Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters 16 (2006) 587–591
Bioorganic & Medicinal Chemistry Letters

The electronic structure of human erythropoietin as an aid in the design of oxidation-resistant therapeutic proteins

Fabio Pichierri*

COE Laboratory, Tohoku University, IMRAM, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

Received 26 July 2005; revised 26 September 2005; accepted 15 October 2005
Available online 10 November 2005

Abstract—The electronic structure of human erythropoietin (*Hu*EPO) has been investigated with the aid of quantum mechanical calculations. The results indicate that the protein is highly polarized and its permanent dipole moment has a magnitude of 471 D. The HOMO of *Hu*EPO is localized on Trp51, which stays in close proximity to Met54. Three oxidation-resistant mutants of *Hu*EPO (W51F, M54V, and W51F-M54V) have been modeled and their electronic structures are compared to that of the native protein. Among them, the W51F mutation is predicted to be the most effective in increasing the oxidation potential of the protein.

Erythropoietin (EPO) is a 166-residue glycoprotein hormone produced by the kidneys. Upon interacting with its target receptor (EPOR) located on the cell's surface, EPO triggers a signaling cascade which involves the phosphorylation of protein kinases that are integrated in the RAS, JAK-STAT, and PI3 pathways.[1] The principal biological role of the EPO hormone is that of controlling the delivery of oxygen to tissues through erythrocytes. As a result, recombinant human EPO (*rHu*EPO) has been successfully employed in the treatment of cancer-related anemia.[2] Furthermore, recent studies indicate that EPO might act as a neurotransmitter in the brain and, hence, it could be used in the treatment of diseases affecting the central nervous system (CNS).[3]

The effective use of *rHu*EPO as a therapeutic agent requires that the protein be resistant against chemical oxidation. Protein oxidation is a damaging process that occurs through the addition of oxygen atoms to methionine residues to yield the corresponding sulfone (SO) and sulfoxide (SO_2) derivatives.[4] Oxidation-resistant mutants of proteins have been obtained through site-directed mutagenesis by replacing methionine (Met, M) with a hydrophobic residue.[5] Other amino acid residues, however, might be involved in the mechanism that is responsible for the oxidative damage of the protein particularly those where the frontier orbitals are localized. This prompted us to investigate the electronic structure of *Hu*EPO with the aid of linear-scaling semi-empirical quantum mechanical methods.[6]

This study employs the solution structure of the mutant MKLysEPO (N24K, N38K, and N83K) determined by Cheetham et al.[7] with the aid of NMR spectroscopy (1BUY). The protein is de-glycosylated and adopts the left-handed four-helix bundle topology shown in Figure 1. The four α helices, denoted to as A, B, C, and D, run in an up-up–down-down direction with two long and one short cross-over loops. Two disulfide bonds are formed, one between Cys7 and Cys161, and one between Cys29 and Cys33, the former connecting helix A to helix D.

Cheetham et al.[7] made three-point mutations and introduced one lysine residue on each glycosylation site at Asn24, Asn38, and Asn83. Furthermore, they also added methionine (M) and lysine (K) to the N-terminus of *rHu*EPO to improve expression yields in *Eco*RI.

To construct our model of *Hu*EPO, we have restored the Asn residues by replacing the corresponding Lys residues. Note that both M and K are not present in the MKLysEPO mutant thus making their deletion unnecessary. Also, His32 and His94 are in the neutral state (the ionized sites are those in the experimental NMR structure). We subjected our model to 1000 cycles of energy minimization by employing the Amber99 force field[8] and the Polak–Ribiere conjugate-gradient method

Keywords: Erythropoietin; Protein-based therapeutics; Anemia; Computational chemistry.
* Tel./fax: +81 22 217 5110; e-mail: fabio@tagen.tohoku.ac.jp

0960-894X/$ - see front matter
doi:10.1016/j.bmcl.2005.10.041

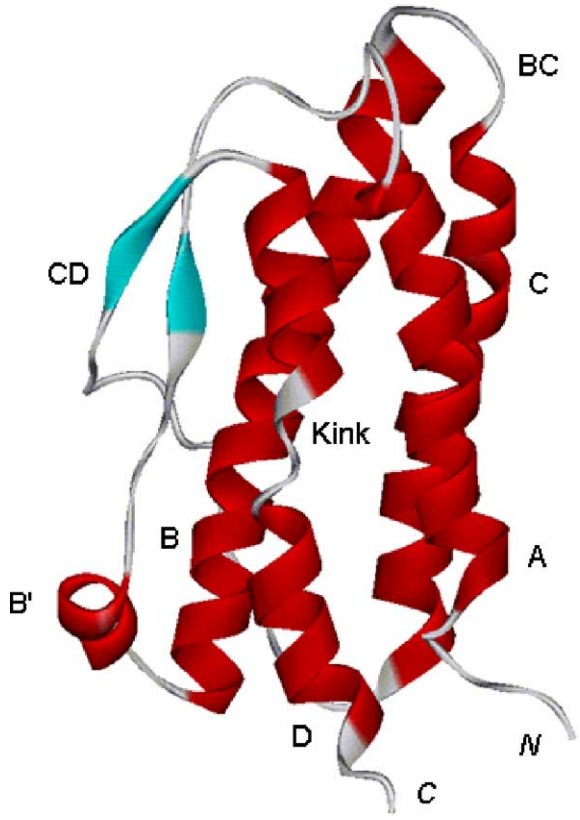

Figure 1. The left-handed four-helix bundle topology of MKLysEPO determined by NMR spectroscopy (PDB code: 1BUY).

as implemented in the HyperChem software package.[9] A distance-dependent dielectric constant was introduced to mimic the effect of the solvent. This procedure ensures that the structure of our atomic model is being relaxed while keeping its conformation as close as possible to that experimentally determined by Cheetham et al.[7]

In the next step, we subjected our *Hu*EPO model to a self-consistent field (SCF) calculation by employing the localized molecular orbital (LMO) approach of Stewart[10] in conjunction with the conductor-like screening model (COSMO) of Klamt and Schüürmann.[11] COSMO is a continuum solvation model in which the effect of the solvent is mimicked by introducing a ε-dependent correction factor into the molecular Hamiltonian. The value $\varepsilon = 78.4$ was used here to simulate (implicitly) the effect of water.

Three semiempirical Hamiltonians based on the NDDO approximation[6] have been employed here, namely Dewar's AM1 method[12] and Stewart's PM3 and PM5 methods,[13] as implemented in the MOPAC2002 software package.[14] A number of computational studies performed by this author on different proteins[15] indicate that the most accurate results are those obtained by combining the LMO-SCF-COSMO approach with the PM5 Hamiltonian.

LMO-SCF-COSMO calculations indicate that the Lewis structure of *Hu*EPO is made of 2635 sigma bonds, 716 lone pairs, 220 normal π-bonds, 17 aromatic ring π-bonds, 22 cations, and 19 anions. The net charge of the native protein is 3+ and there exists 3607 filled levels. Table 1 reports the residue sequence of *Hu*EPO.

Table 2 reports the results of our quantum mechanical calculations utilizing three different NDDO methods. The heat of formation (HOF) computed with all three methods is negative. The energy gap between the highest-occupied (HOMO) and lowest-unoccupied (LUMO)

Table 2. Results of LMO-SCF-COSMO calculations on *Hu*EPO using three different NDDO-based methods

Method[a]	HOF (kcal/mol)	H–L (eV)	Dipole (D)
AM1	−9513	5.982	474
PM3	−10062	5.533	471
PM5	−11987	5.534	471

[a] The dielectric constant of the solvent is set at $\varepsilon = 78.4$.

Table 1. Residue sequence in *Hu*EPO[a]

	1	2	3	4	5	6	7	8	9	10
	ALA+	PRO	PRO	ARG+	LEU	ILE	CYS	ASP−	SER	ARG+
10	VAL	LEU	GLU−	ARG+	TYR	LEU	LEU	GLU−	ALA	LYS+
20	GLU−	ALA	GLU−	ASN	ILE	THR	THR	GLY	CYS	ALA
30	GLU−	HIS	CYS	SER	LEU	ASN	GLU−	ASN	ILE	THR
40	VAL	PRO	ASP−	THR	LYS+	VAL	ASN	PHE	TYR	ALA
50	TRP	LYS+	ARG+	*MET*	GLU−	VAL	GLY	GLN	GLN	ALA
60	VAL	GLU−	VAL	TRP	GLN	GLY	LEU	ALA	LEU	LEU
70	SER	GLU−	ALA	VAL	LEU	ARG+	GLY	GLN	ALA	LEU
80	LEU	VAL	ASN	SER	SER	GLN	PRO	TRP	GLU−	PRO
90	LEU	GLN	LEU	HIS	VAL	ASP−	LYS+	ALA	VAL	SER
100	GLY	LEU	ARG+	SER	LEU	THR	THR	LEU	LEU	ARG+
110	ALA	LEU	GLY	ALA	GLN	LYS+	GLU−	ALA	ILE	SER
120	PRO	PRO	ASP−	ALA	ALA	SER	ALA	ALA	PRO	LEU
130	ARG+	THR	ILE	THR	ALA	ASP−	THR	PHE	ARG+	LYS+
140	LEU	PHE	ARG+	VAL	TYR	SER	ASN	PHE	LEU	ARG+
150	GLY	LYS+	LEU	LYS+	LEU	TYR	THR	GLY	GLU−	ALA
160	CYS	ARG+	THR	GLY	ASP−	ARG				

[a] The underlined residues are those where HOMO (—) and LUMO (- - -) are localized. The Met54 residue is shown in italic.

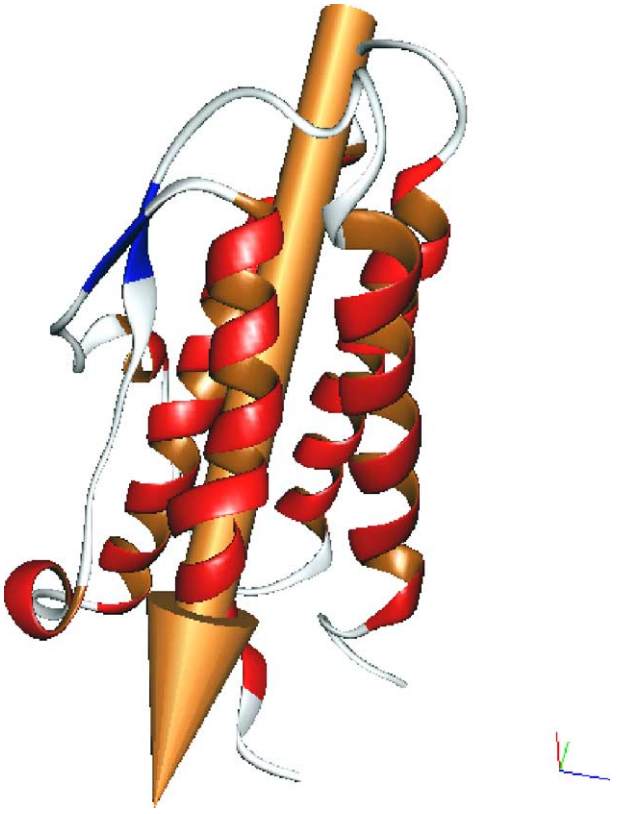

Figure 2. The macrodipole of our *Hu*EPO model as computed with the PM5-COSMO method.

molecular orbitals, H–L, corresponds to 5.534 eV as computed with the PM5 method.[16] The PM3 value is almost of the same magnitude as the PM5 one, while AM1 gives a slightly larger H–L gap of 5.982 eV.

The dipole moment computed with both PM5 and PM3 methods has a magnitude of 471 debye, while AM1 gives a dipole of 474 debye.[17] The computed dipole is comparable in magnitude to that of 492 debye (PM5 value) recently calculated by us for α-chymotrypsin.[15a] Figure 2 shows the dipole moment vector of *Hu*EPO. As it appears from this figure, the dipole vector is aligned almost parallelly with respect to the axis of the four-helix bundle.

The positive side of the dipole vector, which is represented by the arrow's tip, is located on the N- and C-terminals of the protein, while the negative side of the dipole is located on the side of the BC loop. This result indicates that the electronic charge of *Hu*EPO is highly polarized along the direction of the four-helix bundle axis. We predict that the observed polarization of charge might play an important role during the interaction of *Hu*EPO with its target receptor.

Given the up-up–down-down topology of the helices of *Hu*EPO (Fig. 1), we can expect that the vector sum of the dipoles associated to their peptide linkages[18] should be quite small in magnitude. Hence, it is highly likely that the large magnitude of the computed macrodipole arises from the spatial arrangement of the charged residues in the protein. In this regard, we have computed the net charge of each residue in the protein and plotted the results in Figure 3.

As it can be seen from Figure 3 and from Table 1, there exist 39 charged amino acid residues together with the charged carboxyl- and aminotermini of the protein. All in all *Hu*EPO possesses 21 positively charged and 18 negatively charged amino acids. An interesting result of this analysis is that helix D, which is made of 24 residues, is almost entirely positively charged. This suggests the possibility that the small kink that appears in the middle of the helix D (see Fig. 1) could arise from electrostatic effects. Mutants of *rHu*EPO containing one or more negatively charged residues placed on helix D should stabilize it thus providing structurally stable therapeutic agents.

The asterisks in Figure 3 indicate the position of the N-glycosylation sites on the primary structure of the protein.[19] It appears that the three N-glycosylation sites differ from one another. The first, on Asn24, is surrounded by both positive and negative residues. The second, on

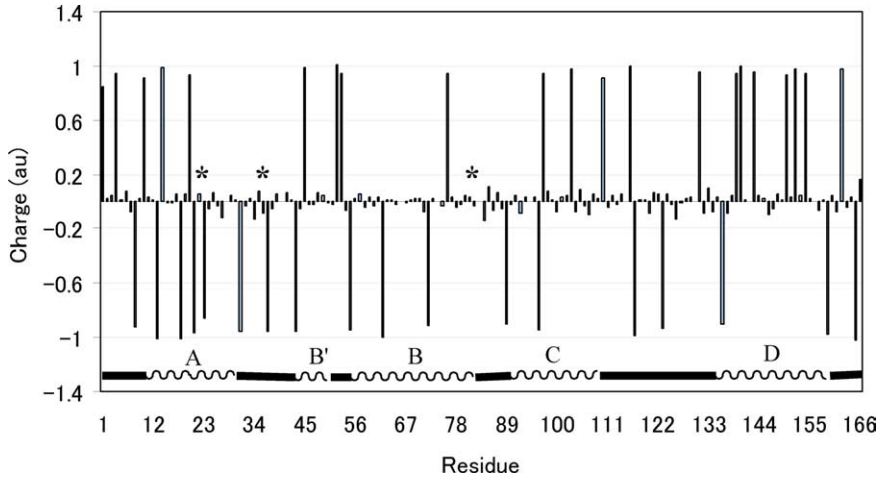

Figure 3. Net charge on the residues of *Hu*EPO as computed with the PM5-COSMO method. The secondary structure of the protein is indicated at the bottom. The asterisks indicate the location of the N-glycosylation sites on Asn24, Asn38, and Asn83.

Asn38, is surrounded by negatively charged residues while the third on Asn83, is located nearby positively charged residues. These differences in charge distribution might affect the properties of the sugar chains that are attached to the protein as a result of post-translational modifications.

Next we analyzed the frontier orbitals of *Hu*EPO. Molecular orbitals (MOs) were obtained from a procedure known as orbital canonicalization[6] which transforms LMOs into MOs. Eight occupied and eight unoccupied MOs spanning the H–L gap were generated and analyzed. Figure 4 displays the HOMO−1, HOMO, LUMO, and LUMO+1 of *Hu*EPO. HOMO−1 (Fig. 4a) is localized on the Trp64 residue, while HOMO (Fig. 4b) is found on the Trp51 residue which belongs to the short B′ helix (see Fig. 1). Interestingly, we notice that Trp51 is in close proximity to Met54 (the distances between CH_3 of Met54 and the carbon atoms of the six-membered ring of Trp51 are in the 3.04–3.90 Å range). This observation suggests that the oxidation of Met54 does require the presence of Trp51 which, owing to the localization of HOMO on it, might act as a reservoir of electrons. Further experimental or theoretical evidence will be needed in order to validate this suggestion.

As far as the virtual (i.e., unoccupied) orbitals of *Hu*EPO are concerned, LUMO (Fig. 4c) is localized on the disulfide bond connecting Cys7 to Cys161 residues which connects helix A to helix D. The next empty level, LUMO+1, is also localized on a disulfide bond, namely that formed between Cys29 and Cys33.

The HOMO−*n* and LUMO+*n* levels, with *n* = 2–8, were also inspected (MO drawings not shown). The residues on which the HOMO−*n* levels are localized are the following: Trp88 (HOMO−2), Trp51 (HOMO−3), Tyr45 (HOMO−4), Trp64 (HOMO−5), Trp88 (HOMO−6), Tyr49 (HOMO−7), and Phe138 (HOMO−8). As far as the LUMO+*n* levels are concerned, they are localized on: the peptide bond connecting Glu31 to His32 (LUMO+2), the sulfur bond connecting Cys7 to Cys161 (LUMO+3), and on the atoms of the protein backbone spanning the residues Leu35 to Asn38 (LUMO+4), Asn83 to Ser85 (LUMO+5), Cys7 to Asp8 (LUMO+6), Leu102 to Ser104 (LUMO+7), and Thr134 to Asp136 (LUMO+8).

With the above information at hand, we have designed two single-point mutants, one in which Met54 is replaced by valine (M54V) and the other in which Trp51 is replaced by phenylalanine (W51F). A third mutant having both Met54 and Trp51 replaced with Val54 and Phe51, respectively, has also been modeled. The energies of the frontier orbitals and the corresponding H–L gap for native *Hu*EPO and the three mutants are reported in Table 3.

It appears that the W51F mutation causes the largest increase in the H–L gap, from 5.534 to 5.843 eV. For all the three mutants, the increase in H–L gap with respect to that of the native protein can be ascribed to the decrease in their HOMO energies, whereas the energies of the LUMO increase only slightly. The linear relationship between HOMO energy and the oxidation potential in molecules[20] suggests that these mutations would increase the oxidation potential of *Hu*EPO in water.

In the present study, we assumed that Trp51 and Met54 are involved in the oxidation of *Hu*EPO. However, other residues such as Tyr49 and Phe48 equally could be oxidized. The availability of experimental EPO-oxidation data will shed further light on this important issue. Furthermore, the conformational properties of *Hu*EPO might be important in modulating the accessibility of O_2 to the residues that are prone to oxidation. One possible way to assess this is by performing long-range molecular dynamics (MD) simulations in a water environment. Finally, it is worth noticing that the three *N*-glycan chains on natural *Hu*EPO are essential for its

Table 3. Frontier orbital energies and H–L gap of native *Hu*EPO and three of its mutants as computed with the PM5-COSMO method[a]

Protein	HOMO (eV)	LUMO (eV)	H–L (eV)
Native	−8.504	−2.970	5.534
W51F	−8.736	−2.893	5.843
M54V	−8.659	−2.884	5.775
W51F, M54V	−8.637	−2.869	5.768

[a] The dielectric constant of the solvent is set at $\varepsilon = 78.4$.

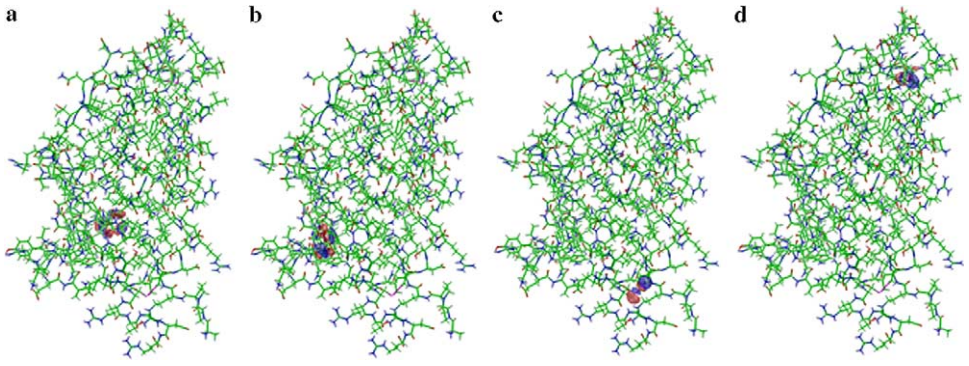

Figure 4. The frontier orbitals (a) HOMO−1 (on Trp64); (b) HOMO (on Trp51); (c) LUMO (on Cys7-Cys161); (d) LUMO+1 (on Cys29-Cys33) of *Hu*EPO as computed with the PM5-COSMO method.

erythropoietic function in vivo.[19,21] The possible effect exerted by these sugar chains on both the electronic structure and conformation of *Hu*EPO could be investigated by combining classical MD simulations with linear-scaling semiempirical quantum mechanical calculations.[22] As far as *recombinant Hu*EPO is concerned, recent experimental studies indicate that its biological activity strongly depends upon the nature of its glycosylation state which, in turn, is highly dependent upon the production process.[21] Modeling studies on such glycosylated proteins do require that the chemical composition of their *N*-glycan chains be fully characterized with the aid of biophysical methods.

In conclusion, we have carried out the first quantum mechanical study of the electronic structure of *Hu*EPO. The results of our LMO-SCF-COSMO calculations indicate that the protein is highly polarized along the direction of the four-helix bundle axis and the magnitude of the permanent macrodipole amounts to 471 debye. The analysis of the frontier orbitals indicates that HOMO is localized on Trp51 which interacts with Met54, the only methionine residue in the protein that is prone to oxidation. From the information obtained from the electronic structure calculations we have designed three oxidation-resistant mutants of *Hu*EPO, two single-point mutants (M54V and W51F) and one double-point mutant. It appears that W51F mutation is more effective than M54V mutation in increasing the oxidation potential of *Hu*EPO thus representing a viable alternative to the commonly exploited replacement of methionine. These results will pave the way toward the development of novel cytokine-based therapeutics that are resistant against oxidation.

Acknowledgments

I thank Fujitsu Ltd. (Tokyo) for providing the MOPAC2002 software package and the XMO graphical user interface. Two anonymous referees are gratefully acknowledged for useful comments and constructive criticisms. Financial support for this work was provided by the 21st century Center of Excellence (COE) project 'Giant Molecules and Complex Systems' of MEXT activated at Tohoku University.

References and notes

1. Sytkowski, A. J., Ed.; *Erythropoietin: Blood, Brain, and Beyond*; John Wiley: New York, 2004.
2. (a) Rosso, R.; Del Mastro, L.; Venturini, M.; Bergaglio, M.; Pasquetti, W.; Garrone, O. *Tumori* **1997**, *83*, S26; (b) Eckardt, K. U. *Curr. Opin. Investig. Drugs* **2001**, *2*, 1081; (c) Oton Sanchez, C.; Penate Gonzalez, G.; Oton Sanchez, L. F. *Med. Clin. (Barc.)* **2005**, *124*, 186.
3. Buemi, M.; Caccamo, C.; Nostro, L.; Cavallaro, E.; Floccari, F.; Grasso, G. *Med. Res. Rev.* **2005**, *25*, 245.
4. Rosenberg, S.; Barr, P. J.; Najarian, R. C.; Hallewell, R. A. *Nature* **1984**, *312*, 77.
5. (a) Strandberg, L.; Lawrence, D. A.; Johansson, L. B.; Ny, T. *J. Biol. Chem.* **1991**, *266*, 13852; (b) Yang, Y.; Jiang, L.; Zhu, L.; Wu, Y.; Yang, S. *J. Biotechnol.* **2000**, *81*, 113; (c) Gustavsson, N.; Kokke, B. P.; Anzelius, B.; Boelens, W. C.; Sundby, C. *Protein Sci.* **2001**, *10*, 1785.
6. Cramer, C. J., Ed.; *Essentials of Computational Chemistry*; John Wiley: Chichester, 2002.
7. Cheetham, J. C.; Smith, D. M.; Aoki, K. H.; Stevenson, J. L.; Hoeffel, T. J.; Syed, R. S.; Egrie, J.; Harvey, T. S. *Nat. Struct. Biol.* **1998**, *5*, 861.
8. Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
9. HYPERCHEM (version 7.51), Hypercube, Gainesville, US.
10. Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133.
11. Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
12. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
13. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
14. Stewart, J. J. P., MOPAC2002, Fujitsu Ltd. (Tokyo).
15. (a) Pichierri, F. *J. Mol. Struct. (THEOCHEM)* **2003**, *664–665*, 197; (b) Pichierri, F. *Biophys. Chem.* **2004**, *109*, 295; (c) Pichierri, F. *Chem. Phys. Lett.* **2005**, *410*, 462.
16. Note that the H–L energy gap computed in the gas phase corresponds to only 0.59 eV (PM5). It appears that COSMO significantly increases the energy of LUMO thus widening the H–L gap with respect to that computed in the gas phase.
17. The dipole moment is that of an ion and is related to the center of mass, see: Buckingham, A. D. *Chem. Soc. (Great Britain) Quart. Rev.* **1959**, *13*, 183, We chose to draw the dipole vector by placing the middle point of the arrow on the center of mass of the protein, see:; Pichierri, F. *J. Mol. Struct. (THEOCHEM)* **2003**, *664–665*, 197.
18. The dipole moment associated to the peptide bond corresponds to about 3.46 debye, see: Hol, W. G. J. *Prog. Biophys. Mol. Biol.* **1985**, *45*, 149.
19. Takeuchi, M.; Kobata, A. *Glycobiology* **1991**, *1*, 337.
20. (a) Xue, Z.-M.; Chen, C.-H. *Electrochim. Acta* **2004**, *49*, 5167; (b) D'Andrade, B. W. et al. *Org. Electron.* **2005**, *6*, 11.
21. Gerngross, T. U. *Nat. Biotechnol.* **2004**, *22*, 1409.
22. For computational studies addressing the effect of sugar chains on the conformation of proteins and peptides see, for example: (a) Zuegg, J.; Gready, J. E. *Glycobiology* **2000**, *10*, 959; (b) Bosques, C. J. et al. *J. Am. Chem. Soc.* **2004**, *126*, 8421.