ELECTRONIC SUPPLEMENTARY INFORMATION (ESI)

Covalent and *non-covalent* binding in vanadium–protein adducts

Valeria Ugone,[a] Daniele Sanna,[a] Simone Ruggiu,[b] Giuseppe Sciortino,*[b,c] and Eugenio Garribba*[b]

[a] *Istituto CNR di Chimica Biomolecolare, Trav. La Crucca 3, I-07040 Sassari, Italy.*

[b] *Dipartimento di Chimica e Farmacia, Università di Sassari, Via Vienna 2, I-07100 Sassari, Italy. E-mail: garribba@uniss.it*

[c] *Institute of Chemical Research of Catalonia (ICIQ), Avgda. Països Catalans 16, 43007 Tarragona, Spain. E-mail: gsciortino@iciq.es*

1. Experimental and computational section

Chemicals. Water was deionized prior to use through the purification system Millipore MilliQ Academic or purchased from Sigma-Aldrich (LC-MS grade), when used for ESI-MS measurements. $V^{IV}O^{2+}$ solutions were prepared from $VOSO_4 \cdot 3H_2O$.[1] Nalidixic acid (1-ethyl-1,4-dihydro-7-methyl-4-oxo-1,8-naphthyridine-3-carboxylic acid, Hnal; code N8878), 1-methyimidazole (MeIm; M50834) and 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES; H3375), lysozyme from hen egg white (Lyz; 62970) and cytochrome *c* from equine hearth (Cyt; C2506) were purchased from Sigma-Aldrich with the highest grade available and used as received. The solid complex $[V^{IV}O(nal)_2(H_2O)] \cdot 2H_2O$ (V-nal) was synthesized according to the procedure established in the literature.[2]

EPR measurements. With Lyz, V-nal were dissolved in ultra-pure water to obtain a V concentration of 7.0×10^{-4} M. HEPES was added as a buffer with a concentration of 1.0×10^{-1} M. The value of pH was raised to 7.4 and, to 1 mL of this solution, Lyz was added to obtain a concentration of 7.0×10^{-4} M and a V-nal/Protein molar ratio of 1/1. Subsequently, various aliquots of the solution with V-nal were added to have a protein concentration of 3.5×10^{-4} M and 2.3×10^{-4} M and a ratio V-nal/Lyz of 2/1 and 3/1 (Lyz). Argon was bubbled through the solutions to ensure the absence of oxygen and avoid the oxidation of the $V^{IV}O^{2+}$ ion.

In the systems with Cyt, the protein solution (1.0×10^{-3} M) was prepared adding $Na_2S_2O_4$ (2.0×10^{-3} M) in order to reduce iron to Fe^{2+} and avoid the possible oxidation of V^{IV} to V^{V}. Subsequently, various aliquots of solutions containing V-nal were added to obtain a metal concentration of 5.0×10^{-4} M and a V-nal/Protein ratio of 1/1 and 2/1. In all the solutions Ar was always bubbled.

The solutions with the model systems containing only V-nal and V-nal/MeIm were prepared according to the data recently published.[2]

EPR spectra were recorded at 120 K with an X-band (9.4 GHz) Bruker EMX spectrometer equipped with a HP 53150A microwave frequency counter. When the samples were transferred into the EPR tubes, the spectra were immediately measured. Signal averaging was used to increase the signal to noise ratio.[3] The microwave frequency was in the range 9.40-9.41 GHz, microwave power was 20 mW, time constant was 81.92 ms, modulation frequency 100 kHz, modulation amplitude 0.4 mT, resolution 4096 points.

ESI-MS measurements. The solutions containing V-nal were prepared dissolving the solid complex V-nal in DMSO to obtain a 1.0×10^{-2} M and diluting subsequently with ultrapure water

(LC-MS grade, Sigma-Aldrich) and with an aliquot of protein stock solution (500 μM) in LC-MS grade water to obtain a final ratio V-nal/Protein of 3/1 and 5/1 and a protein concentration of 5 and 50 μM. Argon was bubbled through the solutions to ensure the absence of oxygen and avoid the oxidation of $V^{IV}O^{2+}$ ion. ESI-MS spectra were recorded immediately after the preparation of the solutions.

Mass spectra in positive-ion mode were obtained on a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ (Thermo Fisher Scientific) mass spectrometer. The solutions were infused at a flow rate of 5.00 μL/min into the ESI chamber. The spectra were recorded in the m/z range 300-4500 at a resolution of 140,000 and accumulated for at least 5 min in order to increase the signal-to-noise ratio. The instrumental conditions used for the measurements were as follows: spray voltage 2300 V, capillary temperature 250 °C, sheath gas 5-10 (arbitrary units), auxiliary gas 3 (arbitrary units), sweep gas 0 (arbitrary units), probe heater temperature 50 °C. Negative-ion mode: spray voltage -1900 V, capillary temperature 250 °C, sheath gas 20 (arbitrary units), auxiliary gas 5 (arbitrary units), sweep gas 0 (arbitrary units), probe heater temperature 14 °C.

The spectra were analyzed by using Thermo Xcalibur 3.0.63 software (Thermo Fisher Scientific) and the average deconvoluted monoisotopic masses were obtained through the Xtract tool integrated in the software.

Molecular Modelling protocol (DFT and docking calculation). The DFT optimized structure of the penta- (*SPY*-5-(12,13)), and hexa-coordinated (*OC*-6-(23,24,32,34)) isomers of $[V^{IV}O(nal)_2]$ and *cis*-$[V^{IV}O(nal)_2(H_2O)]$, have been obtained from ref. 2 (Scheme S1).

Docking calculations were performed with the GOLD 5.8 software[4] on the X-ray structures available in the Protein Data Bank (PDB) of lysozyme (PDB code: 2LYZ[5], and 4C3W[6]) and cytochrome *c* (PDB code: 1HRC[7]). The two X-ray structures of lysozyme show slight differences: 2LYZ is the native protein,[5] while in 4C3W the $V^{IV}O(pic)_2$ moiety is bound to Asp52.[6] The conformational variations could, in principle, bring differences during docking allowing a broader study. To overcome this potential problem, both proteins were studied. The structures were first prepared removing all the small molecules and crystallographic waters.

In the case of *coordinative* (*covalent*) docking the following protocol was applied. i) Proteins were probed for solvent accessible amino acid-containing COO^- donors throughout relative Solvent Excluded Surface (SES)[8] calculations obtaining the following residues as the most accessible: Asp18, Asp48, Asp87, Asp101, Asp119, Glu7 and Glu35 for Lyz, and Asp2, Asp50, Glu4, Glu21, Glu61, Glu62, Glu66, Glu69, Glu90, Glu92 and Glu104 for Cyt. For sake of completeness also His15 for Lyz and His26 for Cyt were taken into account, both the protonation states at δ and ε nitrogens of imidazole ring being considered. ii) The structure of the four

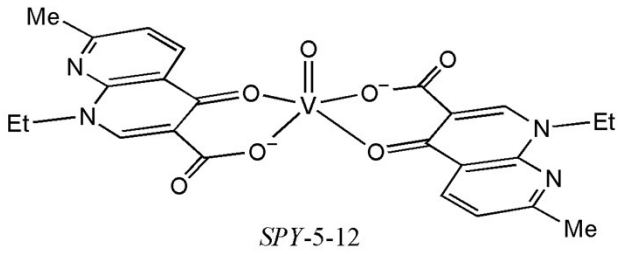
isomers of *cis*-[$VO(nal)_2(H_2O)$] and their enantiomers were prepared from the DFT optimized structures replacing the equatorial leaving H_2O ligand with a dummy hydrogen atom according to the procedure established recently.[9] All dockings were performed building an evaluation sphere with a radius of 10 Å, centered – for each simulation – in the region of interest considering the possible dihedral changes along the ligand aliphatic bonds applying the GOLD algorithm. The side chains flexibility was taken into account considering the GOLD implemented rotamers libraries.[10] Genetic algorithm (GA) parameters have been set to 100 GA runs and a minimum of 100,000 operations. The other parameters of GA were set to default. The solutions were analyzed by means of GaudiView.[11]

Concerning the non-specific *non-covalent* dockings, the methods in the literature were applied to all the *SPY*-5 and *OC*-6 isomers.[12] The simulations were performed on 2[$V^{IV}O(nal)_2$]-Lyz or 2[$V^{IV}O(nal)_2$]-Cyt adducts, coming from the *coordinative* dockings (being two the maximum number of $V^{IV}O(nal)_2$ moieties bound simultaneously with a *covalent* bond to proteins, considering the whole rigid protein, while side chains flexibility was taken into account using the GOLD implemented rotamers libraries[10] for the coordination dockings. Genetic algorithm (GA) parameters have been set to perform a minimum of 100,000 operations during 50 GA runs for coordination, and 100 GA runs for classic dockings. The other parameters of GA were set to default.

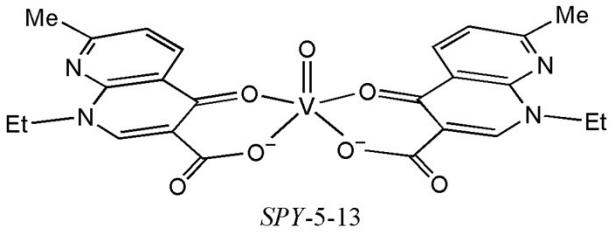
All dockings were performed according to the following strategy: i) with a first *coordinative* docking, the primary binding site has been determined and the adduct [$V^{IV}O(nal)_2$]–Protein identified; ii) a second *coordinative* docking was carried out on the structure [$V^{IV}O(nal)_2$]–Protein to identify the eventual secondary site with a *covalent* bond; iii) an additional *coordinative* docking has been performed on the structure 2[$V^{IV}O(nal)_2$]–Protein without finding (neither with Lyz nor with Cyt) another stable *covalent* adduct; iv) a series of subsequent *non-covalent* dockings were performed on the adduct 2[$V^{IV}O(nal)_2$]–Protein until reaching the total number of moieties suggested by ESI-MS (4 for Lyz and 5 for Cyt).

The scoring (*Fitness* of GoldScore) was evaluated applying the modified versions of GoldScore scoring function, which was validated in previous published papers.[9, 12] The best solutions (binding poses) were evaluated through three main criteria: i) the mean (F_{mean}) and the highest value (F_{max}) of the scoring (*Fitness* of GoldScore) associated with each pose; ii) the population of the cluster containing the best pose; iii) the position in the *Fitness* ranking of the computed cluster.

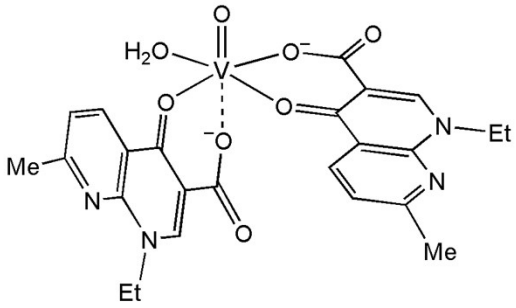
2. Supplementary schemes, tables and figures



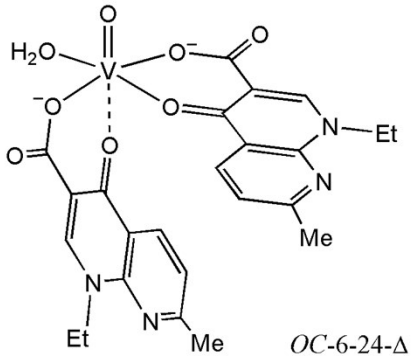
SPY-5-12



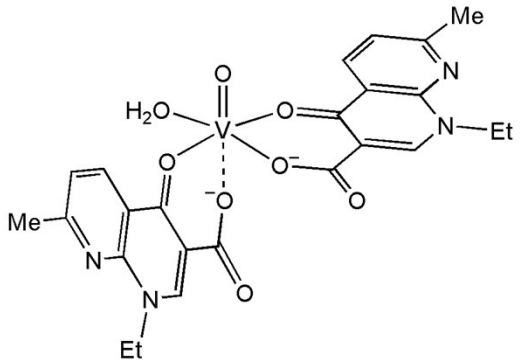
SPY-5-13



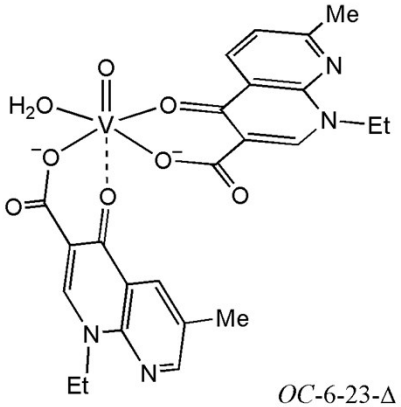
OC-6-34-Δ



OC-6-24-Δ



OC-6-32-Δ



OC-6-23-Δ

Scheme S1. Isomers for the bis-chelated $V^{IV}O$ species of nalidixato ligand (only the enantiomers of the Δ series are shown for simplicity).[2]

Table S1. Main adducts formed after the interaction of $[V^{IV}O(nal)_2(H_2O)]$ with Lyz and Cyt revealed with ESI-MS.

Protein	Adduct	V-nal/protein	Conc. protein / μM	Mass / Da
Lyz	$[V^{IV}O(nal)]$–Lyz	3/1, 5/1	50	14601.9
	$[V^{IV}O(nal)_2]$–Lyz	3/1, 5/1	5, 50	14834.0
	$\{[V^{IV}O(nal) + [V^{IV}O(nal)_2]\}$–Lyz	3/1, 5/1	50	15130.9
	$2[V^{IV}O(nal)_2]$–Lyz	3/1, 5/1	5, 50	15363.1
	$\{[V^{IV}O(nal) + 2[V^{IV}O(nal)_2]\}5$–Lyz	3/1, 5/1	50	15660.1
	$3[V^{IV}O(nal)_2]$–Lyz	3/1, 5/1	5, 50	15892.2
	$\{[V^{IV}O(nal) + 3[V^{IV}O(nal)_2]\}$–Lyz	5/1	50	16190.1
	$4[V^{IV}O(nal)_2]$–Lyz	5/1	50	16423.2
Cyt	$[V^{IV}O(nal)]$–Cyt	3/1, 5/1	50	12656.4
	$[V^{IV}O(nal)_2]$–Cyt	3/1, 5/1	5, 50	12887.5
	$\{[V^{IV}O(nal) + [V^{IV}O(nal)_2]\}$–Cyt	3/1, 5/1	50	13185.5
	$2[V^{IV}O(nal)_2]$–Cyt	3/1, 5/1	5, 50	13418.5
	$\{[V^{IV}O(nal) + 2[V^{IV}O(nal)_2]\}$–Cyt	3/1, 5/1	50	13714.5
	$3[V^{IV}O(nal)_2]$–Cyt	3/1, 5/1	5, 50	13947.6
	$\{[V^{IV}O(nal) + 3[V^{IV}O(nal)_2]\}$–Cyt	3/1, 5/1	50	14243.7
	$4[V^{IV}O(nal)_2]$–Cyt	3/1, 5/1	5, 50	14475.7
	$\{[V^{IV}O(nal) + 4[V^{IV}O(nal)_2]\}$–Cyt	5/1	50	14774.7
	$5[V^{IV}O(nal)_2]$–Cyt	5/1	50	15005.8

Table S2. Spin Hamiltonian parameters measured from EPR for binary and ternary $V^{IV}O$ species of nalidixato ligand.

Species	g_z	$\lvert A_z \rvert$ [a]	Donor set
cis-$[V^{IV}O(nal)_2(H_2O)]$	1.940	173.5	(CO, COO^-); (CO, COO^-); H_2O [b]
$[V^{IV}O(nal)_2]$	1.946	167.7	(CO, COO^-); (CO, COO^-)
cis-$[V^{IV}O(nal)_2(MeIm)]$	1.944	169.1	(CO, COO^-); (CO, COO^-); Imidazole-N [b]
$[V^{IV}O(nal)_2]$–Lyz	1.943	171.2	(CO, COO^-); (CO, COO^-); Asp/Glu-COO^- [b]
$[V^{IV}O(nal)_2]$–Cyt	1.943	171.2	(CO, COO^-); (CO, COO^-); Asp/Glu-COO^- [b]

[a] A values reported in 10^{-4} cm^{-1} units. [b] One CO or COO^- group of nalidixato in the equatorial position.

Table S3. Best docking solutions for *covalent* binding of *cis*-[$V^{IV}O(nal)_2$] to lysozyme and cytochrome *c*.

Protein	Isomer [a]	Donor	d(V–D_{aa}) [b]	F_{max} [c]	F_{mean} [d]	Pop.% [e]	Rank.
Lyz [f]	*OC*-6-32-Λ	Asp52	2.014	56.7	49.3	22.5	I
	OC-6-34-Λ	Asp87	2.310	48.5	45.2	30.3	II
	OC-6-34-Δ	Asp18	2.396	49.3	41.1	22.5	III
	OC-6-34-Δ	Asp101	2.300	45.0	38.1	14.6	IV
Cyt	*OC*-6-24-Δ	Glu21[f]	2.008	54.1	43.0	38.1	I
	OC-6-23-Λ	Glu90[g]	2.360	39.5	34.2	61.9	II

[a] Isomer with the best affinity. [b] Distance between V and the amino acid donors (D_{aa}) reported in Å. [c] GoldScore *Fitness* value obtained for the more stable pose of each *cluster*. [d] Average value of GoldScore *Fitness* for each *cluster*. [e] Percent population computed taking into account all the isomers. [f] Values relative to the pdb structure with code 4C3W which obtained the best affinities in the docking essay.

Table S4. Docking solutions for *non-covalent* binding of $[V^{IV}O(nal)_2]$ (*SPY*-5) and *cis*-$[V^{IV}O(nal)_2(H_2O)]$ (*OC*-6) to lysozyme and cytochrome *c*.

Protein	Isomer	Interactions	F_{max} [a]	F_{mean} [b]	Pop.%[c]	Rank.
Lyz [d]	*SPY*-5-13	VO···Arg21	19.2	17.7	44.1	I
		VO···Ser100				
		VO···Tyr20				
		N(nal) ··Lys96				
	SPY-5-12	COO(nal)···Lys13	14.1	12.7	8	II
		COO(nal)···Arg14				
	OC-6-34-Δ	VO···Arg73	13.3	13.3	0.6	III
		N(nal)···Arg61				
	OC-6-23-Δ	VO···Arg114	13.2	12.9	1.2	IV
		COO(nal)···Arg114				
	OC-6-24-Λ	N(nal)···Tyr23	13.1	13.1	1.3	V
		N(nal)···Arg21				
		COO(nal)···Arg73	10.4	10.4	0.6	VI
Cyt	*OC*-6-32-Λ	COO(nal)···Lys72	19.1	17.5	35.7	I
		COO(nal)···Ala83(CO)				
		Nal(N)···Gln16				
	OC-6-23-Δ	VO···Lys87	18.6	16.8	35.7	I
		VO···Lys86				
		COO(nal)···Lys13				
	OC-6-24-Λ	N(nal)···Lys55	17.0	17.0	7.1	III
	OC-6-24-Δ	COO(nal)···Arg38	11.5	11.5	7.1	IV
	SPY-5-12	COO(nal)···Arg91	9.4	9.4	7.1	V
		N(nal)···Asn70	7.7	7.7	7.1	VI

[a] GoldScore *Fitness* value obtained for the more stable pose of each *cluster*. [b] Average value of GoldScore *Fitness* for each *cluster*. [c] Percentage population of the *cluster*. [d] Values relative to the pdb structure with code 2LYZ.

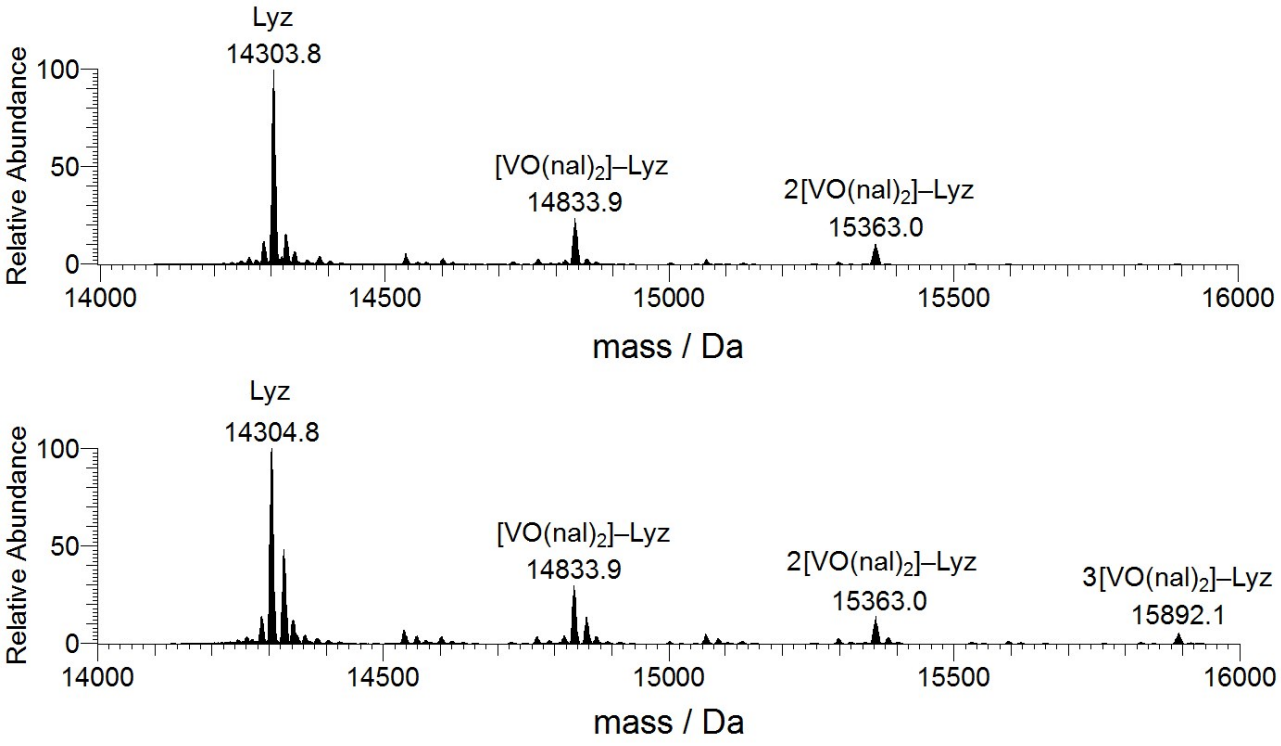


Fig. S1. Deconvoluted ESI-MS spectra recorded on the system containing $[V^{IV}O(nal)_2(H_2O)]$ and lysozyme (5 μM): molar ratios 3/1 (top) and 5/1 (bottom).

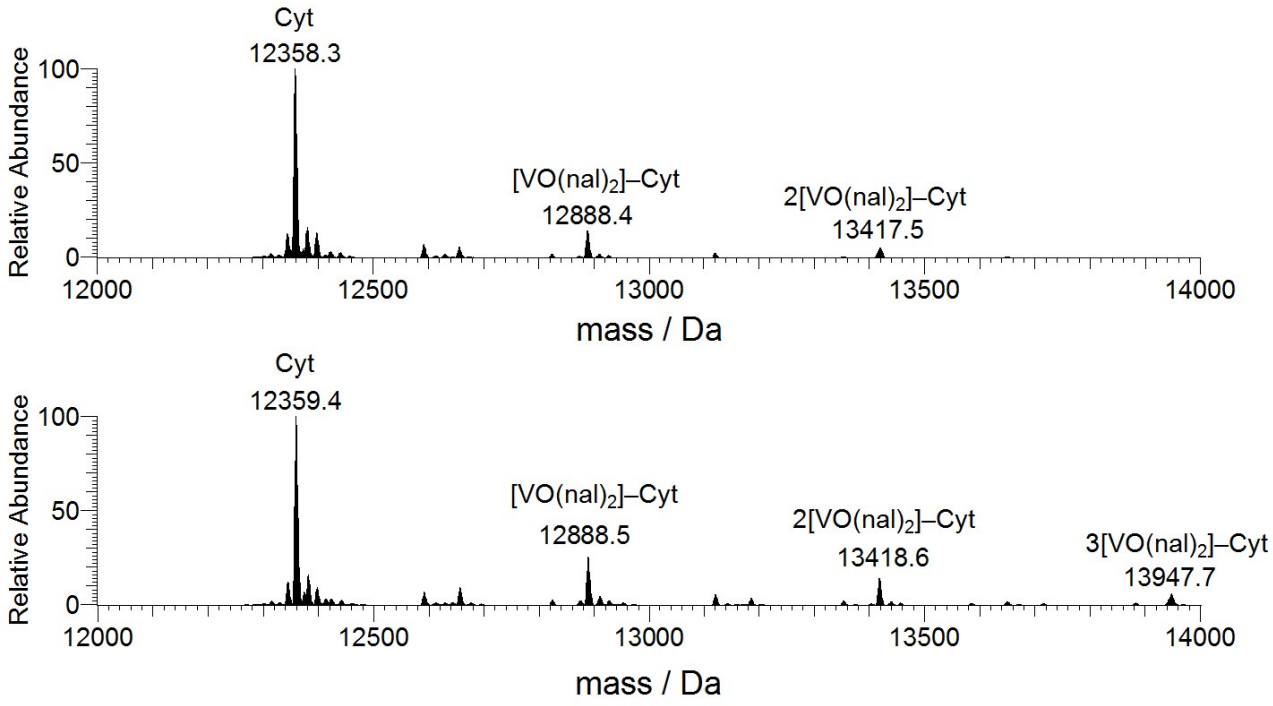


Fig. S2. Deconvoluted ESI-MS spectra recorded on the system containing $[V^{IV}O(nal)_2(H_2O)]$ and cytochrome *c* (5 μM): molar ratios 3/1 (top) and 5/1 (bottom).

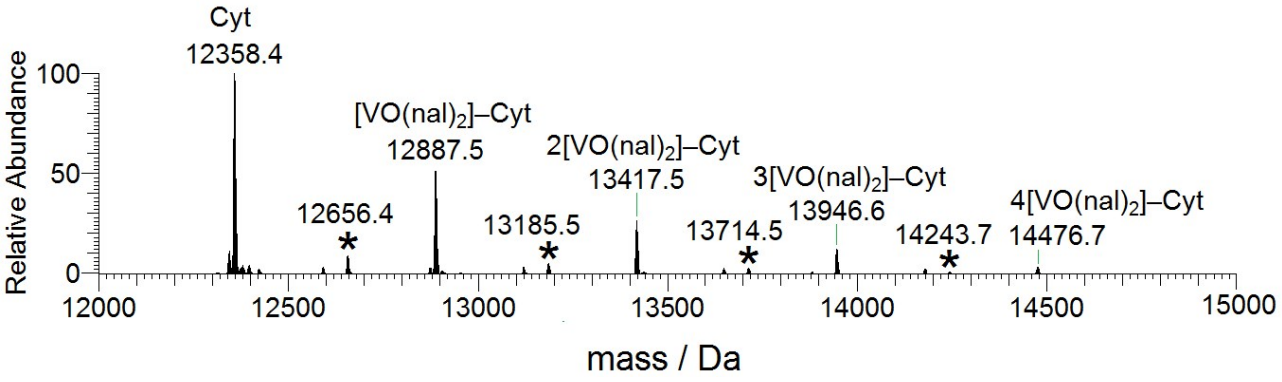


Fig. S3. Deconvoluted ESI-MS spectrum recorded on the system containing $[V^{IV}O(nal)_2(H_2O)]$ and cytochrome *c* (50 μM) with molar ratios 3:1. With the asterisks the minor peaks attributed to $\{[V^{IV}O(nal) + n[V^{IV}O(nal)_2]\}$–Cyt are indicated; their masses fall at *ca*. (12358 + 298 + n × 529) Da.

3. References

1. I. Nagypál and I. Fábián, NMR relaxation studies in solution of transition metal complexes. V. Proton exchange reactions in aqueous solutions of VO^{2+}-oxalic acid, -malonic acid systems, *Inorg. Chim. Acta*, 1982, **61**, 109-113.
2. D. Sanna, V. Ugone, G. Sciortino, P. Buglyo, Z. Bihari, P. L. Parajdi-Losonczi and E. Garribba, $V^{IV}O$ complexes with antibacterial quinolone ligands and their interaction with serum proteins, *Dalton Trans.*, 2018, **47**, 2164-2182.
3. D. Sanna, E. Garribba and G. Micera, Interaction of VO2+ ion with human serum transferrin and albumin, *J. Inorg. Biochem.*, 2009, **103**, 648-655.
4. G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, Development and validation of a genetic algorithm for flexible docking1, *J. Mol. Biol.*, 1997, **267**, 727-748.
5. R. Diamond, Real-space refinement of the structure of hen egg-white lysozyme, *J. Mol. Biol.*, 1974, **82**, 371-391.
6. M. F. A. Santos, I. Correia, A. R. Oliveira, E. Garribba, J. Costa Pessoa and T. Santos-Silva, Vanadium Complexes as Prospective Therapeutics: Structural Characterization of a V^{IV} Lysozyme Adduct, *Eur. J. Inorg. Chem.*, 2014, 3293-3297.
7. G. W. Bushnell, G. V. Louie and G. D. Brayer, High-resolution three-dimensional structure of horse heart cytochrome c, *J. Mol. Biol.*, 1990, **214**, 585-595.
8. M. Connolly, Analytical molecular surface calculation, *J. Appl. Crystallogr.*, 1983, **16**, 548-558.
9. (a) G. Sciortino, D. Sanna, V. Ugone, G. Micera, A. Lledós, J.-D. Maréchal and E. Garribba, Elucidation of Binding Site and Chiral Specificity of Oxidovanadium Drugs with Lysozyme through Theoretical Calculations, *Inorg. Chem.*, 2017, **56**, 12938-12951; (b) G. Sciortino, J. Rodríguez-Guerra Pedregal, A. Lledós, E. Garribba and J.-D. Maréchal, Prediction of the interaction of metallic moieties with proteins: an update for protein-ligand docking techniques, *J. Comput. Chem.*, 2018, **39**, 42-51; (c) G. Sciortino, E. Garribba and J.-D. Maréchal, Validation and Applications of Protein–Ligand Docking Approaches Improved for Metalloligands with Multiple Vacant Sites, *Inorg. Chem.*, 2019, **58**, 294-306.
10. S. C. Lovell, J. M. Word, J. S. Richardson and D. C. Richardson, The penultimate rotamer library, *Proteins: Struct., Funct., Bioinf.*, 2000, **40**, 389-408.
11. (a) J. Rodríguez-Guerra, *Insilichem/gaudiview: Pre-alpha public releas, Zenodo*, 2017; (b) J. Rodríguez-Guerra Pedregal, G. Sciortino, J. Guasp, M. Municoy and J.-D. Maréchal, GaudiMM: A modular multi-objective platform for molecular modeling, *J. Comput. Chem.*, 2017, **38**, 2118-2126.

12. G. Sciortino, D. Sanna, V. Ugone, A. Lledós, J.-D. Maréchal and E. Garribba, Decoding Surface Interaction of $V^{IV}O$ Metallodrug Candidates with Lysozyme, *Inorg. Chem.*, 2018, **57**, 4456-4469.