Fast and Accurate Quantum Crystallography: from Small to Large, from Light to Heavy

Lorraine A. Malaspina,[a] Erna K. Wieduwilt,[a,b] Justin Bergmann,[a,%] Florian Kleemiss,[a,†] Benjamin Meyer,[b,&] Manuel F. Ruiz-López,[b] Rumpa Pal,[a,§] Emanuel Hupf,[a,$] Jens Beckmann,[a] Ross O. Piltz,[c] Alison J. Edwards,[c] Simon Grabowsky,[a,†*] Alessandro Genoni[b*]

[a] *Institut für Anorganische Chemie und Kristallographie, Fachbereich 2 – Biologie/Chemie, Universität Bremen, Leobener Straße 3 and 7, 28359 Bremen, Germany*

[b] *Université de Lorraine, CNRS, Laboratoire LPCT, 1 Boulevard Arago, 57078, Metz, France*

[c] *Australian Nuclear Science and Technology Organisation, Australian Centre for Neutron Scattering, New Illawarra Rd, Lucas Heights NSW 2234, Australia*

Present addresses:

[%] *Lund University, Department of Theoretical Chemistry, Chemical Center, P.O. Box 124, S-22100 Lund, Sweden*

[†] *Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, CH-3012, Bern, Switzerland*

[&] *Laboratory for Computational Molecular Design (LCMD), Institute of Chemical Sciences and Engineering (ISIC), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

[§] *Division of Physics, Faculty of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8571, Japan*

[$] *Department of Chemistry, University of Alberta, 11227 Saskatchewan Drive, Edmonton, Alberta, Canada T6G 2G2*

ABSTRACT

The coupling of the crystallographic refinement method Hirshfeld Atom Refinement (HAR) with the recently constructed libraries of extremely localized molecular orbitals (ELMOs) gives rise to the new quantum-crystallographic method HAR-ELMO. This method is significantly faster than HAR but as accurate and precise, especially concerning the free refinement of hydrogen atoms from X-ray diffraction data, so that the first fully quantum-crystallographic refinement of a protein is presented here. However, the promise of HAR-ELMO exceeds large molecules and protein crystallography. In fact, it also renders possible electron-density investigations of heavy elements in small molecules, such as mercury as presented here, and facilitates to detect and isolate systematic errors from physical effects.

TOC GRAPHICS
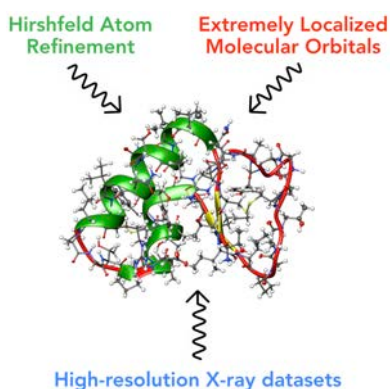
Structure determination via single-crystal X-ray diffraction is undoubtedly the most important technique for the experimental elucidation of the connectivity and three-dimensional structure of small as well as large molecules. Famous examples in the history of science are the X-ray structure determinations of biological molecules such as vitamins[1], DNA[2] or proteins[3]. The method is so successful and widespread that it gave rise to about 30 Nobel Prizes in chemistry, physics and medicine.[3] Despite its paramount importance, it is sometimes forgotten that the refinement of the diffraction data is based on a crude approximation. In fact, the electron densities of the atoms are estimated as spheres (so-called Independent Atom Model, IAM[4]), which yields accurate atomic positions for all atoms except hydrogen, but no direct information on chemical bonding and lone pairs, energies or other properties of the molecules or materials under investigation. Moreover, if hydrogen atom positions are not attainable, intermolecular interactions such as hydrogen bonding or agostic interactions cannot be characterized adequately,[5] thermodynamic properties cannot be estimated,[6] and protonation states of enzymes remain unknown[7]. If electron deformation due to chemical effects is neglected, bond properties in materials or in metalloproteins are in doubt.[8]

To solve such problems, on the experimental side, an enormous technical and financial effort has been made to obtain increasingly higher quality and resolution X-ray diffraction data sets, especially in protein crystallography. More and more dedicated end stations at synchrotron sources around the world provide a rapidly increasing number of high-resolution structures of proteins.[9,10] On the other hand, method development is trying to keep up by providing tools to replace the IAM with refinement techniques that take into account electron deformation and can potentially locate hydrogen atoms. For small molecules, multipole modeling has become a mature technique[11] and provides information on chemical bonding,[12] although refining hydrogen[13,14,15] and heavy atom[16,17] parameters is only possible in special cases. Unfortunately, for large molecules, multipole modeling is not feasible due to limited

computer resources. However, fixed multipole parameters are stored in properly designed databanks[18,19,20] and successfully transferred to the protein structures of choice to perform accurate refinement of coordinates and atomic displacement parameters (ADPs).[21,22]

Introducing quantum-chemical information into X-ray refinements to replace the IAM is an important aspect of quantum crystallography.[23,24] Within this framework, Hirshfeld Atom Refinement (HAR)[25,26] has been shown to be able to locate hydrogen atoms for small molecules from routine X-ray data with the same accuracy and precision as from neutron-diffraction experiments[27] and now is accessible in standard crystallographic refinement software.[28] Since it is based on regularly updated full wavefunction calculations, it is too slow to be applied to proteins or molecules containing heavy elements, the two compound classes that have been addressed above and that bear the most pressing problems of crystallography. For proteins, quantum-based refinements have been realized by utilizing restraints calculated quantum mechanically.[29,30,31]

In this work, we introduce a new method of quantum crystallographic refinement (HAR-ELMO) that combines HAR with the recently constructed libraries of extremely localized molecular orbitals (ELMOs) that rapidly provide wavefunctions of large and complicated molecules[32]. In fact, it has been shown previously that ELMOs for bonds and functional groups exhibit the same transferability from model compounds to proteins (or any other organic framework) as multipoles do, but with the significant advantage of allowing the reconstruction of full wavefunctions for the target systems.[33,34]

Here, we apply the new HAR-ELMO strategy to the amino acid *L*-alanine (*L*-Ala)[35] and the dipeptide glycyl-*L*-alanine (Gly-*L*-Ala)[26,36] to validate it against the original HAR technique and against suitable neutron-diffraction studies. Subsequently, we apply it to the polypeptide Leu-enkephalin,[37] the fibril-forming segment of the human prion protein[38] and to crambin[39] to provide the first full and complete multipole-free quantum crystallographic

refinement of a protein. This opens the door for a new research stream in protein crystallography. To demonstrate the potential applicability of HAR-ELMO in the field of coordination chemistry (e.g., organometallic compounds and metalloproteins), we apply it to the two coordination compounds $(Ph_3P)_2Hg(NO_3)_2$ and $(Ph_3P)_2Hg(CN)_2$, which are the first two fully quantum crystallographic refinements of molecules containing heavy elements.

Results. Before applying the novel HAR-ELMO technique to proteins, we validated it on smaller systems, namely on Gly-*L*-Ala at five different temperatures (12, 50, 100, 150 and 295 K) and on *L*-Ala at three different temperatures (23, 100 and 150 K). The results clearly indicate that approximations introduced by the ELMO approach do not reduce the accuracy and precision of the HAR results (see Figure 1 and Supporting Information). In fact, unlike IAM, C-H and N-H bond distances obtained at HAR-ELMO level are as accurate (average values agree within one sample standard deviation (ssd) as shown in Figure 1a) and as precise (ssd values depicted as error bars in Figure 1a approximately agree) as the HAR and neutron results for all temperatures. This also holds for intermolecular hydrogen-bond parameters (Tables S14 and S15 in the Supporting Information). This means that the hydrogen atoms could be located successfully, but using a significantly lower CPU time compared to traditional HAR (see Table 1 below).

HAR-ELMO even allows obtaining hydrogen Anisotropic Displacement Parameters (ADPs) with precision and accuracy similar to that observed for the bond lengths. ADPs for hydrogen atoms are known to be important for the experimental determination of thermodynamic properties.[6] From visual inspection of Figure 1b, there is a qualitative agreement between the hydrogen and non-hydrogen ADPs obtained through neutron, HAR and HAR-ELMO refinements of Gly-*L*-Ala at 150 K. This is confirmed quantitatively by extensive statistical

5

analyses detailed in the Supporting Information for both compounds at all investigated temperatures.
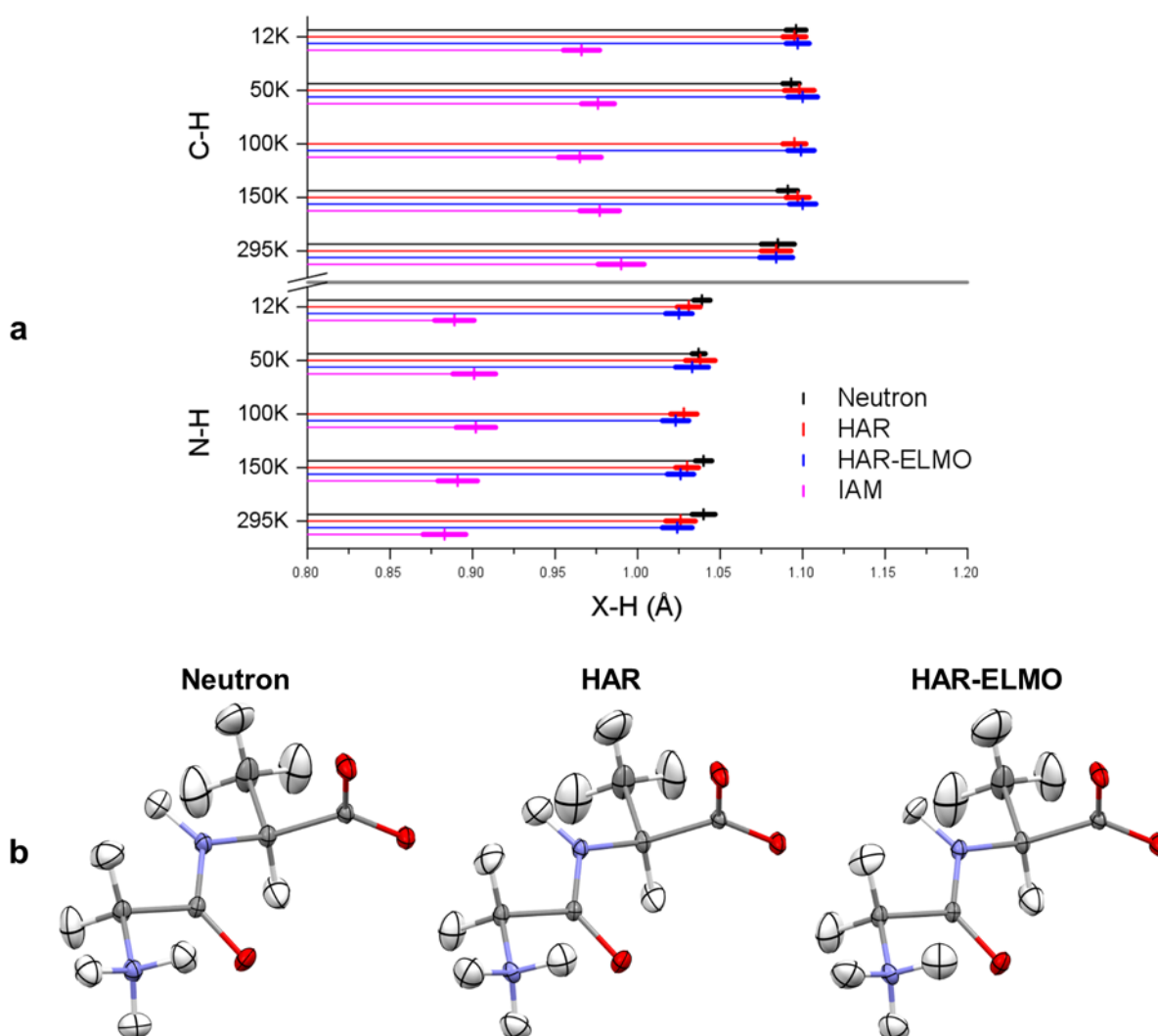

Figure 1. a) Accuracy and precision of bond distances involving hydrogen atoms (X-H) for Gly-*L*-Ala at all temperatures (note that there is no neutron diffraction experiment at 100 K). Estimated accuracy is expressed by a match of the averaged bond lengths with the reference neutron experiment. Estimated precision is expressed by the size of the error bar, which represents the sample standard deviation upon averaging.

b) Visual comparison of refinement results of Gly-*L*-Ala at 150K for neutron, HAR and HAR-ELMO refinements. The remaining comparisons are depicted in the Supporting Information. Mean absolute differences (MADs) relative to the neutron experiment for Gly-*L*-Ala at 150K: $MAD(d_{X-H,IAM}) = 0.128$Å, $MAD(d_{X-H,HAR-ELMO}) = 0.013$Å, $MAD(d_{X-H,HAR}) = 0.008$Å; $MAD(U_{ij,H,HAR-ELMO}) = 0.0079$Å^2, $MAD(U_{ij,H,HAR}) = 0.0078$Å^2.

After validation against amino acids and dipeptides, we applied HAR-ELMO to two polypeptides (Leu-enkephalin and the fibril-forming segment of the human prion protein) as well as to the protein crambin at two different resolutions (d=0.54 and 0.73 Å). All the refinements successfully converged, yielding atomic coordinates and atomic displacement parameters for hydrogen atoms. Only in crambin, the constraint of some atomic parameters mainly belonging to the outermost and flexible loop of the protein was required because of disorder (see the Methods section and the Supporting Information for more details). In Figure 2a the HAR-ELMO refined crambin structure is depicted, constituting the first quantum crystallographic structure of a protein. In Figure 2b, we show deformation density maps for a peptide bond fragment and a disulphide bond of crambin, which highlights the capability of the new HAR-ELMO method to describe bonding and lone-pair valence electron density distributions accurately. Residual and deformation densities for other regions of crambin and of the other two polypeptides can be found in the Supporting Information.
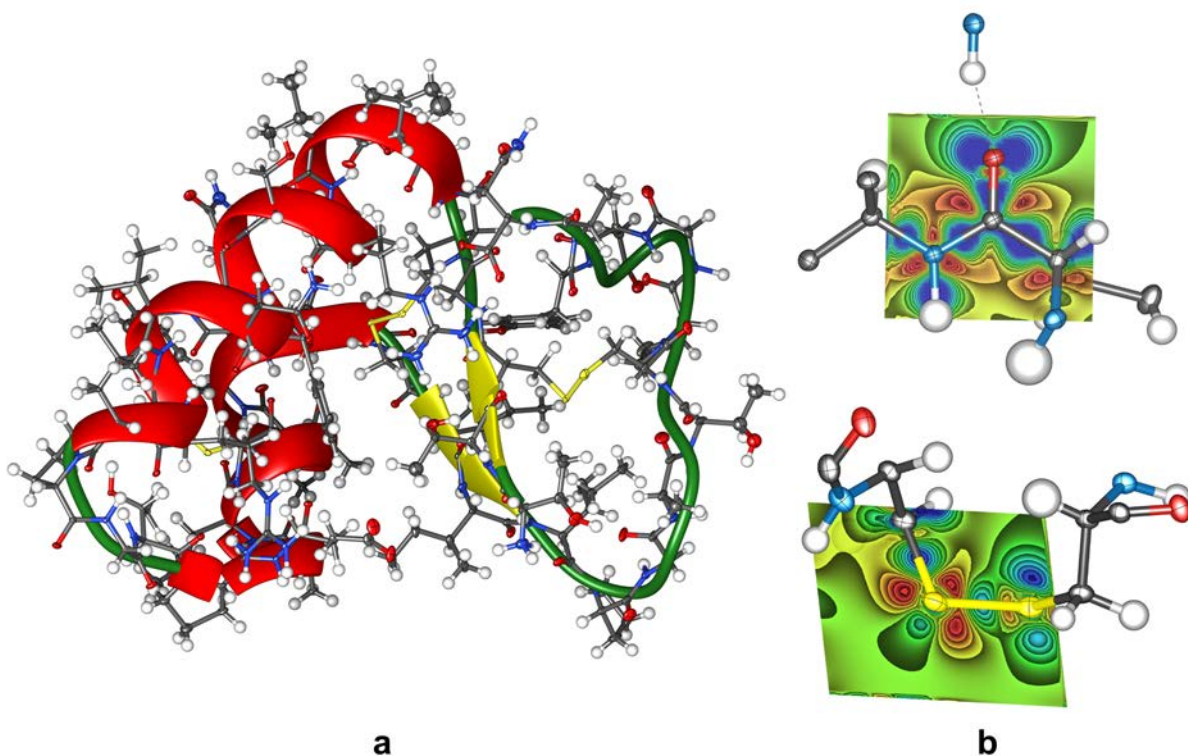
Figure 2. a) Refined protein structure of crambin at d=0.54 Å, not showing the disordered regions. For clarity, all H atoms are drawn with fixed spheres of 0.3 Å radius. Different representations for crambin at both selected resolutions and for the polypeptides are given in the Supporting Information. b) Deformation density maps of crambin in a peptide region (C=O in Leu25) and a disulphide bond (between Cys4 and Cys32). Contour interval: 0.05 eÅ^{-3}, blue=positive, red=negative, green=zero. For a) and b) ellipsoids at 50% probability.

As a consequence of the improved modelling of bonding effects, the novel HAR-ELMO method is able to determine the positions of hydrogen atoms precisely and accurately, also in situations where they could not be detected previously. This can be seen in Figure 3, where, for the fibril-forming segment of the human prion protein, we show the model of a hydrogen bond interaction in the deposited PDB file (2OL9, Figure 3a) and upon HAR-ELMO refinement (Figure 3b). In the deposited model, all the hydrogen atoms are missing although residual densities clearly indicate their presence (and the presence of related hydrogen bonds), while, after detecting them, we could refine them freely exploiting the IAM and HAR-ELMO models. Modeled by IAM, the indicated N-H bond length is too short, whereas, modeled by

HAR-ELMO, it matches the reference neutron bond length of 1.03 Å.[40] Determination of accurate X-H bond lengths allowed us to calculate intermolecular interaction energies[41] between the fibril strands, which are a key to the investigation of the biological relevance of fibrils.[38] Parallel fibril strands are held together with a total interaction energy of 169 kJ/mol, whereby the electrostatic component and the dispersion component are about equal in size. This means that hydrogen bonding does not dominate the interaction inside the fibril. Overall, average bond lengths involving hydrogen atoms in the two polypeptides and crambin are significantly elongated with respect to the IAM refinements, and for all C-H bonds and most N-H and O-H bonds they agree with bond lengths derived from neutron diffraction[40] within a single sample standard deviation (see Figure 3c). For O-H and N-H bonds in crambin, the agreement is more problematic, which is independent of the resolution, but only depending on the data quality of protein crystallographic measurements (compare with Figure 1a for small-molecule crystallography).

The success of the new HAR-ELMO technique with respect to hydrogen atom localization derives from the fact that hydrogen has only one electron, which is always involved in bonding interactions and it cannot be properly described with spherical atomic electron densities, as usually implemented in IAM. Overall, accounting for valence bonding deformations ensures a better agreement of the model with the measured diffraction pattern as long as the quality of the X-ray measurements includes this information. Table 1 shows, in terms of crystallographic R values, that this is the case for Gly-*L*-Ala, *L*-Ala, Leu-enkephalin and the mercury complexes that will be discussed below. For Leu-enkephalin, the improvement is significant (from R_{IAM}=0.0557 to $R_{HAR-ELMO}$=0.0430), and Figure S87 (see Supporting Information) shows that this is due to residual electron density in the bonding and lone-pair regions after IAM. For crambin and the fibril forming segment of the human prion

protein, the quality of the measurements is not sufficient to observe valence electron density (see Figures 3a and 3b), so there is no improvement in R-values, but only in refined structural parameters and in electronic properties contained in the wavefunction.
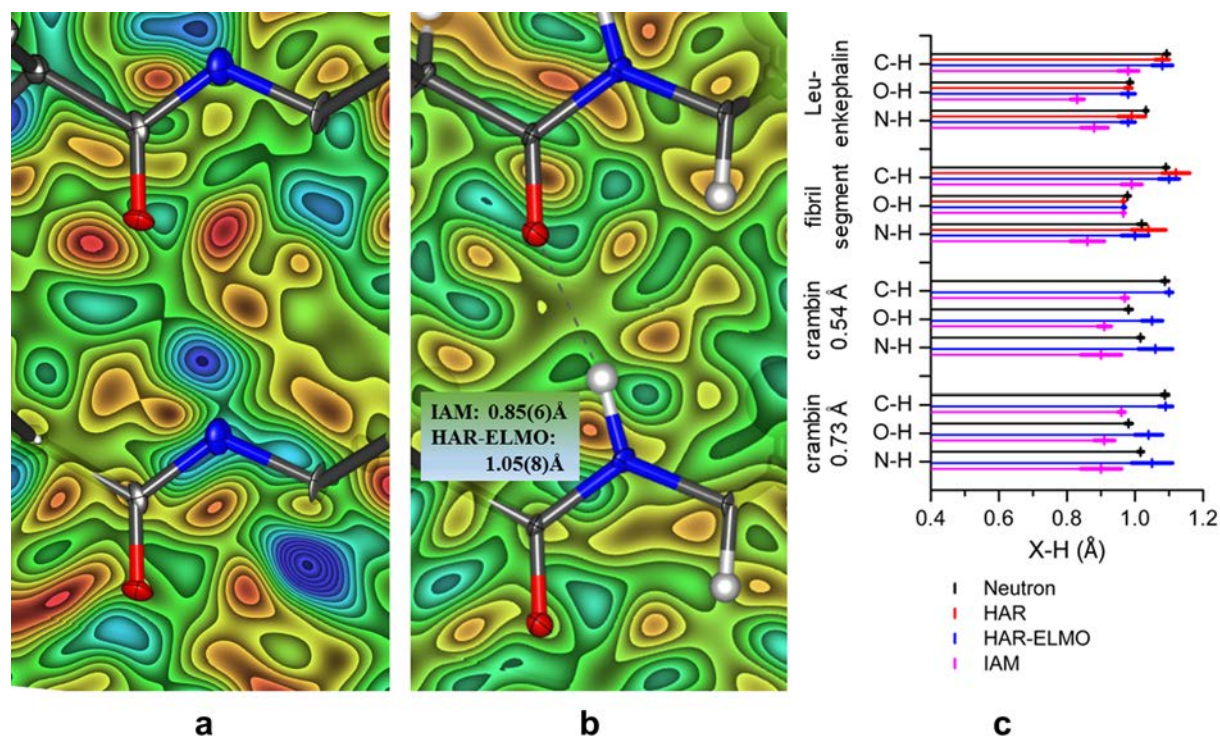

Figure 3. Residual electron density of the investigated fibril forming segment (original PDB structure: 2OL9) in a region of intermolecular interaction between two parallel strands (namely, hydrogen-bond N2-H2... O2=C3 [according to the deposited CIF labels] between residues Asn2 and Ser1): a) IAM without hydrogen atoms as deposited in the original PDB, b) HAR-ELMO with refined hydrogen atoms, including the IAM and HAR-ELMO refined N-H distances. For (a) and (b), contour interval=0.05 eÅ^{-3}, blue=positive, red=negative, green=zero. For a) and b) ellipsoids at 50% probability.

c) Average values and associated standard deviations of C-H, O-H and N-H bond lengths in IAM, HAR (if possible) and HAR-ELMO structures of the fibril-forming segment, Leu-enkephalin and crambin compared to average values from neutron diffraction.[40]

Table 1. CPU wall-clock timing and refinement statistics. The values in brackets refer to the IAM R-values obtained after the refinement with the software TONTO. For Gly-*L*-Ala and *L*-Ala, only the 150K measurements are given as examples (complete details are given in Table S11 of the Supporting Information).

Structure	Method	CPU wall time dd:hh:mm:ss	$R_1(F)$	χ^2
Gly-*L*-Ala (150K)	HAR	00:00:12:57	0.0161	2.336
	HAR-ELMO	00:00:03:36	0.0168	2.725
			(0.0251)	
L-Ala (150K)	HAR	00:00:05:55	0.0202	3.152
	HAR-ELMO	00:00:02:25	0.0210	3.523
			(0.0279)	
Leu-Enkephalin	HAR	00:09:52:00	0.0422	0.505
	HAR-ELMO	00:01:44:17	0.0430	0.545
			(0.0557)	
Fibril forming segment	HAR	01:07:00:00	0.0436	9.903
	HAR-ELMO	00:00:22:56	0.0474	11.913
			(0.0446)	
Crambin (d=0.54Å)	HAR	Impossible	-	-
	HAR-ELMO	09:23:47:53	0.0715	5.004
			(0.0704)	
Crambin (d=0.73Å)	HAR	Impossible	-	-
	HAR-ELMO	06:00:15:16	0.0624	7.672
			(0.0618)	
$(Ph_3P)_2Hg(NO_3)_2$	HAR	00:11:00:00	0.0162	4.232
	HAR-ELMO	00:00:31:47	0.0162	4.400
			(0.0204)	
$(Ph_3P)_2Hg(CN)_2$	HAR	00:13:30:00	0.0187	2.603
	HAR-ELMO	00:01:09:00	0.0187	2.622
			(0.0215)	

The more sensitive χ^2 values in Table 1 show that HAR is always in slightly better agreement with the measured diffraction pattern than HAR-ELMO because of the approximations used in the ELMO procedure, but this difference does not reduce the accuracy of geometrical and derived properties as shown above. This is important since, for crambin, and consequently every larger molecules, HAR is impossible or impractical because it is too computationally demanding requiring the recalculation of the wavefunction after every refinement step (see Table 1). Therefore, HAR-ELMO renders quantum crystallographic refinement possible for

11

protein crystallography. The CPU time depends on the size of the molecule (total number of atomic orbitals used in the calculation) and on the number of structure factors used in the least-squares refinement. For crambin, the number of reflections in the higher-resolution data set (d=0.54Å, 96139 refl.) is more than twice as high as the number in the lower-resolution data set (d=0.73Å, 45265 refl.). The HAR-ELMO CPU time reduces from almost 10 days to about 6 days (see Table 1) from the higher to the lower resolution data set.

The reduction of CPU time from HAR to HAR-ELMO for the two coordination compounds $(Ph_3P)_2Hg(NO_3)_2$ and $(Ph_3P)_2Hg(CN)_2$ is significant (22- and 13-fold reduction, Table 1). Figure 4b also shows that the hydrogen atom positions and ADPs in the phenyl groups of $(Ph_3P)_2Hg(NO_3)_2$ are accurately determined, even with the heavy mercury dominating the scattering. A comparison of refinement results with isotropic hydrogen atom displacement parameters and with $(Ph_3P)_2Hg(CN)_2$ is given in the Supporting Information.

In the IAM refinement (Figure 4b), there is a significant amount of unmodelled physically meaningful residual density in the C-C, C-P, N-O, Hg-O and Hg-P bonds as well as in the Hg core region. The maximum unmodelled electron density peak in the Hg core is +2.1 e$Å^{-3}$. The HAR-ELMO refinement accounts for the deformation of the electron density caused by bonding, so unmodelled features are absent from the residual-density map after HAR-ELMO (Figure 4c). The maximum unmodelled core Hg density reduces to +1.6 e$Å^{-3}$, a significant reduction showing the importance of modelling the aspherical core deformation of heavy mercury. Similar, but less pronounced, features can be seen for $(Ph_3P)_2Hg(CN)_2$, described in the Supporting Information (maximum residual density for this compound being +1.3 e$Å^{-3}$ (IAM) and +1.1 e$Å^{-3}$ (HAR-ELMO)). HAR gives very similar results in terms of residual electron density (+1.7 and +1.1 e$Å^{-3}$ for $(Ph_3P)_2Hg(NO_3)_2$ and $(Ph_3P)_2Hg(CN)_2$, respectively).

a

Min/max res.: -0.843/+2.072 eÅ⁻³

b

Min/max res.: -0.680/+1.632 eÅ⁻³

c

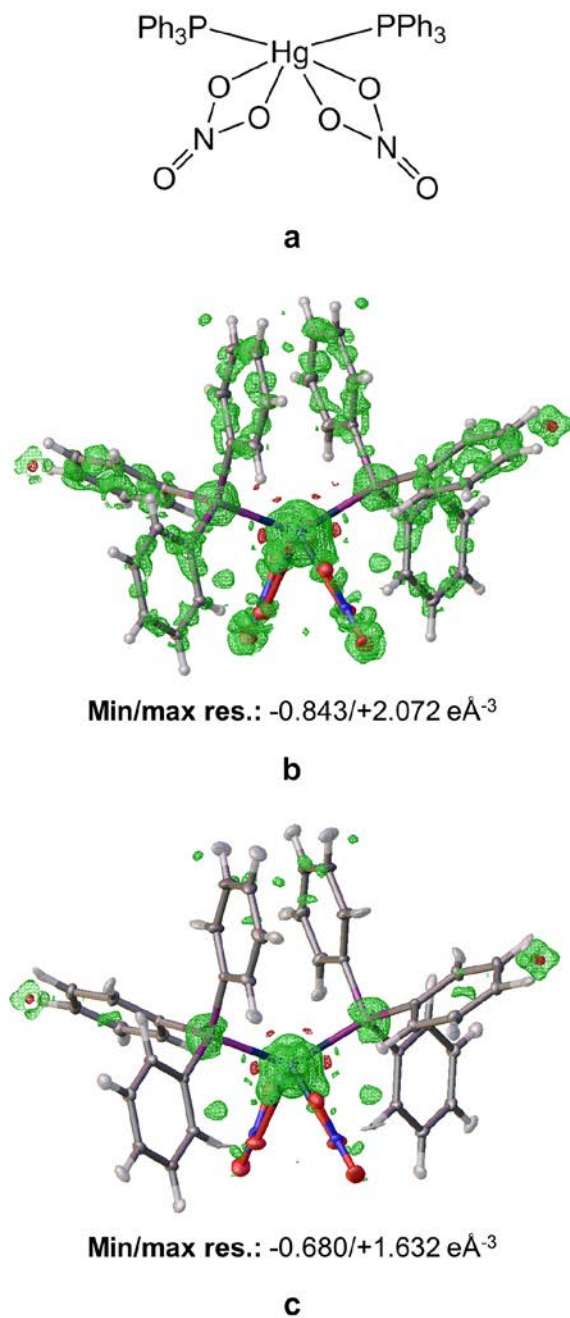**Figure 4.** Refined structures of a) $(Ph_3P)_2Hg(NO_3)_2$ with residual electron density distribution and minimum/maximum value b) after IAM and c) after HAR-ELMO. Isovalue = $\pm$ 0.2 eÅ⁻³, green=positive, red=negative. ADPs at 50% probability level. The two nitro and triphenylphosphine groups are symmetry-equivalent with their counterpart leading to six-fold coordination of Hg. Distances after HAR-ELMO: Hg-P = 2.4502(5)Å, Hg-O = 2.4692(7), Hg-O' = 2.8446(9).

The residual density distribution after HAR-ELMO shown in Figure 4c is not free of systematic effects. Unmodelled effects remain at the Hg and P cores, and there are two ghost peaks that are not due to aspherical bonding deformation. The HAR-ELMO treatment allows examination of these additional effects. The ghost peaks may be Fourier truncation errors that we could not remove in the data reduction procedure. The features at the Hg core are physical effects that the HAR-ELMO model (including anomalous dispersion and absorption correction) did not resolve. The most obvious problems in our theoretical ansatz are the neglect of electron correlation and relativistic effects. Therefore, we performed fully relativistic HARs at the B3LYP-IOTC level[42], which is currently not possible for HAR-ELMO. This refinement took 8 days and 19 hours, and only reduced the positive maximum residual density to 1.15 eÅ^{-3}, still leaving very similar systematic features at the Hg core (see Figure S97, Supporting Information). Other possible effects could be anharmonicity, insufficiently modelled absorption or radiation damage. All these effects can be investigated in detail in future projects because the fast HAR-ELMO refinements allowed us to separate deformation due to bonding from all the other important effects.

Discussion. Considering all the results shown above, the HAR-ELMO structural model includes chemically meaningful bonding and non-bonding features, in contrast to IAM. As long as the quality of the diffraction experiment is high enough to detect such features, the new HAR-ELMO refinement method leads to significantly improved agreements between the model and the measured diffraction data, as enumerated by reduced values of figures of merit (R values, χ^2, residual density distributions). When the quality of the measurements is not high enough, the intrinsic quantum-mechanical nature of the underlying model nevertheless allows the derivation of accurate experimental geometric parameters even in cases where

14

ADPs become non-positive definite in free HAR and free HAR-ELMO refinements, and can be used to obtain properties that are not restricted to those associated with the electron density of the system under consideration. Accessible properties are, for example, bonding descriptors such as the Electron Localizability Indicator (ELI-D)[43], electrostatic potentials (ESPs), non-covalent interaction (NCI) plots[44,45,46], or inter- and intra-molecular interaction energies, all of which are important for rational drug design.[47,48,49] Figure 5 shows the deformation density, ESP and the ELI-D for the entire crambin molecule, calculated rapidly on graphics cards using our own GPU-accelerated software cuQCT, based on the wavefunction and geometry from the last HAR-ELMO step.

For all the mentioned properties, accurate hydrogen atom positions are imperative. This study shows that HAR-ELMO can detect hydrogen atoms and locate them accurately and precisely, even for those large systems for which a fully quantum crystallographic refinement was previously considered prohibitive.

Notwithstanding the promise of the current results, the major drawback associated with protein refinements is solvent and intrinsic disorder. At the moment, there is no refinement engine underlying original quantum crystallography methods such as HAR that can treat this problem in a proper crystallographic way. It will be of utmost importance to develop and implement a novel strategy to refine disorder in HAR-ELMO or couple HAR-ELMO to existing software for protein refinement that already provides the solution. Although nowadays the number of exceptionally high-resolution protein structures is increasing, our study shows that the results of the HAR-ELMO strategy do not depend on the highest resolution (0.85 Å is sufficient), but, rather, on the quality of the data.
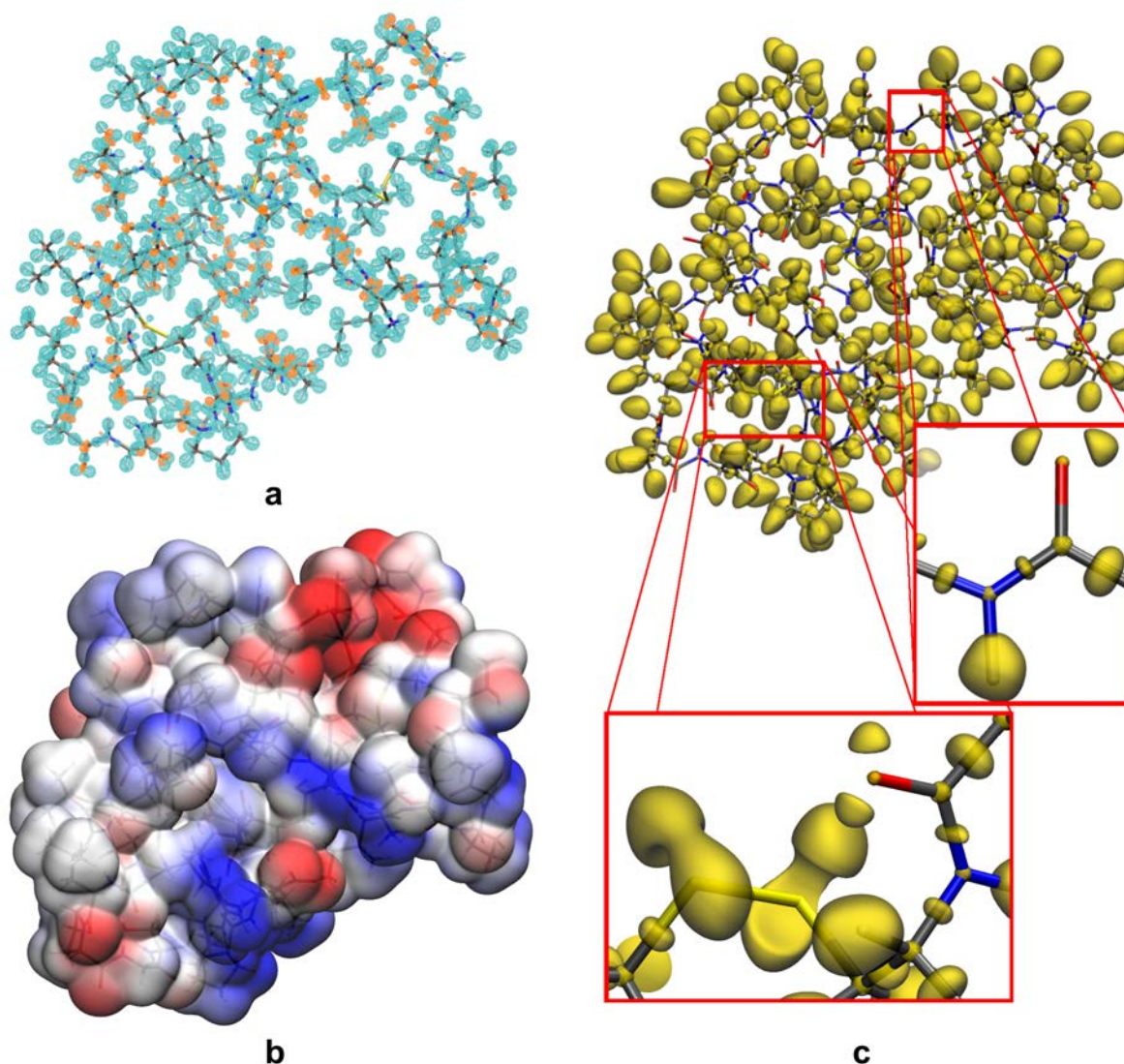
Figure 5. cuQCT GPU-accelerated calculations of molecular properties for crambin based on the final HAR-ELMO geometry and wavefunction: a) isosurfaces of deformation density at ±0.04 eÅ^{-3} (blue = positive, red = negative); b) Electrostatic potential in a range from -0.1 to +0.1 e bohr^{-1} mapped on an electron-density isosurface of 0.001 e bohr^{-3}; c) ELI-D at an isovalue of 2.2 for the main picture and 2.0 for the magnified details. All grids were calculated within 2 hours on a normal desktop computer with a grid-point separation of 0.1 bohr, except for the ESP grid which took about 500 hours at a 0.6 bohr separation. The program cuQCT is a software for the rapid calculation of large and dense grids of molecular properties on GPUs developed by some of us at the Universities of Bremen and Bern.

To improve the current version of the HAR-ELMO technique, on the one hand, we have already envisaged the implementation of a parallelized version of the presented HAR-ELMO procedure, also introducing linear-scaling options for the computation of one-electron

16

density matrices. This will result in a further and drastic reduction of the computational cost. On the other hand, we are also developing a multi-scale HAR-QM/ELMO technique. It will consist in coupling HAR with the QM/ELMO method,[50] a recently developed embedding strategy that describes the most chemically/biologically relevant part of the system (e.g., the active site of a protein or the heavy element in coordination compounds) at a very high QM level (e.g., including dispersion for weak intermolecular interactions in enzyme pockets, or core excitations and relativity for heavy elements), while treating the rest of the system with fixed ELMOs transferred from the databanks. This will increase the accuracy of the results for the regions of particular interest while keeping the computational cost within reasonable limits.

The HAR-ELMO refinement captures aspherical electron-density features, such as bonds, lone pairs and the core deformation of heavy elements, in a rapid and automated way. This allows the separation of such physical effects from unmodelled systematic effects, which are smaller but still significant. These unmodelled effects can be of two different natures: i) effects that are related to the adopted model (e.g., relativity, basis-set dependency and anharmonicity); ii) effects related to the measurement and data reduction strategy. To tackle problem (i) we propose to apply the above-mentioned HAR-QM/ELMO strategy. Furthermore, we can exploit the speed of HAR-ELMO and HAR-QM/ELMO in order to use them as debugging tools to address problem (ii). In this way, changes in data-reduction routines (for instance, absorption correction or mitigation of radiation damage) can directly be mapped on the electron density.

While traditional refinement strategies normally place hydrogen atoms according to valence considerations and not according to measured electron-density peaks and while the pioneering and innovative QM/MM refinements of proteins[29,30,31] are still based on the use of

tabulated spherical atomic form factors, the new HAR-ELMO technique introduces for the first time the explicit quantum mechanical treatment of the electron density asphericity in the computation of structure factors for the refinement of macromolecular crystal structures. This adds unprecedented insights into protein structure determination. Moreover, for the chemical-crystallographic community, a major advantage of the new approach also lies in the future possibility of extending the ELMO libraries to very different classes of compounds, such as protein inhibitors or ligands that regularly occur in syntheses of, for example, metal and covalent organic frameworks (MOFs and COFs). In particular, if ligands are sterically demanding or contain heavy elements in a closed-shell state, resorting to such ELMO libraries will render quantum crystallographic refinements possible for high-throughput investigations of chemical bonding features.

In *conclusion*, two classes of compounds can now be modelled by quantum crystallography thanks to the developments presented here: proteins and molecules containing heavy elements. For large molecules, the treatment of the lightest atoms is crucial; for small molecules, the treatment of the heaviest elements is the biggest challenge. HAR-ELMO is today the only method that allows fully quantum-crystallographic refinements of both such very different compound types on a reasonable timeframe. This reduction of computational cost derives from the introduction of the ELMO approximation in the procedure, but, despite this, accuracy and precision are only marginally affected. In future work, we will exploit the new possibilities in macromolecular chemistry to investigate bonding situations in metalloproteins and for enzyme-ligand binding as well as in chemical crystallography to accurately model bonding to heavy elements and understand unusually close intermolecular contacts, e.g. in halogen and hydrogen bonding motifs.

COMPUTATIONAL AND EXPERIMENTAL METHODS

In this section, all the computational and experimental methods used in the present investigation are briefly presented. More details are given in the Supporting Information.

HAR-ELMO implementation. The HAR-ELMO technique was implemented interfacing the HAR functionalities in the software TONTO[51] with the ELMOdb program associated with the ELMO libraries.[32] The interface between the programs is the lamaGOET graphical user interface (GUI), which reads the initial geometry (PDB, CIF or XYZ file) from a previous IAM refinement and provides it to ELMOdb for the automatic transfer of molecular orbitals. After the transfer, the wavefunction and first-order density matrix of the system are computed and written to a formatted checkpoint file, which is afterwards passed to TONTO by lamaGOET together with the ADPs of the IAM refinement, in order to perform the first HAR iteration. The updated geometry and ADPs are used as input for iteratively repeating the procedure until convergence. Although interfaced, lamaGOET and ELMOdb are currently two independent software. The former is available free of charge at the link: http://www.tinyurl.com/lamaGOET (source code is also available on https://github.com/lomalaspina/lamaGOET). On the contrary, at the moment, the ELMO libraries and the associated ELMOdb program are not publicly available, but they can be obtained upon motivated request to A.G. (Alessandro.Genoni@univ-lorraine.fr). The construction of a dedicated webserver is already envisaged.

ELMO libraries. The used ELMO libraries[32] cover all the functional units of water and of the twenty natural amino acids in all their possible protonation states and forms. The associated ELMOdb program can also read tailor-made ELMOs properly computed to describe fragments of specific molecules/ligands beyond the standard amino acids and the water

molecule. Other than selecting the molecular orbitals associated with the fragments of the system under exam, ELMOdb also automatically handles the rotation of the ELMOs by simply reading the geometry of the target molecule, without the need of defining any local coordinate systems as usually done in multipole refinements. In particular, following a method originally introduced by Philipp and Friesner[52], this is done by the program defining, for each fragment, a proper rotation matrix from the reference frame in the model molecule (on which the ELMOs were originally computed) to the reference frame in the target system (see Supporting Information for more details).

Additional refinement details. For the HAR-ELMO refinements of systems that only involve amino acids and water molecules, we exploited the ELMO library for the 6-311G(d,p) basis-set. For the HAR-ELMO refinements of the two coordination compounds we used tailor-made ELMOs computed on proper model molecules with the DZP-DKH basis-set. For comparison, we performed traditional HARs at the respective same basis-sets and at Hartree-Fock level, except for crambin for which traditional HAR is too time-consuming. In addition, for $(PPh_3)_2Hg(NO_3)_2$ a HAR at fully relativistic infinite-order two-component level (IOTC) was performed with the DZP-DKH basis-set.[42] All refinements were performed in TONTO 18.12.17 github v. 2b3dd9c.

Validation of HAR-ELMO. The experimental X-ray[26] and neutron[36] data of Gly-*L*-Ala were obtained from the literature. For *L*-Ala, the X-ray data at 23K were taken from Destro *et al.*[35] while the data at 100K and 150K were measured using a Bruker Venture D8 diffractometer with a Mo-K$_\alpha$ microsource and a Photon100 CMOS detector to a maximum resolution of 0.45Å(100K) and 0.55Å(150K). Neutron-diffraction data of *L*-Ala at 23K and 150K were measured at the Australian Centre for Neutron Scattering of ANSTO (beamline KOALA)

using the Laue diffraction technique.[53] KOALA is equipped with a large cylindrical image plate, so that a resolution of 0.50Å was reached in the wavelength range 0.750 to 1.500 Å, refined to R-values R(F)= 0.033/0.038. Other than in the Supporting Information, more details are in the Cambridge Structural Database (CSD) under nos. CCDC-1917810-1917812, 1917816-1917818, 1917820-1917822, 1917823-1917837 (X-ray) and CCDC-1917685, 1917813 (neutron).

Protein refinement. The experimental data of crambin were taken from the literature.[39] In the available PDB file, there was an unspecified and unresolved amount of disordered water molecules. Since disorder for HAR, HAR-ELMO or any other refinement belonging to the original definition of quantum crystallography is currently an unsolved problem, we used the SQUEEZE routine[54] to calculate the scattering contributions of the solvent molecules and excluded them from the original structure factors. These structure factors were used for full-matrix least-squares IAM and HAR-ELMO refinements in TONTO at two resolutions, 0.54 and 0.73Å. Parameters of atoms belonging to lower occupancy regions in the reported crambin structure were removed and only the major component atoms were included at full occupancy. Positional and displacement parameters were freely refined for all atoms but those in the disorder regions and those of some atoms belonging to the outermost flexible loop of crambin. For those atoms, positions and displacement parameters were kept fixed at values obtained from an initial unconstrained IAM refinement.

The original structure in the 2OL9 PDB file for the fibril-forming segment of the human prion protein is resolved up to 0.85 Å, contains ordered oxygen sites of water solvent molecules, but no hydrogen atoms.[38] Before the initial IAM refinement, all hydrogen atoms were geometrically placed, and then refined freely in the subsequent HAR and HAR-ELMO refinements. The experimental data of Leu-enkephalin[37] are resolved to d=0.43Å, and all

atoms are present and ordered in the original file. The reference IAM refinement was performed in the software TONTO, and followed by both HAR and HAR-ELMO treatments. The files associated with the polypeptide refinements correspond to CSD depositions CCDC-1917596-1917598, 1917730-1917732.

Small-molecule refinement involving heavy elements. The two coordination compounds $(Ph_3P)_2Hg(NO_3)_2$ and $(Ph_3P)_2Hg(CN)_2$ are known from literature,[55] but were re-synthesized according to the literature procedures, crystallized and measured at the same X-ray setup described above to resolutions of 0.50 and 0.55Å, respectively. IAM, HAR and HAR-ELMO refinements were performed in TONTO. For HAR and HAR-ELMO, two models were adopted, one with anisotropic and one with isotropic hydrogen displacement parameters. $(Ph_3P)_2Hg(NO_3)_2$ is the only compound in this study with half a molecule in the asymmetric unit (Z'=0.5), whereas for all the others Z'=1.

ASSOCIATED CONTENT

Supporting Information

Theoretical bases of the HAR-ELMO technique, details about the validation of the HAR-ELMO method, details and further results of the polypeptides and protein refinements, details and further results of the refinements of the coordination compounds. (PDF)

Crystallographic information files (cif files) and files containing experimental and calculated structure factors (fco files) associated with all the refinements (all the considered models) for all the compounds in this study. (ZIP)

This material is available free of charge via the Internet at http://pubs.acs.org/.

DATA AVAILABILITY

The best-quality models for all compounds obtained in the present study (except those for crambin) have been deposited at the Cambridge Crystallographic Data Centre (CCDC) under the deposition numbers indicated in the methods section. This data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif. All models for all refinements (including the inferior models and including the protein crambin) plus the measured structure factors have been deposited with this journal in addition to the CCDC deposition to ensure reproducibility of the study. The lamaGOET GUI can be obtained free of charge for registered users at http://www.tinyurl.com/lamaGOET. All other programs developed for the present study are available from the corresponding authors of the paper upon motivated request.

AUTHORS INFORMATION

Corresponding Authors

- Alessandro Genoni (A.G.), Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), 1 Boulevard Arago, F-57078 Metz, France; Phone: +33 (0)3 72 74 91 70; Fax: +33 (0)3 72 74 91 87. E-mail: Alessandro.Genoni@univ-lorraine.fr.

- Simon Grabowsky (S.G.), Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, CH-3012, Bern, Switzerland; Phone: +41 31 631 4281. E-mail: simon.grabowsky@dcb.unibe.ch

Authors Contributions

A. G. and S. G. conceived the HAR-ELMO method and secured funding for the project, the latter together with J. Beckmann. L. A. M., A. G. and S. G. designed the study, and A.G. and S. G. supervised it. L. A. M. developed the software lamaGOET to couple HAR to the ELMO libraries with the help of A. G. L. A. M. performed all the structural refinements, obtained the

results for this paper and generated most of the display items. L. A. M. and E. K. W. performed the statistical analyses. E. K. W. compiled the supporting information file and generated Figure 3c. S. G. and A. G. wrote the paper. E. H. synthesized and crystallized both Hg-containing compounds. R. P. measured the data set of $(Ph_3P)_2Hg(NO_3)_2$. J. Bergmann and L. A. M. measured the data set of $(Ph_3P)_2Hg(CN)_2$ and performed the relativistic refinement of $(Ph_3P)_2Hg(NO_3)_2$ together. F. K. developed the code to calculate electrostatic potential, deformation density and electron localizability in cuQCT, carried out the cuQCT calculations and generated Figure 4. J. Beckmann and M. F. R.-L. helped to design the study and interpret the results. B. M. helped to generate some material for the Supporting Information. E. K. W., L. A. M. and R. P. measured the 23K neutron-diffraction data set of L-alanine under the supervision of R. O. P. The 150 K neutron-diffraction data set of *L*-alanine was measured by A. J. E. The 100 and 150 K X-ray data sets of *L*-alanine were measured by L. A. M.

Notes

The authors declare no competing financial interests.

REFERENCES

[1] Hodgkin, D. C.; Kamper, J.; Mackay, M.; Pickworth, J.; Trublood, K. N.; White, J. G. Structure of Vitamin B$_{12}$. *Nature* **1956**, *178*, 64-66.

[2] Watson, J. D; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953***, 171*, 737-738.

[3] Jaskolski, M.; Dauter, Z.; Wlodawer, A. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. *FEBS J.* **2014**, *281*, 3985-4009.

[4] Compton, A. H. The Distribution of Electrons in Atoms. *Nature* **1915**, *95***,** 343–344.

[5] Braga, D.; Grepioni, F.; Biradha, K; Desiraju, G. R. Agostic interactions in organometallic compounds. A Cambridge Structural Database study. *J. Chem. Soc., Dalton Trans.* **1996**, *0*, 3925-3930.

[6] Hoser, A. A.; Madsen, A. Ø. Dynamic quantum crystallography: lattice dynamical models refined against diffraction data. II. Applications to *L*-alanine, naphthalene and xylitol. *Acta Cryst. A* **2017**, *73*, 102-114.

[7] Fisher, S. J.; Blakeley, M. P.; Cianci, M.; McSweeney, S.; Helliwell, J. R. Protonation-state determination in proteins using high-resolution X-ray crystallography: effects of resolution and completeness. *Acta Cryst. D* **2012**, *68*, 800-809.

[8] Lu, Y.; Yeung, N.; Sieracki, N.; Marshall. N. M. Design of functional metalloproteins. *Nature* **2009**, *460*, 855-862.

[9] Petrova, T.; Podjarny, A. Protein crystallography at subatomic resolution. *Rep. Prog. Phys.* **2004**, *67*, 1565-1605.

[10] Ogata, H.; Nishikawa, K.; Lubitz, W. Hydrogens detected by subatomic resolution protein crystallography in a [NiFe] hydrogenase. *Nature* **2015**, *520*, 571-574.

[11] Coppens, P. Charge Density Come of Age. *Angew. Chem. Int. Ed.* **2015**, *44*, 6810-6811.

[12] Stalke, D. *Electron Density and Chemical Bonding I. Experimental Charge Density Studies*; Springer-Verlag: Berlin & Heidelberg, 2012.

[13] Hoser, A. A.; Dominiak, P. M.; Woźniak, K. Towards the best model for H atoms in experimental charge-density refinement, *Acta Cryst. A* **2009**, *65*, 300-311.

[14] Zhurov, V. V.; Zhurova, E. A.; Stash, A. I.; Pinkerton, A. A. Importance of the consideration of anharmonic motion in charge-density studies: a comparison of variable-temperature studies on two explosives, RDX and HMX. *Acta Cryst. A* **2011**, *67*, 160-173.

[15] Dittrich, B.; Lübben, J.; Mebs, M.; Wagner, A.; Luger, P.; Flaig, R. Accurate Bond Lengths to Hydrogen Atoms from Single-Crystal X-ray Diffraction by Including Estimated Hydrogen ADPs and Comparison to Neutron and QM/MM Benchmarks. *Chem. Eur. J.* **2017**, *23*, 4605-4614.

[16] Iversen, B. B.; Larsen, F. K.; Pinkerton, A. A.; Martin, A.; Darovsky, A.; Reynolds, P. A. Accurate charge densities in days – use of synchrotrons, image plates and very low temperatures. *Acta Cryst. B.* **1999**, *55*, 363-374.

[17] Zhurov, V. V.; Zhurova, E. A.; Stash, A. I.; Pinkerton, A. A. Characterization of Bonding in Cesium Uranyl Chloride: Topological Analysis of the Experimental Charge Density. *J. Phys. Chem. A* **2011**, *115*, 13016-13023.

[18] A Domagała, S.; Fournier, B.; Liebschner, D.; Guillot, B.; Jelsch, C. An Improved Experimental Databank of Transferable Multipolar Atom Models – ELMAM2. Construction Details and Applications. *Acta. Cryst. A* **2012**, *68*, 337-351.

[19] Dominiak, P. M.; Volkov, A.; Li, X.; Messerschmidt, M.; Coppens, P. A Theoretical Databank of Transferable Aspherical Atoms and Its Application to Electrostatic Interaction Energy Calculations of Macromolecules. *J. Chem. Theory Comput.* **2007**, *3*, 232−247.

[20] Dittrich, B.; Hübschle, C. B.; Pröpper, K.; Dietrich, F.; Stolper, T.; Holstein, J. J. The Generalized Invariom Database (GID). *Acta Cryst. B* **2013**, *69*, 91-104.

[21] Elias, M.; Liebschner, D.; Koepke, J.; Lecomte, C.; Guillot, B.; Jelsch, C. Hydrogen atoms in protein structures: high-resolution X-ray diffraction structure of the DFPase. *BMC Research Notes* **2013**, *6*, 308.

[22] Pröpper, K.; Holstein, J. J.; Hübschle, C. B.; Bond, C. S.; Dittrich, B. Invariom refinement of a new monoclinic solvate of thiostrepton at 0.64 Å resolution. *Acta Cryst. D* **2013**, *69*, 1530-1539.

[23] Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159-4176.

[24] Genoni, A.; Bučinský, L.; Claiser, N.; Contreras-García, J.; Dittrich, B.; Dominiak, P. M.; Espinosa, E.; Gatti, C.; Giannozzi, P.; Gillet, J.-M.; Jayatilaka, D.; Macchi, P.; Madsen, A. Ø.; Massa, L. J.; Matta, C. F.; Merz, K. M. Jr.; Nakashima, P. N. H.; Ott, H.; Ryde. U.; Schwarz,

K.; Sierka, M.; Grabowsky, S. Quantum Crystallography: Current Developments and Future Perspectives. *Chem. Eur. J.* **2018**, *24*, 10881-10905.

[25] Jayatilaka, D.; Dittrich, B. X-ray structure refinement using aspherical atomic density functions obtained from quantum-mechanical calculations. *Acta Cryst. A* **2008**, *64* 383-393.

[26] Capelli, S.; Bürgi, H.-B.; Dittrich, B.; Grabowsky, S.; Jayatilaka, D. Hirshfeld atom refinement. *IUCrJ* **2014**, *1*, 361-379.

[27] Woińska, M.; Grabowsky, S.; Dominiak, P. M.; Woźniak, K.; Jayatilaka, D. Hydrogen atoms can be located accurately and precisely by X-ray crystallography. *Sci. Adv.* **2016**, *2*, e1600192.

[28] Fugel, M.; Jayatilaka, D.; Hupf, E.; Overgaard, J.; Hathwar, V. R.; Macchi, P.; Turner, M. J.; Howard, J. A. K.; Dolomanov, O. V.; Puschmann, H.; Iversen, B. B.; Bürgi, H.-B.; Grabowsky, S. Probing the accuracy and precision of Hirshfeld atom refinement with HARt interfaced with Olex2. *IUCrJ* **2018**, *5*, 32-44.

[29] Zheng, M.; Reimers, J. R.; Waller, M. P.; Afonine, P. V. Q|R: quantum-based refinement. *Acta Cryst. D* **2017**, *73*, 45-52.

[30] Ryde, U.; Olsen, L.; Nilsson, K. Quantum Chemical geometry optimizations in proteins using crystallographic raw data. *J. Comput. Chem.* **2002**, *23*, 1058-1070.

[31] Yu, N.; Yennawar, H. P.; Merz, K. M. Jr. Refinement of protein crystal structures using energy restraints derived from linear scaling quantum mechanics. *Acta Cryst. D* **2005**, *61*, 322-332.

[32] Meyer, B.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 3. Construction and Preliminary Assessment of the New Databanks. *J. Phys. Chem. A* **2018**, *122*, 8965-8981.

[33] Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability. *J. Chem. Theory. Comput.* **2016**, *12*, 1052-1067.

[34] Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Jelsch, C.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 2. Comparison with the Pseudoatoms Transferability. *J. Chem. Theory. Comput.* **2016**, *12*, 1068-1081.

[35] Destro, R.; Marsh, R. E.; Bianchi, R. A low-temperature (23 K) study of *L*-alanine. *J. Phys. Chem.* **1998**, *92*, 966–973.

[36] Capelli, S. C.; Bürgi, H.-B.; Mason, S. A.; Jayatilaka, D. Glycyl-*L*-alanine: a multi-temperature neutron study. *Acta Cryst. C* **2014**, *70*, 949–952.

[37] Pichon-Pesme, V.; Lecomte, C.; Wiest, R.; Bernard, M. Modeling fragments for the *ab initio* determination of electron density in polypeptides. An experimental and theoretical approach to the electron distribution in Leu-enkephalin trihydrate. *J. Am. Chem. Soc.* **1992**, *114*, 2713-2715.

[38] Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riekel, C.; Eisenberg, D. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007**, *447*, 453–457.

[39] Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Pesme, V.; Blessing, R. H.; Lecomte, C. Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc. Natl. Acad. USA* **2000**, *97*, 3171-3176.

[40] Allen, F. H.; Bruno, I. J. Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data. *Acta Cryst. B* **2010**, *66*, 380–386.

[41] Turner, M. J.; Grabowsky, S.; Jayatilaka, D.; Spackman, M. Accurate and Efficient Model Energies for Exploring intermolecular Interactions in Molecular Crystals. *J. Phys. Chem. Lett.* **2014**, *5*, 4249-4255.

[42] Bučinský, L.; Jayatilaka, D.; Grabowsky, S. Importance of Relativistic Effects and Electron Correlation in Structure Factors and Electron Density of Diphenyl Mercury and Triphenyl Bismuth. *J. Phys. Chem. A* **2016**, *120*, 6650-6669.

[43] Kohout, M. A measure of electron localizability. *Int. J. Quantum Chem.* **2004**, *97*, 651-658.

[44] Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing noncovalent interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498−6506.

[45] Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. NCIPLOT: a program for plotting noncovalent interaction regions. *J. Chem. Theory Comput.* **2011**, 7, 625−632.

[46] Arias-Olivares, D.; Wieduwilt, E. K.; Contreras-García, J.; Genoni, A. NCI-ELMO: A New Method To Quickly and Accurately Detect Noncovalent Interactions in Biosystems. *J. Chem. Theory Comput.* **2019**, DOI: 10.1021/acs.jctc.9b00658.

[47] Morra, G; Genoni, A.; Neves, M. A. C.; Merz, K. M., Jr.; Colombo, G. Molecular Recognition and Drug-Lead identification: What Can Molecular Simulations Tell Us? *Curr. Med. Chem.* **2010**, *17*, 25-41.

[48] Genoni, A.; Pennati, M.; Morra, G.; Zaffaroni, N.; Colombo, G. Ligand selection from the analysis of protein conformational substates: new leads targeting the N-terminal domain of Hsp90. *RSC Adv.* **2012**, *2*, 4268−4282.

[49] Ferraro, M.; D'Annessa, I.; Moroni, E.; Morra, G.; Paladino, A.; Rinaldi, S.; Compostella, F.; Colombo, G. Allosteric Modulators of HSP90 and HSP70: Dynamics Meets Function through Structure-Bases Drug Design. *J. Med. Chem.* **2019**, *62*, 60-87.

[50] Macetti, G.; Genoni, A. Quantum Mechanics/Extremely Localized Molecular Orbital Method: a Fully Quantum Mechanical Embedding Approach for Macromolecules. *J. Phys. Chem. A* **2019**, DOI: 10.1021/acs.jpca.9b08882.

[51] Jayatilaka, D; Grimwood, D. J. Tonto: A Fortran Based Object-Oriented System for Quantum Chemistry and Crystallography. In *Computational Science – ICCS 2003*; Sloot, P. M. A., Abramson, D., Bogdanov, A. V., Dongarra, J. J., Zomaya, A. Y., Gorbachev, Y. E., Eds.; Springer-Verlag: Berlin & Heidelberg, 2003; Chapter 4, pp 142-151.

[52] Philipp, D. M.; Friesner, R. A. Mixed Ab Initio QM/MM Modeling Using Frozen Orbitals and Tests with Alanine Dipeptide and Tetrapeptide. *J. Comput. Chem.* **1999**, *20*, 1468-1494.

[53] Piltz, R. O; Edwards, A. First results from the KOALA neutron Laue instrument *Acta Cryst. A* **2008**, *64*, C187.

[54] Spek, A. L. PLATON SQUEEZE: a tool for the calculation of the disordered solvent contribution to the calculated structure factors. *Acta Cryst. C* **2015**, *71*, 9-18.

[55] Buergi, H. B.; Fischer, E.; Kunz, R. W.; Parvez, M.; Pregosin, P. S. Correlation between NMR coupling constants and molecular structure. Synthesis and ^{31}P NMR measurements of [HgX$_2$(*cis*-Ph$_2$PCH=CHPPh$_2$)] and x-ray crystal structures of [HgBr$_2$(cis-Ph$_2$PCH=CHPPh$_2$)], [Hg(NO$_3$)$_2$(PPh$_3$)$_2$], and [Hg(CN)$_2$(PPh$_3$)$_2$]. *Inorg. Chem.* **1982**, *21*, 1246-1256.