ELECTRONIC SUPPLEMENTARY INFORMATION (ESI)

Computational insight into the interaction of oxaliplatin with insulin

Giuseppe Sciortino,[a,b] José-Emilio Sánchez-Aparicio,[a] Jaime Rodríguez-Guerra Pedregal,[a] Eugenio Garribba,[b] and Jean-Didier Maréchal,*[a]

[a] *Departament de Química, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallés, Barcelona, Spain*

[b] *Dipartimento di Chimica e Farmacia, Università di Sassari, Via Vienna 2, I-07100 Sassari, Italy*

Corresponding author: jeandidier.marechal@uab.cat.

INDEX

1. Supplementary details

2. Supplementary figures

3. Supplementary tables

4. References

1. Supplementary details

Metal centres parametrization

The force field building operations were carried out using the MCPB.py.[1] The MCPB.py workflow can be resumed as follows:

1) Starting from the whole structure of the protein containing the metal-complex, a small model is built. The model contains the metal, its ligands, CH_3R (R = side-chain) to represent amino acids that coordinate the metal with a side-chain donor and $CO{-}CH_3$ or $NH{-}CH_3$ to describe amino acids coordinated with a backbone donor. This small model is optimized and the equilibrium parameters and force constants of the metal center are obtained from frequency calculation.

2) From the same structure a larger model is built containing the metal, its ligands and the amino acid residues involved in the coordination capped by $CO{-}CH_3$ or $NH{-}CH_3$ terminal groups. In the case of two or more amino acids bound to the metal with less than 5 residues between them, the inner amino acids are retained and modelled as glycine. From this large model a RESP[2] fitting is performed and the charge included in the force-field.

Comparative study with respect to unbounded insulin

Aiming to make a structural comparison between the ligand-free insulin and the three binding modes of oxaliplatin, a Principal Component Analysis (PCA) of the two main components relative to each other was carried out for all the MD simulations (Figure S3). Previously, all the frames were aligned taking as reference the X-ray structure of porcine insulin (PDB 1zni[3]). Alpha carbons of the three helices and central loop were selected to perform the PCA analyses, in order to detect relevant changes in the conformation of these zones. The resulting plots were superposed and sixteen representative frames of each conformational zone were selected (provided in ESI files). Comparison of the two most relevant structures of each binding mode are also shown (Figure S3, from “a” to “f”).

Convergence of the MD simulations

Concept of convergence

From our point of view and for our scope, a MD trajectory was considered converged when a full exploration of the conformational space was achieved. In particular, a stable conformation or a pool of relative stable conformations visited for a statistically consistent number of times was considered a convergence indicator.

From this statistical point of view, it is not possible to describe if a simulation is absolutely converged.[4] However, there are several methods in the literature that, through qualitative and visual analyses, can quickly suggest that the simulation has not run enough and can help to achieve a good sampling quality.

Specifically, for each MD trajectory we applied: the traditional RMSD analysis from the initial structure searching if the plot reaches an steady state, an all-to-all RMSD analysis[4] to visually identify the number of transitions between the different sub-states (Figure S4), a cluster counting[5] to evaluate the rate of discovery of new clusters over the trajectory (Figure S6-S8), and a PCA analysis over time[4] to ensure that dynamic transitions between different conformations can occur during the MD (Figure S5),

Alpha carbons of the three helices and the central loop were selected to perform all the studies to avoid excessively flexible parts that distort the results. In the cluster counting study, RMSD metrics with three different cut-offs were used to tell whether a frame belongs to a cluster: 1.5, 2.0 and 2.5 Å (Figure S6, S7 and S8), measured from the frame centre of the cluster.

2. Supplementary figures

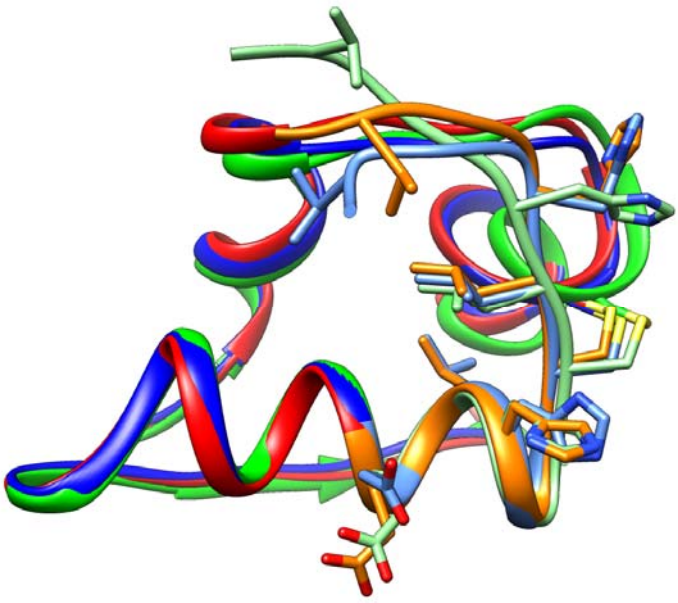

Figure S1. Superposition of the X-ray structure of porcine insulin (PDB 1zni,[3] used as starting structure for docking calculations) in red, human insulin forms (PDB 5ena[6]) in blue and human insulin (PDB 3w7y) in green, used as high-quality reference. B1 peptides are highlighted in orange, light blue and light green, respectively. The relevant binding side chains are also shown. From the comparison of the structures, which report RMSD values lower than 0.55 Å highlighting very small differences, we conclude that the structure selected in this study is a valuable model for studying the interaction with oxaliplatin. RMSD was computed with the algorithm implemented in UCSF Chimera. [7]

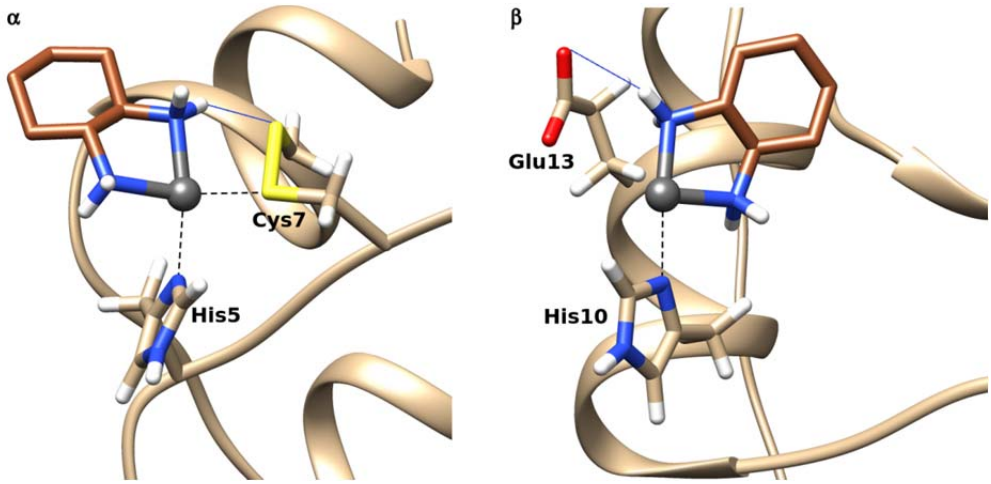

Figure S2. The first two docking proposals for the interaction of $[Pt(dach)]^{2+}$ moiety to insulin: α binding site with the coordination of $Cys7^{B}$ and $His5^{B}$ and β secondary binding site with Pt^{II} coordinated to $His10^{B}$. The hydrogen bonds are also highlighted with solid blue lines.

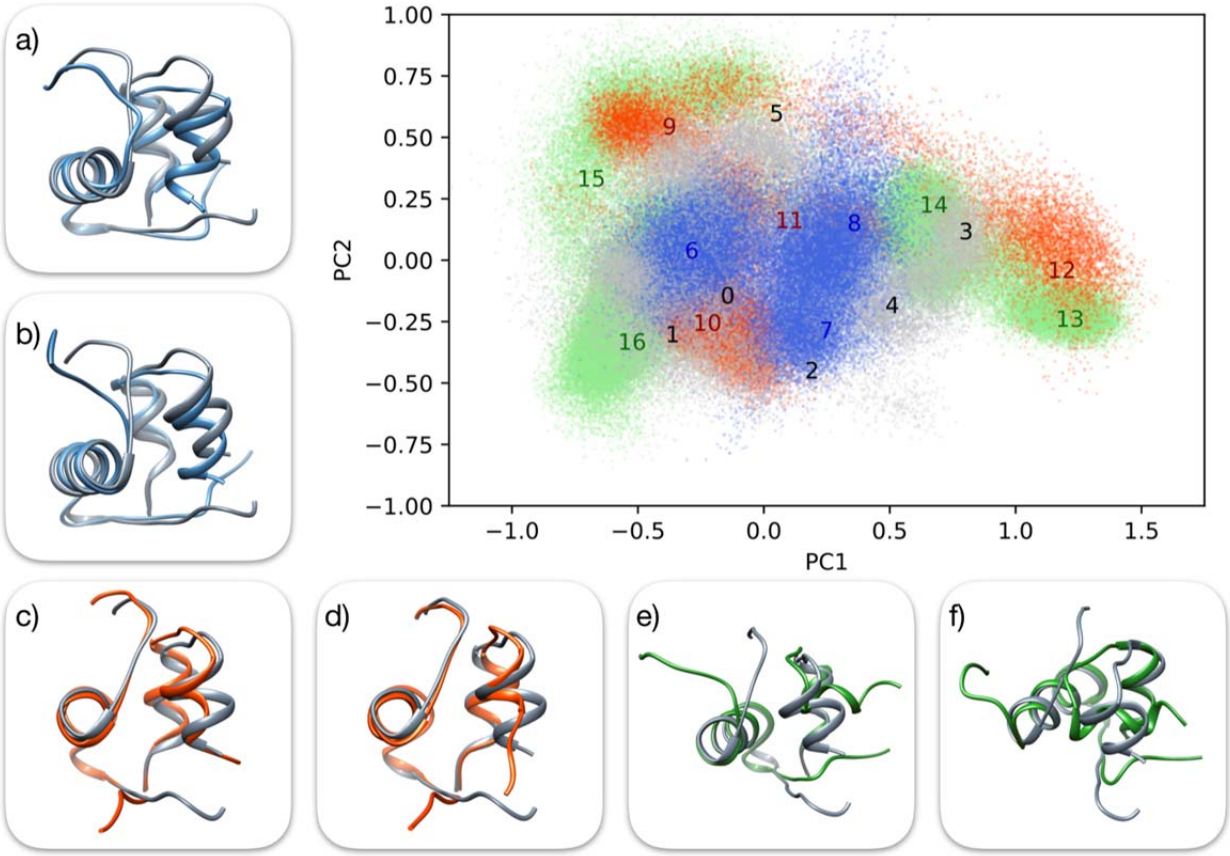

Figure S3. Comparative analysis of the three different binding modes with respect to the unbound state of insulin. Superposition of the PCA analyses (two main components relative to each other) of all the frames of the MDs is shown in the main sub-figure. Ligand-free insulin is represented in grey, binding mode α in blue, binding mode β in red and binding mode γ in green. Several representative frames are numbered from 1 to 16, being marked with 0 the crystallographic structure 1zni.[3] Superposition of some of these representative structures with 1zni are shown in sub-figures from a) to f) (corresponding to representatives 6, 7, 9, 12, 15 and 16, respectively).

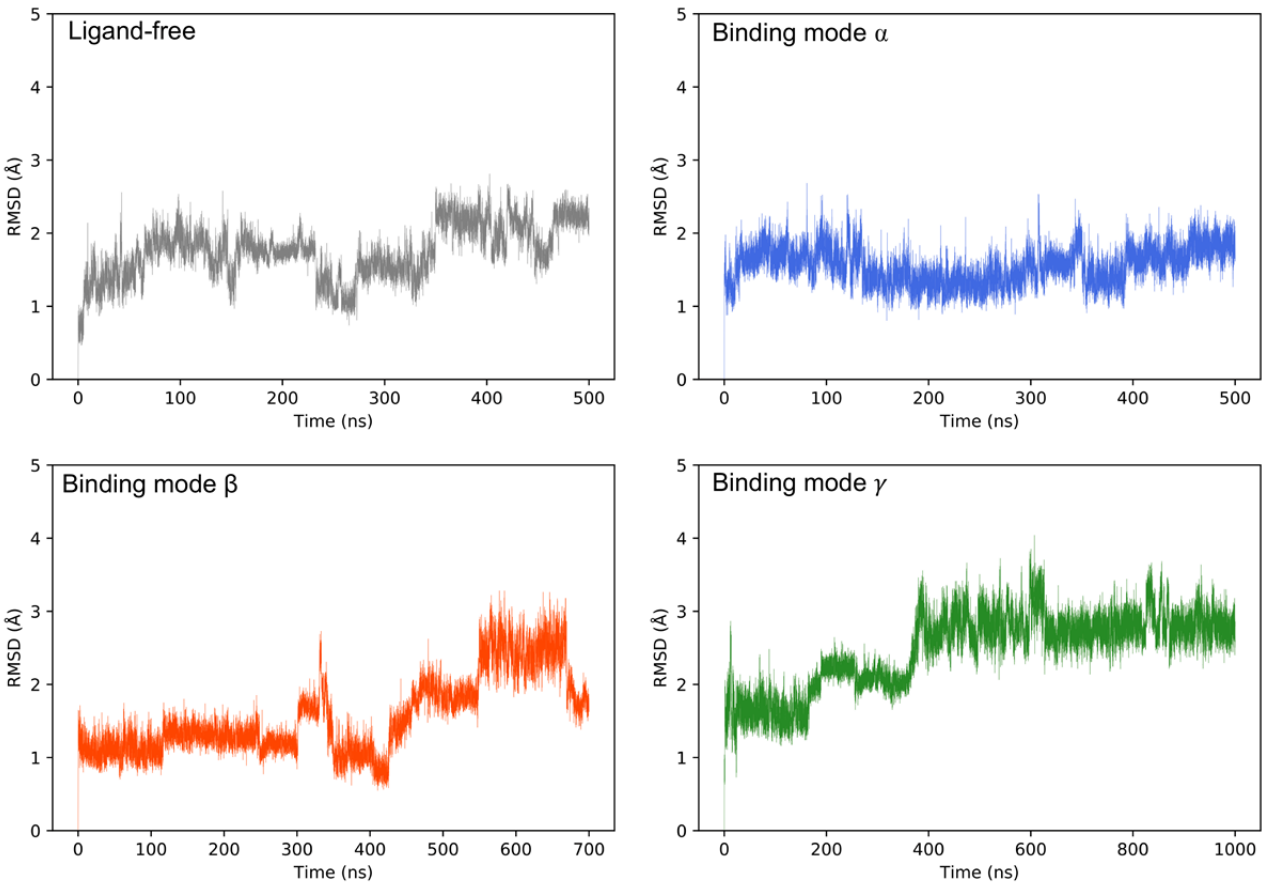

Figure S4. Scalar RMSD analysis from the initial frame calculated along each MD trajectory (ligand-free in grey, binding mode α in blue, binding mode β in red and binding mode γ in green).

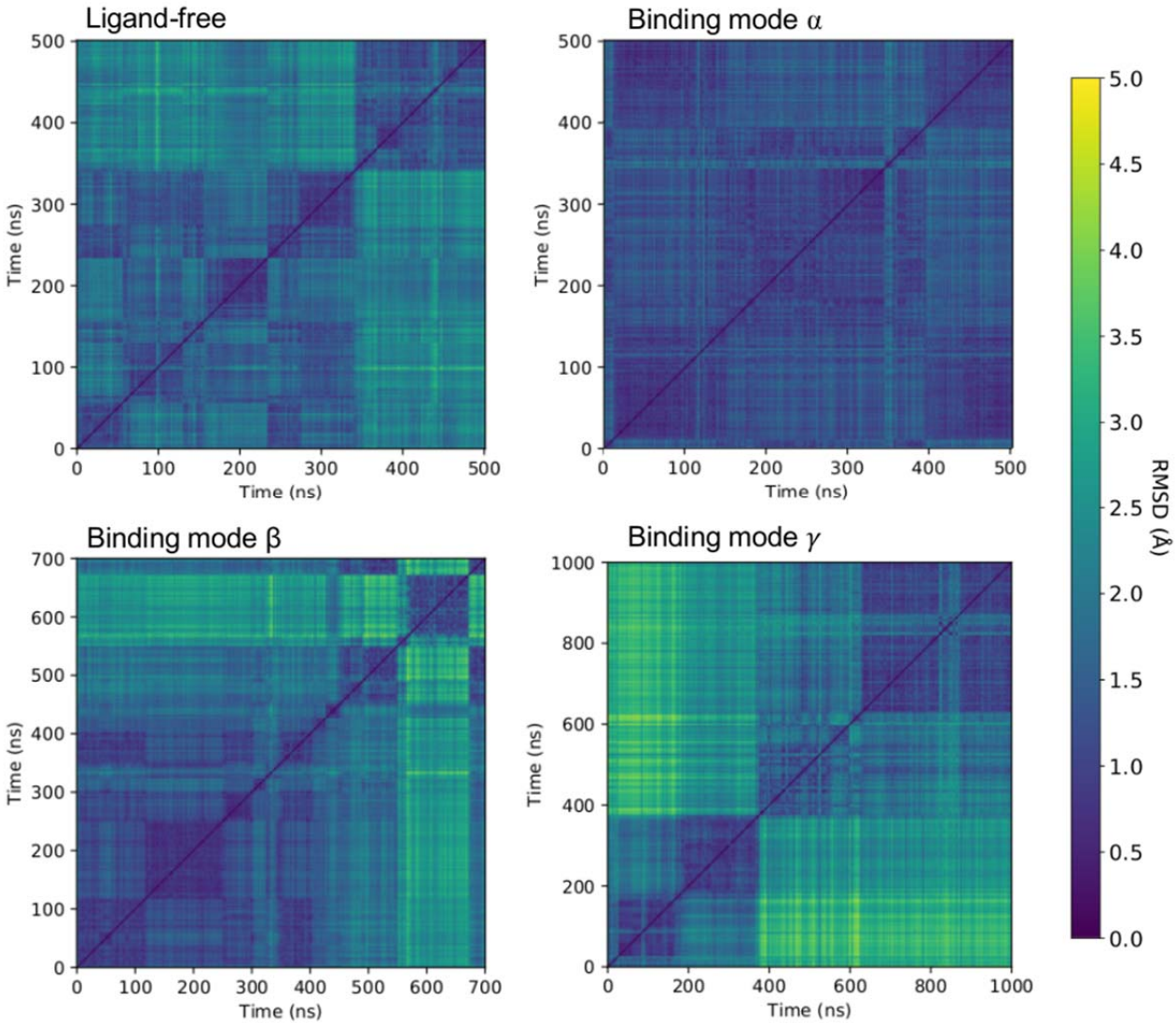

Figure S5. All-to-all frames RMSD study for each MD trajectory. For every frame, the RMSD with respect all other frames is plotted in the indicated colour scale.

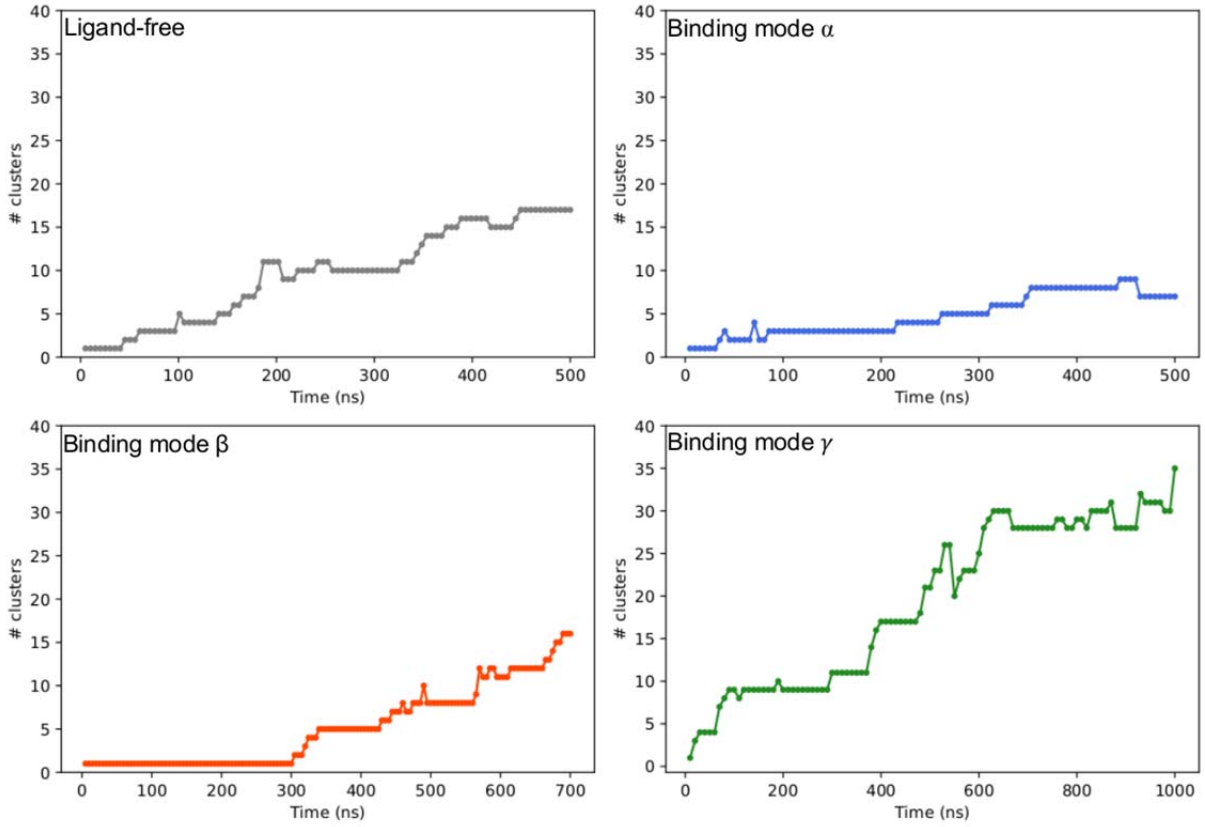

Figure S6. Cluster counting study for each MD trajectory (ligand-free in grey, binding mode α in blue, binding mode β in red and binding mode γ in green) setting the cut-off at 1.5 Å.

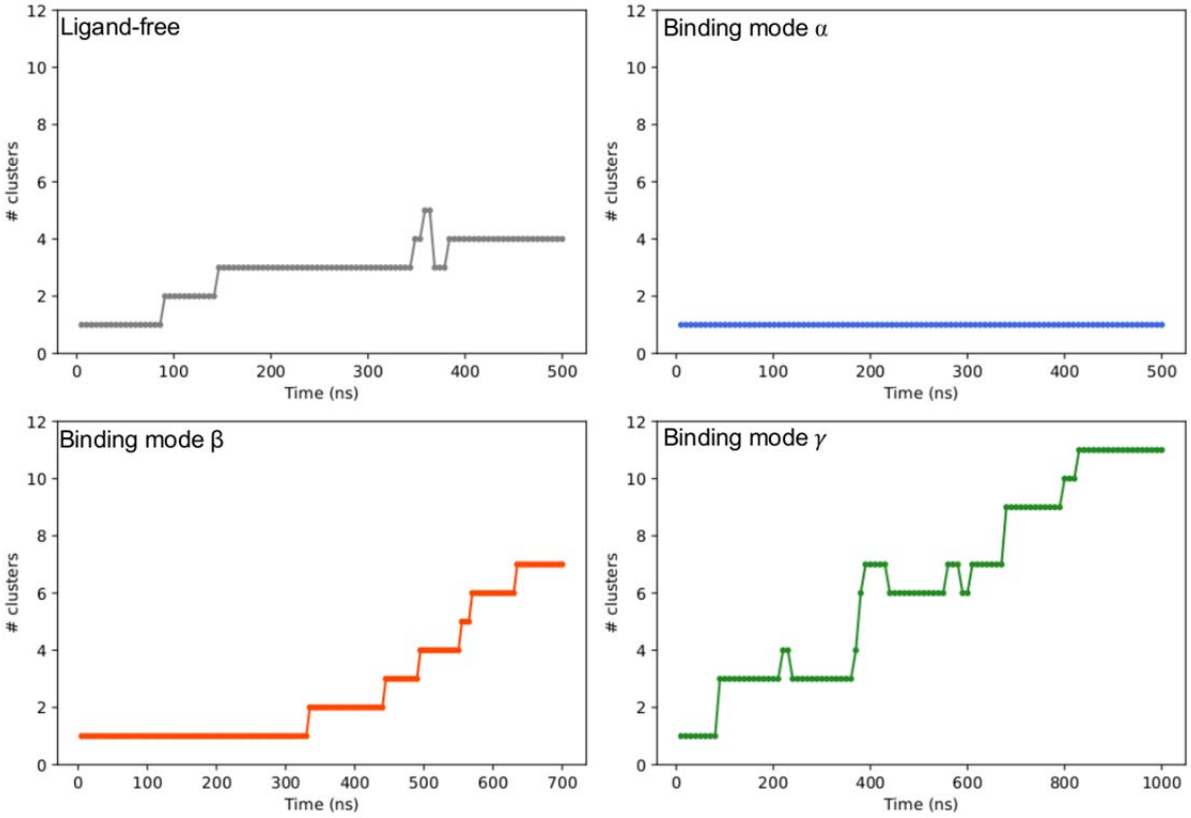

Figure S7. Cluster counting study for each MD trajectory (ligand-free in grey, binding mode α in blue, binding mode β in red and binding mode γ in green) setting the cut-off at 2.0 Å.

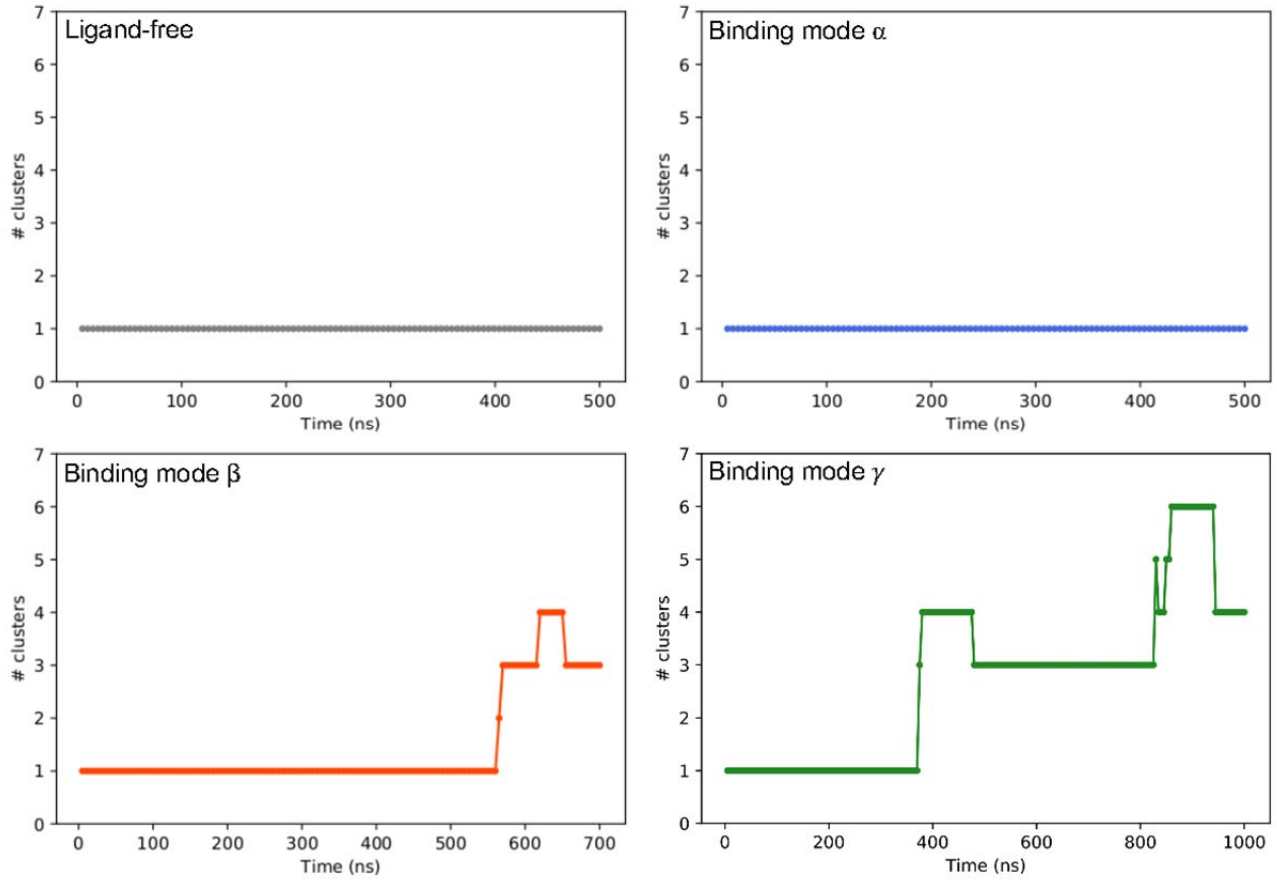

Figure S8. Cluster counting study for each MD trajectory (ligand-free in grey, binding mode α in blue, binding mode β in red and binding mode γ in green) setting the cut-off at 2.5 Å.

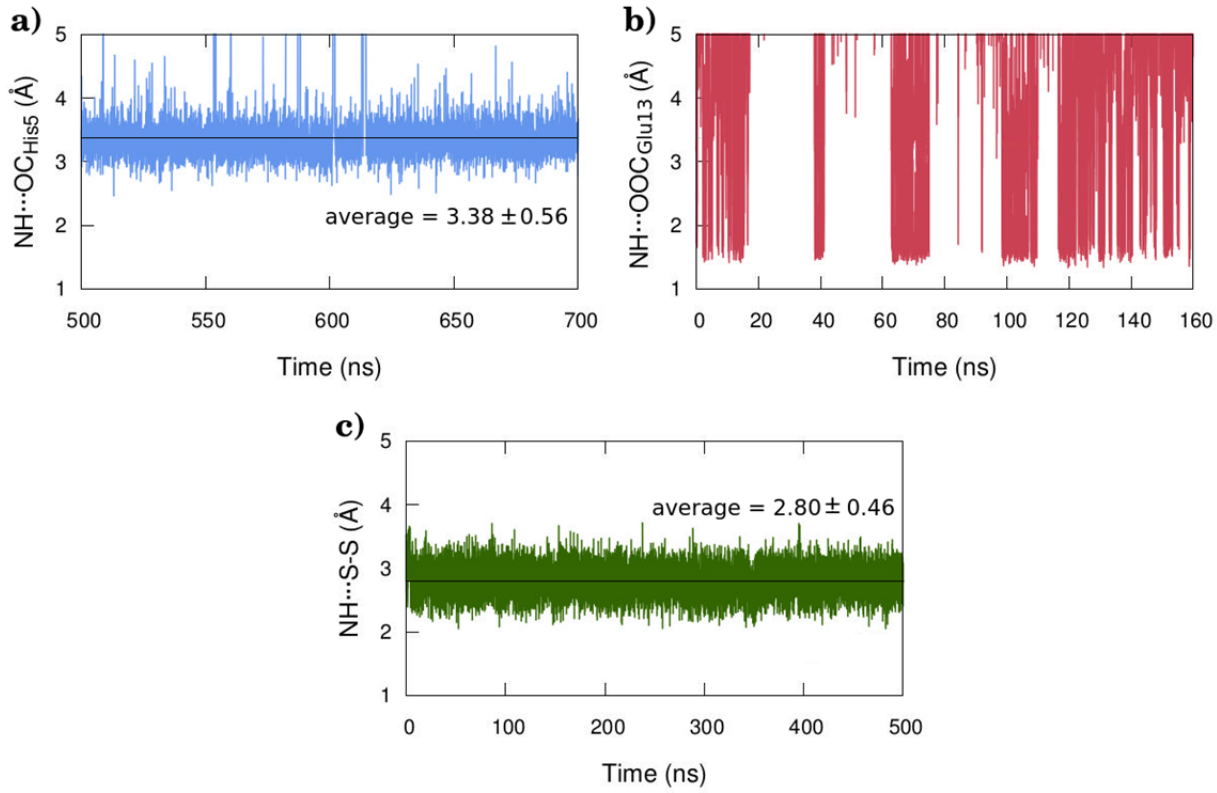

Figure S9. Computed distances along the MD trajectory of the adducts: a) distance between the hydrogen of amino group of dach and carbonyl oxygen of His5 of Pt(dach)(OH)(His10^{B})–insulin (in blue), b) distance between the hydrogen of amino group of dach and the carboxylate oxygen of Glu13 of Pt(dach)(OH)(His10^{B})–insulin (in red), and c) distance between the hydrogen of the amino group of dach and sulfur of Cys7^{A} of Pt(dach)(His5^{B}){Cys7^{B}(-SS-)}–insulin. Timescales only cover periods where the interaction is observed.

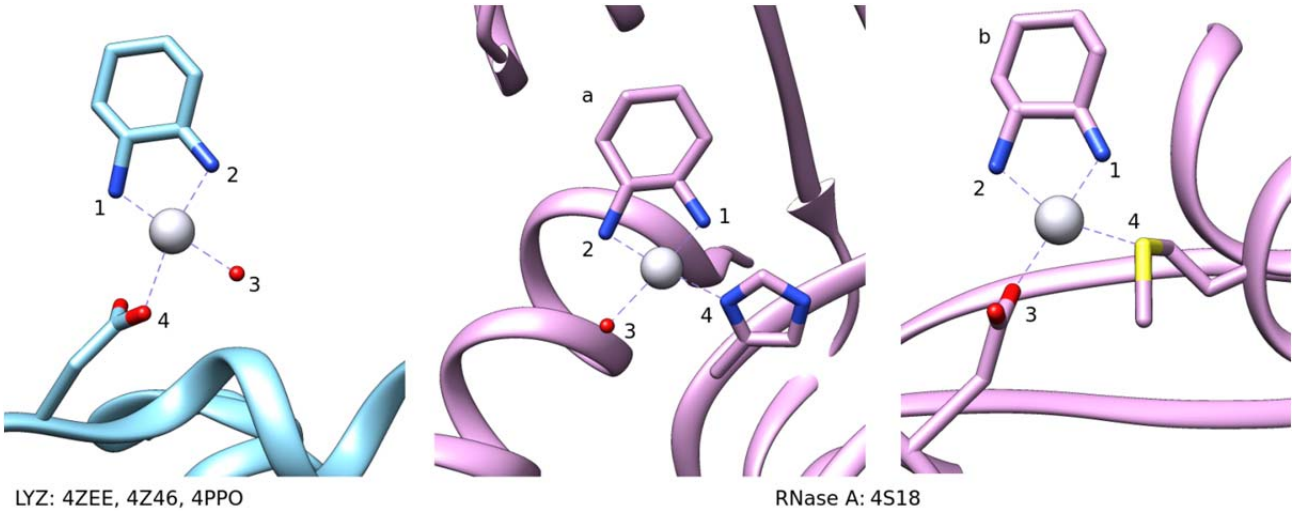

Figure S10. Indices for the binding sites of oxaliplatin found in the X-ray structures of lysozyme (Lyz, 4zee, 4z46, 4ppo) and ribonuclease A (RNase A, 4s18).

3. Supplementary tables

Table S1. Specifications of the solvent boxes built for each binding mode.

Mode	Box (Å)	Waters	Counterions (Na^+)
Binding mode α	55.4 × 50.6 × 57.9	3835	0
Binding mode β	55.4 × 50.6 × 56.4	3734	1
Binding mode γ	55.4 × 50.6 × 57.9	3835	1
Unbound	58.2 × 53.5 × 63.7	4758(+55, XRD)	2

Table S2. Distribution of clusters and most representative frames for binding mode α using the the K-medoids clustering algorithm.[8] The most representative structure of the most populated cluster is reported in boldface.

Cluster	Representative frame	% frames of the cluster
0	42909	8.4
1	6562	10.1
2	3461	7.3
3	39002	12.5
4	**45885**	**22.5**
5	32441	12.5
6	20783	6.4
7	21657	4.2
8	17466	9.5
9	22243	6.7

Table S3. Distribution of clusters and most representative frames for binding mode β using the the K-medoids clustering algorithm.[8] The most representative structure of the most populated cluster is reported in boldface.

Cluster	Representative frame	% frames of the cluster
0	20073	7.3
1	4638	12.1
2	9072	3.7
3	9521	8.1
4	10426	8.1
5	22297	8.9
6	**26110**	**16.4**
7	11694	8.9
8	13982	10.2
9	**32912**	**16.3** [a]

[a] the metal centre is superimposable, the structures differ for the orientation of the 4 last residues orientation of the A chain.

Table S4. Distribution of clusters and most representative frames for binding mode γ using the the K-medoids clustering algorithm.[8] The most representative structure of the most populated cluster is reported in boldface.

Cluster	Representative frame	% frames of the cluster
0	13100	14.0
1	752	2.5
2	32090	5.0
3	52128	8.2
4	28644	8.2
5	19844	7.8
6	43378	14.1
7	61028	4.1
8	**69582**	**24.8**
9	81318	11.3

4. References

1. P. Li and K. M. Merz Jr, Mcpb. Py: A Python Based Metal Center Parameter Builder.*Journal*, 2016.
2. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The Resp Model, *J. Phys. Chem.*, 1993, **97**, 10269-10280.
3. G. Bentley, E. Dodson, G. Dodson, D. Hodgkin and D. Mercola, Structure of Insulin in 4-Zinc Insulin, *Nature*, 1976, **261**, 166-168.
4. A. Grossfield and D. M. Zuckerman, Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations, *Annual reports in computational chemistry*, 2009, **5**, 23-48.
5. L. J. Smith, X. Daura and W. F. van Gunsteren, Assessing Equilibration and Convergence in Biomolecular Simulations, *Proteins: Struct., Funct., Bioinf.*, 2002, **48**, 487-496.
6. K. Mandal, B. Dhayalan, M. Avital-Shmilovici, A. Tokmakoff and S. B. H. Kent, Crystallization of Enantiomerically Pure Proteins from Quasi-Racemic Mixtures: Structure Determination by X-Ray Diffraction of Isotope-Labeled Ester Insulin and Human Insulin, *ChemBioChem*, 2016, **17**, 421-425.
7. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, Ucsf Chimera—a Visualization System for Exploratory Research and Analysis, *J. Comput. Chem.*, 2004, **25**, 1605-1612.
8. Y. Dodge, *Statistical Data Analysis Based on the L1-Norm and Related Methods*, Birkhäuser, 2012.