Supporting Information

Cosolvent sites based discovery of *Mycobacterium tuberculosis* Protein Kinase G inhibitors

Osvaldo Burastero,^{1,2,3,¶} Lucas A. Defelipe,^{1,2,3,¶} Gabriel Gola,^{4,5} Nancy L. Tateosian,^{1,2} Elias D. Lopez,^{1,2} Camila Belen Martinena,^{1,2} Juan Pablo Arcon,^{1,2} Martín Dodes Traian,^{1,2} Diana E.Wetzler,^{1,2} Isabel Bento,³ Xavier Barril,^{6,7} Javier Ramirez,^{4,5} Marcelo A. Marti,^{1,2} Maria M. Garcia-Alai,³ Adrián G. Turjanski^{1,2*}

[¶]Both authors contributed equally to this work

1 Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EGA, Buenos Aires, Argentina.

2 Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EGA, Buenos Aires, Argentina.

3 European Molecular Biology Laboratory Hamburg, Notkestrasse 85, D-22607, Hamburg, Germany.

4 Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EGA, Buenos Aires, Argentina.

5 Unidad de Microanálisis y Métodos Físicos Aplicados a Química Orgánica (UMYMFOR), Facultad de Ciencias Exactas y Naturales, C1428EGA, Universidad de Buenos Aires. CONICET

6 Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, Barcelona 08010, Spain

7 Faculty of Pharmacy and Institute of Biomedicine (IBUB), University of Barcelona, Av.Joan XXIII 27-31, Barcelona 08028, Spain

*Correspondence should be addressed to AGT (adrian@qb.fcen.uba.ar)

1. Figures



Figure S1. Schematic representation of PknG's active site bound to Ligand B1. Dashed lines indicate hydrogens bond while lines around atoms and residues indicate hydrophobic contacts. Black spheres are Carbon atoms, blue spheres Nitrogen atoms, red spheres Oxygen atoms and pink spheres Fluor atoms. Image was generated with LigPlot Plus 2.2.5.¹ Based upon a Kinase ATP pocket description from Traxler *et al.*, 1999.²



Figure S2. PknG kinase activity after incubation with the selected fragments at 2 mM (A), or the selected known kinase inhibitors (GSK kinase inhibitors set) at 40 μ M (B). One and zero units of kinase activity represent respectively 100 % and 0 % of inhibition. Associated Simplified Molecular Input Line Entry System (SMILES) codes of the hits ('F6' and 'gsk4') are provided in the Supplementary Material. Gsk4 targets IkappaB kinase 1 (IKK1, also known as IKK- α).³





Figure S3. Description of the SMARTS pattern used in the tethered docking. Picture was created using the SMARTSviewer online app (<u>https://smarts.plus/smartsview</u>, Copyright: ZBH - Center for Bioinformatics Hamburg).

AutodockBias docked pose superimposed with the solvent sites

Three active compounds (% inhibtion > 50 in end-point assay)



Four (out of sixteen) inactive compounds



Figure S4. Some examples of the final poses of the twenty compounds tested *in vitro* for kinase inhibition after the cosolvents sites-based docking protocol (Section "Biased runs of the compounds obtained after the tethered docking").



Figure S5. Binding affinity estimation from Intrinsic Tryptophan Quenching spectroscopy. Panels A to F: fluorescence versus ligand concentration for compounds T1, T2, S1, S2, B1 and AX20017.



Figure S6. Binding affinity estimation from Intrinsic Tryptophan Quenching spectroscopy. Panels A and B: fluorescence versus ligand concentration for compounds B3 and B4.



Figure S7. Binding affinity estimation from isothermal titration calorimetry for ligand B1 (panel A) and T1 (panel B). For B1, the normalized heats of injection were fitted using only two parameters: DH (enthalpy of binding) and K_d (equilibrium dissociation constant). For T1, it was necessary to add a parameter to correct the ligand concentration (similar to SEDPHAT⁴).



Figure S8. Apparent binding affinity estimation from fluorescence-based melting curves. A) and B) Fraction unfolded versus ligand concentration at 44°C for ligands T1 and B1. C) and D) Observed melting temperature versus ligand concentration for ligands T1 and B1.



Figure S9. Up) Example of one autoradiographic plate showing the decrease of GarA phosphorylation upon addition of Ligand T1. Lanes 1 to 5, and 6 to 10: compound T1 at 50, 5.6, 1.8, 0.6 and 0 μ M. The two bands correspond to GarA, the lower band is His-tagged GarA and the upper one is GarA cut by proteolytic enzymes during purification (confirmed by mass spectrometry). Below) At each ligand concentration (different from zero), the percentage of activity was calculated as one hundred times the quotient between the corresponding band intensity and the band intensity of the protein without ligand. Half inhibitory concentration (IC₅₀) was fitted from the activity versus ligand concentration curve using the equation $Activity(x) = minAct + (100 - minAct) / (1 + 10^{(x-log_{10}(lC_{50}))})$

where x is the logarithm of the ligand concentration, and *minAct* is the minimum observed activity (set to 14). The confidence interval was estimated using the asymptotic t-based approach with a 95 % confidence level.



Figure S10. Thermal shift assay of PknG Δ TPR Δ 73 mutants I292G, V235P and Y234L at different concentrations of B1 (A) or S1 (B) at 100 and 6.25 μ M. The fluorescence (350 nm) raw curves were min-max normalized and smoothed using a median filter with a 2 degrees window. Analysis was done in the MoltenProt online app (spc.embl-hamburg.de). Far UV Circular dichroism experiments were done to confirm that there was not a complete structure loss for mutants V235P and Y234L.



Figure S11. Ligand B1 bound at the ATP binding site of PknG (PDBID 7Q52). Both panels show different views of the Ligand B1 at the ATP binding site with the $(2F_o-F_c)$ maps contoured at 1σ level.



Figure S12. Predicted (violet) versus crystallized structure (pink) for compound B1 using a similar protocol as the one used for the fragments and kinase inhibitors. Briefly, we performed 100 RDock docking runs using as receptor the PDB 4y12 and a pharmacophore restraint located at point (-7.188, 7.214, -27.796) with a radius of tolerance of 0.5 Å and restrain type 'Acc'. The resulting docking poses were clustered based on RMSD, i.e., for each member in a cluster, there is at least one other member (in the same cluster) such that the heavy-atom based RMSD between these two members is less or equal than 1.1 Å. For each cluster, a representative member was chosen by selecting the one with the lowest mRMSD, where mRSMD is the mean of the RMSD against all the other members (in the same cluster). The final pose (shown in this Figure) is the representative member from the lowest energy cluster having more than 10 members.



Figure S13. Intracellular survival of *Mycobacterium tuberculosis* H37Rv (*Mbt*H37Rv) in human macrophages treated with B1 or T1 compounds. Macrophages derived from THP-1 cells line (1×10⁶ cells/mL) were infected with *Mtb*H37Rv (MOI: 10). After 2 h of infection, the culture medium was replaced and cells were cultured with compounds solution (10 and 1 μ M) or RIF (12.5 μ M) for 24 hs. Then, cells were washed and lysed for mycobacterial colony-forming units (CFU/mL) determination. Data are presented as means of bacterial viability CFU/mL \pm standard error of the mean (SEM), * $p \le 0.05$, ** $p \le 0.01$. *P* values were calculated using Welch's ANOVA and the post-hoc dunnett T3 multiple comparison tests. The *p* values were adjusted using the Bonferroni-Holmberg correction. Only the statistical significance of the comparisons against the control is shown.



Figure S14. Intracellular survival of *Mycobacterium tuberculosis* H37Rv (*Mbt*H37Rv) in human macrophages treated with B1, T1, T2, T3, or AX20017 at 10 μ M. Macrophages derived from THP-1 cell lines (1×10⁶ cells/mL) were infected with *Mtb*H37Rv (harvested from plates, MOI: 10). After 2 h of infection, the culture medium was replaced and cells were cultured with compound solution or RIF (12.5 μ M) for 24 hs. Then, cells were washed and lysed for mycobacterial colony-forming units (CFU/mL) determination. Data are presented as means of bacterial viability CFU/mL ± standard error of the mean (SEM). Experiments were done in triplicate.



Figure S15. HPLC traces for compound B1. Purity was > 98 %.



Figure S16. HPLC traces for compound T1. Purity was > 99 %.



Figure S17. HPLC traces for compound S1. Purity was > 95 %.



Figure S18. HPLC traces for compound S2. Purity was > 97 %.



Figure S19. HPLC traces for compound AX20017. Purity was > 96 %.



Figure S20. HPLC traces for compound T2. Purity was > 96 %.



Figure S21. HPLC traces for compound T3. Purity was > 84 %.

2. Tables

Table S1. Data Collection and Refinement Statistics.* Values in parenthesis refer to the highest resolution shell (2.434 - 2.35)

Data collection	
X-ray diffraction source	P14, PETRAIII
Wavelength (Å)	0.9763
Resolution range	51.45 - 2.35 (2.434 - 2.35) *
Space group	C2

Unit cell (Å, °)	77.327 37.015 106.78, 105.5
Total reflections	80226 (8138)
Unique reflections	12265 (1205)
Multiplicity	6.5 (6.8)
Completeness (%)	98.23 (97.80)
Mean I/sigma(I)	11.15 (1.10)
R-merge	0.090 (1.612)
CC1/2	0.998 (0.657)
Refinement	
Reflections used in refinement	12228 (1201)
Reflections used for R-free	625 (59)
R-work	0.224 (0.445)
R-free	0.267 (0.366)
Number of non-hydrogen atoms	2464
macromolecules	2410
Ligands/Na/Fe(II)	22/2/1
solvent	29
Protein residues	312
RMS(bonds)	0.13
RMS(angles)	2.9
Ramachandran favored (%)	96.75
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.39
Clashscore	2.49
Average B-factor (Å2)	78.6
macromolecules	78.9
ligands	64.9
solvent	64.4

3. Methods

Experimental methods

Forward (F) and reverse (R) primers used for mutating $PknG\Delta TPR\Delta 73$.

Y234L

F. 5' GTCGGCTACATCGTGATGGAActcGTCGGCGGGCAATCGCTCAAACGC 3' R. 5' GCGTTTGAGCGATTGCCCGCCGACgagTTCCATCACGATGTAGCCGAC 3'

```
V235P
```

```
F. 5' GTCGGCTACATCGTGATGGAATACcccGGCGGGCAATCGCTCAAACGC 3'
R. 5' GCGTTTGAGCGATTGCCCGCCgggGTATTCCATCACGATGTAGCCGAC 3'
```

```
I292G
F. 5' GCTGACCGAGGAACAGCTCAAGCTGggtGACCTGGGCGCGGGTATCGCGGATC
3'
R. 5' GATCCGCGATACCGCGCCCAGGTCaccCAGCTTGAGCTGTTCCTCGGTCAGC
3'
```

Quality control of the protein sample (his-tagged Pkng Δ TPR Δ 73)

To verify the purity of the protein-containing fractions after passage through the HiLoad 16/600 Superdex 200 pg (Figure S22 A), we run a sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (Figure S22 B).



Figure S22. A) Absorbance at 280 nm versus elution volume. B) SDS-PAGE of the selected fractions. The first nine lanes (after the protein marker) correspond to fractions near the highest peak.

Only the fractions containing significant protein with a MW around 40 kDa were pooled together and concentrated up to 1-4 mg/ml. The protein sample integrity and homogeneity was then assessed with circular dichroism and mass spectrometry. These two experiments confirmed that we had a well folded and homogenous protein (Figure S23).



Figure S23. A) Circular dichroism of 0.6 mg/ml PknG Δ TPR Δ 73 measured in a ChirascanTM CD spectrometer (Applied Photophysics). Milidegrees were converted to molar ellipticity by applying the formula $MRE = m^{\circ} * M/(10 * L * C)$ where m° , M, L and C are respectively the millidegrees, the mean residue weight in g/mol (109.1), the path length in cm (0.1) and the concentration in g/L. B) MALDI-TOF spectrum showing a homogenous sample of PknG Δ TPR Δ 73. The peaks at 40838 Da and 20449 Da correspond to the single charged and double charged ion, respectively. The peaks at 81771 Da (~ 40838*2) and 123515 (~ 40838*3) seem to correspond to the singled charged ion of a dimeric and trimeric state. The peak at 27272 may correspond to the triply charged ion of the dimer state. The MALDI-TOF spectrum was acquired with a MALDI–TOF Voyager DE-STR mass spectrometer (Applied Biosystems). Data processing was done according to the workflow described in the MALDIquant R package.^{5,6} Briefly, raw intensities were square-root transformed and smoothed using a moving average. The baseline was removed using the "SNIP" method and the intensity was calibrated with the "median" method. Peaks were detected using a "half window size" of 50 points.

Purity control of compounds B2, B3 and B4

The purity of compounds B2, B3 and B4 was verified by Nuclear magnetic resonance (NMR) (\mathbb{R} Vitas-M Laboratory). In all cases, purity was confirmed to be > 95 %. A representative spectra is shown below.



Figure S24. NMR spectra provided by ® Vitas-M Laboratory of compound B4. The percentage of impurity was 2 %.

Computational methods

Biased runs of the compounds obtained after the tethered docking

PknG structure was downloaded from the PDB (4y12).⁷ Solvent and ligand molecules were removed, and standard protonation state at physiological pH was assigned to ionizable residues. The receptor was converted into PDBQT format with AutoDockTools (prepare_receptor4.py).⁸ Ligands were converted from SDF into PDBQT format with Open Babel.⁹ The pocket grid was centered at (-0.487, 7.638, -30.062), the spacing between grid points was set at 0.375 Å, and the even number of grid points was 88, 56 and 40 in the x, y and z axis, respectively. Energy maps were computed as usual inside the grid and later modified following the Solvent Site Biased Docking Method (SSBDM)¹⁰. Briefly, for each ligand atom that could be involved in hydrogen bond interactions (OA/NA/HD atom types), an extra energy term is added to the AutoDock scoring function according to eq. 1

$$\Delta G_{SSBDM} = \Delta G_{ConventionalDocking} - RT(\sum_{i=1}^{3} ln (PFP_i) * e^{\sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2} * R_{90,i}^{-1}})$$
(1)

were ΔG_{SSBDM} is the modified scoring function, $\Delta G_{ConventionalDocking}$ is the original scoring function, R is the gas constant, T is the temperature (298K) and the sum iterates over the ethanol-OH sites. PFP is the probe finding probability which was arbitrarily set to 90. (x, y, z) are the grid points coordinates, (x_i, y_i, z_i) are the solvent site coordinates, and $R_{90,i}$ is the solvent site dispersion parameter (arbitrarily set to 1). The term $e^{\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}*R_{90,i}^{-1}}$ implies that the energy decreases in a gaussian fashion. The

 e^{-1} implies that the energy decreases in a gaussian fashion. The solvent site centers were respectively (-7.496, 5.157, -27.644) and (-6.845, 9.576, -26.339) for the two hydrogen bond acceptor sites (OA/NA atom types), and (-7.191, 6.952, -27.825) for the hydrogen bond donor site (HD atom type). A positional correction was applied to reach ideal hydrogen bond geometries according to the protein residues.

A hydrophobic site was taken into account by creating for each ligand aromatic ring a new dummy atom located at the ring center and adding a new grid map for the dummy atom. For this solvent site, the *PFP*was arbitrarily set to 4.8 and the R_{90} to 1.2. The solvent site center was arbitrarily set at (-4.984, 6.339, -27.129).

For each ligand, we performed one hundred independent docking runs using the Lamarckian Genetic Algorithm with a maximum of 27,000 generations or 2,500,000 energy evaluations.¹¹ The resulting poses were clustered using a heavy atom-based root mean square deviation (rmsd) cutoff of 2 Å.

References

- Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: A Program to Generate Schematic Diagrams of Protein-Ligand Interactions. "*Protein Engineering, Design and Selection*." 1995, pp 127–134. https://doi.org/10.1093/protein/8.2.127.
- (2) Strategies toward the Design of Novel and Selective Protein Tyrosine Kinase Inhibitors. *Pharmacol. Ther.* **1999**, *82* (2-3), 195–206.
- (3) The Discovery of 2-Amino-3,5-Diarylbenzamide Inhibitors of IKK-α and IKK-β Kinases. *Bioorg. Med. Chem. Lett.* **2007**, *17* (14), 3972–3977.
- (4) Zhao, H.; Piszczek, G.; Schuck, P. SEDPHAT A Platform for Global ITC Analysis and Global Multi-Method Analysis of Molecular Interactions. *Methods*. 2015, pp 137–148. https://doi.org/10.1016/j.ymeth.2014.11.012.
- (5) Gibb, S.; Strimmer, K. MALDIquant: A Versatile R Package for the Analysis of Mass Spectrometry Data. *Bioinformatics* **2012**, *28* (17), 2270–2271.
- (6) Gibb, S.; Strimmer, K. Mass Spectrometry Analysis Using MALDIquant. *Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry*. 2017, pp 101–124. https://doi.org/10.1007/978-3-319-45809-0_6.
- (7) Lisa, M.-N.; Gil, M.; André-Leroux, G.; Barilone, N.; Durán, R.; Biondi, R. M.; Alzari, P. M. Molecular Basis of the Activity and the Regulation of the Eukaryotic-like S/T Protein Kinase PknG from Mycobacterium Tuberculosis. *Structure* **2015**, *23* (6), 1039–1048.

- (8) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J. Comput. Chem. 2009, 30 (16), 2785–2791.
- (9) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3*, 33.
- (10) Arcon, J. P.; Defelipe, L. A.; Modenutti, C. P.; López, E. D.; Alvarez-Garcia, D.; Barril, X.; Turjanski, A. G.; Martí, M. A. Molecular Dynamics in Mixed Solvents Reveals Protein-Ligand Interactions, Improves Docking, and Allows Accurate Binding Free Energy Predictions. *J. Chem. Inf. Model.* **2017**, *57* (4), 846–863.
- (11) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *Journal of Computational Chemistry*. 1998, pp 1639–1662. https://doi.org/10.1002/(sici)1096-987x(19981115)19:14<1639::aid-jcc10>3.0.co;2-b.