JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Positional Variance in NMR Crystallography

Albert Hofstetter and Lyndon Emsley*

Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Supporting Information

ABSTRACT: We propose a method to quantify positional uncertainties in crystal structures determined by chemical-shift-based NMR crystallography. The method combines molecular dynamics simulations and density functional theory calculations with experimental and computational chemical shift uncertainties. In this manner we find the average positional accuracy as well as the isotropic and anisotropic positional accuracy associated with each atom in a crystal structure determined by NMR crystallography. The approach is demonstrated on the crystal structures of cocaine, flutamide, flufenamic acid, the K salt of penicillin G, and form 4 of the drug 4-[4-(2-adamantylcarbamoyl)-5-*tert*-butylpyrazol-1-yl]benzoic acid (AZD8329). We find that, for the crystal structure of cocaine, the uncertainty corresponds to a positional RMSD of 0.17 Å. This is a factor of 2.5 less than for single-crystal X-ray-diffraction-based structure determination.

Structure−activity relationships play a central role in chemistry, and the possibility to determine three-dimensional molecular structures has been key to molecular science over the past 50 years. Many molecules or materials have been characterized by single-crystal X-ray diffraction (XRD). However, when the sample is a powder, structural characterization remains a challenge. This is important, for example, in pharmaceutical applications, where the determination of the structures of drug polymorphs is an intrinsic part of the development process. Solid-state nuclear magnetic resonance (NMR) can overcome this bottleneck through its sensitivity to the local atomic environment regardless of the degree of long-range order, and much progress has been made with methods that, for example, directly measure dipolar couplings.[1,2] In this area, the combination of solid-state NMR and computational methods has also made tremendous progress over the past decade.[3−7] The scope of this combined approach has rapidly increased, and today there are many examples of structure validation by chemical shift measurements combined with density functional theory (DFT) calculations.[8−18] Recently there have been examples of *de novo* structure determination combining NMR shifts, DFT shift calculations, and crystal structure prediction.[14−17] However, in contrast to XRD methods, there exists no protocol to quantify the positional errors on individual atoms for structures determined by chemical-shift-based NMR crystallography.

Here, we introduce a method, based on molecular dynamics (MD) simulation, DFT, and machine learning methods, to estimate the correlation between the root-mean-square deviation (RMSD) of the experimental and calculated shifts and the variances of atomic positions of individual atoms in structures determined by NMR crystallography, thereby making them directly comparable to structures determined by other methods. The approach is demonstrated on multiple crystal structures recently characterized by NMR crystallography.[13,15,16]

First we generate an ensemble of slightly perturbed crystal structures with MD simulations at finite temperatures. By "slightly perturbed" we refer to structures that remain within the same local minima and do not undergo any significant conformational shifts. The temperature ranges used and the associated computational costs are detailed in the Supporting Information (SI). Predicted ^{1}H and ^{13}C chemical shifts are then calculated for the members of the ensemble using plane wave DFT and the GIPAW[5] method. Given the estimated errors in the measured and predicted chemical shifts, we then correlate this directly with the atomic positions that are compatible with the measured chemical shifts to within the error, yielding a distribution of positions for each atom. The positional distributions are then converted into anisotropic displacement parameters (ADPs),[19] which can be represented by ellipsoids on the determined structure. The results of this process are given in Figure 1 for cocaine, flutamide, flufenamic acid, the drug 4-[4-(2-adamantylcarbamoyl)-5-*tert*-butylpyrazol-1-yl]-benzoic acid (AZD8329),[15] and the K salt of penicillin G.

To obtain the correlation between the chemical shift uncertainty and the ADPs, first the chemical shift RMSD between each structure in an ensemble and a reference structure from the ensemble is calculated. Next the positional deviations between each structure and the reference structure are calculated. For each individual atom the principal axis system (PAS) of the ensemble of positional deviations is determined using principal component analysis (PCA) as detailed in the SI. This results in a scatter plot of the type shown in Figure 2a.

A continuous correlation function is obtained by maximizing the log-likelihood between the correlation points and a Gaussian distribution:

$$G(\langle r_{i,l}\rangle, \langle\delta\rangle) = \frac{1}{\sqrt{2\pi\Sigma_{i,l}^2\langle\delta\rangle^2}} \exp\left\{-\frac{(\langle r_{i,l}\rangle - \mu_{i,l}\langle\delta\rangle)^2}{2\Sigma_{i,l}^2\langle\delta\rangle^2}\right\} \quad (1)$$

where $\langle r\rangle$ denotes the positional deviation, $\langle\delta\rangle$ the chemical shift RMSD, Σ the scaling of the variance, and μ the scaling of the mean. The indices l and i denote the atom and the principal

Received: December 9, 2016
Published: February 1, 2017

DOI: 10.1021/jacs.6b12705
J. Am. Chem. Soc. 2017, 139, 2573−2576

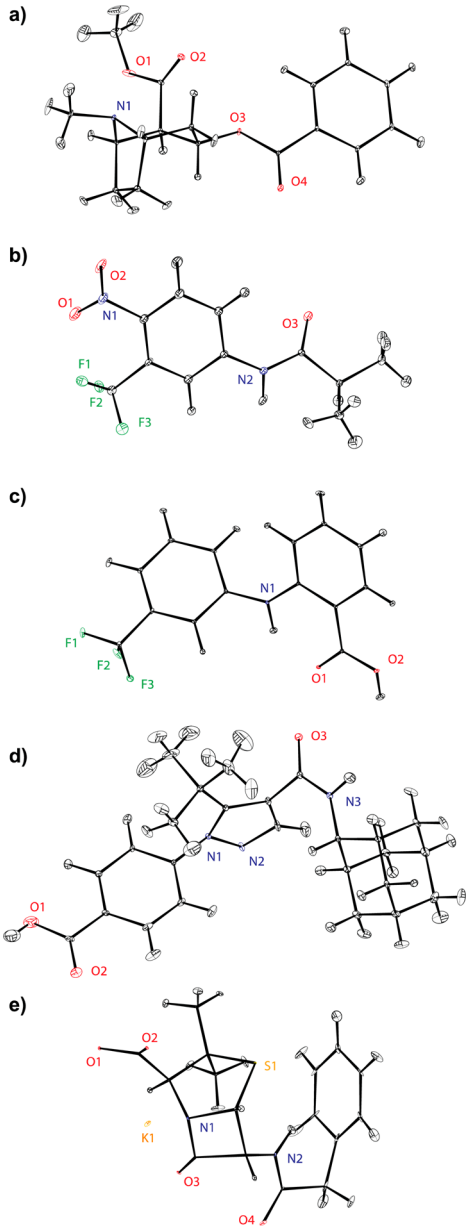

Figure 1. ORTEP plots drawn at the 90% probability level for the NMR-determined crystal structures of (a) cocaine, (b) flutamide, (c) flufenamic acid, (d) AZD8329, and (e) the K salt of penicillin G. The ellipsoids correspond to positions within a ^{1}H chemical shift RMSD of 0.49 ppm.

axis, respectively. The fit parameters are Σ and μ. The detailed procedure is given in the SI. The result of this procedure for the O1 atom of cocaine is shown in Figure 2. Please note that the uncertainty prediction method described here is not limited to the use of a Gaussian distribution function (details in the SI).

The principal values of the ADPs in the PAS are calculated as the mean-square displacements, which for Gaussian distributions is given as the variance, as a function of the chemical shift RMSD,

$$U_{ii,l}^{PAS} = \Sigma_{i,l}^{2}\langle\delta\rangle^{2} \tag{2}$$

The amplitudes of the second-rank tensors describing the ellipsoids at a given probability W are calculated in the PAS, where they are diagonal, as,

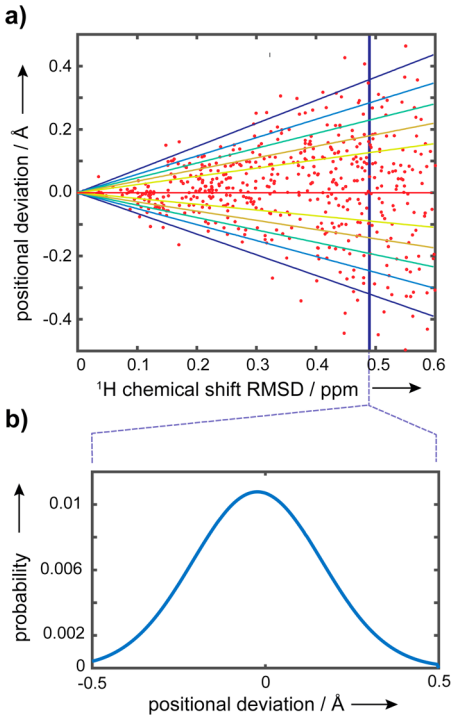

Figure 2. (a) Contour plot of the Gaussian fit of the correlation between the positional displacement (Å) and the ^{1}H chemical shift RMSD (ppm) along one principal axis of the anisotropic displacement tensor for the O1 atom for the cocaine crystal structure. (b) Probability distribution of the positional displacement (Å) for a ^{1}H chemical shift RMSD of 0.49 ppm.

$$T_{ii,l}^{PAS} = p_{i,l}(W, \langle\delta\rangle)^{2} \tag{3}$$

where $p_{i,l}(W,\langle\delta\rangle)$ denotes the Wth percentile of the fitted Gaussian for a chemical shift RMSD $\langle\delta\rangle$. These are the quantities that are usually plotted in so-called ORTEP plots as anisotropic displacement ellipsoids, and this is what is shown in Figure 1.

Note that, for simplicity, or for cases with insignificant anisotropy in the displacements, the second-rank ADP can be replaced by the equivalent isotropic displacement parameter:[20,21]

$$U_{eq}^{l} = \frac{1}{3}(U_{11,l}^{PAS} + U_{22,l}^{PAS} + U_{33,l}^{PAS}) \tag{4}$$

Note also that, from the equivalent isotropic displacement parameters, we can derive a global measurement of the positional uncertainty (U_{eq} and T_{eq}) for the whole structure, which is given as the average of the equivalent isotropic displacement parameters over all N atoms in the structure,

$$U_{eq} = \frac{1}{N}\Sigma_{l=1}^{N}U_{eq}^{l} \tag{5}$$

The radii of the isotropic spheres and of the average isotropic spheres at a certain probability W are calculated analogously to the axes of the anisotropic displacement ellipsoids (eq 3; the formula is detailed in the SI). The isotropic spheres and the average isotropic spheres are shown for cocaine in Figure 3b,c. The average positional RMSD $\langle r_{av}\rangle$ for a given chemical shift RMSD $\langle\delta\rangle$ is then calculated as

$$\langle r_{av}\rangle = \sqrt{3U_{eq}} \tag{6}$$

DOI: 10.1021/jacs.6b12705
J. Am. Chem. Soc. 2017, 139, 2573−2576

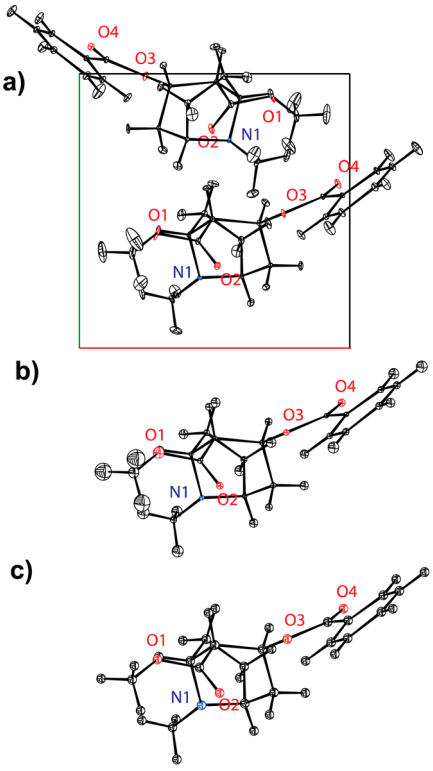

Figure 3. ORTEP plot of the cocaine structure drawn at the 90% probability level. (a) Anisotropic ellipsoids, corresponding to a ^{1}H chemical shift RMSD of 0.49 ppm. (b) Equivalent isotropic spheres, corresponding to a ^{1}H chemical shift RMSD of 0.49 ppm. (c) Average isotropic spheres for a chemical shift RMSD $\langle\delta\rangle$ = 0.49 ppm, corresponding to an average positional RMSD $\langle r_{av}\rangle$ = 0.169 Å.

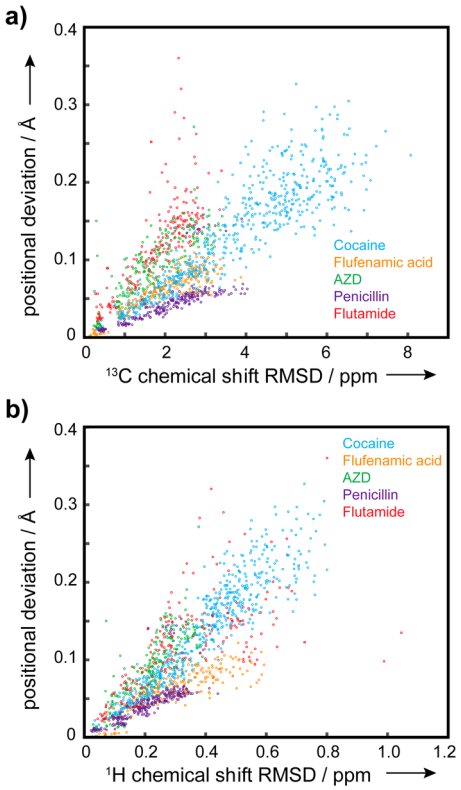

Figure 4. (a) Correlation between positional RMSD (Å) and ^{13}C chemical shift RMSD (ppm) for five ensembles of perturbed crystal structures generated by MD. (b) Correlation between positional RMSD (Å) and ^{1}H chemical shift RMSD (ppm) for five ensembles of slightly perturbed crystal structures.

The factor $\sqrt{3}$ results from the fact that the isotropic displacement parameter is given as in eq 4, while the RMSD is calculated as $\langle r\rangle = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$.

As indicated in Figure 4, we find that the positional RMSD $\langle r_{av}\rangle$ shows an approximately linear correlation with the average chemical shift RMSD $\langle\delta\rangle$ for each of the five structures, but that the slope of the correlation is different for each structure. For example, for a given chemical shift RMSD the structure determined for penicillin has more than a factor two less uncertainty than that for flutamide. This is not surprising. The sources of this variation depend on the rigidity of the molecule, its hybridization, and the electron density gradients in the crystal structure. A detailed investigation of these factors will be the subject of future studies. Also, the positional uncertainty depends on how internal dynamics (such as methyl rotation) is accounted for (detailed in the SI), as one of the main contributors to the positional RMSD. The positional uncertainty presented here should therefore be viewed as an upper limit.

From eqs 2 and 4–6, the correlation between the chemical shift RMSD $\langle\delta\rangle$ and the average positional RMSD $\langle r_{av}\rangle$ is

$$\langle r_{av}\rangle = \sqrt{\frac{1}{N}\Sigma_{i,l}\Sigma_{i,l}^2}\,\langle\delta\rangle = \bar{\Sigma}\langle\delta\rangle \tag{7}$$

For the crystal structure of cocaine we find a direct correlation, $\bar{\Sigma}$ = 0.345. Given an average chemical shift RMSD $\langle\delta\rangle$ = 0.49 ppm, which is the current estimated upper limit for the accuracy in ^{1}H chemical-shift-based crystallography methods,[14] this leads to an average positional RMSD $\langle r_{av}\rangle \approx$ 0.169 Å, corresponding to an average equivalent displacement parameter U_{eq} = 0.0095 Å^2. Compared to other structure determination methods, for example, XRD, which yielded an average positional RMSD $\langle r_{av}\rangle$ = 0.458 Å for the crystal structure of cocaine,[22] we find an increase in positional accuracy by a factor 2.5. It is interesting to note that, for XRD, the positional uncertainty mainly results from the thermal motion of the atoms and is a direct result of the decrease in scattering amplitude due to vibrations. In contrast, in NMR spectroscopy, thermal motion and fast lattice vibrations lead to motional narrowing of the measured signal and, if anything, are likely to increase accuracy; thus, we see that the different techniques naturally have different limits on the positional accuracy.

We remark that the methods used to create the ensemble of structures and to calculate the chemical shifts are important in determining the positional errors. We have evaluated the use of different force fields in the MD simulation, a fixed versus a variable unit cell as discussed in the SI, and we find that they have no significant effect on the uncertainty quantification.

Comparable calculations were also done for an ensemble of perturbed cocaine crystal structures generated by random uncorrelated displacement of the atoms (i.e., this corresponds to systematic uncorrelated bond stretching). For this ensemble the correlation predicts much larger deviations in chemical shift for a given average displacement (see Figure S13 in SI), which would lead to much higher apparent positional accuracy. This is expected, due to the generation of physically improbable structures resulting in an unreasonable electronic density. A possibility to overcome this would be to weight the random structures with a Boltzmann factor based on their calculated

DOI: 10.1021/jacs.6b12705
J. Am. Chem. Soc. 2017, 139, 2573–2576

energy, but this should provide no direct advantage compared to the MD method. The MD method, on the other hand, searches the conformational space more efficiently and implicitly weights the generated structures with a Boltzmann factor. The random displacement method thus severely underestimates the positional errors. The MD ensemble allows for a significantly larger uncertainty in position than the random displacement method for a given chemical shift RMSD, and it is thus a better representation of the uncertainty in positions in the experimentally determined structures. We are currently exploring other methods to generate physically reasonable ensembles, for example, through the exploitation of vibrational modes of the crystal structures.

Finally, it is possible that the choice of the DFT functional might have an influence on the calculated uncertainties. The PBE[23] functional used here is the current standard for the computation of chemical shifts in crystals,[24] and we remark that the systematic error in chemical shift calculations has shown to be similar for different functionals.[25] This systematic error likely results from the difficulty for DFT to correctly describe polar groups and long-range dispersion forces, e.g., H-bonds. However, here we would not be sensitive to this systematic error, but only to any systematic variation within the error, which is likely to be small.

In conclusion, we have introduced a method to quantify positional uncertainties in crystal structures derived from NMR chemical shifts. The structures quantified here were determined by chemical-shift-based NMR crystallography, but in principle structures determined by other methods, e.g., XRD, could be refined with this method. An ensemble of structures around the experimentally determined structure is generated *in silico*, and the predicted chemical shift deviations for this ensemble are compared to the positional deviations. In this way we determine the average positional error of the experimentally determined structure for each atom in the crystal structure. We find that the average positional uncertainty in the five structures studied here yields an RMSD of 0.17 Å, or an average value of the equivalent displacement parameter of 0.0095 Å^2. We find that chemical-shift-based NMR crystallography methods provide a gain in positional accuracy of around a factor 2 compared to XRD structure determination. This is mainly because thermal vibrations are not limiting for chemical-shift-based NMR methods.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b12705.

Calculation details and ORTEP plots for all molecules (PDF)

Structure files in CIF format, containing the displacement parameters corresponding to the ^{1}H and ^{13}C chemical shift RMSD (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*lyndon.emsley@epfl.ch

ORCID

Lyndon Emsley: 0000-0003-1360-2572

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful for financial support from Swiss National Science Foundation Grant No. 200021_160112.

■ REFERENCES

(1) Partridge, B. E.; Leowanawat, P.; Aqad, E.; Imam, M. R.; Sun, H. J.; Peterca, M.; Heiney, P. A.; Graf, R.; Spiess, H. W.; Zeng, X.; Ungar, G.; Percec, V. *J. Am. Chem. Soc.* **2015**, *137* (15), 5210.

(2) Dudenko, D.; Kiersnowski, A.; Shu, J.; Pisula, W.; Sebastiani, D.; Spiess, H. W.; Hansen, M. R. *Angew. Chem., Int. Ed.* **2012**, *51* (44), 11068.

(3) Facelli, J. C.; Grant, D. M.; Michl, J. *Acc. Chem. Res.* **1987**, *20* (4), 152.

(4) de Dios, A.; Pearson, J.; Oldfield, E. *Science* **1993**, *260*, 1491.

(5) Yates, J. R.; Pickard, C. J.; Mauri, F. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *76*, 024401.

(6) Pickard, C. J.; Mauri, F. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2001**, *63* (24), 245101.

(7) Bonhomme, C.; Gervais, C.; Babonneau, F.; Coelho, C.; Pourpoint, F.; Azaïs, T.; Ashbrook, S. E.; Griffin, J. M.; Yates, J. R.; Mauri, F.; Pickard, C. J. *Chem. Rev.* **2012**, *112* (11), 5733.

(8) Salager, E.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. *Phys. Chem. Chem. Phys.* **2009**, *11* (15), 2610.

(9) Ochsenfeld, C.; Brown, S. P.; Schnell, I.; Gauss, J.; Spiess, H. W. *J. Am. Chem. Soc.* **2001**, *123* (11), 2597.

(10) Goward, G. R.; Sebastiani, D.; Schnell, I.; Spiess, H. W.; Kim, H. D.; Ishida, H. *J. Am. Chem. Soc.* **2003**, *125* (19), 5792.

(11) Harris, R. K. *Analyst* **2006**, *131* (3), 351.

(12) Harris, R. K.; Joyce, S. A.; Pickard, C. J.; Cadars, S.; Emsley, L. *Phys. Chem. Chem. Phys.* **2006**, *8* (1), 137.

(13) Mifsud, N.; Elena, B.; Pickard, C. J.; Lesage, A.; Emsley, L. *Phys. Chem. Chem. Phys.* **2006**, *8* (29), 3418.

(14) Salager, E.; Day, G. M.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. *J. Am. Chem. Soc.* **2010**, *132* (8), 2564.

(15) Baias, M.; Dumez, J. N.; Svensson, P. H.; Schantz, S.; Day, G. M.; Emsley, L. *J. Am. Chem. Soc.* **2013**, *135* (46), 17501.

(16) Baias, M.; Widdifield, C. M.; Dumez, J.-N.; Thompson, H. P. G.; Cooper, T. G.; Salager, E.; Bassil, S.; Stein, R. S.; Lesage, A.; Day, G. M.; Emsley, L. *Phys. Chem. Chem. Phys.* **2013**, *15* (21), 8069.

(17) Baias, M.; Lesage, A.; Aguado, S.; Canivet, J.; Moizan-basle, V.; Audebrand, N.; Farrusseng, D.; Emsley, L. *Angew. Chem., Int. Ed.* **2015**, *54* (20), 5971.

(18) Pinon, A. C.; Rossini, A. J.; Widdifield, C. M.; Gajan, D.; Emsley, L. *Mol. Pharmaceutics* **2015**, *12* (11), 4146.

(19) Trueblood, K. N.; Bürgi, H. B.; Burzlaff, H.; Dunitz, J. D.; Gramaccioli, C. M.; Schulz, H. H.; Shmueli, U.; Abrahams, S. C. *Acta Crystallogr., Sect. A* **1996**, *52* (5), 770.

(20) Hamilton, W. C. *Acta Crystallogr.* **1959**, *12* (8), 609.

(21) Willis, B. T. M.; Howard, J. A. K. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1975**, *31* (4), 514.

(22) Hrynchuk, R. J.; Barton, R. J.; Robertson, B. E. *Can. J. Chem.* **1983**, *61* (3), 481.

(23) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865.

(24) Harris, R. K.; Hodgkinson, P.; Pickard, C. J.; Yates, J. R.; Zorin, V. *Magn. Reson. Chem.* **2007**, *45*, S174.

(25) Hill, D. E.; Vasdev, N.; Holland, J. P. *Comput. Theor. Chem.* **2015**, *1051*, 161.

DOI: 10.1021/jacs.6b12705
J. Am. Chem. Soc. 2017, 139, 2573−2576