An approximation of the Cioslowski–Mixon bond order indexes using the AlteQ approach

Elena Salmina · Maria A. Grishina ·
Vladimir A. Potemkin

Abstract Fast and reliable prediction of bond orders in organic systems based upon experimentally measured quantities can be performed using electron density features at bond critical points (J Am Chem Soc 105:5061–5068, 1983; J Phys Org Chem 16:133–141, 2003; Acta Cryst B 61:418–428, 2005; Acta Cryst B 63:142–150, 2007). These features are outcomes of low-temperature high-resolution X-ray diffraction experiments. However, a time-consuming procedure of gaining these quantities makes the prediction limited. In the present work we have employed an empirical approach AlteQ (J Comput Aided Mol Des 22:489–505, 2008) for evaluation of electron density properties. This approach uses a simple exponential function derived from comparison of electron density, gained from high-resolution X-ray crystallography, and distance to atomic nucleus what allows calculating density distribution in time-saving manner and gives results which are very close to experimental ones. As input data AlteQ accepts atomic coordinates of isolated molecules or molecular ensembles (for instance, protein–protein complexes or complexes of small molecules with proteins, etc.). Using AlteQ characteristics we have developed regression models predicting Cioslowski–Mixon bond order (CMBO) indexes (J Am Chem Soc 113(42):4142–4145, 1991). The models are characterized by high correlation coefficients lying in the range from 0.844 to 0.988 dependently on the type of covalent bond, thereby providing a bonding quantification that is in reasonable agreement with that obtained by orbital theory. Comparative analysis of CMBOs approximated using topological properties of AlteQ and experimental electron densities has shown that the models can be used for fast determination of bond orders directly from X-ray crystallography data and confirmed that AlteQ characteristics can replace experimental ones with satisfactory extent of accuracy.

Keywords Covalent bond order · Topological analysis of electron density · Free-orbital approach · X-ray crystallography

Introduction

The modern quantitative description of connectivity in organic molecular systems is inseparably concerned with an estimation of bond orders. Traditionally accepted that values of bond orders are associated with electron density shared by adjacent atoms. Any transformations in molecular systems, for instance, a conversion of reactant(s) to product(s), intermediate formation or complexing by attractive interactions, etc., result in reorganization of electron structures and it turns in changes of bond order magnitudes. Therefore, estimation of orders of bonding is extensively used in studies of reaction mechanisms, complex formation of organic and organometallic compounds, coordination complexes, thermodynamic and vibrational characteristics of chemical bonds, etc. [1–5].

A theoretical bond order definition became available owing to the ingenious idea of Lewis [6, 7], who based upon concepts of those predecessors such as Frankland [8], Kekule [9–11], Couper [12], etc., and atom electron structure [13–16], offered an explanation of chemical

E. Salmina (✉)
Institute of Organic Synthesis, Ural Branch of the Russian Academy of Sciences, 620069 Ekaterinburg, Russia
e-mail: es.salmina@gmail.com

M. A. Grishina · V. A. Potemkin
Pharmaceutical Department, South Ural State Medical University, 454048 Chelyabinsk, Russia

bonding considering a bond as a result of sharing one or several electron pairs by adjacent atoms. This concept has become a foreground of subsequent quantum approaches for calculations of bonding properties [e.g., 17–20] and formed the starting point for practically all bond order determination schemes [21–35]. Thus, a number of authors realized Lewis's sharing electron bond concept in the framework of molecular orbital representation of molecules and so chemical bonding [22–29, 36]. However, this type of bond order determination schemes demands splitting up molecular space into orbitals and so strongly depends on the choice of a basis set for construction of these orbitals.

Another representation of molecular electron structure is offered in the Quantum Theory Atoms in Molecules developed by Bader and co-authors [37]. In the Bader's concept, a molecule is represented as an assemblage of the atomic volumes divided by a series of surfaces through which the gradient vector field of electron density has no flux. Hence, chemical bonding can be considered as a result of sharing two atoms a common interatomic surface. Based upon this type of molecular space partitioning Cioslowski and Mixon [21] proposed a bond order index n_{CM} which is assessed according to the next calculation sequence: (1) determination of the atomic boundaries as zero-flux surfaces that bound atomic basins; (2) calculation of the atomic overlap matrix elements in the molecular orbital basis as the integral of a basis function product over a particular basin (A):

$$\langle \varphi_i \mid \varphi_j \rangle_A = \int_\Omega \varphi_i(r) \varphi_j(r) dr$$

Finally, the bond order n_{CM} is computed by summing corresponding contributions of A and B atoms:

$$n_{CM} = \sum_k n_k^2 \langle \varphi_k \mid \varphi_k \rangle_A \langle \varphi_k \mid \varphi_k \rangle_B, \qquad (1)$$

where φ_k are the localized orbitals, k is the number of the orbital and n_k is the occupancy of the kth orbital. This scheme of bond order assessment "requires neither references to the basis sets used in the expansion of the molecular wave function nor a parameterization involving a set of standard bonds" [21] and is enough successful approach uniting the original Lewis's idea of shared electrons and the quantum determination of their quantity.

However, none of proposed schemes are appropriate for assessment of bond order values directly from experimental electron density because of disability of the latest to be divide into atomic or molecular orbitals and so their contributions to bonding. But measuring of order of bonding in molecular systems directly from experimental data is an attractive unchallenged problem. One of the

appropriate solutions of the problem is an approximation of bond orders determined in the framework of suitable theoretical techniques by experimentally measured quantities [38–41]. Such approximation for CMBOs first has been performed by Howard and Lamarch [39]. They proposed a regression model interrelating CMBOs and topological features of electron density at the bond critical point [37], such as electron density ρ_{BCP}, the Laplacian of electron density $\nabla^2 \rho_{BCP}$ and the Hessian eigenvalues (λ_1, λ_2 and λ_3) (Eq. 2). n_{CM}^{H-L}-Equation was developed for five types of the covalent bonds (C–C, C–O, C–N, C–S and C–P), comprises experimentally measured quantities and has simple mechanistical interpretation what makes it very attractive.

$$n_{CM}^{H-L} = a + b \cdot \rho_c + c \cdot \lambda_3 + d \cdot (\lambda_1 + \lambda_2). \qquad (2)$$

However, subsequent application of the Eq. (2) carried out by Tsirelson et al. [41] has shown that correlation coefficients between original n_{CM} and n_{CM}^{H-L} for the C–C and C–O bonds are only 0.84 and 0.57, respectively. They also found that n_{CM}^{H-L} for the C–C bonds appeared to be systematically higher than n_{CM} computed from the same wave functions exhibiting reasonable correlation only in the limited range $1.1 < n_{CM} < 1.4$. So Tsirelson et al. conducted a systematic study of the Howard-Lamarche regression model and modified coefficients in Eq. (2) what has led to significant improvement of correlation between original and predicted CMBOs.

For CMBO-prediction above-mentioned authors made use of topological characteristics of quantum electron density. In the present paper it is proposed to approximate CMBOs utilizing electron density calculated in the framework of an AlteQ approach [42].

AlteQ is an alternative free-orbital approach which uses the simple empirical electron density function ρ_A^{Alt} (Eq. 3) based on the relationship analysis between the logarithm of experimental electron density derived from the high-resolution X-ray diffraction experiment and the distance to the nuclei [42]:

$$\rho_A^{Alt} = \sum_{i=1}^{n} a_{Ai} \cdot 10^{b_{Ai} \cdot R_A}, \qquad (3)$$

where a_{Ai} and b_{Ai} are parameters for each element, n is the period number of the atom A, and R_A is the distance from the nucleus of atom A. Thus, the AlteQ function associates an electron density at each point of the molecular space with own characteristics of atoms: period number and distance from nucleus. Such simple parameterization of the electron density function allows reconstructing electron distribution very fast and gaining results which are very close to experimental ones. It is especially valuable for studying of connectivity in such molecular aggregates as crystal structures, polymeric and protein systems,

coordination complexes and organometallic compounds where the analysis of electron density topology on the high level of theory (for instance, B3LYP/6-311++G(d,p)) is a very time-consuming procedure. As an example, PC AlteQ allows calculating electron density and its topology for medium-size molecules (about 30 atoms) within a few seconds, for larger molecules (about 80 atoms) within about 1 min, in crystal cells containing 6-8 large-size molecules within an hour. And provided that AlteQ electron density is adjusted to experimental one calculated and experimental data are in a good agreement.

In this paper we have used topological features of AlteQ electron density at bond critical points to generate a set of regression models for prediction of CMBOs (calculated at DFT B3LYP/6-311G(d,p) level of theory) and have discussed a possibility of usage these models and AlteQ *per se* for studying connectivity in molecular systems.

Computational method

For development of regression models a range of closed-shell molecules belonging to the next classes of organic compounds: linear, branched and cyclic alkanes, alkenes, terminal, internal and conjugated alkynes; cumulated and conjugate dienes; halogenated saturated and unsaturated hydrocarbons; primary, secondary and tertiary alcohols containing one, two or three OH-groups; saturated and unsaturated aldehydes, ketones, ketenes, ethers; saturated and unsaturated monoatomic and diatomic carboxylic acids; amides, imides, anhydrides, halogenanhydrides, esters; acetales; amino acids; enols and enamines; primary, secondary and tertiary amines; sulphides, disulphides, sulfones, sulfonic acids, mercaptans, thiocarbonyl compounds; small heterocycles containing one heteroatom; peroxides; nitrates, nitrites; cyanates, isocyanates, thiocyanates; hydrazines, oximes; amino alcohols; monocyclic and fused aromatics was used. In total, 250 molecules were considered. A geometry of each of them was optimized at the DFT B3LYP/6-311G(d,p) level of theory using *GA-MESS* [43]. For all bonds in optimized structures n_{CM}-bond orders were computed using DFT B3LYP/6-311G(d,p) many-electron wavefunctions and AIMPAC software [37, 44]. A comprehensive training set included 286 C–C bonds, 122 C–O bonds, 51 C–N bonds, 19 C–S bonds, 34 C–Cl bonds, 37 C–F bonds and 723 X–H bonds, where X is C, N, O, S atom. Then for each molecule an electron density distribution was calculated using AlteQ and optimized atomic coordinates as input data. Topological analysis of AlteQ electron density was performed in AIMPAC and calculated quantities and their derivatives used as descriptors of covalent bonds in question. Finally, the multiple linear and non-linear regression analyses were

applied to matrix table for construction of CMBO-predictive models. The quality of the models was estimated using the following statistical characteristics: correlation coefficient R (equal to the fraction of variance accounted for by the model); Fisher-statistic F, which represents the statistical reliability of the model and is calculated as the ratio of variances derived by computation with a model under consideration (predicted values) and those of a corresponding zero-order model (the F-statistics depends on both the number of variables in the regression equation and the number of points); and root-mean-square error S. These statistics for developed models are presented in Tables 1, 2.

Results and discussion

Topological characteristics of electron density are widely used for approximation of bond orders estimated by different theoretical techniques to make bond orders "observed quantities" [e.g., 39–41]. However, accurate X-ray crystallographic data are not always available. They can be replaced by quantum mechanical calculations but it would be time-consuming procedure for such molecular systems as crystal fragments, protein–protein/small molecule complexes, chelates, etc. It is conditioned on the cause that a calculation of electron density distribution in such fair-sized molecular systems at the high level of theory demands colossal number of quantum wave-functions and so colossal time. Moreover, it is conjugated with a number of quantum mechanical problems such as the basis set superposition error. AlteQ provides an opportunity to shorten calculation time considerably and gain results closed to experimental ones. This has predetermined its usage in this work and it has been found out that topological features of AlteQ electron density allow predicting n_{CM}-bond order better than quantum ones.

Prediction of n_{CM}-bond order values using AlteQ electron density

The first attempt was concerned with the determination of a unified descriptor of n_{CM}-bond orders. By analogy with the Bader's theory [38] we examined simple exponential (Eq. 4) and quadratic (Eq. 5) dependence of n_{CM}-values on AlteQ electron density at bond critical points ρ_c^{Alt}:

$$n_{CM}^{Alt} = a + b \cdot exp \, \rho_c^{Alt}. \tag{4}$$

$$n_{CM}^{Alt} = a + b \cdot \left(\rho_c^{Alt}\right)^2. \tag{5}$$

The test statistics for the models are shown in Table 1. Based on the obtained results the exponential dependence yields better statistics than the quadratic one and so has

Table 1 Summary of one-parameter exponential and quadratic Cioslowsky–Mixon bond order models

Bonds	$n_{CM}^{Alt} = a + b \cdot exp\, \rho_c^{Alt}$		$n_{CM}^{Alt} = a + b \cdot \left(\rho_c^{Alt}\right)^2$		Correlation coefficient R	Fisher statistic $F(N)$	R. m. s. error, S
	a	b	A	b			
C–C	−0.366	0.489			0.988	11164.1 (286)	0.070
			0.235	0.718	0.987	10622.0 (286)	0.072
					0.980	**458.7 (55)**	**0.040**
C–N	0.181	0.220			0.981	1268.4 (51)	0.102
			0.343	0.379	0.976	976.6 (51)	0.115
					0.950	**221.2 (82)**	**0.070**
C–O	0.486	0.098			0.956	1282.1 (122)	0.081
			0.531	0.176	0.954	1225.6 (122)	0.082
					0.940	**94.7 (45)**	**0.070**
C–F	0.378	0.073			0.501	11.8 (37)	0.054
			0.444	0.114	0.484	10.7 (37)	0.055
C–S	−1.389	1.205			0.950	156.1 (19)	0.180
			0.221	1.662	0.950	158.5 (19)	0.179
C–Cl	0.179	0.397			0.770	45.2 (34)	0.370
			0.713	0.546	0.771	45.4 (34)	0.370
X–H	1.237	−0.086			0.942	5673.4 (723)	0.028
			1.142	−0.132	0.939	5384.1 (723)	0.029

Statistics for the model $n_{CM}^{Alt} = a + b \cdot exp\, \rho_c^{Alt}$. are in the first line; Statistics for the model $n_{CM}^{Alt} = a + b \cdot \left(\rho_c^{Alt}\right)^2$ are in the second line; Statistics obtained by Tsirelson et al. [41] are in bold in the third line; **N** is the number of bonds of a given type

been kept. But the main finding after applying these models was that the quality of the one-parameter exponential model for first-row atoms (Eq. 4) is better than the quality of the three-parameter bond order model suggested by Howard and Lamarche and improved by Tsirelson et al. However, the statistics of exponential model for the C–halogen bonds are unsatisfactorily poor what impels us to consider a series of two-parameter models comprising exponent of AlteQ electron density at bond critical points as the first term (Eq. 6–8). Since the quality of the one-parameter exponential model becomes worse with increasing of bond polarity (Table 1) we have chosen $\nabla^2 \rho_c^{Alt}$ at bond critical points as one of the second descriptors:

$$n_{CM}^{Alt} = a + b \cdot exp\, \rho_c^{Alt} + c \cdot \nabla^2 \rho_c^{Alt}, \qquad (6)$$

$$n_{CM}^{Alt} = a + b \cdot exp\, \rho_c^{Alt} + c \cdot \rho_{vac}^{Alt}, \qquad (7)$$

$$n_{CM}^{Alt} = a + b \cdot exp\, \rho_c^{Alt} + c \cdot \left(\rho_c^{Alt}\rho_{vac}^{Alt}\right)^2, \qquad (8)$$

where ρ_{vac}^{Alt} is the density of vacant orbital at the bond critical point. The parameters and statistics for these models are presented in Table 2. It is seen that quality of two-parameter models is better than one-parameter ones. But only model (6) describes all types of considered covalent bonds equally good whereas models (7–8) show unsatisfactory results for C–F bonds ($R = 0.696$ and $R = 0.660$, respectively). So this model has been kept and its detailed exploration has been carried out.

Model (6) fits well strongly polar (R > 0.98 for C–F) and nonpolar (R > 0.98 for C–C) bonds. Correlation coefficients for the C–O, C–S and X–H bonds are a little less (R > 0.94). The poorest results are observed for C–Cl bonds (R = 0.844). The agreement between theoretical n_{CM} (Eq. 1) and estimated n_{CM}^{Alt} (Eq. 6) bond order values for all considered bond types are shown in Fig. 1.

A training set for C–C bonds contains single, double, triple and aromatic bonds and n_{CM}-bond order values vary from 0.765 to 2.856. The deviation, $\Delta = n_{CM} - n_{CM}^{Alt}$, ranges from $\Delta = -0.372$ (in difluoroacetylene C_2F_2) to 0.250 (C_2H_4) (Fig. 1a). In general, the n_{CM}^{Alt}(C–C) values are underestimated for double bonds adjacent to electron-seeking groups (–OH, –COOH, –CN, –CHO, –Hal, etc.) and overestimated for triple bonds neighboring to the same groups.

A set of C–N bonds incorporates single bonds, double bonds in isothiocyanates and nitrile triple bonds. Observed range of the theoretical bond orders n_{CM} is 0.811–2.594. The most overestimated value n_{CM}^{Alt}(C–N) is observed in SCN–CH$_3$ ($\Delta = -0.421$) and the most underestimated value is predicted for C≡N bond in cyanic acid ($\Delta = 0.193$) (Fig. 1b).

A C–O bond sampling ($n_{CM} = 0.815$–1.864) includes bonds in hydroxyl, carbonyl, ketonic, carboxylic, ether, ester groups and bonds in heterocycles. According to outliers analysis (standard residual >2·σ) all outliers have only positive deviation of n_{CM}^{Alt} from n_{CM} values: the highest

Table 2 Summary of two-parameter Cioslowsky–Mixon bond order models

Bonds (N)	$n_{CM}^{Alt} = a + b \cdot exp\,\rho_c^{Alt} + c \cdot A$			$n_{CM}^{Alt} = a + b \cdot exp\,\rho_c^{Alt} + c \cdot B$			$n_{CM}^{Alt} = a + b \cdot exp\,\rho_c^{Alt} + c \cdot C$			Correlation coefficient R	Fisher statistic F	R. m. s. error S
	a	b	c	a	b	c	a	B	c			
C–C (286)	0.270	0.512	−0.124							0.988	6068.0	0.068
				−0.233	0.475	−0.101				0.989	6152.6	0.067
							−0.317	0.489	−0.050	0.989	6365.6	0.066
C–N (51)	1.899	0.298	−0.246							0.984	720.8	0.096
				0.650	0.204	−0.307				0.987	899.3	0.086
							0.274	0.240	−0.052	0.988	1013.6	0.081
C–O (122)	3.484	0.371	−0.369							0.965	798.8	0.073
				0.838	0.093	−0.220				0.965	814.7	0.072
							0.510	0.113	−0.018	0.964	788.8	0.073
C–F (37)	8.377	1.116	−0.842							0.983	485.3	0.012
				0.043	0.107	0.464				0.696	16.0	0.046
							0.223	0.089	0.220	0.660	13.1	0.048
C–S (19)	1.023	6.089	−2.926							0.966	113.7	0.152
				−0.410	1.035	−0.612				0.974	145.7	0.135
							−1.287	1.293	−0.501	0.978	179.0	0.122
C–Cl (34)	1.968	1.647	−0.909							0.844	38.3	0.032
				0.232	0.393	−0.126				0.853	41.4	0.031
							0.193	0.401	−0.226	0.880	53.0	0.028
X–H (723)	1.553	−0.165	−0.046							0.945	3002.8	0.028
				1.276	−0.071	−0.096				0.958	3992.1	0.024
							1.094	−0.032	−0.029	0.966	4974.3	0.022

Statistics for the model $n_{CM}^{Alt} = a + b \cdot exp\,\rho_c^{Alt} + c \cdot \nabla^2 \rho_c^{Alt}$ are in the first line; Statistics for the model $n_{CM}^{Alt} = a + b \cdot exp\,\rho_c^{Alt} + c \cdot \rho_{vac}^{Alt}$ are in the second line; Statistics for the model $n_{CM}^{Alt} = a + b \cdot exp\,\rho_c^{Alt} + c \cdot \left(\rho_c^{Alt} \cdot \rho_{vac}^{Alt}\right)^2$ are in the third line; **N** is the number of bonds of a given type. In the headers $A = \nabla^2 \rho_c^{Alt}$, $B = \rho_{vac}^{Alt}$ and $C = \left(\rho_c^{Alt} \cdot \rho_{vac}^{Alt}\right)^2$

$\Delta = 0.272$ for double C–O bonds in oxalic acid $C_2H_2O_4$. The maximum negative deviation of n_{CM}^{Alt} from n_{CM} observed in methyl chloroformate (the ester double C–O bond): $\Delta = -0.124$. How it was noted from Fig. 1c points depicting double C–O bonds in the carbonyl and ketonic groups are above the fit line whereas double C–O bonds adjacent to N, O, S atoms, for example, in amides and esters, are under the fit line.

The theoretical bond order values n_{CM} for C–F bonds range from 0.691 to 0.918, for C–Cl bonds range from 0.951 to 1.289. The maximum n_{CM}-value for the given types of C–halogen bonds is observed if halogen atoms are adjacent to triple C≡C bond and n_{CM}^{Alt} for both these bond types are overestimated. Figure 1d illustrates n_{CM} versus n_{CM}^{Alt} for C–F bonds. There can be discerned five subclasses in the figure. Subclass content analysis has revealed that each subclass comprises C–F bonds where the number of *gem*-fluorine atoms is the same. Thus, the first subclass includes only one substance—perfluoromethane; the second—compounds containing trifluoromethyl group; the third—compounds containing Csp3 or Csp2 with two adjacent fluorine atoms; the forth-compounds containing one fluorine atom bonded to Csp3 or Csp2. The fifth subclass encloses F–Csp bonds. And how it is expected, the n_{CM}-bond order for C–F(Cl) bonds decreases with increasing of polarity of contiguous substituents. In Fig. 1e describing n_{CM}^{Alt} versus n_{CM} for C–Cl bonds the most deviating point depicts C–Cl bond in methyl chloroformate.

Single C–S bonds in thiocyanates, sulfonic acids and its derivatives, disulfides, sulfoxides, etc., double C–S bonds in isothiocyanates, thiocarbonyl compounds, carbon disulfide and a triple bond in carbon monosulfide are included in the C–S bond sampling. Comparison of regression coefficients in Eq. 6 (Table 2) for all bond types has shown that contribution of electron density ρ_c^{Alt} and its laplacian $\nabla^2 \rho_c^{Alt}$ to n_{CM}^{Alt} for the C–S bonds is the highest. The observed n_{CM}-bond order values for C–S bonds vary from 0.849–2.778. The maximum overestimated value n_{CM}^{Alt}(C–S) observed for double C–S bonds in carbon disulfide CS_2 ($\Delta = -0.231$) and the maximum underestimated value is predicted for C–S bond in thioformaldehyde CH_2S ($\Delta = 0.285$) (Fig. 1f).
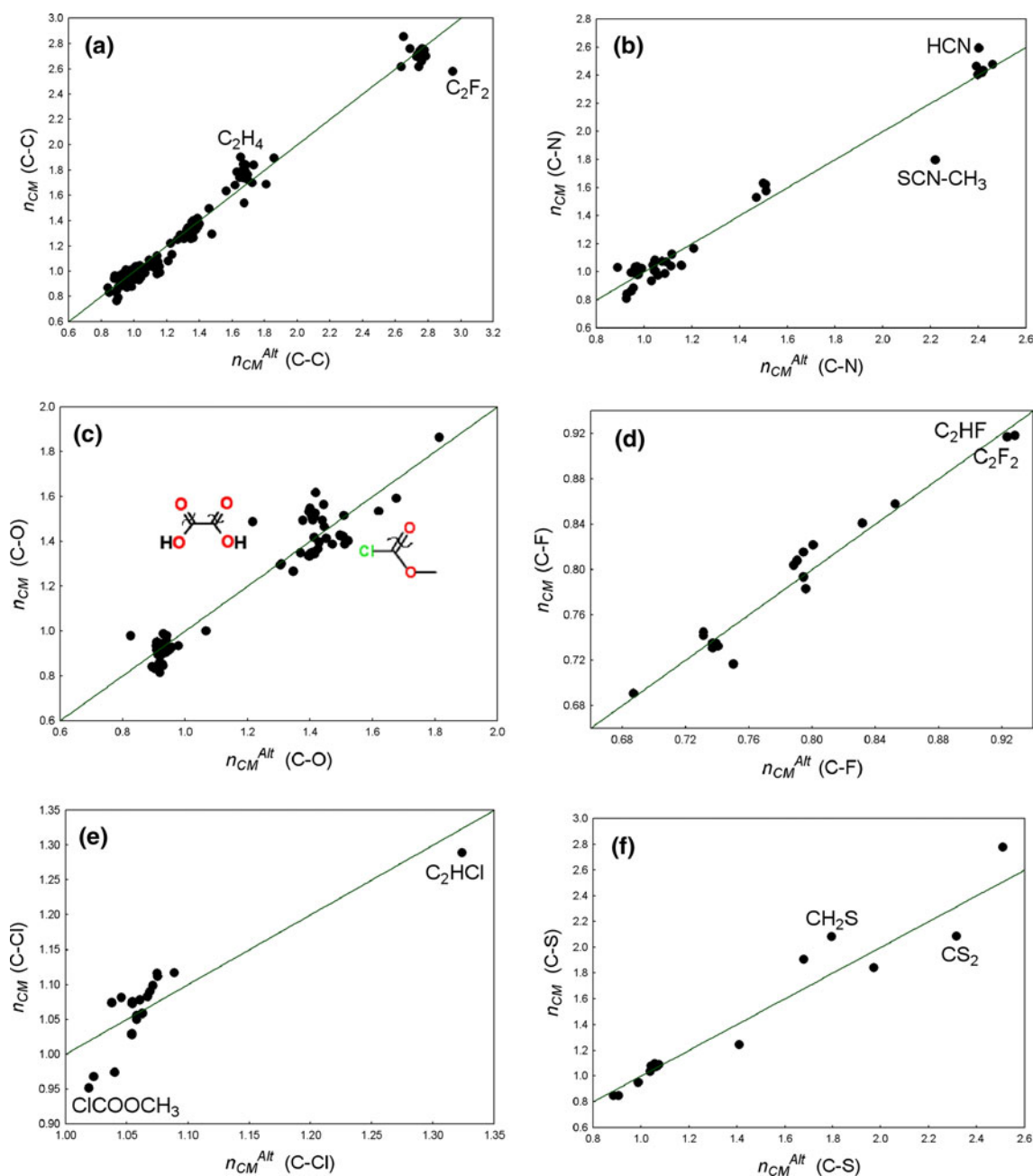
Fig. 1 Plot of the observed Cioslowski–Mixon bond orders n_{CM} versus predicted bond orders n_{CM}^{Alt} for the model $n_{CM}^{Alt} = a + b \cdot exp \, \rho_c^{Alt} + c \cdot \nabla^2 \rho_c^{Alt}$ (Eq. 6) for the next covalent bond types: **a** C–C; **b** C–N; **c** C–O; **d** C–F; **e** C–Cl; **f** C–S

All X–H bonds were combined to form shared training set (X=C, O, N and S). The aim was to unify description of this type of the covalent bonds. The minimum n_{CM}-value is observed for the O–H bond in peroxyacetic acid (0.565) and the maximum—for the S–H bond in β-unsaturated thiol (1.096). According to Fig. 2a, an application of model (6) to shared X–H bond sampling results in formation of well-defined subclasses. Each subclass comprises X–H bonds where X is the same atom. Thus, the first subclass contains O–H bonds. The n_{CM}-bond order values of these bonds are

the smallest ones amid X–H sampling and vary from 0.565 (in peroxyacetic acid CH$_3$COOOH) to 0.703 (in methanol CH$_3$OH). Second and third subclasses incorporate N–H and C–H bonds, respectively. Ranges of n_{CM}- values for these two subclasses overlap and are 0.752 (N–H bonds in the NH$_2$-group of monomethylhydrazine CH$_3$NHNH$_2$)–0.895 (N$_{sp2}$–H bond in α-aminoimine HN = CHNH$_2$) and 0.845 (in trifluoromethane, CHF$_3$)–0.984 (in methane CH$_4$). The n_{CM}-values are the highest for S–H bonds and range from 1.077 (in thiophenol C$_6$H$_5$SH) to 1.096 (in β-unsaturated

thiol C_4H_7SH). These bonds form fourth subclass. So, CMBO values of X–H bonds decrease with increasing of electronegativity of the X atom. However, except of subclass formation there is a bigger problem in application of model (6) to hydrogen-containing bonds, namely Y-axis-paralleled mode of point alignment in each subclass. So, the model is not "sensitive" to variation of bond order values inside the subclasses and it hardly can be applied for X–H bond order prediction in spite of statistically good quality of the model (Table 2). Therefore, we have applied the rest of proposed two-parameter models to X–H bonds. Plots of n_{CM} (X–H) versus n_{CM}^{Alt}(X–H) for models (7) and (8) are respectively presented in Fig. 2b, c. It is noted that, in comparison with model (6) (Fig. 2a), these two models work better for X–H bond order approximation, especially model (8) where tendency of points to orientate along the fitting line is obvious. However, in attempt to improve prediction of order of X–H bonds we have considered three- and four-parameter models. It is quite reasonable accounting that the X–H training set includes four types of the covalent bonds and shared description of them might require more descriptors. Since the model (6) gives poor results we have refused to include ρ_c^{Alt} as

an obligatory descriptor of n_{CM}-bond orders and, contrariwise, relatively good quality of model (8) makes the square of a product $(\rho_c^{Alt} \cdot \rho_{vac}^{Alt})^2$to be involved. Eventually, it has been found that the four-parameter regression equation (9) allows describing X–H bonds as a "single bond type" without explicit subset formation on the plot n_{CM} versus ρ_c^{Alt} (Fig. 2d):

$$n_{CM}^{Alt} = a + b \cdot (\rho_c^{Alt} \cdot \rho_{vac}^{Alt})^2 + c \cdot exp(\nabla^2 \rho_c^{Alt}) + d \cdot \sqrt{\frac{\rho_c^{Alt}}{\lambda_3}} + e \cdot r_c^{max},$$

(9)

where $a = -4.177$, $b = -0.023$, $c = -0.049$, $d = 11.401$ and $e = 3.616$; r_c^{max} is a distance between the nucleus of the atom X and the bond critical point on the X–H atomic interaction path; λ_3 is the principal curvature of the electron density function ρ_c^{Alt} at the bond critical point. Statistics for the model are R = 0.982, F = 4602.5, r.m.s. = 0.016 (the number of described X–H bonds is 723) and show that the model can be used for prediction of X–H bond order values with satisfactory extent of accuracy.
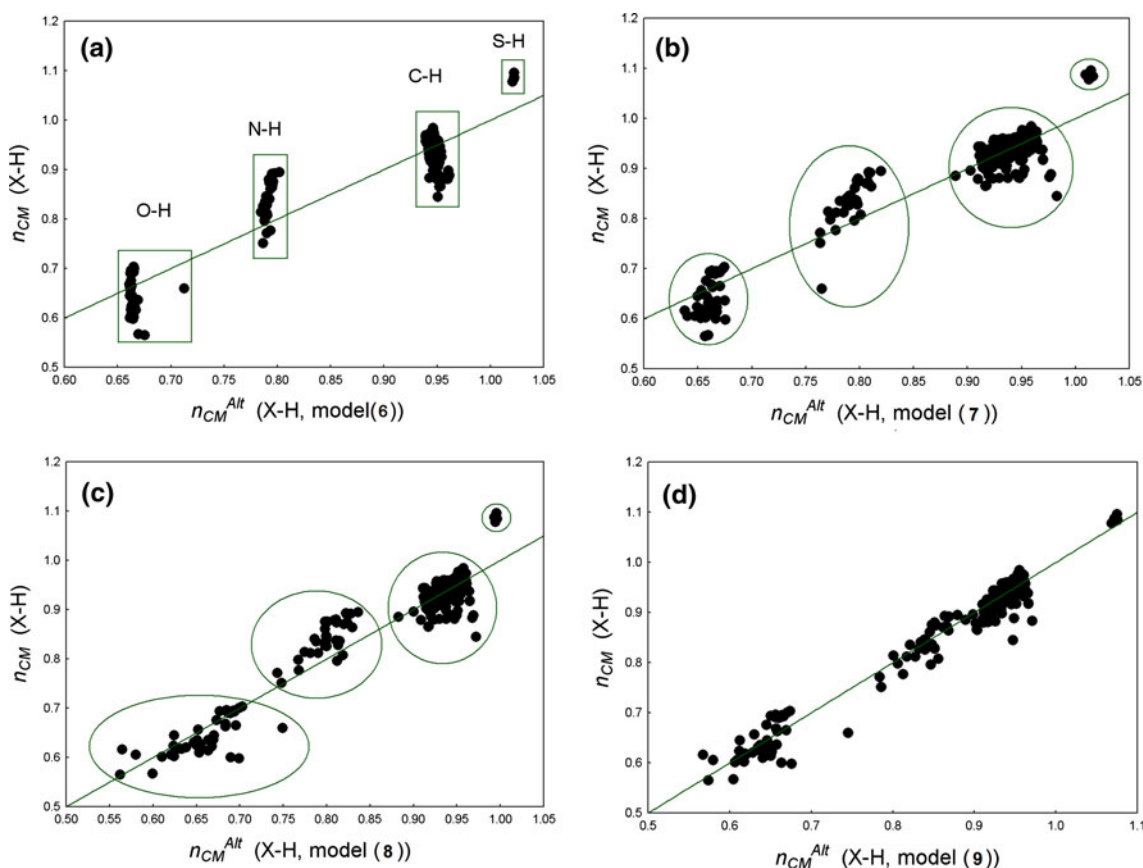

Fig. 2 Plot of the observed Cioslowski–Mixon bond orders n_{CM} versus the predicted bond orders n_{CM}^{Alt} for the X–H bonds, where X is C, N, O and S atom. The number of the model which is used for n_{CM}^{Alt}-bond order calculation is in the X-axis title. Green figures on the plots a–c enclose subclasses containing O–H, N–H, C–H and S–H bonds. Order of the subclasses on the plots a–c are the same

Table 3 Observed Cioslowski–Mixon bond orders n_{CM} and predicted bond orders using AlteQ electon density n_{CM}^{Alt} and experimental electron density n_{CM}^{exp} characteristics for covalent bonds in 5-acetyl-4-ethyl-6-methyl-1,2,3,4-tetrahydropyrimidine-2-thione (AEMTT), ethyl-4,6-dimethyl-2-thioxo-1,2,3,4-tetrohydropyrimidine-5-carboxylate (EDTTC) and 5-acetyl-4,6-dimethyl-1,2,3,4-tetrahydropyrimidine-2-one (ADT)
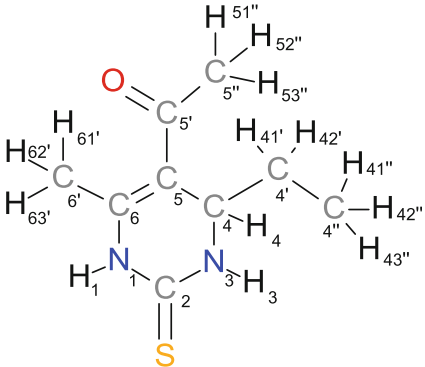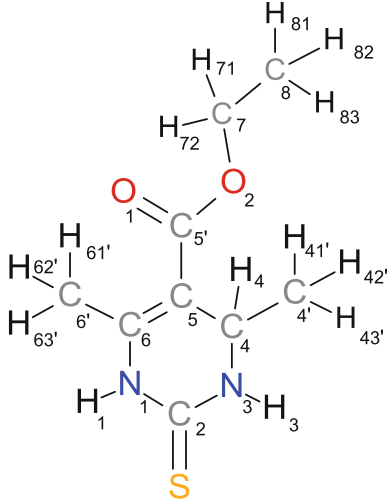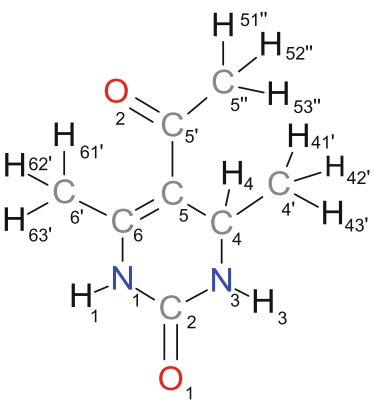
Molecule	Bond	n_{CM}	n_{CM}^{Alt} (Eq. 6)	n_{CM}^{exp} (Eq. 6)	n_{CM}^{Alt} (Eq. 4)	n_{CM}^{exp} (Eq. 4)
5-Acetyl-4-ethyl-6-methyl-1,2,3,4-tetrahydropyrimidine-2-thione (AEMTT)	S–C2	1.520	1.322	10.843	1.524	1.824
	N1–C2	1.047	1.107	2.521	1.151	1.326
	N3–C2	1.166	1.251	2.733	1.303	1.603
	N3–C4	0.901	0.961	2.229	0.929	0.924
	C4–C5	0.954	0.983	1.725	0.978	1.371
	C5–C6	1.518	1.495	2.659	1.511	2.365
	N1–C6	1.047	1.073	2.390	1.098	1.146
	C6–C6′	0.992	1.011	1.719	1.011	1.362
	C6′–H61′	0.939	0.925	0.779	0.919	0.909
	C6′–H62′	0.953	0.925	0.794	0.919	0.917
	C6′–H63′	0.939	0.923	0.793	0.917	0.917
	C5–C5′	1.059	1.107	1.863	1.111	1.526
	O–C5′	1.412	1.283	4.855	1.283	1.234
	C5′–C5″	0.959	1.003	1.683	1.001	1.328
	C5″–H51″	0.949	0.925	0.836	0.919	0.938
	C5″–H52″	0.948	0.925	0.791	0.919	0.917
	C5″–H53″	0.945	0.925	0.811	0.919	0.925
	C4–H4	0.886	0.942	0.731	0.943	0.891
	C4–C4′	0.935	0.933	1.575	0.921	1.205
	C4′–H41′	0.931	0.943	0.824	0.943	0.932
	C4′–H42′	0.938	0.943	0.805	0.943	0.924
	C4′–C4″	0.987	0.960	1.631	0.953	1.269
	C4″–H41″	0.964	0.925	0.842	0.920	0.939
	C4″–H42″	0.963	0.925	0.817	0.919	0.928
	C4″–H43″	0.959	0.925	0.800	0.920	0.921
Ethyl-4,6-dimethyl-2-thioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate (EDTTC)	S–C2	1.520	1.318	11.530	1.518	2.039
	N3–C2	1.166	1.246	2.664	1.296	1.485
	N1–C2	1.043	1.122	2.455	1.164	1.270
	N3–C4	0.905	0.960	2.235	0.930	0.963
	C4–C5	0.960	0.984	1.644	0.978	1.280
	C5–C6	1.532	1.522	2.494	1.538	2.180
	N1–C6	1.055	1.073	2.361	1.101	1.150
	N1–H1	0.762	0.787	0.624	0.796	0.858
	N3–H3	0.781	0.789	0.663	0.799	0.875
	C6–C6′	0.993	1.021	1.718	1.022	1.365
	C6′–H61′	0.940	0.926	0.903	0.920	0.971
	C6′–H62′	0.956	0.924	0.904	0.919	0.972
	C6′–H63′	0.929	0.923	0.910	0.919	0.975
	C5–C5′	1.018	1.101	1.903	1.108	1.567
	O1–C5′	1.295	1.324	5.096	1.333	1.332
	O2–C5′	0.886	0.958	3.954	1.016	0.986
	O2–C7	0.825	0.934	3.655	0.851	0.806
	C7–H71	0.912	0.945	0.807	0.944	0.927
	C7–H72	0.914	0.945	0.830	0.945	0.937
	C7–C8	0.982	0.994	1.704	0.994	1.353
	C8–H81	0.960	0.925	0.821	0.920	0.933
	C8–H82	0.961	0.926	0.850	0.920	0.946
	C8–H83	0.963	0.924	0.866	0.918	0.953
	C4–C4′	0.947	0.943	1.605	0.935	1.232
	C4′–H41′	0.947	0.925	0.858	0.919	0.947
	C4′–H42′	0.961	0.926	0.844	0.920	0.941
	C4′–H43′	0.960	0.925	0.835	0.919	0.937

Table 3 continued

Molecule	Bond	n_{CM}	n_{CM}^{Alt} (Eq. 6)	n_{CM}^{exp} (Eq. 6)	n_{CM}^{Alt} (Eq. 4)	n_{CM}^{exp} (Eq. 4)
	O1–C2	1.263	1.268	4.566	1.294	1.213
	N1–C2	0.907	1.070	2.491	1.136	1.277
	N1–C6	1.060	1.083	2.411	1.125	1.148
	C5–C6	1.541	1.558	2.542	1.574	2.232
	C4–C5	0.945	0.981	1.718	0.980	1.352
	N3–C4	0.904	0.951	2.248	0.950	0.923
	N3–H3	0.792	0.564	0.473	0.264	0.774
	N1–H1	0.777	0.621	0.515	0.403	0.794
	C4–H4	0.922	0.853	0.787	0.808	0.914
	C4–C4′	0.927	0.973	1.549	0.984	1.174
	C4′–H41′	0.974	0.851	0.765	0.807	0.905
	C4′–H42′	0.966	0.853	0.840	0.811	0.939
	C4′–H43′	0.958	0.875	0.831	0.849	0.935
	C5′–C5″	1.057	1.102	1.718	1.105	1.360
	O2–C5′	1.405	1.321	4.935	1.318	1.308
	C5′–C5″	0.943	1.012	1.718	1.022	1.360
	C5″–H51″	0.955	0.878	0.810	0.851	0.924
	C5″–H52″	0.959	0.875	0.757	0.846	0.899
	C5″–H53″	0.957	0.875	0.847	0.847	0.940
	C6–C6′	0.962	1.009	1.757	1.018	1.403
	C6′–H61′	0.970	0.824	0.821	0.759	0.928
	C6′–H62′	0.951	0.870	0.806	0.845	0.922
	C6′–H63‴	0.953	0.845	0.792	0.802	0.916

5-Acetyl-4,6-dimethyl-1,2,3,4-tetrahydropyrimidine-2-one (ADT)

The number of the model using for the n_{CM}^{Alt} and n_{CM}^{exp} bond order calculations is in parenthesis in the headers

Application of developed regression models for bond order prediction from experimental electron density

Application of proposed regression models for bond order prediction directly from low-temperature high-resolution X-ray diffraction experiment has been estimated utilizing data for 5-acetyl-4-ethyl-6-methyl-1,2,3,4-tetrahydropyrimidine-2-thione $C_9H_{13}N_2OS$ [45], ethyl-4,6-dimethyl-2-thioxo-1,2,3,4-tetrohydropyrimidine-5-carboxylate $C_9H_{14}N_2O_2S$ [46] and 5-acetyl-4,6-dimethyl-1,2,3,4-tetrahydropyrimidine-2-one $C_8H_{12}N_2O_2$ (ADT) [47]. For all structures the standard X–H bond distances tabulated in *International Tables for Crystallography* (1995) were employed. Experimental electron density was presented in terms of the Hansen and Coppens [48] multipole model. Multipole parameters used in this work were derived from X-ray diffraction data at 100–120 K by applying the same refinement strategy. Results of application of Eq. 6 to these molecules are shown in Table 3.

It was repeatedly noted in literature that there is a significant difference in the multipole-model-based and exponent-function-based Laplacian of the electron density $\nabla^2\rho$ for the shared atomic interactions [45, 49, 50]. Since Eq. 6 includes the Laplacian $\nabla^2\rho$ as an independent variable discrepancies between bond orders predicted using AlteQ electron density n_{CM}^{Alt} and experimental electron density n_{CM}^{exp} were expected. To estimate an extent of discrepancy contributed by the Laplacian term and taking into account that model (4) works well for all bond types in the molecules we have applied model (4) to the molecules as well (Table 3). Plots of observed Cioslowski–Mixon bond orders (CMBO) n_{CM} versus calculated ones n_{CM}^{exp} for the C–C, C–N and C–O bonds are shown in Fig. 3. Because of insufficient amount of experimental data plots for the C–S and X–H (where X = C, N) bonds are not provided. According to the data point alignment in Fig. 3, in general model (4) predicts n_{CM}-values better than model (6) and the discrepancy between calculated and authentic n_{CM}-values gets smaller with increasing of bond polarity. Thus, for the C–C bonds a constant overestimation of the n_{CM}^{exp}-values in comparison with n_{CM}-values is observed (Fig. 3b) and the best agreement between n_{CM}^{exp} and n_{CM}-bond orders is presented for the C–O bonds (Fig. 3f). For model (6) an opposite situation appears: the maximal difference between n_{CM}^{exp} and n_{CM}-values is observed for the C–O bonds (Fig. 3e) and the minimum one—for the C–C bonds (Fig. 3a). So, with a growth of bond polarity the extent of discrepancy contributed by the Laplacian term $\nabla^2\rho$ is increasing and for the non-polar C–C bonds models (4) and (6) give quite similar results (Fig. 3a, b). The most outlaying point in the Fig. 3 is observed for $C_2–N_1$ bond in

ADT. For the other C–N bond in the urea fragment of tetrahydropyrimidinone (C_2–N_3) such deviation is not registered. According to the presented data $n_{CM}(C_2–N_3) − n_{CM}(C_2–N_1) = 0.121$ when the difference in experimental electron densities ρ_{CP}^{exp} (C_2–N_3) − ρ_{CP}^{exp} (C_2–N_1) is 0.009 a.u. Consideration of other bonds in the molecules has shown that the change in ρ_{CP}^{exp} should be bigger than observed one. Such discrepancy could be explained as follows: the electron-withdrawing carbonyl group in

tetrahydropyrimidinone ring affects equally both nitrogen atoms (N_1, N_3). However electron-donating properties of N_1 are decreased because its lone-pair is conjugated with double bond of the ring. So electron density on the C_2–N_1 bond should be decreased what leads to decreasing of the bond order. For the N_3 atom such mesomeric effect is not possible. Taking into account that n_{CM}-bond orders are calculated for optimized molecular geometry (DFT B3LYP/6-311G(d,p)) in vacuum all electronic effects have
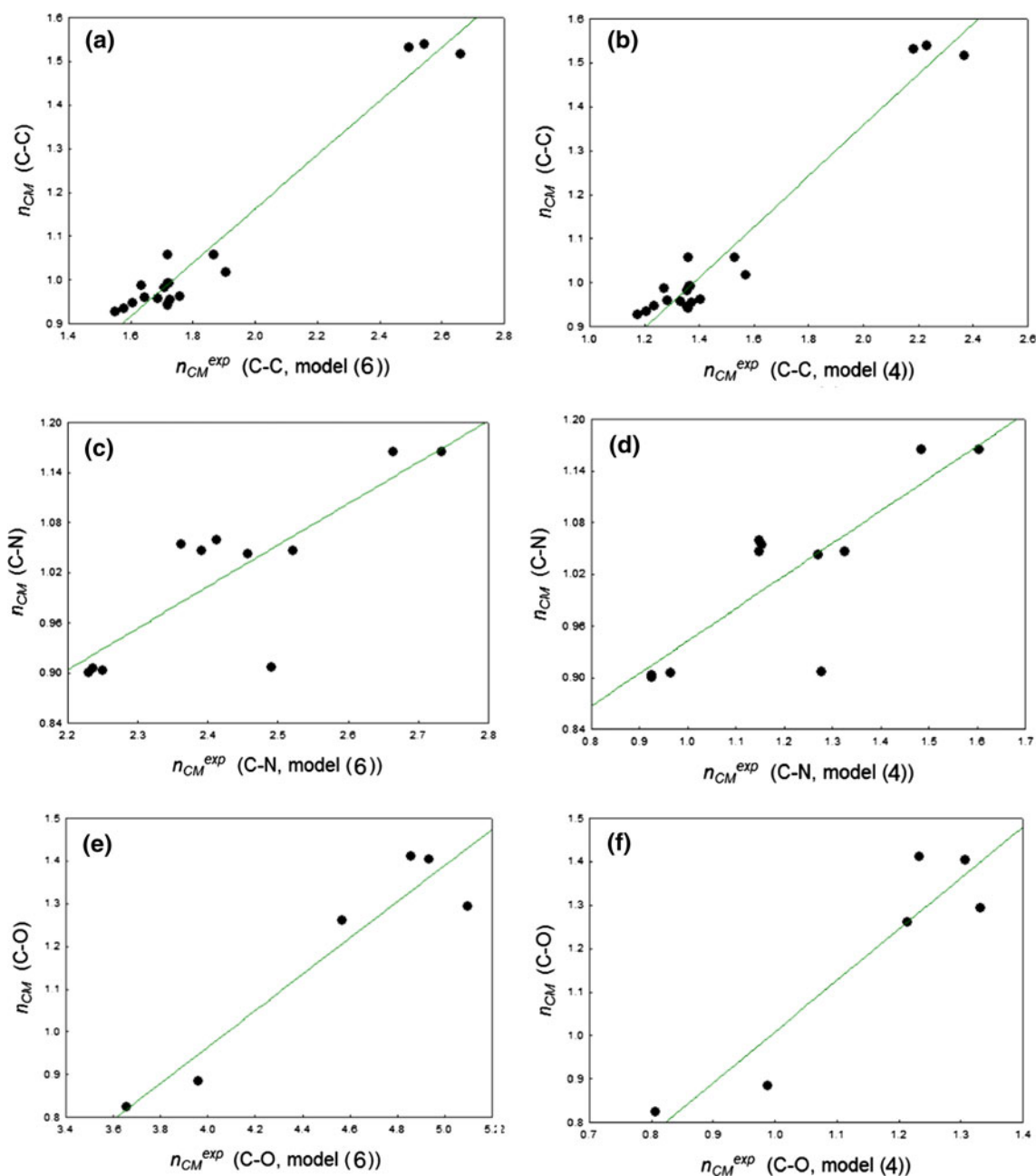

Fig. 3 Plot of the observed Cioslowski-Mixon bond orders n_{CM} versus the bond orders n_{CM}^{exp} predicted from experimental electron density characteristics for the next bond types: **a**, **b** C–C; **c**, **d** C–N; **e**, **f** C–O. The number of the model which is used for n_{CM}^{exp}-bond order calculation is in the X-axis title

an influence upon calculated bond orders. In crystal systems such effects could be leveled by surrounding molecules and the essential difference in electron densities on these bonds do not appear.

Figure 4 shows an agreement between the bond order values calculated from experimental n_{CM}^{exp} and AlteQ n_{CM}^{Alt} electron density characteristics for models (6) and (8). Analysis of Fig. 4 has shown that n_{CM}^{exp} and n_{CM}^{Alt} values calculated using model (6) are well correlated for all bond

types in question what should be expected provided an existent linear correlation between the AlteQ and experimental electron densities. Correlation between n_{CM}^{Alt} and n_{CM}^{exp} using model (8) gets worse with an increasing of bond polarity what one more time confirms the existent discrepancy in experimental and calculated Laplacian values.

Thus, an application of model (8) for bond order determination from experimental electron density features for polar bonds is restricted provided the discrepancy
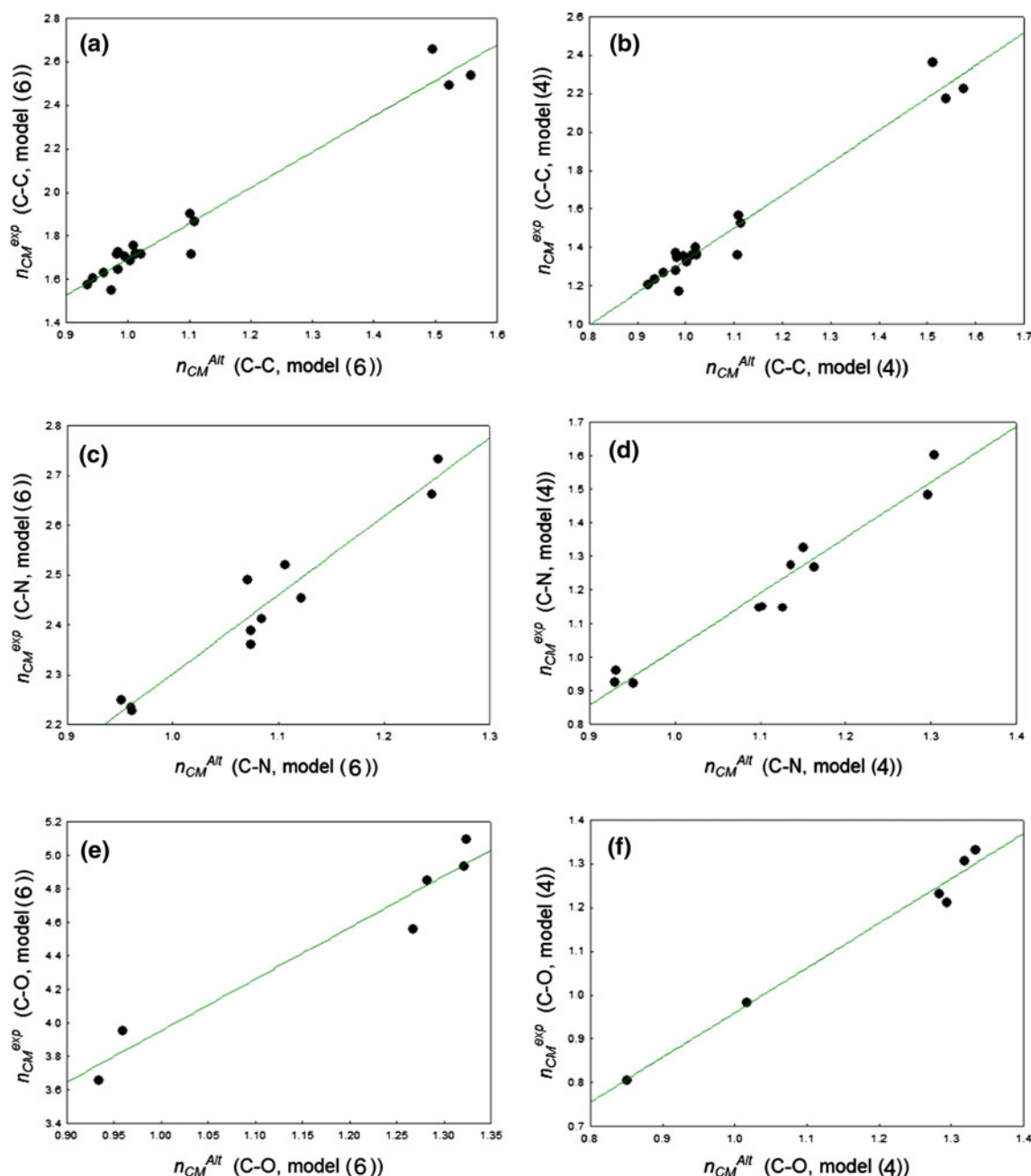

Fig. 4 Plot of the bond orders predicted from experimental electron density characteristics n_{CM}^{exp} versus the bond orders predicted from AlteQ electron density characteristics n_{CM}^{Alt} for the next bond types: **a, b** C–C; **c, d** C–N; **e, f** C–O. The number of the model which is used for the n_{CM}^{exp}- and n_{CM}^{Alt}-bond order calculation is in the X-axis title

between calculated and experimental Laplacian values. Model (6) can be applied for estimation of bond orders from experimental data for polar covalent bonds namely C–O and C–N. At the same time comparative analysis of the data in Table 3 and Fig. 4 has proved that AlteQ electron density characteristics can be used for CMBO approximation. A quality of such prediction statistically relevant and the prediction can be made very fast for large molecular ensembles. Moreover, in the AlteQ workflow routine X-ray diffraction experiment geometries can be utilized as input data what allows making comparative analyses of intra- and intermolecular interaction in crystal systems.

Conclusion

An approximation of bond order values determined in the framework of suitable theoretical technique by experimentally measured quantities is one of the ways to make bond orders "observed quantities". In this study we have proposed two regression models interrelating Cioslowsky-Mixon bond orders n_{CM} (calculated at DFT B3LYP/6-311G(d, p) level of theory) with topological properties of AlteQ electron density. These models allow approximating CMBOs with a high extent of accuracy for the following types of covalent bonds: single, double, triple and aromatic C–C bonds; single, double, triple and aromatic C–N bonds; single, double and aromatic C–O bonds; single, double and triple C–S bonds; C–Cl bonds; C–F bonds and X–H bonds (where X is C, N, O and S). One regression model is utilized for description of non-hydrogen containing covalent bonds and another one approximates X–H bonds. It has been found that the usage of AlteQ electron density characteristics provides a bonding quantification that is in reasonable agreement with that obtained by orbital theory. Even one parameter model comprising exponent of AlteQ electron density at bond critical points ρ_c^{Alt} as an independent variable has better quality (Table 1) than the model proposed in [41] and involved more quantum-chemical characteristics.

Application of proposed regression models for determination of the CMBOs from topological features of experimental electron density has been tested. It was found out that usage of two-parameter regression Eq. (8) is restricted because of the existent discrepancy in the multipole-model-based and exponent-function-based Laplacian of the electron density $\nabla^2\rho$. One-parameter model (6) containing exponent of electron density as an independent variable allows gaining bond orders values which are similar to authentic ones (especially for polar bonds) and close to bond order values calculated from the AlteQ

electron density features. This has proved that topological features of AlteQ electron density can be used for CMBO prediction. Fastness of such prediction makes is attractive for studying connectivity in large molecular ensembles. Moreover, routine X-ray crystallographic geometries can be utilized in the AlteQ workflow as input data what allows making comparative analyses of intra- and intermolecular interaction in crystal systems.

References

1. Lendvay G (1994) J Phys Chem 98:6098–6104
2. Kiralj R, Kojic-Prodic B, Zinic M, Alihodzic S, Trinajstic N (1996) Acta Cryst B 52:823–837
3. Blowers P, Masel RI (1998) J Phys Chem A102:9957–9964
4. Zhan C-G, Wang Q-L, Xiong Y, Chen XJ (2000) J Mol Struct Theochem 531:33–37
5. Nalewajski RF (2000) J Phys Chem A 104:11940–11951
6. Lewis GN (1916) J Am Chem Soc 38:762–785
7. Lewis GN (1923) Valence and structure of atoms and molecules. Chemical Catalogue Co., New York
8. Frankland E (1852) Phil Trans R Soc Lond 142:417–444
9. Kekulé A (1857) Annalen der Chemie und Pharmacie 104:129–150
10. Kekulé A (1858) Annalen der Chemie und Pharmacie 106:129–159
11. Kekulé A (1866) Annalen der Chemie und Pharmacie 137:129–136
12. Couper AS (1858) Annales de Chimie et de Physique 53:469–489
13. Thomson JJ (1897) Phil Mag S 5(44):293–316
14. Bohr N (1913) Phil Mag S 6(26):1–24
15. Bohr N (1913) Phil Mag S 6(26):476–502
16. Bohr N (1913) Phil Mag S 6(26):857–875
17. Burrau Ø (1927) Danske Vidensk Selsk Math-fys Medd 7(14):1–18
18. Burrau Ø (1927) Naturwissenschaften 15:16–17
19. Heitler W, London F (1927) Zeitschrift für Physik 44:455–472
20. Lennard-Jones JE (1929) Trans Faraday Soc 25:668–686
21. Cioslowski J, Mixon ST (1991) J Am Chem Soc 113(42):4142–4145
22. Coulson CA (1939) Proc R Soc Lond A 169:413–428
23. Mulliken RS (1955) J Chem Phys 23:1833–1840
24. Mulliken RS (1955) J Chem Phys 23:1841–1846
25. Mulliken RS (1955) J Chem Phys 23:2338–2342
26. Mulliken RS (1955) J Chem Phys 23:2343–2346
27. Wiberg KB (1968) Tetrahedron 24:1083–1096
28. Mayer I (1983) Chem Phys Lett 97:270–274
29. Mayer I (1986) Int J Quantum Chem 29:477–483
30. Kaufman JJ (1970) Int J Quantum Chem 5:205–208
31. Ham NS (1958) J Chem Phys 29:1229–1231
32. Angyan JG, Loos M, Mayer I (1994) J Phys Chem 98:5244–5248
33. Angyan JG, Rosta E, Surjan PR (1999) Chem Phys Lett 299:1–8
34. Bridgeman AJ, Cavigliasso G, Ireland LR, Rothery JJ (2001) J Chem Soc Dalton Trans 14:2095–2108
35. Mayer I, Knapp-Mohammady M, Suhai S (2004) Chem Phys Lett 389:34–38
36. Lennard-Jones JE (1937) Proc R Soc Lond A 158:280–296

37. Bader RFW (1990) Atoms in molecules: a quantum theory. Oxford University Press, Oxford
38. Bader RFW, Slee TS, Cremer D, Kraka E (1983) J Am Chem Soc 105:5061–5068
39. Howard ST, Lamarche O (2003) J Phys Org Chem 16:133–141
40. Stash AI, Tanaka K, Shiozawa K, Makino H, Tsirelson VG (2005) Acta Cryst B 61:418–428
41. TsirelsonVG, Bartashevich EV, Stash AI, Potemkin VA (2007) Acta Cryst B 63:142–150
42. Potemkin VA, Grishina MA (2008) J Comput Aided Mol Des 22:489–505
43. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA Jr (1993) J Comp Chem 14:1347–1363
44. AIMPAC suite of software applications http://www.chemistry.mcmaster.ca/aimpac/imagemap/imagemap.htm. Accessed April 8, 2013
45. Potemkin VA, Rykounov AA, Bartashevich EV, Pereyaslavskaya ES, Stash AI, Tsirelson VG (2006) IV National Crystal Chemistry Conference, 26–30 June 2006, Russia, Chernogolovka: Book of Abstracts, p 75
46. Tsirelson VG, Stash AI, Potemkin VA, Rykounov AA, Shutalev AD, Zhurova EA, Zhurov VV, Pinkerton AA, Gurskaya GV, Zavodnik VE (2006) Acta Cryst B 62:676–688
47. Potemkin VA, Rykounov AA, Stash AI, Tsirelson VG (2006) IV National Crystal Chemistry Conference, 26–30 June 2006, Russia, Chernogolovka: Book of Abstracts, p 130
48. Hansen NK, Coppens P (1978) Acta Cryst A 34:909–921
49. Zhurova EA, Matta CF, Wu N, Zhurov VV, Pinkerton AA (2006) J Am Chem Soc 128:8849–8862
50. Hibbs DE, Overgaard J, Howard ST, Nguyen TH (2005) Org Biomol Chem 3:441–447