STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, LV, 1, 2010

COMPUTATIONAL DESCRIPTION OF PEPTIDE ARCHITECTURES BASED ON HYDROGEN BONDS

RADU SILAGHI-DUMITRESCU[a]

ABSTRACT. The geometrical structures of systems whose architectures rely entirely on hydrogen bonds, such as α-helical polypeptides, are not always properly described with current computational methods.

Keywords*: peptide, density functional, semiempirical, Hartree-Fock, hydrogen bond*

INTRODUCTION

Computational examination of enzyme mechanisms employing such methods as density functional theory (DFT) or Hartree-Fock (HF) has traditionally been restricted, especially in metalloproteins containing transition metal ions, to smaller-size models of the active sites, due to the large computing resources required.[1-5] In recent years, the use of accurate solvation models and QM/MM-type techniques have allowed for the influence of the protein environment to be taken into account more explicitly, albeit with treatment of the less-relevant parts of the polypeptide at levels of theory inferior to those applied to the active site models.[1-5] Additionally, there have been constant efforts to improve the performance of the 'lower-level' theory levels (semiempirical, molecular mechanics) so that they can be applied not just to the distant polypeptide regions, but also to the actual active site.[6-9] At the current rate of increase in computer performance, it is not unfeasible to consider that in the near future it will be possible to apply DFT or (post-)-HF methods to enzyme models consisting of the entire protein. Already at the present stage, single-point energies have occasionally been computed for entire protein models at higher levels, such as DFT.[9-21] It is in this context that the present study seeks to estimate the performance of some of the commonly used computational models in predicting accurate geometries of polypeptide chains. For some of these methods, this is likely to be a particularly challenging situation. First, polypeptide architectures tend to rely primarily on weak, non-covalent, interactions (mainly hydrogen bonding). Second, those methods requiring parametrization may not necessarily perform efficiently if

[a] *Universitatea Babeş-Bolyai, Facultatea de Chimie şi Inginerie Chimică, Str. Kogălniceanu Nr. 1, RO-400084 Cluj-Napoca, Romania, rsilaghi@chem.ubbcluj.ro*

they have not been parametrized for proteins; given the unexpected structural and reactivity findings occasionally uncovered by experiment, even properly parametrized methods may in principle fail in some cases.

RESULTS AND DISCUSSION

Shown in Figure 1 and Table 1 are the results of geometry optimizations on the decaglycine helix. At the HF/3-21G** level, the computed geometry appears funnel-shaped as opposed to the purely helical starting point. Although 1:1 peptide CO---HN hydrogen bonds appear conserved, they are distorted (decreased C-O-H and N-H-O angles, by up to 47º) and elongated by at least 0.1-0.4 Å in terms of the N---O distance and 0.2-0.5 Å in term of the O-H distance) compared to the canonical structure. It is also notable that the parameters listed in Table 1 for the HF geometry vary significantly between the six peptide bonds involved in CO---HN hydrogen bonding, while there would be no such differences in a canonical helix. This HF geometry to our knowledge has no known equivalent in experimental chemistry.

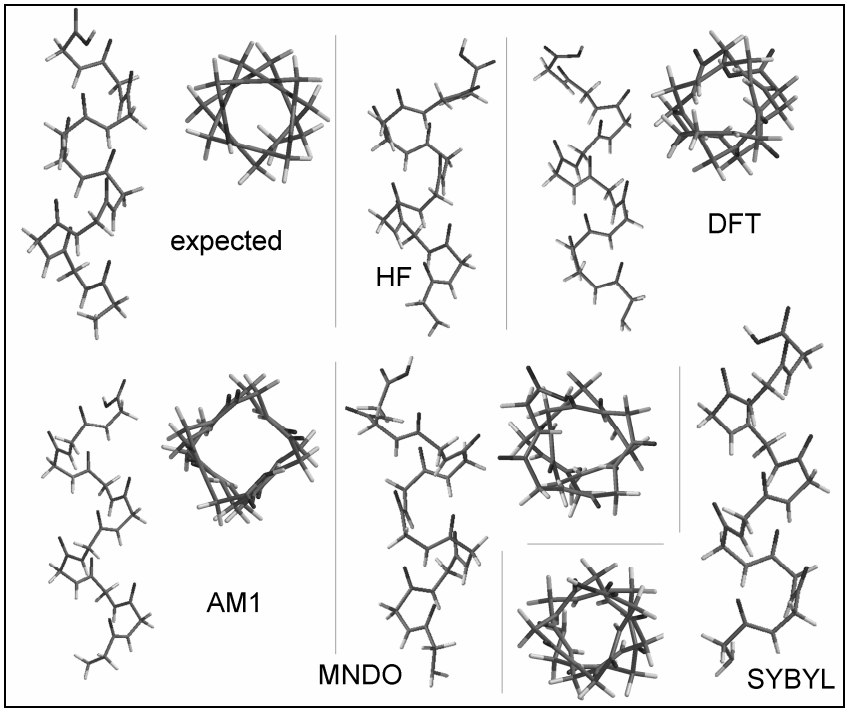

Figure 1. Canonical form of an alpha-helical deca-glycine, compared to geometries obtained by geometry optimization with various computational methods. (side-views and top views).

In the BP86/6-31G** geometry the regularity of the helix is also lost, and the resulting geometry is in-between the canonical form of Figure 1 and the artefactual HF discussed above. Thus, the CO---HN hydrogen bonds feature O-H distances ranging from 2.01 to 2.63 Å, i.e. both irregular and inappropriately long; they are also distinctly less linear (by up to 40º) than in a canonical helix.

Table 1. Key geometrical parameters (distances in Å, angles in º) for the deca-glycine model and its computed geometries at various theory levels. Values are given for each peptide bond starting from the N-terminus of the peptide; the internal parameters of the peptide bonds (C=O, N-H and OC-NH bond lengths) are given as averages since they do not vary significantly (i.e., by more than 0,01 Å) within a model computed with a given method.

Parameter	expected	Sybyl	MNDO	AM1	BP86	HF
O---H	1.72	2.40	3.31	2.33	2.01	2.20
	1.72	2.40	3.18	2.28	2.12	1.94
	1.72	2.50	3.25	2.37	2.09	2.03
	1.72	2.51	3.22	2.43	2.15	2.05
	1.72	2.41	3.23	2.16	2.63	1.93
	1.72	2.45	3.30	2.52	2.67	2.11
HN---O	2.74	3.17	4.30	2.96	2.98	3.16
	2.74	3.26	4.17	3.13	3.13	2.93
	2.74	3.26	4.23	3.15	3.08	2.99
	2.74	3.43	4.19	3.08	3.06	3.00
	2.74	3.40	4.21	3.15	3.36	2.90
	2.74	3.27	4.28	3.04	3.71	2.87
N-H---O	169º	133º	171º	151º	153º	164º
	169º	143º	169º	142º	167º	171º
	169º	133º	167º	134º	160º	161º
	169º	153º	166º	122º	146º	160º
	169º	171º	167º	172º	128º	164º
	169º	139º	166º	112º	167º	132º
C-O---H	168º	134º	146º	145º	121º	155º
	168º	135º	145º	150º	150º	158º
	168º	138º	139º	142º	132º	142º
	168º	134º	137º	145º	137º	152º
	168º	142º	138º	161º	135º	153º
	168º	121º	136º	120º	137º	122º
C=O	1.22	1.22	1.23	1.25	1.24	1.22
N-H	1.03	1.00	1.02	1.00	1.03	0.99
OC-NH	1.32	1.35	1.38	1.38	1.36	1.34

The PM3 (semiempirical) implementation examined here leads to an entirely non-helical geometry (not shown). Figure 1 and Table 1 show that the MNDO method performs somewhat better than the PM3, even though the hydrogen bonds are still elongated by ~2 Å compared to the canonical alpha helix. The AM1 method yields a funnel-type geometry similar to that of HF/3-21G**, and featuring a distinct loss of hydrogen-bond regularity, with the O-H distances at least 0.6 Å longer than in the canonical alpha-helical structure. When using the SYBYL forcefield, the O-H 'hydrogen-bonding' distances were also overestimated by >0.8 Å and the C=O---HN linearity was completely lost.

CONCLUSIONS

No two computational methods yield the same geometry for a deca-glycine helix, and none of these geometries correctly reproduces the canonical alpha-helical structure expected and in fact used as a starting point in the respective computations; the origin of these differences is most likely in the different degrees of accuracy with which these computational methods describe hydrogen bonds (and weak interactions in general); all of the methods tested here distinctly underestimate the strength of the hydrogen bonds, predicting O---H distances larger by up to 2 Å than expected, and N-H---O or CO---H angles by up to 40-50º smaller than expected. The methods employed here are often used in computational studies of large-scale systems; our results appear to point out to intrinsic deficiencies of all these methods in quantitatively describing systems whose architecture relies considerably on weak interactions.

MATERIALS AND METHODS

The present study employs an alpha-helical model in order to estimate and compare the performance of several levels of theory in providing accurate geometries. The methods tested here include Hartree-Fock (HF/3-21G**[22]) density functional (BP86/6-31G**[22]) semiempirical (MNDO[22], AM1[22], PM3[23]) and molecular mechanics (SYBYL[22]) applied with standard convergence criteria as defined in the respective software packages (Spartan[22] and Gaussian98[23] respectively); similar results, not discussed here, were obtained with the newer M05 functional as implemented in Gaussian03.[24] Figure 1 shows the canonical form expected for an alpha-helical stretch of decaglycine. This model is built within the Builder module of the Spartan software package,[22] capped with hydrogen atoms (i.e., C-terminal COOH, N-terminal –NH2) and not subjected to any further modifications. Characteristic for this structure is the regularity imposed by hydrogen bonds formed between C=O and NH groups of different peptide bonds. The length of the helix (free carboxy atom to free amino group) is 15.4 Å, and other parameters are listed

in Table 1. This model was then subjected to geometry optimization at various levels of theory, as detailed above. A zwitterionic version of this helix (capped with ammonium and carboxylate groups, respectively) as well as a 20-alanine helix were also examined with qualitatively similar results, and are not discussed further.

ACKNOWLEDGMENTS

Funding from the Romanian Ministry of Education and Research, (grants PN II 565/2007 and Parteneriate-FLUORODENT) and from the Romanian Academy (63/2007) is gratefully acknowledged.

REFERENCES

1. Noodleman, L.; Lovell, T.; Han, W. G.; Li, J., Himo, F. *Chem Rev*, **2004**, *104*, 459.
2. Siegbahn, P. E.,Blomberg, M. R. A. *Chem. Rev.*, **2000**, *100*, 421.
3. Harris, D. L. *Curr. Opin. Chem. Biol.*, **2001**, *5*, 724.
4. Friesner, R. A.,Dunietz, B. D. *Acc. Chem. Res.*, **2001**, *34*, 351.
5. Gooding, S. R.; Winn, P. J.; Jones, G. A.; Ferenczy, G. G.; Frusher, M. J., Reynolds, C. A. *J. Phys. Chem. A.,* **2006**, *110*, 6487.
6. Silaghi-Dumitrescu, R. *Rev. Chim.,* **2004**, *55*, 304.
7. Tejero, I.; Gonzalez-Lafont, A., Lluch, J. M. *J Comput Chem*, **2007**, *28*, 997.
8. McNamara, J. P.; Sundararajan, M.; Hillier, I. H.; Ge, J.; Campbell, A., Morgado, C. *J Comput Chem*, **2006**, *27*, 1307.
9. McNamara, J. P.; Sundararajan, M., Hillier, I. H. *J Mol Graph Model*, **2005**, *24*, 128.
10. Csontos, J.; Palermo, N. Y.; Murphy, R. F., Lovas, S. *J Comput Chem*, **2008**, *29*, 1344.
11. Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P., Sprik, M. *J Phys Chem B*, **2007**, *111*, 3969.
12. Jensen, K. P. *J Inorg Biochem*, **2006**, *100*, 1436.
13. Sundararajan, M.; Hillier, I. H., Burton, N. A. *J Phys Chem A*, **2006**, *110*, 785.
14. Machonkin, T. E.; Westler, W. M., Markley, J. L. *Inorg Chem*, **2005**, *44*, 779.
15. Vrajmasu, V. V.; Munck, E., Bominaar, E. L. *Inorg Chem*, **2004**, *43*, 4867.
16. Kennepohl, P., Solomon, E. I. *Inorg Chem*, **2003**, *42*, 696.
17. Vrajmasu, V. V.; Bominaar, E. L.; Meyer, J., Munck, E. *Inorg Chem*, **2002**, *41*, 6358.
18. Glaser, T.; Rose, K.; Shadle, S. E.; Hedman, B.; Hodgson, K. O., Solomon, E. I. *J Am Chem Soc*, **2001**, *123*, 442.
19. Sigfridsson, E.; Olsson, M. H., Ryde, U. *Inorg Chem*, **2001**, *40*, 2509.

20. Kloiber, K.; Weiskirchen, R.; Krautler, B.; Bister, K., Konrat, R. *J Mol Biol*, **1999**, *292*, 893.
21. Nayal, M., Di Cera, E. *J Mol Biol*, **1996**, *256*, 228.
22. Spartan 5.0, Wavefunction, Inc., 18401 Von Karman Avenue Suite 370, Irvine, CA 92612 U.S.A.
23. Gaussian 98 (Revision A.1), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian 98 (Revision A.1), Gaussian, Inc., Pittsburgh PA, 1998.
24. Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.