Akademie věd České republiky

Teze disertační práce

k získání vědeckého titulu "doktor věd"

ve skupině věd *chemické vědy*

Simulations of Protein Vibrational Spectra

Komise pro obhajoby doktorských disertací v oboru

fyzikální chemie

Doc. RNDr. Petr Bouř, CSc.

Ústav organické chemie a biochemie, v.v.i., Akademie

věd České republiky, Flemingovo nám. 2, 16610, Praha,

Česká Republika

Summary

The dissertation was written in the Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, and is based on nine publications. They summarize our research dedicated in a large part to the vibrational spectroscopy of peptides and protein. Applications are in particular oriented towards modern polarized spectroscopies (vibrational circular dichroism, Raman optical activity). In addition to the classical infrared absorption and Raman scattering, these techniques provide additional structural information encoded in signs and intensities of spectral peaks. A short time ago commercial spectrometers enabling the polarized measurements started to be available in U. S. A. We think that our science is keeping pace in this respect, as our chiral chemistry has always been strong and such research is conducted on a high level not only in our institute, but also in other places of the Czech Republic.

The work deals with a brief history and overview of modern theoretical and experimental procedures in given in the Introduction. The main part is dedicated to computational methods that we developed to enhance interpretation of biomolecular vibrational spectra. All started in 1993 (*J. A. Chem. So*c., **1993,** *15*, 9602) when we have for the first time

shown that vibrational spectra of all the main secondary peptide structures could be very reliably interpreted with local models and ab initio computational techniques.

Several advanced computational methods are further presented. They involve the Cartesian coordinate transfer techniques for calculations of vibrations for large molecules, partial optimization in normal mode coordinates as a tool for vibrational spectroscopy, simulations of spectra from classical trajectories using the Fourier transform techniques, empirical modeling of the effect of hydration on the amide group, light interaction with giant systems, and the inclusion of the anharmonic effects.

Functions of some of the developed computer programs are also summarized. Hopefully, this will help other researchers attracted to the field and trying to understand physical properties of biomolecules.

5. Kubelka, J.; Bouř, P. *J. Chem. Theory Comput.* **2009**, *5*, 200-207. "Simulation of Vibrational Spectra of Large Molecules by Arbitrary Time Propagation" 0/0 **4.31**

6. Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2003**, *119*, 11253-11262. "Empirical modeling of the peptide amide I band IR intensity in water solution." 68/48 **3.04**

7. Bouř, P.; Michalík, D.; Kapitán, J. *J. Chem. Phys.* **2005**, *122*, 144501 1-9. "Empirical solvent correction for multiple amide group vibrational modes" 22/12 **3.04**

8. Andrushchenko, V.; Bouř, P. *J. Comput. Chem.* **2008**, *29*, 2693-2703. "Circular Dichroism Enhancement in Large DNA Aggregates Simulated by a Generalized Oscillator Model." 1/0 **4.30**

9. Daněček, P.; Kapitán, J.; Baumruk, V.; Bednárová; L. Kopecký, V. Jr.; Bouř, P. *J. Chem. Phys.* **2007**, *126*, 224513. "Anharmonic effects in IR, Raman, and Raman optical activity spectra of alanine and proline zwitterions." 4/0 **3.04**

Peptide and Protein Vibrational Spectroscopy

Vibrational spectroscopic methods established as important experimental tools for studies of peptides and proteins. This was enabled by advances in both the instrumentation, and the interpretation and computational procedures. Experimentally, Fourier transform infrared (FTIR) and high sensitivity Raman techniques are the most frequently used methods. (Barth, 2007;Haris, 2000;Tu, 1982;Williams, 1986).

The IR and Raman bands provide valuable structural insights. However, in biopolymer systems many, often hundreds, of overlapping transitions contribute to the spectra. Therefore, the polarized techniques, such as the vibrational circular dichroism (VCD) and Raman optical activity (ROA), provide a particularly welcome enhanced structural sensitivity through characteristic bandshapes. The polarized intensities arise from the differential response to left- and right-circularly polarized light by chiral structures, and thus are very sensitive to the conformation (Barron and Hecht, 2000;Bouř, Kubelka and Keiderling, 2002).

Commercial VCD and ROA instruments started to be available in the last decade. Nevertheless, we should at least mention that the vibrational spectroscopy of proteins and other biomolecules continues to evolve also in other directions, such

The Methodic Publications

citation with/without autocitation **Impact factor**

1. Bouř, P.; Keiderling, T. A. *J. A. Chem. So*c., **1993**, *15*, 9602-9607. "*Ab initio* Simulations of the Vibrational Circular Dichroism of Coupled Peptides". 63/47 **7.89**

2. Bouř, P.; Sopková, J.; Bednárová, L.; Maloň, P.; Keiderling, T. A. *J. Comput. Chem.* **1997**, *18*, 646-659. "Transfer of Molecular Property Tensors in Cartesian Coordinates: A New Algorithm for Simulation of Vibrational Spectra." 77/34 **4.30**

3. Bour, P.; Keiderling, T. A. *J. Chem. Phys.* **2002**, *117*, 4126-4132. "Partial optimization of molecular geometry in normal coordinates and use as a tool for simulation of vibrational spectra" 19/9 **3.04**

4. Horníček, J.; Kaprálová, P.; Bouř, P. *J. Chem. Phys.* **2007**, *127*, 084502. "Simulations of vibrational spectra from classical trajectories: Calibration with *ab initio* force fields." 2/0 **3.04**

Abbreviations

AAT	atomic axial tensor
APT	atomic polar tensor
CCT	Cartesian coordinate transfer
FF	force field
FTIR	Fourier transform infrared
IR	infrared
LD	linear dichroism
MM	molecular mechanics
QM	quantum mechanics
RFO	rational function optimization
ROA	Raman optical activity
TDC	Transition dipole coupling
VCD	vibrational circular dichroism
VCI	vibrational configuration interaction
VSCF	vibrational self-consistent field

as the ultrafast techniques (Fabian and Neumann, 2004), laser phototriggering (Gregoriou and Braiman, 2005;Vogel and Siebert, 2000) and temperature-jump (Callender and Dyer, 2006;Dyer, Gai and Woodruff, 1998), two-dimensional (2-D) IR (Chung et al., 2007;Woutersen and Hamm, 2002), and utilization of selective isotopic labeling (Bouř and Keiderling, 2005a;Decatur, 2006).

An inseparable part of the success of the vibrational spectroscopy was the development of the interpretation, simulation and computational methods. First theoretical approaches were based on simplified vibrational calculations, but soon quantum mechanical determinations of complete force filed (FF) offered much more powerful means (Pulay, 1995). The Schrödinger equation could be solved for larger systems, quantum mechanical force fields (FF) as well as intensities for IR, Raman, vibrational circular dichroism (Stephens, 1985) and Raman optical activity are now available (Bouř, Kapitán and Baumruk, 2001;Nafie and Freedman, 2000;Polavarapu, 1998;Ruud, Helgaker and Bouř, 2002).

As was discovered in the first study involved in the dissertation (Bouř and Keiderling, 1993), the rigorous ab initio approach could explain not only spectra of small molecules, but also provided a reliable basis for interpretation of peptide and

protein spectra. In particular, all the main peptide secondary structure spectra could be interpreted on the basis of a simplified dipeptide model (*Figure 1*).
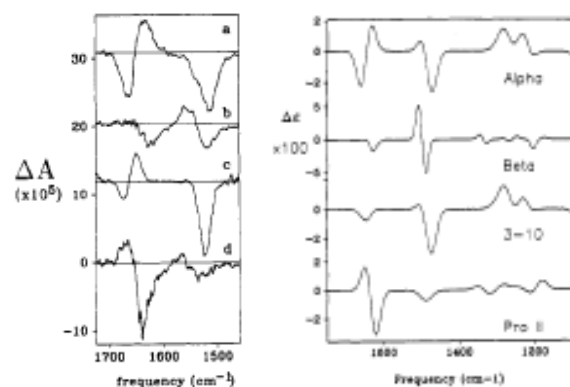

Figure 1. Experimental VCD spectra of proteins rich at the α-helical (a), β-sheet (b), 3_{10}-helical (c) and polyproline II conformations (left) could be interpreted an initio using a signal of a dipeptide model (right). From ref. (Bouř and Keiderling, 1993).

Locality and transfer

   The concept of the locality of the vibrational interactions and modeling of large biomolecules from smaller fragments that can be subjected to accurate ab initio computations was elaborated in several other works

Developed Programs

CCT (Cartesian coordinate tensor transfer)

MCM (Graphical program)

QGRAD (Normal mode optimization)

NEWFTQ (Fourier-based spectral generation)

HYPEP (Electrostatic solvent correction)

TDC (Transition dipole coupling, large molecule extension)

S4 (Anharmonic corrections)

Krummel, A. T., and Zanni, M. T. (2006). *J. Phys. Chem. B* 110, 24720.

Kubelka, J., and Bouř, P. (2009). *J. Chem. Theor. Comput.* 5, 200.

Lee, K. K., Hahn, S., Oh, K. I., Choi, J. S., Joo, C., Lee, H., Han, H., and Cho, M. (2006). *J. Phys. Chem. B* 110, 18834.

Lerman, L. S. (1971). *Proc. Natl. Acad. Sci. U. S. A.* 68, 1886.

Liao, L. W., and Cole, R. D. (1981). *J. Biol. Chem.* 256, 6751.

Loparo, J. J., Roberts, S. T., and Tokmakoff, A. (2006). *J. Chem. Phys.* 125, 194521.

Maa, S., Cao, X., Mak, M., Sadik, A., Walkner, C., Freedman, T. B., Lednev, I., Dukor, R., and Nafie, L. (2007). *J. Am. Chem. Soc.* 129, 12364.

Mankoo, P. K., and Keyes, T. (2006). *J. Chem. Phys.* 124, 204503.

Matsunaga, N., Chaban, G. M., and Gerber, R. B. (2002). *J. Chem. Phys.* 117, 3541.

Nafie, L. A., and Freedman, T. B. (2000). *In* "Circular Dichroism. Principles and Applications. 2nd edition." (N. Berova, K. Nakanishi, and R. W. Woody, eds.), p. 97. Wiley-VCH, New York.

Papoušek, D., and Aliev, M. R. (1982). "Molecular Vibrational/Rotational Spectra", Academia, Prague.

Polavarapu, P. L. (1998). "Vibrational spectra: principles and applications with emphasis on optical activity", Elsevier, Amsterdam.

Pulay, P. (1995). *J. Phys. Chem.* 99, 3093.

Ruud, K., Helgaker, T., and Bouř, P. (2002). *J. Phys. Chem. A* 106, 7448.

Šebek, J., Kejík, Z., and Bouř, P. (2006). *J. Phys. Chem. A* 110, 4702.

Shanno, D. F. (1970). *Math. Comput.* 24, 647.

Simons, J., and Nichols, J. (1990). *Int. J. Quantum Chem., Quantum Chem. Symp.* 24, 263.

Stephens, P. J. (1985). *J. Phys. Chem.* 89, 748.

Tam, C. N., Bouř, P., Eckert, J., and Trouw, F. R. (1997). *J. Phys. Chem.* 101, 5877.

Tam, C. N., Bouř, P., and Keiderling, T. A. (1996). *J. Am. Chem. Soc.* 118, 10285.

Torii, H. (2006). *J. Phys. Chem. A* 110, 9469.

Tu, A. T. (1982). "Raman Spectroscopy in Biology", Wiley, New York.

Vogel, R., and Siebert, F. (2000). *Curr. Opin. Chem. Biol.* 4, 518.

Williams, R. W. (1986). *Methods Enzymol.* 130, 311.

Wilson, E. B., Decius, J. C., and Cross, P. C. (1980). "Molecular vibrations", Dover, New York.

Woutersen, S., and Hamm, P. (2002). *Condens. Matter* 14, R1035.

(Andrushchenko, Wieser and Bouř, 2003;Bouř, 1993;Bouř, 2002;Bouř, 2002-2006;Bouř and Keiderling, 1993;Bouř and Keiderling, 2002;Bouř and Keiderling, 2003;Bouř and Keiderling, 2004;Bouř and Keiderling, 2005b;Bouř, Keiderling and Malon, 1994;Bouř, Kubelka and Keiderling, 2000a;Bouř, Kubelka and Keiderling, 2000b;Bouř, Kubelka and Keiderling, 2002;Bouř et al., 1997;Keiderling et al., 1999). Historically, molecular force field was associated rather with internal coordinates (Wilson, Decius and Cross, 1980). However, introduction of internal coordinates, such as bond lengths, bending angles, torsional (dihedral) angles, out of plane deformations or linear bendings becomes very laborious for larger molecules. Moreover, to deal with the force field diagonalization a set of non-redundant coordinates has to be defined (Papoušek and Aliev, 1982). This is almost impossible for larger systems, especially if the coordinates should have a clear chemical meaning.

As an alternative to the internal coordinates, we suggested a numerically more convenient concept of transferring of molecular properties (tensors) based on the Cartesian coordinates (Bouř et al., 1997). That does not mean that chemical bonds are not used: they are utilized for definition of the rotation matrix used for transformation of the Cartesian

molecular property tensors. This way, the force field (second energy derivatives, Hessian) and intensity tensors (first derivatives of electric and magnetic dipole, and dipolar, quadrupolar and magnetic polarizabilities) can be transferred directly as computed by the quantum mechanical programs. The basic idea of the transfer is illustrated in *Figure 2*
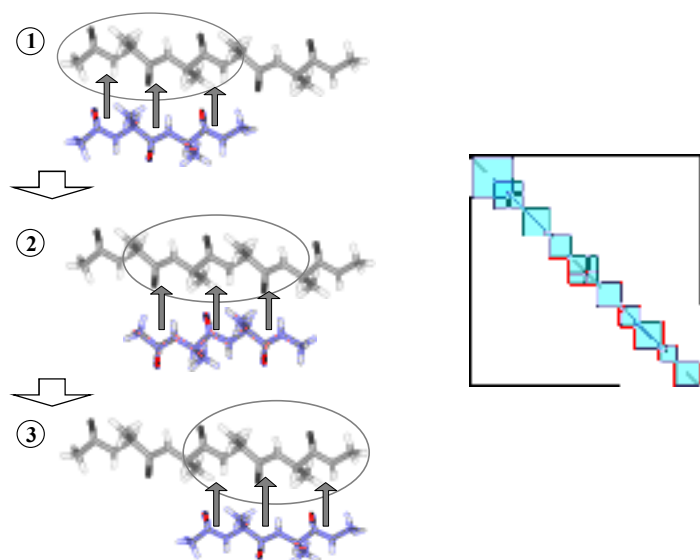

Figure 2. The Cartesian coordinate parameter transfer method applied for a regular peptide. Force field (a square matrix, on the right) of the larger peptide is partially reconstructed by a propagating of a smaller triamide unit.

Phys. 10, 3839.

Chung, H. S., Ganim, Z., Jones, K. C., and Tokmakoff, A. (2007). *Proc. Natl. Acad. Sci. U. S. A.* 104, 14237.

Cowman, M. K., and Fasman, G. D. (1978). *Proc. Natl. Acad. Sci. U. S. A.* 75, 4759.

Daněček, P., and Bouř, P. (2007a). *J. Chem. Phys.* 126, 224513.

Daněček, P., and Bouř, P. (2007b). *J. Comput. Chem.* 28, 1617.

Decatur, S. M. (2006). *Acc. Chem. Res.* 39, 169.

Dračínský, M., Kaminský, J., and Bouř, P. (2009). *J. Phys. Chem.* 130, 094106.

Dyer, R. B., Gai, F., and Woodruff, W. H. (1998). *Acc. Chem. Res.* 31, 709.

Evdokimov, Y. M., Pyatigorskaya, T. L., Kadikov, V. A., Polyvtsev, O. F., Doskocil, J., Koudelka, J., and Varshavsky, Y. M. (1976). *Nucleic Acids Res.* 3, 1533.

Fabian, H., and Neumann, D. (2004). *Methods* 34, 28.

Fasman, G. D., Schaffhausen, B., Goldsmith, L., and Adler, A. (1970). *Biochemistry* 9, 2814.

Fletcher, R. (1970). *Comput. J.* 13, 317.

Gersanovski, D., Colson, P., Houssier, C., and Fredericq, E. (1985). *Biochim. Biophys. Acta* 824, 313.

Gnanakaran, S., and Hochstrasser, R. M. (2001). *J. Am. Chem. Soc.* 123, 12886.

Goldfarb, D. (1970). *Math. Comput.* 24, 23.

Gosule, L. C., and Schellman, J. A. (1976). *Nature* 259, 333.

Gregoriou, V. G., and Braiman, M. S. (2005). CRC, Boca Raton, FL.

Hahn, S., Lee, H., and Cho, M. (2004). *J. Chem. Phys.* 121, 1849.

Ham, S., Kim, J. H., Kochan, L., and Cho, M. (2003). *J. Chem. Phys.* 118, 3491.

Haris, P. I. (2000). *In* "Infrared Analysis of Peptides and Proteins: Principles and Applications. ACS Symposium Series." (B. Ram Singh, ed.), p. 54. ACS, Washington DC.

Horníček, J., Kaprálová, P., and Bouř, P. (2007). *J. Chem. Phys.* 127, 084502.

Kapitán, J., Hecht, L., and Bouř, P. (2008). *Phys. Chem. Chem. Phys.* 10, 1003.

Kapitán, J., Johannessen, C., Bouř, P., Hecht, L., and Barron, L. D. (2009). *Chirality*, submitted.

Keiderling, T. A., Silva, R. A. G. D., Decatur, S. M., and Bouř, P. (1999). *In* "Spectrocopy of Biological Molecules: New Directions" (J. Greve, G. J. Puppels, and C. Otto, eds.), p. 63. Kluwer AP, Dortrecht.

Khadake, J. R., and Rao, M. R. (1995). *Biochemistry* 34, 15792.

10. References

Andrushchenko, V., and Bouř, P. (2008). *J. Comput. Chem.* 29, 2693.
Andrushchenko, V., Wieser, H., and Bouř, P. (2003). *J. Phys. Chem. B* 91, 12623.
Banerjee, A., Adams, N., Simons, J., and Shepard, R. (1985). 89, 52.
Barron, L. D., and Hecht, L. (2000). *In* "Circular dichroism, principles and applications" (K. Nakanishi, N. Berova, and R. W. Woody, eds.), p. 667. Wiley-VCH, New York.
Barth, A. (2007). *Biochim. Biophys. Acta* 1767, 1073.
Bouř, P. (1993). Academy of Science of the Czech Republic, Prague, Czech Republic.
Bouř, P. (1994a). *J. Phys. Chem.* 98, 8862.
Bouř, P. (1994b). Academy of Sciences, Prague.
Bouř, P. (2002). *Chem. Phys. Lett.* 365, 82.
Bouř, P. (2002-2006). Academy of Sciences, Prague, Prague.
Bouř, P. (2004). *J. Chem. Phys.* 121, 7545.
Bouř, P. (2005). *Collect. Czech. Chem. Commun.* 70, 1315.
Bouř, P., and Bednárová, L. (1995). *J. Phys. Chem.* 99, 5961.
Bouř, P., Kapitán, J., and Baumruk, V. (2001). *J. Phys. Chem. A* 105, 6362.
Bouř, P., and Keiderling, T. A. (1993). *J. Am. Chem. Soc.* 115, 9602.
Bouř, P., and Keiderling, T. A. (2002). *J. Chem. Phys.* 117, 4126.
Bouř, P., and Keiderling, T. A. (2003). *J. Chem. Phys.* 119, 11253.
Bouř, P., and Keiderling, T. A. (2004). *J. Mol. Struct. (THEOCHEM)* 675, 95.
Bouř, P., and Keiderling, T. A. (2005a). *J. Phys. Chem. B* 109, 5348.
Bouř, P., and Keiderling, T. A. (2005b). *J. Phys. Chem. B* 109, 23687.
Bouř, P., Keiderling, T. A., and Malon, P. (1994). *In* "23rd European Peptide Symposium" (H. L. L. Maia, ed.), p. 517. ESCOM, Leiden.
Bouř, P., Kubelka, J., and Keiderling, T. A. (2000a). *to be submitted*.
Bouř, P., Kubelka, J., and Keiderling, T. A. (2000b). *Biopolymers* 53, 380.
Bouř, P., Kubelka, J., and Keiderling, T. A. (2002). *Biopolymers* 65, 45.
Bouř, P., Michalík, D., and Kapitán, J. (2005). *J. Chem. Phys.* 122, 144501.
Bouř, P., Sopková, J., Bednárová, L., Maloň, P., and Keiderling, T. A. (1997). *J. Comput. Chem.* 18, 646.
Bowman, J. M. (1978). *J. Chem. Phys.* 68, 608.
Brauer, B., Chaban, G. M., and Gerber, R. B. (2004). *Phys Chem. Chem. Phys.* 6, 2543.
Broyden, C. G. (1970). *J. Inst. Math. Appl.* 6, 76.
Callender, R., and Dyer, R. B. (2006). *Chem. Rev.* 106, 3031.
Choi, J. H., Cheon, S., Lee, H., and Cho, M. (2008). *Phys. Chem. Chem.*

Test computations proved that the transfer method not only enables to calculate spectra of molecules of virtually any size, but for suitably chosen fragmentation the results are almost indistinguishable form benchmark full-quantum computations.

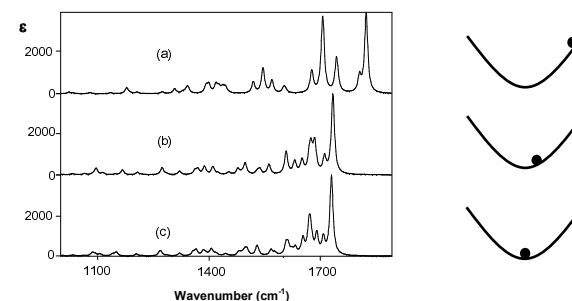**Normal Mode Partial Optimization**

Figure 3. Calculated IR spectra of a tripeptide documenting the normal mode optimization: (a) Non-optimized geometry provides unreasonable spectra, while (b) that obtained with the normal modes provides most of the modes nearly relaxed and the spectral shape is very close to (c) that of a fully-optimized structure.

The fragmentation of large molecules brings about the problem of geometrical constraints. These have to be chosen

very carefully, as the harmonic vibrational frequency calculations are in principle applicable only to molecules whose geometry corresponds to the minimum of the potential energy surface. As a common approach, backbone torsional motions are constrained for peptides, since they do not significantly affect the high frequency, stretching and bending modes of interest for spectral analysis, whose coordinates were fully optimized.

However, such an approach is dependent on the initial choice of the coordinates and geometry and the spectral effects cannot be controlled. Therefore, we found it more convenient to fix the low-frequency normal mode coordinates directly (Bouř, 2002;Bouř, 2005;Bouř and Keiderling, 2002). For linear peptides the method is roughly equivalent to the torsional constraints (Fig. 3), but is also applicable for more complicated systems and complexes. The normal mode method was implemented as the qgrad.f program, so that it can communicate with usual *ab initio* program packages. The algorithm is simple, but allows for a robust optimization and comprises the modern convergence tools as the continuous BFGS update of the Hessian (Broyden, 1970;Fletcher, 1970;Goldfarb, 1970;Shanno, 1970), or the rational function optimization (RFO) (Banerjee et al., 1985;Simons and Nichols,

that are now being derived from careful peptide and protein vibrational spectroscopic studies start to touch some very basic questions in biology, such as the mechanisms governing the protein folding.

anharmonic (C, C', C") spectra (300 K, B3LYP/CPCM/6-31++G**, 5000 VCI states, 5 modes frozen), and the experiment (D, D', D").

Conclusions

It is now fully possible to compute vibrational spectra for moderately sized peptides to an experimentally useful degree of accuracy using a method free of empirical parameterization, except for the fairly universal DFT functionals. These developments reflect continuing advances in quantum chemical computational methodology, and in widely available computer capabilities, but also special computational procedures that we were trying to develop. Present analyses have gone beyond frequency correlation to address apparent band shapes due to the dispersion of intensity in the exciton coupled modes characteristic to the polymeric species. Modern methods permit to address the effects of solvation and structural fluctuation in ways previously viewed as unachievable. These developments have made the computational simulation an integral part of experimental vibrational spectroscopy.

As an additional reward, we get a more correct physical understanding of the interactions that govern the peptide and protein spectral features. Finer details of spectra and structure

1990). The program was later extended to comprise both the chemical coordinate and normal mode control (Bouř, 2005).

Generation of Spectra Without Matrix Diagonalization

For some system we found it convenient to pursue the Fourier transform techniques instead to directly diagonalize the force field matrix. Although the standard procedures are powerful means for simulating the spectra, they may become prohibitively lengthy for larger systems. Also, molecular dynamics effects (flexibility, temperature) are difficult to incorporate. A renewed interest in the generation of the spectra *via* Fourier transformations appeared also in a connection with combined quantum mechanics-molecular mechanics (QM/MM) methods (Hahn, Lee and Cho, 2004;Mankoo and Keyes, 2006), and numerous new spectroscopic techniques including the vibrational circular dichroism (VCD). Also two-dimensional (2D) infrared spectroscopy is quite complex and often conveniently utilizes variously modified Fourier techniques (Gnanakaran and Hochstrasser, 2001;Krummel and Zanni, 2006;Lee et al., 2006;Loparo, Roberts and Tokmakoff, 2006;Torii, 2006).

To investigate the possibilities of the Fourier time-frequency transformation for the vibrational polarized

spectroscopy, we reformulated the theory in terms of correspondence between the classical and quantum variables (Horníček, Kaprálová and Bouř, 2007). For example, the transition integral of a normal mode coordinate Q_j between vibrational states |0> and |1> needed for the spectral intensities can be obtained as

$$\langle 0|Q_j|1\rangle = \sqrt{\frac{\hbar\omega_j}{2kT}\langle Q_j^2\rangle_{Classical}} \ ,$$

where $\hbar$ is the Planck constant, ω_j frequency, k the Boltzmann constant, T temperature, the classical average in the brackets can be obtained from Newtonian molecular dynamics. This allowed us to simulate exactly IR, Raman, VCD and ROA spectra of all harmonic systems.

In practical tests, nevertheless, the method exhibited some limitations. The accuracy of the frequencies thus obtained was dependent on the integration time step. A more serious restriction came from the need to distribute evenly the kinetic energy between all degrees of freedom (normal modes). But this could be solved by averaging of many trajectories (Figure 4).

escapes to the modeling. Combination bands are sometime predicted qualitatively correctly, too (Kapitán, Hecht and Bouř, 2008), but the limited expansion and the harmonic oscillator basis set used are probably too crude for exact computations in this case.
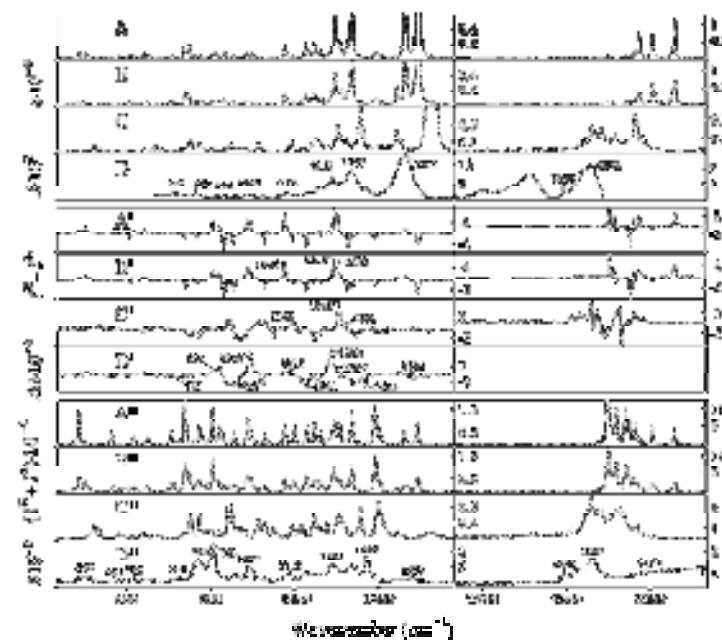

Figure 9. The effect of Boltzmann conformer averaging on absorption (A-D), ROA (A'-D') and Raman (A"-D") proline spectra: harmonic approximation (A, A', A"), Boltzmann-weighted conformer averaging of the harmonic (B, B', B") and

factor for such cases. For our tests we systematically increased the number the states and found that spectral improvement of higher frequency modes can be achieved if we neglect the coupling to the lower-frequency ones. Nevertheless, more has to be done in the future for reliable spectral simulations

When these methods were applied to the proline and alanine zwitterions, only limited improvement with respect to the harmonic case was observed within the lower frequency region (Daněček and Bouř, 2007a). For proline, we can see the effects in the Raman and Raman optical activity spectra in *Figure 9*. Observed behavior of the anharmonic corrections appears, however, quite general (Kapitán et al., 2009): For the low-frequency modes ($\sim<1000$ cm^{-1}) it is more profitable to make a Boltzmann averaging of the lowest-energy conformers, than to undergo a quantum anharmonic averaging. In the medium spectral region ($\sim$500-2000 cm^{-1}) current DFT functionals and solvent models are not accurate enough to profit from the anharmonicities, and only occasional improvements occur. The correction is essential for the highest-frequency modes (> 2000 cm^{-1}), particularly for those involving the C-H, N-H, O-H etc. stretching modes. But here the anharmonic character is so strong that we often achieve a correct overall frequency shift only. The fine mode splitting
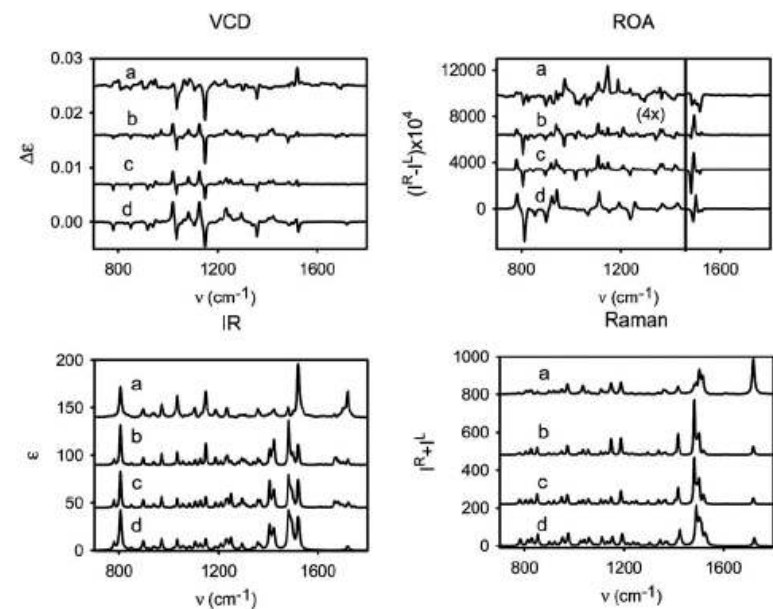

Figure 4. Test of the Fourier method on VCD, ROA, absorption, and Raman spectra of α-pinene: convergence of the simulations with 1 (a), 10 (b) and 1000 (c) trajectories to the ab initio result (d).

Another practical example how the Fourier techniques can be used to generate spectra for very large molecules is a propagation of the S-vectors in an arbitrary (fictitious) time (Kubelka and Bouř, 2009). Our original algorithm can circumvent direct or iterative procedures for finding the eigenvalues and eigenfunctions of the force field matrix by a

direct generation of the spectrum. Precise relative absorption and circular dichroism peak intensities with correct ABS/CD ratios can be obtained. Moreover, if compared with other diagonalization routines (Figure 5), we find that the methods scales favorably with molecular size and for large molecules it becomes the fastest method of the spectral generation.
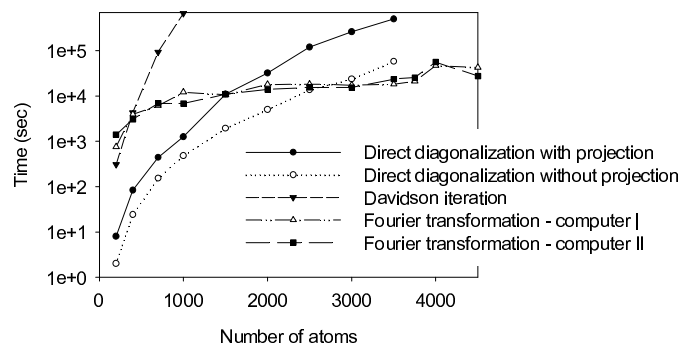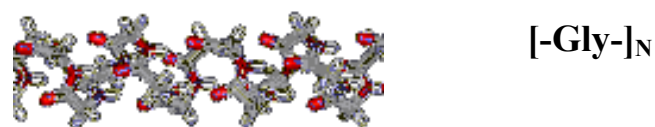
[-Gly-]$_N$

Figure 5. The dependence of the time (in seconds, on a logarithmic scale) needed to obtain spectra using the force field matrix on the number of atoms in the (Gly)$_N$ polymer.

$$E_n^{(2)} = \sum_{m \neq n} \frac{|W_{nm}|^2}{E_n - E_m}, \text{ with } W_{nm} = \langle n|W|m \rangle, W \text{ is the anharmonic}$$

part of the potential, E_n are unperturbed energies. The division by the energy difference is numerically unstable because of the random degeneracies. Simple treatment based on separation of the degenerate and non-degenerate states was proposed previously (Matsunaga, Chaban and Gerber, 2002). We explored a differently modified algorithm, using

$$E_n^{(2)} = \frac{1}{2} \sum_{m \neq n} \left[E_m - E_n + W_{mm} - W_{nn} \pm \sqrt{(E_m - E_n + W_{mm} - W_{nn})^2 + 4|W_{nm}|^2} \right]$$

where the + sign holds for $E_n > E_m$ and − sign for $E_n < E_m$. This formula significantly improves numerical stability, provides exact solutions for the degenerate case, and converges to the standard formula for small perturbations.

Nevertheless, the proper variational vibrational configuration interaction (VCI) is still the most accurate and universal method. If the wavefunction is expressed as a sum of harmonic oscillator functions, we obtain solution of the Schrödinger equation directly by the Hamiltonian diagonalization. As this method is limited by the size of the matrix that must be diagonalized, number of the harmonic states must be restricted. In addition, the speed of the diagonalization, mostly scaled as M^3, becomes the limiting

and the potential as a sum of effective potentials,
$V(Q_1,...,Q_M) \rightarrow \sum_{i=1}^{M} v_i(Q_i)$. Under the assumption of the separability the Schrödinger equation divides into a sum of effective one-dimensional equations and can be easily solved in the harmonic oscillator basis. Due to the dependence of the effective potentials on the resultant wavefunction, the solving has to be repeated "self-consistently" until the energy values (e_i) stabilize.

For excited states we have investigated behavior of two VSCF schemes. First, only the lowest-energy (ground) states (ψ_i) were used for determining the effective potential (which is referred to as gVSCF), while in the second approximation ("eVSCF") excited states were included in the averaging. Thus, in the gVSCF method, resultant set of excited states is orthogonal, but not self-consistent. On the contrary, in eVSCF the self-consistency holds at the expense of the orthogonality. The gVSCF computation requires only one set of the self-consistent iterations for the ground state, while this has to be repeated for each excited state in eVSCF. In test computations, eVSCF provided slightly superior results to gVSCF.

Alternatively to VSCF, standard second-order perturbation theory can be added to the harmonic energies as

Empirical solvent correction

Modeling protein vibration, we should obviously involve the protein dynamics and interaction with the solvent. In practice, this becomes very difficult for the notoriously time-demanding *ab initio* methods. Not only the system with a molecule of interest and the solvent quickly becomes prohibitively big for a calculation at a reasonable level, but the proper averaging of the temperature motion often requires to take into account many, often hundreds, different geometries. Therefore, empirical methods still provide an attractive alternative to the brute force *ab initio* computations.

It should be noted that for the electronic excitations local methods cannot be used or provide very misleading results (Šebek, Kejík and Bouř, 2006). For vibrations the associated phenomena do not involve an extensive electronic transfer and related to absolute vacuum values the solvent influence causes comparatively small changes in the vibrational properties. By other words, the solvent influence can be thought of as a small perturbation to a vacuum system. This is convenient because vacuum results accessible at a higher level might be still usable, and because the solvent correction might be extremely fast.

A simpler solvent treatment for the amide group was

first proposed by Cho and coworkers (Ham et al., 2003). An empirical frequency correction due to the solvent induced electric field was applied to the *ab initio* amide I frequencies (ω_0) obtained from a vacuum calculation. Note, that the amide I mode is the most useful for peptide IR and VCD spectroscopy, and also very affected by the environment; therefore, such a correction was quite desirable. In the original formulation (Ham et al., 2003) the amide I stretching frequency ω (in solvent) linearly relates to the vacuum frequency (ω_0) and to the electrostatic solvent potential φ_i, as

$$\omega = \omega_0 + \sum_{k=1}^{N} b_i \varphi_i .$$

The potential is measured at N amide group atoms (*e.g.* CNHCOC) and b_i are constants. In a following work, Choi and Cho make the method more flexible (Choi et al., 2008), but solely for amide I. Meanwhile, testing the potential method, we found that the correction can be generalized to any chromophore (any number of atoms) and applied also to the intensity tensors (Bouř, 2004;Bouř and Keiderling, 2003). The coefficients b_i can be obtained by a fit to *ab initio* computations involving explicit solvent molecules. For intensity parameters, it is more convenient to express the fitting coefficients in a

we proposed to separate the rotational degree of freedom (the methyl rotation) and diagonalize the Hamiltonian including rigid rotor basis functions (Kapitán, Hecht and Bouř, 2008) .

Fortunately, quite often, the anharmonic character of the potential can be treated in a perturbative way, as another term in the Taylor expansion. Energy third (cubic) and some fourth (quartic) derivatives can be calculated relatively easily by numerical differentiation of the harmonic force field, either in Cartesian or normal mode coordinates.

Even with this simplest anharmonic treatment, we arrive to the complicated problem of diagonalizing the Hamiltonian. As discussed in ref. (Daněček and Bouř, 2007b) there are three common procedures (vibrational self-consistent field, perturbation and vibrational configuration interaction) that are differently numerically stable and for real system often do not converge to same energies.

In the vibrational self-consistent field (VSCF) model (Bowman, 1978;Brauer, Chaban and Gerber, 2004) the wavefunction was expressed as a product of one-dimensional parts,

$$\Psi(Q_1,...,Q_M) \approx \prod_{i=1}^{M} \psi_i(Q_i),$$

Anharmonic Corrections

For large molecules anharmonic corrections are not usually calculated because of the computer limitations. However, as the precision of calculated vibrational frequencies increases, the harmonic approximation may one day become a bottleneck in the simulation and interpretation methods. Therefore, in the past, we tried to explore the correction possibilities at least for smaller systems as well as to draw some conclusions also for peptides and proteins (Bouř, 1994a;Bouř, 1994b;Bouř and Bednárová, 1995;Daněček and Bouř, 2007a;Daněček and Bouř, 2007b;Dračínský, Kaminský and Bouř, 2009;Kapitán, Hecht and Bouř, 2008;Tam et al., 1997;Tam, Bouř and Keiderling, 1996).

Contrary to the harmonic approximation, computation of the anharmonic energies of molecules is not a black box method. Every system behaves differently. Leaving the harmonic Hamiltonian, a plethora of additional terms in the potential appears, including purely vibrational terms, as well as those responsible for the vibrational-rotational interaction (Papoušek and Aliev, 1982;Wilson, Decius and Cross, 1980). For very flexible molecules, such as the toluene, perturbation treatment of the anharmonicities is not possible. In such case

local coordinate system, yielding the corrected atomic polar tensor.

We have developed a set of the b_i parameters for N-methylacetamide in water clusters by referencing to DFT computed force field results (Bouř, 2004;Bouř and Keiderling, 2003). The approach is especially suitable to combined QM/MM modeling, as the correction is fast and allows for averaging of large ensembles of geometries (Figure 6). The electrostatic correction can reproduce *ab initio* results. Moreover, the changes observed in α-helical peptide spectra under deuteration were correctly predicted for the first time by this method (Bouř, Michalík and Kapitán, 2005). Using the electrostatics, the complicated IR and VCD spectral patterns of β-hairpins can be better understood in terms of the electrostatic interactions between the solvent and the amide groups, or in terms of shielding of some groups by the side chains (Bouř and Keiderling, 2005b).
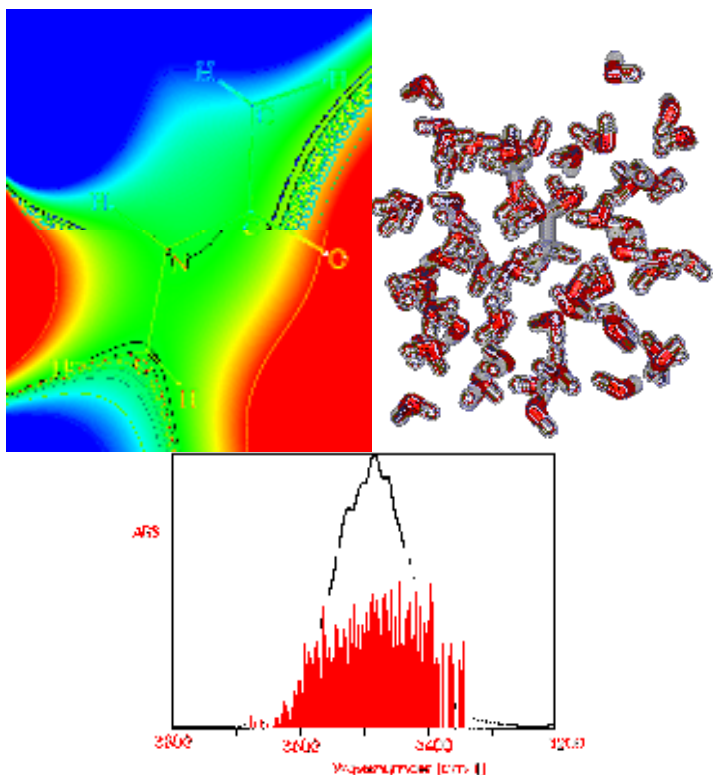
Figure 6. Calculated average solvent electrostatic potential (red-positive, blue-negative) for N-methylacetamide in water. By averaging many MD configurations, we get inhomogeneous band width, as for the absorption spectrum in the bottom.
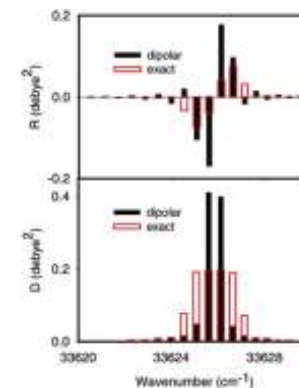

Figure 7. Patterns of rotational (R) and dipolar (D) strengths obtained for an ideal helix 300-nm long by the generalized formulae and the usual dipolar approximation.

Figure 8. Transition dipole contributions to selected vibrational states: For the regular helix state number 6 is not spectrally active, whereas delocalized states 100 and 101 provide large dipolar and rotational strengths. Similarly in the nucleosome (right) partially delocalized mode 2' has larger spectral signals than the localized state 73' (Andrushchenko and Bouř, 2008).

$$D = \sum_i c_{Ei}{}^2 \mu_i{}^2 + 3\sum_{i<j} c_{Ei} c_{Ej} \boldsymbol{\mu}_i \cdot \left(\left(\mathbf{1} - \mathbf{r}_{0ij}\mathbf{r}_{0ij}\right) j_0\left(kr_{ij}\right) - \left(\mathbf{1} - 3\mathbf{r}_{0ij}\mathbf{r}_{0ij}\right) \frac{j_1\left(kr_{ij}\right)}{kr_{ij}} \right) \cdot \boldsymbol{\mu}_j$$

$$R = -\frac{3c}{2} \sum_{i<j} c_{Ei} c_{Ej} \boldsymbol{\mu}_i \cdot \mathbf{r}_{ij} \times \boldsymbol{\mu}_j \frac{j_1\left(kr_{ij}\right)}{r_{ij}}.$$

Finally, the expressions have the pleasant property to converge to the common dipolar and rotational strength for small-system (large light wavelength) limit. For example, for the TDC approximation, we obtain

$$D_{k\to 0} = \sum_i c_{Ei}{}^2 \mu_i{}^2 + 2\sum_{i<j} c_{Ei} c_{Ej} \boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j$$

and

$$R_{k\to 0} = -\frac{ck}{2} \sum_{i<j} c_{Ei} c_{Ej} \mathbf{r}_{ij} \cdot \boldsymbol{\mu}_j \times \boldsymbol{\mu}_i ,$$

i.e. the classical textbook intensities.

In principle, larger molecules thus provide quite a different intensity pattern than expected from the standard dipolar theory (Figure 7). In reality, many of the "short wavelength" effects disappear because of limited resolution of spectrometers. For regular structures, however, large-scale phonon-like excitations are possible, and these are responsible for the unusual enhancement of CD and VCD intensities (Figure 8).

Large Systems, ψ-CD

For large systems another complication arises: their size can be comparable with the wavelength of the light, and standard light absorption/scattering theories are not applicable. Therefore, we extended the theory (Andrushchenko and Bouř, 2008) and applied it to nucleosome and DNA particles.

Historically, this phenomenon was first encountered for DNA, but also later experiments done for very large biopolymers suggest that they may behave in some aspects differently than smaller ones. For example, first spectroscopic characterisation of condensed DNA was done by Lerman and coworkers (Lerman, 1971). They noticed that DNA condensed by a combination of neutral polymers and high salt concentration produces atypical electronic circular dichroism (ECD) spectra intensity of which was enhanced if compared to standard B-DNA signal. These atypical ECD spectra were called ψ-type CD (psi for "polymer and salt induced",(Lerman, 1971). Later other DNA condensations, *e.g.* those achieved by multivalent metal ions or H1 histone, produced similar enhancements (Cowman and Fasman, 1978;Evdokimov et al., 1976;Fasman et al., 1970;Gersanovski et al., 1985;Gosule and Schellman, 1976;Khadake and Rao, 1995;Lerman, 1971;Liao

and Cole, 1981). Similar effect was found for protein fibrils later (Maa et al., 2007), although large proteins systems are obviously rarer than for DNA.

To describe the absorption of light on systems comparable in size with the wavelength, we have to generalize the theory. Its starts from the Fermi golden rule for the time derivative of transition probability between states G and E,

$$\frac{dP_{GE}}{dt} \approx \frac{2\pi}{\hbar^2}\left|\left\langle G\left|V_0\right|E\right\rangle\right|^2 \delta(\omega_E - \omega_G - \omega),$$

where $V(t) = V_0^\dagger e^{i\omega t} + V_0 e^{-i\omega t}$ is a perturbation potential dependent on the angular frequency ω, $\hbar$ is the Planck constant, δ is the Dirac delta-function. We can use the standard semiclassical approach ("classical field and quantum molecule"), but the usual multipole expansion is not possible.

After much algebra, we arrived to fairly simple formula for generalized dipolar and rotational strengths not restricted by molecular size. In particular, the dipole (D) and rotational (R) strengths are given by

$$D = \frac{\hbar}{2\omega}\left[\sum_\lambda \left(\sum_\alpha S_{\lambda\alpha}^i \mathbf{P}_{\lambda\alpha}\right)^2 + 3 \sum_{\lambda<\lambda',\alpha,\alpha'} S_{\lambda\alpha}^i S_{\lambda'\alpha'}^i \mathbf{P}_{\lambda\alpha}\mathbf{P}_{\lambda'\alpha'}\cdot\cdot\right.$$

$$\left.((1 - \mathbf{R}_{0\lambda\lambda'}\cdot\mathbf{R}_{0\lambda\lambda'})j_0(kR_{\lambda\lambda'}) - (1 - 3\mathbf{R}_{0\lambda\lambda'}\cdot\mathbf{R}_{0\lambda\lambda'})\frac{j_1(kR_{\lambda\lambda'})}{kR_{\lambda\lambda'}})\right]$$

and

$$R = i\hbar \sum_{\lambda,\alpha,\alpha'} S_{\lambda\alpha}^i S_{\lambda'\alpha'}^i \mathbf{P}_{\lambda\alpha}\cdot\mathbf{M}_{\lambda\alpha'} + \frac{3}{4}\sum_{\lambda>\lambda',\alpha,\alpha'} S_{\lambda\alpha}^i S_{\lambda'\alpha'}^i \{\frac{c}{\omega}\mathbf{P}_{\lambda\alpha}\mathbf{P}_{\lambda'\alpha'}\cdot\cdot\mathbf{R}_{0\lambda\lambda'}$$

$$\times \mathbf{1}j_1(kR_{\lambda\lambda'}) + [(\mathbf{O}_{\lambda\alpha}\mathbf{P}_{\lambda'\alpha'} + \mathbf{O}_{\lambda'\alpha'}\mathbf{P}_{\lambda\alpha})\cdot\cdot\mathbf{R}_{0\lambda\lambda'}\mathbf{1}\times\mathbf{R}_{0\lambda\lambda'}R_{\lambda\lambda'}^{-1}$$

$$+ 2\hbar i(\mathbf{P}_{\lambda\alpha}\mathbf{M}_{\lambda'\alpha'} + \mathbf{P}_{\lambda'\alpha'}\mathbf{M}_{\lambda\alpha})\cdot\cdot\mathbf{R}_{0\lambda\lambda'}\mathbf{R}_{0\lambda\lambda'}](\frac{3j_1(kR_{\lambda\lambda'})}{kR_{\lambda\lambda'}} - j_0(kR_{\lambda\lambda'}))$$

$$+ 2\hbar i(\mathbf{P}_{\lambda\alpha}\cdot\mathbf{M}_{\lambda'\alpha'} + \mathbf{P}_{\lambda'\alpha'}\cdot\mathbf{M}_{\lambda\alpha})(1 - \frac{j_1(kR_{\lambda\lambda'})}{kR_{\lambda\lambda'}})\}$$

where the symbols are in detailed explained in the dissertation. Here, we want to point out the most important aspects. Firstly, it was possible for the first time to derive expressions for absorption and VCD peak areas beyond the dipolar approximation. Also, averaging for isotropic samples (solutions) was possible via the Bessel functions of the zero's and first kind (j_0, j_1).

Within the simplified transition dipole coupling (TDC) mechanism we obtain analogous generalized strength for an ensemble of oscillators (Andrushchenko and Bouř, 2008):