A test of ligand field molecular mechanics as an efficient alternative to QM/MM for modelling metalloproteins: the structures of oxidised type I copper centres†

Robert James Deeth

Received (in Cambridge, UK) 23rd March 2006, Accepted 25th April 2006
First published as an Advance Article on the web 16th May 2006
DOI: 10.1039/b604290b

Ligand Field Molecular Mechanics based on homoleptic model systems delivers accurate, unbiased geometries of complete mononuclear blue copper proteins about four orders of magnitude faster than comparable QM/MM calculations.

Computer modelling of metalloproteins presents significant challenges to the computational chemist.[1] The sheer size and conformational complexity of proteins preclude a complete treatment *via* quantum mechanics (QM) and we rely instead on classical techniques such as molecular mechanics (MM) and molecular dynamics (MD). Conventional MM/MD is successful at modelling systems comprising the lighter elements but has problems with heavier elements, notably transition metals (TMs), especially if the metal centre is electronically 'difficult'.[2]

The oxidised Type I 'blue' copper centre is arguably the hardest site for conventional MM. The d^9 configuration causes large electronic effects which have a significant impact on structure. These problems are exacerbated for Type I centres in that the geometry is highly distorted and comprises a ligand set with few precedents in small model compounds. However, it is also one of the most studied metalloprotein active sites.[3,4] Type I copper proteins thus form an excellent test bed for the new methodologies.

A typical Type I centre has the copper strongly ligated by two histidine nitrogens and a cysteinate (CYS) sulfur plus a weakly bound fourth ligand, often a methionine (MET) (Fig. 1). The three strongly bound groups form a more or less trigonal plane with the electronic structure dominated by the Cu–S_{CYS} interaction.[5,6] The

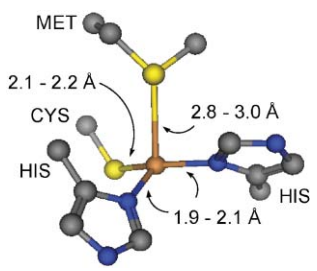

Fig. 1 Active site detail for typical Type I blue copper centre showing approximate structure and bond length ranges observed experimentally.

Department of Chemistry, University of Warwick, Coventry, UK CV4 7AL. E-mail: r.j.deeth@warwick.ac.uk; Fax: 024 7652 4112; Tel: 024 7652 3187

† Electronic supplementary information (ESI) available: Computational details, AMBER94 and LFMM parameters plus coordinates of all the molecules discussed herein. See DOI: 10.1039/b604290b

site continues to attract experimental and theoretical interest[7] with one of the major themes being the so-called 'entatic state'.

The entatic state[8] (or induced rack[9]) model asserts that the protein enforces a geometry commensurate with the role of the active site. Since Type I copper centres are involved in electron transport, this entails the protein binding the metal in a configuration, and with a donor set, designed to minimise the reorganisation energy associated with the change in metal oxidation state. Since oxidised Cu(II) prefers nitrogen donors in a planar arrangement while reduced Cu(I) prefers a tetrahedron of sulfur ligands, the Type I site can be viewed as some intermediate compromise between the demands of each state.

Ryde *et al.*[10] used quantum chemical calculations based on density functional theory (DFT) to estimate the entatic strain energy. The metal–ligand interactions were captured by using imidazole (im) as a model for histidine, methyl thiolate (SMe) as a model for deprotonated CYS and dimethyl sulfide (DMS) as a model for MET. To his, and many others', surprise, the structure of the fully relaxed Type I model $[Cu(im)_2(SMe)(DMS)]^{n+}$ (n = 1 or 0) was not significantly different to the 'in protein' model, especially for the oxidised form, leading to the conclusion that the metal centre in plastocyanin is not strained by the protein.

This highly controversial result has spawned a vigorous debate which hinges on the role of the protein.[4] Hence, computational studies have tended to try and model the entire system. Given the 'plasticity' of Cu(II) and its tendency to display dramatic electronic effects (*e.g.* Jahn–Teller distortions), one needs a flexible, general treatment around the metal centre and most workers have opted to use quantum mechanics. Since QM is too expensive for the whole protein, this inevitably leads to hybrid QM/MM calculations.[7,11,12] One exception is an attempt by Comba and Remenyi[13] to develop an empirical MM force field for Type I centres but their approach was highly restrictive and specific. The great attraction of QM/MM is its lack of bias in that it will simultaneously give the right results for a Type I copper centre and a simpler system like $[Cu(im)_4]^{2+}$. In contrast, the Comba–Remenyi force field would fail for $[Cu(im)_4]^{2+}$ and only works for a restricted subset of Type I centres which conform to the choice of training set.

QM/MM appears to work well for metal centres.[14] However, any method which relies on quantum mechanics, even if only for a small part of the whole system, is, compared to a fully classical approach, relatively slow. Consequently, the number and type of calculations that can be achieved in a reasonable time is limited.

Instead, we require the flexibility and generality of quantum mechanics combined with the speed of molecular mechanics. This

communication demonstrates how these goals are met using the ligand field molecular mechanics (LFMM) model.

The LFMM approach[15] extends conventional MM by explicitly incorporating the ligand field stabilisation energy (LFSE). The observed structure is then a compromise between ligand–ligand repulsions and the d electronic stabilisation.

We first introduced the LFMM method over a decade ago[16] and the concept has since been adopted by others.[17–19] More recently, the LFMM model has been applied to a range of Cu(II) complexes[20,21] where it has been demonstrated that DFT-like accuracy can be achieved, even for subtle problems such as the Jahn–Teller effect, around 30 000 times faster. However, these studies were restricted to relatively small molecules with nitrogen donors only. Extension of the LFMM to copper metalloproteins represents a significant step forward.

A critical design feature of the LFMM approach is generality. It is assumed that the Cu–N bond to, say, an imidazole group is essentially independent of whether the imidazole is isolated or happens to be part of a histidine connected to a protein backbone. This behaviour is akin to what is expected from QM but contrasts with the approach used by Comba and Remenyi.

The first step in the LFMM parameterisation is to develop parameters for simple, model homoleptic complexes using DommiMOE,[22] our LFMM-extended version of the Molecular Operating Environment.[23] The MOE implementation of AMBER94 provides the 'organic' part of the force field (FF).

$Cu–N_{im}$ parameters are already available for the Merck Molecular Force Field (MMFF94).[20] Even though the functional forms used in MMFF94 and AMBER94 are different, the same LFMM treatment is embedded in all of MOE's force fields. As a test, the X-ray crystal structures of $[Cu(im)_4](N\text{-tosylvalinate})_2$ (CSD Refcode GADGOH) and $[Cu(1,3,5Me_3\text{-}im)_4](ClO_4)_2$ (CSD Refcode BUXDUT) yield Cu–N distances of 1.99–2.00 Å *vs.* the (AMBER94) LFMM value of 1.99 Å. This is the first application of the LFMM to second-row donors like sulfur. Parameterisation is based exclusively on DFT data since $[Cu(SMe)_4]^{2-}$ and $[Cu(DMS)_4]^{2+}$ are unknown experimentally. It is well established that DFT provides an excellent structural tool for modelling coordination complexes.[24]

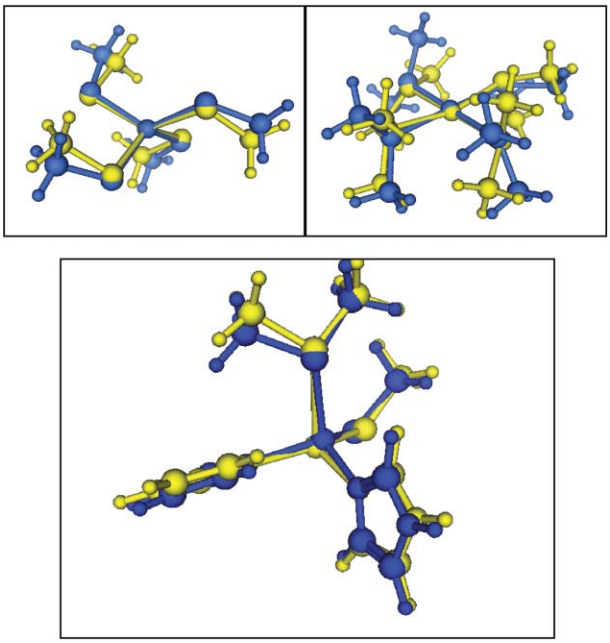

Fig. 2 Overlay of LFMM (yellow) and DFT (blue) structures: top left, $[Cu(SMe)_4]^{2-}$, top right, $[Cu(SMe_2)_4]^{2+}$ and bottom, the Type I active site model $[Cu(im)_2(SMe)(SMe_2)]^{+}$.

Fig. 2 displays overlays of LFMM and reference structures for the two homoleptic sulfur-ligated models plus the Type I active site model $[Cu(im)_2(SMe)(SMe_2)_2]^{+}$. The Cu–S distances of the former two species agree to better than 0.02 Å. For the Type I model complex, the LFMM delivers three key results. Firstly, the Cu–S_{MET} distance is 3.0 Å which is 0.64 Å longer than the Cu–S distance of 2.36 Å in the reference homoleptic complex. Conventional MM has to assign the Cu–S_{MET} interaction a long 'reference' bond length in order to force a result which is automatic in the LFMM (and QM—the DFT protocol used here gives 2.59 Å). Secondly, the Cu–S_{CYS} distance in the Type I model decreases significantly (~0.15 Å) compared to the homoleptic model. Thirdly, the LFMM gives a somewhat more T-shaped

Table 1 Comparison of selected bond lengths (Å) and bond angles (°) for five proteins containing an oxidised Type I copper centre

	Distance to Cu/Å				L–Cu–L′ angle/°					
PDB code	S_{CYS}	N_b[a]	N_l[a]	S_{MET}	N_b/N_l	S_{CYS}/N_b	S_{CYS}/N_l	S_{CYS}/S_{MET}	S_{MET}/N_b	S_{MET}/N_l
1AAC: X-ray	2.11	1.95	2.03	2.90	104	136	112	111	84	100
1AAC: MM[b]	2.23	2.08	2.06	2.80	101	133	113	117	84	102
1AAC: LFMM	2.14	2.03	2.01	3.00	97	152	106	108	82	107
1BQK: X-ray	2.13	1.95	1.92	2.71	100	135	114	107	87	107
1BQK: MM[b]	2.23	2.07	2.06	2.76	102	133	113	108	84	113
1BKQ: LFMM	2.18	2.04	2.03	2.53	99	140	104	100	87	133
1RCY: X-ray	2.26	2.04	1.89	2.88	105	127	119	106	85	106
1RCY: MM[b]	2.22	2.07	2.06	2.78	102	133	104	114	83	124
1RCY: LFMM	2.17	2.04	2.07	2.52	98	133	111	105	88	105
1KDJ: X-ray	2.26	1.93	2.07	2.92	107	126	118	107	81	110
1KDJ: MM[b]	2.21	2.07	2.07	2.72	112	119	120	110	80	107
1KDJ: LFMM	2.18	2.04	2.05	2.51	104	136	106	98	83	135
6PAZ: X-ray	2.13	2.01	2.07	2.90	98	135	118	108	85	106
6PAZ: MM[b]	2.22	2.05	2.07	2.75	104	122	122	108	82	111
6PAZ: LFMM	2.16	2.07	2.05	2.57	98	137	107	104	88	126

[a] N_b = Backbone histidine; N_l = loop histidine. [b] FF results reported by Comba and Remenyi. 1AAC: amicyanin from *Paracoccus denitrificans*; 1BQK: pseudoazurin from *Achromobacter cycloclastes*; 1RCY: rusticyanin from *Thiobacillus ferrooxidans*; 1KDJ: plastocyanin from *Dryopteris crassirhizoma*; 6PAZ: pseudoazurin from *Alcaligenes. faecalis*.

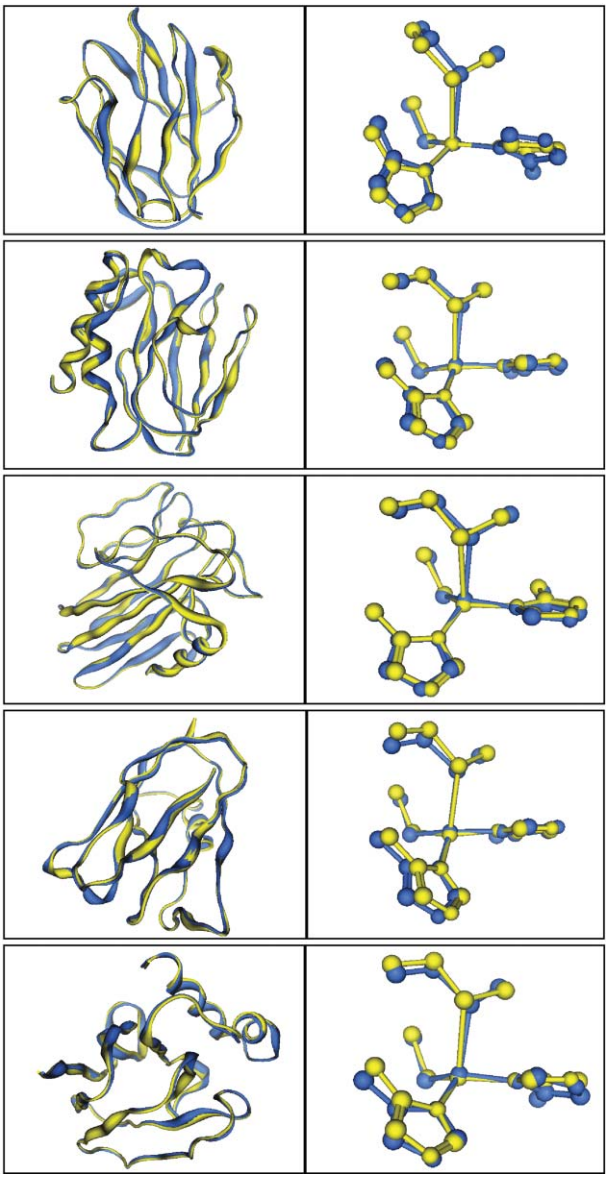

Fig. 3 Overlay of X-ray (yellow) and LFMM (blue) structures. Figures on the left show the superposition of the protein backbones while those on the right show the optimal overlay of the trigonal [CuNNS] set. The proteins are given in the same order as Table 1

geometry for the three strongly bound ligand (the larger N–Cu–S_{CYS} angle is 153° by LFMM *vs.* 125° by DFT) but continues to reproduce the *trans* influence of the S_{CYS} donor which leads to a pronounced lengthening of the pseudo-*trans* Cu-N distance (2.07 Å LFMM, 2.10 Å DFT) compared to the pseudo-*cis* Cu-N bond (1.99 Å LFMM, 2.03 Å DFT).

The LFMM performance is equally impressive for the five complete proteins studied by Comba and Remenyi (Table 1). LFMM-optimised structures of the complete proteins which include a surrounding sheath of water molecules at least 5 Å thick, reproduce both the protein structure and the details of the active site geometry. Rms overlays of observed (yellow) and LFMM (blue) structures for both the protein backbones and the active site detail are shown in Fig. 3. The full protein backbone atom overlays have rms deviations less than 0.42 Å and, given the inherently larger errors in protein crystallography, the calculated active site geometries are good. The Cu–S_{MET} contacts appear to vary substantially but Ryde has shown[10] that the energetic consequences of these differences is of minor importance. Importantly, the LFMM does not oblige the methione to be at some predetermined distance unlike conventional MM which forces all the bond lengths in a narrow range. Since the methionine is not strongly bound to the metal, other factors will influence its position and may also explain why Cu–N–S_{MET} angles involving the loop histidine occasionally differ by 20–25° from the X-ray values.

In conclusion, the LFMM has 'quantum-like' behaviour and delivers accurate structures of oxidised blue copper proteins at least 6000 times faster than QM/MM. Furthermore, the LF part of the calculation probes the local M–L interactions and can be used to compute other properties such as EPR parameters. The latter are important for distinguishing sites which give rhombic *g*-value patterns from those which give axial spectra. A fully-detailed report of LFMM studies for these and other blue copper proteins, including LFMM estimates of protein strain energies and 'virtual mutagenesis' calculations comparing the LFMM performance with the latest QM/MM results on rusticyanin,[7] is in preparation.

Notes and references

1 A. Ghosh, *Curr. Opin. Chem. Biol.*, 2003, **7**, 110–112.
2 P. Comba and R. Remenyi, *Coord. Chem. Rev.*, 2003, **238**, 9–20.
3 E. I. Solomon, R. K. Szilagyi, S. D. George and L. Basumallick, *Chem. Rev.*, 2004, **104**, 419–458.
4 H. B. Gray, B. G. Malmstrom and R. J. P. Williams, *J. Biol. Inorg. Chem.*, 2000, **5**, 551–559.
5 U. Ryde, M. H. M. Olsson, B. O. Roos and A. C. Borin, *Theor. Chem. Acc.*, 2001, **105**, 452–462.
6 E. I. Solomon, M. D. Lowery, D. E. Root and B. L. Hemming in, *Advances in Chemistry*, ed. H. H. Thorpe and V. L. Pecoraro, OUP, Oxford, 1995, vol. **246**, pp. 121–164.
7 M. L. Barrett, I. Harvey, M. Sundararajan, R. Surendran, J. F. Hall, M. J. Ellis, M. A. Hough, R. W. Strange, I. H. Hillier and S. S. Hasnain, *Biochemistry*, 2006, **45**, 2927–2939.
8 B. L. Vallee and R. J. P. Williams, *Proc. Natl. Acad. Sci. USA*, 1968, **59**, 498–505.
9 B. G. Malmstrom, *Eur. J. Biochem.*, 1994, **223**, 711–718.
10 U. Ryde, M. H. M. Olsson, K. Pierloot and B. O. Roos, *J. Mol. Biol.*, 1996, **261**, 586–596.
11 A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 425–443.
12 P. Comba, A. Lledos, F. Maseras and R. Remenyi, *Inorg. Chim. Acta*, 2001, **324**, 21–26.
13 P. Comba and R. Remenyi, *J. Comput. Chem.*, 2002, **23**, 697–705.
14 U. Ryde, *Curr. Opin. Chem. Biol.*, 2003, **7**, 136–142.
15 R. J. Deeth, *Coord. Chem. Rev.*, 2001, **212**, 11–34.
16 V. J. Burton, R. J. Deeth, C. M. Kemp and P. J. Gilbert, *J. Am. Chem. Soc.*, 1995, **117**, 8407–8415.
17 J. P. Piquemal, B. Williams-Hubbard, N. Fey, R. J. Deeth, N. Gresh and C. Giessner-Prettre, *J. Comput. Chem.*, 2003, **24**, 1963–1970.
18 M. B. Darhovskii, M. G. Razumov, I. V. Pletnev and A. L. Tchougreeff, *Int. J. Quantum Chem.*, 2002, **88**, 588–605.
19 S. M. Woodley, P. D. Battle, C. R. A. Catlow and J. D. Gale, *J. Phys. Chem. B*, 2001, **105**, 6824–6830.
20 R. J. Deeth and L. J. A. Hearnshaw, *Dalton Trans.*, 2006, 1092–1100.
21 R. J. Deeth and L. J. A. Hearnshaw, *Dalton Trans.*, 2005, 3638–3645.
22 R. J. Deeth, N. Fey and B. J. Williams-Hubbard, *J. Comput. Chem.*, 2005, **26**, 123–130.
23 MOE, Molecular Operating Environment 2005.06, 2003 edn, Chemical Computing Group, Montreal, Montreal, 2005.
24 M. R. Bray, R. J. Deeth, V. J. Paget and P. D. Sheen, *Int. J. Quantum Chem.*, 1997, **61**, 85–91.