Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

# Supplementary Information

## 1 Methodology

Our approach builds upon a series of prior studies exploring quantum entanglement and electronic correlations in small biomolecular systems such as dipeptides[1–4]. In this work, we scale the framework to a full protein, employing MI and its coarse-grained form to quantify the correlation landscape within insulin.

### 1.1 Mutual Information from Reduced Density Matrices

At the core of our method is the quantification of electronic correlations using von Neumann entropy. For a bipartite quantum system described by a pure state with the Hilbert space of the subsystem and environment spanned by the states $|n\rangle$ and $|e\rangle$ respectively,

$$|\Psi\rangle = \sum_{n,e} \Psi_{n,e} |n\rangle \otimes |e\rangle, \tag{1}$$

we obtain the reduced density matrix (RDM) of the subsystem by tracing out the environment:

$$\rho = \mathrm{Tr}_{|e\rangle} \left( |\Psi\rangle\langle\Psi| \right). \tag{2}$$

From the eigenvalues $\omega_\alpha$ of this RDM, the entanglement entropy of a single orbital is calculated as:

$$s_i^{(1)} = -\sum_{\alpha=1}^{4} \omega_{\alpha,i} \ln \omega_{\alpha,i}, \tag{3}$$

while the joint entropy of a pair of orbitals with 4 possible electron configurations $\{ |-\rangle, |\uparrow\rangle, |\downarrow\rangle, |\updownarrow\rangle \}$ is defined as:

$$s_{ij}^{(2)} = -\sum_{\alpha=1}^{16} \omega_{\alpha,ij} \ln \omega_{\alpha,ij}. \tag{4}$$

These entropies are then used to compute the mutual information:

$$I_{ij} = s_{ij}^{(2)} - s_i^{(1)} - s_j^{(1)}, \tag{5}$$

which captures both quantum and classical correlations between orbitals $i$ and $j$. This provides a resolution-independent means of identifying electronically significant interactions in the molecular system.

### 1.2 From Orbital to Atomic and Residue-Level Correlations

To interpret MI in a chemically and biologically meaningful way, we coarse-grain the orbital-based values to the atomic and residue levels. The AMI is defined as the sum over all orbital pairs belonging to atoms $A$ and $B$:

$$I_{AB} \approx \sum_{i \in A} \sum_{j \in B} I_{ij}. \tag{6}$$

Further coarse-graining yields fragment mutual information (FMI), which aggregates atomic contributions over residues $X$ and $Y$:

$$I_{XY} \approx \sum_{A \in X} \sum_{B \in Y} I_{AB}. \tag{7}$$

This hierarchy, from orbital to atomic to residue scale, enables the analysis of electronic structure across multiple levels of resolution, from covalent bonds to long-range intra-protein interactions.

### 1.3 System Preparation and Fragmentation

The 3D structure of insulin (PDB ID: 3I40)[5] was obtained from the Protein Data Bank. The structure was protonated and energy minimized using the steepest descent algorithm implemented in GROMACS[6,7] with the OPLS-AA force field[8,9] and TIP3P water model. To reduce the computational cost associated with the full system (782 atoms), the protein was divided into 51 overlapping spherical fragments centered on $\alpha$-carbon atoms using a primary radius of 5.0 Å; additional fragment radii of 4.0 and 6.0 Å were also investigated to assess the trade-off between computational cost and completeness of long-range correlation recovery. All spherical cuts were capped with $NH_2$/COOH groups and subjected to position-restrained energy minimization implemented in GROMACS with the OPLS-AA force field and the TIP3P water model to satisfy valencies and remove atomic overlaps. The purpose of this procedure was to preserve the structure of each fragment as closely as possible to that of the full protein while resolving steric artifacts introduced at the cut boundaries. Each fragment was subjected to DFT calculations using ORCA 6.0.1[10] with the $\omega$B97M-V functional[11] and 6-31G(d) basis set[12]. Prior to the quantum mechanical calculations, all solvent molecules and ions were removed. AMI values were then extracted from the fragment wavefunctions. Contributions from the capping atoms were excluded before stitching overlapping fragments into the reconstructed full-protein AMI matrix.

To validate the accuracy of our divide-and-conquer approach, we conducted a full DFT calculation on the entire insulin protein without fragmenting the structure. Using the same functional and basis set, we extracted the complete wavefunction and computed the MI matrix from the full-system RDMs. This full-protein MI map served as a benchmark to evaluate the consistency and fi-

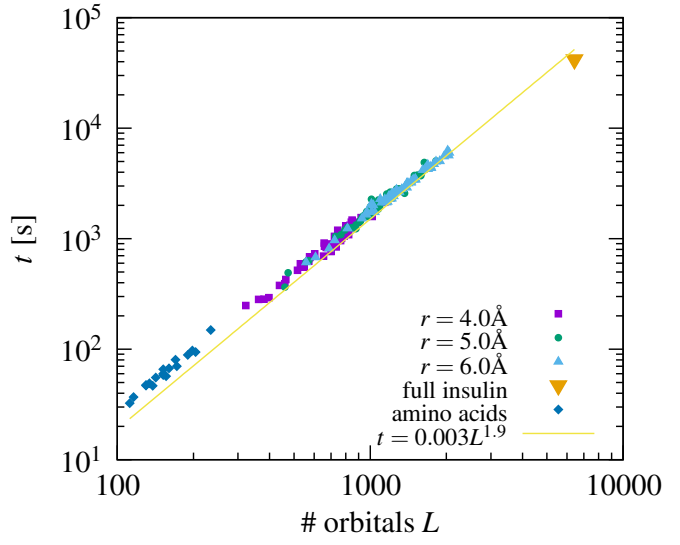


Fig. S1 CPU wall timings $t$ (measured in s) as a function of the number of orbitals $L$ in the ORCA 6.0.1 DFT computation at the $\omega$B97M-V/6-31G(d) level of theory employing 1CPU nodes.

delity of the stitched correlation data assembled from the overlapping structural cuts.

Small quantitative differences in AMI/FMI values between the stitched and full-protein matrices may arise from the position-restrained minimization procedure. For example, the Glu17–Arg22 salt bridge appears slightly stronger in the 5 and 6 Å stitched matrices than in the full-protein reference. This difference can be traced to a modest rotation of the Glu17 carboxylate side chain (approximately 19°) toward the Arg22 amino group in the corresponding spherical fragment. The RMSD between the original protein structure and the relevant spherical fragment (centered at the $C_\alpha$ of Glu17) after position-restrained minimization is 0.18 Å, indicating that these deviations are minor and localized. The average weighted RMSD values for fragment radii of 4, 5, and 6 Å are 0.18, 0.15, and 0.17 Å, respectively. For transparency, All data associated with this study including both weighted and unweighted RMSD values for all individual fragments relative to the full-protein structure, are publicly available on GitHub (https://github.com/QuNB-Repo/Insulin.git) and the UNB Dataverse (https://doi.org/10.25545/OZX7DP); the accompanying README file has been updated to document the folder structure and project tree.

We recorded computational timings on all DFT runs at the described $\omega$B97M-V/6-31G(d) level of theory[11,12] with ORCA 6.0.1[10] on 1CPU nodes on the ACENET high-performance compute cluster of the Digital Research Alliance Canada. As presented in Figure S1, we extracted timings for the 51 $r = 4.0 - 6.0$Å fragments, the full insulin protein, as well as all single 20 capped amino acids to set a reference. We observed a polynomial scaling described by the $t = 0.003L^{1.9}$ curve in the Figure, testimony of a highly performant DFT engine behind the ORCA software. A summary of the timings can be found in Table 1 of the paper, and the output files can be downloaded at the Dataverse[13]

## 2 Supplementary Figures and Tables

Fig. S2 provides SPAWN plots generated from the 5 Å stitched AMI matrix obtained through fragment-wise DFT calculations with the one derived from the full-protein DFT calculation. The figure is a complementary view of the insulin protein as Fig. 3 in the main text. Again, these plots confirm strong correlations between cysteine pairs forming disulfide bonds (Cys6–Cys11, Cys7A–Cys7B,Cys20A–Cys19B).

The table S1 lists 31 hydrogen bonds, with donor–acceptor distances between 2.7–3.3 Å, validating the signals observed in the heat map.

## Notes and references


1 K. Boguslawski and P. Tecmer, Int. J. Quantum Chem., 2015, **115**, 1289–1295.

2 K. Boguslawski, P. Tecmer, O. Legeza and M. Reiher, J. Phys. Chem. Lett., 2012, **3**, 3129–3135.

3 K. Boguslawski, P. Tecmer, G. Barcza, O. Legeza and M. Reiher, J. Chem. Theory Comput., 2013, **9**, 2959–2973.

4 M. J. Moghadam, K. Boguslawski, R. Doucet, Ö. Legeza, P. Tecmer and S. De Baerdemacker, chemrxiv., 2024.

5 V. Timofeev, R. Chuprov-Netochin, V. Samigina, V. Bezuglov, K. Miroshnikov and I. Kuranova, Struct. Biol. Crystallogr. Commun., 2010, **66**, 259–263.

6 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, Softw. X, 2015, **1**, 19–25.

7 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, J. Comput. Chem., 2005, **26**, 1701–1718.

8 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, J. Am. Chem. Soc., 1996, **118**, 11225–11236.

9 W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc., 1988, **110**, 1657–1666.

10 F. Neese, WIREs Comput. Mol. Sci., 2012, **2**, 73–78.

11 N. Mardirossian and M. Head-Gordon, J. Phys. Chem., 2016, **144**, year.

12 R. Ditchfield, W. J. Hehre and J. A. Pople, J. Chem. Phys., 1971, **54**, 724–728.

13 M. Javaheri Moghadam, R. Mulder and S. De Baerdemacker, Divide and Correlate: Mapping Electronic Correlations in Proteins via Local Cut-Wise Reconstruction, 2025, `https://doi.org/10.25545/OZX7DP`.

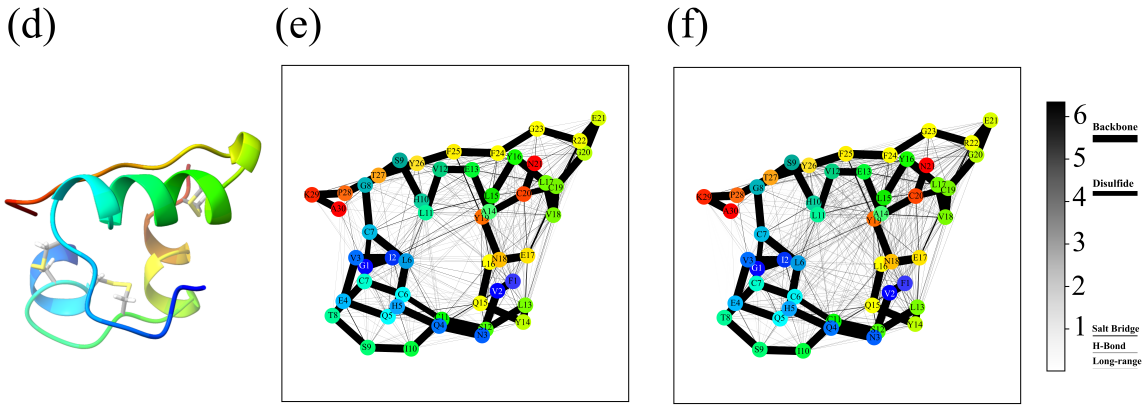


Fig. S2 Coarse-grained visualization of electronic correlations in insulin using FMI. (d) Back view of the insulin protein structure, with residues colored by sequence index using the rainbow gradient. Disulfide bonds are depicted as yellow sticks. (e) shows SPAWN plots reconstructed from 51 overlapping structural fragments using 5 Å cut-off, while (f) presents reference plots from a full-system DFT calculation. In all SPAWN plots, edges represent FMI values between residues, and edge thickness corresponds to correlation strength.

Table S1 List of 31 hydrogen bonds in the insulin protein structure. Each entry includes donor and acceptor residues and their interatomic distances in (Å).

| No. | Donor | Acceptor | Dist. (Å) | No. | Donor | Acceptor | Dist. (Å) |
|---|---|---|---|---|---|---|---|
| 1 | GLN 5 (A) | GLY 1 (A) | 3.168 | 17 | LEU 6 (B) | CYS 6 (A) | 2.723 |
| 2 | CYS 6 (A) | ILE 2 (A) | 2.828 | 18 | HIS 10 (B) | CYS 7 (B) | 3.150 |
| 3 | CYS 7 (A) | VAL 3 (A) | 2.824 | 19 | LEU 11 (B) | GLY 8 (B) | 3.132 |
| 4 | THR 8 (A) | VAL 3 (A) | 2.937 | 20 | VAL 12 (B) | GLY 8 (B) | 3.149 |
| 5 | CYS 11 (A) | GLN 4 (B) | 2.935 | 21 | GLU 13 (B) | SER 9 (B) | 2.887 |
| 6 | GLN 15 (A) | SER 12 (A) | 3.164 | 22 | ALA 14 (B) | HIS 10 (B) | 3.159 |
| 7 | LEU 16 (A) | SER 12 (A) | 2.998 | 23 | LEU 15 (B) | LEU 11 (B) | 2.977 |
| 8 | GLU 17 (A) | LEU 13 (A) | 3.003 | 24 | TYR 16 (B) | VAL 12 (B) | 2.923 |
| 9 | ASN 18 (A) | GLN 15 (A) | 3.215 | 25 | LEU 17 (B) | GLU 13 (B) | 2.974 |
| 10 | TYR 19 (A) | LEU 16 (A) | 3.001 | 26 | VAL 18 (B) | ALA 14 (B) | 2.897 |
| 11 | CYS 20 (A) | GLU 17 (A) | 3.234 | 27 | CYS 19 (B) | LEU 15 (B) | 2.915 |
| 12 | ASN 21 (A) | GLY 23 (B) | 2.968 | 28 | GLY 20 (B) | TYR 16 (B) | 3.222 |
| 13 | GLN 4 (B) | PHE 1 (B) | 2.889 | 29 | ARG 22 (B) | CYS 19 (B) | 3.256 |
| 14 | HIS 5 (B) | CYS 7 (A) | 3.112 | 30 | ARG 22 (B) | CYS 19 (B) | 3.319 |
| 15 | ARG 22 (B) | GLU 17 (A) | 3.183 | 31 | GLY 23 (B) | GLY 20 (B) | 3.122 |
| 16 | PHE 25 (B) | TYR 19 (A) | 2.998 | | | | |