Amino Acids (1994) 7: 19–26

Amino
Acids

Libraries of atomic multipole moments for precise modeling of electrostatic properties of amino acids

Minireview Article

W. Sokalski

Institute of Physical and Theoretical Chemistry, Wrocław, Poland

Accepted September 9, 1993

Summary. Contemporary theoretical models used in describing electrostatic properties of amino acids in polypeptides rely usually on atomic point charges. Recently noted defects of such models in reproducing protein folding originate from the inadequate representation of the electrostatic term, in particular inability of atomic charges to account for local anisotropy of molecular charge distribution. Such defects could be corrected by multicenter multipole moments derived directly from any high quality quantum chemical wavefunctions. This is illustrated by comparison of monopole and multipole electrostatic interactions between some amino acids within glutathione S-transferase.

High quality Point Charge Models (PCM) can be derived analytically from multipole moment databases. Preliminary results suggest that torsional potentials are controlled by electrostatic interactions of atomic multipoles.

Examples illustrating various uses of multicenter multipole moment databases of protein building blocks in modeling various properties of amino acids and polypeptides have been described, including calculation of molecular electrostatic potentials, electric fields, interactions between amino acid residues, estimates of pK_a shifts and changes in catalytic activity induced by amino acid substitutions in mutated enzymes.

Keywords: Amino acids – Atomic multipoles – Electrostatic effects

Introduction

The structure and many properties of proteins are controlled by interactions of their amino acid residues. Intramolecular interactions due to bond stretching, bending and rotation as well as nonbonded interactions are frequently modeled by various simple empirical force fields applied in molecular mechanics or

dynamic calculations for proteins. However, recent studies by Gilson et al. (1988) and by Roterman et al. (1989a) show that the use of alternative force fields may sometimes yield different results, at variance with experiment. This may be due to arbitrary analytical form assumed for all components of the force field and lack of possibility to verify parameterization of individual terms. More detailed analysis of this problem by Roterman et al. (1989b) indicates that the main source of these differences originates from inadequate representation of the electrostatic term. This is due to strongly anisotropic nature of atomic charge distribution in molecules, neglected in typical force fields employing isotropic atomic point charges only. One of the possible ways to correct this problem is to supplement the standard atomic monopole (point charge) model by higher terms in atomic multipole expansion (i.e., atomic dipoles, quadrupoles, etc.) derived directly from quantum chemical wave functions. The corresponding techniques are outlined in the Methods section. The importance of higher atomic multipoles has been convincingly demonstrated by Buckingham et al. (1985) by precise prediction of angular characteristics of many van der Waals complexes. In addition, atomic multipole moments for larger functional groups are reasonably transferable between different molecules (see Liang et al., 1986; Price et al., 1992). This allows to construct multipole models for large biopolymers from data collected for smaller building blocks.

Next sections briefly review currently available atomic multipole databases for amino acids and various possible applications of multicenter multipole moments in modeling electrostatic properties of protein.

Methods

In principle the detailed description of the molecular charge distribution is contained in the accurate quantum chemical wavefunction. However, such calculations require use of well saturated basis sets with polarization functions and including of correlation effects (see Sokalski and Sawaryn, 1987). In practice for a long time such precise nonempirical calculations in extended basis sets have been limited to molecules smaller than most amino acids (less than about 200 atomic orbitals). Thanks to significant progress in computational methods of quantum chemistry as well advances in computer hardware present limit for ab initio calculations has been shifted to about 1800 atomic orbitals (see Ciosłowski, 1993). This way the accurate nonempirical wave functions for amino acids and small peptides started to be routinely available only very recently. However, due to the extensive size of wavefunctions permanent storing such huge files is not practical. On the other hand the properties calculated in conventional calculations are frequently limited to atomic charges and lowest molecular multipole moments. Unfortunately neither atomic charges nor low order molecular multipole moments are sufficient to yield precise description of local anisotropy of molecular charge distribution and fail to describe molecular interactions at distances close to equilibrium geometry. Fortunately as it has been shown by Rein (1973) the divergence of molecular multipole expansion could be overcomed by partitioning molecules into smaller additive segments: atoms, bonds, etc. and extracting higher multipole moments of these fragments from molecular

wavefunction. These segmental moments add up to global moments in cumulative fashion allowing much more precise representation of fine details of local charge distribution. On the same time the higher atomic multipoles (dipoles, quadrupoles, octupoles, hexadecapoles, etc.) supplement in natural way any inherently arbitrary definition of atomic charge (Sokalski et al., 1992). In addition atomic multipoles constitute compact extract from quantum chemical wavefunction suitable for creating corresponding databases. Among several different methods for calculation of multicenter multipole moments the most frequently applied techniques for biomolecules are Overlap Multipole Expansion Procedure (OMTP) by Goldblum et al. (1979), Distributed Multipole Analysis (DMA) proposed by Stone (1981) and Cumulative Atomic Multipole Moments (CAMM) described by Sokalski and Poirier (1983). According to comparative study of Spackman (1986) CAMM and DMA techniques yield numerically equivalent results, although their features, analytical form and derivation are entirely different. DMA moments are generated for pairs of primitive Gausian orbitals in the form of spherical tensors and then transformed to chosen expansion centers. In the case of CAMM, atomic charges are augmented by higher cartesian moments reproducing in cumulative manner corresponding global molecular moments. This unique feature allows to derive CAMM for any arbitrary definition of atomic charge (Sokalski et al., 1992).

It has to be mentioned that atomic multipoles can be decomposed into off-atom point charges. Such Point Charge Models (PCM) can be derived analytically (Sokalski et al., 1987) or by approximate fitting procedure proposed by Ferenczy et al. (1991). Quite popular alternative procedure involves fitting atomic charges to grid of quantum chemical potentials calculated around molecule (see Williams, 1991). Although such Potential Derived charges (PD) include effectively some higher multipole moments, they still fail to represent local anisotropies and are inherently arbitrary and approximate (Sokalski et al., 1992). In contrast to atomic multipole calculations, calculation of PD charges becomes very costly for large molecular systems and results are less transferable (see Chipot et al., 1993).

Recent advances in high resolution X-ray crystallography by Lecomte et al. (1992) made possible satisfactory comparisons between properties calculated from theoretical and experimental atomic multipole expansions for several peptides. However, such experimental studies require exceptional accuracy and are still quite rare. Therefore it seems probable that at least for some time high quality theoretical calculations (ab initio LCAO MO SCF, MP2, MRD-CI, etc.) will be the main source of information regarding charge distribution in aminoacids and other related molecules.

Atomic multipole databases for amino acids

The chronology of the development of nonempirical databases of molecular charge distribution for aminoacids and blocked peptides has been presented in Table 1. The earliest complete libraries of nonempirical atomic charges for amino acids, precursors of multipole databases, have been published by Clementi et al. (1977) and Sheridan et al. (1981) and later by Bellido et al. (1989). However,

Table 1. Nonempirical databases describing charge distributions in amino acids and blocked peptides

Authors	Theory	Basis set	Multipole level
Clementi et al. (1977)	SCF	minimal	q
Sheridan et al. (1981)	SCF	minimal	q
Sokalski et al. (1989)	SCF	minimal	CAMM(q, μ, Θ)
Bellido et al. (1989)	SCF	DZ	q
Price et al. (1991)	SCF	3-21G	DMA(q, μ, Θ, Ω)
Sokalski et al. (1992)	SCF	3-21G	CAMM(q, μ, Θ, Ω, Φ)
Sokalski et al. (1993)[†]	MP2	6-31G*	CAMM(q, μ, Θ, Ω, Φ)

[†] N-acetyl, N′-methylamide-blocked derivatives of glycine, alanine, cysteine, threonine, leucine, lysine and serine

as explained in Introduction, atomic charges alone are not always sufficient for more demanding applications. Therefore parallel with the progress in the efficiency of computational methods and hardware one may notice continuing quest for better quality databases. The first atomic multipole (CAMM) database for aminoacids has been published by Sokalski et al. (1989), whereas Faerman et al. (1990) and Price et al. (1992) published first (DMA) database for blocked amino acid residues. Earlier systematic studies for smaller systems (Sokalski and Sawaryn, 1987) indicate that for more precise representation of molecular charge distribution requires inclusion of correlation effects and the use of extended basis sets with polarization functions. Therefore our efforts have been recently directed to collecting correlated CAMM for N-acetyl, N′-methylamide-blocked derivatives of several smaller amino acids in 6-31G* basis set (Sokalski et al., 1993).

Atomic multipole applications

One of the most frequently calculated molecular properties are Electrostatic Molecular Potentials (EMP), representing the unit point charge interaction energy with the given molecule. This quantity calculated for example on the van der Waals surface illustrates unique characteristics of molecule closely related with its specific recognition and binding to receptors and active sites. However, direct calculations of EMP from wavefunction of large molecules could be very costly. Alternatively, popular atomic charge models frequently yield qualitatively incorrect results. This problem can be effectively corrected by the use of higher atomic multipoles (Sokalski and Poirier, 1983; Price et al., 1989; Sokalski and Sneddon, 1991; Sokalski et al., 1993). Analogous role of higher atomic multipoles has been observed in calculations of electrostatic fields (Price et al., 1991).

The crucial role of the quality of electrostatic term in structural predictions has been already noted in calculations for small molecular complexes (Buckingham et al., 1985) and for polypeptides (Roterman et al., 1989b). In order to examine the role of neglected higher multipole moments relative electrostatic interaction energies between cysteine 219 and various neutral amino acids lo-

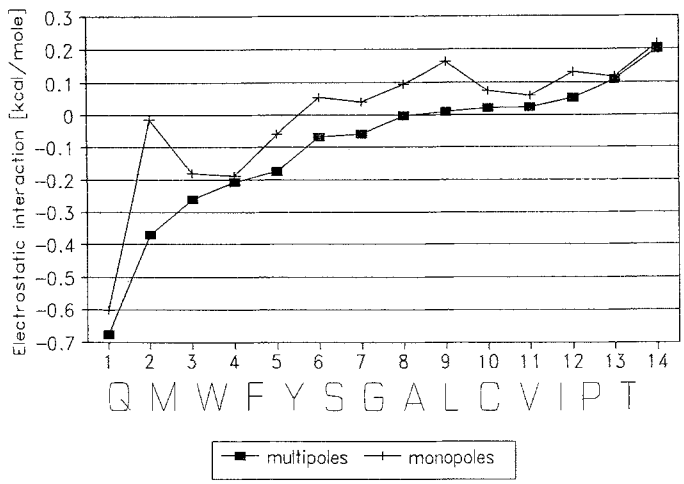

Fig. 1. Electrostatic interaction energies between CYS 219 and various neutral aminoacids denoted by one letter codes substituting position 45 in glutathione transferase

cated in position 45 of 3-3 isosyme of glutathione S-transferase (Ji et al., unpublished) have been calculated. The results presented on Fig. 1 illustrate that use of atomic charges (monopoles) only may lead to qualitative errors in relative stabilities of different pairs of aminoacids. This may explain observed differences between structural predictions obtained from various force fields (Roterman et al., 1989a).

Another remarkable application of CAMM is related to nonempirical modeling of torsional potentials. Preliminary results (Sokalski et al., 1991a) for S-S bridges typical for cystine cross-linkage indicate that the location of minima in corresponding torsional potential is determined mainly by electrostatic interactions. For longer bonds such intramolecular interactions can be modeled by atomic multipoles located on two atoms defining axis of rotation only. This observation may be used in entirely nonempirical derivation of torsional potentials and rational development of new more efficient procedures for conformational analysis of polypeptides.

One other promising CAMM application is modeling changes in catalytic activity in mutated enzymes resulting from amino acid substitutions by site directed mutagenesis. This is possible by decomposing activation barrier lowering into additive electrostatic, exchange, induction and dispersion components (Sokalski, 1985). Dominant electrostatic nature of the transition state differential stabilization permits to model via CAMM catalytic effects for systems as large as mutated enzymes. Within such approach we examined the locations of the catalytically active amino acid residues in tyrosyl *t*-RNA synthetase (Sokalski et al., 1991a). The theoretical results coicide with corresponding data from site-directed mutagenesis experiments (Fersht, 1987) and indicate catalytic nature of charged amino acids constituting conservative amino acid substitution patterns HIGH, KMSKS and GSDQ, characteristic of class I of the aminoacyl *t*-RNA synthetases.

Analogous approach has been applied to estimate pK shifts of HIS 64 in several mutated subtilisines (Sokalski et al., 1989). Again the theoretical results

agreed reasonably with corresponding vacuum scaled experimental data (Sternberg et al., 1987).

Both above described applications require access to database of atomic multipoles for amino acids. Higher atomic multipole moments play important role in calculations related to catalytic effects due to particularly large deviations from spherical charge distribution for transition states.

Conclusions

Atomic multipoles compensate deficiencies of conventional atomic charge models in calculations of molecular electrostatic potentials, electric fields and intermolecular electrostatic interaction energies at distances close to van der Waals radii. This allows precise and cost effective modeling of electrostatic effects responsible for specific recognition and binding of biomolecules.

Atomic multipoles seem to be the key ingredient determining location of the minima of torsional barriers.

Databases of atomic multipoles for amino acids and peptides can be used in modeling optimal catalytic activity and pK shifts resulting from amino acid substitution in mutated enzymes.

Acknowledgements

Part of this work has been sponsored by the State Committee for Scientific Research and the Polish-American Maria Skłodowska-Curie Joint Fund II in cooperation with the National Institute of Standards and Technology under Project MEN/NIST-91-53.

The author is grateful to Dr. X. Ji, Dr. G. L. Gilliland and Dr. M. Krauss for transmitting crystal structure of 3-3 isosyme of glutathione S-transferase before publication.

References

Bellido MN, Rullmann JAC (1989) Atomic charge models for polypeptides derived from ab initio calculations. J Comp Chem 10: 479–487

Buckingham AD, Fowler PW (1985) A model for geometries of Van der Waals complexes. Can J Chem 63: 2018–2025

Chipot C, Angyan JG, Ferenczy GG, Scheraga HA (1993) Transferable net atomic charges from a distributed multipole analysis for the description of electrostatic properties. A case study of saturated hydrocarbons. J Phys Chem 97: 6628–6636

Ciosłowski J (1993) Ab initio calculations on large molecules: methodology and application. Rev Comp Chem 4: 1–33

Clementi E, Cavallone F, Scordamaglia R (1977) Analytical potentials from "ab initio" computations for the interaction between biomolecules. 1. Water with amino acids. J Am Chem Soc 99: 5531–5545

Faerman CH, Price SL (1990) A transferable distributed multipole model for the electrostatic interactions of peptides and amides. J Am Chem Soc 112: 4915–4926

Ferenczy GG (1991) Charges derived from distributed multipole series. J Comp Chem 12: 913–917

Fersht AR (1987) Dissection of the structure and activity of the enzyme. Proceedings of the R.A. Welch Foundation, Conf Chem Res, Houston 31, pp 158-182

Gilson MK, Honig BH (1988) Energetics of charge-charge interactions in proteins. Proteins: Struct Funct Gen 3: 32–52

Goldblum A, Perahia D, Pullman A (1979) Use of the overlap multipole expansion for approximating molecular electrostatic potentials. Int J Quantum Chem 15: 121–129

Ji X, Gilliland GL, Armstrong RN (1994) Crystal structure of 3-3 isozyme of glutathione S-transferase complex with glutathione (to be published)

Lecomte C, Ghermani N, Pichon-Pesne V, Souhassou M (1992) Experimental electron density and electrostatic properties of peptides by high resolution X-ray diffraction. J Mol Struct (Theochem) 255: 241–260

Liang JY, Lipscomb WN (1986) Transferabilities of atomic multipoles in atomic multipole expansions. J Phys Chem 90: 4246–4253

Price SL, Richards NGJ (1991) On the representation of electrostatic fields around ab initio charge distributions. J Comput Aided Mol Des 5: 41–54

Price SL, Stone AJ (1992) Electrostatic models for polypeptides: can we assume transferability? J Chem Soc, Faraday Trans II 88: 1755–1763

Price SL, Harrison RJ, Guest MF (1989) An ab initio distributed multipole study of the electrostatic potential around an undecapeptide cyclosporin derivative and a comparison with point charge electrostatic models. J Comp Chem 10: 552–567

Price SL, Faerman CH, Murray CW (1991) Towards accurate transferable electrostatic models for polypeptides: a distributed multipole study of blocked amino acid residue charge distributions. J Comp Chem 12: 1187–1197

Rein R (1973) On physical properties and interactions of polyatomic molecules with application to molecular recognition in biology. Adv Quantum Chem 7: 335–396

Roterman IK, Gibson KD, Scheraga HA (1989a) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. I. Conformational predictions for the tandemly repeated peptide $(Asn\text{-}Ala\text{-}Asn\text{-}Pro)_9$. J Biomolec Struct Dyn 7: 391–419

Roterman IK, Lambert MH, Gibson KD, Scheraga HA (1989b) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. II. ϕ Ψ maps for N-acetyl alanine N′-methyl amide: comparisons, contrasts and simple experimental tests. J Biomolec Struct Dyn 7: 421–453

Sheridan RP, Allen LC (1981) The active site electrostatic potential of human carbonic anhydrase. J Am Chem Soc 103: 1544–1550

Sokalski WA (1985) The physical nature of catalytic activity due to the molecular environment in terms of intermolecular interaction theory: derivation of simplified models. J Mol Catal 30: 395–410

Sokalski WA, Poirier RA (1983) Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. Chem Phys Lett 98: 86–92

Sokalski WA, Sawaryn A (1987) Correlated molecular and cumulative atomic multipole moments. J Chem Phys 87: 526–534

Sokalski WA, Sawaryn A (1992) Cumulative multicenter multipole moment databases and their applications. J Mol Struct (Theochem) 256: 91–112

Sokalski WA, Sneddon S (1991) Efficient method for the generation and display of electrostatic potential surfaces from ab-initio wavefunctions. J Mol Graphics 9: 74–77

Sokalski WA, Hariharan PC, Kaufman JJ (1987) Library of cumulative atomic multipole moments. I. Nucleic acid bases. Int J Quantum Chem: Quantum Biol Symp 14: 111–126

Sokalski WA, Maruszewski K, Hariharan PC, Kaufman JJ (1989) Library of cumulative atomic multipole moments. II. Neutral and charged amino acids. Int J Quantum Chem: Quantum Biol Symp 16: 119–164

Sokalski WA, Barak D, Shibata M, Rein R (1991a) Catalytic activity of aminoacyl tRNA synthetases and its implications for the origin of life I. Aminoacyl adenylate formation in tyrosyl tRNA synthetase. J Mol Evolution 33: 405–411

Sokalski WA, Lai J, Luo N, Sun S, Shibata M, Rein R (1991b) Ab initio study of the electrostatic multipole nature of torsional potentials in CH_3SSCH_3, CH_3SSH and HOOH. Int J Quantum Chem: Quantum Biol Symp 18: 61–71

Sokalski WA, Shibata M, Ornstein RL, Rein R (1992) Cumulative atomic multipole moments complement any atomic charge model to obtain more accurate electrostatic properties. J Comp Chem 13: 883–887

Sokalski WA, Keller DA, Ornstein RL, Rein R (1993) Multipole correction of atomic monopole models of molecular charge distribution. I. Peptides. J Comp Chem 14: 970–976

Stone AJ (1981) Distributed multipole analysis, or how to describe a molecular charge distribution. Chem Phys Lett 83: 233–239

Spackman M (1986) A simple quantitative model of hydrogen bonding. J Chem Phys 85: 6587–6601

Sternberg MJ, Hayes FRF, Russell AJ, Thomas PG, Fersht AR (1987) Prediction of electrostatic effects of engineering of protein charges. Nature 330: 86–88

Williams DE (1991) Net atomic charge and multipole models for the ab initio molecular electric potential. Rev Comp Chem 2: 219–271

Authors' address: Dr. W. A. Sokalski, Institute of Physical and Theoretical Chemistry, I-30 Wyb. Wyspiańskiego 27, PL-50-370 Wrocław, Poland.

Received January 2, 1993