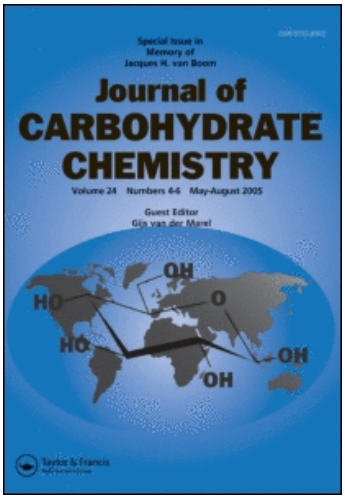

Journal of Carbohydrate Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713617200

The Consistent Force Field. 5. PEF95SAC: Optimized Potential Energy Function for Alcohols and Carbohydrates

Jesper Fabricius[a]; Soren B. Engelsen[b]; Kjeld Rasmussen[a]
[a] Department of Chemistry, The Technical University of Denmark, Lyngby, Denmark [b] Department of Dairy and Food Science, Food Technology, The Royal Veterinary and Agricultural University, Frederiksberg C, Denmark

To cite this Article Fabricius, Jesper , Engelsen, Soren B. and Rasmussen, Kjeld(1997) 'The Consistent Force Field. 5. PEF95SAC: Optimized Potential Energy Function for Alcohols and Carbohydrates', Journal of Carbohydrate Chemistry, 16: 6, 751 — 772
To link to this Article: DOI: 10.1080/07328309708006539
URL: http://dx.doi.org/10.1080/07328309708006539

J. CARBOHYDRATE CHEMISTRY, 16(6), 751-772 (1997)

THE CONSISTENT FORCE FIELD. 5. PEF95SAC: OPTIMIZED POTENTIAL ENERGY FUNCTION FOR ALCOHOLS AND CARBOHYDRATES

Jesper Fabricius,[a] Søren B. Engelsen[b] and Kjeld Rasmussen[a*]

[a]The Technical University of Denmark, Department of Chemistry, Building 207, DK-2800 Lyngby, Denmark
[b]The Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Technology, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

Received June 3, 1996 - Final Form March 18, 1997

ABSTRACT

A new empirical molecular mechanics force field, PEF95SAC, has been developed for model alcohols and carbohydrates using the optimization program CFF (The Consistent Force Field). It can be transferred to most scientific and commercial molecular mechanics programs. From a total set of 142 parameters, 110 were previously optimized on experimental data for hydrocarbons and ethers, and 32 related to alcohols and carbohydrates are optimized in this work. Data included in the optimization database encompass molecular structures in gaseous and crystalline phases, unit cells, and vibrational spectra. Other categories of data such as equilibrium constants are used for comparison after optimization. Calculations on the crystalline phase utilize convergent lattice summation. Bond deformation is described with Morse functions. Resulting calculated properties are of such quality that PEF95SAC should be useful in a broader context, and be applicable to gases as well as solids. The *ab initio* charges used in the force field are quite similar to those used by most established water potentials so that also realistic condensed phase molecular dynamics simulations are also within the scope.

INTRODUCTION

In this work we report the development of a molecular mechanics force field for application to alcohols and carbohydrates. The new force field can be implemented in all scientific and commercial molecular mechanics programs having a minimum degree of flexibility in the choices of potential energy functions and a clear separation between the program and the potential energy functions and their parameters. Fortunately we also see other molecular mechanics program developers include facilities such as the open force field approach and force field editors.

The molecular mechanics program used in this work, CFF,[1-3] can be regarded as an advanced force field editor.[4] It can optimize a molecular mechanics force field to reproduce a set of experimental data using a least-squares approach. This requires the relationship between conformation and the experimental property to be well understood. Experimental data used in the development of PEF95SAC include vibrational frequencies, unit cell parameters, and crystalline and gaseous molecular structures. In this empirical approach the experimental data are subjected to a mathematical model which derives structure and other properties from a potential energy function and then seeks to optimize the parameters of the potential energy function with respect to a selected set of experimental data. In general, high-precision experimental data on conformational properties are much too scarce to make use of model-free multivariate approaches such as partial least squares regression or neural networks. Modelling is further confounded by the multiple minima problem.

Developing molecular mechanics force fields for alcohols and carbohydrates is very troublesome due to the presence of the flexible and hydrogen-bonding hydroxyl groups. As a result of the many possible orientations of these groups, and because hydrogen-bonding can be satisfied in many different associations, the multiple minima problem is always a dominant difficulty. Time averaged spectral gas-phase data will have to be calculated from complete Boltzman distributed ensembles. In this work we have chosen to work with explicit environment variables (crystal lattices) for the optimization of data measured in condensed phase and with only stable group frequencies for optimization of alcohols in gas phase. Another prominent complexity of alcohol and carbohydrate structures is the assortment of energetic, structural and kinetic effects at the anomeric center, recently reviewed by Tvaroska and Bleha.[5] In this work we have dealt with the anomeric problem by introducing a special

type of atom, the anomeric carbon atom, and parametrized it separately. We have previously used this approach with good results for aliphatic ethers.[6]

Because of the difficulties mentioned, which are typical for carbohydrates and because of the often poor performance of the general force fields bundled with the most popular molecular mechanics programs, a series of refined carbohydrate force fields have been published in the recent years. For MM2[7,8] a refined parameter set was developed by Tvaroska and Pérez.[9] For CHARMM[10] a refined carbohydrate force field was published by Brady and coworkers[11] and later, using the extended atom approach for the hydroxyl groups, by Grootenhuis and Haasnoot[12a] and, very recently, the force field CHEAT95 was published by Kouwijzer and Grootenhuis.[12b] For AMBER[13] three refined force fields have been published by Homans,[14] Glennon *et al.*[15] and Woods *et al.*[16] For SYBYL™ (Tripos Associates) an ongoing carbohydrate refinement is carried out by Pérez and coworkers.[17,18] For DISCOVER™ (Biosym Technologies Inc.) Sarko and coworkers[19] describe modelling of carbohydrate crystal structures with a customized version of the programs. Among the most popular molecular mechanics force fields, only MM3[20] has not been subjected to re-parametrization by the carbohydrate community. This is undoubtedly because of its very good performance in carbohydrate applications, but perhaps also because its in-source special treatment of anomeric effects.

The force field presented in this work can be regarded as a continuation of our previous efforts leading to the force field PEF91L,[21] but it has been totally revised and reoptimized in a stepwise manner on a functional group basis. First, it was reoptimized on the aliphatic hydrocarbons,[22] then on aliphatic and alicyclic ethers[16] with a changed electrostatic profile, and now on alcohols and carbohydrates. The force field is consistently optimized in an iterative least squares process on non-redundant experimental data with environment variables taken explicitly into account. Apart from this, it differs from other molecular mechanics force fields mainly by using Morse potentials for bonded interactions and by using small hydrogen atoms of low van der Waals energy. The latter were originally adjusted[22] by letting the carbons in *n*-octane fill out the unoccupied space in the crystal lattice and then subsequently letting the hydrogens grow until no further improvement was observed in the optimization cycle with respect to lattice energies, internal structure and vibrational

frequencies. Since then, both the aliphatic carbon and hydrogen van der Waals parameters have been left invariant.

PARAMETRIZATION

Development of the new parameter set was carried out with the Lyngby version[2,3,4] of the Consistent Force Field[1] (CFF) program. The strategy of optimization was described before;[23] it departed from the treatment of fractional charges, as described[6] in the work on ethers. The functional form of the potential energy was the same as used previously.[4,6,22,23]

INITIAL PARAMETER VALUES

Initial parameters were transferred from the set PEF91L[21] where possible. The bond dissociation energy D_e for alcohol oxygen needed for the Morse potential was set to the rounded value 418.4 kJmol^{-1} (100 kcalmol^{-1}). D_e is insensitive to the data used and did not change during optimization.

For non-bonded interactions, several initial sets of parameter values were tried, and crystal stuctures were analyzed prior to selection of one set for optimization. Selection of charge parameters is discussed in a later section. Valence angle parameters θ_0 were kept fixed at the tetrahedral angle except for those involving lone pairs at the apical atom.

EXPERIMENTAL DATA

For water and alcohols, the following data were used in optimization: The r_e structure of water,[24] the r_z structures of methanol [25] and *tert*-butyl alcohol (calculated [23] from the published[26] r_g structure); Argon matrix vibrational spectra of water and its two isotopomers,[27] and of water dimer;[28] gas phase and Argon matrix vibrational spectra of methanol and methanol-d_1,[29,30] and of ethanol and ethanol-d_1.[31,32]

For saccharides, neutron diffraction data of the following were used in optimization: β-L-arabinopyranose,[33] β-L-lyxopyranose,[34] α-L-xylopyranose,[35] methyl β-D-xylopyranoside,[33] methyl β-D-ribopyranoside,[36] α-L-rhamnose monohydrate,[37] methyl β-D-galactopyranoside,[38] methyl β-D-galactopyranoside monohydrate,[38] methyl α-D-mannopyranoside,[39] methyl α-D-glucopyranoside,[39] methyl α-D-altropyranoside,[40] α-D-glucopyranose,[41] α-L-sorbopyranose,[42] β-D-fructopyranose,[43] sucrose,[44] and β-maltose monohydrate.[45]

In order to test the validity of the parameter set we calculated some structures which were not used in the optimization: 1,2,4,5/3,6-cyclohexanehexol (*muco*-inositol, a 6-ring alcohol), the X-ray structure;[46] xylitol (a straight-chain alcohol), the X-ray structure;[47] butane-1,2,3,4-tetrol (*meso*-erythritol), the neutron structure[48] at 23 K; β-D-glucopyranosyl-(1→6)-β-D-glucopyranose (gentiobiose, a highly flexible disaccharide).[49]

PROCEDURE

Prior to the optimizations, the subroutine for calculation of fractional atomic charges from the charge parameter values was adjusted on selected alcohols, to best fit *ab initio* calculations of Wu and Sandler,[50] and the charge parameters were then kept fixed. This procedure proved successful in the work on ethers.[6] For this reason the fractional charges in PEF95SAC are significantly different from those of the previous set PEF91L,[6] where charge parameters were also optimized, with dipole moments included in the experimental database. No specific modelling of lone pairs is used.

Charge allocation is performed through the use of induction factors taking into account the local atomic environment,[3,6,23] and overall charge neutralization is made by subtraction of excess charge. Charge allocation for carbon atoms attached to alcohol oxygen atoms is the same as for carbon atoms attached to ether oxygen atoms,[6] induction factors for C-O and C'-O being the same as for C-O' and C'-O'. For hydroxy hydrogen atoms, charge is assigned by $q(H) = -d^H \times q(O)$ where d^H is the distribution factor for hydroxy hydrogen and q(O) is the charge parameter value for this atom type. Charge assignment is therefore independent of any charge parameter for hydroxy hydrogen. The best fit to *ab initio* results was obtained with $d^H = 0.63$ and q(O) = q(O'), that is, charge parameter of alcohol oxygen the same as for ether oxygen.The partial charges for selected alcohols are shown in Table 1.

Fitting parameters for van der Waals interactions is usually a very slowly converging process. The presence of many crystal structures of flexible molecules such as saccharides makes this fitting cumbersome where very small steps in parameter space are required to ensure convergence in the optimization. Each saccharide molecule has a large number of local minima; for a typical monosaccharide (an aldohexopyranose) the theoretical number of minima is 729, not considering any conformers resulting from a flexible ring shape. Even small adjustments of one or more parameter values lead to changes in the local minimum

Table 1. Fractional charges of alcohols.

		ab initio[50]	PEF95SAC[a]	PEF91L[21]
Methanol	C	-0.041	-0.014	-0.318
	H(C)	+0.110[b]	+0.119	+0.147
	O	-0.654	-0.663	-0.199
	H(O)	+0.339	+0.347	+0.076
Ethanol	C(me)	-0.357	-0.434	-0.415
	H(C)	+0.115[b]	+0.123	+0.142
	C(O)	+0.108	+0.125	-0.169
	O	-0.660	-0.659	-0.204
	H(O)	+0.335	+0.351	+0.071
Propan-1-ol	C(me)	-0.350	-0.432	-0.418
	H(C)	+0.112[b]	+0.125	+0.139
	C(C)	-0.225	-0.269	-0.255
	C(O)	+0.121	+0.128	-0.164
	O	-0.668	-0.657	-0.207
	H(O)	+0.338	+0.353	+0.069
Ethane-1,2-diol	C	+0.087[c]	+0.113	-0.159
	H(C)	+0.117[c]	+0.111	+0.144
	O	-0.674[c]	-0.672	-0.202
	H(O)	+0.353[c]	+0.338	+0.073

a. As calculated by the program from parameters in Table 2. b. Average values.
c. Average values due to inclusion of treatment of intramolecular hydrogen bonds.

positions and the shape of the potential energy surface. During the optimization process the iteratively altered conformational landscape can lead to a local minimum different from the original for the particular molecule or crystal. In the case of saccharides the exocyclic hydroxy groups tend to swap between *anti* and *gauche* positions during minimizations due to the relatively small energy barriers. Such swappings are without dramatic consequences when isolated molecules are treated. The situation is more critical when a crystal is treated. For such complex molecules there are often several possible ways to pack the molecules without changing space group, but unit cell dimensions will often become significantly different in cell lengths and angles. Such situations cause large jumps in the value of the sum of squared differences,[2,4] and thus divergence in the optimization algorithm. Due to the

flexible saccharide structure such interrupts appeared a number of times during this work. When a divergence happened, the optimization process was restarted from the last iteration before the interrupt occurred, excluding the molecule which caused the interrupt for some iterations, then it was included again. The problems mentioned were particularly frequent for the disaccharides and the compounds containing crystal water.

Van der Waals parameters were not optimized on lattice dimensions only. A representive number of intramolecular nonbonded distances which more precisely express the forces between nonbonded atoms were included in the process. The parameters for hydroxy hydrogen are mainly fitted on H'--O distances, such as between an oxygen atom in the saccharide molecule and a hydrogen atom in water for two monohydrated saccharides (α-L-rhamnose and β-maltose). Using van der Waals parameters of the same magnitude as in our previous set PEF91L[21] resulted in overestimated distances for O--O and O--H' as well. To overcome this problem, the optimization process was supplemented interactively by manual estimates of a number of parameters, and by using only nonbonded distances in the set of observables during some cycles of iteration. When a reliable set of values was found, the other data (unit cell dimensions and internal coordinates) were again activated in the final series of refinements.

RESULTS AND DISCUSSION

FINAL PARAMETERS

The final parameter set is presented in Table 2; the functional forms of the potential energy are given in previous publications.[3,4,6,22]

The final van der Waals parameters for hydroxy hydrogen need some comments. The optimization was done on the functional form $V = A/r^{12} - B/r^6$ using one-atom parameters. This is efficient, and the only form accepted in the CFF crystal program. The resulting parameter set is a best solution to the problem, given the functional form and the experimental data selected. It is *a* best, not *the* best; several sets can be obtained, of equal or very close validity. The one presented is the selected "best". It is not very pictorial, and a transformation to the (ε,r^*) form is illuminating. We derive, for H'---H', $\varepsilon = 0.00872$ Jmol^{-1} and $r^* = 586$ pm, corresponding to a van der Waals radius for H' of 293 pm. This is the result of mainly the B parameter for H' which optimized into such a low value that two H' should experience

Table 2. Parameters for the saccharide force field.[a]
Values marked with asterisk are optimized in this work.

Bond parameters	D_e	α	b_0
C - C	369.0	0.022545	151.85
C - C'	369.0	0.022545	150.11
C - O	376.6	0.022396*	141.19*
C - O'	376.6	0.021558	140.03
C' - O	376.6	0.022396*	139.39*
C' - O'	376.6	0.021558	138.99
C - H	425.1	0.018139	109.50
C' - H	425.1	0.018139	109.50
O - H'	418.4	0.023502*	95.71*
Valence angle parameters		K_θ	θ_0
C - C - C		110.8551	T_d
C' - C - C		110.8551	T_d
O' - C - C		173.1952	T_d
O - C - C		173.1952*	T_d
O - C - C'		173.1952*	T_d
O - C' - C		173.1952*	T_d
O' - C - C'		173.1952	T_d
C - C - H		111.0844	T_d
C' - C - H		111.0844	T_d
O' - C - H		127.2297	T_d
O - C - H		127.2297*	T_d
O - C' - H		127.2297*	T_d
H - C - H		94.4058	T_d
C - C' - C		110.8551	T_d
O' - C' - C		173.1952	T_d
C - C' - H		111.0844	T_d
O' - C' - O'		152.1209	T_d
O - C' - O'		152.1209*	T_d
O - C' - O		152.1209*	T_d
O' - C' - H		127.2297	T_d
H - C' - H		94.4058	T_d
C - O' - C		109.5392	102.14
C - O' - C'		109.5392	105.02
C' - O' - C'		109.5392	102.50
C' - O - H'		127.8117*	106.04*
C - O - H'		127.8117*	106.04*
H' - O - H'		118.4794*	104.45*

Torsion angle parameters	K_ϕ	Torsion angle parameters	K_ϕ
C - C - C - C	6.7224	O - C' - C - O'	19.6744*
C - C - C - H	1.0493	C' - C - C - H	1.0493
C - C - C - O	2.7422*	C' - C - C - O	2.7422*
C - C - C - O'	2.7422	C - C - C' - O	2.7422*
H - C - C - H	0.9272	C' - C - C - C	6.7224
O - C - C - H	0.0004	C - C' - C - H	1.0493
O' - C - C - H	0.0004	H - C' - C - H	0.9272
O - C - C - O	19.6744*	O' - C' - C - H	0.0004
C' - C - O - H'	0.0004	C - C' - C - C	6.7224
C - C - O - H'	0.0004	C - O' - C - H	1.4920
C - C' - O - H'	0.0004	C - O' - C - C	3.4125
O - C' - O - H'	0.0004	C' - O' - C - H	1.4920
O' - C' - O - H'	0.0004	C' - O' - C - C	3.4125
H - C - O - H'	0.9891*	C - O' - C' - H	10.6696
H - C' - O - H'	0.9891*	C - O' - C' - C	3.4125
H - C' - C - O'	0.0004	C - O' - C' - O'	1.3795*
H - C' - C - C	1.0493	C' - O' - C' - O	1.3795*
O' - C' - C - C	2.7422	C' - O' - C' - H	10.6696
O - C - C - O'	19.6744*	C' - O' - C' - C	3.4125
O' - C - C - O'	19.6744*	C' - O' - C' - O'	1.3795
O' - C' - C - O	19.6744*	C - O' - C' - O	1.3795

van der Waals and Coulomb parameters	A_i	B_i	ε	r^*	e_i
C --	71174.94	2076.96	30.304	19.094	0.0934
C' --	71174.94	2076.96	30.304	19.094	0.0934
O --	13180.24*	1094.97*	45.483	16.039	-0.6200
O' --	33430.81	2010.14	60.433	16.927	-0.6200
H --	4742.74	153.50	2.484	18.767	0.1627
H' --	3764.60*	26.52*	0.0934	24.198	0.3906

a. Units: D_e, kJ mol^{-1}; α, pm^{-1}; b_o, pm; K_θ, J mol^{-1} $^{\circ -2}$; θ_0, $^\circ$, K_ϕ, kJ mol^{-1}; A, (J mol^{-1} pm^{12})$^{0.5}$ · 10^{-12}; B, (J mol^{-1} pm^6)$^{0.5}$ · 10^{-6}; ε, (J mol^{-1})$^{0.5}$; r^*, pm$^{0.5}$; e_i are the elementary charge parameters; T_d the tetrahedral angle [=arccos (-⅓) ≈ 109.5°]. C' designates anomeric carbon, O' ether oxygen, and H' hydroxy hydrogen.
The one-atom non-bonded parameters are presented in the (A,B) form used in the calculations as well as in the (ε,r^*) form. The combination rules are $A_{ij} = A_iA_j$, $B_{ij} = B_iB_j$; $\varepsilon_{ij} = \varepsilon_i\varepsilon_j$, $r^*_{ij} = r^*_i\, r^*_j$ which is unusual.

Table 3. Gas phase structures of water and alcohols [a]

Compound	Ref.	Internal	Exp.	PEF95SAC
Water r_e	24	OH'	95.7(1)	95.7
		H'OH'	104.5(1)	104.5
Methanol r_z	25	OC	142.8(5)	141.7
		OH'	97.5(3.0)	96.0
		COH'	107.6(2.7)	108.2
tert-Butanol r_z	26	OC	144.4(1.2)	140.7
		OH'	98.3(4.2)	95.8

a. Units: pm and degree. All internal coordinates are numbered individually.

essentially no London attraction. In this representation the hydroxy hydrogens are regarded as protons almost stripped of electrons.

Taken as isolated data, these parameter values seem unrealistic, but they function well *in the consistently optimized set* where they are used in conjunction with the full set of parameters.

Lennard - Jones parameters alone do not account for the entire non-bonded interactions. The Coulomb terms play an essential part, and more so for the rather heavily charged atoms O and H'.

GAS PHASE STRUCTURES

Very few gas phase structures were used in this optimization, as ED structures of alcohols of sufficient perfection are not found. The r_e structure of water is perfectly reproduced, see Table 3; for the two small alcohols methanol and *tert*-butanol the r_z structure is reproduced within the experimental accuracy except for the CO bond which is calculated too short. The OH' bond length and the COH' angle are not determined to great precision in gas phase ED.

The Consistent Force Field will deliver a structure of the type employed in the optimization of the parameters of the energy functions which is why we use the r_z structure as the most acceptable approximation to the r_e structure.[23]

Table 4. Vibrational frequencies (in cm^{-1}) of water and alcohols.

Compound	Ref.	Internal	Exp.	PEF95SAC
Water	27	OH' as b_1	3734	3734
		OH' ss a_1	3638	3680
		H'OH' a_1	1589	1597
Water-d_1	27	OH' a'	3687	3708
		OD' a'	2710	2699
		H'OD' a'	1397	1397
Water-d_2	27	OD' as b_1	2772	2736
		OD' ss a_1	2658	2664
		D'OD' a_1	1174	1164
Water *dimer*	28	OH'	3714	(3731)
		OH'	3698	(3725)
		OH'	3626	(3680)
		OH'	3548	(3672)
		H'OH'	1618	(1607)
		H'OH'	1600	(1597)
Methanol	29,	OH' str a'	3682	3677
	30	COH' bend a'	1336	1279
		CO str a'	1033	1058
		ϕ a''	295	286
Methanol-d_1	29,	OD st a'	2718	2679
	30	COD' bend a'	864	879
		ϕ a''	215	227
Ethanol	31,	OH' str a'	3676	3693
	32	COH' bend a'	1241	1297
		CO str a'	1089	1096
		CC/CO str a'	885	939
		OCC bend a'	419	407
		ϕ a'' (-O-C-)	201	211
Ethanol-d_1	31,	OD' str a'	2713	2691
	32	COD' bend a'	896	844

VIBRATIONAL SPECTRA

Vibrational spectra of the gaseous molecules are neatly reproduced, as seen from Table 4. Reproduction in the bending region does not suffer from the absence of cross-terms so we do not encounter this problem as we did for the alkanes.[22]

The cumulated error in all 24 frequencies used in optimization is -20 cm^{-1}; the largest error is -57 cm^{-1}; the average numerical error is 22 cm^{-1}; and the root mean square error is 6

cm^{-1}. The good accuracy, especially of the low frequencies, is important for valid predictions of thermodynamic properties.

CRYSTAL STRUCTURES

The new parameter set PEF95SAC is based on the most extensive optimization on crystal structures so far done for saccharides. Using convergent lattice summation,[2] intermolecular interactions are taken into account explicitly and are not included implicitly in the parameter values. Table 5 collects a selection of data from the crystal structures of those saccharides which are used in the optimization. In general, all CO, C'O, CO', C'O' and OH' bond lengths, angles including -O-, and similarly relevant torsions or non-bonded distances are included in the optimization, together with the unit cell dimensions.

The cumulated error in C'O bond length is -0.5 pm, the largest error is -1.1 pm, the average numerical error is 0.7 pm, and the root mean square error is 0.3 pm. The cumulated error in CO bond length is -10.6 pm, the largest error is -2.5 pm, the average numerical error is 6.5 pm, and the root mean square error is 0.2 pm. The cumulated error in unit cell volume is 419.9 $\times 10^6$ p^3, the largest error is 63.2 *10 pm^3, the average numerical error is 29.2 *10 pm^3, and the root mean square error is 8.8 *10 pm^3.

The validity of PEF95SAC was checked by calculation of the structures of one cyclic and two straight-chain alcohols and one disaccharide which were not included in the optimization process. Selected data for torsions and non-bonded distances, and unit cell dimensions, are shown in Table 6.

The cumulated error in CO bond length is 10.2 pm, the largest error is 2.0 pm, the average numerical error is 0.7 pm, and the root mean square error is 0.3 pm. The torsions are well represented, with the exception of a few involving hydrogen bonding. The cumulated error in unit cell volume is 2.8 $\times$ 10 pm^3, the largest error is -57.1 $\times$ 10 pm^3, the average numerical error is 33.0 $\times$ 10 pm^3, and the root mean square error is 23.0 $\times$ 10 pm^3. The corresponding figures for crystal structures included in the optimization are 419.9, 63.2, 29.2 and 8.8 $\times$ 10 pm^3, indicating a much stronger bias.

The reproduction of the conformation of erythritol and of structural details of the very flexible disaccharide gentiobiose is noteworthy. As stated above, the optimization gave a systematic error in volume, which with one exception is calculated too small (see Table 5);

Table 5. Crystal structures of saccharides. Experimental data from neutron diffraction.

Compound	Ref.	Quantity	Exp.	PEF95SAC
β-L-Arabino-	33	H'O1C'1H	-53.2 (11)	-44.9
pyranose		H'O2C2H	20.7 (11)	47.7
Orthorhombic		H'O3C3H	-21.0 (11)	-27.9
$P2_12_12_1$		H'O4C4H	27.7 (11)	46.4
Z=4		C4O'5C'1O1	60.7 (6)	63.3
		A	651.4 (6)	622.0
		B	1945.3 (12)	1812.7
		C	483.6 (3)	540.0
		Vol	612.8	608.9
β-L-Lyxopyranose	34	O'5C'1O1H'	-78.2 (6)	-56.3
Orthorhombic		A	959.9 (9)	932.6
$P2_12_12_1$		B	1036.5 (9)	1068.4
Z=4		C	652.4 (6)	621.4
		Vol	649.1	619.2
α-L-Xylopyranose	35	O'5C'5O5H'	-104.7 (6)	-75.1
Orthorhombic		A	922.7 (3)	1051.3
$P2_12_12_1$		B	1264.3 (9)	1297.7
Z=4		C	561.9 (6)	440.5
		Vol	655.5	600.9
Methyl β-D-Xylo-	33	O'5C'1O'1C6	-72.1 (6)	-67.2
pyranoside		A	787.7 (6)	752.1
Monoclinic		B	693.3 (9)	629.5
$P2_1$		C	774.8 (6)	848.9
Z=2		β	113.38(23)	117.49
		Vol	388.4	356.5
Methyl β-D-Ribo-	36	O'5C'1O'1C6	-67.7 (12)	-64.9
pyranoside		O...O	276.1 (15)	286.8
Orthorhombic		H'...O	195.9 (24)	228.0
$P2_12_12_1$		A	575.3 (15)	582.3
Z=4		B	1998.6 (15)	1896.5
		C	641.3 (15)	645.2
		Vol	737.4	712.5

(continued)

Table 5. Continued.

α-L-Rhamnose monohydrate Monoclinic $P2_1$ Z=2	37	H'O1C'1H	18.2 (6)	51.1
		H'O2C2H	48.4 (6)	44.6
		H'O3C3H	46.9 (6)	52.3
		H'O4C4H	-22.5 (6)	-30.5
		H'4...Ow	182.0 (9)	224.1
		A	790.1 (9)	780.3
		B	792.2 (9)	757.5
		C	667.0 (6)	703.8
		β	95.52(11)	97.77
		Vol	415.6	412.2
Methyl β-D-Galacto-pyranoside Orthorhombic $P2_12_12_1$ Z=4	38	O'5C'1O'1C7	-77.1 (6)	-57.4
		H'O2C2H	-21.7 (11)	-32.2
		H'O3C3H	145.8 (11)	82.1
		H'O4C4H	-23.7 (11)	22.1
		O6C6C5O'5	63.5 (6)	58.3
		H'O6C6H	133.8 (11)	167.0
		H'O6C6H	15.7 (11)	47.8
		A	777.9 (6)	834.8
		B	853.5 (6)	745.5
		C	1313.1 (15)	1318.4
		Vol	871.8	818.6
Methyl α-D-Galacto-pyranoside, monohydrate Orthorhombic $P2_12_12_1$ Z=4	38	O'5C'1O'1C7	64.0 (6)	60.2
		H'O2C2H	5.9 (11)	42.9
		H'O3C3H	-10.3 (11)	-40.5
		H'O4C4H	25.6 (11)	37.8
		O6C6C5O'5	62.3 (6)	62.1
		H'O6C6H	-178.0 (11)	-170.9
		H'O6C6H	-59.4 (11)	-53.4
		A	612.2 (3)	565.3
		B	748.6 (6)	753.0
		C	2119.8 (18)	2236.9
		Vol	971.5	952.3
Methyl α-D-Manno-pyranoside Orthorhombic $P2_12_12_1$ Z=4	39	O'5C'1O'1C7	61.0 (-)	62.4
		O6C6C5O'5	-65.0 (-)	-61.5
		A	942.9 (9)	851.9
		B	931.5 (9)	953.1
		C	1005.5 (9)	1041.0
		Vol	883.1	845.2

Methyl α-D-Gluco-	39	O'5C'1O'1C7	63.0 (-)	61.3
pyranoside		O6C6C5O'5	74.0 (-)	73.0
Orthorhombic		A	1131.1 (9)	1092.2
$P2_12_12_1$		B	1478.1 (12)	1260.1
Z=4		C	528.1 (6)	616.5
		Vol	882.9	848.4
Methyl α-D-Altro-	40	O'1...O3	295.2 (10)	267.8
pyranoside		A	748.6 (12)	730.0
Orthorhombic		B	909.8 (21)	980.0
$P2_12_12_1$		C	1333.0 (21)	1190.3
Z=4		Vol	907.9	851.5
α-D-Gluco-	41	O6C6C5O'5	70.2 (6)	63.9
pyranose		A	1036.6 (2)	1028.0
Orthorhombic		B	1485.1 (5)	1529.8
$P2_12_12_1$		C	497.5 (1)	485.3
Z=4		Vol	765.9	763.1
α-L-Sorbo-	42	O'6C'2C1O1	78.1 (-)	60.8
pyranose		A	654.5 (6)	672.8
Orthorhombic		B	1806.2	1894.6
$P2_12_12_1$		C	631.0 (6)	561.7
Z=4		Vol	745.9	716.0
β-D-Fructo-	43	O'6C'2C1O1	-61.4 (-)	-63.2
pyranose		A	919.1 (6)	931.9
Orthorhombic		B	1004.6 (6)	973.8
$P2_12_12_1$		C	809.5 (6)	795.6
Z=4		Vol	747.4	722.0
Sucrose	44	O6C2C'1O'1	54.8 (6)	58.1
Monoclinic		O1'...O2	278.1 (10)	305.5
$P2_1$		H'1'...O2	185.1 (9)	224.2
Z=2		O6'...O'5	285.0 (10)	292.5
		H'6'...O'5	189.5 (9)	240.9
		A	1086.3 (2)	1045.6
		B	870.5 (1)	923.5
		C	775.9 (1)	750.4
		β	102.95 (6)	107.24
		Vol	715.0	692.0

(continued)

Table 5. Continued.

β-Maltose	45	O'5C'1O1H'	-100.8 (6)	-72.9
monohydrate		H'O2'C2'C'1'	75.7 (6)	60.2
Monoclinic		H'O3'C3'C2'	-75.3 (6)	-75.6
$P2_1$		O6'C6'C5'O'5'	-62.5 (6)	-69.2
Z=2		H'O6'C6'C'5	70.0 (6)	86.1
		H'O2C2C'1	-48.3 (6)	-55.4
		H'O3C3C2	-138.8 (6)	-86.8
		H'O4C4C3	65.5 (6)	65.5
		O6C6C6O'5	59.2 (6)	62.1
		H'O6C6C5	-70.8 (6)	-70.1
		C5O'5C'1O1	-179.9 (6)	-179.9
		O3'...H'2	183.6 (20)	238.9
		H'6'...Ow	176.2 (20)	245.4
		A	486.6 (6)	488.8
		B	1507.7 (18)	1640.0
		C	1070.2 (15)	1009.9
		β	97.07 (6)	95.74
		Vol	779.2	805.4

this was found also in previous work,[6] but is not substantiated by the check in Table 6. The unit cell of erythritol is no longer tetragonal after minimization in PEF95SAC. We usually run minimizations so that only the angles of the crystal system are kept fixed while axes are allowed to change, as are, indeed, all internal degrees of freedom. Axis lengths could have been locked, but the present procedure allows a more thorough check of the parameter set and gives better hints for future improvements. The discrepancies indicate that such attempts should concentrate on non-bonded parameters for hydrogen. According to our experience, this is worthwhile only in a major effort involving not just new parameters but a different approach, in particular another treatment of Coulomb interactions.

The quality of PEF95SAC is further illustrated in Fig. 1 which shows the packing of the unit cell of *meso*-erythritol.

ENERGY DIFFERENCES AND BARRIERS

Energy barriers were calculated for three small alcohols. For methanol we found 4.02 $kJmol^{-1}$ which compares well with 4.45 $kJmol^{-1}$ deduced by Lees and Baker[51] from microwave

Table 6. Crystal structures of alcohols and gentiobiose. Experimental data from X-ray and neutron diffraction.

Compound	Ref.	Quantity	Exp.	PEF95ALC
1,2,4,5/3,6-	46	H'OCH	45.3 (-)	61.3
Cyclohexanehexol		CCCC	-54.5 (6)	-58.4
(Muco-inositol)		H'OCH	-54.5 (-)	20.3
Monoclinic		CCCC	47.6 (6)	59.6
$P2_1/c$		H'OCH	-53.5 (-)	-41.3
Z=4		CCCC	-47.5 (6)	-59.1
x-ray		H'OCH	-27.4 (-)	28.9
		CCCC	54.9 (6)	57.4
		H'OCH	23.7 (-)	27.4
		CCCC	-61.8 (6)	-57.1
		H'OCH	19.8 (-)	-41.2
		CCCC	61.0 (6)	57.7
		O...O	279.0 (6)	283.8
		O...O	279.8 (6)	283.5
		O...O	299.3 (6)	257.0
		O...O	280.8 (6)	284.5
		O...O	280.9 (6)	282.7
		A	672.7 (3)	684.0
		B	949.1 (6)	919.2
		C	1308.5 (6)	1304.0
		β	119.52 (6)	119.58
		Vol	727.0	713.0
Xylitol	47	H'OCC	-100.6 (-)	-163.7
Orthorhombic		OCCC	-175.0 (-)	-175.6
$P2_12_12_1$		H'OCH	-14.2 (-)	-0.8
Z=4		CCCC	176.1 (-)	-178.8
x-ray		H'OCH	18.0 (-)	52.1
		CCCC	-70.1 (-)	-62.1
		H'OCH	21.6 (-)	28.0
		CCCO	173.2 (-)	-179.4
		H'OCC	19.8 (-)	76.6
		A	829.1 (6)	899.1
		B	897.0 (6)	910.0
		C	897.0 (15)	824.3
		Vol	667.1	674.4

(continued)

Table 6. Continued.

Erythritol	48	H'OCC	-80.2 (1)	-69.1
Tetragonal		OCCO	62.9 (1)	64.5
$I4_1/a$		H'OCH	39.2 (1)	33.1
Z=8		A	1271.3 (5)	1339.5
neutron		B	1271.3 (5)	1375.7
		C	674.7 (2)	622.8
		Vol	1090.5	1147.6
Gentiobiose	48	C1'-O6-C6	113.3	114.1
Orthorhombic		ϕ	63.2	62.8
$P2_12_12_1$		ψ	-156.3	-158.9
Z=4		ω	-177.9	173.1
x-ray		-C5-C6-	g'g	g'g
		H1'---HR	240	236
		H1'---HS	312	329
		A	886.93	845.2
		B	2284.60	2283.5
		C	720.11	729.0
		Vol	1459.2	1406.9

spectra. For ethanol the *anti* conformer was found 1.34 kJmol^{-1} above the *gauche*, with barriers $a \rightarrow g$ 2.2 and $g \rightarrow a$ 3.3 kJmol^{-1}. Kakar and Seibt[52] found a barrier of 5.0 kJmol^{1} from analysis of the microwave spectrum of the *gauche* conformer. For 2-propanol, the *gauche* conformer is 0.58 kJmol^{-1} above the *anti*, with barriers $a \rightarrow g = 6.3$ and $g \rightarrow a = 5.7$ kJmol^{-1}. Kondo and Hirota[53] found a barrier of 7.0 kJmol^{-1} from analysis of the microwave spectrum. The four conformers of 1-propanol have the following relative energies: *ag*, 0.00; *aa*, 0.40; *gg*, 4.21; *ga*, 5.33 kJmol^{-1}.

THERMODYNAMIC PROPERTIES

Conformational equilibrium constants for all isomers of methylcyclohexanol were measured by Subbotin *et al.*[54] by NMR at 193 K. We calculated differences in ΔG at 200 K; the results are given in Table 7.

Calculation of ΔG took into account all vibrational degrees of freedom but not translational and rotational, as comparison was to be made with measurements in solution.

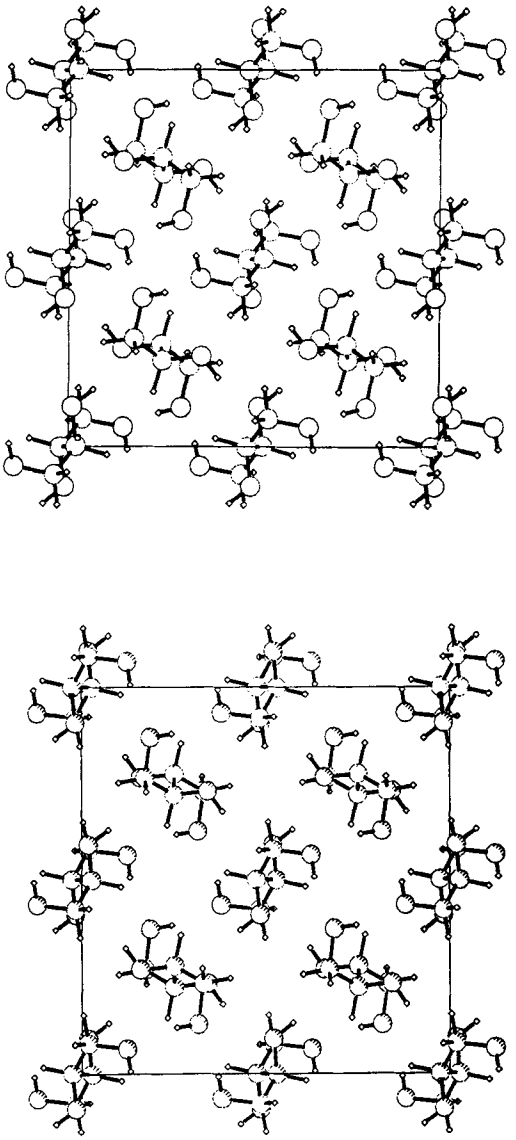

Fig. 1. Packing of *meso*-erythritol. Above: from the Cambridge Structural Database. Below: energy-minimized in PEF95SAC.

Table 7. Conformers and conformational equilibria in methylhexanol.

Isomer	Conformer	ΔG	ΔG_{rel}	K_{calc}	K_{exp} [52]
1-OH-1-me	*eq-ax*	3.98	3.98		
	ax-eq	0.00	0.00	11	2
1-OH-2-me	*eq-eq*	173.51	0.00	4491	>50
	ax-ax	187.67	14.16		
	eq-ax	186.11	5.53		
	ax-eq	180.58	0.00	28	4
1-OH-3-me	*eq-eq*	62.74	0.00	0.4×10^6	>50
	ax-ax	84.25	21.51		
	eq-ax	73.29	4.45		
	ax-eq	68.84	0.00	15	13
1-OH-4-me	*eq-eq*	87.14	0.00	12×10^3	>50
	ax-ax	102.79	15.65		
	eq-ax	96.95	1.20		
	ax-eq	95.75	0.00	2	7

All trends in the experimental data are reproduced. The quantitative discrepancies may be due to neglect of intermolecular interactions in the liquid phase.

An independent example of the utility of PEF95SAC in an extensive modelling can be seen in the following paper on the potential energy surface and NMR properties of β-lactose.[55]

REFERENCES

1. S. Lifson and A. Warshel, *J. Chem. Phys.*, **49**, 5116 (1968).
2. S.R. Niketic and K. Rasmussen, *The Consistent Force Field: A Documentation, Lecture Notes in Chemistry*, Vol. 3, Springer, Berlin, 1977. Kjeld Rasmussen, *Potential Energy Functions in Conformational Analysis. Lecture Notes in Chemistry*, Vol. 37, Springer, Berlin, 1985.
3. K. Rasmussen, S.B. Engelsen, J. Fabricius and B. Rasmussen, *NATO ASI Series C* **406**, 381 (1993).
4. S.B. Engelsen, J. Fabricius and K. Rasmussen, *Comput. Chem.*, **18**, 397 (1994).
5. I. Tvaroska and T. Bleha, *Adv. Carbohydr. Chem. Biochem.*, **47**, 45 (1989).
6. J. Fabricius, S.B. Engelsen and K. Rasmussen, *New J. Chem.*, **19**, 1123 (1995).
7. (a) N.L. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977) (b) N.L. Allinger, S. H.-M. Chang, D.H. Glaser and H. Hönig, *Israel J. of Chem.*, **20**, 51 (1980).

8. U. Burkert and N.L. Allinger, *Molecular Mechanics*, ACS Monograph No. **177**, American Chemical Society, Washington DC, 1982.
9. I. Tvaroska and S. Pérez, *Carbohydr. Res.*, **149**, 389 (1986).
10. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).
11. S.N. Ha, A. Giammona, M. Field, and J.W. Brady. *Carbohydr. Res.*, **180**, 207 (1988).
12. (a) P.D.J. Grootenhuis and C.A.G. Haasnoot, *Mol. Simul.*, **10**, 75 (1993).
(b) M.L.C.E. Kouwijzer and P.D.J. Grootenhuis, *J. Phys. Chem.*, **99**, 13426 (1995).
13. S.J. Weiner, P.A. Kollman, D.T. Nguyen and D.A. Case, *J. Comput. Chem.*, **7**, 230 (1986).
14. S. Homans, *Biochemistry*, **29**, 9110 (1990).
15. T.M. Glennon, Y.-J. Zheng, S.M. Le Grand, B.A. Shutzberg and K.M. Merz, *J. Comput. Chem.*, **15**, 1019 (1994).
16. R.J. Woods, R.A. Dwek, C.J. Edge and B. Fraser-Reid, *J. Phys. Chem.*, **99**, 3832 (1995).
17. A. Imberty, K.D. Hardman, J.P. Carver and S. Pérez, *Glycobiology*, **1**, 631 (1991).
18. S. Pérez, C. Meyer and A. Imberty, *Modelling of Biomolecular Structures and Mechanisms*, A. Pullman et al., Eds., Kluwer Academic Publishers, Doordrecht, 1995, p 425 .
19. A. Sarko, C.H. Chen, B.J. Hardy and F. Tanaka, in *Computer Modeling of Carbohydrate Molecules*, A. D. French and J. Brady, Eds., ACS Symposium Series **430**, 1990, p. 345.
20. (a) N.L. Allinger, Y.H. Yuh and J.-H. Lii, *J. Am. Chem. Soc.*, **111**, 8551 (1989);
(b) N.L. Allinger, M. Rahman and J.-H. Lii, *J. Am. Chem. Soc.*, **112**, 8293 (1990).
21. S.B. Engelsen and K. Rasmussen, *Int. J. Biol. Macromol.*, **15**, 56 (1993).
22. S.B. Engelsen, J. Fabricius and K. Rasmussen, *Acta Chem. Scand.*, **48**, 553 (1994).
23. S.B. Engelsen, J. Fabricius and K. Rasmussen, *Acta Chem. Scand.*, **48**, 548 (1994).
24. *Tables of Interatomic Distances and Configuration in Molecules and Ions*, Spec. Publ. **18**, Suppl. 1956-59. The Chemical Society, London, 1965 , p M37s.
25. T. Iijima, *J. Mol. Struct.*, **212**, 137 (1989).
26. A. Suwa, H. Ohta and S. Konaka, *J. Mol. Struct.*, **172**, 275 (1988).
27. G. P. Ayers and A. D. E. Pullin, *Chem. Phys. Lett.* **29**, 609 (1974).
28. A. J. Tursi and E. R. J. Nixon, *Chem. Phys.*, **52**, 1521 (1970).
29. A. Serrallach, R. Meyer and H. H.J. Günthard, *Mol. Spectrosc.*, **52**, 94 (1974).
30. V. Stern, C. Belorgeot, J. Kachmarsky and K. D.J. Möller, *Mol. Spectrosc.*, **67**, 244 (1977).
31. J. P. Perchard and M. L. Josien, *J. Chim. Phys.*, **65**, 1834 (1968).
32. A. J. Barnes and H. E. Hallam, *Trans. Faraday Soc.*, **66**, 1932 (1970).
33. S. Takagi and G. A. Jeffrey, *Acta Cryst.*, **B33**, 3033 (1977).
34. S. Nordenson, S. Takagi and G. A. Jeffrey, *Acta Cryst.*, **34**, 3809 (1978).
35. S. Takagi and G. A. Jeffrey, *Acta Cryst.*, **B35**, 1482 (1979).
36. V. J. James, J. D. Stevens and F. H. Moore, *Acta Cryst.*, **B34**, 188 (1978).
37. S. Takagi and G. A. Jeffrey, *Acta Cryst.*, **B34**, 2551 (1978).
38. S. Takagi and G. A. Jeffrey, *Acta Cryst.*, **B35**, 902 (1979).
39. G. A. Jeffrey, R. K. McMullan and S.Takagi, *Acta Cryst.*, **B33**, 728 (1977).

40. B. J. Poppleton, G. A. Jeffrey and J. B. G. Williams, *Acta Cryst.*, **B31**, 2400 (1975).
41. G. M. Brown and H. A. Levy, *Acta Cryst.*, **B35**, 656 (1979).
42. S. Nordenson, S. Takagi and G. A. Jeffrey, *Acta Cryst.*, **B35** 1005 (1979).
43. S. Takagi and G. A. Jeffrey, *Acta Cryst.*, **B33**, 3510 (1977).
44. G. M. Brown and H. A.Levy, *Acta Cryst.*, **B29**, 790 (1973).
45. M. E. Gress and G. A. Jeffrey **B33**, *Acta Cryst.*, **B33**, 2490 (1977).
46. D. C. Craig and V. J. James, *Cryst. Struct. Comm.*, **8**, 629 (1979).
47. C. Ceccarelli, G. A. Jeffrey and R. K. McMullan, *Acta Cryst.*, **B36**, 3079 (1980).
48. H. S. Kim and G. A. Jeffrey, *Acta Cryst.*, **B25**, 2607 (1969).
49. D. C. Rohrer, A. Sarko, T. L. Bluhm and Y. N. Lee, *Acta Cryst.*, **B36**, 650 (1980).
50. H. S. Wu and S. I. Sandler, *Ind. Eng. Chem. Res.*, **30**, 881 (1991).
51. R. M. Lees and J. G. Baker, *J. Chem. Phys.*, **48**, 5299 (1968).
52. R. K. Kakar and P. J. Seibt, *J. Chem. Phys.*, **57**, 4060 (1972).
53. S. Kondo and E. Hirota, *J. Mol. Spectrosc.*, **34**, 97 (1970).
54. O. A. Subbotin, N. M. Sergeyev, V. N. Chlopkov, N. G. Nishikova and Yu. G. Bundel, *Org. Magn. Reson.*, **13**, 259 (1980).
55. S.B. Engelsen and K. Rasmussen, following paper.