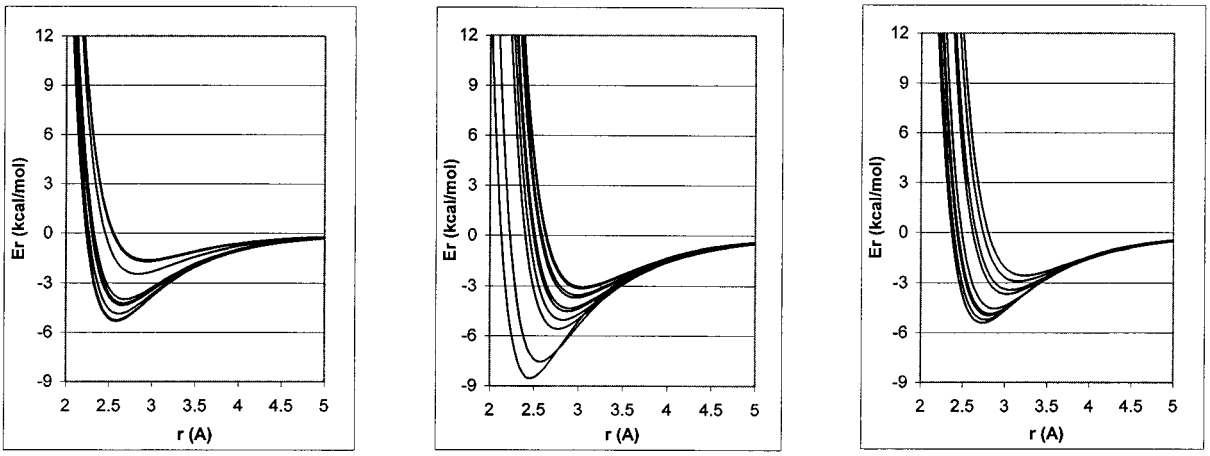

Figure 2. H-bond potentials for 'OH. . . O' (1–13 in Table 2), 'O. . . N' (13–32) and 'NH. . . N' (33–58) complexes.

ergies (E_m) depend on optimum H-bond distances (r_m) for the different types of H-bonds found in different complexes.

At least one attempt has been made to estimate the relationship between X-ray data of H-bonding complexes and thermodynamic parameters [30] by extensively surveying the Cambridge Crystal Structural Database. Group probability values were then calculated by counting the number of H-bonds involving the group divided by the number of times the group appears in the database for which there are also complementary groups present in a crystal. The group-probability values from this crystal structure survey and the equilibrium constants from the solute studies generally ranked the H-bond groups similarly. There are at least two reasons why a more precise relationship between the X-ray data and the thermodynamic parameters was not found. First, it may be that the complexes studied had inconsistent (including non-optimal) configurations of H-bond partners. Second, ranges instead of exact values were used for the equilibrium constants of the functional groups in specific H-bonding complexes.

In the present study we selected specific complexes to assess the relationship between the length of H-bonds and their energies. Optimum H-bonding conditions occur when there is a linear arrangement of the donor *heavy* atom, hydrogen and acceptor atoms and of the acceptor nucleus, electron pair and hydrogen atom. Thus, we selected from the Cambridge Structural Database [31] 58 'ideal' H-bonding complexes in which the angles of the donor *heavy* atom, hydrogen and acceptor atoms were in the range of 173° to 187° and angles of the acceptor nucleus, the lone pair of electrons and the hydrogen atom were in the range of 170° to 190°. There were 13 such 'O-H. . . O' complexes, 19 'O-H. . . N' or 'N-H. . . O' complexes, and 26 'N-H. . . N' complexes among the 58 'ideal' representatives.

Because there are significant differences between the covalent and van der Waals radii of oxygen and nitrogen atoms, we decided to look for relationships between optimum energy (E_m) and optimum distances (r_m) in those three subsets. Such relationships are not expected to be linear because energy values approach zero as distances increase. In this study we used a sigmoid function:

$$E_m = k_4/(1 + 10^{k_5 + k_6 r_m}). \quad (5)$$

Values of k_4 in Equation (5) may be fixed, and are limited by the maximum values for the enthalpies. Table 3 contains values for k_5 and k_6, and the statistical parameters derived from the correlation of E_m with r_m for selected values of k_4. The correlations range from reasonably good for 'O-H. . . O' systems to unimpressive for 'N-H. . . N' systems. The reason for this modest correlation (low R's) may be related to the sloping shape of the energy minima. As a result a rather broad distance range can be realized for an optimum H-bonding energy value.

Rearranging Equation (5) and simplifying the constants gives Equation (6):

$$r_m = k_7 \log[(k_4 - E_m)/(E_m)] + k_8. \quad (6)$$

Using Equation (6), we can now calculate r_m from E_m values. For these correlations, the coefficient values k_7 and k_8 are also presented in Table 3. Here,

Table 5. The calculated energy values for bases interactions in double helix of RNA in A form.

NN	Donor	C_d	Acceptor	C_a	$E_m$calc. (thiswork) (kcal/M)	E_m calc.(as in [21]) (kcal/mol)	R_m calc. (this work)(A)	R_m (as in [21])(A)	R (A)	E_r (this work) (kcal/mol)	E_r (as in [21] (kcal/mol)	E_{hb}[a] (thiswork) kcal/mol)	E_{hb}[a] (as in [21]) (kcal/mol)
1	N	−1.29	O	1.78	−3.06	−2.80	3.057	3.00	2.833	−2.45	−2.51	−2.18	−2.23
2	N	−1.83	N	1.99	−4.84	−2.00	2.840	3.20	2.953	−4.69	−1.53	−4.64	−1.53
3	N	−1.29	O	1.78	−3.06	−2.80	3.057	3.00	2.724	−1.35	−1.80	−1.20	−1.61
4	N	−1.83	N	1.19	−2.90	−2.00	3.186	3.20	2.909	−2.01	−1.31	−1.96	−1.27
5	N	−1.29	O	2.06	−3.53	−2.80	2.995	3.00	3.078	−3.48	−2.76	−2.81	−2.23
6	N	−1.29	O	2.06	−3.53	−2.80	2.995	3.00	3.078	−3.48	−2.76	−2.81	−2.23
7	N	−1.83	N	1.19	−2.90	−2.00	3.186	3.20	2.909	−2.01	−1.31	−1.96	−1.27
8	N	−1.29	O	1.78	−3.06	−2.80	3.057	3.00	2.724	−1.35	−1.80	−1.20	−1.61
9	N	−1.29	O	1.78	−3.06	−2.80	3.057	3.00	2.833	−2.45	−2.51	−2.18	−2.23
10	N	−1.83	N	1.99	−4.84	−2.00	2.840	3.20	2.954	−4.69	−1.53	−4.63	−1.53
										$\Sigma E_r = -27.96$	$\Sigma E_r = -19.82$	$\Sigma E_{hb} = -25.55$	$\Sigma E_{hb} = -17.73$

[a]E_{hb} is hydrogen bond energy calculated with account of angle dependence of hydrogen bond atoms arrangement [21].

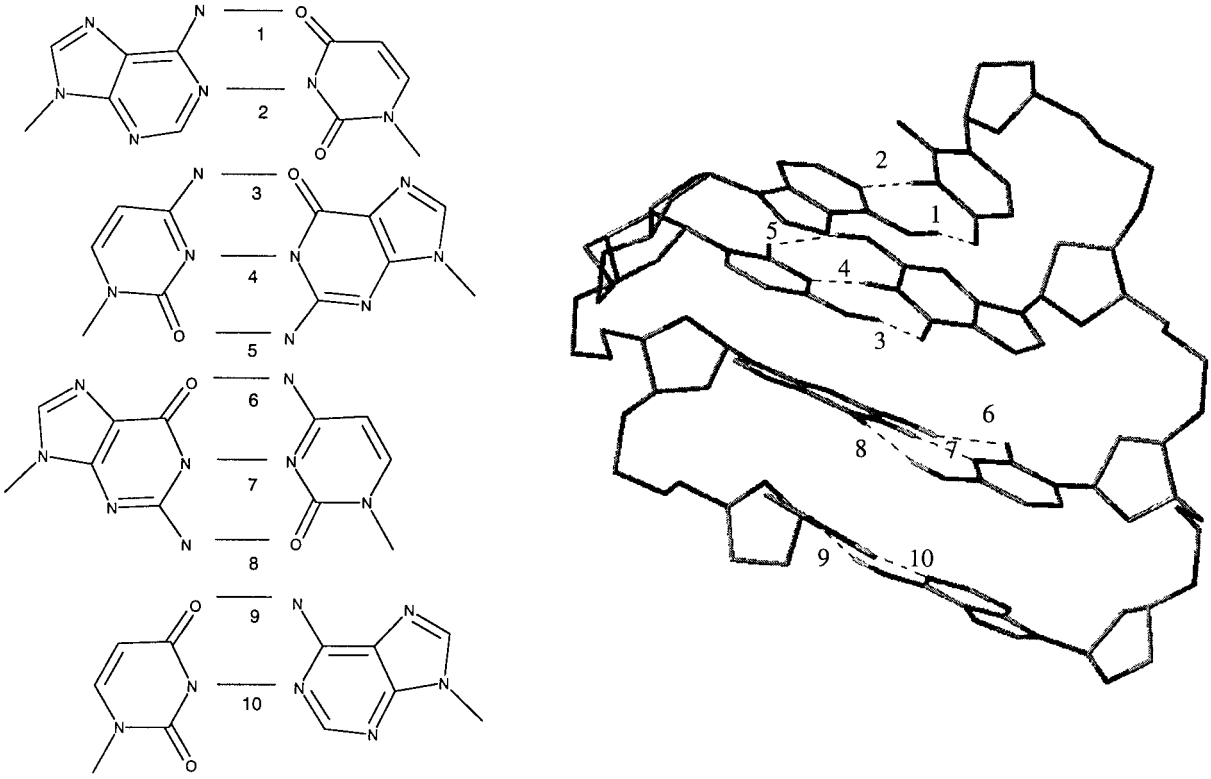

Figure 3. Set of hydrogen bonds found in a short fragment of A-form RNA double helix.

r_m is the optimum distance between acceptor atom and the donor heavy atoms for the corresponding optimum energy value. Under the scenarios of 'O-H. . . O', 'O. . . H. . . N', and 'N-H. . . N', optimum distance values (r_m) for selected optimum energy values (E_m) at selected k_4 values are presented in Table 4. Regardless of the H-bond type or the k_4 value selected, r_m always decreases with increasing E_m. At a particular E_m, r_m for the 'O-H. . . O' bond type is always smaller than that of the corresponding 'O. . . H. . . N' type. At low values for E_m, the r_m for the 'O. . . H. . . N' bond type is smaller than that of the corresponding 'N-H. . . N' type. However, as E_m rises in value, this latter difference diminishes, and eventually r_m of the 'N-H. . . N' bond type becomes smaller than that of the corresponding 'O. . . H. . . N' type. *Thus, for each specific pair of atoms participating in an H-bond, its H-bonding potential can be calculated on the basis of Equations 3, 4 and 6*. Examples of potentials for 'O-H. . . O', 'O-H. . . N' ('N-H. . . O') and 'N-H. . . N' for concrete complexes of small molecules are presented in Figure 2 and in Tables 2 and 4.

Hence, a scheme to calculate the H-bonding energies of atoms in a three-dimensional arrangement can be carried out by the following steps:

(1) Use HYBOT's acceptor and donor enthalpy factors to estimate the optimum H-bonding energy by means of Equation 4.

(2) Estimate the optimum H-bonding distance corresponding to the calculated energy optimum by means of Equation 6.

(3) Use Equation 3 or other functions to take into account the distance and angle deviations of the atoms involved.

An example of such calculation for a model system is presented in Figure 3 and Table 5. This represents the base-pair interactions in a short A-form RNA double helix. There are significant differences between the energies calculated by this approach and those calculated in accordance with [21]. The largest difference is in the case of 'N-H. . . N' (bond 2 on Figure 3). The experimental H-bond energy value in a similar complex in the HYBOT database, C5H5N. . . H-N(COOEt)C6H5, is equal to −5.09 kcal/mol, and is in complete accord with the calculated value (−4.84 kcal/mol).

These calculations were carried out by means of 3-D HYBOT in UNIX, a new program, by considering only H-bonding interactions and Sybyl [32]. We believe it is possible that this approach to H-bonding energy calculations will not be too difficult to integrate with existing and future force-field schemes to produce a superior method to estimate intermolecular interactions in any three dimensional framework.

Conclusion

A new approach has been proposed to more accurately estimate the optimum energy of H-bonding interactions in molecular modelling. The approach is based on the use of H-bond acceptor and donor enthalpy factor values calculated by means of program HYBOT, and the use of a sigmoid relationship between optimal H-bond distances and optimum energies.

Acnowledgements

The authors highly appreciate valuable comments given by Dr J.W. McFarland and Dr A.S. Ivanov during the preparation of this paper. The study was partly supported by Russian Ministry of Sciences (grant No. 509-32(00)-P) and by Russian Foundation of Fundamental Research (grant No. 01-04-49689).

References

1. Raevsky, O.A. and Novikov, V.V., Khim-Pharm. Z., 16 (1982) 583 (in Russian). Chem. Abst., 97 (1982) 33128d.
2. Raevsky, O.A., Avidon V.V. and Novikov, V.P., Khim-Pharm. Z., 16 (1982) 968 (in Russian). Chem. Abst., 97 (1982) 181323d.
3. Raevsky, O.A., In Hadzi, D. and Jerman Blazic, B. (eds), QSAR in Drug Design and Toxicology, Elsevier, Amsterdam, 1987, pp. 31–36.
4. Raevsky, O.A., Grigor'ev, V.Ju., Kireev, D.B. and Zefirov, N.S., Quant. Struct.-Act. Relat., 11 (1992) 49.
5. a. Raevsky, O.A., Grigor'ev, V.Ju., Trepalin, S.V., HYBOT program package, Registration by Russian State Patent Agency No. 990090 of 26.02.99.
   b. HYBOT program package for UNIX or Windows 95/98/2000/NT. Molpro Project, Chernogolovka, Moscow region, Russia (raevsky@ipac.ac.ru). In USA and Canada Reckon.dat.consulting, Lyme, CT (reckon.dat@attglobal.net)
6. Raevsky, O.A., In van de Waterbeemd, Testa, B. and Folkers, G. (eds), Computer-Assisted Lead Finding and Optimization, Willey-VCH, Weinheim, 1997, pp. 369–378.
7. Raevsky, O.A., J. Phys. Org. Chem., 10 (1997) 404.
8. Raevsky, O.A., Schaper, K.-J. and Seydel, J.K., Quant. Struct.-Act. Relat., 14 (1995) 433.
9. Raevsky, O.A. and Schaper, K.-J., Eur. J. Med. Chem., 33 (1998) 799.
10. Raevsky, O.A., Schaper, K.-J., Waterbeemd, H. and McFarland J.W., In Guntertofte, K. and Jorgensen, F.S. (eds), Kluwer Academic/Plenum Publishers, New York, NY, 2000, pp. 221–227.

11. Raevsky, O.A., Fetisov, V.I., Trepalina, E.P., McFarland J.W. and Schaper, K.-J., Quant. Struct.-Act. Relat., 19 (2000) 366.
12. Raevsky, O.A., SAR & QSAR Environ. Res., 12 (2001), 367.
13. Siebel, G.L. and Kollmann, P.A., In Hanch, C., Sammes, P.G., Taylor, J.B. and Ramsden, C.A. (eds), Comprehensive Medicinal chemistry, Vol. 4, Pergamon Press, Oxford, 1990, pp. 125–138.
14. Liljefors, T., In Kubinyi, H., Folkers, G. and Martin, Y.C. (eds), 3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity, Kluwer/Escom, Dordrecht, 1998, pp. 3–17.
15. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
16. Rognan, D., Scopozza, L., Folkers, G. and Daser, A., Biochemistry, 33 (1994) 11476.
17. Taylor, N.R. and von Itzstein, M., J. Med. Chem., 34 (1994) 616.
18. Klebe, G., J. Mol. Biol., 237 (1994) 212.
19. Oprea, T.I., Marshall, G.R., In Kubinyi, H., Folkers, G. and Martin, Y.C. (eds), 3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity, Kluwer/Escom, Dordrecht, 1998, pp. 35–61.
20. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
21. Boobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M. and Wade, R.C., J. Med. Chem., 32 (1989) 1083.
22. Wade, R.C., Clark, K. and Goodford, P.J., J. Med. Chem., 36 (1993) 140.
23. Wade, R.C., Clark, K. and Goodford, P.J., J. Med. Chem., 36 (1993) 148.
24. Goodford, P.J., J. Chemometrics, 10 (1996) 107.
25. Vedani, A. and Dunitz, J.D., J. Am. Chem. Soc., 107 (1985) 7653.
26. Nilsson, L. and Karplus, M., J. Comput. Chem., 7 (1986) 591.
27. Caron, G., Rey, S., Ermondi, G., Crivori, P., Guillard, P., Carrupt, P.-A., and Testa, B., In Testa, B., van de Waterbeemd, H., Folkers, G. and Guy, R. (eds), Pharmokinetic Optimization in Drug Research, Weinheim, Willey-VCH, 2001, pp. 513–524.
28. Abraham, M.H., Chem. Soc. Rev., 22(1993) 73.
29. Kroeger, M.K. and Drago, R.S., J. Am. Chem. Soc., 103 (1981), 3250.
30. Mills, J.E.J and Dean, P.M., J. Comput. Aid. Mol. Des., 10 (1996) 607.
31. Cambridge Structural Database, version 5.20, October 2000.
32. Sybyl 6.7, Tripos Inc., 1699 South Hanley Road, St Louis, Missouri, 63144, USA.