Supplementary Information

Does a Fast Nuclear Magnetic Resonance Spectroscopy- and X-ray Crystallography Hybrid Approach Provide Reliable Structural Information of Ligand-Protein Complexes? A Case Study of Metalloproteinases.

Johan Isaksson^{a}; Susanne Nyström^a; Dean Derbyshire^a; Hans Wallberg^a; Tatiana Agback^a; Helena Kovacs^b; Ivano Bertini^{c,d}, Andrea Giachetti^{c,f}, Claudio Luchinat^{c,e}*

^aMedivir AB, PO Box 1086, SE-141 22 Huddinge, Sweden,

^bBruker BioSpin AG, Industriestrasse 26, CH-8117 Fällanden, Switzerland

^cMagnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy

^dDepartment of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

^eDepartment of Agricultural Biotechnology, University of Florence, Via Maragliano, 75–77, 50144 Florence, Italy

^fProtEra S.r.l., Via delle Idee, 22, 50019 Sesto Fiorentino, Italy

* To whom correspondence should be addressed. Phone, +46-8-546 832 27; fax, +46-8-546 831 99; email, johan.isaksson@medivir.se

Contents

METHODOLOGY DETAILS.....	3
CRYSTALLISATION AND X-RAY STRUCTURE ANALYSIS.....	3
NMR DATA COLLECTION.	4
REFERENCES	7
LEGENDS.....	8
TABLES	9
FIGURES	11

Methodology details

Crystallisation and X-ray Structure analysis.

Crystals of the catalytic domain of human MMP12 in a complex with the hydroxamic acid inhibitor: CGS27023A, were produced using the ‘hanging-drop’ vapour diffusion method¹. Protein was concentrated in 2 stages: initially to 1 mg/ml (at which point the inhibitor was added to give a final concentration of 0.75 mM); then subsequently to a final concentration of 5 mg/ml. Optimal crystals were grown at 18 °C using a ratio of 1.5 µl ‘complex’ to 1 µl reservoir containing: 100 mM imidazol (pH 6.2), and 1.1 -1.6 M Sodium Acetate. Brick shaped crystals grew to approximately 75 µm x 100 µm x 300 µm after 2 weeks. The addition of 1 mM 1,3-diamino-propane was found to improve crystal quality somewhat, which suffered quite badly from heterogeneous nucleation and intergrowth. Indeed, multiple lattices could be observed on test diffraction images, even with the best crystals.

A complete high resolution dataset was collected at the Synchrotron Radiation Source (Daresbury, Cheshire, UK) station 14.2 ($\lambda = 0.9801 \text{ \AA}$). Crystals were found to belong to the monoclinic spacegroup $P2_1$ with unit cell dimensions: $A = 38.160 \text{ \AA}$; $B = 99.939 \text{ \AA}$; $C = 79.423 \text{ \AA}$; and $\beta = 96.59^\circ$. Matthews coefficient calculations² suggested a probable crystal packing with four molecules in the asymmetric unit. Despite an obvious second lattice, data was successfully integrated, scaled and ‘truncated’ to $2.0 \text{ \AA}$ using MOSFLM, SCALA, and TRUNCATE respectively³ (Table S1). The structure was solved using PHASER⁴ with a model derived from existing in-house coordinates. This confirmed the

presence of four molecules with the 'log likelihood gain' rising from 143 (for the placement of one molecule) to 2391 (for all four). Furthermore, the first electron density map was of sufficient quality to allow interpretation of features removed in the initial model – the model used extended from 109-260 only, and all Zinc and calcium atoms were removed.

The use of non-crystallographic symmetry averaging using DM⁵ greatly improved the early electron density maps. Subsequent rounds of re-modelling and refinement, COOT⁶ and REFMAC5⁷ respectively, resulted in the unequivocal interpretation of all four copies of both protein and inhibitor. The final structure returns respectable 'R' values of $R_{\text{working}} = 17.87\%$; and $R_{\text{free}} = 23.68\%$ respectively, despite obvious repercussions of the systematic errors arising from the multiple lattices in the diffraction pattern. Coordinate analysis was carried out by using PROCHECK⁸. The final coordinates and structure factors have been deposited into the PDB, accession code: 2W0D.

NMR data collection.

All experiments were part of the standard Topspin 2.1 pulse program package and were performed at 22 °C, using Bruker Avance II spectrometers operating at 500, 600, 700 and 900 MHz ¹H-frequency, equipped with cryogenically cooled inverse triple resonance probes.

For the assignment of the ^1H -, ^{13}C - and ^{15}N - resonances of the protein two-dimensional (2D) ^1H - ^{15}N -HSQC and the following 3-dimensional (3D) spectra were acquired at 600MHz: ^{15}N -edited NOESY, HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB, HCCH-TOCSY, (H)CCH-TOCSY, HC(C)H-COSY, H(CCCO)NH, HBHANH and HBHA(CO)NH. Where appropriate, the versions relying on the WATERGATE method⁹ for the suppression of the H_2O signal were employed.

For the assignment of the ligand resonances in the bound conformation, 2D $^{12}\text{C}/^{12}\text{C}$ -filtered NOESY and 2D $^{12}\text{C}/^{12}\text{C}$ -filtered TOCSY experiments were acquired. These spectra were typically recorded with 128 scans per increment, 200 complex points in indirect and 2k complex points in acquisition dimension. The spectral window was 20 ppm in both dimensions and the carrier frequency was set on the water frequency. The mixing time was 125ms in $^{12}\text{C}/^{12}\text{C}$ -filtered NOESY and ROESY and 80 ms in $^{12}\text{C}/^{12}\text{C}$ -filtered TOCSY experiment.

The protein-ligand interactions were monitored by means of ω_2 - ^{15}N -edit/ ω_1 - ^{15}N , ^{13}C -filtered (^{15}N -X-filtered) and ω_2 - ^{13}C -edit/ ω_1 - ^{15}N , ^{13}C -filtered (^{13}C -X-filtered) 3D NOESY¹⁰, and a modified 3D simultaneous ^{13}C - and ^{15}N -X-filtered NOESY experiment. In these experiments the ligand proton frequencies appear in the F1-dimension, protein heteronuclei in F2-dimension and protein protons in the acquisition dimension. The 3D spectra were recorded with 32 transients per increment. The number of complex points for ^1H was 2k in the acquisition dimension and 108 in the t_1 -dimension. The number of complex points in the heteronuclear indirect dimension was 58 for ^{13}C -correlated spectrum, 40 for the ^{15}N -correlated spectrum and 48 for the simultaneous experiment. The ^1H spectral window was set to 20 ppm while the carrier was on the water frequency,

the ^{13}C spectral window was set to 140 ppm with the carrier frequency set at 70 ppm, and the ^{15}N spectral window was set to 36 ppm with the ^{15}N -frequency set at 120 ppm.

In order to assess the inter-molecular NOE build-up rates a series of the 2D ^1H - ^1H planes of the 3D ^{13}C -X-filtered NOESY experiment were recorded at 600 MHz using mixing times of 25, 50, 75, 100, 125 and 150 ms. For the collection of the NOE based structural constraints the 3D X-filtered NOESY experiments were recorded with mixing times of 50 ms, 75 ms and 125 ms of which the 75 ms spectra was used to extract interaction points. In all the X-filtered experiments adiabatic pulses with matched frequency sweep were employed for proper filtering of the ^{13}C -multiplicities¹⁰. For this purpose, smoothed adiabatic CHIRP¹¹ pulses were used, where the ^{13}C -frequency was swept through 60 kHz in 2.098 ms at 500 MHz, in 1.751 ms at 600 MHz and in 1.502 ms at 700 MHz and through 80 kHz in 1.5 ms at the 900 MHz instrument.

REFERENCES

1. McPherson, A., *Preparation and Analysis of Protein Crystals*. John Wiley: New York, 1982.
2. Matthews, B. W., *J. Mol. Biol.* **1968**, *33*, 491-497.
3. The CCP4 (Collaborative Computational Project, Number 4) suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **1994**, *50*, 760-763.
4. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., *J. Appl. Cryst.* **2007**, *40*, 658-674.
5. Cowtan, K., Joint CCP4 and ESF-EACBM. *Protein Crystallography* **1994**, *31*, 34-38.
6. Emsley, P.; Cowtan, K., *Acta Crystallogr. D. Biol. Crystallogr.* **2005**, *60*, 2126-2132.
7. Murshudov, G.; Vagin, A.; Dodson, E., Application of Maximum Likelihood Refinement. *Proc. of Daresbury Study Weekend - Refinement of Protein Structures* **1996**.
8. Laskowski, A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereorechemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283-291.
9. Piotto, M.; Saudek, V.; Sklenar, V., *J. Biomol. NMR* **1992**, *2*, 661-665.
10. Zwahlen, C.; Legault, P.; Vincent, S. J. F.; Greenblatt, J.; Konrat, R.; Kay, L. E., *J. Am. Chem. Soc.* **1997**, *119*, (6711-6721).
11. Böhlen, J.-M.; G., B., *J. Magn. Reson. Ser A* **1993**, *102*, 293-301.

LEGENDS

Figure S1. RMSD plots of all docked crystal structures (A-O) using only HN contacts (left column) and both HN and HC contacts (right column).

Figure S2. RMSD vs. nOe energy plot of the test calculation comparing the three clusters minimized with integrated restraints based on an arbitrary reference distance (**red**, 0.5 Å flat well and **yellow**, 1.0 Å flat well) to the clusters minimized with classified restraints (**green**).

TABLES

Table S1. Data collection and refinement statistics

Crystal Data		
Space group		$P2_1$
Cell dimensions (Å; °)		38.160 x 99.939 x 79.423; $\beta = 96.59$
Wavelength (Å)		0.9801
No. of observed reflections		157346
No. of unique reflections!		38877
Resolution limits (Å)		31.1 – 2.0 (2.11 – 2.0)*
Completeness (%)		97.4 (87.1)*
Redundancy		4.0 (3.8)*
$^1R_{\text{merge}}$ (%)		9.8 (19.6)*
$^2R_{\text{pim}}$ (%)		5.7 (11.6)
$\ \sigma(I) \ $		10.3 (6.0)
$^1R_{\text{merge}} = \sum (I_i - I_{\text{mean}}) / \sum (I_i)$ $^2R_{\text{pim}(hkl)} = \sqrt{(1/(n-1)) \cdot \sum (I_i - I_{\text{mean}}) / \sum (I_i)}$		
* Values in parentheses are for the data in the outermost shell.		
Refinement Statistics		
3R -factor (%)		19.9
4 Free R -factor (%)		24.8
rmsd in bond lengths (Å)		0.010
rmsd in bond angles (°)		1.243
Ramachandran plot (%) 626 residues:	Most favored	94.2
	Additional allowed	5.3
	Generously allowed	0.6
	Disallowed	0.0
G-factors: Dihedrals: -0.19 Covalent: 0.50 Overall: 0.07		
$^3R\text{-factor} = \sum (F_{\text{obs}} - k F_{\text{calc}})^2$ $^4\text{Free } R\text{-factor} = \sum (F_{\text{obs}} - k F_{\text{calc}}) / \sum_t F_{\text{obs}} $ where t = test set		

Table S2. RMSD summary of the negative control MD refinements without application of noe restraints for one structure from each MMP group (MMP12: ref, MMP9: 1GKC, MMP13: 1ZTQ and MMP1: 1FBL). The controls consistently show that the restraints are responsible for steering the ligand and are not a result of a general MD minimization.

	xplor_CN	xplor_N	autodock	control
MMP12 (ref)	0,43	0,45	1,22	0,71
	0,43	0,45	0,76	0,69
	0,57	0,59	1,62	1,29
MMP9 (1GKC)	0,68	0,69	2,17	1,94
	0,67	0,69	2,01	1,87
	0,73	0,68	1,12	1,29
MMP13 (1ZTQ)	0,74	0,74	2,07	1,54
	2,31	2,33	4,91	4,96
	2,17	2,17	3,05	3,1
MMP1 (1FBL)	0,64	0,63	0,79	0,96
	0,56	0,56	1,23	1,13
	1,23	1,23	1,46	1,53

Table S3. RMSD and nOe energy summary of the three negative control MD refinements for the Reference MMP12 structure. Weak, medium, strong classified restraints were compared to integrated restraints and relaxed integrated restraints. The accuracy is not improved when using integrated restraints and relaxation of the restraints energies does not lead it in the right direction.

	Classes	Integrated normal	Integrated relaxed
nOe energy	91	746	306
	88	717	282
	89	753	309
RMSD	1,10	1,52	1,57
	0,40	1,10	1,16
	1,10	1,52	1,57

FIGURES

(A)

(B)

(C)

(D)

(E)

(F)

(G)

(H)

(I)

(J)

(K)

(L)

(M)

(N)

(O)

Figure S1. RMSD plots of all docked crystal structures (A-O) using only HN contacts (left column) and both HN and HC contacts (right column).

Figure S2. RMSD vs. nOe energy plot of the test calculation comparing the three clusters minimized with integrated restraints based on an arbitrary reference distance (**red**, 0.5 Å flat well and **yellow**, 1.0 Å flat well) to the clusters minimized with classified restraints (**green**).