

Electric charge and salting in/out effects on glucagon's dipole moments and polarizabilities using the GruPol database

Raphael F. Ligorio,^a Rasmus H. M. Gehle,^a Leonardo H. R. Dos Santos^b and Anna Krawczuk^{a*}

Received 22 November 2024

Accepted 6 February 2025

Edited by A. Martín Pendás, Universidad de Oviedo, Spain

This article is part of a collection of articles on Quantum Crystallography, and commemorates the 100th anniversary of the development of Quantum Mechanics.

Keywords: quantum crystallography; proteins; distributed polarizability approach; dipole moments; solvation effects.

Supporting information: this article has supporting information at journals.iucr.org/b

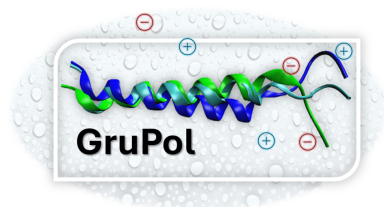
^aGeorg-August-Universität Göttingen, 37077 Göttingen, Germany, and ^bDepartment of Chemistry, Federal University of Minas Gerais, Avenida Pres. Antônio Carlos 6627, Belo Horizonte, Minas Gerais, 31270-901, Brazil. *Correspondence e-mail: anna.krawczuk@uni-goettingen.de

This work demonstrates the use of the GruPol database to predict the functional group dipole moments and polarizabilities of glucagon in the presence of NaCl, simulating an electric charge distribution on the protein's backbone. A new feature of the database allows for the inclusion of ions on the protein backbone, effectively simulating a protein salt and predicting the impact on electrical properties. Glucagon was selected as a proof-of-concept molecule due to its relatively small chain, which enabled benchmarking against quantum mechanical calculations. Firstly, we simulated 70 different ionic configurations, varying the number of Na⁺ and Cl⁻ ions from zero to four NaCl moieties. Additionally, we investigated the effects of solvation under two distinct conditions: one involving just the peptide and water, and the other also including NaCl at a concentration of approximately 4.2 mol L⁻¹. Regarding the *ab initio* results, GruPol showed good accuracy, with an angular direction error of around 10° and a 15% difference in the magnitude of the dipole moments. However, the error in polarizability values was higher, most likely due to the lack of an augmented basis set in the *ab initio* quantum calculations (M06-HF/cc-pVDZ). The database entries were generated using the same functional along with the aug-cc-pVDZ basis set. In solution, a high ionic concentration lowered the overall dipole moment, while the main components of polarizability increased.

1. Introduction

Understanding protein stability, reactivity and overall functionality is a central objective for many researchers, as these insights can elucidate the specific roles of proteins in medical conditions (Gebauer *et al.*, 2021; Gonzalez *et al.*, 2020; Hu *et al.*, 2022), inform the design of artificial catalysts (Li *et al.*, 2021) and even guide the engineering of novel proteins (Lovelock *et al.*, 2022). Many of a protein's characteristics are intrinsically linked to its electrical and electrostatic properties, including dipole moment and polarizability (the latter referring to a molecule's ability to produce a dipole moment in response to an externally applied electric field). These features are key determinants of intra- and intermolecular interaction energies, thus being crucial in predicting protein–ligand binding affinities (Goel *et al.*, 2020; Vascon *et al.*, 2020), solubility (Vascon *et al.*, 2020) and lattice energies (Ma *et al.*, 2023; Spackman, 2018).

Proteins are highly sensitive to their environment. A particular behaviour may only be detectable under specific conditions, such as a particular pH value, which can significantly change the charges, and hence electrical properties, on the protein's backbone (Kim *et al.*, 2024; de Resende *et al.*,



2024). It has also been well established that altering the salt concentration in a protein solution can either enhance or reduce the protein's solubility (Duan & Wang, 2024) (known as salting in or salting out, respectively), affect its activity (Martin del Campo *et al.*, 2023) or facilitate its crystallization process (Majeed *et al.*, 2003). These observations suggest that accurate assessment of a protein's properties requires careful consideration of the experimental conditions. For instance, cell organelles are not found in pure aqueous solutions; they exist in the cytosol, an environment containing various substances including salts. Due to the vast number of proteins, which are not easy to crystallize, the wide range of experimental variables and the lack of accurate experimental techniques, it is impractical to measure all possible combinations. Therefore, the ability to predict these properties accurately through theoretical approaches is of significant value.

With the use of *ab initio* quantum theoretical methods, it is possible to calculate electrical properties and reach the desired accuracy. However, a challenge remains, as most proteins contain a large number of atoms, making *ab initio* methods generally computationally expensive and often difficult to converge. To overcome these limitations, a database approach is frequently employed. This method relies on the transferability of electron density or wavefunction among functional groups within similar chemical environments, which can then be used to calculate electrical properties. For smaller fragments, such as individual atoms or recurring functional groups, the electron density is calculated and stored in the database for future use. Large molecules can then be broken down into these fragments, allowing the molecular properties to be reconstructed using the database. Several databases employ this approach, each with distinct objectives. For example, the MATTS database (Kumar *et al.*, 2019) utilizes aspherical pseudo-atoms to estimate electrostatic potential maps of small relevant biomolecules. The ELMAM2 database (Domagala & Jelsch, 2008), with similar purposes, emphasizes resource efficiency for faster predictions. The Generalized Invariom Database (GID) (Dittrich *et al.*, 2013) applies the Hansen–Coppens multipole model to transfer the electron density of functional groups (building blocks) across similar systems. A recent study (Treger *et al.*, 2023) demonstrated the prediction of refractive indices for metal–organic frameworks through a building block approach, reconstructing the framework from its individual moieties.

Over the years, we have accumulated extensive experience in exporting electrical properties across similar systems in small molecules (Krawczuk *et al.*, 2014; Ligorio *et al.*, 2021; Dos Santos *et al.*, 2015; Dos Santos & Macchi, 2016; Jabluszewska *et al.*, 2020; Rodrigues *et al.*, 2023). Building on this expertise, a general database for polarizabilities was introduced by Ernst *et al.* (2019), based on the concept of atoms in molecules proposed by Bader (Laidig & Bader, 1990) and further refined by Keith (2007). This database concept has been expanded (Ligorio *et al.*, 2022; Ligorio *et al.*, 2023) and we recently launched GruPol, which focuses on the electrical properties of proteins (Ligorio *et al.*, 2024), consisting of the 20 most common amino acid residues. Our approach is not

limited to polarizability; we have also incorporated group dipole moments and electrostatic potentials into the database. Additionally, GruPol employs a dipole interaction model (Applequist, 1977) to predict changes in these properties due to solvation, acknowledging the biological relevance of these interactions. GruPol focuses on predicting the properties of large molecules, as demonstrated in Fig. 1, where the apoprotein glutamine amidotransferase is used as an example due to its substantial size, consisting of four chains, nearly 2000 residues and approximately 30000 atoms. The molecule's four chains are shown, each represented in a different colour. The dipole moment of each chain is illustrated with an arrow, while each building block is represented by its own polarizability tensor.

Due to the possibility of multiple protonation states in large proteins, resulting from the numerous ionizable residues, GruPol also incorporates a scheme to predict changes in dipole moments when simulations at different pH levels are required. Now, given the biological significance of neighbouring ions in the vicinity of the protein's backbone, either altering the protein structure (Baldauf *et al.*, 2013) or changing its charge distribution (Lindman *et al.*, 2006), it seemed natural to extend our model to include their contribution to the electrical properties. Therefore, this paper focuses on benchmarking our latest approach to account for ions interacting with the protein and how they influence its dipole moment and polarizability. However, due to the size of glutamine amidotransferase, performing quantum calculations is impractical. For the purpose of guaranteeing good benchmarking, we opted for a smaller peptide, *i.e.* glucagon with 29 residues, as our test system. Its relatively small size allows for feasible *ab*

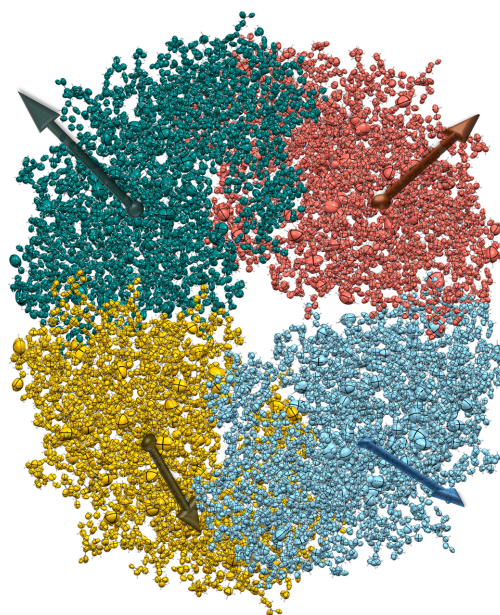


Figure 1 Building block polarizabilities and chain dipole moments for all four segments of glutamine amidotransferase (PDB refcode 1ao8). These properties are visualized with their original directions but with reduced magnitudes to fit the figure. Details and numerical values are provided in the supporting information.

initio quantum mechanical calculations, enabling us to validate the new GruPol approach. As a further step, we simulated the same peptide in two distinct aqueous environments, with and without the presence of ions, thus allowing us to evaluate the changes in glucagon's electrical properties due to a solvated medium.

2. Methodology

We utilized the glucagon peptide to investigate the presence of ions and their influence on the electrostatic properties of proteins, such as dipole moments and dipole polarizabilities. This peptide comprises 29 amino acid residues, including eight primarily ionizable groups: three ASP residues, two ARG residues, one LYS residue, and the terminal groups CTER ($-\text{COO}^-$) and NTER ($-\text{NH}_3^+$). Atomic coordinates of the peptide without additional ions were obtained from PDB entry 1gcn (Sasaki *et al.*, 1975) and kept frozen during the calculations. Missing H atoms were added using the *CHARMM-GUI* feature (Jo *et al.*, 2008). To maintain neutrality, Na^+ and Cl^- ions were added in pairs along the protein's backbone, coordinating to the ionizable residues, meaning that there is no excess of chlorine ions over sodium ions, or *vice versa*. All 70 possible configurations were taken into account, considering the complementarity between positive and negative charges: 16 with one pair, 36 with two pairs, 16 with three pairs and one with zero or four pairs. The properties of the ion-free compound were calculated assuming the terminal zwitterion state for which the database's entries were created. Importantly, for all calculations, except the case without NaCl, all ionizable residues, when not coordinated to an ion, were considered charged, meaning that acidic residues (ASP and CTER) were deprotonated while basic ones (ARG, LYS and NTER) were protonated. Maintaining overall molecular neutrality is crucial as it eliminates the origin dependence of the dipole moment.

We named the compounds containing up to three pairs of ions using a six-digit code: A, B, C and D for Na^+ coordinated to the ASP and CTER residues, and X, Y, Z and W for Cl^- coordinated to LYS, ARG and NTER. Fig. 2 illustrates the positions of each ionizable group. For instance, if two pairs of ions were coordinated at residues A, C, Z and W, the code would be AC-ZW-. For the case without ions, we used the name 'standard', and when four ions were present the name was ABCDXYZW. Each compound was assigned a number: 1 for the standard, 2–17 for compounds with one NaCl pair, 18–53 for those with two pairs, 54–69 for those with three pairs and 70 for ABCDXYZW. The association of the symbols and numbering is provided in the supporting information.

Ab initio calculations were conducted to obtain molecular dipole moments and polarizabilities. The results were then benchmarked against the GruPol database (Ligorio *et al.*, 2024). This first step was undertaken to validate the new approach used by GruPol to account for the presence of the ions on the protein's backbone. Additionally, to investigate the impact of solvation on the above-mentioned properties, glucagon was placed in a water box under two conditions: one

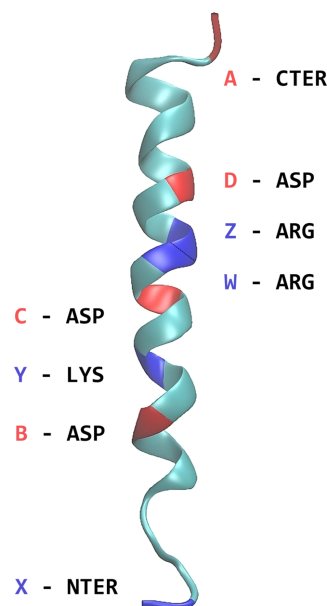


Figure 2

Representation of the glucagon molecule, highlighting the position of each ionizable residue. Negative residues are designated as A, B, C and D and are marked in red, while positive ones are labelled as X, Y, Z and W and indicated in blue.

without salt and the other with NaCl, where the ions were placed randomly. Further molecular dynamics simulations were performed to grasp the inherent flexibility of the macromolecule and its impact on the studied properties, as well as for properly coordinating the ions to the peptide, since they were initially randomly distributed in the water box.

2.1. *Ab initio* quantum calculations

Ab initio quantum calculations were performed using the *GAUSSIAN16* software (Frisch *et al.*, 2016) at the M06-HF/cc-pVDZ level of theory. This choice of the density functional theory (DFT) functional was based on our previous research, where we benchmarked DFT functionals against coupled-cluster singles or doubles (CCSD) calculations for the atomic and molecular polarizabilities of organic molecules (Ligorio *et al.*, 2020). The basis set selection was influenced by the size of the peptide. While the aug-cc-pVDZ basis set typically offers more reliable results due to its extended spatial coverage, thus increasing the polarizability by nearly 20% compared with the cc-pVDZ basis set, convergence issues with the large number of atoms in glucagon necessitated the use of the non-augmented version. Importantly, at least for smaller peptides, the differences in dipole moments between the two basis sets seem to be minimal (Ligorio *et al.*, 2024).

2.2. Including charges in GruPol predictions

The charge q of a specific ionizable residue was determined using the Henderson–Hasselbalch model, described by equation (1) for acidic groups and equation (2) for alkaline residues:

Table 1

pK_a values used to calculate charges of ionizable residues.

CTER and NTER refer to the terminal COO^- and NH_3^+ groups, respectively.

Group	Charge	pK_a
GLU/ASP	–	4.4
CYS	–	8.5
TYR	–	10.1
ARG	+	12.5
LYS	+	10.6
HIS	+	6.6
CTER	+	4.0
NTER	–	8.0

$$q_{\text{neg}} = \frac{10^{\text{pH}-pK_a}}{1 - 10^{\text{pH}-pK_a}}, \quad (1)$$

$$q_{\text{pos}} = \frac{1}{1 - 10^{\text{pH}-pK_a}}. \quad (2)$$

Charges are applied to the centre of ‘mass’ (here, masses are replaced by atomic numbers) of the individual ionizable building block, rather than to the centre of the entire residue. The values of pK_a (Mellor *et al.*, 2011) used are given in Table 1.

It is important to note that the terminal groups show a charge with a counter-intuitive sign because they were previously classified as charged groups when the database entries were created. For example, one might expect the carboxylic CTER group to carry a negative charge. However, a positive charge is added because the database creation process has already accounted for the charged group. For these groups, we calculate the expected charge at a given pH and make adjustments based on an initial zwitterion model with a +1 charge for the NTER group and –1 for the CTER group. Thus, if the NTER group has an effective charge of +0.6 at a certain pK_a , then instead of the original +1 we apply a compensating charge of –0.4 to achieve the intended charge dipole moment. Benchmarking against *ab initio* quantum calculations was performed by setting the pH to 7.3, which is the isoelectric point for glucagon as estimated by GruPol and which aligns well with the already reported literature value (Joshi *et al.*, 2000). At this pH, the charges assigned to ASP, LYS and ARG residues, and to the terminal groups, showed magnitudes close to 1. The other ionizable residues indicated in Table 1 presented minimum charges, apart from a single HIS residue with charge of approximately +0.3. It is important to note that, in the standard GruPol calculation, where no ions were included, charges are not added onto the protein backbone. This is equivalent to simulating the terminal zwitterion species. This was done to assess the intrinsic deviation of GruPol in estimating the properties of glucagon before applying the new salting in or salting out approach.

2.2.1. Property corrections

The arrangement of charges along the protein backbone produces a dipole moment referred to as the charge dipole moment. The dipole moment arising from the polarization of electron density due to chemical bonding is known as the core

dipole moment. The total dipole moment of the protein is the summation of these two components, the charge dipole moment and the core one,

$$\boldsymbol{\mu}_{\text{total}} = \boldsymbol{\mu}_{\text{core}} + \boldsymbol{\mu}_{\text{charge}}, \quad (3)$$

where $\boldsymbol{\mu}_{\text{charge}}$ is given by

$$\boldsymbol{\mu}_{\text{charge}} = \frac{1}{2} (|q_{\text{pos}}| + |q_{\text{neg}}|) \mathbf{R}. \quad (4)$$

Here, \mathbf{R} is a vector with its origin at the centre of the negative charges pointing to the centre of positive charges, given by

$$\mathbf{R} = \frac{1}{q_{\text{pos},\text{total}}} \sum |q_{\text{pos}}| \mathbf{r}_{\Lambda} - \frac{1}{q_{\text{neg},\text{total}}} \sum |q_{\text{neg}}| \mathbf{r}_{\Omega}, \quad (5)$$

and \mathbf{r} denotes the position of the centre of charge of each building block Λ or Ω that possesses a positive or negative charge, respectively.

The core dipole moment is adjusted by accounting for the electric field \mathbf{F} originating in the presence of these charges, where q represents either a positive or a negative charge,

$$\mathbf{F} = \frac{q}{|r|^2} \hat{r}, \quad (6)$$

and $|r|$ is the modulus of the vector connecting a given charge to a particular building block. \hat{r} is a unitary vector with its origin at the charge, pointing to a given building block Ω . The total electric field experienced by a building block is the summation of each individual electric field \mathbf{F} . The correction on the core dipole moment is then given by

$$\boldsymbol{\mu}_{\text{core},\Omega} = \boldsymbol{\mu}_{\Omega}^0 + \boldsymbol{\alpha}_{\Omega} \sum \mathbf{F}, \quad (7)$$

where $\boldsymbol{\alpha}_{\Omega}$ is the polarizability of a given building block Ω , and the superscript 0 indicates the initial dipole moment in the absence of any correction.

Note that, as reported in our previous studies (Jabluszewska *et al.*, 2020; Ligorio *et al.*, 2024), the presence of charges on the protein’s backbone does not significantly impact either the molecular polarizability or the corresponding values of its constituent building blocks, at least for small peptides. For this reason, the presence of charges on ionizable residues has no impact on GruPol’s polarizabilities.

2.2.2. Presence of ions

To identify ions linked to ionizable residues, GruPol uses a distance threshold of 2.5 Å for chloride ions binding to LYS and ARG residues and the NTER group (H–Cl bond distance), and 3.5 Å for sodium ions associated with ASP, GLU and the CTER group (O–Na bond distance). When an ion is detected, the corresponding charged residue becomes neutral and no longer contributes to the charge dipole moment [equation (3)]. However, both the ionic charge and the residue charge continue to influence the core dipole moments of the other building blocks. In the second step, after correcting the core dipole moments due to the presence of charges on the protein’s backbone, a dipole interaction model (ADIM) (Appelquist, 1977; Thole, 1981; Ligorio *et al.*, 2021) is employed to correct both polarizabilities and dipole moments

due to the presence of the ions. The model assumes that the total electric field within each atomic basin is the result of both the external field and the field generated by the collective dipoles from neighbouring sites (Guillaume & Champagne, 2005; Mkhadmeh *et al.*, 2009),

$$\mathbf{F} = \mathbf{F}^{\text{ext}} - \sum_{\Lambda \neq \Omega} \mathbf{T}_{\Omega\Lambda} \boldsymbol{\mu}_{\Lambda}. \quad (8)$$

$\mathbf{T}_{\Omega\Lambda}$ represents the field tensor between the atomic basins Ω and Λ ,

$$\mathbf{T}_{\Omega\Lambda} = \frac{-3}{r_{\Omega\Lambda}^5} \begin{pmatrix} x_{\Omega\Lambda}^2 - \frac{r_{\Omega\Lambda}^2}{3} & x_{\Omega\Lambda} y_{\Omega\Lambda} & x_{\Omega\Lambda} z_{\Omega\Lambda} \\ x_{\Omega\Lambda} y_{\Omega\Lambda} & y_{\Omega\Lambda}^2 - \frac{r_{\Omega\Lambda}^2}{3} & y_{\Omega\Lambda} z_{\Omega\Lambda} \\ x_{\Omega\Lambda} z_{\Omega\Lambda} & y_{\Omega\Lambda} z_{\Omega\Lambda} & z_{\Omega\Lambda}^2 - \frac{r_{\Omega\Lambda}^2}{3} \end{pmatrix}. \quad (9)$$

$x_{\Omega\Lambda} = (x_{\Omega} - x_{\Lambda})$ represents the difference in the Cartesian x coordinate between the basins Ω and Λ , and $r_{\Omega\Lambda}$ denotes the corresponding interatomic distance. Using the total electric field and the original polarizability tensors, we can recompute the dipole moments,

$$\boldsymbol{\mu}_{\Omega} = \boldsymbol{\mu}_{\Omega}^0 + \boldsymbol{\alpha}_{\Omega} \left(\mathbf{F}^{\text{ext}} - \sum_{\Lambda \neq \Omega} \mathbf{T}_{\Omega\Lambda} \boldsymbol{\mu}_{\Lambda} \right). \quad (10)$$

Finally, corrections on polarizabilities can be obtained by rearranging equation (10) in the form

$$\begin{pmatrix} \boldsymbol{\mu}_{\Omega} \\ \boldsymbol{\mu}_{\Lambda} \\ \vdots \\ \boldsymbol{\mu}_{\Gamma} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_{\Omega}^{-1} & \mathbf{T}_{\Omega\Lambda} & \cdots & \mathbf{T}_{\Omega\Gamma} \\ \mathbf{T}_{\Lambda\Omega} & \boldsymbol{\alpha}_{\Lambda}^{-1} & \cdots & \mathbf{T}_{\Lambda\Gamma} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}_{\Gamma\Omega} & \mathbf{T}_{\Gamma\Lambda} & \cdots & \boldsymbol{\alpha}_{\Gamma}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{F}^{\text{ext}} \\ \mathbf{F}^{\text{ext}} \\ \vdots \\ \mathbf{F}^{\text{ext}} \end{pmatrix}. \quad (11)$$

Summation of each row of equation (11) provides the corrected building block polarizabilities.

The close proximity of two polarizable sites can result in unrealistic values of dipole moments and polarizabilities. This occurrence is known as polarization catastrophe and was addressed by Thole (1981). To address this issue, a damping function is applied to the distance tensor, which here takes the form

$$\tau_{\Omega\Lambda} = 1 - \left(1 + \frac{s_{\Omega\Lambda} r_{\Omega\Lambda}}{2} \right) \exp(-s_{\Omega\Lambda} r_{\Omega\Lambda}),$$

$$s_{\Omega\Lambda} = \frac{1}{b(\alpha_{\text{iso},\Omega}, \alpha_{\text{iso},\Lambda})^{1/6}}. \quad (12)$$

In this context, the function $\tau_{\Omega\Lambda}$ scales the distance tensor $\mathbf{T}_{\Omega\Lambda}$ to decrease its magnitude at short distances. The coefficient b is an adjustable parameter chosen to match results from *ab initio* quantum calculations or experimental data, with a commonly used value of 2.6 (Lemkul *et al.*, 2016; Litman *et al.*, 2022), which was employed in this work as well. It is important to note that building blocks within the same molecule, such as the glucagon peptide, do not interact with each other, as these interactions were already accounted for during the development of the database itself. Therefore, changes in the dipole moment are exclusively attributed to environmental effects, which may or may not involve ionic solutions.

2.3. Molecular dynamics simulations

An initial cubic water box with a side length of approximately 84 Å was placed around the glucagon molecule under two different conditions. In the first setup, only the peptide and water molecules were included. In the second setup, NaCl units were also added, with an initial concentration of 3 mol L⁻¹. After simulation, the final concentration of NaCl was found to be around 4.2 mol L⁻¹.

The protein backbone was chosen in such a way to ensure its neutral terminal zwitterion state, as well as charged ASP, LYS and ARG residues, thus simulating the isoelectric point. An equilibration step was performed in the NVT ensemble, comprising 125 000 steps with a time interval of 1 fs. Temperature control was achieved using the Nosé–Hoover (Hoover, 1985) thermostat set to 303.15 K. Following this, molecular dynamics simulations were performed in the NPT ensemble, controlling both temperature and pressure at 303.15 K and 1 atm, respectively (Berendsen *et al.*, 1984). The simulation was carried out for a total of 200 000 molecular frames with a time interval of 2 fs, corresponding to an overall simulation time of 400 ps. Geometries were extracted every 2 ps. The first 100 conformers were discarded to ensure proper volume accommodation after the equilibration steps. The 100 final geometries were used as input for GruPol database. All molecular dynamics simulations were performed utilizing the CHARMM additive force field (Version 46b1; Brooks *et al.*, 2009), whereas input files and water box placement were done employing the CHARMM-GUI feature (Jo *et al.*, 2008). CHARMM was chosen for its dedicated focus on macromolecules, making it particularly suitable for this study, as well as for its accessibility, being freely available in its basic form.

Corrections due to the chemical environment were made using the dipole interaction model described earlier. To achieve a more isotropic environment for GruPol calculations, a cutoff of 8 Å was applied. Essentially, this process involves creating an ellipsoid around the peptide, shrinking it to the smallest acceptable size that still encompasses the entire protein, and then enlarging the ellipsoid by the specified cutoff distance. For a detailed description of this approach, the reader is referred to our latest paper on GruPol functionalities (Ligorio *et al.*, 2024). Note that all the ions were kept after cutting the water box, in order to ensure proper charge balance in the entire system. Non-bonded ions exert a smaller influence due to the weaker electric field they produce on the protein, given their greater distance. This effect cannot be neglected and is indeed considered in GruPol, especially because other negative and positive sites are present. For example, even if the ions are not bound, such as when a sodium ion is near the oxygen atom of a peptide bond, they still have a potentially significant impact.

3. Results and discussion

3.1. *Ab initio* versus GruPol

Dipole moments calculated using GruPol closely align with those obtained from M06-HF/cc-pVDZ, with an average

deviation of 15% in magnitude and 10° in angular direction, as shown in Fig. 3(a). This deviation is similar to the error observed with GruPol when previously benchmarking the database without ions (Ligorio *et al.*, 2024), indicating that the proposed model effectively corrects dipole moments influenced by ion presence. Fig. 3(b) depicts the significant fluctuations in dipole moments that are observed when ions are present, which are naturally dependent on their positions along the protein backbone. It is noteworthy that the average dipole moment, as shown in Fig. 3(c), systematically decreases with an increasing number of ions in the protein backbone. This trend can be attributed to the overall reduction in the number of charged sites and a more homogeneous distribution of charges around the protein backbone. Detailed numerical values for dipole moments and polarizabilities across all 70 configurations are provided in the supporting information, along with the corresponding numbering and naming based on ion positions.

Although the presence of ions bound to a given ionizable residue eliminates its contribution to the charge dipole moment, the significance of correcting the core dipole with these charges is demonstrated in Table 2. As shown, the salt BCDYZW (number 69), which includes three pairs of NaCl ions not associated with terminal groups, exhibits similar dipole moments to standard calculations involving only the terminal zwitterion state. This observation suggests that the presence of ions induces changes in the properties similar to hydrogen atoms in acidic groups, or neutralizes the extra hydrogen atom in basic groups. Despite the comparable dipole moments between the two scenarios (ions and H atoms), *ab initio* calculations show that the ion-containing species possess a slightly higher μ . This difference could also be predicted with GruPol. The core dipole correction is responsible for this effect, as simply removing the charges of the ionizable residues from μ_{charge} without taking into account μ_{core} would result in the same dipole moment as the standard calculation. GruPol was not only effective in correcting the magnitude of the dipole moment but also all of its components.

Before discussing the performance of the database to predict polarizabilities, it is important to highlight that GruPol

Table 2

Comparison of dipole moments obtained using GruPol and *ab initio* quantum calculations for peptide 69, BCDYZW (which includes three pairs of coordinated ions, excluding the terminal groups), with those from a standard GruPol calculation, peptide 1.

Results are given in atomic units (a.u.).

	GruPol				M06-HF/cc-pVDZ			
	μ_x	μ_y	μ_z	$ \mu $	μ_x	μ_y	μ_z	$ \mu $
BCDYZW (69)	106.9	12.9	-4.9	107.8	123.7	22.3	-1.7	125.7
Standard (1)	105.3	11.1	-5.8	106.0	123.4	20.6	-4.6	125.2

was built using the M06-HF/aug-cc-pVDZ level of theory. Unfortunately, despite extensive efforts, we were unable to achieve SCF convergence for any of the peptides with this basis set, which is more accurate than its non-augmented counterpart, particularly for evaluating polarizabilities (Ligorio *et al.*, 2020). Yet, even with the less demanding cc-pVDZ basis set, convergence issues persisted for three peptides: B--Z--, B--W-- and BC-ZW-. To overcome these limitations, the YQC approach, which is implemented in GAUSSIAN16, was chosen. In short, the YQC method improves SCF convergence by starting with steepest descent steps and then switching to the regular SCF method, only using a more complex quadratic approach if needed. It is slower but more reliable than standard methods, especially for large molecules. Although the dipole moments obtained were consistent with those from other calculations, the polarizability values were exceedingly high and unrealistic, and consequently were discarded.

As anticipated, the principal components of polarizability tensors calculated using *ab initio* quantum methods exhibit values up to nearly 30% lower than those obtained with GruPol. However, a linear increase in these components could be predicted with the addition of ions, as illustrated in Fig. 4. While the α_{11} and α_{22} components show average deviations of approximately 20–25%, the α_{33} component exhibits a significantly lower deviation of 6–7%. This can be attributed to the approximate alignment of the α_{33} component with the molecular axis of glucagon. The superposition of the atomic basis set in this direction creates an effect similar to that of diffuse

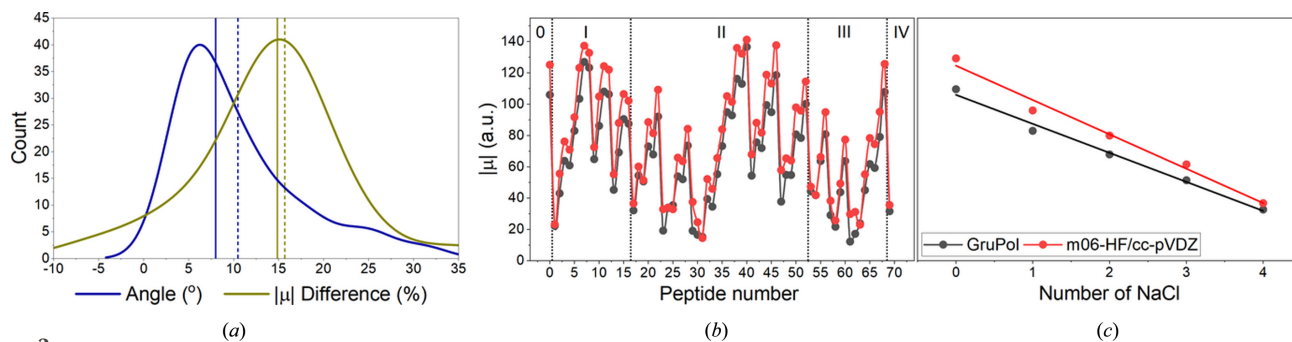


Figure 3

(a) Differences in angle and magnitude of dipole moments as predicted by GruPol and *ab initio* quantum calculations at the M06/cc-pVDZ level. The vertical solid lines represent the median deviations, while the dashed lines show the mean deviations. (b) Individual dipole moment magnitudes for each of the 70 calculations, varying from zero to four NaCl units within the protein backbone. (c) Average dipole moments as a function of the number of ions, with solid lines indicating the linear regression. The case involving zero ions bound was calculated assuming the terminal zwitterion state, where all other ionizable residues are in their neutral form.

functions. In contrast, components perpendicular to the helix are considerably lower due to the limited spatial coverage resulting from the absence of diffuse functions, as seen in Fig. 5.

3.2. Molecular dynamics simulation

To evaluate the properties obtained after molecular dynamics simulation, we first calculated polarizabilities and dipole moments for the different conformers of the protein without using ADIM and then employing the dipole interaction model. This approach enabled us to isolate and understand the impact of intrinsic geometric variations on the electric properties, since the chemical environment is responsible for changing both the properties themselves and the molecular geometries. When evaluating the influence of the chemical environment in the ion-free simulation, the application of the interaction model led to an increase in dipole moments, as shown in Fig. 6(a). Note that, throughout the entire simulation, the curves with and without usage of the interaction model (black and red, respectively) can be almost superimposed, differing by a translation of 5–10 a.u.

The presence of ions decreased μ , either when ADIM was utilized or not, with the former situation presenting significantly lower values. These findings are consistent with the results visualized in Fig. 3(c), demonstrating that the presence of ions attached to ionizable residues may reduce the dipole moment. Interestingly, in the presence of ions, the dipole moment follows a distinct pattern: during the first half of the simulation, it was approximately 50% percent of the value observed in the latter half. This behaviour can be understood by the number of bound ions throughout the simulation, as shown in Fig. 6(d). Initially, an average of two pairs of NaCl were bound, which later decreased to one pair, corroborating

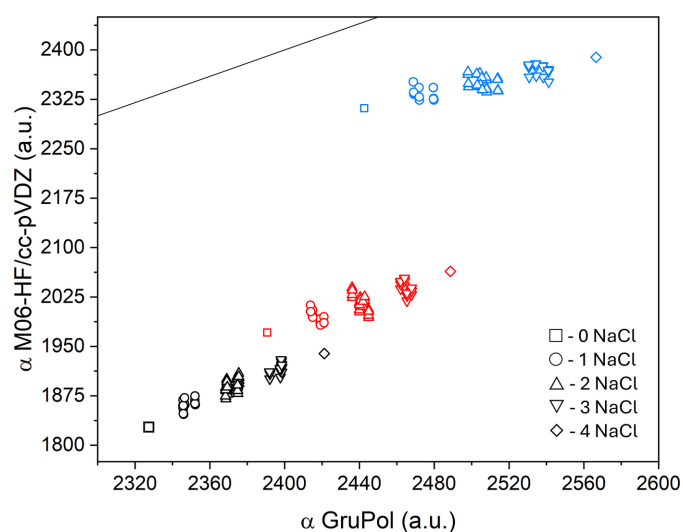


Figure 4
Comparison of the principal components of polarizabilities obtained from GruPol and *ab initio* quantum calculations at the M06-HF/cc-pVDZ level of theory. Black points represent α_{11} , red points α_{22} and blue points α_{33} . The solid line represents the ideal correspondence between the database and the quantum calculations.

the results observed previously. When analysing both curves without ADIM (red and green), a clear difference of 10–20 a.u. can be observed, suggesting that the simulations may have resulted in intrinsically distinct geometries due to the presence of ions.

Polarizability results showed that when ADIM was not employed, the isotropic values (α_{iso}) remained nearly constant, with only minor fluctuations of a few atomic units. When the chemical environment was considered, the polarizability values increased by nearly 2% in the simulation without ions. In the presence of NaCl, the inclusion of ions, particularly chloride ones, significantly increased α , with values of around 30 a.u. for each ion bound. The presence of Na^+ ions had a negligible effect on polarizability, given its isotropic value of 0.3 a.u., a consequence of the very contracted nature of its electron density. For this reason, the patterns observed in Figs. 6(b) and 6(d) are similar, particularly when comparing the blue curve in Fig. 6(b) (ADIM + ions) with the purple curve in Fig. 6(d) (number of chloride ions). In terms of polarizability anisotropy ($\Delta\alpha$), ADIM does not affect its value in the simulation with NaCl, as demonstrated in Fig. 6(c), where both curves overlap. In contrast, the simulation without ions produced higher $\Delta\alpha$ values, regardless of whether ADIM was applied. Nevertheless, the chemical environment contributed to an increase in anisotropy compared to the ADIM-free scenario.

A closer examination of the properties obtained, (i) for the molecule extracted from the experimental crystal structure (PDB 1gcn), *i.e.* the initial coordinates employed for the

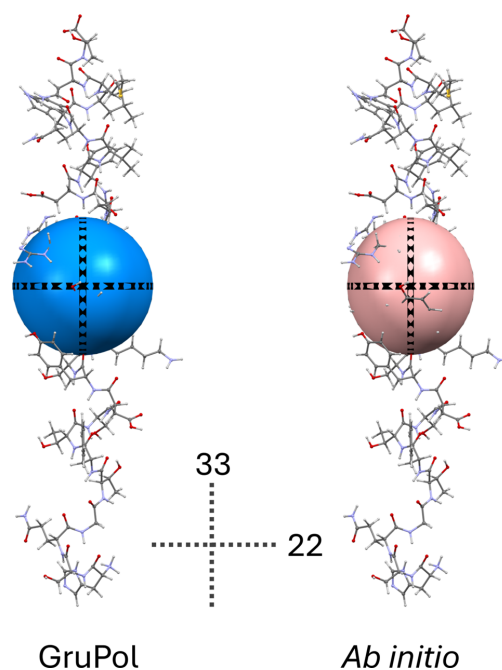


Figure 5
Molecular polarizabilities of glucagon obtained using standard GruPol and *ab initio* quantum calculations at the M06-HF/cc-pVDZ level (no ions coordinated). The polarizability tensor values have been reduced by a factor of 100. The labels 22 and 33 refer to the components of the polarizability tensor.

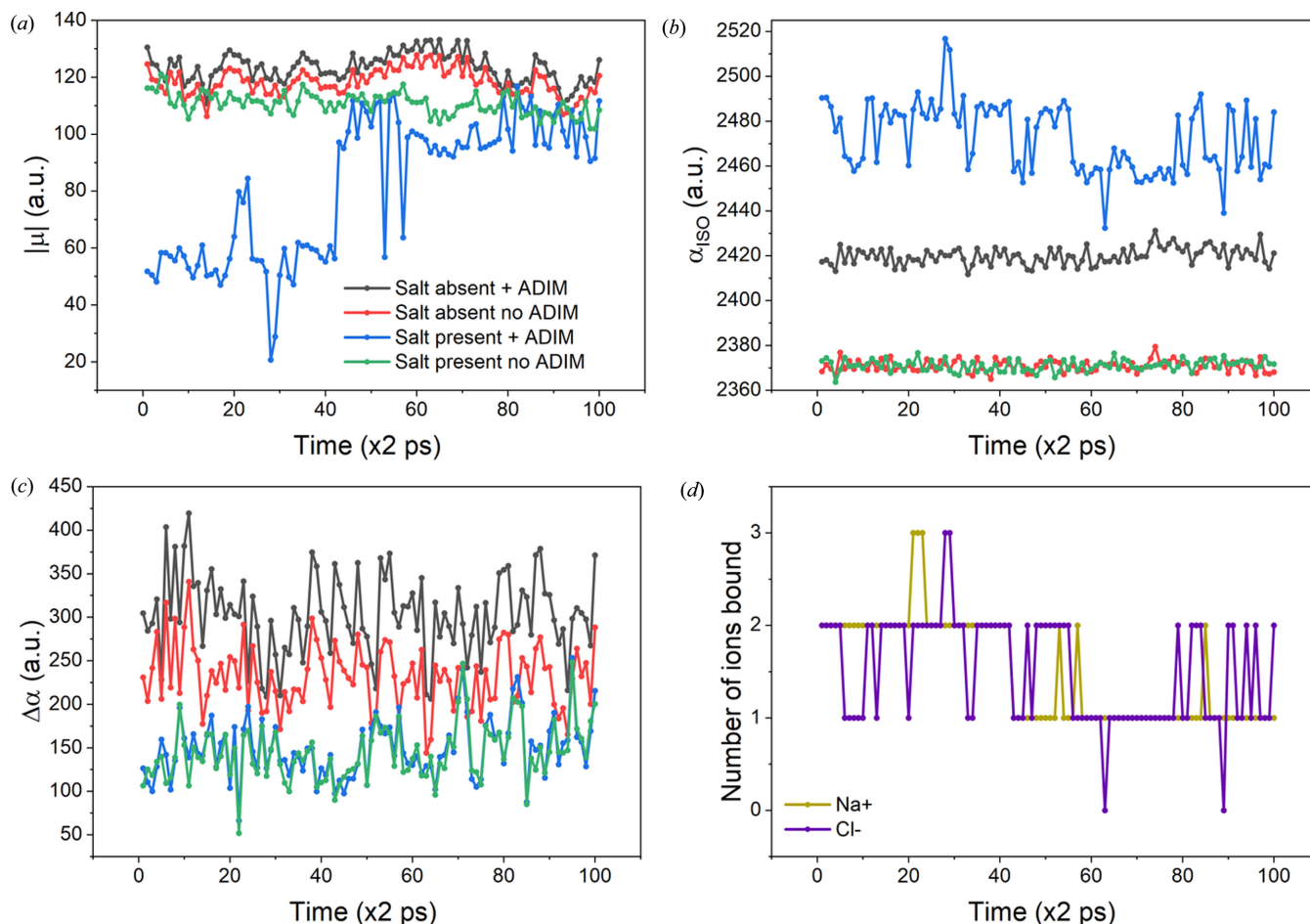


Figure 6 (a) Dipole moments, (b) isotropic polarizabilities, (c) anisotropy of polarizability and (d) number of ions coordinated to the protein’s backbone during the molecular dynamics simulation. Isotropic polarizability is defined as $\alpha_{\text{iso}} = \text{Tr}(\alpha)/3$ and anisotropy of polarizability is given by the formula $\Delta\alpha = \left(\frac{1}{2} \{3 \text{Tr}(\alpha^2) - [\text{Tr}(\alpha)]^2\}\right)^{1/2}$.

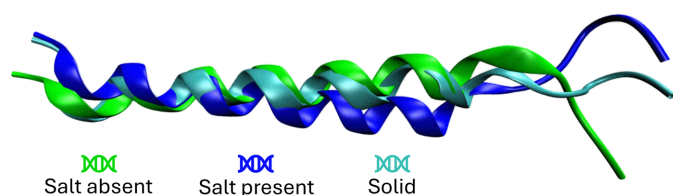


Figure 7 Representations of the glucagon molecule obtained at frame 29 from both molecular simulations and the solid-state structure.

molecular dynamics simulations, and (ii) those derived from the simulations themselves, reveals that the presence of salt results in properties closely resembling those of the solid-state geometry. Fig. 7 provides a comparison of geometries across the three possible scenarios, with the snapshots taken from the simulation at 58 ps, corresponding to frame 29. The comparison of properties was conducted without the use of ADIM, focusing solely on the effects of geometric changes, exhibiting less fluctuation throughout the dynamics. For this reason, neither crystal field effects nor solvent were considered. Frame 29 was selected since at this frame, in the three

Table 3 Polarizabilities and dipole moments calculated for glucagon using the solid-state geometry and at frame 29 (58 ps) extracted from simulations conducted both with and without the presence of NaCl.

In this analysis, the dipole interaction model was not applied, allowing us to focus on the changes in these properties resulting solely from intrinsic geometric variations across these three scenarios. Despite the absence of environmental interactions via ADIM, charges were included on the protein backbone, assuming a pH of 7.3.

	α_{iso}	$\Delta\alpha$	$ \mu $
Salt absent	2369.2	237.3	117.2
Salt present	2368.8	148.4	111.5
Solid	2386.9	102.0	94.3

scenarios, the properties were close to the average value across the entire simulation. These results indicate that the presence of salt stabilizes the protein, making it less susceptible to geometric changes during the simulation. Numerical values of the properties for the entire simulation are given in the supporting information; data only for one simulation step, frame 29, are given in Table 3.

4. Conclusions

Electrical charges on a protein backbone play a crucial role in accurately determining its electrical and optical properties, such as dipole moments and polarizabilities. For this reason, the GruPol database was developed with the goal of providing fast yet accurate predictions of these properties. The database enables the inclusion of charges in the backbone by selecting a specific pH, which then assigns charges to ionizable residues. In the presented update of the database, we have incorporated the option to include ions on the protein backbone, either as a salt in the solid state or within an aqueous environment.

To demonstrate the effects of salt and/or electrical charges in general, we selected the glucagon molecule and benchmarked its dipole moments and polarizabilities against *ab initio* quantum calculations, accounting for 70 different possibilities for ion arrangement along the protein's backbone. We have showed that GruPol yields small errors for the dipole moment, averaging 15% in magnitude and 10° in angular direction. These error values are comparable with those obtained when benchmarking GruPol without ions or charges on the protein backbone. We observed that, in general, the more ions are linked to the protein, the lower its dipole moment tends to be, depending naturally on their specific positions on the protein structure. In the case of polarizability behaviour, GruPol exhibits a linear correlation with quantum calculations, although the database tends to result in higher values, primarily due to the absence of diffuse basis sets in the present quantum calculations. This discrepancy underscores the need to use the database when accurate properties are required, especially given the size limitations of proteins.

The second step of this study involved solvating the protein in two distinct aqueous media: one containing only the protein and water molecules, and the other including including Na⁺ and Cl⁻ ions as well. We observed that the inclusion of ions consistently reduced the overall dipole moment of the protein throughout the simulation, as well as the anisotropy of the polarizability. The reduced dipole moment can be attributed to two factors: (i) intrinsic geometric changes induced by the medium, and (ii) the presence of ions bound to the protein, which decreases the overall charge on the backbone. In simulations with ions present, the observed properties were more closely aligned with those from the solid-state structure, suggesting that the geometry was more stable and less prone to changes in the presence of salt.

Acknowledgements

Open access funding enabled and organized by Projekt DEAL.

Funding information

We gratefully acknowledge the Polish high-performance computing infrastructure PLGrid (HPC Centre: ACK Cyfronet AGH) for providing computer facilities and support within computational grant No. PLG/2024/017677 to Anna

Krawczuk. Leonardo H. R. Dos Santos acknowledges financial support from the Brazilian agencies FAPEMIG (project No. APQ-01465-21) and CNPq (project No. 305330/2023-3).

References

- Applequist, J. (1977). *Acc. Chem. Res.* **10**, 79–85.
- Baldauf, C., Pagel, K., Warnke, S., von Helden, G., Kokscho, B., Blum, V. & Scheffler, M. (2013). *Chem. A Eur. J.* **19**, 11224–11234.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). *J. Chem. Phys.* **81**, 3684–3690.
- Brooks, B. R. III, Brooks, C. L. III, Mackerell, A. D. Jr, Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caffisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M. & Karplus, M. (2009). *J. Comput. Chem.* **30**, 1545–1614.
- Dittrich, B., Hübschle, C. B., Pröpper, K., Dietrich, F., Stolper, T. & Holstein, J. J. (2013). *Acta Cryst.* **B69**, 91–104.
- Domagała, S. & Jelsch, C. (2008). *J. Appl. Cryst.* **41**, 1140–1149.
- Dos Santos, L. H. R., Krawczuk, A. & Macchi, P. (2015). *J. Phys. Chem. A*, **119**, 3285–3298.
- Dos Santos, L. H. R. & Macchi, P. (2016). *Crystals*, **6**, 43.
- Duan, C. & Wang, R. (2024). *ACS Cent. Sci.* **10**, 460–468.
- Ernst, M., Dos Santos, L. H. R., Krawczuk, A. & Macchi, P. (2019). *Understanding Intermolecular Interactions in the Solid State: Approaches and Techniques*, edited by D. Chopra, ch. 7, pp. 211–242. London: The Royal Society of Chemistry.
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, J. A. Jr, Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B. & Fox, D. J. (2016). *GAUSSIAN 16*, Revision C. 01. Gaussian Inc., Wallingford, Connecticut, USA.
- Gebauer, F., Schwarzl, T., Valcárcel, J. & Hentze, M. W. (2021). *Nat. Rev. Genet.* **22**, 185–198.
- Goel, H., Yu, W., Ustach, V. D., Aytenfisu, A. H., Sun, D. & Mackerell, A. D. (2020). *Phys. Chem. Chem. Phys.* **22**, 6848–6860.
- Gonzalez, L. L., Garrie, K. & Turner, M. D. (2020). *Biochim. Biophys. Acta*, **1867**, 118677.
- Guillaume, M. & Champagne, B. (2005). *Phys. Chem. Chem. Phys.* **7**, 3284–3289.
- Hoover, W. G. (1985). *Phys. Rev. A*, **31**, 1695–1697.
- Hu, C., Yang, J., Qi, Z., Wu, H., Wang, B., Zou, F., Mei, H., Liu, J., Wang, W. & Liu, Q. (2022). *MedComm*, **3**, e161.
- Jabłuszczyńska, A., Krawczuk, A., Dos Santos, L. H. R. & Macchi, P. (2020). *ChemPhysChem*, **21**, 2155–2165.
- Jo, S., Kim, T., Iyer, V. & Im, W. (2008). *J. Comput. Chem.* **29**, 1859–1865.
- Joshi, A. B., Rus, E. & Kirsch, L. E. (2000). *Int. J. Pharm.* **203**, 115–125.

- Keith, T. A. (2007). *The Quantum Theory of Atoms in Molecules*, edited by C. F. Matta & R. J. Boyd, ch. 3, pp. 61–94. Weinheim: Wiley-VCH.
- Kim, J., Qin, S., Zhou, H.-X. & Rosen, M. K. (2024). *J. Am. Chem. Soc.* **146**, 3383–3395.
- Krawczuk, A., Pérez, D. & Macchi, P. (2014). *J. Appl. Cryst.* **47**, 1452–1458.
- Kumar, P., Gruza, B., Bojarowski, S. A. & Dominiak, P. M. (2019). *Acta Cryst.* **A75**, 398–408.
- Laidig, K. E. & Bader, R. F. W. (1990). *J. Chem. Phys.* **93**, 7213–7224.
- Lemkul, J. A., Huang, J., Roux, B. & MacKerell, A. D. Jr (2016). *Chem. Rev.* **116**, 4983–5013.
- Li, Y., Gomez-Mingot, M., Fogeron, T. & Fontecave, M. (2021). *Acc. Chem. Res.* **54**, 4250–4261.
- Ligorio, R. F., Grosskopf, P., Dos Santos, L. H. R. & Krawczuk, A. (2024). *J. Phys. Chem. B*, **128**, 7954–7965.
- Ligorio, R. F., Krawczuk, A. & Dos Santos, L. H. R. (2020). *J. Phys. Chem. A*, **124**, 10008–10018.
- Ligorio, R. F., Krawczuk, A. & Dos Santos, L. H. R. (2021). *J. Phys. Chem. A*, **125**, 4152–4159.
- Ligório, R. F., Rodrigues, J. L., Krawczuk, A. & Dos Santos, L. H. R. (2023). *J. Comput. Chem.* **44**, 745–754.
- Ligorio, R. F., Rodrigues, J. L., Zuev, A., Dos Santos, L. H. R. & Krawczuk, A. (2022). *Phys. Chem. Chem. Phys.* **24**, 29495–29504.
- Lindman, S., Xue, W., Szczepankiewicz, O., Bauer, M. C., Nilsson, H. & Linse, S. (2006). *Biophys. J.* **90**, 2911–2921.
- Litman, J. M., Liu, C. & Ren, P. (2022). *J. Chem. Inf. Model.* **62**, 79–87.
- Lovelock, S. L., Crawshaw, R., Basler, S., Levy, C., Baker, D. D. H., Hilvert, D. & Green, A. P. (2022). *Nature*, **606**, 49–58.
- Ma, C. Y., Moldovan, A. A., Maloney, A. G. & Roberts, K. J. (2023). *J. Pharm. Sci.* **112**, 435–445.
- Majeed, S., Ofek, G., Belachew, A., Huang, C. C., Zhou, T. & Kwong, P. D. (2003). *Structure*, **11**, 1061–1070.
- Martin del Campo, M., Gómez-Secundino, O., Camacho-Ruiz, R. M., Mateos Díaz, J. C., Müller-Santos, M. & Rodríguez, J. A. (2023). *Biochim. Biophys. Acta*, **1868**, 159380.
- Mellor, B. L., Khadka, S., Busath, D. D. & Mazzeo, B. A. (2011). *Protein J.* **30**, 490–498.
- Mkadmh, A. M., Hinchliffe, A. & Abu-Awwad, F. M. (2009). *J. Mol. Struct. Theochem*, **901**, 9–17.
- Resende, L. F. T. de, Basilio, F. C., Filho, P. A., Therézio, E. M., Silva, R. A., Oliveira, O. N., Marletta, A. & Campana, P. T. (2024). *Int. J. Biol. Macromol.* **259**, 129142.
- Rodrigues, J. L., Ligorio, R. F., Krawczuk, A., Diniz, R. & Dos Santos, L. H. R. (2023). *J. Mol. Model.* **29**, 49.
- Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J. & Blundell, T. (1975). *Nature*, **257**, 751–757.
- Spackman, M. A. (2018). *CrystEngComm*, **20**, 5340–5347.
- Thole, B. T. (1981). *Chem. Phys.* **59**, 341–350.
- Treger, M., König, C., Behrens, P. & Schneider, A. M. (2023). *Phys. Chem. Chem. Phys.* **25**, 19013–19023.
- Vascon, F., Gasparotto, M., Giacomello, M., Cendron, L., Bergantino, E., Filippini, F. & Righetto, I. (2020). *Comput. Struct. Biotechnol. J.* **18**, 1774–1789.