Supporting Information

Selecting a Stable Solid Form of Remdesivir Using Microcrystal Electron Diffraction and Crystal Structure Prediction

Sivakumar Sekharan,*† Xuetao Liu,†b Zhuocen Yang,b Xiang Liu,b Li Deng,b Shigang Ruan,b Yuriy Abramov,a GuangXu Sun,b Sizhu Li,b Tian Zhou,b Baime Shi,b Qun Zeng,b Qiao Zeng,b Chao Chang,b Yingdi Jinb and Xuekun Shib

Table of Contents (total pages 24)

I. Experimental Methods and Analysis

II. Crystal Structure Prediction Workflow

III. References
I. Experimental Methods and Analysis

Sample preparation

The active pharmaceutical ingredient (API) powder of remdesivir forms II and IV were purchased from Guangzhou Dreampharm CLT., with experimental form II X-ray powder diffraction (XRPD) matching the patent publication.\textsuperscript{1} We used the solvent information from the patent publication to optimize the crystallization steps and obtain high crystallinity sample of experimental form IV. The EMS (Electron Microscopy Sciences) lacey carbon grid with super thin carbon layer coated was directly dropped into the sample crystal powder. The extra sample attached to the grid was blown away using the washing ear ball.

Data collection

The sample grids were quickly frozen in liquid nitrogen, assembled in Gatan cryo-transfer holder 626 and then loaded into JEOL F200 equipped with the Gatan OneView detector, respectively.\textsuperscript{2} Diffraction tilt series of crystals were collected using the continuous rotation method, whose rate was 0.5 degree/second.\textsuperscript{3-6} The detector distance used was 890 mm and the tilt range was 40 degrees. The total dose per tilt series was 1 electron and the exposure time was 80 seconds.

Data analysis and structure determination

Diffraction frames collected from 10 crystals of remdesivir form II and 25 crystals of remdesivir form IV have been individually indexed and integrated with XDS, then merged to a single data set for each form, respectively.\textsuperscript{2,7,8} The high-quality diffraction data collected from remdesivir form II enabled us to solve its crystal structure by direct method with SHELXT. The data quality of remdesivir form IV is relatively poorer but can still be solved with SHELXT.\textsuperscript{9} The solved models of both forms were then refined against the merged data set with SHELXL as in the pipeline of the structure refinement in the single-crystal X-ray diffraction.\textsuperscript{10,11} Each model was refined by using full matrix least squares on $F^2$ minimization. All non-hydrogen atoms were refined anisotropically,
and the positions of all hydrogen atoms were calculated geometrically and refined using the riding model.

**Figure S1.** A typical diffraction frame for form II.
Figure S2. A typical diffraction frame for form IV.
Table S1. Crystallographic data and refinement parameters of remdesivir form II

| Sample code     | remdesivir form II                        |
|-----------------|-------------------------------------------|
| Empirical formula | C<sub>27</sub>H<sub>35</sub>N<sub>6</sub>O<sub>8</sub>P |
| Formula weight  | 601.55 g·mol<sup>-1</sup>                 |
| Temperature     | cryo-temperature                          |
| Wavelength      | 0.02508 Å                                  |
| Crystal system, space group | Monoclinic, P2₁ (No.4)                      |
| Unit cell dimensions | a = 10.21(4) Å  
                        | b = 12.49(14) Å  
                        | c = 10.85(10) Å  
                        | α = 90°                     
                        | β = 100.7(6)°                
                        | γ = 90°                     |
| Volume          | 1495(22) Å<sup>3</sup>                    |
| Z, calculated density | 2, 1.336 g/cm<sup>3</sup>                |
| F(000)          | 234                                        |
| Resolution for data collection | 0.9 Å                               |
| Limiting indices | -11 ≤ h ≤ 11 
                        | -13 ≤ k ≤ 13                        
                        | -12 ≤ l ≤ 11                        |
| Reflection collected / Independent reflections | 11574 / 3562 [R<sub>int</sub> = 0.2297]          |
| Data completeness | 91%                                            |
| Refinement method | Full-matrix least-squares on F<sup>2</sup> |
| Data / restraints / parameters | 3562 / 1 / 360                          |
| Goodness-of-fit on F<sup>2</sup> | 1.042                                  |
| Final R indices [I ≥ 2sigma(I)] | R<sub>1</sub> = 0.1609                  |
| Final R indices [all data] | R<sub>1</sub> = 0.2331, wR<sub>2</sub> = 0.4225 |
| Largest diff. peak and hole | 0.242 / -0.226 e.Å<sup>-3</sup>               |
Table S2. Crystallographic data and refinement parameters of remdesivir form IV

| Sample code | Remdesivir form IV |
|-------------|--------------------|
| Empirical formula | C\textsubscript{27} H\textsubscript{35} N\textsubscript{6} O\textsubscript{8} P |
| Formula weight | 601.55 g·mol\textsuperscript{-1} |
| Temperature | cryo-temperature |
| Wavelength | 0.02508 Å |
| Crystal system, space group | Monoclinic, P2\textsubscript{1} (No.4) |
| Unit cell dimensions | 
\begin{align*}
a &= 10.03(7) \text{ Å} \\
b &= 12.20(20) \text{ Å} \\
c &= 11.44(18) \text{ Å} \\
\alpha &= 90° \\
\beta &= 104.4(7)° \\
\gamma &= 90° 
\end{align*} |
| Volume | 1356(33) Å\textsuperscript{3} |
| Z, calculated density | 2, 1.476 g/cm\textsuperscript{3} |
| F(000) | 235 |
| Resolution for data collection | 0.955 Å |
| Limiting indices | 
\begin{align*}
-10 \leq h \leq 10 \\
-12 \leq k \leq 12 \\
-11 \leq l \leq 11 
\end{align*} |
| Reflection collected / Independent reflections | 19547 / 3133 [R\textsubscript{int} = 0.4016] |
| Data completeness | 96% |
| Refinement method | Full-matrix least-squares on F\textsuperscript{2} |
| Data / restraints / parameters | 31330 / 8 / 360 |
| Goodness-of-fit on F\textsuperscript{2} | 1.033 |
| Final R indices [I ≥ 2sigma(I)] | R\textsubscript{1} = 0.1593 |
| Final R indices [all data] | R\textsubscript{1} = 0.2346, wR\textsubscript{2} = 0.4042 |
| Largest diff. peak and hole | 0.171 / -0.197 e.Å\textsuperscript{-3} |

The structures with unit-cell constants obtained directly from MicroED and the structures with unit-cell constants adjusted with XRPD indexing results are compared. The calculated root-mean-square distance (RMSD) between the MicroED and CSP structures of both remdesivir form II and form IV are given in Table S3, with the low values strongly indicating that the experimental and predicted structures are almost
identical. Adjusting the unit-cell constants by XRPD indexing results can further reduce the RMSD values by nearly 50%.

Table S3. The RMSD values between MicroED and CSP structures.

| Crystal Structures               | Before adjusting unit-cell constants (Å) | After adjusting unit-cell constants of both MicroED and CSP structures to XRPD indexing results (Å) |
|----------------------------------|------------------------------------------|--------------------------------------------------------------------------------------------------|
| remdesivir form II vs. X2        | 0.441 (15/15)                            | 0.298 (15/15)                                                                                   |
| remdesivir form IV vs. X1        | 0.368 (15/15)                            | 0.183 (15/15)                                                                                   |

Table S4. Positional parameters and equivalent isotropic displacement parameters of non-H atoms in remdesivir form II unit-cell.

| ATOM  | X       | Y       | Z       | U_{eq}   |
|-------|---------|---------|---------|----------|
| N2    | 0.6723(6) | 0.8010(12) | 0.7250(10) | 0.064(5) |
| N3    | 0.7390(8) | 0.7847(13) | 0.8476(9)  | 0.069(5) |
| C00K  | 0.8775(9) | 0.7846(15) | 0.8749(9)  | 0.079(7) |
| N00B  | 0.9494(6) | 0.8008(15) | 0.7797(11) | 0.067(6) |
| C004  | 0.8828(9) | 0.8171(14)| 0.6572(10) | 0.076(7) |
| C00H  | 0.7442(9) | 0.8172(14)| 0.6298(8)  | 0.068(6) |
| C3    | -0.0802(9)| 0.4431(13)| 0.7948(15) | 0.084(9) |
| C00V  | -0.1023(13)| 0.4879(13)| 0.6754(14) | 0.084(7) |
| C00Q  | -0.2257(16)| 0.5321(16)| 0.6253(14) | 0.095(8) |
| C00R  | -0.3272(12)| 0.5315(18)| 0.6946(17) | 0.118(11)|
| C015  | -0.3052(11)| 0.4867(16)| 0.8140(17) | 0.097(9) |
| C00W  | -0.1817(12)| 0.4425(14)| 0.8641(14) | 0.117(13)|
| P001  | 0.1564(8)  | 0.3920(10)| 0.7947(10) | 0.056(3) |
| O1    | 0.2621(17) | 0.877(2)  | 0.664(2)   | 0.090(7) |
| N1    | 0.9505(19) | 0.826(2)  | 0.572(2)   | 0.084(7) |
| C1    | 0.5302(16) | 0.823(2)  | 0.567(2)   | 0.079(8) |
| O2    | 0.0261(14) | 0.3992(19)| 0.850(2)   | 0.085(6) |
| O002  | 0.2144(18) | 0.4988(18)| 0.8403(18) | 0.077(6) |
| C2    | 0.4591(16) | 0.781(2)  | 0.762(2)   | 0.067(6) |
| O3    | 0.053(3)   | 0.160(2)  | 0.867(2)   | 0.102(8) |
| O4    | 0.164(3)   | 0.039(5)  | 0.762(9)   | 0.28(4)  |
| C003  | 0.4760(15) | 0.8548(15)| 0.8672(19) | 0.047(4) |
| N4    | 0.245(2)   | 0.3011(16)| 0.8793(16) | 0.067(5) |
| C4    | 0.390(2)   | 0.151(3)  | 0.905(3)   | 0.104(10)|
| O5    | 0.4656(13) | 0.6829(16)| 0.8157(17) | 0.062(4) |
| O006  | 0.2020(18) | 0.786(2)  | 0.8656(19) | 0.103(9) |
| C00C  | 0.5527(14) | 0.7995(17)| 0.6840(17) | 0.049(5) |
| C00D  | 0.2636(17) | 0.196(2)  | 0.831(3)   | 0.090(8) |
| N00F  | 0.4949(16) | 0.915(2)  | 0.9450(19) | 0.070(6) |
Table S5. Anisotropic displacement factor coefficients for non-H atoms of remdesivir form II

| Atom | U_{11}   | U_{22}   | U_{33}   | U_{23}   | U_{13}   | U_{12}   |
|------|----------|----------|----------|----------|----------|----------|
| N2   | 0.054(9) | 0.066(15) | 0.074(12) | 0.041(10) | 0.015(9) | 0.006(9) |
| N3   | 0.021(6) | 0.096(17) | 0.093(14) | 0.000(11) | 0.013(7) | -0.007(8) |
| C00K | 0.049(9) | 0.08(2)   | 0.09(2)   | 0.044(14) | -0.016(10) | 0.001(11) |
| N00B | 0.064(9) | 0.062(15) | 0.078(12) | 0.046(10) | 0.020(9) | 0.031(9) |
| C004 | 0.028(8) | 0.09(2)   | 0.124(19) | 0.029(14) | 0.038(11) | -0.005(10) |
| C00H | 0.020(7) | 0.10(2)   | 0.073(13) | 0.019(11) | -0.008(8) | 0.011(9) |
| C3   | 0.044(11) | 0.08(2)   | 0.13(2)   | 0.077(16) | 0.033(12) | 0.009(10) |
| C00V | 0.073(13) | 0.09(2)   | 0.10(2)   | -0.020(14) | 0.030(13) | -0.009(13) |
| C00Q | 0.095(16) | 0.08(2)   | 0.11(2)   | -0.001(14) | 0.023(16) | 0.020(15) |
| C00R | 0.090(18) | 0.08(3)   | 0.19(3)   | -0.04(2) | 0.04(2) | 0.038(17) |
| C015 | 0.059(12) | 0.11(3)   | 0.13(2)   | 0.001(18) | 0.050(14) | 0.036(14) |
| C00W | 0.030(9) | 0.17(3)   | 0.15(3)   | -0.09(2) | 0.020(13) | -0.006(13) |
| P001 | 0.033(4) | 0.078(9) | 0.056(7) | -0.028(5) | 0.008(4) | -0.001(5) |
| O1   | 0.045(8) | 0.12(2)   | 0.095(15) | 0.025(13) | -0.009(9) | -0.006(11) |
| N1   | 0.052(9) | 0.11(2)   | 0.095(15) | -0.016(12) | 0.023(10) | -0.025(11) |
| C1   | 0.030(8) | 0.11(2)   | 0.082(17) | 0.030(13) | -0.017(9) | 0.006(10) |
| O2   | 0.031(7) | 0.089(16) | 0.143(18) | 0.010(12) | 0.040(9) | -0.004(9) |
| O002 | 0.073(10) | 0.080(17) | 0.069(13) | -0.019(9) | -0.009(9) | 0.027(10) |
| C2   | 0.029(8) | 0.065(19) | 0.099(16) | 0.031(13) | -0.009(9) | 0.013(9) |
| O3   | 0.031(9) | 0.11(2)   | 0.076(19) | 0.010(13) | 0.000(13) | 0.017(16) |
| C003 | 0.041(8) | 0.033(13) | 0.065(13) | -0.012(10) | 0.009(8) | -0.005(7) |
| N4   | 0.098(12) | 0.058(14) | 0.046(9) | -0.030(8) | 0.018(8) | -0.019(11) |
| C4   | 0.062(12) | 0.09(3)   | 0.15(3)   | 0.016(16) | 0.000(14) | 0.042(14) |
| O5   | 0.035(6) | 0.075(14) | 0.078(11) | 0.017(9) | 0.019(7) | 0.004(7) |
| O006 | 0.053(9) | 0.17(3)   | 0.091(14) | 0.047(14) | 0.015(10) | 0.071(14) |
| C00C | 0.029(8) | 0.066(14) | 0.057(12) | 0.012(9) | 0.018(8) | -0.012(8) |
### Table S6. Bond lengths of remdesivir form II

| Atom   | Atom   | Length (Å)   | Atom   | Atom   | Length (Å)   |
|--------|--------|--------------|--------|--------|--------------|
| N2     | C00C   | 1.218(18)    | C1     | C00C   | 1.28(3)      |
| N2     | N3     | 1.39         | C1     | N00B   | 1.43(3)      |
| N2     | C00H   | 1.39         | O002   | C00P   | 1.37(3)      |
| N3     | C00K   | 1.39         | C2     | O5     | 1.35(3)      |
| C00K   | N00B   | 1.39         | C2     | C00C   | 1.41(3)      |
| N00B   | C004   | 1.39         | C2     | C003   | 1.46(3)      |
| C004   | N1     | 1.26(2)      | C2     | C00L   | 1.55(2)      |
| C004   | C00H   | 1.39         | O3     | C00Y   | 1.21(4)      |
| C00H   | C00P   | 1.30(3)      | O3     | C012   | 1.46(4)      |
| C3     | O2     | 1.26(2)      | O4     | C00Y   | 1.34(6)      |
| C3     | C00V   | 1.39         | C003   | N00F   | 1.12(3)      |
| C3     | C00W   | 1.39         | N4     | C00D   | 1.44(4)      |
| C3     | P001   | 2.500(15)    | C4     | C00D   | 1.49(3)      |
| C00V   | C00Q   | 1.39         | O5     | C00N   | 1.45(3)      |
| C00V   | P001   | 2.97(2)      | O006   | C00G   | 1.31(4)      |
| C00Q   | C00R   | 1.39         | C00D   | C00Y   | 1.44(5)      |
| C00R   | C015   | 1.39         | C00G   | C00N   | 1.45(3)      |
| C015   | C00W   | 1.39         | C00G   | C00L   | 1.58(3)      |
| P001   | O00X   | 1.36(3)      | C00N   | C00U   | 1.42(3)      |
| P001   | O002   | 1.50(3)      | C00T   | C010   | 1.46(4)      |
| P001   | O2     | 1.565(19)    | C00T   | C013   | 1.68(8)      |
| P001   | N4     | 1.63(3)      | C00Z   | C010   | 1.30(3)      |
| P001   | C00U   | 2.50(3)      | C00Z   | C011   | 1.62(4)      |
| P001   | C00D   | 2.69(4)      | C010   | C012   | 1.56(4)      |
| O1     | C00L   | 1.24(3)      |        |        |              |

### Table S7. Bond angles of remdesivir form II

| Atom | Atom | Atom | Angle (°) | Atom | Atom | Atom | Angle (°) |
|------|------|------|-----------|------|------|------|-----------|
|      |      |      |           |      |      |      |           |
| C00C | N2  | N3  | C3   | P001 | C00V | 27.8(3) |
|------|-----|-----|------|------|------|---------|
| C00C | N2  | C00H | 111.3(12) | C00D | P001 | C00V | 137.5(7) |
| N3   | N2  | C00H | 120 | C00C | C1 | C00P | 108.1(17) |
| C00K | N3  | N2  | 120 | C3   | O2  | P001 | 123.8(18) |
| N3   | C00K | N00B | 120 | C00U | O002 | P001 | 120.1(16) |
| C004 | N00B | C00K | 120 | O5   | C2  | C00C | 115.2(18) |
| N1   | C004 | C00H | 121.5(13) | O5   | C2  | C003 | 104(2) |
| N1   | C004 | N00B | 118.4(13) | C00C | C2  | C003 | 112(2) |
| C00H | C004 | N00B | 120 | O5   | C2  | C00L | 102.5(17) |
| C00P | C00H | C004 | 134.1(12) | C00C | C2  | C00L | 115.9(18) |
| C00P | C00H | N2   | 105.9(12) | C003 | C2  | C00L | 106.2(19) |
| C004 | C00H | N2   | 120 | C00Y | O3  | C012 | 116(3) |
| O2   | C3  | C00V | 126.4(13) | N00F | C003 | C2  | 175(2) |
| O2   | C3  | C00W | 113.6(13) | C00D | N4  | P001 | 122.0(17) |
| C00V | C3  | C00W | 120 | C2   | O5  | C00N | 116.7(15) |
| O2   | C3  | P001 | 31.3(10) | N2   | C00C | C1  | 109.8(14) |
| C00V | C3  | P001 | 95.1(8)  | N2   | C00C | C2  | 122.1(17) |
| C00W | C3  | P001 | 144.9(7) | C1   | C00C | C2  | 128.0(16) |
| C00Q | C00V | C3   | 120 | C00Y | C00D | N4  | 113(2) |
| C00Q | C00V | P001 | 177.1(5) | C00Y | C00D | C4  | 114(3) |
| C3   | C00V | P001 | 57.1(7)  | N4   | C00D | C4  | 108(2) |
| C00R | C00Q | C00V | 120 | C00Y | C00D | P001 | 103.5(17) |
| C00Q | C00R | C015 | 120 | N4   | C00D | P001 | 30.9(11) |
| C00W | C015 | C00R | 120 | C4   | C00D | P001 | 135(2) |
| C015 | C00W | C3   | 120 | O006 | C00G | C00N | 115(2) |
| O00X | P001 | O002 | 118.1(16) | O006 | C00G | C00L | 106(3) |
| O00X | P001 | O2   | 119.5(13) | C00N | C00G | C00L | 107.3(16) |
| O002 | P001 | O2   | 97.6(12)  | O1   | C00L | C2  | 119(2) |
| O00X | P001 | N4   | 109.0(14) | O1   | C00L | C00G | 119(2) |
| O002 | P001 | N4   | 106.9(13) | C2   | C00L | C00G | 100.3(15) |
| O2   | P001 | N4   | 104.3(13) | C00U | C00N | C00G | 113(2) |
| O00X | P001 | C00U | 95.7(13)  | C00U | C00N | O5  | 106.8(16) |
| O002 | P001 | C00U | 28.5(9)   | C00G | C00N | O5  | 103(2) |
| O2   | P001 | C00U | 125.7(12) | C00H | C00P | C1  | 104.6(18) |
| N4   | P001 | C00U | 100.8(11) | C010 | C00T | C013 | 111(3) |
| O00X | P001 | C3   | 100.1(10) | O002 | C00U | C00N | 105.3(18) |
| O002 | P001 | C3   | 95.4(10)  | O002 | C00U | P001 | 31.5(10) |
| O2   | P001 | C3   | 24.8(9)   | C00N | C00U | P001 | 136.5(14) |
| N4   | P001 | C3   | 127.9(10) | O3   | C00Y | O4  | 130(4) |
| C00U | P001 | C3   | 118.5(9)  | O3   | C00Y | C00D | 121(3) |
| O00X | P001 | C00D | 83.6(14)  | O4   | C00Y | C00D | 109(3) |
| O002 | P001 | C00D | 129.6(11) | C010 | C00Z | C011 | 112(3) |
| O2   | P001 | C00D | 110.4(11) | C00Z | C010 | C00T | 122(3) |
| N4   | P001 | C00D | 27.1(9)   | C00Z | C010 | C012 | 114(3) |
| C00U | P001 | C00D | 113.7(9)  | C00T | C010 | C012 | 104(2) |
| C3   | P001 | C00D | 126.9(7)  | O3   | C012 | C010 | 106(3) |
| O00X | P001 | C00V | 76.9(11)  | N4   | P001 | C00V | 152.2(9) |
| O002 | P001 | C00V | 92.8(11)  | C00U | P001 | C00V | 105.6(9) |
| O2   | P001 | C00V | 52.7(10)  |      |      |      |      |
Table S8. Torsion angles of remdesivir form II

| Atom | Atom | Atom | Atom | Angle (°) | Atom | Atom | Atom | Atom | Angle (°) |
|------|------|------|------|----------|------|------|------|------|----------|
| C00C | N2   | N3   | C00K | 177(2)   | C003 | C2   | C00C | N2   | 55(3)    |
| C00K | N00B | C004 | N1   | -176.4(17)| C00L | C2   | C00C | N2   | 176(2)   |
| N1   | C004 | C00H | C00P | -7(3)    | O5   | C2   | C00C | C1   | 120(3)   |
| N00B | C004 | C00H | C00P | 177(2)   | C003 | C2   | C00C | C1   | -121(3)  |
| N1   | C004 | C00H | N2   | 176.3(17)| C00L | C2   | C00C | C1   | 1(4)     |
| C00C | N2   | C00H | C00P | 4.7(18)  | P001 | N4   | C00D | C00Y | -78(3)   |
| N3   | N2   | C00H | C00P | -177.9(18)| P001 | N4   | C00D | C4   | 155.9(19)|
| C00C | N2   | C00H | C00P | -177.4(17)| O5   | C2   | C00L | O1   | 163(2)   |
| O2   | C3   | C00V | C00Q | 179(2)   | C00C | C2   | C00L | O1   | -71(3)   |
| P001 | C3   | C00V | C00Q | -179.6(10)| C003 | C2   | C00L | O1   | 54(3)    |
| O2   | C3   | C00V | P001 | -1.4(16) | O5   | C2   | C00L | C00G | 31(3)    |
| C00W | C3   | C00V | P001 | 179.6(10)| C00C | C2   | C00L | C00G | 157(2)   |
| O2   | C3   | C00W | C015 | -179(2)  | C003 | C2   | C00L | C00G | -78(2)   |
| P001 | C3   | C00W | C015 | 179.3(18)| O006 | C00G | C00L | O1   | -35(3)   |
| C00V | C3   | O2   | P001 | 3(3)     | C00N | C00G | C00L | O1   | -158(2)  |
| C00W | C3   | O2   | P001 | -178.3(14)| O006 | C00G | C00L | C2   | 97(2)    |
| O00X | P001 | O2   | C3   | -42(3)   | C00N | C00G | C00L | C2   | -26(3)   |
| O002 | P001 | O2   | C3   | 86(2)    | O006 | C00G | C00N | C00U | 140(2)   |
| N4   | P001 | O2   | C3   | -164(2)  | C00L | C00G | C00N | C00U | -103(2)  |
| C00U | P001 | O2   | C3   | 81(3)    | O006 | C00G | C00N | O5   | -105(2)  |
| C00D | P001 | O2   | C3   | -136(2)  | C00L | C00G | C00N | O5   | 12(3)    |
| C00V | P001 | O2   | C3   | -1.3(14) | C2   | O5   | C00N | C00U | 130(2)   |
| O00X | P001 | O002 | C00U | -42(2)   | C2   | O5   | C00N | C00G | 10(3)    |
| O2   | P001 | O002 | C00U | -171.0(19)| C004 | C00H | C00P | C1   | 179.2(16)|
| N4   | P001 | O002 | C00U | 81(2)    | N2   | C00H | C00P | C1   | -2(2)    |
| C3   | P001 | O002 | C00U | -146.1(17)| C00C | C1   | C00P | C00H | -1(3)    |
| C00D | P001 | O002 | C00U | 64(2)    | P001 | O002 | C00U | C00N | 174.1(17)|
| C00V | P001 | O002 | C00U | -118.3(18)| C00G | C00N | C00U | O002 | -71(3)   |
| O00X | P001 | N4   | C00D | -22(2)   | O5   | C00N | C00U | O002 | 175.8(17)|
| O002 | P001 | N4   | C00D | -150.3(18)| C00G | C00N | C00U | P001 | -67(3)   |
| O2   | P001 | N4   | C00D | 106.9(19)| O5   | C00N | C00U | P001 | -179.7(17)|
| C00U | P001 | N4   | C00D | -121.7(17)| C012 | O3   | C00Y | O4   | 8(8)     |
| C3   | P001 | N4   | C00D | 98.5(18) | C012 | O3   | C00Y | C00D | 174(3)   |
| C00V | P001 | N4   | C00D | 76(2)    | N4   | C00D | C00Y | O3   | -1(4)    |
| C00C | C2   | O5   | C00N | -154.2(18)| C4   | C00D | C00Y | O3   | 122(3)   |
| C003 | C2   | O5   | C00N | 83(2)    | P001 | C00D | C00Y | O3   | -32(3)   |
| C00L | C2   | O5   | C00N | -27(3)   | N4   | C00D | C00Y | O4   | 168(5)   |
| N3   | N2   | C00C | C1   | 177.1(17)| C4   | C00D | C00Y | O4   | -69(5)   |
| C00H | N2   | C00C | C1   | -6(2)    | P001 | C00D | C00Y | O4   | 137(5)   |
| N3   | N2   | C00C | C2   | 1(3)     | C011 | C00Z | C010 | C010 | -48(4)   |
| C00H | N2   | C00C | C2   | 177.9(18)| C011 | C00Z | C010 | C012 | 78(3)    |
| C00P | C1   | C00C | N2   | 4(3)     | C013 | C00T | C010 | C00Z | -66(5)   |
| C00P | C1   | C00C | C2   | -179(2)  | C013 | C00T | C010 | C012 | 164(3)   |
| O5   | C2   | C00C | N2   | -64(3)   | C00Y | O3   | C012 | C010 | -116(3)  |
| C00T | C010 | C012 | O3   | -177(3)  | C00Z | C010 | C012 | O3   | 48(3)    |
Table S9. Positional parameters and equivalent isotropic displacement parameters of non-H atoms in remdesivir form IV unit-cell.

| ATOM  | X       | Y       | Z       | U<sub>eq</sub> |
|-------|---------|---------|---------|---------------|
| N3    | 0.3408(8) | 0.8373(12) | 0.2722(13) |
| C00C  | 0.2710(11) | 0.8494(12) | 0.3622(10)  |
| C5    | 0.1280(11) | 0.8532(11) | 0.3324(11)  |
| N00D  | 0.0548(8)  | 0.8449(13) | 0.2124(12)  |
| C6    | 0.1246(10) | 0.8328(15) | 0.1223(10)  |
| N5    | 0.2676(11) | 0.8290(14) | 0.1522(12)  |
| C00F  | 1.0529(16) | 0.4647(16) | 0.229(2)     |
| C1    | 1.0793(19) | 0.5100(19) | 0.344(2)     |
| C2    | 1.196(2)   | 0.5742(18) | 0.387(2)     |
| C00L  | 1.2858(16) | 0.5931(17) | 0.314(2)     |
| C00V  | 1.2594(18) | 0.548(2)   | 0.199(2)     |
| C00I  | 1.143(2)   | 0.4836(19) | 0.1567(18)  |
| O002  | 0.758(2)   | 0.509(2)   | 0.166(3)     |
| N2    | 0.0646(17) | 0.8668(15) | 0.411(3)     |
| O3    | 0.692(4)   | 0.098(2)   | 0.293(3)     |
| C3    | 0.651(4)   | 0.1727(18)| 0.233(6)     |
| O4    | 0.530(2)   | 0.225(6)   | 0.226(3)     |
| O004  | 0.8142(18) | 0.368(2)   | 0.337(2)     |
| N4    | 0.522(2)   | 0.9759(17)| 0.055(3)     |
| C4    | 0.5422(16) | 0.915(2)  | 0.128(2)     |
| O5    | 0.5271(17) | 0.7209(18)| 0.156(3)     |
| C8    | 0.369(2)   | 0.861(2)  | 0.462(3)     |
| C9    | 0.4891(18) | 0.8548(18)| 0.425(4)     |
| C10   | 0.146(3)   | 0.204(5)  | 0.431(6)     |
| C11   | 0.650(3)   | 0.561(2)  | 0.181(7)     |
| C009  | 0.699(3)   | 0.208(3)  | 0.115(3)     |
| C00A  | 0.753(2)   | 0.7491(16)| 0.199(3)     |
| C00H  | 0.697(2)   | 0.810(2)  | 0.296(2)     |
| C00N  | 0.5543(13) | 0.8205(19)| 0.226(2)     |
| C00O  | 0.6378(16) | 0.662(2)  | 0.138(3)     |
| C00R  | 0.195(4)   | 0.106(2)  | 0.124(4)     |
| C00S  | 0.4651(17) | 0.835(2)  | 0.298(2)     |
| O00T  | 0.7936(18) | 0.805(2)  | 0.121(2)     |
| C00U  | 0.277(4)   | 0.159(2)  | 0.405(5)     |
| C00W  | 0.334(3)   | 0.217(3)  | 0.310(3)     |
| C00X  | 0.467(3)   | 0.158(2)  | 0.312(3)     |
| C00Y  | 0.830(4)   | 0.130(4)  | 0.113(5)     |
| P1    | 0.8110(10) | 0.3964(10)| 0.2137(14)   |
| O1    | 0.9527(15) | 0.3922(16)| 0.174(2)     |
| N1    | 0.7230(19) | 0.3106(16)| 0.120(2)     |
| O2    | 0.771(2)   | 0.8963(18)| 0.342(4)     |
| C010  | 0.252(3)   | 0.2160(18)| 0.178(4)     |
### Table S10. Anisotropic displacement factor coefficients for non-H atoms of remdesivir form IV

| Atom | U₁₁   | U₂₂   | U₃₃   | U₂₃   | U₁₃   | U₁₂   |
|------|-------|-------|-------|-------|-------|-------|
| N3   | 0.085(18) | 0.074(13) | 0.021(11) | -0.003(10) | -0.022(11) | 0.016(11) |
| C00C | 0.087(18) | 0.069(15) | 0.08(2) | 0.020(14) | 0.017(17) | 0.015(12) |
| C5   | 0.023(9) | 0.049(11) | 0.10(2) | -0.021(11) | 0.036(12) | 0.015(7) |
| N00D | 0.045(9) | 0.097(15) | 0.056(15) | 0.003(13) | -0.016(10) | -0.008(9) |
| C6   | 0.096(16) | 0.16(2) | 0.006(11) | -0.011(16) | 0.018(12) | 0.016(18) |
| N5   | 0.029(8) | 0.110(17) | 0.15(3) | 0.024(17) | 0.067(14) | 0.006(10) |
| C00F | 0.050(15) | 0.12(2) | 0.16(4) | 0.08(2) | 0.013(19) | -0.021(14) |
| C1   | 0.17(3) | 0.20(4) | 0.11(3) | -0.12(3) | 0.09(3) | -0.09(3) |
| C2   | 0.078(17) | 0.14(3) | 0.15(4) | -0.06(3) | 0.04(2) | -0.060(19) |
| C00L | 0.23(5) | 0.09(2) | 0.11(4) | 0.03(2) | 0.02(4) | 0.07(3) |
| C00V | 0.062(18) | 0.13(3) | 0.53(11) | -0.02(4) | 0.14(4) | -0.06(2) |
| C00I | 0.14(3) | 0.16(3) | 0.08(3) | -0.02(2) | 0.00(2) | -0.06(2) |
| O002 | 0.079(14) | 0.13(2) | 0.10(2) | -0.003(18) | 0.032(17) | -0.018(15) |
| N2   | 0.023(8) | 0.065(12) | 0.20(3) | -0.016(16) | 0.027(14) | 0.011(7) |
| O3   | 0.16(3) | 0.073(18) | 0.11(3) | -0.027(16) | 0.06(2) | -0.031(16) |
| C3   | 0.16(3) | 0.002(10) | 0.51(10) | 0.02(3) | 0.14(5) | 0.023(14) |
| O4   | 0.051(11) | 0.59(9) | 0.040(15) | -0.09(4) | 0.035(12) | -0.07(3) |
| O004 | 0.065(10) | 0.101(14) | 0.046(14) | -0.005(13) | 0.004(10) | -0.008(10) |
| N4   | 0.077(14) | 0.067(13) | 0.062(19) | 0.046(13) | 0.003(13) | -0.012(10) |
| C4   | 0.022(9) | 0.12(2) | 0.035(16) | 0.023(17) | -0.005(10) | 0.008(10) |
| O5   | 0.035(9) | 0.094(15) | 0.12(3) | -0.031(15) | 0.012(13) | -0.016(9) |
| C8   | 0.045(12) | 0.11(2) | 0.10(2) | -0.02(2) | -0.048(15) | 0.025(12) |
| C9   | 0.014(8) | 0.085(15) | 0.18(4) | 0.026(19) | 0.049(15) | 0.016(9) |
| C10  | 0.06(2) | 0.35(8) | 0.42(11) | 0.08(7) | 0.05(4) | 0.13(4) |
| C11  | 0.048(16) | 0.031(12) | 0.82(16) | -0.01(3) | 0.04(4) | 0.038(11) |
| C009 | 0.067(15) | 0.13(3) | 0.09(3) | 0.01(2) | -0.022(17) | -0.057(16) |
| C00A | 0.051(12) | 0.040(10) | 0.10(2) | 0.020(14) | 0.004(14) | -0.017(9) |
| C00H | 0.061(14) | 0.12(2) | 0.051(18) | 0.040(17) | 0.010(13) | 0.014(15) |
| C00N | -0.001(7) | 0.107(16) | 0.072(18) | -0.055(15) | 0.003(9) | -0.007(8) |
| C00O | 0.024(10) | 0.116(19) | 0.048(18) | 0.034(15) | -0.005(10) | 0.016(11) |
| C00R | 0.18(3) | 0.074(19) | 0.21(5) | 0.00(2) | 0.08(4) | -0.06(2) |
| C00S | 0.005(9) | 0.14(2) | 0.045(15) | 0.026(17) | -0.013(10) | 0.010(10) |
| O00T | 0.037(9) | 0.138(19) | 0.087(18) | 0.048(17) | -0.007(11) | -0.006(11) |
| C00U | 0.25(5) | 0.074(19) | 0.31(8) | -0.05(3) | 0.22(6) | -0.01(3) |
| C00W | 0.18(5) | 0.24(7) | 0.32(10) | 0.07(6) | -0.17(6) | 0.11(5) |
| C00X | 0.13(3) | 0.11(2) | 0.05(2) | -0.036(18) | 0.00(2) | -0.03(2) |
| C00Y | 0.10(2) | 0.18(4) | 0.17(5) | -0.02(3) | 0.02(3) | 0.05(2) |
| P1   | 0.038(5) | 0.081(8) | 0.079(12) | 0.030(8) | 0.004(7) | 0.005(6) |
| O1   | 0.037(7) | 0.091(12) | 0.061(14) | 0.019(11) | 0.022(9) | 0.016(9) |
| N1   | 0.061(11) | 0.068(13) | 0.083(19) | 0.063(13) | 0.034(13) | 0.018(10) |
| O2   | 0.072(12) | 0.062(12) | 0.26(5) | 0.017(18) | 0.07(2) | -0.028(11) |
| C010 | 0.20(4) | 0.045(15) | 0.60(13) | 0.00(3) | 0.30(7) | 0.050(18) |
### Table S11. Bond lengths of remdesivir form IV

| Atom   | Atom   | Length (Å) | Atom   | Atom   | Length (Å) |
|--------|--------|------------|--------|--------|------------|
| N3     | C00S   | 1.21(2)    | O004   | P1     | 1.45(3)    |
| N3     | C00C   | 1.39       | N4     | C4     | 1.11(3)    |
| N3     | N5     | 1.39       | C4     | C00N   | 1.59(4)    |
| N3     | C9     | 2.01(4)    | O5     | C00O   | 1.38(3)    |
| C00C   | C8     | 1.31(3)    | O5     | C00N   | 1.44(4)    |
| C00C   | C5     | 1.39       | C8     | C9     | 1.38(4)    |
| C5     | N2     | 1.23(3)    | C9     | C00S   | 1.44(4)    |
| C5     | N00D   | 1.39       | C10    | C00U   | 1.52(3)    |
| N00D   | C6     | 1.39       | C11    | C00O   | 1.32(5)    |
| C6     | N5     | 1.39       | C11    | P1     | 2.55(4)    |
| C00F   | O1     | 1.37(3)    | C009   | N1     | 1.28(4)    |
| C00F   | C1     | 1.39       | C009   | C00Y   | 1.63(5)    |
| C00F   | C00I   | 1.39       | C009   | P1     | 2.69(5)    |
| C00F   | P1     | 2.53(2)    | C00A   | O00T   | 1.27(3)    |
| C1     | C2     | 1.39       | C00A   | C00H   | 1.55(4)    |
| C1     | P1     | 3.07(3)    | C00A   | C00O   | 1.59(4)    |
| C2     | C00L   | 1.39       | C00H   | O2     | 1.32(4)    |
| C00L   | C00V   | 1.39       | C00H   | C00N   | 1.46(3)    |
| C00V   | C00I   | 1.39       | C00N   | C00S   | 1.37(3)    |
| O002   | C11    | 1.30(4)    | C00R   | C010   | 1.53(3)    |
| O002   | P1     | 1.52(4)    | C00U   | C00W   | 1.53(3)    |
| O3     | C3     | 1.16(6)    | C00W   | C00X   | 1.52(3)    |
| C3     | O4     | 1.36(5)    | C00W   | C010   | 1.53(4)    |
| C3     | C009   | 1.60(7)    | P1     | O1     | 1.60(2)    |
| O4     | C00X   | 1.53(6)    | P1     | N1     | 1.60(3)    |

### Table S12. Bond angles of remdesivir form IV

| Atom   | Atom   | Atom   | Angle (°) | Atom   | Atom   | Atom   | Angle (°) |
|--------|--------|--------|-----------|--------|--------|--------|-----------|
| C00S   | N3     | C00C   | 120.1(17) | O00T   | C00A   | C00O   | 112(3)    |
| C00S   | N3     | N5     | 119.9(16) | C00H   | C00A   | C00O   | 106.3(19) |
| C00C   | N3     | N5     | 120      | O2     | C00H   | C00N   | 122(2)    |
| C00S   | N3     | C9     | 45.1(14)  | O2     | C00H   | C00A   | 113(2)    |
| C00C   | N3     | C9     | 75.2(13)  | C00N   | C00H   | C00A   | 97(2)     |
| N5     | N3     | C9     | 164.5(9)  | C00S   | C00N   | O5     | 112(2)    |
| C8     | C00C   | C5     | 135.4(18) | C00S   | C00N   | C00H   | 112(2)    |
| C8     | C00C   | N3     | 104(2)    | O5     | C00N   | C00H   | 104.7(19) |
| C5     | C00C   | N3     | 120      | C00S   | C00N   | C4     | 112(2)    |
| N2     | C5     | C00C   | 120.8(16) | O5     | C00N   | C4     | 105(2)    |
| N2     | C5     | N00D   | 119.1(17) | C00H   | C00N   | C4     | 110.2(17) |
| C00C   | C5     | N00D   | 120      | C11    | C00O   | O5     | 116(3)    |
| C5     | N00D   | C6     | 120      | C11    | C00O   | C00A   | 118(3)    |
| N5     | C6     | N00D   | 120      | O5     | C00O   | C00A   | 96(2)     |
| C6     | N5     | N3     | 120      | N3     | C00S   | C00N   | 130(2)    |
| O1     | C00F   | C1     | 129.7(15) | N3     | C00S   | C9     | 98(2)     |
| Atom | Atom | Atom | Atom | Angle (°) | Atom | Atom | Atom | Atom | Angle (°) |
|------|------|------|------|----------|------|------|------|------|----------|
| C00S | N3   | C00C | C8   | 3(2)     | O00T | C00A | C00H | C00N | 86(3)    |
| N5   | N3   | C00C | C8   | -176.5(17)| C00O | C00A | C00H | C00N | -41(2)   |
| C9   | N3   | C00C | C8   | 0.1(5)   | C00O | O5   | C00N | C00S | -143(3)  |
| C00S | N3   | C00C | C5   | 179.3(19)| C00O | O5   | C00N | C00H | -21(3)   |
| N5   | N3   | C00C | C5   | 0       | C00O | O5   | C00N | C4   | 95(3)    |
| C9   | N3   | C00C | C5   | 176.6(11)| O2   | C00H | C00N | C00S | -79(3)   |
| C8   | C00C | C5   | N2   | -3(3)   | C00A | C00H | C00N | C00S | 158(2)   |
| N3   | C00C | C5   | N2   | -178.1(15)| O2 | C00H | C00N | O5   | 159(3)   |
| C8   | C00C | C5   | N00D | 175(2)  | C00A | C00H | C00N | O5   | 36(2)    |
| N3   | C00C | C5   | N00D | 0       | O2   | C00H | C00N | C4   | 47(3)    |

Table S13. Torsion angles of remdesivir form IV
|   | C5 | C00D | C6  | C00A | C00H | C00N | C4  |   |
|---|----|------|-----|------|------|------|-----|---|
| N2 |   | C5   | N00D| C6   | 178.2(15) | C00A | C00H | C00N | C4  | -76(2) |
| C00C| C5 | N00D | C6  | 0    | C00A | C00H | C00N | C4  | -76(2) |
| C5 | N00D| C6   | N5  | 0    | P1   | C00A | C00H | C00N | C4  | -76(2) |
| N00D| C6 | N5   | N3  | 0    | C00A | C00H | C00N | C4  | -76(2) |
| C00S| N3 | N5   | C6  | -179.3(19) | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C00C| N3 | N5   | C6  | 0    | C00A | C00H | C00N | C4  | -76(2) |
| C9 | N3 | N5   | C6  | -168(4) | C00A | C00H | C00N | C4  | -76(2) |
| O1 | C00F| C1   | C2  | 174(2) | C00A | C00H | C00N | C4  | -76(2) |
| C00I| C00F| C1   | C2  | 0    | C00A | C00H | C00N | C4  | -76(2) |
| P1 | C00F| C1   | C2  | -169.9(12) | C00A | C00H | C00N | C4  | -76(2) |
| O1 | C00F| C1   | P1  | -16.4(14) | C00A | C00H | C00N | C4  | -76(2) |
| C00I| C00F| C1   | P1  | 169.9(12) | C00A | C00H | C00N | C4  | -76(2) |
| C00F| C1 | C2   | C00L| 0    | N5   | N3   | C00S | C00N | C4  | -76(2) |
| P1 | C1 | C2   | C00L| -55(5) | N5   | N3   | C00S | C00N | C4  | -76(2) |
| C1 | C2 | C00L| C00V| 0    | C00A | C00H | C00N | C4  | -76(2) |
| C2 | C00L| C00V| C00I| 0    | N5   | N3   | C00S | C00N | C4  | -76(2) |
| C00L| C00V| C00I| C00F| 0    | N5   | N3   | C00S | C00N | C4  | -76(2) |
| C00F| C00F| C00I| C00V| 174.8(18) | C00A | C00H | C00N | C4  | -76(2) |
| C1 | C00F| C00I| C00V| 0    | C00A | C00H | C00N | C4  | -76(2) |
| P1 | C00F| C00I| C00V| 164.4(19) | C00A | C00H | C00N | C4  | -76(2) |
| O3 | C3 | C00X| C00X| -2(7) | C00A | C00H | C00N | C4  | -76(2) |
| C009| C3 | C00X| C00X| 160(3) | C00A | C00H | C00N | C4  | -76(2) |
| C5 | C00C| C8   | C9  | -175.8(14) | C00A | C00H | C00N | C4  | -76(2) |
| N3 | C00C| C8   | C9  | 0(2)  | C00A | C00H | C00N | C4  | -76(2) |
| C00C| C8 | C9   | C00S| -2(3) | C00A | C00H | C00N | C4  | -76(2) |
| C00C| C8 | C9   | N3  | 0.1(15) | C00A | C00H | C00N | C4  | -76(2) |
| P1 | O002| C11  | C00O| 174(3) | C00A | C00H | C00N | C4  | -76(2) |
| O3 | C3 | C009| N1   | -133(5) | C00A | C00H | C00N | C4  | -76(2) |
| O4 | C3 | C009| N1   | 64(4)  | C00A | C00H | C00N | C4  | -76(2) |
| O3 | C3 | C009| C00Y| -9(6)  | C3    | C00A | C00H | C00N | C4  | -76(2) |
| O4 | C3 | C009| C00Y| -171(3) | C11  | O002 | P1   | C00A | C00H | C00N | C4  | -76(2) |
| O3 | C3 | C009| P1   | -115(5) | C11  | O002 | P1   | C00A | C00H | C00N | C4  | -76(2) |
| O4 | C3 | C009| P1   | 83(3)  | C11  | O002 | P1   | C00A | C00H | C00N | C4  | -76(2) |
| O00T| C00A| C00H| O2   | -43(3) | C11  | O002 | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C00O| C00A| C00H| O2   | -170(2) | C11  | O002 | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C00Y| C009| N1   | P1   | -67(5) | C11  | O002 | P1   | C00A | C00H | C00N | C4  | -76(2) |
| O004| P1 | N1   | C009| -35(3) | C1    | C00A | C00H | C00N | C4  | -76(2) |
| O002| P1 | N1   | C009| -166(3) | C00A | C00H | C00N | C4  | -76(2) |
| O1 | P1 | N1   | C009| 90(3)  | O004  | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C00F| P1 | N1   | C009| 94(3)  | O002  | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C11| P1  | N1   | C009| -142(3) | N1    | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C1 | P1  | N1   | C009| 78(4)  | C11   | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C00X| C00W| C010| C00R| 70(4)  | C009  | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C00U| C00W| C010| C00R| -47(4) | C1    | P1   | C00A | C00H | C00N | C4  | -76(2) |
| C3 | C009| N1   | P1   | 52(4)  | C00A | C00H | C00N | C4  | -76(2) |
II. Crystal Structure Prediction Workflow

The crystal structure prediction (CSP) workflow (Figure S3) starts with a 2D molecular structure of an active pharmaceutical ingredient (API), with the goal of predicting and ranking its crystal polymorphs. An extensive conformation analysis of the molecular structure is performed by scanning over the flexible dihedral angles of the molecule with 20° steps. For each flexible torsion scan, we obtain a potential energy profile from high precision DFT calculations with the B3LYP/6-31G* basis set\textsuperscript{12,13} using the Gaussian16 software package.\textsuperscript{14}

![Figure S3. XtalPi’s crystal structure prediction workflow](image)

Generally, one or two stable conformations are used to start the screening process by varying the values of flexible torsion angles of the starting conformation to create potential energy profiles, which are used to initially parameterize the force field (stage 1) that characterizes the potential energy of the molecule. An indication of the quality of the force field can be obtained by measuring the correlation between the force field itself and the QM values. The energy correlation between the force field and QM calculations for the crystal structures in the low-energy region is a good indication to show the accuracy of
the specific force field. The RMSE between E-MM and E-QM of structures in the low energy region for a tailor-made force field is usually 3-6 kJ/mol even for some complex systems, e.g., salts and zwitterions. Moreover, in the MD sampling process, we monitor the RMSD, volume, potential energy and other properties of the structure to determine whether the structure in the MD sampling is unstable and if the sampling is uniform and converged. By using a tailor-made force field generated for each input candidate structure and a global search algorithm, such as heuristic particle swarm optimization\textsuperscript{15} or stochastic Monte Carlo simulations\textsuperscript{16} a large set of crystal structures are generated (stage 2) to allow for an accelerated convergence toward the globally enumerated crystal landscape.\textsuperscript{17} Then, the generated crystal structures were filtered using an energy filter that adjusted based on the energy correlation of the tailor-made force field, and the high energy crystal structures are filtered out. The remaining crystal structures are clustered using ultrafast AI-based technologies both to remove redundant configurations and to build statistics on our sampling of crystal space groups and molecular conformers. An in-house modified version of the k-medoids cluster method with a distance threshold instead of the k number was adopted for the clustering of the crystal structures (stage 3).

The representative structures from each cluster were selected for further ranking using higher accuracy methods (stage 4). The crystal structures are first ranked by a semi-empirical tight binding density functional algorithm, followed by high-precision density functional level of theory with dispersion energy correction (DFT-D).\textsuperscript{18} The final energy ranking of the polymorphs (stage 5) was performed at high precision DFT-D, optPBE-vdW,\textsuperscript{19} level of theory as implemented in the VASP software package.\textsuperscript{20} The accuracy of the reported calculated relative lattice energies is about 1.5 kJ/mol.\textsuperscript{21-23}

Since CSP is an iterative workflow, the force field gets more and more accurate after every iteration to reproduce a number of geometrical and energetic properties of crystal structures. The clustering statistics are used to improve crystal structure sampling until convergence is reached. To measure convergence, we perform our searches using two independent CSP algorithms simultaneously. Each of these CSP searches uses a different algorithm to generate the crystal structures and at every iteration of the CSP
The search converges when each algorithm finds the same low-energy polymorphs over a number of successive iterations. We also consider the rate at which we find new low energy structures. For example, if the total number of crystal structures in the low energy region remains the same then we conclude that the search has converged. This procedure ensures that an exhaustive search of the low energy regions of the crystal energy landscape is performed. Finally, after convergence, free energy calculations on a subset of low-energy polymorphs are performed to determine the solid form free energy over a range of temperatures (stage 6).21

Since the majority of the workload of CSP is a series of independent calculations, it can be made efficient if you have access to a large number of computers. Consequently, we have developed a cloud-based algorithm that can take advantage of a large number of resources on the amazon cloud server. The algorithm works by spawning a series of secure virtual private clouds (VPC) at computing centers across the globe and then it optimizes the workload of each VPC according to the number of compute nodes available, the cost, and the wait time. The VPCs are synced and within 1 hour a million-core high performance cluster (HPC) can be assembled, which, allows us to screen billions of crystal structures on the fly and perform a truly exhaustive search. Screening this many crystal structures has a number of indirect benefits, for example, the force field can be trained and parameterized on a large number of crystal structures, which, in turn, improves both the quality of our search and the energy ranking. Another benefit is the data can be used to train a number of AI and machine learning algorithms to make the CSP search and energy ranking more efficient. Also, by using the combination of a large computing cluster, special enhancing sampling strategies and in-house development structure generation algorithm, we can overcome a lot of complexity issues and push the boundaries of what can be done with CSP. In particular, we can do CSP for more flexible molecules with higher values of Z prime (Z´), and complicated multicomponent crystals.

From CSP calculations, we obtain a crystal energy landscape from which one can learn a lot about the crystal structures of a given API. In particular, ranking of the experimental
structures in the landscape is crucial for de-risking the solid form selection of the API. To identify the experimental structures in the polymorph landscape, we calculate XRDs of the predicted structures and compare them with the experimental XRDs for validation. If there is an experimental single crystal structure available, then we also overlay the predicted crystal structure with the experimental structure and measure their similarity with RMSD_{15} calculations. To discriminate between similar and distinct crystal structures, RMSD_{15} calculations were used where if the RMSD is less than a certain threshold (<1.0 Å) and there are 15 common molecules for 15 packing shell size, then the two structures are considered to be the same. If we find that the experimental structure corresponds to the lowest energy polymorph in the landscape (ranked X1) and there is a large energy gap between X1 and all other polymorphs, then we can confidently conclude that the experimental structure is the most stable polymorph for the given API. However, in many cases, there are multiple low energy polymorphs that are close in energy to X1. According to available data in the literature, 98.5% of predicted polymorph lattice energy differences are less than 10 kJ/mol, and within 5 kJ/mol between observed polymorphs. Absolute lattice energy represents the energy needed to form a crystal from infinitely separated formula units such as ions, molecules, or atoms. The 0 kJ/mol is not an absolute energy, but it is defined as the energy of the lowest energy crystal structure in the landscape.

In situations where there are a number of low energy polymorphs that are close in energy to X1 it is difficult to definitively determine which is the most stable polymorph using the information from the CSP landscape. At the same time, it is incomplete information because these energy landscapes correspond to the relative stability of polymorphs at 0 K. For many systems the relative stability of polymorphs changes with temperature. Therefore, for CSP to be truly predictive relative free energies of crystal polymorphs must be determined at ambient temperature.

There are a number of methods that have been used for this purpose such as the harmonic approximation or quasi-harmonic approximation (QHA). However, while these methods can work well at low temperatures, they do not always work well at higher temperatures, especially in case of flexible molecules like remdesivir. Therefore, we use
a fundamentally different approach for our free energy calculations known as the pseudo-supercritical path (PSCP) method.\textsuperscript{22,23} It takes all the conformation effects into account based on a given force field, which is more accurate, but not faster than QHA. This method uses a thermodynamic cycle in conjunction with MD to compute the free energy for each polymorph at a specific reference temperature. In this method, each polymorph is converted through a series of transformations into a reference state, called an Einstein atomic crystal\textsuperscript{25-27} and the details are described elsewhere.\textsuperscript{21} In short, each polymorph can be converted to this reference state and by summing up all of the free energy contributions along this pathway, we can calculate the free energy of the polymorphs from the CSP landscape. We perform this procedure for a given set of low energy polymorphs (for example, X1, X2, X3, X4, X5) to calculate the relative free energies of the polymorphs at a specific reference temperature. We also perform a series of additional MD simulations across a range of temperatures to calculate the relative free energies across this temperature range. This gives us information about how stability of the polymorphs changes as a function of temperature, which is key to predict the most stable polymorph at any given temperature. This protocol has been recently shown to rapidly identify the most stable forms over a wide range of temperatures and can also reveal enantiotropic transitions.\textsuperscript{21}

III. References

1. Brak, K.; Carra, E. A.; Heumann, L. V.; Larson, N. Crystalline forms of (S)-2-ethylbutyl 2-(((S)-((2R,3S,4R,5R)-5- (4-aminopyrrolo[2,1-f] [1,2,4]triazin-7-yl)-5-cyano-3,4-dihydroxytetrahydrofuran-2-yl)methoxy)(phenoxy) phosphoryl)amino)propanoate. US 2018/0346504 A1, December 6, 2018.
2. Nannenga, B. L.; Gonen, T., MicroED: a versatile cryoEM method for structure determination. \textit{Emerg Top Life Sci} 2018, 2 (1), 1-8.
3. Nannenga, B. L.; Shi, D.; Leslie, A. G. W.; Gonen, T., High-resolution structure determination by continuous-rotation data collection in MicroED. \textit{Nature Methods} 2014, 11 (9), 927-930.
4. Huang, Z.; Grape, E. S.; Li, J.; Inge, A. K.; Zou, X., 3D electron diffraction as an important technique for structure elucidation of metal-organic frameworks and covalent organic frameworks. *Coordination Chemistry Reviews* **2021**, *427*, 213583.
5. Hattne, J.; Reyes, F. E.; Nannenga, B. L.; Shi, D.; de la Cruz, M. J.; Leslie, A. G. W.; Gonen, T., MicroED data collection and processing. *Acta Crystallographica Section A* **2015**, *71*(4), 353-360.
6. Roslova, M.; Smeets, S.; Wang, B.; Thersleff, T.; Xu, H.; Zou, X., InsteaDMatic: towards cross-platform automated continuous rotation electron diffraction. *Journal of Applied Crystallography* **2020**, *53*(5), 1217-1224.
7. Kabsch, W., XDS. *Acta Crystallographica Section D* **2010**, *66*(2), 125-132.
8. Cichocka, M. O.; Angstrom, J.; Wang, B.; Zou, X.; Smeets, S., High-throughput continuous rotation electron diffraction data acquisition via software automation. *Journal of Applied Crystallography* **2018**, *51*(6), 1652-1661.
9. Sheldrick, G., SHELXT - Integrated space-group and crystal-structure determination. *Acta Crystallographica Section A* **2015**, *71*(1), 3-8.
10. Sheldrick, G., Crystal structure refinement with SHELXL. *Acta Crystallographica Section C* **2015**, *71*(1), 3-8.
11. Hubschle, C. B.; Sheldrick, G. M.; Dittrich, B., ShelXle: a Qt graphical user interface for SHELXL. *Journal of Applied Crystallography* **2011**, *44*(6), 1281-1284.
12. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *Journal of Physical Chemistry* **1994**, *98*, 11623–11627.
13. Hehre, W. J.; Ditchfeld, R.; Pople, J. A., Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *Journal of Chemical Physics* **1972**, *56*, 2257.
14. Gaussian 16, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang,
W.; Hada, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.

15. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In Proceedings of ICNN’95 - International Conference on Neural Networks; IEEE, 2002; Vol. 4, pp 1942–1948 vol.4.

16. Gamerman, D.; Lopes, H. F. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition, 2nd ed.; Chapman & Hall/CRC: Philadelphia, PA, 2006.

17. Zhang, P.; Wood, G. P. F.; Ma, J.; Yang, M.; Liu, Y.; Sun, G.; Jiang, Y. A.; Hancock, B. C.; Wen, S. Harnessing Cloud Architecture for Crystal Structure Prediction Calculations. Cryst. Growth Des. 2018, 18 (11), 6891–6900.

18. Grimme, S. Accurate Description of van Der Waals Complexes by Density Functional Theory Including Empirical Corrections. J. Comp. Chem. 2004, 25 (12), 1463–1473.

19. Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab-initio total-energy calculations using a plane-wave basis. Phys. Rev. B: Condens. Matter Mater. Phys. 1996, 54 (16), 11169.

20. Hafner, J. Ab-initio Simulations of Materials Using VASP: Density-functional Theory and Beyond. J. Comput. Chem. 2008, 29 (13), 2044–2078.

21. Yang, M.; Dybeck, E.; Sun, G.; Peng, C.; Samas, B.; Burger, V. M.; Zeng, Q.; Jin, Y.; Bellucci, M. A.; Liu, Y.; Ma, J.; Jiang, A.; Hancock, B. C.; Wen, S.; Wood. G. P. F. Prediction of the Relative Free Energies of Drug Polymorphs above Zero Kelvin. Cryst. Growth Des. 2020, 20 (8), 5211–5224.

22. Dybeck, E. C.; Scheiber, N. P.; Shirts, M. R. Effects of a More Accurate Polarizable Hamiltonian on Polymorph Free Energies Computed Efficiently by Reweighting Point-Charge Potentials. J. Chem. Theory Comput., 2016, 12 (8), 3491–3505.
23. Dybeck, E. C.; Abraham, N. S.; Schieber, N. P.; Shirts, M. R. Capturing Entropic Contributions to Temperature-Mediated Polymorphic Transformations through Molecular Modeling. *Cryst. Growth Des.* **2017**, *17*, 1775–1787.

24. Hoja, J.; Ko, H.-Y.; Neumann, M. A.; Car, R.; DiStasio, R. A.; Tkatchenko, A. Reliable and Practical Computational Description of Molecular Crystal Polymorphs. *Sci. Adv.* **2019**, *5*, No. eaau3338.

25. Frenkel, D.; Ladd, A. J. New Monte Carlo Method to Compute the Free Energy of Arbitrary Solids. Application to the Fcc and Hcp Phases of Hard Spheres. *J. Chem. Phys.* **1984**, *81*, 3188–3193.

26. Frenkel, D.; Smit, B. Understanding Molecular Simulation: From Algorithms to Applications; Elsevier (formerly published by Academic Press): **2002**; Vol. 1, pp 1–638.

27. Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268.