Veera Krasnenko[1]
Alan H. Tkaczyk[1]
Eric R. Tkaczyk[1]
Ödön Farkas[2]
Koit Mauring[1]
[1] *Institute of Physics, University of Tartu, Riia 142, 51014 Tartu, Estonia*

[2] *Department of Organic Chemistry, Eötvös Loránd University, P. O. Box 32, H-1518 Budapest 112, Hungary*

Vibrations-Determined Properties of Green Fluorescent Protein

Received 12 June 2004; revised 12 February 2005; accepted 14 February 2005
Published online 9 March 2005 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.20264

Abstract: *The physicochemical characteristics of the green fluorescent protein (GFP), including the thermodynamic properties (entropy, enthalpy, Gibbs' free energy, heat capacity), normal mode vibrations, and atomic fluctuations, were investigated. The Gaussian 03 computational chemistry program was employed for normal mode analysis using the AMBER force field. The thermodynamic parameters and atomic fluctuations were then calculated from the vibrational eigenvalues (frequencies) and eigenvectors. The regions of highest rigidity were shown to be the β-sheet barrel with the central α-helix, which bears the chromophore. The most flexible parts of the GFP molecule were the outlying loops that cover the top and bottom of the β-barrel. This way, the balance between rigidity and flexibility is maintained, which is the optimal relationship for protein stability in terms of Gibbs' free energy. This dual-schemed structure satisfies the requirements for GFP function. In this sense, the structure of GFP resembles a nanoscale drum: a stiff cylinder with flexible vibrating end(s).* © 2005 Wiley Periodicals, Inc. Biopolymers 78: 140–146, 2005

This article was originally published online as an accepted preprint. The ''Published Online'' date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

Keywords: *green fluorescent protein; molecular mechanics calculation; AMBER force field; normal modes; density of vibrational states; thermodynamic parameters; thermal fluctuations*

INTRODUCTION

Only a decade has passed since the discovery of the green fluorescent protein (GFP) present in the jellyfish *Aequorea victoria*, but already it has become a widely studied protein and commonly used tool in biochemistry and cell biology. This protein includes an effectively emissive fluorophore, which promotes

Correspondence to: Koit Mauring; email: mauring@fi.tartu.ee
Contract grant sponsor: Estonian Science Foundation. Contract grant number: 5546

its widespread use. High-resolution crystal structures of GFP offer a good basis to computationally model its spectroscopic properties.[1]

Usually, proteins function due to their flexibility and ability to transition between conformational states. In contrast, the situation with GFP is different. Its main function, to emit bright fluorescence, requires rigidity to prevent fluorescence quenching due to isomerization in the excited state.[2,3] When the chromophore is removed from its natural environment, this rigidity is lost, and fluorescence is quenched.[4] Also, the fluorescence of the synthetic GFP chromophore in vitro is extremely weak,[5] as is that of denatured GFP.[4] The chromophore is both protected from quenching agents and also stiffened by hydrogen bonds to GFP's barrel-like structure. A light-induced dark state has also been discovered in single-molecule experiments.[6,7]

Protein dynamics can be treated as a transition between minima of the potential energy landscape. Since proteins have a high number of degrees of freedom, the potential energy landscape is highly multidimensional. It is possible to investigate the transitions between minima by molecular dynamics (MD) methods. MD has provided valuable information about GFP on the transitions between conformational states and the dynamics of hydrogen bonding.[8] The MD approach is still limited by today's computational efficiency. Many of the processes occurring in proteins have time scales inaccessible to such MD calculations, since current computers can model a protein for only a few nanoseconds. Another approach to gain insight into the dynamics of the protein is to calculate vibrations in harmonic approximation and get detailed information about fluctuations of every single atom in the protein. Present computational capabilities permit the calculation of normal modes of proteins only by molecular mechanics (MM) methods and are not sufficient to allow full quantum mechanical modeling. The MM method has been applied to proteins for calculating vibrational normal modes[9] such as the hinge-bending mode,[10] which is crucial for the enzymes to operate. In the case of the barrel-like GFP molecule, these types of normal mode are not to be expected.

The low-frequency modes represent the collective dynamics of the protein backbone.[11] Visualization programs allow each vibrational mode to be viewed and easily assigned. Computation of the vibrations of GFP yields information about thermal fluctuations and rigidity. We have also calculated thermodynamic parameters, which control the equilibrium between the different conformations of the protein. It is well known that chemical processes are governed by Gibbs' free energy. The calculation of the enthalpic and entropic components and their temperature dependence yields insight into the thermodynamic background of two conformations of a GFP variant cyan fluorescent protein revealed by NMR spectroscopy.[12] The normal mode analysis provides the data on temperature dependence of the heat capacity, which is important for assessing the thermal stability of the protein.

METHODS

We applied the computational chemistry program Gaussian 03 to calculate the vibrational normal modes of GFP.[13] The Hessian was mass weighted, transformed into internal coordinates, and diagonalized. Next, the frequencies, reduced masses, force constants, and displacements were calculated. The calculation required more than 4 gigabytes of RAM. The atomic coordinates of GFP were obtained from the Protein Data Bank file 1EMB.[14] The geometry was optimized with tight optimization criteria: the maximum component and root mean square (RMS) of the forces were $4 \cdot 10^{-6}$ and $2.5 \cdot 10^{-6}$ Hartree/Bohr, the maximum and RMS of the displacement for the next optimization step were $1.5 \cdot 10^{-5}$ and $1 \cdot 10^{-5}$ Bohr, respectively.

The hydrogens were added by the GaussView program.[13] The charges of the chromophore atoms were calculated by the density functional method up to the B3LYP/6-31G(d) level. The molecular mechanics force field AMBER96 was applied,[15] and the force field parameters for the chromophore were adopted from Reuter et al.[16] Reuter et al. used the CHARMM program where the energy is calculated analogously with the AMBER force field. The energy consists of six terms: chemical bonds, bond angles, bond twisting, van der Waals, electrostatic, and hydrogen bonds:

$$\begin{aligned} E_{\text{total}} = & \sum_{\text{bonds}} K_r (r_i - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta_i - \theta_{\text{eq}})^2 \\ & + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ & + \sum_{\substack{\text{nonbond} \\ \text{\_VDW}}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right] \\ & + \sum_{\substack{\text{nonbond} \\ \text{\_el.}}} \frac{q_i q_j}{\varepsilon R_{ij}} + \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right] \end{aligned} \tag{1}$$

where K_r, K_θ, V_n, A_{ij}, B_{ij}, C_{ij}, and D_{ij} are constants; r_i is the length of the bond and r_{eq} is the equilibrium distance; θ_i is the bond angle value; θ_{eq} is the equilibrium angle value; ϕ and γ are dihedral angles; n is the periodicity; R_{ij} is distance between the two atoms; $q_i q_j$ are the charges of the atoms; and ε is effective dielectric constant.[15]

Multiple thermodynamic parameters—heat capacity C_p, Gibbs' free energy G, enthalpy H, and entropy S—were cal-

FIGURE 1 Low-frequency normal mode $v = 7\ \mathrm{cm}^{-1}$ of GFP visualized as vectors of vibrational amplitudes. The hydrogen atoms are omitted.

culated from the normal mode frequencies v_i according to Eqs. (2)–(5)[17]:

$$C_{\mathrm{p\_vibr}} = k\sum_{i=1}^{3N-6}\frac{(hv_i/kT)^2}{2(ch(hv_i/kT)-1)} \tag{2}$$

$$G_{\mathrm{vibr}} = kT\sum \ln\left[1-e^{-(hv_i/kT)}\right] \tag{3}$$

$$S_{\mathrm{vibr}} = k\sum\left\{-\ln\left[1-e^{-(hv_i/kT)}\right]+\frac{(hv_i/kT)e^{-(hv_i/kT)}}{1-e^{-(hv_i/kT)}}\right\} \tag{4}$$

$$H_{\mathrm{vibr}} = kT\sum\frac{(hv_i/kT)e^{-(hv_i/kT)}}{1-e^{-(hv_i/kT)}} \tag{5}$$

RESULTS AND DISCUSSION

GFP contains 226 amino acids with 3799 atoms (including hydrogens) totaling 11,391 normal modes. Vibrational frequencies were calculated in harmonic approximation, and all hindered rotations were treated as harmonic oscillators. The torsional barriers are high in comparison with room temperature thermal energy kT, and the discrepancy introduced by harmonic approximation is acceptable.

The vibrational frequencies span from 6 to 3750 cm^{-1}. All low-frequency vibrational modes have a collective character. As an example, the second lowest-frequency normal mode is visualized in Figure 1. The molecule vibrates as a quasirigid body, which can be revealed from the distribution of the vectors of similar amplitude all over the molecule.

To obtain a smooth spectrum of vibrations, normal modes were sorted into bins by 10 cm^{-1}. The most important region of the normal mode spectrum is below 150 cm^{-1} since these frequencies contribute the most to high amplitude fluctuations. The low frequency maximum of the distribution of vibrational states is situated at 75 cm^{-1} (Figure 2), whereas between 250 and 1250 cm^{-1} the density is almost flat.

The fingerprint region of vibrations has two prominent peaks at 1430 and 1700 cm^{-1} and some minor maxima in the vicinity. Typically, the amide band in a protein backbone is characterized by three groups of vibrational frequencies in the range of 1300–1650 cm^{-1}. There are almost no vibrations in the range of 1800–2800 cm^{-1}. The peaks between 2800 and 3750 cm^{-1} belong to hydrogen vibrations: C—H modes have frequency of 3000–3100 cm^{-1}, whereas O—H is situated in the 3400–3500 cm^{-1} range.

As we can see from Figure 2, the majority of the vibrational frequencies have energy larger than 207 cm^{-1}, the thermal energy at room temperature. By virtue of the Boltzmann factor $e^{-(hv_i/kT)}$, the vibrational levels above $v = 0$ are populated only for lower frequency modes, which provide the largest contribution to the heat capacity, entropy, and enthalpy. In

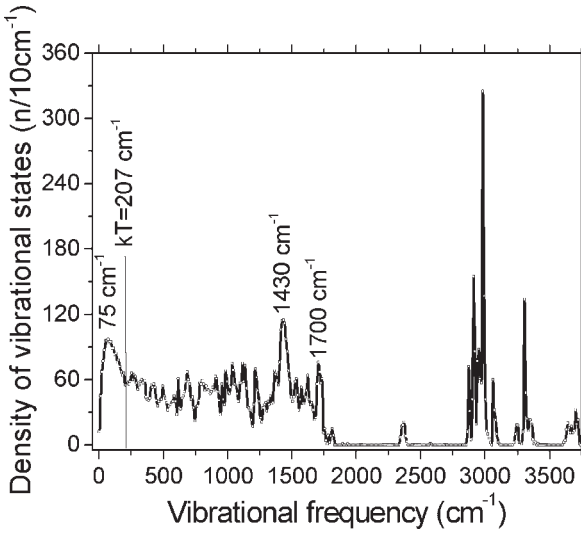

FIGURE 2 Histogram of 11391 normal modes of the GFP molecule. Number of normal modes in each interval of 10 cm^{-1} is shown.

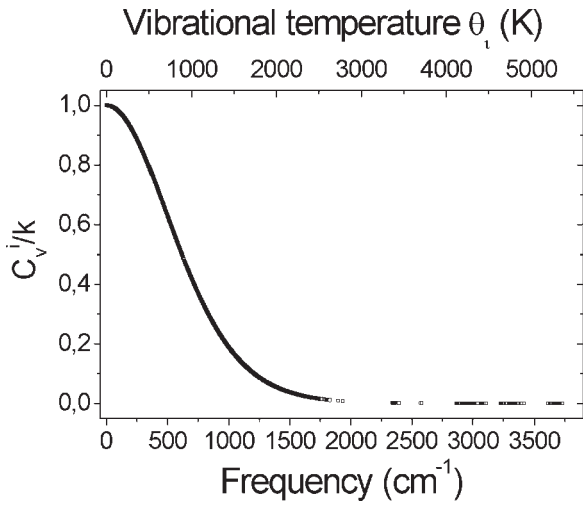

FIGURE 3 Contribution of vibrational modes to GFP heat capacity at room temperature.

Figure 3, the heat capacities of individual vibrational modes at room temperature are depicted. Modes with frequencies <375 cm^{-1} contribute half of the total heat capacity while those with frequencies >1500 cm^{-1} give only 1.5%.

The temperature dependence of heat capacity (Figure 4) is almost linear starting from 100 K. Since there is no experimental data on GFP thermodynamic properties, we compared our heat capacity with the experimental data of Di Lorenzo et al. on β-bactoglobulin.[18] The latter have similar shape, although in the region of 300 K, their heat capacity exhibits slight concavity, while our calculation results in convex form. The computed temperature dependence of the GFP heat capacity is in agreement with the general

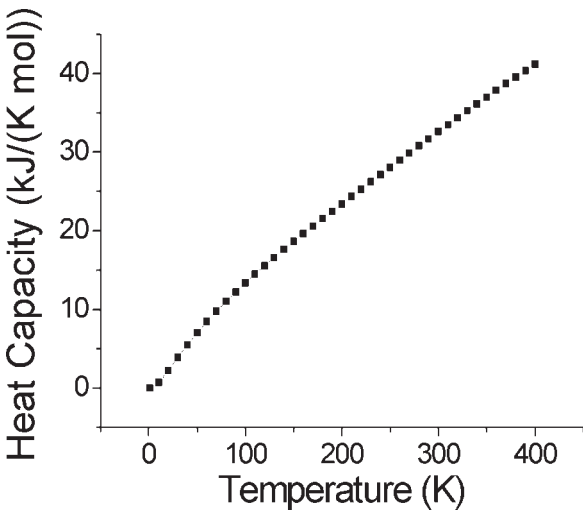

FIGURE 4 The dependence of heat capacity on temperature.

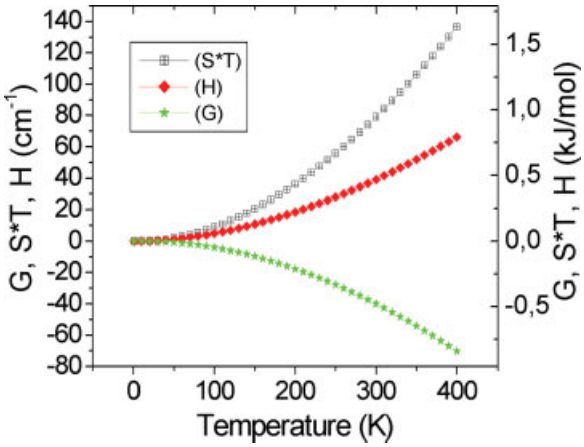

FIGURE 5 Temperature dependence of thermodynamic parameters of GFP (divided by number of vibrations).

thermal dependence trends observed for heat capacities of globular proteins such as bovine β-lactoglobulin. At 300 K, the heat capacity according to our calculations is 32.6 kJ/(K mol), whereas Di Lorenzo et al. report experimental values of 28.7 kJ/(K mol) for bovine β-lactoglobulin,[18] for example.

The vibrational spectrum gives the opportunity to calculate multiple thermodynamic parameters, which have prime importance for understanding protein thermal stability. Gibbs' free energy is the state function that governs the transitions between protein conformational states.

As seen from Figure 5, the entropy component S * T of G and enthalpy both grow with temperature. The entropic component increases more steeply than enthalpic, which is beneficial for stability, although the exponents are similar ($S * T \sim T^{1.92}$ and $H \sim T^{1.84}$). The graph for G decreases monotonically, indicating the stabilization of protein at room temperature with respect of cryogenic temperatures. It is well known that proteins denature at low temperatures (at $T < 273$ K), but also at high temperatures. Our graph reveals the entropic driving force of the former process, whereas it does not describe the denaturation resulting from heating, since at higher temperatures the anharmonicity of vibrations should be taken into account. It should also be noted that our calculations ignore water–protein interactions and thus are not intended for a rigorous description of the denaturation process.

The RMS fluctuations of atoms are described by:

$$\langle \Delta \mathbf{R}^2 \rangle_{\text{har}} = \frac{kT}{m} \sum_{j=1}^{3N-6} \frac{|\mathbf{a}_j|^2}{\omega_j^2} \tag{6}$$

where k is the Boltzmann constant, T is the absolute temperature, N is the number of atoms, ω_f is the

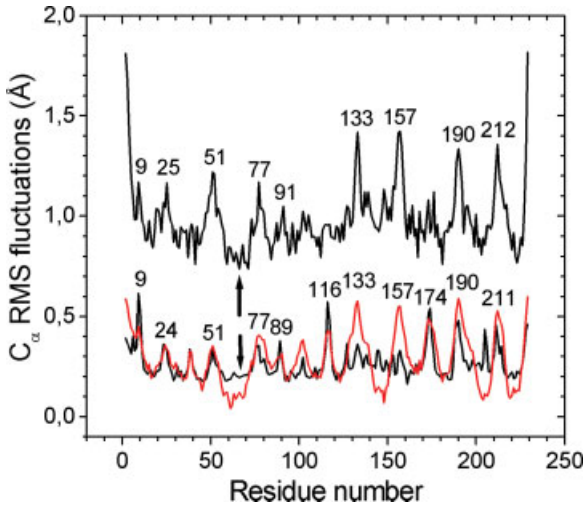

FIGURE 6 RMS fluctuations of GFP backbone C_α atoms. Upper curve: experimental displacements according to temperature factors in the X-ray structure file 1EMB.pdb. Lower curves: fluctuations calculated from eigenvectors of normal modes (black line) and experimental displacements according to temperature factors in the X-ray structure file 1GFL.pdb (red line, shifted down by 0.65 Å). The arrows indicate the location of the chromophore.

frequency of the *j*th normal mode, and $\mathbf{a}_j$ is the vector of the projections of the *j*th normal mode onto the Cartesian components of the vector **R** for the atom of interest.[19]

The experimental fluctuations can be determined from the X-ray structure temperature factors B_i in the Protein Data Bank file according to[20]:

$$\langle u_i^2 \rangle_{\text{exp}}^{1/2} = (3B_i/8\pi^2)^{1/2} \tag{7}$$

The RMS fluctuations of the backbone C_α atoms are depicted in Figure 6. The agreement between the experimental and calculated values is fair. Almost every peak in the experimental data has a counterpart on the calculated curve, although the magnitude of the calculated fluctuations is smaller. The computation yields high fluctuations for residues around 116 and 174 while X-ray data (1EMB) indicate considerably smaller level of mobility. On the other hand, the calculated fluctuations of these residues have excellent agreement with those revealed from the X-ray data of the GFP WT dimer (PDB file 1GFL[21]). The molecular dynamics data of Reuter et al.[16] confirm that the WT dimer data agree much better with the calculation. The average of C_α fluctuations is 0.986 Å for the X-ray data and 0.266 Å for the normal mode data. The most flexible residues are located in the loops connecting the β-strands as shown in Figure 7. This is to be expected, because the side chains of these residues point radially outward from the protein, preventing stiffening due to binding. This is in agreement with the MD data on GFP[8] findings of Dauber-Osguthorpe et al., who report that low-frequency high amplitude motions are concentrated into surface hairpin loops.[22] In contrast, the residues involved in the β-sheets of the GFP cylinder have small fluctuations due to tight hydrogen-bonding between the strands of β-barrel. Our findings indicate that the regions of regular secondary structure (β-strands and α-helices) have smaller RMS displacements, as reported by Gō et al.[23] The chromophore and the α-helix carrying it are well fastened to the surrounding amino acids and show considerable rigidity. The rigidity of the chromophore is crucial for it to emit luminescence, as mentioned previously.

An interesting question about GFP photophysics is the role of the His148 residue. Seifert et al.[12] have revealed by NMR and ^{19}F relaxation measurements a slow exchange process between two conformations of cyan fluorescent protein (having a tryptophan ring in the chromophore). They found almost complete entropy–enthalpy compensation for the slow exchange process between two states and proposed the His148 residue is responsible for the conformational mobility. We calculated the vibrations and thermodynamic parameters of the GFP (1EMG) conformer with the His148 residue

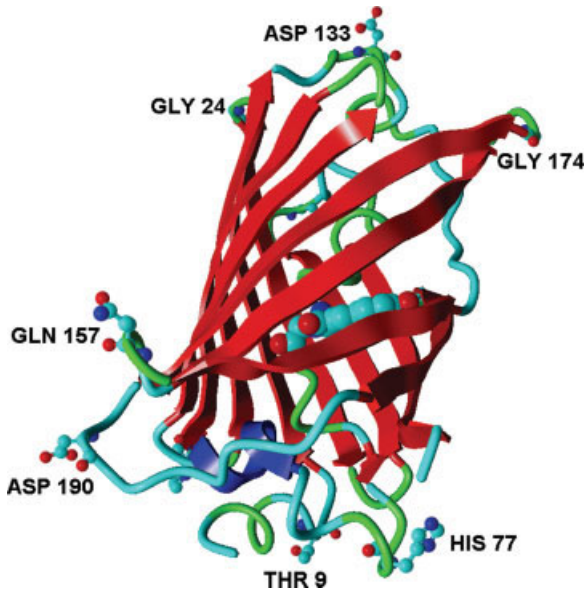

FIGURE 7 The backbone of GFP. The most fluctuating residues are labeled and shown as ball and stick models.

pointing outward from the GFP barrel (His148 isomerization was performed by the GaussView program[13]). The results indicate higher entropy for the conformer due to softening of the protein structure. With a temperature increase, the equilibrium shifts in favor of this high-entropic conformer. The calculation yields only partial entropy–enthalpy compensation: the difference in enthalpy $\Delta H = 1.4$ kJ/mol and $T\Delta S = 8.5$ kJ/mol. The absence of a higher degree of compensation can be understood in terms of the restricted computational model: surrounding solvent molecules could not be included into the calculation because of computational cost.

CONCLUSIONS

Physicochemical characteristics important for GFP's function as a fluorescent marker were investigated utilizing computational data. Thermodynamic parameters were calculated from vibrational normal mode analysis, based on MM calculation using the AMBER force field. The computed temperature dependence of GFP heat capacity is in agreement with the experimentally observed thermal dependence of protein heat capacity.[18] The vibrational analysis of the GFP conformer with the His148 residue pointing outward from the GFP barrel indicates that this conformational transition is characterized by entropy–enthalpy compensation in line with the published NMR experimental data.

The calculated fluctuations from MM-based normal mode analysis are in accordance with the experimentally observed fluctuations obtained from Protein Data Bank temperature factors. The regions of highest rigidity were shown to be the β-sheet barrel with central a α-helix, which bears the chromophore. The most flexible parts of the GFP molecule were identified by normal mode analysis as outlying loops that cover the top and bottom of the β-barrel. This structural observation is in agreement with the known requirements for GFP function: the α-helix and β-barrel provide the chromophore with the necessary stiffness for it to emit fluorescence. On the other hand, the external loops are allowed to fluctuate with greater magnitude, serving as a vibrational energy accumulator. This way, the balance between rigidity and flexibility is maintained, which is the optimal relationship for protein stability in terms of Gibbs' free energy. This dual-schemed structure satisfies the requirements for GFP function. In this sense, the structure of GFP resembles a nanoscale drum: a stiff cylinder with flexible vibrating ends.

We thankfully acknowledge the support from the Estonian Science Foundation under grant number 5546. AHT is grateful for support from a Fulbright Grant (U.S. Student Program).

REFERENCES

1. Tsien, R. Y. Annu Rev Biochem 1998, 67, 509–544.
2. Voityuk, A. A.; Michel-Beyerle, M.-E.; Rösch N. Chem Phys Lett 1998, 296, 269–276.
3. Weber, W.; Helms, V.; McCammon, J.; Langhoff, P. W. Proc Natl Acad Sci USA 1999, 96, 6177–6182.
4. Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. I. Proc Natl Acad Sci USA 1996, 93, 13617–13622.
5. Mandal, D.; Tahara, T.; Webber, N. M.; Meech, S. R. Chem Phys Lett 2002, 358, 485–501.
6. Dickson, R.; Cubbit, A.; Tsien, R.; Moerner, W. Nature 1997, 388, 355–358.
7. Moerner, W. E. J Chem Phys 2002, 117, 10925–10937
8. Helms, V.; Straatsma, T. P.; McCammon, J. A. J Phys Chem B 1999, 103, 3263–3269.
9. Brooks, B.; Karplus, M. Proc Natl Acad Sci USA 1983, 80, 6571–6575.
10. Brooks, B.; Karplus, M. Proc Natl Acad Sci USA 1985, 82, 4995–4999.
11. Levitt, M.; Sander, C.; Stern, P. S. Int J Quant Chem: Quant Biol Symp 1983, 10, 181–199.
12. Seifert, M. H. J.; Ksiazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A. J Am Chem Soc 2002, 124, 7932–7942.
13. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, Ö.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 03, Gaussian, Inc., Pittsburgh, PA, 2003.
14. Brjec, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormö, M.; Remington, S. J. Proc Natl Acad Sci USA 1997, 94, 2306–2311.
15. Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. In Computer Simulation of Biomolecular Systems; Wilkinson, A., Weiner, P., Van Gunsteren, W., Eds.; Elsevier: New York, 1997; Vol 3A, pp 83–96.

16. Reuter, N.; Lin, H.; Thiel, W. J Phys Chem B 2002, 106, 6310–6321.
17. Volkenstein, M.; Eljashevits, M.; Stepanov, B. Kolebanija molekul; Gostechizdat: Moscow, 1949; T.2, Chap 27.
18. Di Lorenzo, M.; Zhang, G.; Pyda, M.; Lebedev, B.; Wunderlich, B. J Polym Sci B: Polym Phys 1999, 37, 2093–2102.
19. Brooks, B. R.; Janežič, D.; Karplus, M. J Comput Chem 1995, 16, 1522–1542.
20. James, R. W. The Optical Principles of the Diffraction of X-Rays; Ox Bow: Woodbridge, 1962.
21. Yang, F.; Moss, L. G.; Phillips, G. N., Jr. Nat Biotechnol 1996, 14, 1246–1251.
22. Dauber-Osguthorpe, P.; Osguthorpe, D. J.; Stern, P. S.; Moult J. J Comput Phys 1999, 151, 169–189.
23. Gō, N.; Noguti, T.; Nishikawa, T. Proc Natl Acad Sci USA 1983, 80, 3696–3700.